Crude oil price prediction using temporal fusion transformer model

He, Kaijian; Zheng, Linyuan; Yang, Qian; Wu, Chiwai; Yu, Yi; Zou, Yingchao

Published in:
Procedia Computer Science

Published: 01/01/2023

Document Version:
Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

Published version (DOI):
10.1016/j.procs.2023.08.070

Publication details:
He, K., Zheng, L., Yang, Q., Wu, C., Yu, Y., & Zou, Y. (2023). Crude oil price prediction using temporal fusion transformer model. *Procedia Computer Science*, *221*, 927-932. https://doi.org/10.1016/j.procs.2023.08.070

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 221 (2023) 927–932

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2023)

Crude oil price prediction using temporal fusion transformer model

Kaijian He[a], Linyuan Zheng[a], Qian Yang[a], Chiwai Wu[b,c], Yi Yu[d], Yingchao Zou[a,*]

[a]*College of Tourism, Hunan Normal University, Changsha 410081, China*
[b]*School of Tourism Management, Macao Institute for Tourism Studies, Macao, China*
[c]*Department of Marketing, City University of Hong Kong, Hong Kong, China*
[d]*School of Foreign Languages, Hunan University, Changsha, 410082, China*

Abstract

In this paper, we applied the temporal fusion transformer model to the crude oil price movement modeling and forecasting. The temporal fusion transformer model has been adopted in the crude oil price forecasting model using the attention mechanism, to capture the different level of autocorrelations among observations in the crude oil prices. Empirical evaluation of the transformer based multi-horizon ahead crude oil price forecasting model has been conducted. Experiment results show that the introduction of transformer model in the forecasting process has improved the forecasting accuracy significantly at longer time horizon.

Keywords: Crude oil price forecasting; Temporal Fusion Transformer model; ARIMA

1. Introduction

Crude oil is one of the most important energy commodities that underpin the economic operations of the modern society [1]. The constant evaluations and forecasting of the price movement in this key commodities market is essential to the smooth operations of different levels of industries across the entire economy [2]. For example, the budget planning for the transportation industry requires the accurate forecasting of the crude oil price as the essential energy input. But its forecasting is inherently difficult since far too many factors could potentially affect its market demand and supply [3]. And the price fluctuations in the crude oil market has become larger and the crude oil price movement become increasingly more complicated [4, 5]. That poses higher demand for the forecasting accuracy of the crude oil forecasting models.

Over the years, numerous different models have been developed to improve the crude oil forecasting accuracy. These development efforts mainly follow two trends, i.e. econometric and time series models and Artificial Intelligence models. The econometric and times series approach have imposed restrictive assumptions on the link between the crude oil price changes and various influencing factors such as the external shocks and past observations [4, 6, 7]. For example, [8] proposed the diffusion indices with nonlinearity to model the quadratic terms

*Corresponding author. Tel.:+86-18511075982
Email address: evangeline1203@hunnu.edu.cn (Yingchao Zou)

