

## Article

# Forecasting Crude Oil Prices with Major S&P 500 Stock Prices: Deep Learning, Gaussian Process, and Vine Copula

Jong-Min Kim <sup>1,\*</sup>, Hope H. Han <sup>2</sup> and Sangjin Kim <sup>3</sup><sup>1</sup> Statistics Discipline, University of Minnesota at Morris, Morris, MN 56267, USA<sup>2</sup> School of Business Administration, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Korea; hopehan@unist.ac.kr<sup>3</sup> Department of Management and Information Systems, Dong-A University, Busan 49236, Korea; skim10@dau.ac.kr

\* Correspondence: jongmink@morris.umn.edu; Tel.: +1-320-589-6341

**Abstract:** This paper introduces methodologies in forecasting oil prices (Brent and WTI) with multivariate time series of major S&P 500 stock prices using Gaussian process modeling, deep learning, and vine copula regression. We also apply Bayesian variable selection and nonlinear principal component analysis (NLPCA) for data dimension reduction. With a reduced number of important covariates, we also forecast oil prices (Brent and WTI) with multivariate time series of major S&P 500 stock prices using Gaussian process modeling, deep learning, and vine copula regression. To apply real data to the proposed methods, we select monthly log returns of 2 oil prices and 74 large-cap, major S&P 500 stock prices across the period of February 2001–October 2019. We conclude that vine copula regression with NLPCA is superior overall to other proposed methods in terms of the measures of prediction errors.

**Keywords:** oil prices; S&P 500; multivariate time series; Gaussian process model; vine copula; Bayesian variable selection; functional principal component analysis; nonlinear principal component analysis

**MSC:** 62-07

**Citation:** Kim, J.-M.; Han, H.H.; Kim, S. Forecasting Crude Oil Prices with Major S&P 500 Stock Prices: Deep Learning, Gaussian Process, and Vine Copula. *Axioms* **2022**, *11*, 375. <https://doi.org/10.3390/axioms11080375>

Academic Editors: Palle E.T. Jorgensen and Adrian Petrusel

Received: 2 June 2022  
Accepted: 27 July 2022  
Published: 29 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Global monetary policies developed in response to the COVID-19 crisis and the 2022 Russia–Ukraine war have resulted in high crude oil prices, causing economic inflation and a bear market rally in 2022.

The relationship between the price of crude oil and the stock market has been a main research topic in economics and finance. The relationship between oil prices and the stock market, specifically in terms of forecasting stock returns and analyzing volatilities using oil prices, has been studied in [1,2]. Since oil-sensitive stocks have strong forecasting power on crude oil prices [3], we want to examine how oil prices can be affected by the most influential stock prices in this study. Some interesting statistical methods to predict oil prices have been proposed, such as investor attention (constructed by the Google search volume index [4]), the LASSO machine learning method [5], and copula dependence structures between oil prices, exchange rates, and interest rates [6]. In this study, we want to employ deep learning, Gaussian process modeling, and vine copula regression methods to predict the oil prices with the most influential stock prices.

Deep learning has been utilized to forecast stock prices in [7], and the comparison of stock-price prediction models using pre-trained neural networks has been performed in [8]. The LSTM and ARIMAX algorithms were employed in [9] to analyze the impact of sentiment analysis in stock market prediction. Gaussian process regression methods and extensions for stock market prediction have been studied in [10]. Stock prediction

using Gaussian process regression has been studied in [11]. The amount of training data required for deep learning [11] and the choosing of hyperparameters can make the method difficult to use. The response of a Gaussian process model needs to be normally distributed if the hyperparameters are fixed. So, we propose an alternative forecasting model (vine copula regression) to predict oil-price returns using US stock returns. The copula method does not need assumptions, such as normality, linearity, and independence of errors. Additionally, vine copula can explain a flexible multivariate dependence structure. To show our proposed method's superiority over the deep learning and Gaussian process models, we apply accuracy measures to deep learning, Gaussian process models, and vine copula regression models. This study also examines whether there are firms that are highly influential to oil prices. To do this, we use Bayesian variable selection and nonlinear principal component analysis for forecasting crude oil prices (Brent and WTI).

This paper is organized as follows: Section 2 presents the data description and summary, Section 3 gives an overview of the statistical models for forecasting, the illustrated comparison study of the proposed methods is presented in terms of the measures of errors in Section 4, and the discussion is presented in Section 5.

## 2. Summary Statistics

The sample contains the monthly log returns of Brent Crude, Western Texas Intermediate (WTI), and 74 major S&P 500 stock prices from February 2003 to October 2019. (Variable names for our sample data can be found in Appendix A.) The reason for choosing monthly log returns over daily log returns in this paper is our attempt to eliminate the noise from small economic factors, such as political news. The 74 stocks were selected based on the size of their market capitalization. Figure 1 plots the log returns for the 2 oil prices and 74 stock prices, along with a functional mean equation line of sample log returns. We observed a co-movement among oil and stock returns in Figure 1. This is already a well-known phenomenon. Our sample was collected from the Yahoo Finance website (<https://finance.yahoo.com/>) (accessed on 7 November 2020). We converted prices to log returns throughout our analyses.

