

Article

Forecasting Crude Oil Risk Using a Multivariate Multiscale Convolutional Neural Network Model

Yingchao Zou and Kaijian He * 

College of Tourism, Hunan Normal University, Changsha 410081, China

* Correspondence: paulhekj@hunnu.edu.cn; Tel.: +86-155-1028-5982

Abstract: In light of the increasing level of correlation and dependence between the crude oil markets and the external influencing factors in the related financial markets, we propose a new multivariate empirical decomposition convolutional neural network model to incorporate the external influence of financial markets such as stock market and exchange market in a multiscale setting into the modeling of crude oil market risk movement. We propose a multivariate empirical model decomposition to analyze the finer details of interdependence among risk movement of different markets across different time horizons or scales. We also introduce the convolutional neural network to construct a new nonlinear ensemble algorithm to reduce the estimation bias and improve the forecasting accuracy. We used the major crude oil price data, stock market index, and the euro/United States dollar exchange rate data to evaluate the performance of the multivariate empirical model decomposition convolutional neural network model. The combination of both the multivariate empirical model decomposition and the convolutional neural network model in this paper has produced the risk forecasts with significantly improved risk forecasting accuracy.

Keywords: crude oil price; value-at-risk; multivariate empirical mode decomposition (MEMD) model; multi-scale analysis; convolutional neural network model

MSC: 62M10; 91B84; 68T07



Citation: Zou, Y.; He, K. Forecasting Crude Oil Risk Using a Multivariate Multiscale Convolutional Neural Network Model. *Mathematics* **2022**, *10*, 2413. <https://doi.org/10.3390/math10142413>

Academic Editors: Olimpia Neagu and Mihaela Simionescu

Received: 8 June 2022

Accepted: 6 July 2022

Published: 11 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

With the deregulation wave spreading across the crude oil markets around the world and the widespread usage of energy derivatives, we have witnessed closer interaction and linkage between the crude oil markets and the global financial markets. Numerous empirical studies have reported the intensifying financialization of the crude oil markets, and the complicated relationship between crude oil markets, the stock market, and the exchange market [1,2]. For example, Ji [3] has discovered that before the 2008 global financial crisis, the crude oil markets were affected by the speculation factors in the short term and the fundamental factors in the long term. However, after 2008, the dependence between crude oil markets and influencing factors was embodied in the spillover effect between the crude oil markets and other financial markets such as the stock market, foreign exchange market, and commodities market. Du and He [4] found that there is a significant spillover effect between the stock markets and crude oil markets. Chen and Lv [5] used the extreme value theory to analyze the extreme correlation between the crude oil markets and Chinese stock markets. They confirmed that the extreme correlation reached its peak during the financial crisis. de Truchis and Keddad [6] used the fractal co-integration and the copula method to identify the volatility dependence between the crude oil price and the Japanese yen, subject to the influence of the financial crisis. Husain et al. [7] found that there is significant correlation and spillover effect between the crude oil price, stock index, and the metal price. However, the integration of these financial markets as the exogenous variables (factors) into the crude oil risk analysis and forecasting has been a difficult research

problem in the literature, which attracted significant attention [8–11]. From a data analysis and forecasting perspective, the dependence between the crude oil markets and financial factors such as stock index and exchange rate are highly complex to capture in the modeling process. The investigations into the forecasting model incorporating the influences of these financial factors such as stock index and exchange rate are critical to the improvement of the risk forecasting accuracy.

There are a series of risk measures such as standard deviation, value at risk (VaR), and conditional value at risk (CVaR) in the literature. Among them, VaR is the most widely used one, which defines the tail loss at the pre-specified confidence level and investment time horizon [12]. It summarizes and reports the downside risk level as the critical information of the asset distribution that investors are concerned about. In this paper, VaR is used to measure the risk level that investors are most concerned about. The estimation and modeling of risk measures employ parametric models such as econometric and time series models, as well as nonparametric models. These models in the typical VaR estimation would rely on the historical data itself to infer the distribution and the estimation of the appropriate quantile. The difference between these two approaches is the additional assumptions imposed on the historical data in the parametric approach [12]. The autoregressive moving average-generalized autoregressive conditional heteroskedasticity (ARMA-GARCH) model is a parametric approach that imposes strict assumptions on the underlying distribution and model parameters. For example, Fiszeder et al. [13] proposed the range-based dynamic conditional correlation (DCC) GARCH model to estimate VaR in commodity and equity markets. Aloui and Mabrouk [14] found the fractionally integrated GARCH (FIAPARCH) model with skewed Student *t* distribution provided better forecasting accuracy compared to the benchmark model. Bedowska-Sojka [15] showed through comprehensive model evaluation that the GARCH model still represents one of the most robust models to estimate VaR with intraday data, better than the popular models such as the realized volatility model. Nonparametric models such as Monte Carlo and historical simulation models use the historical data changes to infer the future risk exposure and have been used extensively in the industry due to their assumption-free approach [16–20].

