Applied Energy 164 (2016) 366–379

Contents lists available at ScienceDirect

Applied Energy

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

Forecasting the term structure of crude oil futures prices with neural networks ☆

Jozef Baruník [a,b,*], Barbora Malinská [a]

[a] Institute of Economic Studies, Charles University, Opletalova 26, 110 00 Prague, Czech Republic
[b] Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodarenskou Vezi 4, 182 00 Prague, Czech Republic

HIGHLIGHTS

- We analyse term structure of crude oil markets.
- New model for forecasting based on neural networks is proposed.
- We show that even basic architecture of neural models performs very well against benchmarking models.

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 23 November 2015
Accepted 28 November 2015
Available online 22 December 2015

Keywords:
Term structure
Nelson–Siegel model
Dynamic neural networks
Crude oil futures

ABSTRACT

The paper contributes to the limited literature modelling the term structure of crude oil markets. We explain the term structure of crude oil prices using the dynamic Nelson–Siegel model and propose to forecast oil prices using a generalized regression framework based on neural networks. The newly proposed framework is empirically tested on 24 years of crude oil futures prices covering several important recessions and crisis periods. We find 1-month-, 3-month-, 6-month- and 12-month-ahead forecasts obtained from a focused time-delay neural network to be significantly more accurate than forecasts from other benchmark models. The proposed forecasting strategy produces the lowest errors across all times to maturity.

1. Introduction

Modelling and forecasting the term structures of commodity markets is attractive from an academic perspective and valuable for producers, speculators, and risk managers. Generally, the term structure illustrates expectations of the future development of the corresponding market. Notwithstanding the importance of the subject, there are almost no relevant studies forecasting commodity term structures. In this paper, we introduce a novel framework for forecasting the term structure of crude oil futures prices. We propose to couple dynamic neural networks with the Nelson–Siegel model to obtain precise forecasts of crude oil futures prices.

Crude oil is essential to the world's economies from an industrial perspective because it is a vital production input and its price is driven by distinct demand and supply shocks. Shifts in the price of oil are, to varying extents, driven by aggregate or precautionary demand related to market anxieties concerning the availability of future oil supplies. As the demand for crude oil, which is not as dependent on price as it is on income [1], continues to rise and supply is likely to decline (because crude oil is a limited resource), the literature agrees that the future development of crude oil prices will be highly volatile and hence uncertain future [2]. The main reasons that the crude oil market is one of the most volatile in the world are the growing demand for a supply that is highly dependent on the behaviour of politically and economically unstable countries, crude oil demand and production are highly correlated with the occurrence of exogenous events such as military conflicts and natural catastrophes, and the presence of speculators [3].

As the crude oil futures market is one of the most developed markets based on trading volume, understanding the behaviour of its term structure is crucial. However, studies modelling and forecasting the term structure of petroleum markets are scarce

☆ Support from the Czech Science Foundation under the P402/12/G097 DYME – "Dynamic Models in Economics" project is gratefully acknowledged. The research leading to these results received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007–2013/ under REA grant agreement number 609642.

* Corresponding author at: Institute of Economic Studies, Charles University, Opletalova 26, 110 00 Prague, Czech Republic.
E-mail address: barunik@utia.cas.cz (J. Baruník).

http://dx.doi.org/10.1016/j.apenergy.2015.11.051
0306-2619/

(see Lautier [4] for review), and most of researchers focus on directly forecasting speculative prices [5] or researching the market's efficiency [6–8]. Similar to interest rate models, there are two approaches to modelling the term structure of petroleum commodities. As a natural candidate for the state variable in a one-factor model, the spot price is modelled as geometric Brownian motion [9] or a mean-reverting process [10]. Subsequent contributions consider the convenience yield as a second state variable in a two-factor model [10]. Alternatively, Gabillon [11] employs the long-term price as the second state variable. While both of these approaches assume a constant interest rate, which implies that future spot prices and forward prices are the same, Cortazar and Schwartz [12] proposes a three-factor model.

A relatively series of contributions to the literature explaining commodity futures prices uses the approach of Diebold and Li [13], originally applied to model yield curves. Motivated by similarities in the stylized facts between commodity markets and interest rate markets, the dynamic Nelson–Siegel model is a natural candidate for this task. Among the relatively limited number of contributions on the subject, Karstanje et al. [14] examine the co-movement of factors driving commodity futures curves and their shapes by adopting the framework of the dynamic Nelson–Siegel model [13]. To study the joint dynamics of the factors driving commodity futures curves, Nomikos and Pouliasis [15] uses a multiple-regime framework. Almansour [16] model the futures term structure of crude oil and natural gas markets with switching regimes, and Heidorn et al. [17] regress futures curve factors extracted from the dynamic Nelson–Siegel model on fundamental and financial traders. While the dynamic Nelson–Siegel model explains the dynamics of factors underlying the term structure of commodity prices, the extant literature offers no suggestions for predicting future price developments, with the only exception being Grønborg and Lunde [18]. In their original work, Diebold and Li [13] propose the use of a simple autoregressive time series model to successfully forecast the dynamics of term structure factors and, hence, prices in the interest rate market. We hypothesize that factors in commodity markets may contain further nonlinear dependencies, which need to be modelled to obtain precise forecasts. Therefore, it would be sensible to apply more general methods that do not require restrictive assumptions concerning the underlying structure of factors.