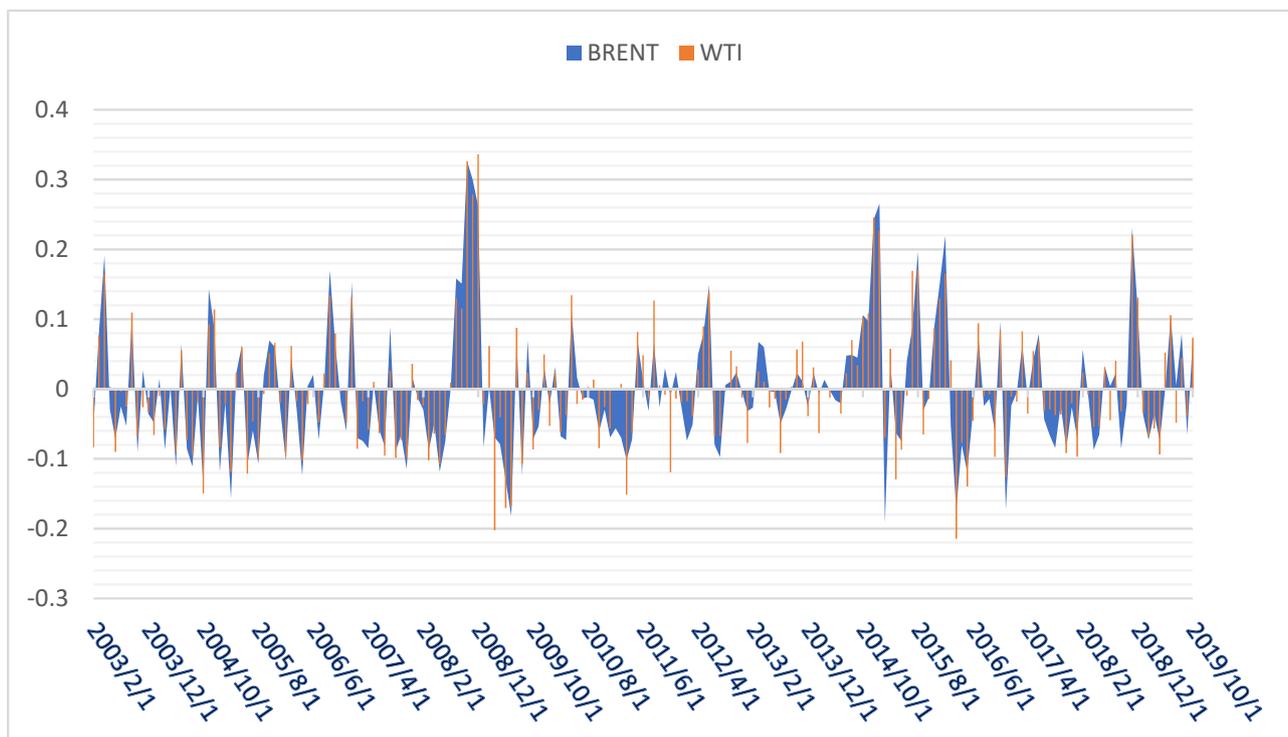


Figure 1. Brent and WTI Monthly Log Returns.

Table 1 displays the summary statistics for the oil and stock price monthly log returns. We observed that Brent and WTI have similar distributional properties: the log returns of Brent and WTI prices are positively skewed with fat tails, while the average log returns of the Brent and WTI prices are close to zero.

**Table 1.** Summary statistics.

	Mean	Median	Minimum	Maximum	St.D	Skewness	Kurtosis
BRENT	−0.003	−0.015	−0.192	0.326	0.089	0.968	4.476
WTI	−0.002	−0.014	−0.215	0.336	0.088	0.897	4.728

We could expect that there would be a more prominent relationship among crude oil and major S&P 500 stock prices. We used February 1 as the beginning of the log return monthly data because of the 3 January 2003 base log return difference.

Let  $S_t$  be a price time series at time  $t$ . For a log return series,  $r_t = \log\left(\frac{S_t}{S_{t-1}}\right)$ . Each of the datasets was given a new variable known as “log returns”. We summarized the descriptive statistics for the BRENT and WTI log return data, such as mean, skewness, and kurtosis, as well as 5 summary statistics in Table 1.

In Table 1, it can be observed that the standard deviations of the log returns of BRENT and WTI are about the same. The values of skewness in the log returns of BRENT and WTI in this period are positive, such that oil prices will increase in the future. In addition, the values of kurtosis in the log returns of BRENT and WTI are greater than 3, meaning that they have heavy tails compared to a normal distribution. Figure 1 shows the time trend of the Brent and WTI monthly log returns over the given period. If we look at the 2008 economic crisis period, as shown in Figure 1, the log returns of Brent and WTI were very high in the first half of 2008, and then they suddenly dropped to very low values in the second half of 2008. Because investors feared the tightening of monetary policy, a slowing economy, and an intensifying trade war between the U.S. and China in December 2018, the S&P 500 fell more than 9%, causing the log returns of Brent and WTI to be low.

### 3. Statistical Methods

#### 3.1. Gaussian Process (GP) Model

The first forecasting method we used was the Gaussian process (GP) model, which leads to a supervised learning method aimed at solving regression and probabilistic classification problems. The GP models have been popular for use in studying uncertainty quantification (UQ), which is the science of the quantitative characterization and reduction of uncertainties in both computational and real-world applications. Eight software packages for fitting Gaussian processes to various functions have been compared based on the root-mean-square error of predictions over the input space [12].

A Gaussian process (GP) is a random process where any point  $x \in R^d$  is assigned a random variable  $f(x)$  and the joint distribution of a finite number of these variables  $p(f(x_1), \dots, f(x_N))$  is Gaussian:  $p(f|X) = N(f|\mu, K)$ , where  $f = (f(x_1), \dots, f(x_N))$ ,  $\mu = (m(x_1), \dots, m(x_N))$  and  $K_{ij} = \kappa(x_i, x_j)$ .  $m$  is the mean function, and it is common to use  $m(x) = 0$  as GPs are flexible enough to model the mean.  $\kappa$  is a positive definite kernel function or covariance function. We recommend reading [13] to understand the GP model. We constructed the GP model to forecast the next month’s oil price returns. We used the *GauPro* function from the ‘*GauPro*’ R package [14]. The GP is a stochastic process where every finite linear combination of random variables has a multivariate normal distribution [15]. The GP model accurately predicts and reports standard errors for predictions as well. There are two main parameters in the GP. The theta determines how strong the correlation is between points in each parameter. The nugget is a smoothing parameter that allows for noise and improves computational stability [12].

### 3.2. Copulas

The second forecasting method we used was the copula method, which does not require any assumptions, such as independence or normality. Additionally, by using the copula method, we could avoid multicollinearity and heteroscedasticity issues when we performed the regression analysis. This is the reason that the copula method has been popular in economics and finance. Modeling for the way that corporate bond yield spreads are affected by explanatory variables, such as equity volatility, interest rate volatility,  $r$ , slope, rating, liquidity, coupon rate, and maturity, was studied by [16]. The dependence at the mean of the joint distribution by using the Gaussian copula marginal regression method and the dependence structure at the tails by using various copula functions was also studied by [16]. Recently, the impacts of COVID-19 on the dependence structure of the stock market were considered by Gaussian copula regression modeling in [17], and size anomalies in U.S. bank stock returns were investigated by using panel copula in [18].