10.1016/j.procs.2023.08.070

in the original predictors and found the improved forecasting performance. [9] showed that the shrinkage forecasts with different nonvonvex penalties and Huber loss function demonstrate the improved forecasting accuracy in the short term time horizon. [3] based on the assumption that potential structural breaks lead to parameter instability used four strategies to address the problem parameter instability, including rolling window, regime switching model, time-varying parameter model, and the time-dependent weighted least squares. The experiment results show that the forecast accuracy of the crude oil price prediction model can be improved. [10] found that ARIMA model produced international crude oil prices forecasts with good accuracy. Time series and economic models remain one of the most powerful tools that we have available to model and forecast crude oil time series movement. They serve as the benchmark models in the literature. Meanwhile, Artificial Intelligence models mainly follow different data-driven self-adaptive approach. Although the early development in the Artificial Intelligence such as neural network only mimics roughly the main information processing in the human brain and has achieved limited success [11]. The recent Artificial Intelligence models, such as deep learning models hasvetaken a more elaborated approach such as more abstract feature extraction and long memory mechanism to extract and model nonlinear data patterns and have demontrated promising results [12, 13, 14]. For example, spatial data features have been captured in the CNN model, Long Short Term Memory (LSTM) model [15, 16]. Correlation properties have been captured in the transformer model. [17] proposed a novel hybrid model which combined improved slope-based method (ISBM) ,empirical mode decomposition (EMD) and feed-forward neural network (FNN) method. He decomposed the time series of Brent crude oil prices using ISBM-based EMD method, and then inputted them into the feed-forward neural network for prediction.The study shows that the method proposed in this study outperformed other benchmark methods. [18] considered crude oil price forecasting using Google Trends and proposed a tree-based stacking ensemble method. Experiment results showed that after the introduction of Google Trends, the prediction accuracy of most models is improved, and the multiple methods has strong nonlinear time series prediction ability which is better than several popular single-model machine learning methods in terms of prediction accuracy. [19] constructed a fusion prediction and trading model of crude oil future price under multiple timeframe, and interpreted the model by using SHAP approach. The study show that the method proposed in this study outperformed other benchmark methods in terms of accuracy, profitability, and trading risk. [20] proposed a novel hybrid model for predicting WTI and Brent crude oil future prices based on the decomposition-prediction-ensemble-error correction framework, named VILIG, which uses an improved golden jackal optimization algorithm to optimize the hyperparameters of the LSTM. Despite the success of artificial intelligence models such as the deep learning model in different disciplines and applications, the deep learning model based time series forecasting have shown inconsistent and the mixture results in terms of forecasting performance improvement. It remains an open question that how the rapid development of deep learning model can contribute to the better modeling and forecasting of the time series forecasting.

In the literature, there are different ways to construct innovative network structure with different latest deep learning model to achieve better modeling and forecasting. For example, [21] combined Bidirectional long short-term memory network (BiLSTM), attention mechanism and Convolution Neural Network, the time series of crude oil prices are decomposed and reconstructed, and input the decomposed subsequence into the BiLSTM-Attention -CNN deep learning combination model for prediction. Temporal Fusion Transformer is one of the latest development in the transformer field. Although transformer has become very popular and prominent in the recent applications in different disciplines such as the large scale language model, natural language processing, image processing, etc. Its application in the time series forecasting, in general are restricted and limited. As one of the latest development in the Artificial Intelligence field, temporary fusion transformer model is specially designed to model a series of data characteristic among different transformer models and has been introduced into the crude oil price forecasting field [22]. We have proposed a new crude oil price forecasting model based on the temporal fusion transformer model. We have used the mainstream WTI crude oil price to evaluate the forecasting performance of this new model. And therefore we contribute to the forecasting literature by exploring the recent new deep learning model in capturing the essential data features and modeling the nonlinearity in the data.

We organize the rest of the paper as follows. Section 2 illustrates the details of the proposed temporal fusion transformer based crude oil forecasting model. Section 3 reports the results from the empirical studies applying the proposed model to the forecasting exercises of the daily crude oil data in the WTI markets. Section 4 provides some summarizing remarks.

2. Methodology

Given the complex relationship between the crude oil price changes over time, the forecasting of the crude oil price can be modeled as the nonlinear autoregressive process as in Eq. (1).

$$\hat{y}_{t+h} = f(y_1, y_2, \ldots, y_t) \tag{1}$$

Where y_t is the past observations at time t. If we make the restrictive linear assumption for function f, it will reduce to the simple benchmark AR model. But when this assumption is relaxed to the nonlinear assumption, various recent models can be introduced. It assumes that the past observations have different levels of correlations with the current price. The structure of the correlations are modeled using the encoder decoder model and attention mechanism in the temporary fusion transformer models. Different from the traditional artificial intelligence models, the temporal fusion transformer model employs the attention mechanisms that attributes weights as the attention to different observations in the data, as well as a series of canonical deep neural network component to capture different data features.

We propose a new temporal fusion transformer based forecasting model to model the nonlinear function f. TFT is the recent develoment of deep learning model field that takes full advantage of various deep learning models to model the time series data. More specifically it takes a block structure that consists of different stages of time series processing such as variable selection networks, static covariate encoder, interpretable multi-head attention, temporal fusion decoder, quantile output [22]. TFT uses the gated mechanism known called Gated Residual Network (GRN) as the building block for different layers in the network [22].