A natural candidate for the forecasting task are neural networks, which can be understood as a generalized non-linear regression tool. Concisely, neural networks are semi-parametric non-linear models, which are able to approximate any reasonable function [19,20]. While the number of models using machine learning is growing rapidly in the academic literature, applications in energy markets are rather limited. Among the few works from recent increase in contributions to the literature, neural networks are applied to predict fuel consumption [21] and day-ahead electricity prices [22], model energy demand in the residential sector in the United States [23], or quantify patterns in the co-movement between futures and spot prices [24]. Several works use neural networks to forecast energy prices [25–32]. Contributing to this strand of the literature, we are the first to employ this approach for forecasting term structures.

The contribution of this work is twofold. First, we contribute to the rare literature studying the term structure of commodity prices by providing new results from the application of the dynamic Nelson–Siegel modelling strategy to crude oil futures markets for long period 1990–2014. Second, we propose the use of a time-delay neural network to forecast the term structure factors identified by the dynamic Nelson–Siegel model. Using this framework, we forecast the term structure of crude oil futures prices successfully over the 1-month, 3-month, 6-month and 12-month forecasting horizons.

2. Data

2.1. Raw data

The dataset consists of monthly closing prices of West Texas Intermediate (WTI) futures contracts,[1] traded on the New York Mercantile Exchange (NYMEX). Each contract expires three trading days prior the 25th calendar day of the month preceding the month of delivery.[2] In total, we analyse 396 monthly historical (already delivered) and to-date undelivered contracts – 12 contracts per year with delivery months in the period beginning in 1990. The undelivered contracts included in the dataset are contracts stipulating delivery in November and December 2014 and 24 contracts with delivery in the two subsequent years (2015 and 2016).

The main reason for using data beginning in 1990 is that the maximum time to maturity for contracts before this date was nine months, while later during the period considered, this duration increased to more than six years. Hence to avoid potentially large risk and inaccuracies stemming from data extrapolation, we consider only data after the year 1990. The choice of the monthly frequency is primarily driven by the fact that contracts with longer times to maturity were traded rather infrequently in the first half of the studied period. In addition, Baumeister et al. [33] find monthly data to have equal predictive power to that of daily data.

Table 1 presents an example of actual data to illustrate the structure and dimension of the dataset. To associate each futures price observation with its corresponding time to maturity, it is necessary to first determine the exact expiry date of each contract. Then, the difference between the expiry and observation dates provides us with the remaining days to maturity. Table 1 captures the end-of-month futures prices of three different (in this case consecutive) contracts with delivery in August, September and October 2003. For example, at the end of February 2001, CLQ2003 and CLU2003 contracts were traded. On February 28, 2001, it was possible to enter into a contract with delivery in August 2003 at a futures price of USD 21.72 per barrel. The time to maturity in this case (τ) was 625 trading days.

2.2. Reorganized data

After combining the days to maturity with each observed futures price quotation, the dataset should be formed into a matrix with a number of rows equal to number of days included in analysis and a number of columns equal to number of analysed maturities.

Time series captured in Table 2 are reorganized into constant-maturity futures prices. WTI crude oil futures are delivered and expire with one-month regularity; therefore, futures prices with exactly 30, 60, or 90 days to maturity are not traded every day. The literature suggests several approaches to interpolating the prices to obtain the desired data format. Diebold and Li [13] use linear interpolation for constant maturity, while Holton [34] prefers cubic splines interpolation.[3] In our work, we follow the approach of Holton [34] and use cubic spline interpolation. Fig. 1 illustrates the reorganized constant-maturity futures prices we employ, plotted against the daily evolution of the spot price.

Due to the long period we consider, which includes several turbulent periods, we present the term structure in separate periods to better highlight the rich dynamics (Fig. 2).

[1] Available at https://www.quandl.com/c/futures/cme-wti-crude-oil-futures.

[2] Full specification of WTI futures contracts is available at http://www.cmegroup.com/trading/energy/files/en-153_wti_brochure_sr.pdf.

[3] For a detailed discussion of the interpolation methods for curve construction with application to yield curve modelling, see Hagan and West [35].

Table 1
Example of futures prices and corresponding maturities for contracts traded between February 28, 2001 and May 31, 2001 for different contracts. CME product code *CL* is used for a WTI futures contract, and the letters *Q*, *U*, and *V* denote delivery in August, September and October, respectively.