In the meantime, the semi-parametric approaches have emerged as the alternatives to combine the advantages of both parametric and nonparametric models in VaR research. New models in this category have demonstrated promising positive performance improvement and have attracted increasing attention. These models include extreme value theory, regime switching models, multiscale analysis, etc. [21]. Among them, multiscale models such as the empirical mode decomposition (EMD) are the latest development and have attracted increasing levels of attention in the risk measurement field. For example, Biage [22] used the wavelet analysis to model the stock volatility at different frequency scales and construct more accurate VaR estimates. Cifter [23] proposed a wavelet-based extreme value theory to estimate more accurate VaR. Zhu et al. [24] used the EMD model to construct the multiscale data structure for the carbon price and identify the extreme events. That way they improved the VaR estimation accuracy with better risk modeling. He et al. [25] identified the transient and extreme risk factors in the multiscale data domain extracted with the variational mode decomposition (VMD) model. Empirical studies conducted in the crude oil markets have demonstrated the evidence that the separate modeling of risk of different characteristics has produced VaR with higher-level reliability and accuracy. Zou et al. [26] have used the convolutional neural network (CNN) to consider risk forecasts at different scales to improve the overall risk estimation accuracy. The performance of the proposed model has been evaluated and verified using the trading data in the crude oil markets. Given the success of new techniques such as EMD in the modeling of univariate risk measures, their recent extension to the multivariate case can be introduced into the multivariate risk analysis and modeling. The widely popular deep learning model can also contribute to accurate forecast risk forecasting. Limited attempts in the literature may include recent research using the high dimensional multiscale analysis models such as bivariate EMD to estimate VaR for a portfolio consisting of multiple assets in metal markets, crude oil

markets, etc. For example, He et al. [27] proposed the portfolio value at risk model based on bivariate EMD and the Copula technique and applied it to the precious metal market. Experiment results showed improved risk forecasting accuracy.

The purpose of this study is to apply the multivariate empirical mode decomposition (MEMD) model and the CNN model to investigate the multiscale nonlinear relationship between crude oil markets and financial markets such as foreign exchange markets and stock markets using crude oil price data. We also propose the MEMD-CNN VaR model to forecast the downside risk in the crude oil markets. This model incorporates the interaction and multiscale dependence between the crude oil markets and financial markets into the multiscale risk estimation process. We further constructed the nonlinear ensemble forecasting model based on the CNN. An empirical study has been conducted using the major crude oil price, euro/United States dollar (EUR/USD) exchange rate, and Dow Jones Industrial Average (DJIA) stock index data as the typical test case to evaluate the risk forecasting performance of the proposed model.

The major contribution of this paper is the proposition of the MEMD-CNN VaR model. It quantifies the impacts of external influences on financial sectors such as the stock market and exchange market over different investment time horizons in the MEMD-CNN VaR model. The MEMD-CNN VaR model is unique in that we have introduced the MEMD model to capture and model the correlation and multiscale dependence in the high dimensional multiscale risk structure and risk movement in the crude oil markets. Deep learning models such as the CNN model have also been introduced to construct a nonlinear ensemble algorithm to integrate individual forecasts across different scales and estimate the nonlinearity in the risk measure changes. To the best of our knowledge, very few approaches have been developed to address the modeling of risk measures in the multivariate multiscale setting, taking into account the external influences from the other financial markets. The work in this paper conducts an exploratory investigation into the use of multivariate multiscale models such as MEMD and deep learning models such as CNN to improve the VaR estimation accuracy.

We organize the rest of the paper as follows. In Section 2, we present details of the methodology used in the paper. This includes a brief introduction to the theory and algorithms for MEMD and CNN. Then we explain the theory and algorithm for the MEMD-CNN model in detail. Results from the experiments conducted to empirically evaluate the performance of the proposed model against the benchmark model have been reported in Section 3. Section 4 concludes and makes some summarizing remarks.

2. Methodology

2.1. Multivariate Empirical Mode Decomposition and Convolutional Neural Network

MD model is an extension of the traditional univariate EMD model in multivariate data analysis [28–31]. The problem with the traditional EMD model is that the decomposed intrinsic mode functions (IMFs) differ in the maximum number, the fluctuation range, and the center frequency of modes when the EMD model is applied to the analysis of high dimensional data [29]. This results in the well-known mode mixing problem. This causes trouble for the interpretation of the statistical and economic meaning of the decomposed IMFs. The recently proposed MEMD model extended the volatility concept in the EMD model into the rotation concept in the higher dimensional space. In MEMD analysis, the multivariate data is viewed as the sphere in the geometric space. The decomposed IMFs constitute different surfaces for the spheres, all of which rotate around the center of the sphere [28,30,31]. The rotation speed can be used to distinguish the surfaces between spheres. Therefore, during the analysis, the correlation and dependence between different elements of the multivariate matrix data are incorporated into the rotation concept during the analysis. The decomposition process for the IMFs would reflect the correlation between them.

The MEMD algorithm is illustrated as follows.

1. Sample from $n - 1$ dimensional sphere.

2. Using the directional vector x^{θ_k} and the input $v(t)_{t=1}^T$, calculate projection set vector $p^{\theta_k}(t)_{k=1}^K$.
3. For projection vector $p^{\theta_k}(t)_{k=1}^K$, determine the time point $t_i^{\theta_k}$ that corresponds to the maximal projection point.
4. Use the interpolation method to calculate the multivariate surface $e^{\theta_k}(t)_{k=1}^K$ over $[t_i^{\theta_k}, v(t_i^{\theta_k})]$.
5. Calculate the average of the curve surface $m(t)$, $m(t) = \frac{1}{K} \sum_{k=1}^K e^{\theta_k}(t)$.
6. Check the termination criteria on $d(t) = x(t) - m(t)$. If the termination criteria is satisfied, repeat the above steps on $x(t) - d(t)$. If the termination criteria is not satisfied, repeat the above steps on $d(t)$.