A  $d$ -dimensional copula  $C$  is a  $d$ -variate distribution function on the unit hypercube  $[0,1]^d$  with uniform marginal distribution functions. Sklar's theorem [19] provides a link between multivariate distributions and their associated copulas. It states that, for every multivariate random vector  $X = (X_1, \dots, X_d)' \sim F$  with marginal distribution functions  $F_1, \dots, F_d$ , there exists a copula  $C$  associated with  $X$ , such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

This decomposition of the multivariate distribution into its margins and its associated copula is unique when  $X$  is absolutely continuous. Its marginals are  $U_j = F_j(X_j)$ ,  $j = 1, \dots, d$ . The  $U_j$  are then uniformly distributed, and their joint distribution function is the copula  $C$  associated with  $X$ . The Gaussian copula,  $t$ -copula, and Archimedean copulas are popularly used in finance and economics. The Gaussian copula is constructed from a multivariate normal distribution by using the probability integral transform, and the  $t$ -copula is the copula that underlies the multivariate Student's  $t$ -distribution. The most-used Archimedean copulas are the Clayton copula, the Frank copula, and the Gumbel copula. The Clayton copula is used to look at the negative tail dependence, whereas the Gumbel copula is used for the positive tail dependence. The Frank copula is a symmetric Archimedean copula with no tail dependence. Refer to [20,21] for a detailed examination of copulas, including examples of parametric copulas, especially bivariate copulas. All numerical calculations in this study are performed using the programming language R and the package VineCopula [22]. Vine copula is a flexible multivariate dependence structure model. Researchers have used vine copula in economics and statistics. Using VineCopula for the Granger causality test in mean has been proposed by [23]. We also recommend reading that paper to understand vine copula.

### 3.3. Deep Learning

The third forecasting method we used was the deep learning and neural network model. Deep learning and neural network models for quality control research of count data were developed by [24].

Deep learning is the most promising research trajectory for big and complex data analysis. Based on the idea of imitating the interactions between brain neurons, researchers developed deep learning and neural network methods. For detailed explanations of deep learning and neural network, we recommend reading [24]. To perform deep learning data analysis, we used the R package deepnet [25], which trains single or multiple hidden layers in neural networks using the back propagation (BP) neural network algorithm, which is a multi-layer feedforward network trained according to the error back propagation algorithm. BP is one of the commonly used neural network models. The BP is used to regulate the weight value and threshold value of the network to achieve the minimum error sum of square. For training data, we used 2 hidden layers (30, 30) of neurons, and the activation function of the hidden unit is "sigm" for the logistic function. The function of the output

unit is “*sigm*”, and other conditions are set as default in the R package. We forecast the Brent and WTI variables with the given stock data by using the ‘*nn.predict*’ command in the R package “*deepnet*”.

### 3.4. Bayesian Variable Selection

The Bayesian variable selection method is an efficient statistical method for selecting the most influential explanatory variables. So, we used the Bayesian variable selection method of the objective Bayesian model proposed by [26]. The R package *BayesVarSel*, developed by [27], was used for data analysis in this paper. We used a Gibbs sampling scheme to determine the optimal model for the data set. We also set the possible prior distribution for regression parameters within each model as ‘Constant’ and set possible prior distribution over the model space as ‘gZellner.’ The number of iterations was 10,000 times after the 100th number of iterations at the beginning of the Markov Chain Monte Carlo (MCMC) that were dropped.

### 3.5. Nonlinear PCA

Kernel principal component analysis (PCA), a nonlinear PCA method, was developed by [28]. If we use a kernel as described in [28], we know that this procedure exactly corresponds to standard PCA in a high-dimensional feature space, so we do not need to perform expensive computations in that space. To extract five principal components in high-dimensional feature spaces using kernel PCA, we used the ‘*kernlab*’ R package [29], which provides the most popular kernel functions. We used the Gaussian Radial Basis kernel function with a hyperparameter:  $\sigma = 0.2$ , which is the inverse kernel width for the radial basis kernel function.

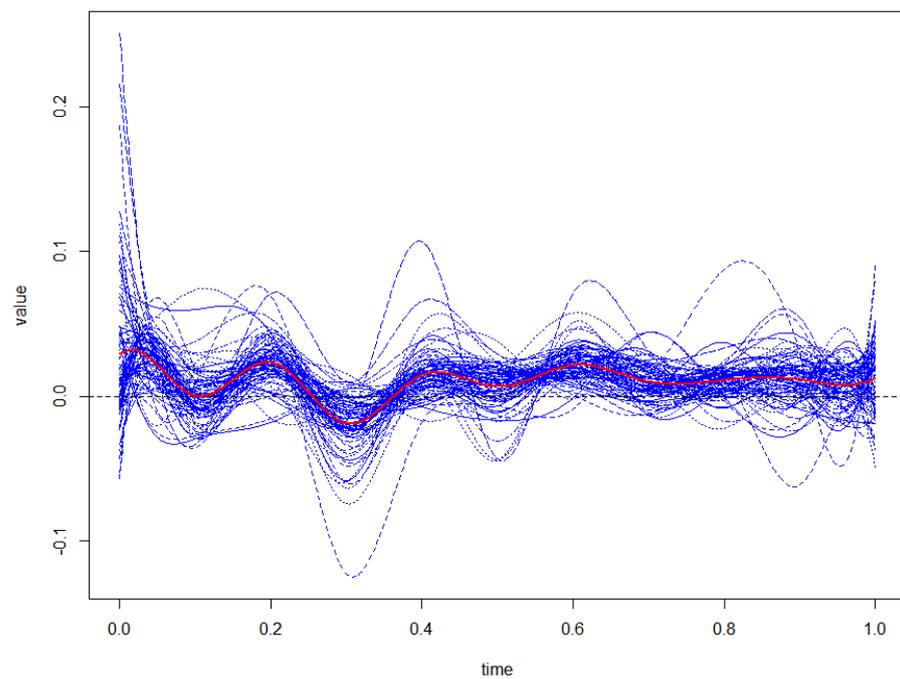
## 4. Data Analysis

First, we want to visualize the relationship between the crude oil and stock data. Functional data analysis is a popular big data dimension-reduction method for time-course data. Functional principal component analysis (FPCA) is an effective clustering visualization analysis for time-course data. It provides a much more informative way of examining the sample covariance structure than does PCA, and it is an effective statistical method for explaining the variance of components because of the use of nonlinear eigenfunctions. The PCA only shows the clustering pattern of the whole data at a certain year or certain time, but FPCA is the more suitable method for showing the clustering pattern of the time series oil data over the given period.