Contract date	CLQ2003		CLU2003		CLV2003	
	Settle	τ	Settle	τ	Settle	τ
February 28, 2001	21.72	625	21.62	646	–	–
March 31, 2001	22.76	602	22.70	623	–	–
April 30, 2001	23.46	582	23.35	603	–	–
May 31, 2001	23.57	559	23.45	580	23.33	603

Table 2
Example of a dataset reorganized to exhibit constant time to maturity.

Date	Days to maturity (τ)							
	30	60	90	120	150	180	210	...
February 28, 2001	27.48	27.36	26.99	26.60	26.21	25.84	25.48	...
March 31, 2001	26.50	26.59	26.43	26.20	25.94	25.68	25.43	...
April 30, 2001	28.74	28.89	28.53	28.07	27.60	27.19	26.78	...
May 31, 2001	28.49	28.42	28.14	27.78	27.38	27.00	26.59	...

Fig. 2(a) illustrates the term structure dynamics in the period from 1990 to 2004. Prices are relatively steady with slight downward shift during the Asian crisis of the 1990s with a dramatic change after the year 2000 due to the energy crisis. Some authors attribute the increase in futures prices to speculators and sudden declines in oil reserves, whereas others disprove these arguments [36,37]. The subsequent period begins with steady dynamics and reflects a reasonable increase in futures prices across all maturities beginning in 2005 (Fig. 2(b)). The calm period is interrupted by turbulent period circa 2008, when crude oil prices exceeded USD 100 per barrel.

Fig. 2(c) provides a more detailed illustration of the rich dynamics during the period of analysis. Military conflicts in Nigeria (including attacks on oil pipelines) and tensions between Iran and Israel and the consequent fear of an oil crisis accelerated the increase in oil prices to unprecedented levels. Political unrest in the Middle East combined with a sharp depreciation of the U.S. Dollar resulted in further frequent and significant horizontal shifts in the term structure. The global financial crisis drove the WTI term structure back under USD 100 per barrel, and the data exhibit a horizontal shift upwards at the end of 2009, driven by the complicated political environment in the Middle East – conflicts in the Gaza Strip.

Increasing, decreasing and humped shapes in the term structure can be observed during the most recent five-year period, as illustrated in Fig. 2(d). The WTI term structure experienced a strong upward horizontal shift during 2011 caused by political unrest in Egypt and Libya combined with the weak U.S. Dollar. Another steep upward shift in 2012 had also has a political explanation – the danger that Iran would close the Strait of Hormuz in response to sanctions imposed on the country because of its nuclear programme.[4] Finally, the Greek bailout and a Chinese economy stimulated by an increased money supply contributed to rise of crude oil prices.

2.3. Stylized facts regarding the term structure

The above discussion reveals that the crude oil futures term structure takes many shapes, which are essentially similar to the yield curves of government bonds, although the data are fundamentally different. The similarities between the two are discussed in detail by Grønborg and Lunde [18], who compare five stylized facts concerning government bond yield curves [13] to the stylized facts regarding crude oil term structures. The discussion is important because we rely on the dynamic Nelson–Siegel approach [13] to model the term structure.

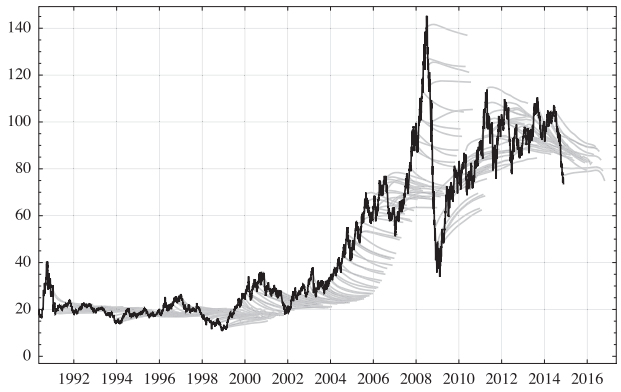

Fig. 1. Reorganized dataset: monthly term structures of crude oil futures prices plotted against daily spot prices for the period 1990–2014.

The main stylized facts concerning the yield curves are: (1) on average, the yield curve is increasing in time to maturity and concave; (2) it exhibits various shapes over time – upward or downward sloping, humped, and inverted humped; (3) the "near" end of the yield curve is much more volatile than the "far" end; (4) yield dynamics are persistent, while spread dynamics are much less persistent; and (5) long rates are more persistent than short rates.