In the MEMD model, firstly, the higher dimensional data are projected into the direction vectors in the lower dimensional space. The triple cubic spline method is used to perform interpolation on the direction vectors at extreme points and construct the projection surface in different directions. The accuracy of the constructed surface would increase with more directions from which the projection has been performed. Mainstream projection algorithms include the uniform sampling method for direction vectors, and pseudo-Monte Carlo sampling series [28].

Depending on the number of layers, the neural network can be classified as a shallow neural network or deep neural network [32]. The seminal universal approximation theory originally shows that the shallow neural network with one hidden layer can approximate any bounded continuous function with arbitrary lower errors [33]. Although this theoretical result is often cited to support the belief that the shallow neural network is capable of approximating the nonlinearity in data, in practice the number of parameters to estimate would grow exponentially when the complexity in the nonlinear data grows. This would result in a lower level of efficiency in training, and pose the upper bound on the depth of the neural network in its early days of development. When the structure of the network moves from the shallow one to the deeper one with continuously growing computational power, the neural network structure and training algorithm has been designed to reduce the number of parameters to estimate and improve the training efficiency. The universal approximation theorem has also been extended to the deep neural network case to show that the deep neural network is theoretically equivalent to the shallow neural network in approximating nonlinear continuous functions. In addition, in practice, the deep neural network would have a much higher level of efficiency in function approximation than the shadow neural network [34,35]. For example, if the objective function consists of functions with local characteristics and hierarchical characteristics, the deep neural network can significantly lower the number of parameters to estimate and avoid the curse of dimensionality problem [34]. Another interesting finding in recent research is that the deep learning models would have better generalization when the model significantly overfits the data during the deep learning training process. The generalization and the forecasting accuracy of the models can be improved simply by increasing the depth of the neural network and the number of neurons [36]. Mainstream deep learning algorithms refer to a series of algorithms such as the CNN for image processing, the long short-term memory (LSTM) network for natural language processing, and the deep belief network [37–39].

Inspired by the human visual processing system where neurons are responsive to local information, a CNN is a specially designed deep neural network that focuses on the extraction of space invariant data characteristics during the data training and modeling [37,40,41]. The CNN network uses different layers to process information. For the higher dimensional time series data, the neurons in the convolutional layer use the kernel function or the filters to perform the linear convolutional process on the input data. Neurons at the same layer share the filters, i.e., they use the same filters to produce a feature map of the represented data characteristic [41]. Filters may differ across different layers so that the feature maps generated by different layers correspond to different data characteristics. Therefore, the CNN can represent different data characteristics with the combination of different features and different convolutional layers. The typical layers in the CNN model include the in-

put layer, the hidden layer, the pooling layer, and convolutional layer [32]. Each layer contains the corresponding neurons. CNN model has used the weight sharing and local responses during the network structure design and training. To process different data with different dimensions, the convolutional layer could be classified into the one-dimensional convolution layer, two-dimensional convolution layer, and higher-dimensional convolution. The one-dimensional convolution layer is especially suitable for the processing of natural language data and time series data. The two-dimensional convolution layer is the most popular one. It targets two-dimensional data such as images. Currently, the theoretical and mathematical theory for the convolutional layer is still in its early stage of development. The network structure and parameter design are usually determined based on the research of the empirical method.

Given input X , the feature map data at layer l is as in Equation (1) [42].

$$X^{l+1} = f(W^{l+1} \otimes dX^l + b^{l+1}) \tag{1}$$

where W is the weight matrix, b is the bias matrix, and f is the activation function.

Neurons in the pooling layers perform the pooling operation such as the maximum pooling and minimum pooling on the input data. The pooling function is shown in Equation (2).

$$d(X) = T_{r \in R}(X_{i \times T+r}) \tag{2}$$

where T is the pooling function, such as the max pooling or averaging pooling.

Fully connected layer use the transfer function to process the input X as in Equation (3).

$$f(X) = \max(0, X) \tag{3}$$

Typical activation functions include sigmoid, rectification linear unit (ReLU). ReLU is defined as the maximum value between 0 and X [42].

2.2. MEMD-CNN VaR Model

Given the random variable $Y = (y_1, y_2, \dots, y_T)$ as the crude oil price, $X = (x_1, x_2, \dots, x_T)$ and $Z = (z_1, z_2, \dots, z_T)$ as the financial variable data, construct the high dimensional matrix $W = [X' \ Y' \ Z']$. W is the $T \times 3$ matrix.

MEMD model is used to decompose W into C up to scale S , as in Equation (4).

$$W = \sum_{s=1}^S C_s + \epsilon \tag{4}$$

where $C = [C'_y \ C'_x \ C'_z]$, $C_y = (c_{y,1}, c_{y,2}, \dots, c_{y,T})$, $C_x = (c_{x,1}, c_{x,2}, \dots, c_{x,T})$, $C_z = (c_{z,1}, c_{z,2}, \dots, c_{z,T})$.

