Energy Economics 94 (2021) 105080

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

Neural network prediction of crude oil futures using B-splines

Sunil Butler [a], Piotr Kokoszka [a], Hong Miao [a,*], Han Lin Shang [b]

[a] Colorado State University, Fort Collins, CO, United States of America
[b] Macquarie University, Sydney, Australia

ARTICLE INFO

Article history:
Received 8 June 2020
Received in revised form 7 December 2020
Accepted 18 December 2020
Available online 28 December 2020

Keywords:
Crude oil futures
Term structure
Neural network
Splines
Functional data
Model confidence set

ABSTRACT

We propose two ways to improve the forecasting accuracy of a focused time-delay neural network (FTDNN) that forecasts the term structure of crude oil futures. Our results show that a convergence based FTDNN makes consistently more accurate predictions than the fixed-epoch FTDNN in Barunik and Malinska (2016). Further, we suggest using basis splines (B-splines), instead of Nelson-Siegel functions, to fit the term structure curves. The empirical results show that the B-spline expansions lead to consistently better 1 and 3 months ahead predictions compared to the convergence based FTDNN. We also explore conditions under which the B-spline based approach may be better for longer-term predictions.

1. Introduction

Crude oil is widely considered to be one of the most important commodities for the world's economy. As a result, academia and industry have paid particular attention to the accurate prediction of crude oil prices. However, forecasting the price of crude oil is challenging. Given there is no broadly accepted forecasting methodology, new approaches are being explored.

Forecasting the term structure of crude oil futures is an important branch of the crude oil price forecasting literature. Term structure of crude oil futures show predict the crude oil futures prices with different expiration, it also models the directions and relations between different future contracts. Accurately forecasting the term structure can help speculators to construct profitable crude oil future portfolios, hedges to determine maturities of hedging contracts, oil producers to optimize productivity, oil refineries and governments to decide levels of inventories. This paper considers several ways to improve a neural network approach that previous studies have have shown to outperform many other popular models in forecasting term structures of crude oil futures.

Due to the crude oil market's complexity, many attempts at modeling crude oil futures price movements have failed to consistently outperform basic time-series approaches, such as the random walk and AR (1) models. In general, crude oil price forecasting models fall into four broad categories: statistical, fundamental and/or economic factors, speculation based approaches, and, more recently, machine learning approaches.

A common econometric approach in empirical and theoretical work on the oil price is rooted in the seminal co-integrating framework of Engle and Granger (1987). Since then, there has been considerable academic effort literature has devoted to investigating and developing forecasting oil price levels (i.e., Pindyck, 1999; Radchenko, 2005). However, disappointingly, as documented by several studies (e.g., Hamilton, 2009; Alquist and Kilian, 2010; Alquist et al., 2013), none of these forecasting methods has been particularly successful when compared with the naive no-change forecast. These studies suggest that changes in the oil price are inherently difficult to predict and indicate that the current oil price might be the best forecast of the future price. Mathematically, this is consistent with a random walk model.

More recently, some studies show that combinations of statistical models perform better than simple approaches. Nademi and Nademi (2018) uses a semiparametric Markov switching AR-ARCH model to forecast the prices of OPEC, WTI, and Brent crude oils. The empirical results show the proposed model forecasts crude oil prices more accurately than ARIMA and GARCH models. Chen et al. (2018) proposes a hybrid grey wave forecasting model, which combines Random Walk (RW)/ARMA to forecast multi-step-ahead crude oil prices. The empirical results demonstrate that the model dominates ARMA and RW in terms of correct direction prediction. Chai et al. (2018) proposes a novel forecast combination approach that captures a variety of fluctuation features in crude oil data series, including change points, regime-switching, time-varying determinants, trend decomposition of high-frequency sequences, and the possible nonlinearity of model

* Corresponding author.
E-mail address: hong.miao@colostate.edu (H. Miao).

https://doi.org/10.1016/j.eneco.2020.105080

setting. de Albuquerquemello et al. (2018) shows that a Self-Exciting Threshold Auto-regressive model performs better than most of the oil price prediction methods.

Another stream of literature is predicated on the assumption that crude oil prices are driven by dynamic changes of a large set of factors. The influence of different factors tends to vary from relatively short-lived to long term effects extending over many years. A great deal of research has studied the predictive power of oil futures prices, oil inventories, oil production, macroeconomic fundamentals, product spreads, exchange rates, stock market dynamics, and economic policy uncertainty, among others (c.f., Andreasson et al., 2016; Baumeister and Kilian, 2012, 2015; Chatrath et al., 2016; Kilian, 2009; Basak and Pavlova, 2016; Singleton, 2014). Miao et al. (2017) considers six categories of factors and utilizes the Least Absolute Shrinkage and Selection Operator (LASSO) regression technique to improve the forecasting accuracy of crude oil prices. Zhang et al. (2018) uses an iterated combination approach to examine oil price prediction with 18 macroeconomic variables and 18 technical indicators. The model outperforms standard combination approaches for both in- and out-of-sample. In general, models incorporating factors outperform a random walk model in many situations.

More recent studies show some improvements have been made in the oil price models by expanding the range of explanatory factors and better integrating them (i.e., Dees et al., 2007; Ewing and Thompson, 2007; Kaufmann and Ullman, 2009). In addition, the role of speculation and forward-looking behavior in crude oil prices has been carefully examined (c.f. Kaufmann and Ullman, 2009, Sornette et al., 2009, Alquist and Kilian, 2010, Kaufmann, 2011, Kilian and Murphy, 2014, and Miao et al., 2018, among others). In general, forecasting models based on economic fundamentals work better at short horizons up to 3 months. In contrast, models based on the spread of refined product prices relative to the price of crude oil work better at longer horizons between 12 and 24 months (Baumeister and Kilian, 2015). Baumeister and Kilian (2015) presents a combination approach with six different models. The combination approach improves the forecasting performance in comparison to the no-change forecast.

Neural networks began to be seen as a legitimate approach to modeling crude oil futures in 2009 when Kulkarni and Haidar (2009) explored the general use of artificial neural networks (ANNs) for modeling crude oil futures. They present a model based on a multilayer feed-forward neural network to forecast crude oil spot price direction in the short-term, up to three days ahead. They suggest that a dynamic model of 13 lags is optimal to forecast spot price direction for the short-term. Since then, many competing neural network approaches have been developed and compared. Most notably, Hu et al. (2012) compare an Elman recurrent neural network (ERNN), a multilayer perceptron (MLP), and a recurrent fuzzy neural network (RFNN). They find that the RFNN model outperforms the other two neural networks in forecasting crude oil futures prices. Modifications to neural networks have also been explored; for example, the basic ANN was shown by Mahdiani and Khamehchi (2016) to benefit from the inclusion of a genetic matching algorithm in predicting the future prices of crude oil.

In the past several years, a significant number of combined models have been proposed in the neural network literature. Huang and Wang (2018) proposes a model that combines a wavelet neural network (WNN) with a random time effective function. The empirical results demonstrate that the proposed model has higher accuracy in predicting crude oil price fluctuations. Wang et al. (2020) demonstrates that a multi-granularity heterogeneous combination approach based on an artificial bee colony outperforms not only individual competitive benchmarks but also single-granularity heterogeneous and multi-granularity homogeneous approaches in forecasting crude oil prices. Ramyar and Kianfar (2019) develop a multilayer perceptron (MLP) neural network incorporating the impacts of monetary policy and other major drivers of crude oil prices. They conclude that the proposed MLP neural network can more accurately predict crude oil prices than a VAR model. Tang et al. (2018*b*) show randomized-algorithm-based decomposition-ensemble learning models are efficient and fast, relative to popular single models. Ding (2018) proposes a novel decompose-ensemble prediction process that combines ensemble empirical mode decomposition (EEMD) and an artificial neural network (ANN) and shows the EEMD-based model outperforms the empirical mode decomposition (EMD) model. Tang et al. (2018*a*) develop a non-iterative learning paradigm without an iterative training process to address the time-consuming nature and parameter sensitivity of the emerging decomposition ensemble models. Abdollahi (2020) finds that a hybrid model consisting of complete ensemble empirical mode decomposition, a support vector machine, particle swarm optimization, and Markov-switching generalized autoregressive conditional heteroskedasticity outperforms other models in forecasting crude oil prices.