Moreover, the term structure of crude oil is vulnerable to political decisions and conflicts, and hence its shape often changes not only in terms of horizontal shifts but also in actual shape. To document its ability to exhibit a wide variety of shapes, we borrow Fig. 6 from Section 3.1.2, documenting four days with different shapes of the analysed curve as illustrative examples. At the end of November 1990, we observe a smoothly decreasing term structure (Fig. 6(a)). In May 1999, the curve does not exhibit any smoothness, and its behaviour is unclear. Fig. 6(c) depicts a clear and increasing curve, and the most recent example (Fig. 6(d)) reveals that the data can exhibit humped curves.

Probably the most specific feature of crude oil future markets is backwardation.[5] Hotelling [38] postulates that the equilibrium price of non-renewable resources such as crude oil, which is equal to net marginal revenue, increases over time at the interest rate. However, the key factor distinguishing Hotelling's theory from theories of backwardation on crude oil markets is uncertainty [39]. As Haubrich et al. [40] argues, the market should exhibit the opposite situation – contango. Crude oil futures prices should exceed crude oil spot prices, as the opportunity cost is equal to the interest rate and storage costs make crude oil stocks disadvantageous. The convenience yield justifies the presence of backwardation in commodity markets. Storing a commodity implies not only costs but also benefits. The convenience yield can be understood as the "*. . .flow of services that accrues to an owner of the physical commodity but not to an owner of a contract for future delivery of the commodity* . . ." [9]. The discounted marginal convenience yields the present value, thus entailing equal backwardation appearing on the market, implying exogenously determined backwardation. One can introduce oil production as a call option to make it endogenous [39]. Lautier [4] proposes an alternative explanation by making an analogy between

[4] Approximately 20% of the crude oil traded worldwide passes through the Strait according to the U.S. Energy Information Administration; see http://www.eia.gov/countries/analysisbriefs/World_Oil_Transit_Chokepoints/wotc.pdf.

[5] Backwardation is a situation in which future prices are lower than spot prices.

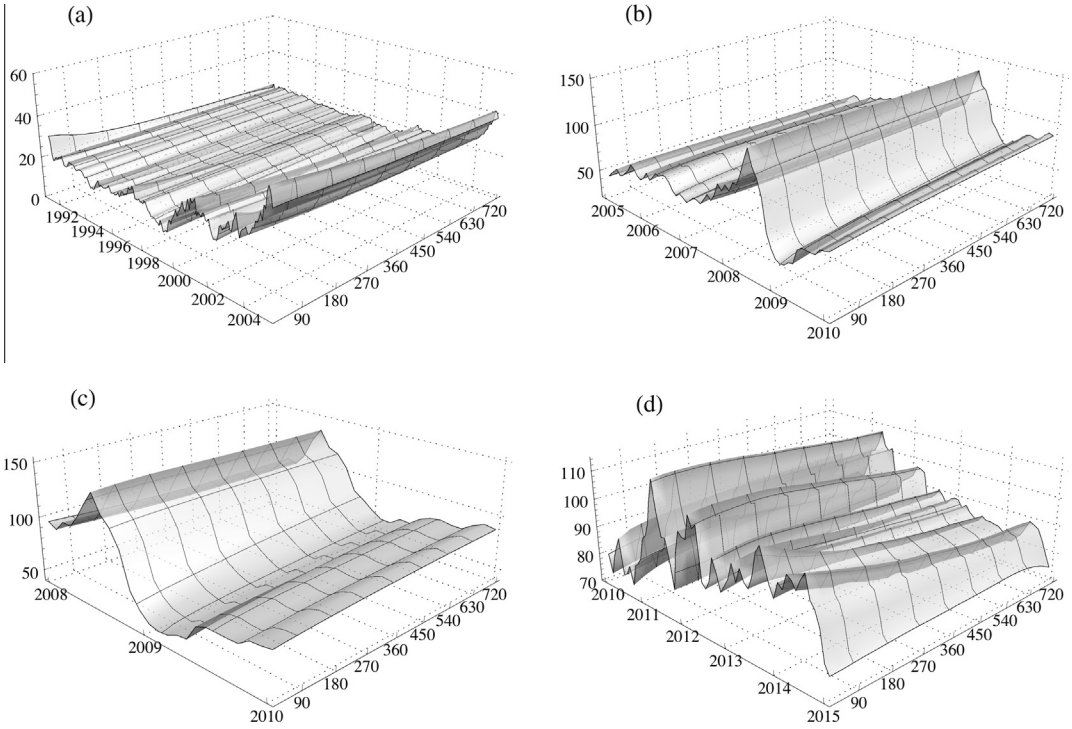

Fig. 2. Term structure of WTI futures prices for the period (a) 1990–2004, (b) 2005–2009, (c) 2008–2009, and (d) 2010–2014. Dates, days to maturity, and futures prices are on the corresponding $\{x, y, z\}$ axes.

the convenience yield and coupons or dividends linked to bonds and stocks, respectively.