For the decomposed modes C , the ARMA model is used to estimate the conditional mean forecasts $U = [U'_y \ U'_x \ U'_z]$, $U_y = (\mu_{y,1}, \mu_{y,2}, \dots, \mu_{y,T})$, $U_x = (\mu_{x,1}, \mu_{x,2}, \dots, \mu_{x,T})$, $U_z = (\mu_{z,1}, \mu_{z,2}, \dots, \mu_{z,T})$, and the GARCH model is used to estimate the conditional standard deviation forecasts $G = [G'_y \ G'_x \ G'_z]$, $G_y = (\sigma_{y,1}, \sigma_{y,2}, \dots, \sigma_{y,T})$, $G_x = (\sigma_{x,1}, \sigma_{x,2}, \dots, \sigma_{x,T})$, $G_z = (\sigma_{z,1}, \sigma_{z,2}, \dots, \sigma_{z,T})$.

Suppose that there is a nonlinear relationship between the conditional mean forecasts U and the returns Y , and the nonlinear relationship f_μ is estimated as in Equation (5).

$$y_t = f_\mu(U_t) \tag{5}$$

f_μ can be estimated using nonlinear estimation models. For example, f_μ is estimated in the CNN model as the network structure and parameters in CNN models.

Suppose that there is nonlinear relationship between the conditional standard deviation forecasts G and the historical standard deviation forecasts σ_t , the nonlinear function f_σ is estimated as in Equation (6).

$$\sigma_t = f_\sigma(G_t) \quad (6)$$

f_σ can be estimated using nonlinear estimation models. For example, f_σ is estimated in the CNN model as the network structure and parameters in CNN models.

With the estimated functions f_μ and f_σ , we can calculate the forecasts $\hat{\mu}$ and $\hat{\sigma}$ using Equations (5) and (6). The VaR is calculated using parametric method as in Equation (7).

$$VaR(a, h, t) = -h\hat{\mu}_t + \sqrt{h}\hat{\sigma}_t Z_\alpha \quad (7)$$

where $a = 1 - cl$, Z_α is the quantile value for the standard normal distribution, h is the holding period, and $VaR_{a,h,t}$ is the VaR at the confidence level cl .

2.3. Data Collection

In this paper, we have used the West Texas Intermediate (WTI) crude oil price in the U.S., Brent crude oil price in the United Kingdom (UK), the EUR/USD exchange rate, and the DJIA index to construct the empirical dataset to evaluate the risk forecasting accuracy of the proposed MEMD-CNN model against the benchmark models such as the ARMA-GARCH model and VMD-CNN model. Numerous research demonstrated the spillover effect between the crude oil markets and other financial markets such as foreign exchange market, stock market, etc. [4–7,43–45]. The EUR/USD exchange rate and DJIA index are selected and included in this dataset. The EUR/USD exchange rate and the DJIA index represent some of the most frequently quoted exchange rates and stock indices in the world, respectively [46,47]. The data were obtained from Quandl which is a central data warehouse that provides a comprehensive collection of financial and economic datasets. In this paper, we constructed the dataset that covers the daily closing price from 2 January 2003 to 19 February 2019. This resulted in the collection of 4066 daily observations in the WTI market, 4104 daily observations in the Brent market, 4048 daily observations in EUR/USD exchange rate data, and 4051 daily observations in the DJIA index. To facilitate the empirical model evaluations, the dataset is divided into three parts, such as the training set, the validation set, and the test set based on the 70:30 ratio [48]. The first 70% of the dataset is reserved as the training set to estimate the model parameters for the econometric models, the middle 21% of the dataset is reserved as the validation set to estimate the hyperparameter for the deep learning models, and the rest 9% of the dataset is used as the test set to conduct the out-of-sample model evaluation. The risk forecasting performance of the MEMD-CNN VaR models and benchmark models are evaluated using the Kupiec backtesting procedure, i.e., the popular unconditional coverage test [49].

3. Empirical Results

In this section, we report the results from a series of experiments conducted to analyze the data characteristics and evaluate the risk forecasting performance of the MEMD-CNN model and benchmark models such as the ARMA-GARCH model and VMD-CNN model.

Firstly, we calculated the descriptive statistics and statistical tests for the crude oil markets, foreign exchange markets, and stock markets in this paper. Results are reported in Table 1.

From the results in Table 1, it can be seen that the risk exposure for the EUR/USD and DJIA index is at a high level. There is generally significant volatility and fluctuations in all crude oil markets except for the EUR/USD exchange rate, indicated by the positive standard deviation of the market return. This observation suggests that there is a significant risk exposure, with a high probability of extreme and transient event occurrence. Interestingly, the risk exposure for the DJIA index is significantly larger than the EUR/USD exchange rate. Since the kurtosis value for the DJIA is at 13.7536, which is much larger than the

kurtosis value for the EUR/USD at 5.6425, there is also a high correlation between the crude oil price and EUR/USD exchange rate. The correlation coefficients are larger than 0.65. The correlation between the crude oil markets and the DJIA index is at a low level, less than 0.05.

Table 1. Descriptive statistics and statistical tests.