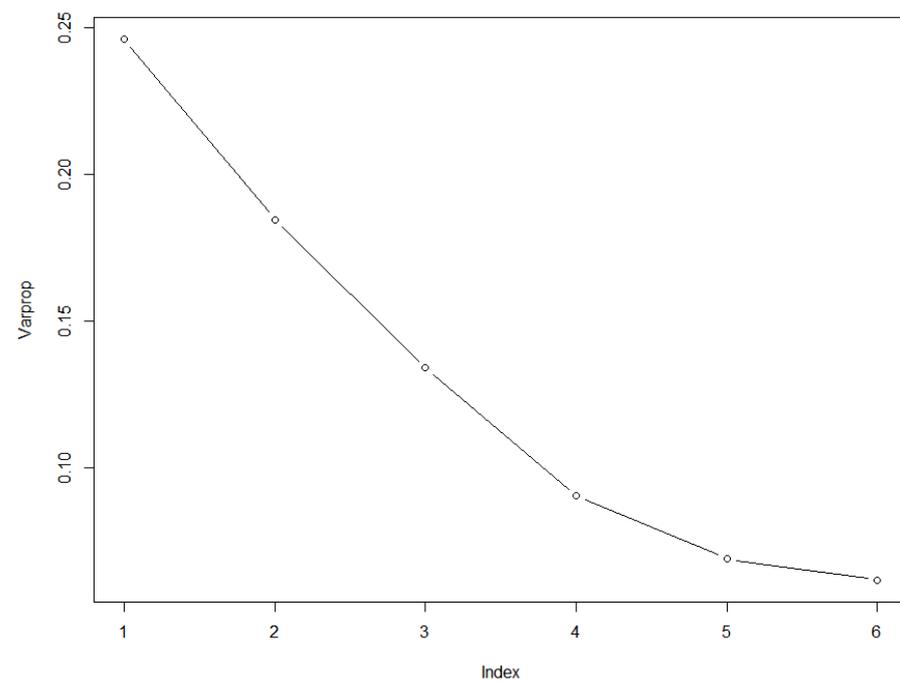
Figure 2 plots the log returns of the 2 oil prices and 74 stock prices, along with a functional mean equation line of sample log returns. We observed a co-movement among oil and stock returns, as shown in Figure 2. We then investigated the relationship in predicting oil price returns using stock returns.

We also performed a functional principal component analysis (FPCA) by using the FDA R package to determine factors (i.e., principal components) that explain the relationship between crude oil and stock prices. Figure 3 shows a variance proportion of total variations in individual stock and oil price returns as explained by each principal component. Each component explains the percentage contribution to the whole density variation. The first principal component accounts for 24.6%, the second component explains 18.5%, and the third component accounts for 13.4% of the whole variance proportion of the FPCA. Note that the first 3 principal components account for 56.5% of the whole variability.

Through visualizations, we illustrated the relationship among the 2 major crude oil prices and the major S&P 500 stock prices. From the 2D FPCA plot in Figure 4, we were able to classify the 2 crude oil and 74 major S&P 500 stock prices into 4 groups.



**Figure 2.** A total of 76 Monthly Log Returns (blue color) and the functional mean equation line (red color).



**Figure 3.** Variance proportions of FPCA. Note: Index: principal components (1st, 2nd, etc.); Varpro: variance proportions.

The 2D FPCA plot captures a limited view of the clusters among major stock price and oil price returns. For a detailed visualization of the relationship between the 2 crude oil and 74 stock prices, we have provided a 3D FPCA plot with the first 3 main harmonics (principal components) in Figure 5. From Figure 5, we can observe that most of the major stock returns are clustered together, implying a co-movement of those return series in our sample period.

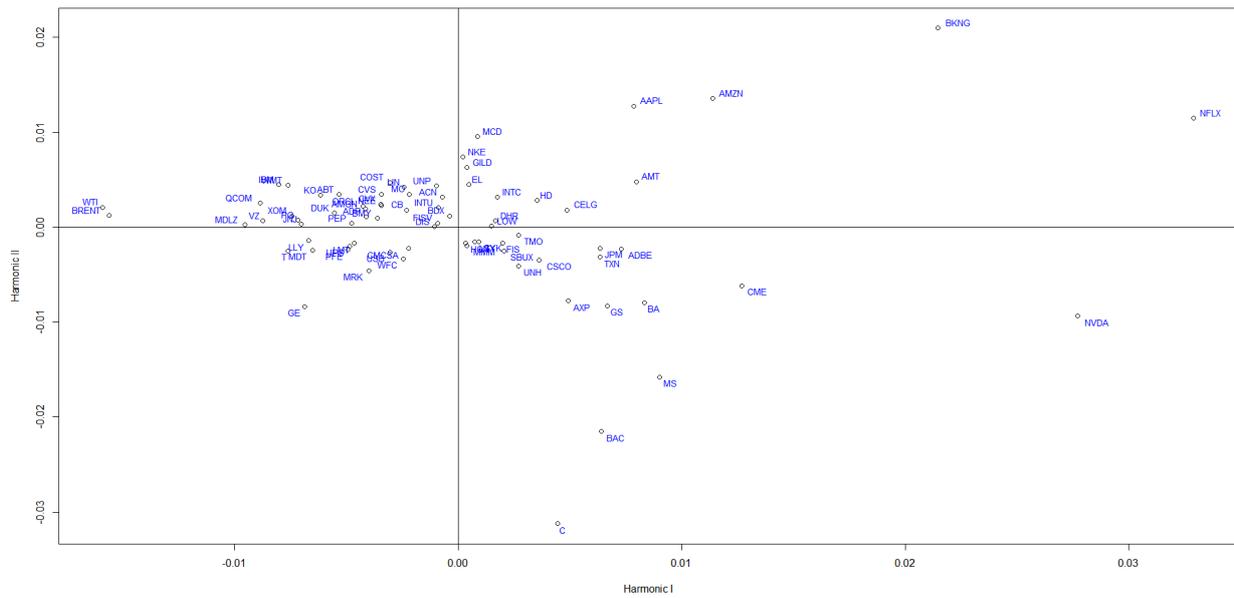


Figure 4. The 2D FPCA. Note: Harmonic I: 1st component; Harmonic II: 2nd component.

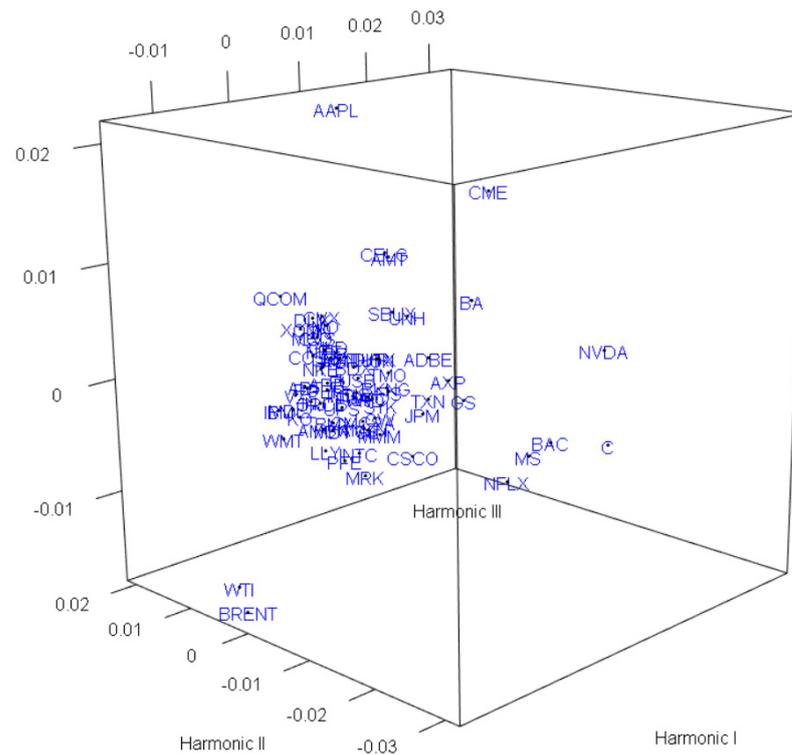


Figure 5. The 3D FPCA. Note: Harmonic I: 1st component; Harmonic II: 2nd component; Harmonic III: 3rd component.