A new approach may be superior for a specific time interval but later shown to be inferior for other time intervals. It is therefore important to consider many time intervals that incorporate various economic conditions and different volatility and trend levels in commodities and capital markets in evaluating various approaches. Barunik and Malinska (2016) shows that the Nelson-Siegel model, coupled with a focused time-delay neural network (FTDNN), outperforms several commonly used time-series approaches in modeling and predicting the term structure of crude oil futures. Their model predicts maturities curves that are first fit using a cubic spline interpolation followed by fitting Nelson-Siegel (Nelson and Siegel, 1987) curves to reduce the dimension of multivariate time series to be forecast. This study is the first study to employ neural networks in the forecasting of the term structure, forecasting crude oil futures prices with 24 different maturities and 1, 3, 6, and 12 months ahead. Thus, the forecasting approach is more thoroughly tested for different time intervals than most previous models, which only focus on forecasting one series.

In this paper, we improve the general approach of Barunik and Malinska (2016) in two ways. We propose replacing the three Nelson-Siegel curves, designed for forecasting bond yields, with a more flexible set of basis functions, which consists of B-splines. We propose a neural network convergence criterion for the resulting scalar time series of coefficients, which applies to each series individually. We demonstrate that the convergence based FTDNN forecasts the term structure more accurately than the fixed epoch FTDNN (Barunik and Malinska, 2016) in all periods for 1, 3, 6, and 12 months ahead. Furthermore, replacing the Nelson-Singel curves with B-spline leads to 1 and 3 months ahead of predictions compared to convergence based forecasts, which are themselves superior to the original approach of Barunik and Malinska (2016). We base our conclusions on applying both methods to many time intervals reflecting various economic and market conditions. We explore situations under which our modifications lead to superior forecasts over longer time horizons.

This paper thus contributes to the methodology for crude oil futures prediction based on neural networks. Studies in this field typically fall into two categories. They either compare the utilities of several different neural networks with unexplored general utility or compare a single neural network to basic time series models to demonstrate utility. This paper studies a neural network with established superiority over many other models and proposes effective improvements. In this way, it sheds light on one potential direction for moving the field forward by advancing the best approach among successful neural networks rather than adding to the already established viable models.

The remainder of the paper is organized as follows. In Section 2, we describe the data we use and previous prediction methods to which our approach is compared. After explaining the methods of comparing the various methods in Section 3.1, we present the details of our approach in Sections 3.2 and 3.3. Our findings are described and analyzed in Section 4, and summarized in Section 5.

2. Data and previous methodology

Before discussing the specifics of our neural network approach, it is important to describe the structure of the data we use, and the general framework of a functional data analysis (FDA) applied to model crude oil futures curves.

The original data set consists of continuous contracts of West Texas Intermediate (WTI) crude oil futures,[1] traded at the New York Mercantile Exchange (NYMEX), part of the Chicago Mercantile Exchange (CME) Group. The Quandl CHRIS continuous contracts data are at a daily frequency. For our purpose, we extract the last observations of each month in the data set. At any point, we only take the nearest 36 contracts (roughly up to 36 months to expiration). The data are from January 1990 to September 2019. The data start in 1990 because that is when the contract expiration expanded to 6 years, and thus, one can observe consecutive 24 months contracts. A sample of price curves over one year is shown in Fig. 1.

To be consistent with Barunik and Malinska (2016), we forecast crude oil prices 30, 60, 90, …, 720 days at monthly frequency. Thus at each month n, we have a curve X_n observed at maturities τ_j, measured in days. Consequently, the data have the form

$$X_n(\tau_j), \quad 1 \le n \le 357, \quad \tau_j = 30, 60, 90, \ldots, 720, \tag{2.1}$$

where n is the end of month date and τ_j is the time to maturity. Data sets with such structure have been studied in Functional Data Analysis (FDA). Kearney and Shang (2020) proposes a functional time series based method to model and forecast oil futures. The out-of-sample exercise provides strong support for the adoption of this approach, because of the superior set of models in all considered instances. Rather than points or vectors, statistical analysis objects are curves, or functions, as shown in Fig. 1. Curves evolve in time as whole observation units. Their levels and shapes evolve in a complex way. To reduce the dimension of the curves, a common approach is to approximate them as $X_n(\tau) \approx \sum_{m=1}^{M} \beta_{mn} v_m(\tau)$, with a relatively small M and suitable functions v_m. In a statistical analysis that follows, one works with a low dimensional vectors $\boldsymbol{\beta}_n = [\beta_{1n}, \ldots, \beta_{Mn}]^T$.[2]

The work of Barunik and Malinska (2016) falls into this general paradigm. For each month n, the curve X_n is fitted using the Nelson-Siegel model developed by Nelson and Siegel (1987). The Nelson-Siegel has been empirically shown to be effective for the modeling term structure of interest rates. In our context, the Nelson-Siegel model has the following form:

$$X_n(\tau) = \beta_{0n} + \beta_{1n}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{2n}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \varepsilon_n(\tau), \quad \tau \in \{30, 60, 90, \ldots, 720\}. \tag{2.2}$$

It thus corresponds to the expansion introduced above with $M = 3$ and the basis curves $v_m = v_{m,\lambda}$ equal to the three Nelson-Siegel factors ($v_1(\tau) = 1$). The value of λ and the coefficients $\beta_{0m}, \beta_{1m}, \beta_{2m}$ are estimated by a three-stage procedure described in Diebold and Li (2006).

The three-stage process obtains three scalar time series, $\{\beta_{0n}\}$, $\{\beta_{1n}\}$, and $\{\beta_{2n}\}$ or, equivalently, a multivariate time series $\left\{[\beta_{0n}, \beta_{1n}, \beta_{2n}]^{\mathrm{T}}\right\}$, all with the time index n. Various prediction methods can be applied to these series. The methods we study are summarized in Table 1. It describes how the one-step ahead predictions are obtained, that is, how to compute predictions of coefficients at month $n+1$ from those at month n and earlier months. Analogous, well-known, formulas exist for h-step ahead prediction. They can be found in many time series textbooks, for example Brockwell and Davis (2003).

[1] Available at https://www.quandl.com/data/CHRIS/CME_CL1-Crude-Oil-Futures-Continuous-Contract-1-CL1-Front-Month

[2] For a concise and accessible introduction to the field of FDA see Ramsay and Silverman (2005), and Kokoszka and Reimherr (2017).

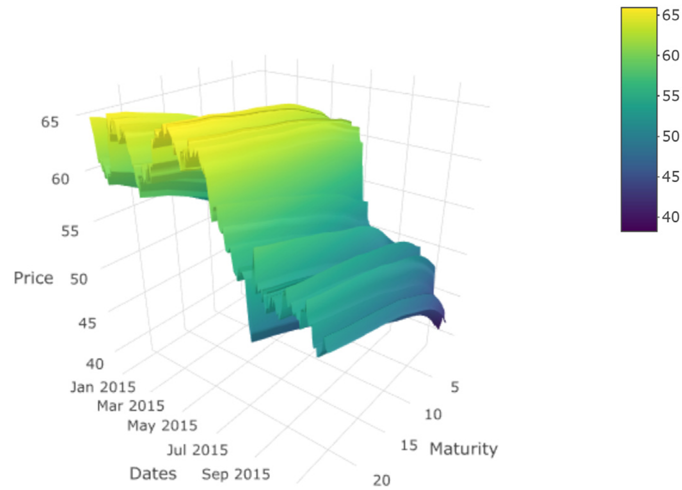

Fig. 1. A 3D plot of fitted daily oil futures curves from January 2 to December 31, 2015.

Barunik and Malinska (2016) propose using to use the Focused Time-Delay Neural Network (FTDNN) for forecasting of the coefficients of future term structure curves. FTDNN is a form of artificial neural networks (ANN). SeeBarunik and Malinska (2016) for a simple description of FTDNN. Following Barunik and Malinska (2016), we use one layer to more closely match the structure of the other models used for comparison.