The Gated Residual Network (GRN) is defined as in Eq. (2) [22].

$$\begin{aligned} \mathrm{GRN}_{\omega}(\boldsymbol{a}, \boldsymbol{c}) &= \mathrm{LN}\left(\boldsymbol{a} + \mathrm{GLU}_{\omega}\left(\boldsymbol{\eta}_1\right)\right), \\ \boldsymbol{\eta}_1 &= \boldsymbol{W}_{1,\omega}\boldsymbol{\eta}_2 + \boldsymbol{b}_{1,\omega} \\ \boldsymbol{\eta}_2 &= \mathrm{ELU}\left(\boldsymbol{W}_{2,\omega}\boldsymbol{a} + \boldsymbol{W}_{3,\omega}\boldsymbol{c} + \boldsymbol{b}_{2,\omega}\right), \end{aligned} \tag{2}$$

where *ELU* is the Exponential Linear Unit activation function, *LN* is the normalization layer.

The variable selection layer selects the most important variables for the forecasting of the crude oil prices. The multi head attention layer attaches the weighting importance to different observations on the historical data. The attention mechanism scale value is defined in Eq. (3) [22].

$$Attention(Q, K, V) = A(Q, K)V \tag{3}$$

Where A is the normalization function.

The temporal fusion decoder models the complex temporal features through a combination of a series of layers such as locality enhancement layer, static enhancement layer, temporal self-attention layer, and the additional feed-forward layer to model the remaining nonlinearity in the data [22].

Quantile output layer calculates the output of the TFT, which produces different quantiles of the probability distribution. In terms of forecasting, the 50% quantile forecasts is used to represent the forecasts.

3. Experiment results

To evaluate the model performance, we have collected the latest crude oil price data in the Benchmark WTI crude oil market to conduct the empirical studies. The time period for the data spans from 27 April 2021 to 24 April 2023, which contains 500 daily observations. The data source is the Energy Information Administration, Department of Energy, U.S. The dataset is divided into both training set and test set based on the commonly used 60-40 ratio. Training set is used to estimate the model parameters while crude oil forecasts made are compared to the samples in the test set. We evaluate and compare the forecasting performance of the proposed temporal fusion transfer models. And the various benchmarks such as Random Walk and the ARIMA models are used as a benchmark model in the model evaluation process.

We calculated the descriptive statistics of the crude oil price and reports the general descriptive statistics in table 1.

Table 1. Descriptive statistics and statistical tests

Statistics	Mean	Standard Deviation	Skewness	Kurtosis	p_{JB}	p_{BDS}
p	84.3474	14.3694	0.7575	-0.3064	0	0

It's clear from table 1 that the crude oil price demonstrate volatile and nonstationary behavior. The volatility is significant with the standard deviation value 14.3694. There is positive skewness value at 0.7575. The kurtosis value is significantly different from the cutoff value 3. Both statistical tests are rejected, indicating that the distribution significantly deviates from the normal distribution and linearity.

We further report the results from the forecasting exercises conducted to evaluate the forecasting accuracy of the temporal fusion transformer model, as well as the benchmark Random Walk (RW) and ARIMA model, in table 2.

Table 2. Forecasting performance of Temporal Fusion Transformer model using out-of-sample dataset

Models	MSE_3	MSE_5	MSE_7	MSE_{10}	MSE_{14}
RW	6.5337	15.4263	25.3665	39.5985	51.0061
$ARIMA$	5.9512	12.9316	22.4510	39.36	61.1181
TFT	32.2424	28.5835	25.6408	20.8357	15.4962

Where $MSE_i, i = 3, 5, 7, 10, 14$ refer to MSE at day 3, 5, 7, 10 and 14. Results in table 2 show some interesting findings. We can see that the performance of the temporal fusion transformer model is superior to the performance of the benchmark RW and ARIMA model when it is used to make forecasts over the long term. But contrary to our common belief, the forecasting accuracy for the temporal fusion transformer model is inferior to the forecasting performance of the benchmark models. This is a surprising finding, consistent with the recent research in the empirical literature.