3. Modelling the term structure

As motivated by the previous analysis, the crude oil term structure is similar to fixed income securities, and hence a common modelling approach can employed for both markets.[6] The most successful approach used in the recent literature to model and forecast yield curves is that of Diebold and Li [13]. This model is a dynamic representation of the Nelson–Siegel model [41], and in a recent contribution, is successfully applied to crude oil markets by Grønborg and Lunde [18]. In contrast to affine general equilibrium models, which assume a concrete functional relationship for the yield curve, this class of models does not stem from any theoretical assumptions and is based solely on the parameterization of curve shapes. Generally, curve-fitting models using standard statistical methods perform better in curve fitting and forecasting relative to affine models [42].

3.1. Dynamic Nelson–Siegel model

To model the term structure of crude oil futures prices, we use the dynamic Nelson–Siegel model [13]. The adoption of this framework is motivated by several considerations. First, other classes of models, such as no-arbitrage or affine general equilibrium models, fail when applied to forecasting. As Sarker et al. [43] notes, no-arbitrage models focus on a cross-sectional fitting of the yield curve at particular point in time, which implies that the model fails to capture yield curve dynamics. Affine models capture time-series dynamics but omit proper cross-sectional fit at a given time. Second, the functional specification of yield curves provided by Nelson and Siegel [41] is able to model the diverse shapes observable in extant markets. Third, the model provides intuitive parameters, which are straightforward to explain and interpret. Further, Bliss [44] demonstrates that the Nelson–Siegel model outperforms other methods in yield curve estimation, and Diebold and Li [13] show that the Nelson–Siegel model is able to replicate stylized facts regarding yield curves. Conversely, Duffle and Kan [45] concludes that yield curves estimated by affine general equilibrium models, such as Vasicek or Cox-Ingersoll-Ross mode, do not conform to the behaviour observed in markets.

Diebold and Li [13] propose forecasting the yield curve using the time series of three yield curve components formulated in the Nelson–Siegel model. In this framework, the dynamics of the term structure of crude oil futures prices are described by the following equation:

$$p_t(\tau) = \beta_{0t} + \beta_{1t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \quad (1)$$

where $p_t(\tau)$ is the price of crude oil futures at time $t = 1, \ldots, T$ with time to maturity $\tau = 30, 60, 90, \ldots, 720$, and β_{0t}, β_{1t}, and β_{2t} are interpreted as the coefficients on the level, slope and curvature factors, respectively. The level factor is a long-term component, as the values of the factor are constant over the entire period and at all maturities. The slope factor is a short-term component, as long as it decays exponentially at rate λ_t. Finally, the curvature factor is regarded as a medium-term component, as it increases for medium-term maturities and then decays over the longest maturities.

Fig. 3 presents estimated loadings of the factors as a function of time to maturity. The plot uses a fixed decay $\lambda_t = \lambda = 0.0058$ that is empirically determined in the next section.

The level factor on β_{0t} is constant for all the maturities and, hence, impacts equally futures prices at all maturities. A change in the level factor entails a horizontal shift in the term structure

[6] There are simplifying assumptions for the crude oil term structure models – there are no frictions, taxes, or transaction costs on the market, trading is continuous, lending and borrowing rates are equal, short sales are unconstrained and markets are complete [4].

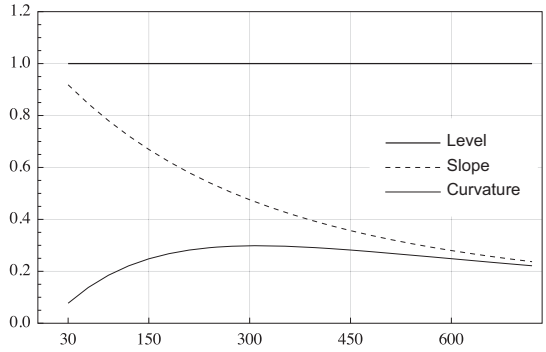

Fig. 3. Loadings of Nelson–Siegel latent factors of the term structure.

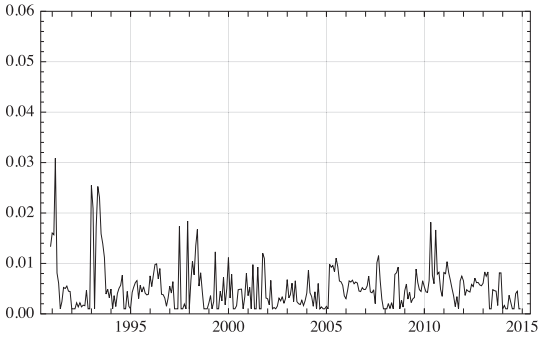

Fig. 4. Time series of λ_t.

and thus will affect prices at all maturities in the same way. The loading on the slope factor is decreasing from one (zero time to maturity) to zero as maturity goes to infinity. Note that Fig. 3 plots maturities beginning with 30 days. Compared with the curvature factor, the slope factor is higher for shorter maturities, which confirms that β_{1t} is a rather short-term factor, i.e., affects prices associated with shorter maturities to a greater extent. Conversely, the curvature factor converges to zero as time to maturity approaches zero, and infinity, while β_{2t} has highest loadings for medium maturities with maximum at time to maturity equal to $1/\lambda$.