The observations presented in Figures 4 and 5 motivated our research to investigate whether there is a more prominent relationship between crude oil prices and major S&P 500 stock prices.

We also selected the most influential stocks in relation to crude oil price returns using the Bayesian variable selection method of the objective Bayesian model proposed by [26]. Our empirical analysis restricted our attention to the stock prices of 74 firms which are considered major stocks in the S&P 500. We performed Bayesian variable selection for the BRENT and WTI oil price returns separately. Interestingly, five covariates were selected for

each of the Brent and WTI oil price returns based on inclusion probabilities. The inclusion probabilities included the highest posterior probability and the median probability in the Bayesian model selection

CB, HD, HON, LIN, and PG returns were critical factors in determining Brent oil price returns. Consistently, HD, HON, LIN, PG, and UNP returns were critical factors in determining WTI oil price returns. The inclusion probabilities of all of the covariates for the Brent and WTI oil price returns are displayed in Tables 2 and 3.

**Table 2.** Bayesian Variable Selection: Brent price as a dependent variable.

	Prob.	HPM	MPM
CB	0.54		*
HD	0.61	*	*
HON	0.65	*	*
LIN	0.73	*	*
PG	0.56	*	*

\* means “included”. This table presents the inclusion probabilities of five selected stocks using the Bayesian variable selection method, where the dependent variable is the Brent price. HPM stands for the highest posterior probability model, and MPM stands for the median probability model. Probabilities are estimated based on visited models.

**Table 3.** Bayesian Variable Selection: WTI price as a dependent variable.

	Prob.	HPM	MPM
HD	0.58		*
HON	0.57		*
LIN	0.61	*	*
PG	0.44	*	
UNP	0.68	*	*

\* means “included”. This table presents the inclusion probabilities of five selected stocks using the Bayesian variable selection method, where the dependent variable is the WTI price. HPM stands for the highest posterior probability model, and MPM stands for the median probability model. Probabilities are estimated based on visited models.

Following ref. [28] (See Section 3), we extract the first 5 principal components, using kernel PCA, to examine the power of forecasting oil price returns. We have used the first five principal components in this paper. The kernel function was used in training and predicting. This parameter can be set to any function of class kernel, which computes a dot product between two vector arguments. The corresponding component eigenvalues were 0.034684495, 0.009930120, 0.007794364, 0.006680240, and 0.005580485. Eigenvalues are used to find the proportion of the total variance explained by the components.

We also performed Gaussian process modeling to forecast oil prices in time  $t+1$  using S&P 500 stock prices in time  $t$ . To perform the analysis, we separated the training and forecast data from our sample, where February 2001–March 2017 was the training set for the stock data, and March 2001–April 2017 was the training set for the oil data. Test data for the stocks were collected from April 2017 to September 2019, and test data for the Brent and WTI prices were collected from May 2017 to October 2019. First, we considered all 74 major S&P 500 stock prices in our model and performed the forecasting for BRENT and WTI separately. Before we applied data to the GP model, we used the Bayesian variable selection method to select the five most influential stocks in relation to the Brent and WTI oil price returns. Then, we performed oil price forecasting using the Gaussian process model with the 5 covariates we found, as discussed in Section 3 and as seen in Tables 2 and 3. Tables 4 and 5 show the GP for Brent with Covariates (CB, HD, HON, LIN, PG) and the GP for Brent with Covariates (HD, HON, LIN, PG, UNP).

**Table 4.** Reduced Models.

	Theta
CB	6.1
HD	0.87
HON	2.07
LIN	9.58
PG	2.51
Nugget = 0.244	
RMSE = 0.088	
N = 170	

GP for Brent with Covariates (CB, HD, HON, LIN, PG).

**Table 5.** Reduced Models.

	Theta
HD	24.8
HON	8.08
LIN	25
PG	112
UNP	2.86
Nugget = 0.546	
RMSE = 0.0797	
N = 170	

GP for Brent with Covariates (HD, HON, LIN, PG, UNP).

We also compared the forecasting accuracy of the GP, deep learning, and vine copula methods by employing two measures for predictive accuracy.

We denoted the predicted values and actual values ( $y_t$  and  $\hat{y}_t$ ), and  $t = 1, 2, \dots, n$ . ( $n$  = the total number of test dataset).

Root-mean-square (prediction) error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \tag{1}$$

where we define the quadratic loss function to be  $LOSS1 = (y_t - \hat{y}_t)^2$ .

Mean absolute error deviation (MAD):

$$MAD = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \tag{2}$$

where we define the L1 loss function to be  $LOSS2 = |y_t - \hat{y}_t|$ .

The metric errors, such as the MAD and RMSE, were used to analyze the performance of the methods. The mean absolute error is not sensitive to outliers as they are weighted less than the other observations when comparing actual and predicted values. The root-mean-square error takes bias and variance into account, but it normalizes the units. Each method also produces plots based on the actual and predicted price returns for visualization purposes.

Table 6 shows the RMSE and MAD for forecasting BRENT and WTI log return prices. From Table 6, we can see that the RMSE and MAD decrease when we use the 5 selected covariates as compared to when we include all 74 covariates. We can also observe that forecasting Brent log return prices with major stock data using vine copula regression with NLPCA is superior to other methods, and that forecasting WTI log return prices with major stock data using Gaussian process and vine copula regression with NLPCA is superior to other methods in terms of the RMSE and MAD.