4. Conclusions

In this paper, we have proposed a new temporal fusion transformer based crude oil forecast model to forecast the crude oil changes. The newly proposed TFT based crude oil forecasting model has been evaluated along with the benchmark models using the real daily crude oil prices. Results confirm the performance gain from the introduction of the new TFT based crude oil forecasting models. Our work provides some interesting insights into the potential long term forecasting accuracy improvement from the use of the latest development in the deep learning models, and the potential of the deep learning models to mine and capture the mixture of complicated data features in the crude oil prices.

Acknowledgments

The work described in this paper was supported by a grant from National Natural Science Foundation of China (Grant no. 72271089), Hunan Provincial Natural Science Foundation of China (Grant no. 2022JJ30401).

References

[1] X. Gong, K. Guan, L. Chen, T. Liu, C. Fu, What drives oil prices? a markov switching var approach, Resources Policy 74 (2021) 102316. doi:https://doi.org/10.1016/j.resourpol.2021.102316.
URL https://www.sciencedirect.com/science/article/pii/S0301420721003263

[2] T. Xiong, Y. Bao, Z. Hu, Beyond one-step-ahead forecasting: Evaluation of alternative multi-step-ahead forecasting models for crude oil prices, Energy Economics 40 (2013) 405–415. doi:https://doi.org/10.1016/j.eneco.2013.07.028.
URL https://www.sciencedirect.com/science/article/pii/S0140988313001746

[3] Y. Wang, X. Hao, Forecasting the real prices of crude oil: What is the role of parameter instability?, Energy Economics 117 (2023) 106483. doi:https://doi.org/10.1016/j.eneco.2022.106483.
URL https://www.sciencedirect.com/science/article/pii/S0140988322006120

[4] H. Miao, S. Ramchander, T. Wang, D. Yang, Influential factors in crude oil price forecasting, ENERGY ECONOMICS 68 (2017) 77–88. doi:10.1016/j.eneco.2017.09.010.

[5] G. Li, S. Yin, H. Yang, A novel crude oil prices forecasting model based on secondary decomposition, Energy 257 (2022) 124684. doi:https://doi.org/10.1016/j.energy.2022.124684.
URL https://www.sciencedirect.com/science/article/pii/S0360544222015870

[6] F. Cheng, T. Li, Y. ming Wei, T. Fan, The vec-nar model for short-term forecasting of oil prices, Energy Economics-doi:https://doi.org/10.1016/j.eneco.2017.12.035.
URL http://www.sciencedirect.com/science/article/pii/S0140988318300082

[7] A. Nademi, Y. Nademi, Forecasting crude oil prices by a semiparametric markov switching model: Opec, wti, and brent cases, Energy Economics 74 (2018) 757 – 766. doi:https://doi.org/10.1016/j.eneco.2018.06.020.
URL http://www.sciencedirect.com/science/article/pii/S014098831830238X

[8] Y. Zhang, M. He, D. Wen, Y. Wang, Forecasting crude oil price returns: Can nonlinearity help?, Energy 262 (2023) 125589. doi:https://doi.org/10.1016/j.energy.2022.125589.
URL https://www.sciencedirect.com/science/article/pii/S0360544222024756

[9] L.-M. Xing, Y.-J. Zhang, Forecasting crude oil prices with shrinkage methods: Can nonconvex penalty and huber loss help?, Energy Economics 110 (2022) 106014. doi:https://doi.org/10.1016/j.eneco.2022.106014.
URL https://www.sciencedirect.com/science/article/pii/S0140988322001852

[10] M. Zolfaghari, S. Gholami, A hybrid approach of adaptive wavelet transform, long short-term memory and arima-garch family models for the stock index prediction, Expert Systems with Applications 182 (2021) 115149. doi:https://doi.org/10.1016/j.eswa.2021.115149.
URL https://www.sciencedirect.com/science/article/pii/S095741742100590X