3. Proposed methodology

3.1. Method of comparison

To compare forecasting approaches, we proceed as follows. For a sample period of Y_{total} years, we select an initial period of Y_{initial} years, ($12Y_{\text{initial}}$ months). We first predict the term curve in month $12Y_{\text{initial}} + 1$. Then we use the curves at months up to the month $12Y_{\text{initial}} + 1$ to predict the curve at month $12Y_{\text{initial}} + 2$. We can proceed in this manner until we predict the term curve at month $12Y_{total}$. For each maturity τ_j, we obtain $12(Y_{\text{total}} - Y_{\text{initial}})$ prediction errors. This allows us to compute the MAE (Mean Absolute Error) for one-month-ahead of predictions. We proceed analogously for 3, 6, and 12-month-ahead predictions. In our numerical evaluations, we set $Y_{initial} = 4$, i.e., the first training data to be 48 observations.

Table 1
Benchmark times-series prediction methods used for comparison with neural networks.

Method	Description
Random Walk	For each $m = 1,2,3$, the predicted value of $\beta_{m(n+1)}$ is equal to β_{mn}.
Random Walk with Drift	For each $m = 1,2,3$ the predicted value of $\beta_{m(n+1)}$ is equal to $\beta_{mn} + \widehat{\delta}_m$, where δ_m is the drift parameter δ_m in the regression $\beta_{mk} = \beta_{m(k-1)} + \delta_m + \varepsilon_{mk}$, $k = 1, 2, \ldots, n$.
AR(1)	For each $m = 1,2,3$, the predicted value of $\beta_{m(n+1)}$ is equal to $\widehat{\varphi}_m(\beta_{mn} - \widehat{\mu}_m) + \widehat{\mu}_m$, where $\widehat{\varphi}_m$ and $\widehat{\mu}_m$ are MLEs in the autoregressions, $\beta_{mk} - \mu_m = \varphi_m(\beta_{m(k-1)} - \mu_m) + \varepsilon_{mk}$, $k = 1, 2, \ldots, n$.
AR(3)	This method is analogous to the AR(1) method, but uses autoregressive models of order 3: $\beta_{mk} - \mu_m = \sum_{l=1}^{3}\varphi_{ml}(\beta_{m(k-l)} - \mu_m) + \varepsilon_{mk}$, $k = 1, 2, \ldots, n$.
VAR(1)	This method predicts all components of the vector $\boldsymbol{\beta}_n = [\beta_{0n}, \beta_{1n}, \beta_{2n}]^T$ based on vector autoregression $\boldsymbol{\beta}_k - \boldsymbol{\mu} = \Phi(\boldsymbol{\beta}_{k-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_k$, $k = 1, 2, \ldots, n$. The 3×3 matrix Φ is estimated by maximum likelihood, and all components of $\boldsymbol{\beta}_{n+1}$ are predicted simultaneously by $\widehat{\Phi}(\boldsymbol{\beta}_n - \widehat{\boldsymbol{\mu}}) + \widehat{\boldsymbol{\mu}}$.

Due to the inherent variability of crude oil prices, models that effectively identify relevant trends and patterns in the data may still be subject to some random variation. Therefore, two models with similar abilities to identify trends perform relative to each other randomly. Furthermore, a model that more effectively identifies trends in the data may be outperformed on a few measurements by another model purely due to random variation. Therefore, we further evaluate our model's performance using the model confidence set (MCS) algorithm developed by Hansen et al. (2011). The MCS selects the best model set from all the model candidates considered. It repeats the procedure of: 1) testing the null hypothesis that all the models have the same forecast errors; 2) eliminating the model with the highest MAE until the null hypothesis can not be rejected.

3.2. Improving the FTDNN by epoch selection

In this section, we propose two improvements to the FTDNN approach based on the Nelson-Siegel model (2.2). We do not change the functional model, nor its estimation, but change how the neural network runs.

Barunik and Malinska (2016) uses an FTDNN with fixed-epochs design with about 500 epochs and a batch size of 20. In their estimation, each coefficient is predicted with the same number of model-fitting iterations. This is disadvantageous from a computational perspective because coefficients converge at different rates depending on the coefficient and the nature of the data.

In general, to fit models, iteratively requires an inherent balancing act. While an underfitted model does not adequately interpret the training data, an overfitted model can begin to take on characteristics specific to the training data that do not appear in other data examples, such as particular manifestations of mostly random error. Models fitted with a fixed number of iterations can be either overfitted or underfitted depending on the size or volatility of the training data. Convergence-based models can also be overfitted or underfitted depending on the choice of convergence criteria. Assuming reasonable similarities in the data, however, the degree of fitness for convergence-based models is largely consistent. In other words, convergence criteria that are assessed to be reasonable for one training data set will likely be reasonable among all comparable training data sets. This standardization also mitigates errors introduced by judgment variation among researchers. It reduces total computation requirements, as models using fixed iterations must be tested and adjusted each new time the model is used.

We modify the Barunik and Malinska (2016) approach by calculating the loss at each iteration from the training data set. After sufficient experimentation, we determined that it is nearly optimal to stop the neural network iteration when the loss function drops below a threshold $\delta = 0.00001$. The batch size is set to be the number of months in a given data.

Consider a sequence of coefficients β_{mn}, where the month index n refers months, and $m = 0,1,2$. These are the series of coefficients estimated by the three-stage procedure described in Section 2. Suppose the initial time period consists of 36 months and we are interested in predictions h months ahead. To train the neural network, we use the values $\{\beta_{mn}, 1 \le n \le 36 - h\}$ as inputs, and the values $\{\beta_{mn}, h + 1 \le n \le 36\}$ as outputs. At each epoch e, the neural network constructs outputs $\{\beta^{\star}_{mn}(e),\ h + 1 \le n \le 36\}$. The quadratic loss function is computed as:

$$L(e) = \frac{1}{36-h}\sum_{n=h+1}^{36}\left(\beta^{\star}_{mn}(e) - \beta_{mn}\right)^2.$$

Iteration stops if $L(e) - L(e+1) < \delta$.

3.3. The B-spline method

Barunik and Malinska (2016) first approximate the data for each month with cubic splines (an ‘interpolation’ step) and then fit the Nelson-Siegel model, as explained in Section 2. We propose to fit the data with a small number of B-spline functions in one step without using the Nelson-Siegel model. Prautzsch et al. (2002) shows that B-splines can serve as basis functions for all other spline functions. Thus any fitted spline can be constructed from B-splines, and therefore using B-splines is as versatile an approach as possible in terms of fitted splines. The potential benefits of our approach are twofold: 1) by using one rather than two smoothing steps, our approach retains more the original structure of the data; 2) the B-spline basis is fairly universal, while the Nelson-Siegel model was explicitly developed for yield curves.

A B-spline basis is constructed by dividing the data interval into subintervals. In our case, this is the interval of maturities, that is [30,720], or 1 to 24 months. The points separating the subintervals are called *breaks*. Over each subinterval, the spline is a polynomial. The highest power occurring in this polynomial is called its *degree*. All polynomials on a B-spline basis have the same degree. The *order* of a B-spline basis is its *degree* plus one. At each breakpoint, neighboring polynomials are constrained to have a certain number of matching derivatives. The number of basis functions in a B-spline basis must satisfy the condition: the number of basis functions = order + number of interior breaks. Fig. 2 shows a B-spline basis consisting of four functions. There is one interior breakpoint, and each function coincides with a polynomial of degree 2 (a parabola) on the two disjoint intervals (0,12) and (12,24). At the breakpoint, the first derivatives coincide. Functions like those shown in Fig. 2 replace the three Nelson-Siegel factors in our approach.

Utilizing the “elbow method” from nonparametric statistics, we select five spline basis functions for all the periods except for the relatively stable period 2012–2018, four functions are used. This process is illustrated in Fig. 3. We compute the sum of squared errors (SSE) for each month for a fit resulting from using a specific number of basis functions, starting with two functions. The SSE decreases rapidly for the crude oil futures curves, in a very similar manner for all months. The “elbow method” suggests using the number of bases such that beyond which the SSEs form an almost flat pattern. An example of resulting fits is shown in Fig. 4.

The model then proceeds exactly as in Section 2, except that Eq. (2.2) is replaced by

$$X_n(\tau) = \sum_{k=1}^{K}\beta_{mn}B_m(\tau) + \varepsilon_n(\tau), \quad \tau \in \{30, 60, 90, \ldots, 720\}.$$

where the $B_1, B_2, \ldots, B_K$ are the functions shown in Fig. 2 for $K = 4$. In contrast to the Nelson-Siegel model, there is no need to find the parameter λ, since the B-spline functions do not depend on any parameters.