**Table 6.** RMSE and MAD for forecasting BRENT and WTI log return prices.

Method	Deep Learning			Gaussian Process			Vine Copula	
RMSE	ALL	BVS	NLPCA	ALL	BVS	NLPCA	BVS	NLPCA
Brent	0.087	0.090	0.088	0.100	0.078	0.077	0.079	<b>0.072</b>
WTI	0.085	0.084	0.084	0.086	0.080	0.073	0.077	<b>0.069</b>
MAD	ALL	BVS	NLPCA	ALL	BVS	NLPCA	BVS	NLPCA
Brent	0.075	0.078	0.076	0.080	0.060	<b>0.058</b>	0.060	0.060
WTI	0.072	0.071	0.072	0.073	0.064	0.059	0.061	<b>0.057</b>

Table 7 shows a 95% Confidence Interval for the Loss functions (LOSS1 and LOSS2) of the BRENT and WTI log return prices. From Table 7, we can observe that the width and center of the 95% Confidence Interval for the Loss functions (LOSS1 and LOSS2) of the BRENT and WTI log return prices using vine copula regression with NLPCA is smaller than that of other methods. We can confirm that forecasting BRENT and WTI log return prices with major stock data using vine copula regression with NLPCA is superior to other methods.

**Table 7.** The 95% Confidence Interval for the Loss functions (LOSS1 and LOSS2) of BRENT and WTI log return prices.

LOSS	Deep Learning		
LOSS1	ALL	BVS	NLPCA
Brent	(0.0048, 0.0106)	(0.0048, 0.0107)	(0.0049, 0.0107)
WTI	(0.0044, 0.0098)	(0.0042, 0.0096)	(0.0042, 0.0096)
LOSS2	ALL	BVS	NLPCA
Brent	(0.0588, 0.0926)	(0.0591, 0.0929)	(0.0591, 0.0930)
WTI	(0.0549, 0.0884)	(0.0540, 0.0870)	(0.0540, 0.0871)
LOSS	Gaussian Process		
LOSS1	ALL	BVS	NLPCA
Brent	(0.0048, 0.0151)	(0.0023, 0.0099)	(0.0025, 0.0101)
WTI	(0.0040, 0.0109)	(0.0022, 0.0095)	(0.0029, 0.0109)
LOSS2	ALL	BVS	NLPCA
Brent	(0.0566, 0.1025)	(0.0414, 0.0793)	(0.0483, 0.0826)
WTI	(0.0550, 0.0904)	(0.0462, 0.0794)	(0.0488, 0.0856)
LOSS	Vine Copula		
LOSS1	ALL	BVS	NLPCA
Brent	NA	(0.0018, 0.0108)	<b>(0.0022, 0.0082)</b>
WTI	NA	(0.0019, 0.0101)	<b>(0.0019, 0.0076)</b>
LOSS2	ALL	BVS	NLPCA
Brent	NA	(0.0402, 0.0796)	<b>(0.0448, 0.0753)</b>
WTI	NA	(0.0434, 0.0793)	<b>(0.0418, 0.0714)</b>

### 5. Discussion

This study is the first to investigate the predictability of oil prices using S&P 500 stock prices by using vine copula regression. We found that the BVS suggests that five stocks have the largest impact on each oil price. The selected companies are related to the energy/chemical industry or to the large retail industry. The important, selected companies for the Brent log returns are Chubb Limited (CB) (a Switzerland-based holding

insurance company), Home Depot (HD) (a home improvement retailer), Honeywell International Inc. (HON) (a software–industrial company that operates through four segments: Aerospace, Honeywell Building Technologies, Performance Materials and Technologies, and Safety and Productivity Solutions), Linde plc (LIN) (an industrial gas company), and the Procter & Gamble Company (PG) (focused on providing branded consumer packaged goods to consumers across the world). The important, selected companies for the WTI log returns are HD, HON, LIN, PG, and Union Pacific Corporation (UNP), which is a railroad operating company in the United States. We also found that forecasting Brent log return prices with major stock data using vine copula regression with NLPCA is superior to other methods in terms of the RMSE and MAD. We also found that forecasting WTI log return prices with major stock data using GP and vine copula regression with NLPCA is superior to other methods in terms of the RMSE and MAD. In conclusion, the stock prices of both the energy/chemical industry and the large retail industry are effective in forecasting oil prices. This study contributes to forecasting oil prices by using a vine copula regression with selected stock prices. In future research, we will consider the prices of commodities, such as gold, silver, copper, platinum, natural gas, wheat, corn, and soybeans in relation to important financial indices, including the Consumer Price Index, inflation, and cryptocurrency prices, using the GP, deep learning, and vine copula methods.

**Author Contributions:** Conceptualization, J.-M.K. and S. Kim; methodology, J.-M.K.; software, J.-M.K.; validation, J.-M.K. and H.H.H.; formal analysis, J.-M.K.; investigation, J.-M.K. and H.H.H.; resources, J.-M.K.; data curation, S.K.; writing—original draft preparation, J.-M.K., H.H.H. and S.K.; writing—review and editing, J.-M.K., H.H.H. and S.K.; visualization, J.-M.K.; supervision, J.-M.K.; project administration, S.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank the three anonymous, respected referees for their suggestions, which have improved the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Variable names.

Variable	Name
BRENT	Brent Crude
WTI	Western Texas Intermediate
AAPL	Apple, Inc.
ABT	Abbott Laboratories
ACN	Accenture Plc
ADBE	Adobe, Inc.
ADP	Automatic Data Processing, Inc.
AMGN	Amgen, Inc.
AMT	American Tower Corp.
AMZN	Amazon.com, Inc.
AXP	American Express Co.
BA	The Boeing Co.
BAC	Bank of America Corp.
BDX	Becton, Dickinson & Co.
BKNG	Booking Holdings, Inc.

Table A1. Cont.

Variable	Name
BMY	Bristol-Myers Squibb Co.
C	Citigroup, Inc.
CB	Chubb Ltd.
CELG	Celgene Corp.
CMCSA	Comcast Corp.
CME	CME Group, Inc.
COST	Costco Wholesale Corp.
CSCO	Cisco Systems, Inc.
CVS	CVS Health Corp.
CVX	Chevron Corp.
DHR	Danaher Corp.
DIS	The Walt Disney Co.
DUK	Duke Energy Corp.
EL	The Estée Lauder Companies, Inc.
FIS	Fidelity National Information Services, Inc.
FISV	Fiserv, Inc.
GE	General Electric Co.
GILD	Gilead Sciences, Inc.
GS	The Goldman Sachs Group, Inc.
HD	The Home Depot, Inc.
HON	Honeywell International, Inc.
IBM	International Business Machines Corp.
INTC	Intel Corp.
INTU	Intuit, Inc.
JNJ	Johnson & Johnson
JPM	JPMorgan Chase & Co.
KO	The Coca-Cola Co.
LIN	Linde Plc
LLY	Eli Lilly & Co.
LMT	Lockheed Martin Corp.
LOW	Lowe's Cos., Inc.
MCD	McDonald's Corp.
MDLZ	Mondelez International, Inc.
MDT	Medtronic Plc
MMM	3M Co.
MO	Altria Group, Inc.
MRK	Merck & Co., Inc.
MS	Morgan Stanley
NEE	NextEra Energy, Inc.
NFLX	Netflix, Inc.
NKE	NIKE, Inc.
NVDA	NVIDIA Corp.
ORCL	Oracle Corp.
PEP	PepsiCo, Inc.
PFE	Pfizer Inc.
PG	Procter & Gamble Co.
QCOM	QUALCOMM, Inc.
SBUX	Starbucks Corp.
SYK	Stryker Corp.
T	AT&T, Inc.
TMO	Thermo Fisher Scientific, Inc.
TXN	Texas Instruments Incorporated
UNH	UnitedHealth Group, Inc.
UNP	Union Pacific Corp.
UPS	United Parcel Service, Inc.
USB	U.S. Bancorp
UTX	United Technologies Corp.