3.1.1. Decay parameter

The most important element in the Nelson–Siegel class of models is the parameter λ_t, which determines the exponential decay. Low values of the parameter imply a slower decay of the resulting curve and vice versa. Empirically, the choice of the value of λ_t represents a trade-off between fitting the near and far ends of the term structure. Higher values of the parameter result in a better fit of the functional form in the case of short maturities. Conversely, lower values improve the fit for longer maturities [13]. The decay parameter also defines the maturity at which the loading on the medium-term curvature factor β_{2t} is maximized.

In addition, λ_t governs the actual nature of above-defined relationship. If we allow λ_t to evolve dynamically over time, we obtain a nonlinear problem, which is computationally much more demanding. While authors in the yield curve literature often regard 2- or 3-year times to maturity as medium-term maturity and use this assumption to fix $\lambda_t = \lambda$ for all times $t = 1, \ldots, T$, it is infeasible in case of crude oil futures. The literature on the crude oil term structure does not provide any well reasoned suggestions regarding medium-term maturities on oil markets, and there is almost no reference for the proper choice of λ, as modelling the term structure of crude oil markets using the Nelson–Siegel family of models is not fully explored in the literature.

A different approach employs nonlinear least squares estimation of all four parameters in Eq. (1), i.e., $\beta_{0t}, \beta_{1t}, \beta_{2t}$ and λ_t for all t. The main problem with such an approach is that λ_t may be unstable due to unexpected jumps. While the model will fit the data very well, its predictive power deteriorates [46].

We find the optimal values of λ_t by minimizing the sum of squared errors of Nelson–Siegel approximations of the WTI futures term structure for each observed point in time. Fig. 4 illustrates the estimates. To facilitate optimization, we restrict the values to correspond to maturities between 0 and 1000 days. While λ determines the reciprocal value for the number of days to maturity at which the medium-term (i.e., curvature) factor is maximized, searching for the optimal $1/\lambda_t$ outside this interval is superfluous.

We observe that λ_t is unstable for the crude oil futures data and exhibits no clear pattern. Consequently, allowing for a dynamic λ_t means that successful predictions are hardly possible. Therefore, we find a single optimal value of λ by minimizing sum of squared errors of the Nelson–Siegel approximation of the WTI term structure over the whole period as follows:

$$\lambda^* = \underset{\lambda \in \Theta}{\operatorname{argmin}} \sum_{t=1}^{289} \sum_{i=1}^{24} \left(p_t(\tau_i) - \widehat{p}_t(\tau_i; \beta_{0t}, \beta_{1t}, \beta_{2t}, \lambda) \right)^2 \tag{2}$$

where 289 is total number of observed points in time, and 24 is number of analysed constant maturities (from 30 to 720 days). The resulting value of $\lambda^* = 0.0058$, implying a reciprocal value of $1/\lambda^*$ equal to 173.4551, thereby yielding an acceptable value of medium term maturity.[7] The optimization result is in line with the literature. Grønborg and Lunde [18], who analyses oil futures (albeit in a different period) arrived at a λ equal to 0.005.

3.1.2. Level, slope, and curvature estimates

Having set the optimal value of λ^*, we proceed with in-sample estimation of the set of β_t coefficients on latent factors. For all times t, the parameters are obtained from ordinary least squares (OLS) fit across maturities as follows:

$$\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^{24} \left(p_t(\tau_i) - \beta_0 - \beta_1 \left(\frac{1 - e^{-\lambda^* \tau_i}}{\lambda^* \tau_i} \right) - \beta_2 \left(\frac{1 - e^{-\lambda^* \tau_i}}{\lambda^* \tau_i} - e^{-\lambda^* \tau_i} \right) \right)^2 \tag{3}$$

where $p_t(\tau_i)$ is the WTI futures price at time t with time to maturity τ_i. This procedure yields a time series of three β-coefficients, with a length of 289 values.

Estimates of the β_t coefficients are plotted in Fig. 5. At first glance, the behaviour of β_{0t} – the level coefficient – attracts attention. An increase in the level coefficient over full period of observation corresponds to a general increase in crude oil prices. The slope and curvature coefficients seem to generally be more stable. The slope factor fluctuates around zero in the first part of the sample and then becomes positive until 2008, meaning that the resulting term structure is downward sloping. After 2008, the slope coefficient jumps to large negative values and remains negative for the following two years, implying an upward-sloping term structure. The most recent period beginning in 2011 is characterized by positive values, which implies a decreasing term structure. Diebold and Li [13] propose forecasting the factor loadings using autoregressive and vector-autoregressive models, with a random walk as the benchmark. One of the directly observable features of the factor loadings is their non-stationarity. Stationarity is rejected for the level factor, and the two remaining factors are on the border. Whereas this makes further autoregressive analysis

[7] The maximum observed time to maturity in our period reached less than 2000 days = approx. 6 years, which is much less compared with the 30 years in the case of the U.S. yield curve. In such a case, authors tend to regard 2–3 years as medium maturity.