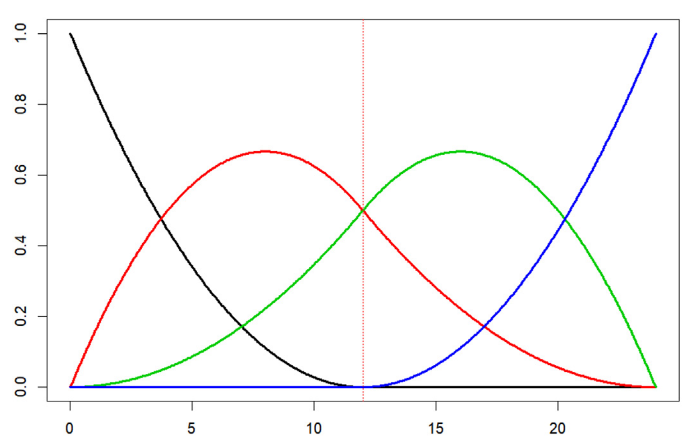

Fig. 2. Basis consisting of four B-splines of order four on the interval [0,24]. The first function is a parabola on (0,12) and is equal to zero on (12,24). For the last function, this is reversed.

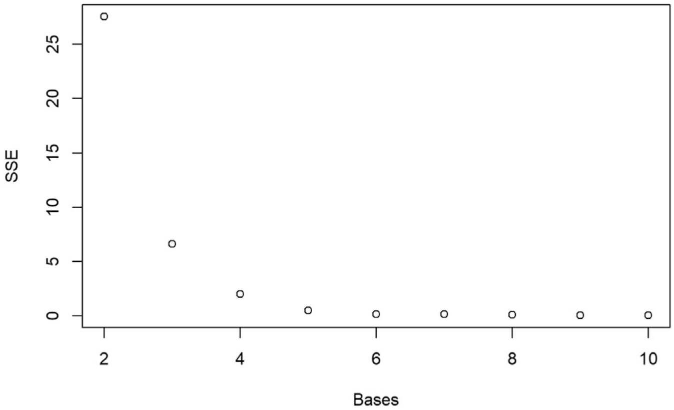

Fig. 3. The SSEs for after fitting B-spline bases with different numbers of functions for crude oil futures prices on January 2, 2015. In this case, four bases are used to fit the model.

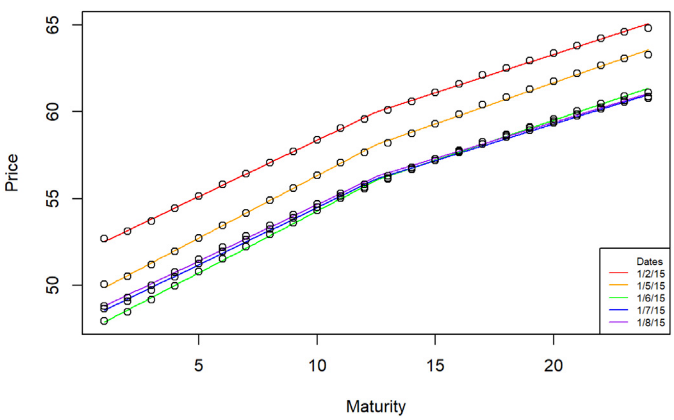

Fig. 4. B-spline curves with four coefficients fitted to crude oil futures prices between January 2 and January 8, 2015.

Once the coefficients β_{mn} are estimated, any of the prediction methods discussed in Section 2, can be used, not just the neural network approach.

Chantziara and Skiadopoulos (2008) employs Principal Components Analysis (PCA) to forecast the daily term structure of petroleum futures. The results show that the retained principal components have approximately similar forecasting powers to VAR models. The proposed B-spline basis functions use pre-determined basis functions to fit any function; as long as the number of B-spline basis functions is large enough, it can fit any function reasonably well. In contrast, PCA basis functions are purely data-driven. They aim to find a direction that maximizes the variance of the original data. PCA is designed to capture the main mode of variation in the data, and PCA basis functions are not designed to predict future functions (i.e., Kargin and Onatski, 2008). PCA basis functions are not designed to minimize the finite-sample prediction error, such as mean absolute error. Therefore, the proposed B-spline basis approach has advantages over PCA for our purpose.

4. Analysis of results

In this section, we compare the results obtained from various prediction methods on several different time intervals.

4.1. Convergence based FTDNN is superior to fixed FTDNN

We first compare the Barunik and Malinska (2016) fixed epoch neural network with the convergence criteria based FTDNN model. Although the returns on accuracy diminish significantly past 800 epochs, we use 1000 epochs on the fixed-epoch model. We do this to ensure that the fixed-epochs network typically runs longer than the convergence based network, so the convergence based implementation's superiority cannot be attributed to longer computation time.

To systematically investigate the methods, we analyze all seven-year-long intervals starting in September of each year from 1993 to 2012. We illustrate our findings by reporting the results for two intervals starting in September 2005 and 2012. Tables 2 and 3 show the relevant results for the periods starting in September 2005 and September 2012. All the results in Table 2 and 3 show MAE relative to the benchmark model AR(1). Table 2 shows that the convergence based FTDNN model outperforms the fixed epoch model on all 96 forecasts. For these seven years, FTDNN (conv) is the best model for 90 out of 96 cases, whereas AR(1) shows a more accelerated performance in six cases. The FTDNN (conv) model is always in the model confidence set. The results in Table 3 are even clearer. For the 2012 to 2019 period, the FTDNN (conv) model presents the most accurate prediction among the three candidate models in all 96 cases. Despite generally running for a shorter amount of time, the convergence implementation is consistently more accurate. Typically, the first coefficient runs for around 400 epochs; the second for 600, and the third for up to 1200. This pattern suggests that the fixed epoch model's inferiority could be

Table 2
MAE (Mean Absolute Error) and Model Confidence Set (MCS) for the interval between September 2005 and August 2012. This table reports the MAE and MCS for the focused time-delay neural network relative to (FTDNN) to the AR(1) model for the sub-period between September 2005 and August 2012. Models selected as a 'best' model by the model confidence set (MCS) for each prediction length and maturity are bold. The model with the best performance for each length of prediction and maturity is highlighted.

MAE	Time to maturity																							
	30	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480	510	540	570	600	630	660	690	720
FTDNN (fix)																								
1 M	1.00	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
3 M	**0.94**	**0.94**	**0.94**	**0.94**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**
6 M	0.73	0.73	0.74	0.74	0.75	0.76	0.76	0.76	0.77	0.77	0.77	0.77	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
12 M	**0.42**	**0.42**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.42**	**0.42**	**0.42**	**0.41**	**0.41**	**0.41**	**0.40**	**0.40**	**0.40**	**0.39**
AR (1)																								
1 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
3 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
FTDNN (conv)																								
1 M	**1.02**	**1.01**	**1.00**	**0.99**	**0.98**	**0.97**	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**	**0.97**	**0.97**	**0.98**	**0.98**	**0.98**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
3 M	**0.89**	**0.89**	**0.89**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.89**	**0.89**	**0.89**	**0.89**	**0.89**	**0.88**	**0.88**	**0.88**
6 M	**0.57**	**0.58**	**0.58**	**0.59**	**0.60**	**0.60**	**0.60**	**0.61**	**0.61**	**0.61**	**0.61**	**0.62**	**0.62**	**0.62**	**0.62**	**0.62**	**0.62**	**0.62**	**0.61**	**0.61**	**0.61**	**0.61**	**0.61**	**0.61**
12 M	**0.40**	**0.40**	**0.40**	**0.40**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.38**	**0.38**	**0.38**	**0.37**	**0.37**	**0.37**	**0.37**	**0.36**	**0.36**	**0.36**	**0.35**

Table 3
MAE (Mean Absolute Error) and Model Confidence Set (MCS) for the interval between September 2012 and August 2019. This table reports the MAE and MCS for the focused time-delay neural network relative to (FTDNN) to the AR(1) model for the sub-period between September 2012 and August 2019. Models selected as a ‘best’ model by the model confidence set (MCS) for each prediction length and maturity are bold. The model with the best performance for each length of prediction and maturity is highlighted.