Table A1. Cont.

Variable	Name
VZ	Verizon Communications, Inc.
WFC	Wells Fargo & Co.
WMT	Walmart, Inc.
XOM	Exxon Mobil Corp.

Table A2. Sector Information for Companies.

Grp.	Symbol	Security	GICS Sector	GICS Sub Industry	
1	AAPL	Apple Inc.	Information Technology	Technology Hardware, Storage and Peripherals	Brent
1	AMT	American Tower Corp.	Real Estate	Specialized REITs	
1	AMZN	Amazon.com Inc.	Consumer Discretionary	Internet and Direct Marketing Retail	
1	BKNG	Booking Holdings Inc	Consumer Discretionary	Internet and Direct Marketing Retail	Brent
1	CELG	Celgene	Health Care	Biotechnology	WTI removed: 21 November 2019
1	DHR	Danaher Corp.	Health Care	Health Care Equipment	
1	EL	Estee Lauder Cos.	Consumer Staples	Personal Products	WTI
1	GILD	Gilead Sciences	Health Care	Biotechnology	
1	HD	Home Depot	Consumer Discretionary	Home Improvement Retail	Brent, WTI
1	INTC	Intel Corp.	Information Technology	Semiconductors	
1	LOW	Lowe's Cos.	Consumer Discretionary	Home Improvement Retail	
1	MCD	McDonald's Corp.	Consumer Discretionary	Restaurants	
1	NFLX	Netflix Inc.	Communication Services	Movies and Entertainment	
1	NKE	Nike	Consumer Discretionary	Apparel, Accessories, and Luxury Goods	
2	ABT	Abbott Laboratories	Health Care	Health Care Equipment	
2	ACN	Accenture plc	Information Technology	IT Consulting and Other Services	
2	ADP	Automatic Data Processing	Information Technology	Internet Services and Infrastructure	
2	AMGN	Amgen Inc.	Health Care	Biotechnology	
2	BDX	Becton Dickinson	Health Care	Health Care Equipment	Brent, WTI
2	BMJ	Bristol-Myers Squibb	Health Care	Health Care Distributors	Brent
2	CB	Chubb Limited	Financials	Property and Casualty Insurance	
2	COST	Costco Wholesale Corp.	Consumer Staples	Hypermarkets and Super Centers	
2	CVS	CVS Health	Health Care	Health Care Services	
2	CVX	Chevron Corp.	Energy	Integrated Oil and Gas	Brent, WTI
2	DIS	The Walt Disney Company	Communication Services	Movies and Entertainment	
2	DUK	Duke Energy	Utilities	Electric Utilities	Brent
2	FISV	Fiserv Inc	Information Technology	Data Processing and Outsourced Services	
2	IBM	International Business Machines	Information Technology	IT Consulting and Other Services	WTI
2	INTU	Intuit Inc.	Information Technology	Application Software	
2	JNJ	Johnson & Johnson	Health Care	Pharmaceuticals	
2	KO	Coca-Cola Company	Consumer Staples	Soft Drinks	
2	LIN	Linde plc	Materials	Industrial Gases	Brent, WTI
2	MDLZ	Mondelez International	Consumer Staples	Packaged Foods and Meats	
2	MO	Altria Group Inc	Consumer Staples	Tobacco	Brent, WTI

Table A2. Cont.

Grp.	Symbol	Security	GICS Sector	GICS Sub Industry	
2	NEE	NextEra Energy	Utilities	Multi-Utilities	
2	ORCL	Oracle Corp.	Information Technology	Application Software	
2	PEP	PepsiCo Inc.	Consumer Staples	Soft Drinks	WTI
2	PG	Procter & Gamble	Consumer Staples	Personal Products	Brent, WTI
2	QCOM	QUALCOMM Inc.	Information Technology	Semiconductors	Brent, WTI
2	UNP	Union Pacific Corp	Industrials	Railroads	WTI
2	VZ	Verizon Communications	Communication Services	Integrated Telecommunication Services	
2	WMT	Walmart	Consumer Staples	Hypermarkets and Super Centers	WTI
2	XOM	Exxon Mobil Corp.	Energy	Integrated Oil and Gas	
3	CMCSA	Comcast Corp.	Communication Services	Cable and Satellite	
3	GE	General Electric	Industrials	Industrial Conglomerates	
3	LLY	Lilly (Eli) & Co.	Health Care	Pharmaceuticals	Brent
3	LMT	Lockheed Martin Corp.	Industrials	Aerospace and Defense	
3	MDT	Medtronic plc	Health Care	Health Care Equipment	
3	MRK	Merck & Co.	Health Care	Pharmaceuticals	Brent, WTI
3	PFE	Pfizer Inc.	Health Care	Pharmaceuticals	WTI
3	T	AT&T Inc.	Communication Services	Integrated Telecommunication Services	
3	UPS	United Parcel Service	Industrials	Air Freight and Logistics	
3	USB	U.S. Bancorp	Financials	Diversified Banks	
3	WFC	Wells Fargo	Financials	Diversified Banks	
4	ADBE	Adobe Systems Inc	Information Technology	Application Software	
4	AXP	American Express Co	Financials	Consumer Finance	
4	BA	Boeing Company	Industrials	Aerospace and Defense	
4	BAC	Bank of America Corp	Financials	Diversified Banks	
4	C	Citigroup Inc.	Financials	Diversified Banks	
4	CME	CME Group Inc.	Financials	Financial Exchanges and Data	
4	CSCO	Cisco Systems	Information Technology	Communications Equipment	Brent, WTI
4	FIS	Fidelity National Information Services	Information Technology	Data Processing and Outsourced Services	Brent, WTI
4	GS	Goldman Sachs Group	Financials	Investment Banking and Brokerage	Brent
4	HON	Honeywell Int'l Inc.	Industrials	Industrial Conglomerates	Brent
4	JPM	JPMorgan Chase & Co.	Financials	Diversified Banks	
4	MMM	3M Company	Industrials	Industrial Conglomerates	
4	MS	Morgan Stanley	Financials	Investment Banking and Brokerage	WTI
4	NVDA	Nvidia Corporation	Information Technology	Semiconductors	Brent, WTI
4	SBUX	Starbucks Corp.	Consumer Discretionary	Restaurants	
4	SYK	Stryker Corp.	Health Care	Health Care Equipment	
4	TMO	Thermo Fisher Scientific	Health Care	Life Sciences Tools and Services	
4	TXN	Texas Instruments	Information Technology	Semiconductors	
4	UNH	United Health Group Inc.	Health Care	Managed Health Care	WTI
4	UTX	United Technologies	Industrials	Aerospace and Defense	