Markets	Mean $\times 10^{-5}$	Standard Deviation	Skewness	Kurtosis	p_{JB}	p_{BDS}	ρ_{wti}	ρ_{brent}
r_{WTI}	13.8424	0.0238	-0.0229	7.4014	0.001	0	1	0.9742
r_{Brent}	18.9063	0.0215	0.0882	7.4790	0.001	0	0.9742	1
$r_{EUR/USD}$	2.0736	0.0061	0.1	5.6425	0.001	0	0.7391	0.6677
r_{DJIA}	20.7152	0.0108	-0.1548	13.7536	0.001	0	-0.0212	0.0446

Notes: p_{JB} is p value for the Jarque-Bera (JB) test of normality, p_{BDS} is p value for BDS test of independence.

In our experiment, we also include the VMD-CNN model as one of the benchmark models. The difference between the VMD model and the MEMD model is that the VMD model is a one-dimensional multiscale analysis technique. It decomposes multiple crude oil prices one by one, compared to the simultaneous decomposition process in the MEMD model. The purpose of the inclusion of this model in the empirical study is to evaluate the necessity of the introduction of the MEMD technique in the MEMD-CNN model.

Since the performance of the CNN is very sensitive to the hyperparameters, in our paper, the optimal network structure for the CNN is calculated using the greedy search method. The parameter to be optimized includes the number of kernels, i.e., the depth of the network. We calculate the risk forecasting accuracy for the CNN with different network parameters using the validation data set. The basic model parameters for the CNN model are as follows: the rolling window is 250; the validation data set is 252 for the WTI market and 252 for the Brent market; the decomposition scale is set to 8; the kernel shape is set to 2; the number of kernels is up to 8.

We adopted the network structure of the classic Alex network and made some minor changes to the original network structure, such as the last layer changed from the classifier to the regression layer. The CNN model structure is as follows: input data layer-convolutional layer-ReLU-normalization-pooling-convolutional layer-ReLU-normalization-pooling-convolutional layer-ReLU-pooling-convolutional layer-ReLU-pooling-fully link-regression layer.

We use the validation data to calculate the in-sample forecasting accuracy with the MEMD-CNN model. Results for both markets are reported in Table 2.

Table 2. In-sample forecasting performance for the MEMD-CNN VaR model using the validation data.

Markets	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
$MSE_{WTI,95\%,\times 10^{-4}}$	2.6890	6890	2.6893	2.6882	2.6892	2.7165	2.6906	2.6891
$MSE_{WTI,97.5\%,\times 10^{-4}}$	3.1898	3.1898	3.1901	3.1891	3.1903	3.2194	3.1858	3.1817
$MSE_{WTI,99\%,\times 10^{-4}}$	3.8816	3.8817	3.8820	3.8812	3.8824	3.9141	3.8712	3.8644
$MSE_{Brent,95\%,\times 10^{-4}}$	2.0657	2.0664	2.0660	2.0751	2.0871	2.0706	2.1003	2.0818
$MSE_{Brent,97.5\%,\times 10^{-4}}$	2.5292	2.5300	2.5295	2.5342	2.5481	2.5357	2.5637	2.5457
$MSE_{Brent,99\%,\times 10^{-4}}$	3.1654	3.1663	3.1657	3.1651	3.1813	3.1740	3.1995	3.1828

It is clear that the forecasting accuracy of risk measures changes as the number of features changes. The reliability of the risk measure changes as the number of filters increases accordingly. We did not observe the strict monotonic increasing or decreasing relationship between the risk forecasting accuracy and the number of filters. In this paper, we use the mean squared error (MSE) as the selection criteria. Since MSE serves as the indication of risk forecasting accuracy in the literature, we argue that the smaller MSE would suggest more accurate risk measures. Based on MSE minimization, the optimal network structure for the MEMD-CNN model is determined as follows: for the WTI market,

the filter size is 2×2 , the number of filters is 4, and the pool size is 1×1 . For the Brent market, the filter size is 2×2 , the number of filters is 1, and the pool size is 1×1 .

In-sample forecasting accuracy results using VMD-CNN model for both markets are reported in Table 3.

Table 3. In-sample forecasting performance for VMD-CNN VaR model using the validation data.

Markets	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)	(2, 7)	(2, 8)
$MSE_{WTI,95\%, \times 10^{-4}}$	2.6890	2.6889	2.6894	2.6923	2.6940	2.7587	2.6818	2.7465
$MSE_{WTI,97.5\%, \times 10^{-4}}$	3.1898	3.1897	3.1904	3.1941	3.1943	3.2612	3.1715	3.2347
$MSE_{WTI,99\%, \times 10^{-4}}$	3.8817	3.8816	3.8824	3.8874	3.8856	3.9551	3.8506	3.9121
$MSE_{Brent,95\%, \times 10^{-4}}$	2.0702	2.0706	2.0712	2.0705	2.0995	2.1311	2.1261	2.1318
$MSE_{Brent,97.5\%, \times 10^{-4}}$	2.5350	2.5355	2.5362	2.5354	2.5582	2.5975	2.5832	2.5965
$MSE_{Brent,99\%, \times 10^{-4}}$	3.1728	3.1734	3.1743	3.1733	3.1888	3.2371	3.2120	3.2343

We observe similar patterns in the results using the VMD-CNN model as those using the MEMD-CNN model. The accuracy and reliability of risk measures also change when the number of filters increases. Based on MSE minimization, the optimal network structure for the VMD-CNN model is determined as follows: for the WTI market, the filter size is 2×2 , the number of filters is 7, and the pool size is 1×1 . For the Brent market, the filter size is 2×2 , the number of filters is 1, and the pool size is 1×1 .