MAE	Time to maturity																							
	30	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480	510	540	570	600	630	660	690	720
FTDNN (fix)																								
1 M	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**
3 M	**0.95**	**0.97**	**0.98**	**0.99**	**1.00**	**1.00**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.94**
6 M	**0.73**	**0.73**	**0.73**	**0.73**	**0.73**	**0.73**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.73**	**0.73**	**0.73**	**0.73**
12 M	**0.52**	**0.52**	**0.53**	**0.54**	**0.54**	**0.55**	**0.55**	**0.55**	**0.55**	**0.55**	**0.55**	**0.55**	**0.54**	**0.54**	**0.54**	**0.54**	**0.53**	**0.53**	**0.53**	**0.53**	**0.52**	**0.52**	**0.52**	**0.51**
AR(1)																								
1 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
3 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
6 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
FTDNN (conv)																								
1 M	**0.95**	**0.96**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.97**	**0.97**	**0.96**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.93**	**0.93**	**0.93**	**0.92**	**0.92**	**0.92**	**0.92**
3 M	**0.93**	**0.94**	**0.96**	**0.97**	**0.97**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.93**	**0.93**	**0.92**	**0.92**	**0.92**	**0.91**	**0.91**	**0.91**
6 M	**0.70**	**0.69**	**0.69**	**0.69**	**0.69**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.67**	**0.67**	**0.67**	**0.67**	**0.66**	**0.66**	**0.66**	**0.66**
12 M	**0.44**	**0.45**	**0.46**	**0.47**	**0.48**	**0.49**	**0.49**	**0.50**	**0.50**	**0.50**	**0.51**	**0.51**	**0.51**	**0.51**	**0.51**	**0.51**	**0.50**	**0.50**	**0.50**	**0.50**	**0.50**	**0.50**	**0.50**	**0.49**

due to insufficient convergence of the third coefficient and/or overfitting of the first two coefficients.

4.2. FTDNN with spline outperformance FTDNN with Nelson-Siegel

As expected, the proposed spline-based algorithm outperforms the Nelson-Siegel model. Representative results are reported in Tables 4 to 6. We used the following approach to generate the results in Tables 4 to 6. After running all candidate models, we select the best non-FTDNN method as a benchmark for comparing the convergence based FTDNN (n-s) method of Barunik and Malinska (2016) with to our FTDNN (spline) approach. The best traditional method is determined as the one with the most predictions contained in the Model Confidence Set. It is either AR(1) (n-s) or VAR(1) (n-s), meaning that either the components-wise AR(1) modeling or vector autoregression is applied to the Nelson-Siegel coefficients. Interestingly, in all cases, the AR (1) model outperforms the VAR(1) specification and is thus used as the benchmark model.

To enhance the information in our tables, we also included the spline versions of these methods, that is., AR(1) (spline). These are generally not competitive. It appears that using more coefficients, generally four vs. three, leads to less predictive accuracy when applying traditional time series predictors. Thus, our discussion focuses on comparing the FTDNN (n-s) method, improved by applying convergence criteria, to the FTDNN (spline) method with the epoch convergence criteria. In this way, we isolate the impacts of convergence criteria and the B-spline approach.

The results in Tables 4 to 6 show that the spline-based AR(1) model performs the best for 1-month-ahead forecast for 13 to 21 months for the sub-sample 2012 to 2019 in Table 5. For all other forecasting cases, the FTDNN based model outperforms the best traditional models.

In general, the results in the tables indicate that the FTDNN (spline) consistently outperforms FTDNN (n-s) for 1- and 3-month-ahead predictions. Furthermore, the B-spline FTDNN's error is lower by a similar margin across many different subsets of data, implying that the B-spline FTDNN is fundamentally and systematically better at prediction.

Table 4
MAE (Mean Absolute Error) and Model Confidence Set (MCS) for the interval between September 2005 and August 2012. This table reports the relative MAE for the focused time-delay neural network compared to the AR(1) model for dates between September 2005 and August 2012. Models selected as a ‘best’ model by the model confidence set (MCS) for each prediction length and maturity are bold. The model with the best performance for each length of prediction and maturity is highlighted.

MAE	Time to maturity																							
	30	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480	510	540	570	600	630	660	690	720
FTDNN (conv, n-s)																								
1 M	**1.00**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.02**	**1.02**	**1.02**	**1.02**	**1.02**	**1.02**	**1.02**	**1.02**
3 M	**0.94**	**0.94**	**0.94**	**0.94**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**
6 M	0.73	0.73	0.74	0.74	0.75	0.76	0.76	0.76	0.77	0.77	0.77	0.77	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
12 M	**0.42**	**0.42**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.43**	**0.42**	**0.42**	**0.42**	**0.41**	**0.41**	**0.41**	**0.40**	**0.40**	**0.40**	**0.39**
AR(1) (n-s)																								
1 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
3 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
6 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
FTDNN (conv, spline)																								
1 M	**1.00**	**1.00**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
3 M	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**	**0.91**	**0.91**	**0.91**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**
6 M	**0.62**	**0.63**	**0.63**	**0.64**	**0.65**	**0.65**	**0.65**	**0.65**	**0.65**	**0.65**	**0.65**	**0.65**	**0.65**	**0.65**	**0.64**	**0.63**	**0.63**	**0.62**	**0.62**	**0.61**	**0.60**	**0.58**	**0.57**	**0.55**
12 M	**0.49**	**0.49**	**0.49**	**0.49**	**0.49**	**0.49**	**0.49**	**0.49**	**0.48**	**0.47**	**0.46**	**0.46**	**0.45**	**0.44**	**0.43**	**0.43**	**0.42**	**0.42**	**0.42**	**0.42**	**0.41**	**0.40**	**0.40**	**0.39**
AR(1) (spline)																								
1 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**
3 M	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.95**	**0.95**	**0.95**	**0.95**	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.95**	**0.94**	**0.94**	**0.93**	**0.92**
6 M	1.17	1.18	1.18	1.19	1.19	1.19	1.19	1.18	1.17	1.16	1.15	1.14	1.14	1.13	1.12	1.12	1.11	1.07	0.99	0.89	0.78	0.65	0.56	0.55
12 M	1.22	1.21	1.21	1.22	1.24	1.24	1.25	1.25	1.24	1.23	1.22	1.21	1.21	1.20	1.20	1.19	1.18	1.14	1.06	0.94	0.79	0.62	0.46	0.39

Table 5
MAE (Mean Absolute Error) and Model Confidence Set (MCS) for the interval between September 2007 and August 2014. This table reports the relative MAE for the focused time-delay neural network compared to the AR(1) model for dates between September 2007 and August 2014. Models selected as a 'best' model by the model confidence set (MCS) for each prediction length and maturity are bold. The model with the best performance for each length of prediction and maturity is highlighted.