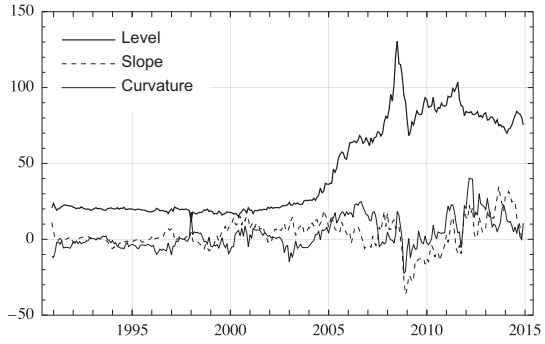

Fig. 5. Estimated coefficients from the dynamic Nelson–Siegel model of crude oil futures for the period 1990–2014. Level – β_{0t}, slope – β_{1t}, and curvature – β_{2t}.

problematic, it forms part of our motivation for the use of neural networks, which do not require the assumption stationary time series. In addition, the factors can contain nonlinearities, which are not captured by simple linear time series analysis.

Before we turn to the main part of the analysis, forecasting, we illustrate the fit of the dynamic Nelson–Siegel model to crude oil futures in Fig. 6. The term structures are generally fitted with a high degree of accuracy for all curve shapes. Similar to Diebold and Li [13], in the case of a term structure with multiple local extremes (as during May 1999), the approximation is not as accurate.

4. Forecasting the term structure with neural networks

To obtain the future term structure forecasts from the dynamic Nelson–Siegel model, Diebold and Li [13] propose forecasting individual β_t coefficients using linear autoregressive (AR) and Vector AR (VAR) models. In this work, we propose to forecast the individual coefficients on factor loadings using artificial neural networks. The motivation is straightforward, as the β_t coefficients are not stationary for crude oil futures and may further contain nonlinear dependence. Linear models are not able to capture these features well, and hence we hypothesize that our proposed approach will yield more accurate forecasts. Similarly to Diebold and Li [13], a forecast of futures price at forecast horizon h will be calculated as follows:

$$\widehat{p}_{t+h}(\tau) = \widehat{\beta}_{0,t+h} + \widehat{\beta}_{1,t+h}\left(\frac{1-e^{-\lambda^*\tau}}{\lambda^*\tau}\right) + \widehat{\beta}_{2,t+h}\left(\frac{1-e^{-\lambda^*\tau}}{\lambda^*\tau} - e^{-\lambda^*\tau}\right), \quad (4)$$

where $\widehat{\beta}_{i,t+h}$ are coefficients to be predicted. Both the AR and VAR models that Diebold and Li [13] use for prediction are developed to capture linear features of the time series. Hence, in using them to forecast coefficients on factor loadings, one assumes that they are generated by linear processes. This is not the case for artificial neural networks (ANNs), as ANNs do not require any assumptions regarding the statistical properties of the underlying series for their proper application. ANNs may be understood as a generalization of these classical approaches, which allow us to model different types of nonlinearities in the data.

Although neural networks, which imitate neural processing in brain activation, are primarily associated with biological systems and have ben successfully applied in numerous fields, such as pattern recognition and medical diagnostics, many econometricians argue that the approach is a black box. This, combined with the fact that one must make arbitrary decisions concerning the implementation of the network, e.g., the number of hidden layers, the choice of transformation functions, the number of neurons, neural networks are not commonly used for financial time series modelling, and we are pioneering their use in term structure forecasting.