[11] K. He, C. Xie, S. Chen, K. K. Lai, Estimating var in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network, Neurocomputing 72 (16) (2009) 3428–3438, financial Engineering Computational and Ambient Intelligence (IWANN 2007). doi:https://doi.org/10.1016/j.neucom.2008.09.026.
URL https://www.sciencedirect.com/science/article/pii/S0925231209001799

[12] Y. Zou, L. Yu, K. He, Forecasting crude oil risk: A multiscale bidirectional generative adversarial network based approach, Expert Systems with Applications 212 (2023) 118743. doi:https://doi.org/10.1016/j.eswa.2022.118743.
URL https://www.sciencedirect.com/science/article/pii/S0957417422017614

[13] K. He, L. Ji, C. W. D. Wu, K. F. G. Tso, Using sarima-cnn-lstm approach to forecast daily tourism demand, Journal of Hospitality and Tourism Management 49 (2021) 25–33. doi:https://doi.org/10.1016/j.jhtm.2021.08.022.
URL https://www.sciencedirect.com/science/article/pii/S1447677021001431

[14] Y. Zou, L. Yu, G. K. Tso, K. He, Risk forecasting in the crude oil market: A multiscale convolutional neural network approach, Physica A: Statistical Mechanics and its Applications 541 (2020) 123360. doi:https://doi.org/10.1016/j.physa.2019.123360.
URL https://www.sciencedirect.com/science/article/pii/S0378437119318795

[15] F.-C. Chen, M. R. Jahanshahi, NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naive Bayes Data Fusion, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS 65 (5) (2018) 4392–4400. doi:10.1109/TIE.2017.2764844.

[16] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, S. Zafeiriou, End-to-End Multimodal Emotion Recognition Using Deep Neural Networks, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 11 (8) (2017) 1301–1309. doi:10.1109/JSTSP.2017.2764438.

[17] T. Fang, C. Zheng, D. Wang, Forecasting the crude oil prices with an emd-isbm-fnn model, Energy 263 (2023) 125407. doi:https://doi.org/10.1016/j.energy.2022.125407.
URL https://www.sciencedirect.com/science/article/pii/S0360544222022897

[18] Q. Qin, Z. Huang, Z. Zhou, C. Chen, R. Liu, Crude oil price forecasting with machine learning and google search data: An accuracy comparison of single-model versus multiple-model, Engineering Applications of Artificial Intelligence 123 (2023) 106266. doi:https://doi.org/10.1016/j.engappai.2023.106266.
URL https://www.sciencedirect.com/science/article/pii/S0952197623004505

[19] S. Deng, Y. Zhu, S. Duan, Y. Yu, Z. Fu, J. Liu, X. Yang, Z. Liu, High-frequency forecasting of the crude oil futures price with multiple timeframe predictions fusion, Expert Systems with Applications 217 (2023) 119580. doi:https://doi.org/10.1016/j.eswa.2023.119580.
URL https://www.sciencedirect.com/science/article/pii/S0957417423000817

[20] J. Wu, J. Dong, Z. Wang, Y. Hu, W. Dou, A novel hybrid model based on deep learning and error correction for crude oil futures prices forecast, Resources Policy 83 (2023) 103602. doi:https://doi.org/10.1016/j.resourpol.2023.103602.
URL https://www.sciencedirect.com/science/article/pii/S0301420723003136

[21] Y. Lin, K. Chen, X. Zhang, B. Tan, Q. Lu, Forecasting crude oil futuresÂ prices using bilstm-attention-cnn model with wavelet transform, Applied Soft Computing 130 (2022) 109723. doi:https://doi.org/10.1016/j.asoc.2022.109723.
URL https://www.sciencedirect.com/science/article/pii/S1568494622007724

[22] B. Lim, S. Ã. ArÄ±k, N. Loeff, T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, International Journal of Forecasting 37 (4) (2021) 1748–1764. doi:https://doi.org/10.1016/j.ijforecast.2021.03.012.
URL `https://www.sciencedirect.com/science/article/pii/S0169207021000637`