MAE	Time to maturity																							
	30	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480	510	540	570	600	630	660	690	720
FTDNN (conv,n-s)																								
1 M	**0.97**	**0.97**	**0.97**	**0.98**	**0.98**	**0.98**	**0.98**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.93**	**0.93**	**0.93**	**0.93**	**0.93**	**0.92**
3 M	**0.85**	**0.85**	**0.86**	**0.87**	**0.87**	**0.88**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.88**	**0.88**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.86**
6 M	0.71	0.69	0.68	0.67	0.65	0.65	0.64	0.64	0.64	0.65	0.66	0.67	0.68	0.69	0.70	0.70	0.71	0.73	0.74	0.75	0.76	0.77	0.78	0.79
12 M	**0.73**	**0.67**	**0.62**	**0.57**	**0.54**	**0.51**	**0.48**	**0.46**	**0.44**	**0.42**	**0.41**	**0.40**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.40**	**0.40**	**0.39**	**0.39**
AR(1) (n-s)																								
1 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
FTDNN (conv, spline)																								
1 M	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**	**0.94**	**0.93**	**0.93**	**0.93**	**0.92**	**0.92**	**0.92**	**0.92**	**0.92**	**0.93**	**0.93**	**0.93**	**0.92**	**0.92**	**0.92**	**0.91**	**0.91**	**0.90**	**0.89**
3 M	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.82**	**0.82**	**0.83**	**0.83**	**0.83**	**0.84**	**0.84**	**0.84**	**0.84**	**0.84**	**0.83**	**0.83**	**0.83**	**0.82**	**0.82**
6 M	**0.75**	**0.72**	**0.71**	**0.68**	**0.66**	**0.63**	**0.60**	**0.57**	**0.55**	**0.54**	**0.52**	**0.51**	**0.50**	**0.49**	**0.48**	**0.48**	**0.47**	**0.48**	**0.49**	**0.51**	**0.52**	**0.54**	**0.57**	**0.59**
12 M	**0.77**	**0.73**	**0.69**	**0.65**	**0.62**	**0.59**	**0.57**	**0.55**	**0.53**	**0.51**	**0.48**	**0.46**	**0.42**	**0.40**	**0.39**	**0.38**	**0.37**	**0.42**	**0.40**	**0.39**	**0.38**	**0.37**	**0.38**	**0.37**
AR(1) (spline)																								
1 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.00	1.00	1.00	0.99	0.98
3 M	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.02	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.00	0.99
6 M	1.34	1.33	1.32	1.31	1.31	1.31	1.30	1.29	1.28	1.27	1.25	1.24	1.22	1.20	1.18	1.17	1.16	1.15	1.14	1.14	1.13	1.13	1.13	1.13
12 M	1.80	1.77	1.74	1.72	1.72	1.72	1.71	1.70	1.68	1.65	1.63	1.60	1.57	1.55	1.52	1.50	1.48	1.47	1.45	1.44	1.42	1.40	1.39	1.38

Table 6
MAE (Mean Absolute Error) and Model Confidence Set (MCS) for the interval between September 2012 and August 2019. This table reports the relative MAE for the focused time-delay neural network compared to the AR(1) model for dates between September 2012 and August 2019. Models selected as a 'best' model by the model confidence set (MCS) for each prediction length and maturity are bold. The model with the best performance for each length of prediction and maturity is highlighted.

MAE	Time to maturity																							
	30	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480	510	540	570	600	630	660	690	720
FTDNN (conv, n-s)																								
1 M	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**	**1.01**
3 M	**0.95**	**0.97**	**0.98**	**0.99**	**1.00**	**1.00**	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.95**	**0.95**	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.94**
6 M	**0.70**	**0.69**	**0.69**	**0.69**	**0.69**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.68**	**0.67**	**0.67**	**0.67**	**0.67**	**0.66**	**0.66**	**0.66**	**0.66**
12 M	0.52	0.52	0.53	0.54	0.54	0.55	0.55	0.55	0.55	0.55	0.55	0.55	0.54	0.54	0.54	0.54	0.53	0.53	0.53	0.53	0.52	0.52	0.52	0.51
AR(1) (n-x)																								
1 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
3 M	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
6 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
12 M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
FTDNN (conv, spline)																								
1 M	**0.95**	**0.97**	**0.99**	**1.00**	**1.02**	**1.02**	**1.03**	**1.04**	**1.04**	**1.04**	**1.04**	**1.03**	**1.03**	**1.03**	**1.02**	**1.02**	**1.02**	**1.02**	**1.01**	**1.01**	**1.01**	**1.00**	**0.99**	**0.99**
3 M	**0.95**	**0.96**	**0.96**	**0.97**	**0.97**	**0.97**	**0.96**	**0.96**	**0.95**	**0.95**	**0.94**	**0.94**	**0.94**	**0.94**	**0.93**	**0.93**	**0.92**	**0.92**	**0.92**	**0.91**	**0.91**	**0.90**	**0.90**	**0.90**
6 M	**0.74**	**0.74**	**0.74**	**0.74**	**0.74**	**0.73**	**0.72**	**0.72**	**0.71**	**0.71**	**0.70**	**0.70**	**0.69**	**0.68**	**0.68**	**0.67**	**0.66**	**0.65**	**0.64**	**0.63**	**0.62**	**0.61**	**0.60**	**0.59**
12 M	**0.50**	**0.49**	**0.47**	**0.47**	**0.46**	**0.46**	**0.46**	**0.47**	**0.47**	**0.47**	**0.47**	**0.47**	**0.46**	**0.46**	**0.45**	**0.44**	**0.43**	**0.42**	**0.41**	**0.40**	**0.39**	**0.38**	**0.37**	**0.35**
AR(1) (spline)																								
1 M	**0.99**	**1.00**	**1.00**	**1.00**	**1.01**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**1.00**	**1.01**	**1.01**	**1.02**
3 M	**0.99**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**	**1.01**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**
6 M	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.96
12 M	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.97	0.97	0.96	0.96	0.95	0.94

For 6-month-ahead predictions, the FTDNN (spline) model shows better performance for all three sub-sample periods except for the 1-month maturity for the period 2017 to 2014 and 1-month to 14-month maturity cases for the period 2012 to 2019. That is of the 72 cases, the FTDNN (spline) model outperforms in 57 cases. For the 12-month forward prediction, the FTDNN (spline) and FTDNN (n-s) models have roughly the same accuracy. The FTDNN (n-s) model outperforms in all 24 cases (1-month to 2-year maturities) from 2005 to 2011. For the interval 2007 to 2014, the FTDNN (n-s) is better for 1- to 15-month maturities, and the FTDNN (spline) performs better for 16 to 24-month maturities. For the most recent 7-year interval, 2012 to 2019, the FTDNN (spline) model consistently outperforms the FTDNN (n-s) model for all maturities.

In summary, the FTDNN (spline) model shows more accuracy in forecasting crude oil time structure curves over the FTDNN (n-s) model.

There is some inconsistency regarding the relative performance of these models across different time frames. However, an extensive review of the data and its relationship to the models' performances yield some insight. To understand the discussion better, it is important to reflect on the structure of the observation $X_n(\tau_j)$, where n denotes month, and τ_j is maturity. The data can be volatile in many ways. If we fix τ_j, then there is volatility across n, which is considered temporal volatility. Furthermore, for different months n, the term curves could exhibit different volatility levels as functions of the maturities τ_j. In some months, the futures prices might change more with maturity than in other months. It is useful to keep these considerations in mind as we proceed with our discussion.

Figs. 5 and 6 show the plots of crude oil prices curves with dates between August 2007 and September 2014 purchased for 12 months out. In Fig. 5, the lines connect each price with the price 6 months later; in Fig. 6, the lines connect each price with the price 12 months later. Several details in Figs. 5 and 6 are noteworthy. First, the 12 month apart divisions seem to skip over several of the more dramatic shocks to the crude oil prices, particularly between 2007 and 2009. Even outside of these, there appears to be a less volatile relationship between prices 12 months apart than between prices 6 months apart. This is possibly due to some seasonality in the data or other general underlying structure that manifests once the prices have stabilized. We can measure the volatility of the changes in price by taking the slope of each of these connecting lines and taking the standard deviation of these slopes, effectively measuring the variability of the changes over a certain number of months. The volatility is higher for the 6-month-ahead predictions both on the whole subset and on subsets of the data divided into 2-year intervals. An opposite phenomenon is observed in prices between September 2012 and August 2019 shown in Figs. 7 and 8. Particularly due to the sharp drop between 2014 and 2015, we see a more dramatic average change between prices 12 months apart than between prices 6 months apart. Even outside of this drastic change, we see a greater average change between points 12 months apart. This may be due to the apparent positive drift of the data between 2016 and 2019. Across this subset, the slopes of the segments in Fig. 8 show a substantially higher volatility than the slopes of the segments in Fig. 7.

4.3. Further discussions

When we look back to the relative performance of the FTDNN (spline) and the FTDNN (n-s) models, it seems that the temporal volatility of the change in prices over intervals of 6 months and 12 months can reveal some reason for the difference in performance. The FTDNN (spline) model performs better when the prediction lead interval has higher volatility, whereas the FTDNN (n-s) model has superior performance if the prediction lead interval has lower volatility. Extensive testing of other chaotic elements of time series, such as the degree of non-stationarity or the frequency and magnitude of shocks, yields some information with respect to the relative performances of each model, but only this rule is consistent among all subsets of data we considered. This is intuitive and consistent with the features of the models. The Nelson-Siegel model was specifically designed to model interest rate structures. The interest rate structures are much less volatile than the term structure of crude oil futures. When the term structure is stable and less volatile, the FTDNN (n-s) model should perform better.