With the optimal network structure determined above, we further evaluate the risk forecasting accuracy using the out-of-sample test data with 360 daily observations in both markets. Performance measures are reported in Table 4.

Table 4. Out-of-sample model performance using the test data.

Model	$N_{95\%}$	$N_{97.5\%}$	$N_{99\%}$	$p_{95\%}$	$p_{97.5\%}$	$p_{99\%}$	$MSE_{95\%, \times 10^{-3}}$	$MSE_{97.5\%, \times 10^{-3}}$	$MSE_{99\%, \times 10^{-3}}$
ARMA – GARCH _{WTI}	22	10	6	0.3492	0.7401	0.0054	1.4780	1.8740	2.4289
VMD – CNN _{WTI}	29	13	9	0.0142	0.2050	0.0163	1.4298	1.7568	2.2135
MEMD – CNN _{WTI}	17	9	6	0.8072	1	0.2460	1.4295	1.7562	2.2128
ARMA – GARCH _{Brent}	24	14	4	0.1665	0.1181	0.8351	1.1621	1.5143	2.0077
VMD – CNN _{Brent}	17	7	2	0.8072	0.4826	0.3549	1.1050	1.4067	1.8299
MEMD – CNN _{Brent}	17	7	2	0.8072	0.4826	0.3549	1.1049	1.4066	1.8297

From results in Table 4, it is clear that the risk forecasting reliability and accuracy of the MEMD-CNN models have improved significantly in both WTI and Brent crude oil markets compared to the benchmark models in terms of the Kupiec backtesting procedure and MSE. The benchmark ARMA-GARCH model has failed to pass the statistical test at a 99% confidence level with a p value of 0.0054, less than the cutoff value of 0.05 in the WTI market. In terms of the number of exceedance, the proposed model achieves a much lower value than the benchmark model. This suggests that the VaR estimates from the proposed model are more conservative and provide better risk coverage, which corresponds to the lower number of cases of default and financial distress. In the meantime, MSE for the MEMD-CNN model decreases uniformly compared to the benchmark ARMA-GARCH model, as well as the VMD-CNN model. For example, at a 95% confidence level, MSE decreases from 1.1621 for the ARMA-GARCH model to 1.1049 for the MEMD-CNN model, and MSE decreases marginally from 1.1050 for the VMD-CNN model to 1.1049 for MEMD-CNN model. The forecasting accuracy improvement is more significant when the performance of the MEMD-CNN model is compared to the benchmark ARMA-GARCH model.

It is interesting to observe from experiment results that the choice of different multi-scale models would affect the forecasting accuracy of the ensemble models. In general, the VaR from the proposed multivariate MEMD-CNN models is more accurate than VaR from univariate VMD-CNN models. VMD-CNN VaR model fails to pass the statistical tests at 95% and 99% confidence levels in the WTI market. Meanwhile, their forecasts from the

multivariate models pass all the confidence levels in both markets. Based on the ensemble theory, the forecasting accuracy improvement suggest the reduction in the variance in the individual ensemble member forecasts and improvement in the heterogeneity of the ensemble members. The decomposed components from multivariate models serving as the ensemble members have a higher degree of heterogeneity. Our experiment results suggest that the introduction of the CNN-based nonlinear ensemble model can effectively reduce the estimation bias from the multiscale decomposition and risk forecasts. It would lead to improvement in forecasting accuracy and stability.

4. Conclusions

In this paper, we have applied the higher dimensional MEMD model to analyze the multiscale data characteristics of risk changes in the crude oil markets, as well as the interaction between the crude oil markets, the DJIA index, and the EUR/USD exchange rate. We have further constructed a crude oil price VaR forecasting model based on MEMD and CNN models. Risk movements at the individual scale are projected into the future using the ARMA-GARCH model. The risk forecast matrix is used as input to the CNN-based nonlinear ensemble forecasting model to predict the future risk movement. Empirical studies have been conducted using the crude oil price data, the EUR/USD exchange rate data, and stock index data as the typical test case. A comprehensive model performance comparison has been conducted in terms of risk forecasting reliability. Results show that the proposed multivariate multiscale ensemble model forecasts risk more accurately than the benchmark models. The deep learning model is especially helpful in providing robust risk forecasts.

Based on the analysis of the results, the work in this paper has two implications for the risk estimation field, in terms of testing of existence of multiscale data features and model construction. Firstly, the influence of the external risk factors for the crude oil markets can be incorporated with consideration of its fine multiscale risk structure. The potential performance improvement requires the introduction of new multivariate multiscale models, in addition to the existing multiscale models. This is expected to result in more accurate feature extraction. Secondly, there is clearly a nonlinear relationship between the movement of external influencing factors across different scales and the crude oil price movement, as indicated by the success of the deep learning based modeling approach to model it in an assumption-free model-free approach. Although it is preferable to derive the exact analytic form of these nonlinear data characteristics and functions, in practice the nonlinear function is often too complex to be computationally feasible in the mainstream approaches. Future research would be directed toward the continuous searches for the analytic form of the nonlinear function, as well as the emergence of a new design of innovative deep neural network structure for its better approximation.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z. and K.H.; writing—original draft preparation, Y.Z.; writing—review and editing, K.H.; funding acquisition, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China grant number 71671013, Hunan Provincial Natural Science Foundation of China grant number 2022JJ30401, and the Humanities and Social Sciences Youth foundation of the Ministry of Education of China grant number 16YJC790026.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: <https://www.quandl.com> (accessed on 6 June 2022).