## References

1. Reboredo, J.C.; Ugolini, A. Quantile dependence of oil price movements and stock returns. *Energy Econ.* **2016**, *54*, 33–49. [[CrossRef](#)]
2. Narayan, P.K.; Gupta, R. Has oil price predicted stock returns for over a century? *Energy Econ.* **2015**, *48*, 18–23. [[CrossRef](#)]
3. Chen, S.-S. Forecasting crude oil price movements with oil-sensitive stocks. *Econ. Inq.* **2014**, *52*, 830–844. [[CrossRef](#)]

4. Han, L.; Lv, Q.; Yin, L. Can investor attention predict oil prices? *Energy Econ.* **2017**, *66*, 547–558. [[CrossRef](#)]
5. Zhang, Y.; Ma, F.; Wang, Y. Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? *J. Empir. Financ.* **2019**, *54*, 97–117. [[CrossRef](#)]
6. Kim, J.-M.; Jung, H. Dependence Structure between Oil Prices, Exchange Rates, and Interest Rates. *Energy J.* **2018**, *39*, 259–280. [[CrossRef](#)]
7. Kamalov, F.; Smail, L.; Gurrib, I. Stock price forecast with deep learning. In Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA), Online, 8–9 November 2020; pp. 1098–1102.
8. Anand, C. Comparison of Stock Price Prediction Models using Pre-trained Neural Networks. *J. Ubiquitous Comput. Commun. Technol. (UCCT)* **2021**, *3*, 122–134.
9. Sharma, A.; Tiwari, P.; Gupta, A.; Garg, P. Use of LSTM and ARIMAX Algorithms to Analyze Impact of Sentiment Analysis in Stock Market Prediction. In *Intelligent Data Communication Technologies and Internet of Things. Lecture Notes on Data Engineering and Communications Technologies*; Hemanth, J., Bestak, R., Chen, J.Z., Eds.; Springer: Singapore, 2021; Volume 57. [[CrossRef](#)]
10. Chen, Z. Gaussian Process Regression Methods and Extensions for Stock Market Prediction. Ph.D. Thesis, University of Leicester, Leicester, UK, 2017.
11. Bisht, A.; Chahar, A.; Kabthiyal, A.; Goel, A. Stock Prediction using Gaussian Process Regression. In Proceedings of the 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 29–31 March 2022; pp. 693–699. [[CrossRef](#)]
12. Erickson, C.B.; Ankenman, B.E.; Sanchez, S.M. Comparison of Gaussian process modeling software. *Eur. J. Oper. Res.* **2018**, *266*, 179–192. [[CrossRef](#)]
13. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and Analysis of Computer Experiments. *Stat. Sci.* **1989**, *4*, 409–423. [[CrossRef](#)]
14. Erickson, C. GauPro: Gaussian Process Fitting. R Package Version 0.2.2. 2017. Available online: <https://CRAN.R-project.org/package=GauPro> (accessed on 19 March 2021).
15. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Available online: [www.GaussianProcess.org/gpml](http://www.GaussianProcess.org/gpml) (accessed on 19 March 2021).
16. Kim, J.-M.; Kim, D.H.; Jung, H. Modeling Non-normal Corporate Bond Yield Spreads by Copula. *N. Am. J. Econ. Financ.* **2020**, *53*, 101210. [[CrossRef](#)]
17. Kim, J.-M.; Jung, H. The impacts of COVID-19 on the dependence structure of the stock market. *Appl. Econ. Lett.* **2022**, 1–6. [[CrossRef](#)]
18. Kim, J.-M.; Jung, H.; Yang, B. A revisit to size anomalies in U.S. bank stock returns by panel copula. *Appl. Econ. Lett.* **2022**, *29*, 750–754. [[CrossRef](#)]
19. Sklar, M. *Fonctions de Répartition À N Dimensions et Leurs Marges*; Université Paris 8: Paris, France, 1959.
20. Joe, H. *Multivariate Models and Dependence Concepts*; Chapman & Hall: London, UK, 1997.
21. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: New York, NY, USA, 2006.
22. Nagler, T.; Schepsmeier, U.; Stoeber, J.; Brechmann, E.C.; Graeler, B.; Erhardt, T. VineCopula: Statistical Inference of Vine Copulas. In *R Package, VineCopula*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
23. Jang, H.; Kim, J.-M.; Noh, H. Vine Copula Granger Causality in Mean. *Econ. Model.* **2022**, *109*, 105798. [[CrossRef](#)]
24. Kim, J.-M.; Ha, I.D. Deep Learning-Based Residual Control Chart for Count Data. *Qual. Eng.* **2022**, *34*, 370–381. [[CrossRef](#)]
25. Rong, X. Deep Learning Toolkit in R. In *R Package, Deepnet*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
26. Bayarri, M.J.; Berger, J.O.; Forte, A.; Garcia-Donato, G. Criteria for Bayesian model choice with application to variable selection. *Ann. Stat.* **2012**, *40*, 1550–1577. [[CrossRef](#)]
27. Garcia-Donato, G.; Forte, A.; Vergara-Hernández, C. Bayes Factors, Model Choice and Variable Selection in Linear Models. In *R Package, BayesVarSel*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
28. Schölkopf, B.; Smola, A.; Müller, K.-R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Comput.* **1998**, *10*, 1299–1319. [[CrossRef](#)]
29. Karatzoglou, A.; Smola, A.; Hornik, K.; National ICT Australia; Maniscalco, M.A.; Teo, C.H. Kernel-Based Machine Learning Lab. In *R Package, Kernlab*; R Foundation for Statistical Computing: Vienna, Austria, 2019.