The consistently superior performance of the B-spline FTDNN for 1- and 3-month-ahead predictions is seemingly inconsistent with this observation, as one would expect 1- and 3-month predictions to offer smaller absolute changes than 6- and 12-month predictions. However, further investigation reveals some intuitive consistency behind this. Several explanations for the apparent relationship between the models' performances and the data's volatility present themselves. While these explanations can most accurately be described as informed speculation due to the 'black box' nature of neural networks, they are nonetheless compelling and may motivate further exploration. The most plausible explanation is that the degree of smoothing resulting from the Nelson-Siegel FTDNN, which results in high errors on the maturities curves, limits its ability to model volatile situations. Degrees of smoothing can be intuitively linked to the generality with which an underlying relationship is explored. In more stable situations where the crude oil prices may have some underlying structure, this could be an asset to the Nelson-Siegel FTDNN. In more volatile situations where there is less underlying structure and more chaos, a model that is more responsive to short-term trends and sharp changes, such as our B-spline model, could be preferable. This explanation is compelling because it also explains the superior performance of the B-spline FTDNN for 1- and 3-month-ahead predictions, as changes in price over short periods

Fig. 5. Crude oil prices between September 2007 and August 2014 connected to the price 6 months later.

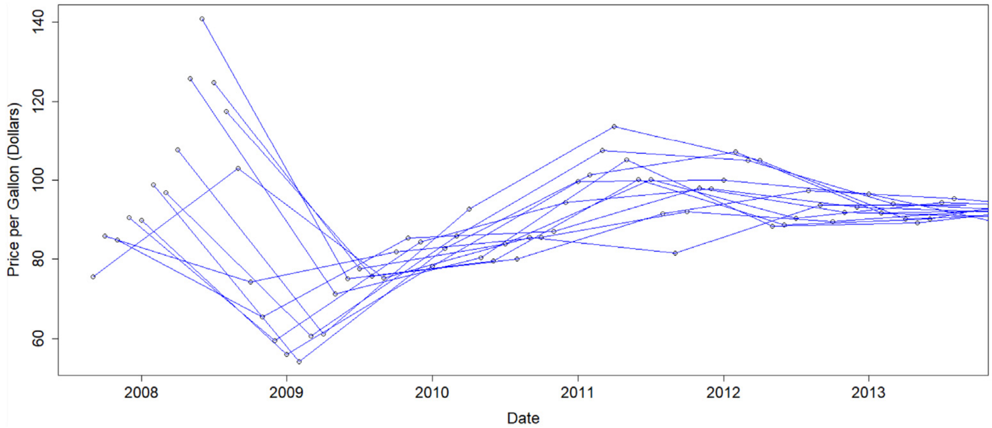

Fig. 6. Crude oil prices between September 2007 and August 2014 connected to the price 12 months later.

of time are less likely to be due to a long-term underlying evolution and more likely to result from short term patterns. By this logic, it is sensible that the B-spline FTDNN's lesser degree of smoothing and a greater number of coefficients would give it an advantage in predicting short term changes in the price.

Another possible explanation is that the convergence criteria, which were optimized over all lengths of predictions, should have been tailored for predictions each number of months ahead. Since the B-spline FTDNN consistently converges more quickly than the Nelson-Siegel FTDNN, the computation time required for the Nelson-Siegel FTDNN to properly fit more volatile data may be higher than suggested by initial tests. Also, more coefficients, each of which are concentrated only on a range of maturities, mean that the FTDNN (spline) approach more effectively customizes the convergence criteria for specific maturities. For example, the first B-spline function in Fig. 2 is not zero only for maturities between 1 and 12 months and is the largest for 1- to 5-month maturities. The second function emphasizes maturities between 5 and 12 months. By contrast, each Nelson-Siegel curve spans the whole range of maturities. We note that in modeling yields on corporate bonds, alternatives to the classical Nelson-Siegel model exist, for example Christensen et al. (2011) or Yallup (2012), but they all involve functions for which the domain is the whole range of maturities rather than localized ranges.

In summary, it is justified to state that the B-spline FTDNN is the superior approach for forecasting crude oil futures prices 1–3 months ahead. A heuristic assessment should be made for predictions further ahead to determine how volatile the prediction lead interval is expected to be. If oil prices are generally expected to be stable or continue to follow some observed consistent trend, then the convergence based Nelson-Siegel FTDNN should be used. If shocks or destabilizing elements are possible, or if prices are already in a volatile state, then the B-spline FTDNN should be used.

5. Conclusions

This paper proposes two directions for improving the forecasting accuracy of the model developed by Barunik and Malinska (2016) to forecast the term structure of crude oil futures. First, the paper suggests evaluating the convergence of each estimated coefficient individually to determine the amount of time (number of time periods) instead of the fixed term approach used in Barunik and Malinska (2016). We demonstrate that the convergence-based modification makes consistently more accurate predictions with similar or shorter computation time.

Second, we propose using basis splines (B-splines) instead of the three Nelson-Siegel functions to fit the term structure curves. The empirical results show that the B-spline curves are consistently better for 1- and 3-month-ahead predictions. We also explore circumstances under which the B-spline may be better for longer-term predictions.

In general, both FTDNN approaches are superior to common time series methods. For an FTDNN using the Nelson-Siegel model, running the model until our convergence criteria have been met is superior to fixing the number of terms in relation to both computation time and accuracy. Using a B-spline instead of the Nelson-Siegel model to fit maturities curves produces more accurate estimates for 1- and 3-month-ahead predictions. For 6- and 12-month-ahead predictions, the B-spline FTDNN handles volatility better. In contrast, the Nelson-Siegel FTDNN

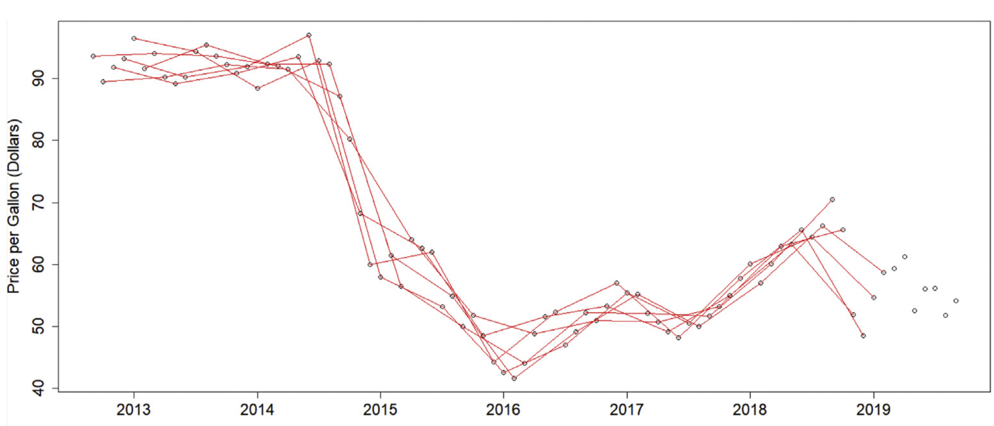

Fig. 7. Crude oil prices between September 2012 and August 2019 connected to the price 6 months later.

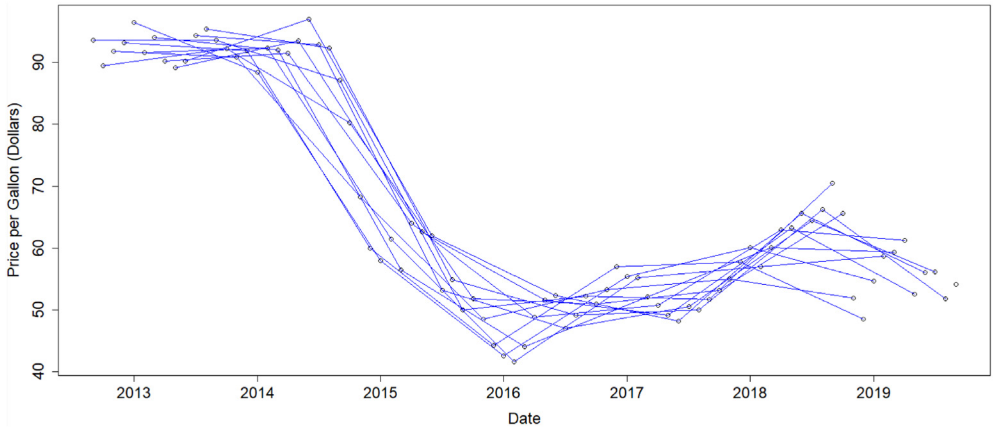

Fig. 8. Crude oil prices between September 2012 and August 2019 connected to the price 12 months later.

handles stability better, meaning that the choice of which model to use should be based on informed speculation regarding future crude price volatility.