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

References

1. Hoque, M.E.; Low, S.W. Impact of Industry-Specific Risk Factors on Stock Returns of the Malaysian Oil and Gas Industry in a Structural Break Environment. *Mathematics* **2022**, *10*, 199. [\[CrossRef\]](#)
2. Lin, A.J.; Chang, H.Y. Volatility Transmission from Equity, Bulk Shipping, and Commodity Markets to Oil ETF and Energy Fund—A GARCH-MIDAS Model. *Mathematics* **2020**, *8*, 1534. [\[CrossRef\]](#)
3. Ji, Q. System analysis approach for the identification of factors driving crude oil prices. *Comput. Ind. Eng.* **2012**, *63*, 615–625. [\[CrossRef\]](#)
4. Du, L.; He, Y. Extreme risk spillovers between crude oil and stock markets. *Energy Econ.* **2015**, *51*, 455–465. [\[CrossRef\]](#)
5. Chen, Q.; Lv, X. The extreme-value dependence between the crude oil price and Chinese stock markets. *Int. Rev. Econ. Financ.* **2015**, *39*, 121–132. [\[CrossRef\]](#)
6. de Truchis, G.; Keddad, B. On the risk comovements between the crude oil market and U.S. dollar exchange rates. *Econ. Model.* **2016**, *52*, 206–215. [\[CrossRef\]](#)
7. Husain, S.; Tiwari, A.K.; Sohag, K.; Shahbaz, M. Connectedness among crude oil prices, stock index and metal prices: An application of network approach in the USA. *Resour. Policy* **2019**, *62*, 57–65. [\[CrossRef\]](#)
8. Zhang, Y.J.; Wang, J.L. Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. *Energy Econ.* **2019**, *78*, 192–201. [\[CrossRef\]](#)
9. Bonnier, J.B. Forecasting crude oil volatility with exogenous predictors: As good as it GETS? *Energy Econ.* **2022**, *111*, 106059. [\[CrossRef\]](#)
10. Chen, Y.; Qiao, G.; Zhang, F. Oil price volatility forecasting: Threshold effect from stock market volatility. *Technol. Forecast. Soc. Chang.* **2022**, *180*, 121704. [\[CrossRef\]](#)
11. Ivanovski, K.; Hailemariam, A. Forecasting the dynamic relationship between crude oil and stock prices since the 19th century. *J. Commod. Mark.* **2021**, *24*, 100169. [\[CrossRef\]](#)
12. Dowd, K. *Measuring Market Risk*; John Wiley & Sons Inc.: Chichester, UK; Hoboken, NJ, USA, 2005; p. 390.
13. Fiszeder, P.; Faldzinski, M.; Molnar, P. Range-based DCC models for covariance and value-at-risk forecasting. *J. Empir. Financ.* **2019**, *54*, 58–76. [\[CrossRef\]](#)
14. Aloui, C.; Mabrouk, S. Value-At-Risk Estimations of Energy Commodities Via Long-Memory, Asymmetry and Fat-Tailed Garch Models. *Energy Policy* **2010**, *38*, 2326–2339. [\[CrossRef\]](#)
15. Bedowska-Sojka, B. Is intraday data useful for forecasting VaR? The evidence from EUR/PLN exchange rate. *Risk Manag. J. Risk Crisis Disaster* **2018**, *20*, 326–346. [\[CrossRef\]](#)
16. Fries, C.P.; Nigbur, T.; Seeger, N. Displaced relative changes in historical simulation: Application to risk measures of interest rates with phases of negative rates. *J. Empir. Financ.* **2017**, *42*, 175–198. [\[CrossRef\]](#)
17. Liu, X. Historical simulation with component weight and ghosted scenarios. *J. Risk* **2015**, *18*, 1–25. [\[CrossRef\]](#)
18. Tzeng, Y.Y.; Beaumont, P.M.; Okten, G. Time Series Simulation with Randomized Quasi-Monte Carlo Methods: An Application to Value at Risk and Expected Shortfall. *Comput. Econ.* **2018**, *52*, 55–77. [\[CrossRef\]](#)
19. Hsieh, M.H.; Liao, W.C.; Chen, C.L. A Fast Monte Carlo Algorithm for Estimating Value at Risk and Expected Shortfall. *J. Deriv.* **2014**, *22*, 50–66. [\[CrossRef\]](#)
20. Hong, L.J.; Hu, Z.; Liu, G. Monte Carlo Methods for Value-at-Risk and Conditional Value-at-Risk: A Review. *ACM Trans. Model. Comput. Simul.* **2014**, *24*, 1–37. [\[CrossRef\]](#)
21. Echaust, K.; Just, M. Value at Risk Estimation Using the GARCH-EVT Approach with Optimal Tail Selection. *Mathematics* **2020**, *8*, 114. [\[CrossRef\]](#)
22. Biage, M. Analysis of shares frequency components on daily value-at-risk in emerging and developed markets. *Phys. A Stat. Mech. Its Appl.* **2019**, *532*, 121798. [\[CrossRef\]](#)
23. Cifter, A. Value-at-risk estimation with wavelet-based extreme value theory: Evidence from emerging markets. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 2356–2367. [\[CrossRef\]](#)
24. Zhu, B.; Ye, S.; He, K.; Chevallier, J.; Xie, R. Measuring the risk of European carbon market: An empirical mode decomposition-based value at risk approach. *Ann. Oper. Res.* **2019**, *281*, 373–395. [\[CrossRef\]](#)
25. He, K.; Tso, G.K.; Zou, Y.; Liu, J. Crude oil risk forecasting: New evidence from multiscale analysis approach. *Energy Econ.* **2018**, *76*, 574–583. [\[CrossRef\]](#)
26. Zou, Y.; Yu, L.; Tso, G.K.; He, K. Risk forecasting in the crude oil market: A multiscale Convolutional Neural Network approach. *Phys. A Stat. Mech. Its Appl.* **2019**, *541*, 123360. [\[CrossRef\]](#)
27. He, K.; Zha, R.; Chen, Y.; Lai, K.K. Forecasting Energy Value at Risk Using Multiscale Dependence Based Methodology. *Entropy* **2016**, *18*, 170. [\[CrossRef\]](#)
28. Rehman, N.; Mandic, D.P. Multivariate empirical mode decomposition. *Proc. R. Soc. Math. Phys. Eng. Sci.* **2010**, *466*, 1291–1302. [\[CrossRef\]](#)
29. Rilling, G.; Flandrin, P.; Goncalves, P.; Lilly, J.M. Bivariate Empirical Mode Decomposition. *IEEE Signal Process. Lett.* **2007**, *14*, 936–939. [\[CrossRef\]](#)
30. Mandic, D.P.; Rehman, N.U.; Wu, Z.H.; Huang, N.E. Empirical Mode Decomposition-Based Time-Frequency Analysis of Multivariate Signals. *IEEE Signal Process. Mag.* **2013**, *30*, 74–86. [\[CrossRef\]](#)