Credit Author Statement

All the four authors contributed equally to this work.

Declaration of Competing Interest

All the four authors contributed equally to this work.

Acknowledgments

This research was partially supported by NSF grants DMS–1923142 and DMS–1914882 at Colorado State University. The authors would like to acknowledge insightful comments from two reviewers, which led to a much-improved manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2020.105080.

References

Abdollahi, H., 2020. A novel hybrid model for forecasting crude oil price based on time series decomposition. Appl. Energy 267, 115035.

Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? J. Appl. Econ. 25 (4), 539–573.

Alquist, R., Kilian, L., Vigfusson, R.J., 2013. Forecasting the price of oil. In: Elliott, G., Timmermann, A. (Eds.), “Handbook of Economic Forecasting”, Vol. 2, Elsevier B.V, pp. 427–507.

Andreasson, P., Bekiros, S., Nguyen, D.K., Uddin, G.S., 2016. Impact of speculation and economic uncertainty on commodity markets. Int. Rev. Financ. Anal. 43, 115–127.

Barunik, J., Malinska, B., 2016. Forecasting the term structure of crude oil futures prices with neural networks. Appl. Energy 164, 366–379.

Basak, S., Pavlova, A., 2016. A model of financialization of commodities. J. Financ. 71 (4), 1511–1556.

Baumeister, C., Kilian, L., 2012. Real-time forecasts of the real price of oil. J. Bus. Econ. Stat. 30 (2), 326–336.

Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: a forecast combination approach. J. Bus. Econ. Stat. 33 (3), 338–351.

Brockwell, P.J., Davis, R.A., 2003. Introduction to Time Series and Forecasting. Springer, New York.

Chai, J., Xing, L.M., Zhou, X.Y., Zhang, Z.G., Li, J.X., 2018. Forecasting the WTI crude oil price by a hybrid-refined method. Energy Econ. 71, 114–127.

Chantziara, T., Skiadopoulos, G., 2008. Can the dynamics of the term structure of petroleum futures be forecasted? Evidence from major markets. Energy Econ. 30 (3), 962–985.

Chatrath, A., Miao, H., Ramchander, S., Wang, T., 2016. An examination of the flow characteristics of crude oil: evidence from risk-neutral moments. Energy Econ. 54, 213–223.

Chen, Y., Zhang, C., He, K., Zheng, A., 2018. Multi-step-ahead crude oil price forecasting using a hybrid grey wave model. Physica A: Statistical Mechanics and its Applications 501, 98–110.

Christensen, J.H.E., Diebold, F.X., Rudebusch, G.D., 2011. The affine arbitrage–free class of Nelson-Siegel term structure models. J. Econ. 164, 4–20.

de Albuquerquemello, V.P., de Medeiros, R.K., da N’obrega Besarria, C., Maia, S.F., 2018. Forecasting crude oil price: Does exist an optimal econometric model? Energy 155, 578–591.

Dees, S., di Mauro, F., Pesaran, M.H., Smith, L.V., 2007. Exploring the international linkages of the euro area: a global VAR analysis. J. Appl. Econ. 22 (1), 1–38.

Diebold, F.X., Li, C., 2006. Forecasting the term structure of government bond yields. J. Econ. 130, 337–364.

Ding, Y., 2018. A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting. Energy 154, 328–336.

Engle, R.F., Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55 (2), 251–276.

Ewing, B.T., Thompson, M.A., 2007. Dynamic cyclical comovements of oil prices with industrial production, consumer prices, unemployment, and stock prices. Energy Policy 35 (11), 5535–5540.

Hamilton, J.D., 2009. Causes and consequences of the oil shock of 2007-08. Brook. Pap. Econ. Act. 40, 215–283.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hu, J.W.-S., Hu, Y.-C., Lin, R.R.-W., 2012. Applying neural networks to prices prediction of crude oil futures. Mathematical Problems in Engineering, p. ID959040.

Huang, L., Wang, J., 2018. Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. Energy 151, 875–888.

Kargin, V., Onatski, A., 2008. Curve forecasting by functional autoregression. J. Multivar. Anal. 99 (10), 2508–2526.

Kaufmann, R.K., 2011. The role of market fundamentals and speculation in recent price changes for crude oil. Energy Policy 39 (1), 105–115.

Kaufmann, R.K., Ullman, B., 2009. Oil prices, speculation, and fundamentals: interpreting causal relations among spot and futures prices. Energy Econ. 31 (4), 550–558.

Kearney, F., Shang, H.L., 2020. Uncovering predictability in the evolution of the WTI oil futures curve. Eur. Financ. Manag. 26 (1), 238–257.

Kilian, L., 2009. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. Am. Econ. Rev. 99 (3), 1053–1069.

Kilian, L., Murphy, D.P., 2014. The role of inventories and speculative trading in the global market for crude oil. J. Appl. Econ. 29 (3), 454–478.

Kokoszka, P., Reimherr, M., 2017. Introduction to Functional Data Analysis. CRC Press, Boca Raton.

Kulkarni, S., Haidar, I., 2009. Forecasting model for crude oil price using artificial neural networks and commodity futures prices. International Journal of Computer Science and Information Security 2, 1.

Mahdiani, M.R., Khamehchi, E., 2016. A modified neural network model for predicting the crude oil price. Intellectual Economics 10, 71–77.

Miao, H., Ramchander, S., Wang, T., Yang, D., 2017. Influential factors in crude oil price forecasting. Energy Econ. 68, 77–88.

Miao, H., Ramchander, S., Wang, T., Yang, J., 2018. The impact of crude oil inventory announcements on prices: evidence from derivatives markets. J. Futur. Mark. 38 (1), 38–65.

Nademi, A., Nademi, Y., 2018. Forecasting crude oil prices by a semiparametric Markov switching model: OPEC, WTI, and Brent cases. Energy Econ. 74, 757–766.

Nelson, C.R., Siegel, A.F., 1987. Parsimonious modeling of yield curves. J. Bus. 60 (4), 473–489.

Pindyck, R.S., 1999. The long-run evolutions of energy prices. Energy Journal 20 (2), 1–27.

Prautzsch, H., Boehm, W., Paluszny, M., 2002. Bezier and B-Spline Techniques. Springer, Berlin Heidelberg.

Radchenko, S., 2005. Oil price volatility and the asymmetric response of gasoline prices to oil price increases and decreases. Energy Econ. 27 (5), 708–730.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis. Springer, New York.

Ramyar, S., Kianfar, F., 2019. Forecasting crude oil prices: a comparison between artificial neural networks and vector autoregressive models. Comput. Econ. 53 (2), 743–761.

Singleton, K., 2014. Investor flows and the 2008 boom/bust in oil prices. Manag. Sci. 60 (2), 300–318.

Sornette, D., Woodard, R., Zhou, W.X., 2009. The 2006–2008 oil bubble: evidence of speculation, and prediction. Physica A: Statistical Mechanics and its Applications 388 (8), 1571–1576.

Tang, L., Wu, Y., Yu, L., 2018a. A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting. Appl. Soft Comput. 70, 1097–1108.

Tang, L., Wu, Y., Yu, L., 2018b. A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting. Energy 157, 526–538.

Wang, J., Zhou, H., Hong, T., Li, X., andWang, S., 2020. A multi-granularity heterogeneous combination approach to crude oil price forecasting. Energy Econ. 91, 104790.

Yallup, P.J., 2012. Models of the yield curve and the curvature of the implied forward rate function. J. Bank. Financ. 36, 121–135.

Zhang, Y., Ma, F., Shi, B., Huang, D., 2018. Forecasting the prices of crude oil: an iterated combination approach. Energy Econ. 70, 472–483.