31. Park, C.; Looney, D.; ur Rehman, N.; Ahrabian, A.; Mandic, D.P. Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 10–22. [[CrossRef](#)]
32. Sezer, O.B.; Ozbayoglu, A.M. Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Appl. Soft Comput.* **2018**, *70*, 525–538. [[CrossRef](#)]
33. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257. [[CrossRef](#)]
34. Poggio, T.; Mhaskar, H.; Rosasco, L.; Miranda, B.; Liao, Q. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. *Int. J. Autom. Comput.* **2017**, *14*, 503–519. [[CrossRef](#)]
35. Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Netw.* **2017**, *94*, 103–114. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. In Proceedings of the ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
37. Liu, C.; Hou, W.; Liu, D. Foreign Exchange Rates Forecasting with Convolutional Neural Network. *Neural Process. Lett.* **2017**, *46*, 1095–1119. [[CrossRef](#)]
38. Chen, Y.; Zhang, C.; He, K.; Zheng, A. Multi-step-ahead crude oil price forecasting using a hybrid grey wave model. *Phys. A Stat. Mech. Its Appl.* **2018**, *501*, 98–110. [[CrossRef](#)]
39. Tzirakis, P.; Trigeorgis, G.; Nicolaou, M.A.; Schuller, B.W.; Zafeiriou, S. End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 1301–1309. [[CrossRef](#)]
40. Hinton, G.E.; Osindero, S.; Teh, Y.W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* **2014**, *18*, 1527–1554. [[CrossRef](#)]
41. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
42. Zhao, Z.; Zhang, Y.; Deng, Y.; Zhang, X. ECG authentication system design incorporating a convolutional neural network and generalized S-Transformation. *Comput. Biol. Med.* **2018**, *102*, 168–179. [[CrossRef](#)]
43. Brahmašre, T.; Huang, J.C.; Sissoko, Y. Crude oil prices and exchange rates: Causality, variance decomposition and impulse response. *Energy Econ.* **2014**, *44*, 407–412. [[CrossRef](#)]
44. Belgacem, A.; Creti, A.; Guesmi, K.; Lahiani, A. Volatility spillovers and macroeconomic announcements: Evidence from crude oil markets. *Appl. Econ.* **2015**, *47*, 2974–2984. [[CrossRef](#)]
45. Coleman, L. Explaining crude oil prices using fundamental measures. *Energy Policy* **2012**, *40*, 318–324. [[CrossRef](#)]
46. Shoven, J.; Sialm, C. The Dow Jones Industrial Average: The Impact of Fixing Its Flaws. *J. Wealth Manag.* **2000**, *3*, 9–18. [[CrossRef](#)]
47. Antonakakis, N.; Chatziantoniou, I.; Gabauer, D. The impact of Euro through time: Exchange rate dynamics under different regimes. *Int. J. Financ. Econ.* **2020**, *26*, 1375–1408. [[CrossRef](#)]
48. Zou, H.; Xia, G.; Yang, F.; Wang, H. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing* **2007**, *70*, 2913–2923. [[CrossRef](#)]
49. Kupiec, P. Techniques for Verifying the Accuracy of Risk Measurement Models. *J. Deriv.* **1995**, *3*, 73–84. [[CrossRef](#)]