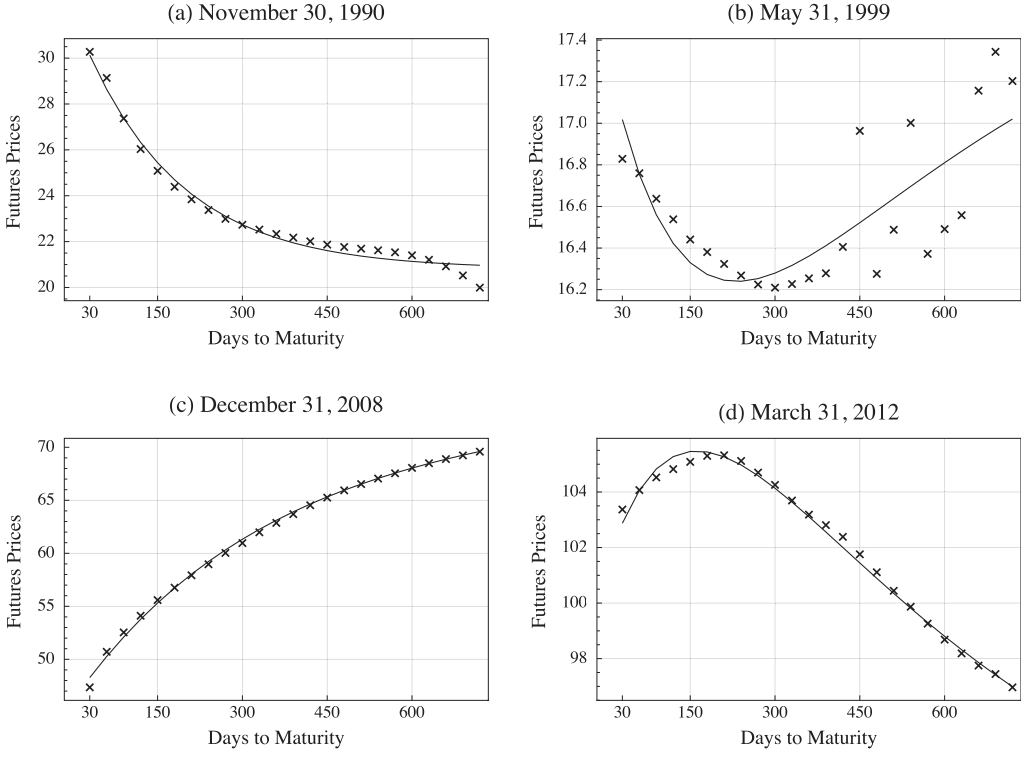

Fig. 6. Examples of the term structures of futures contracts on crude oil (indicated by ×) fitted by the Dynamic Nelson–Siegel model (indicated by solid line): (a) November 30, 1990, (b) May 31, 1999, (c) December 31, 2008, and (d) March 31, 2012.

Abandoning these concerns, we use the neural network as a generalized nonlinear regression, which is able to describe the complex patterns in the time series of curvature parameters. As in other linear or nonlinear methods, a neural network relates a set of input variables, say lags of time series, to output – in our case the forecast. The only difference between network and other models is that the approximating function uses one or more so-called hidden layers, in which the input variables are squashed or transformed by a special function.

The most widely used ANN in financial applications with one hidden layer [20] is the feed-forward neural network. The general feed-forward or multi-layered perceptron (MLP) network we use for forecasting the $\widehat{\beta}_{t+h}$ coefficient can be described by the following equations:

$$\widehat{\beta}_{t+h} = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k \Lambda(n_{k,t}) \tag{5}$$

$$\Lambda(n_{k,t}) = \frac{1}{1+e^{-n_{k,t}}} \tag{6}$$

$$n_{k,t} = \omega_{k,0} + \sum_{i=0}^{m+1} \omega_{k,i} \widehat{\beta}_{t-i} \tag{7}$$

with k^* neurons $n_{k,t}$ and $\omega_{k,i}$ representing a coefficient vector or *weights* vector to be identified. The variable $n_{k,t}$, consisting of $m+1$ lags of the time series being forecast, is squashed by the hyperbolic tangent transfer function and becomes a neuron $\Lambda(n_{k,t})$. Next, the set of k^* neurons are combined linearly with the vector of coefficients $\{\gamma_k\}_{k=1}^{k^*}$ to form the final output, which is the forecast of the $\widehat{\beta}_{t+h}$ coefficient on factor loadings from the dynamic Nelson–Siegel model. The general feed-forward network is the workhorse of the neural network modelling approach in the finance industry, as nearly all researchers begin with this network as the first alternative to linear models.

Note that AR is a simple special case within this framework if the transformation $\Lambda(n_{k,t})$ is skipped (i.e., $\Lambda(n_{k,t}) = n_{t,k}$) and one neuron that contains a linear approximation function is used. Therefore, in addition to classical linear models, there are neurons that process the inputs to improve the predictions.

To be able to approximate the target function, the neural network must be able to *"learn"*. The process of learning is defined as the adjustment of weights using a learning algorithm. The main goal of the learning process is to minimize the sum of the prediction errors for all training examples. The training phase is thus an unconstrained nonlinear optimization problem, where the goal is to find the optimal set of parameter weights by solving the following minimization problem:

$$\min\{\Psi(\omega) : \omega \in \mathbb{R}^n\}, \tag{8}$$

where $\Psi : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable error function. There are several ways of minimizing $\Psi(\omega)$, but in essence, we are searching for the gradient $G = \nabla\Psi(\omega)$ of function Ψ, which is the vector of the first partial derivative of the error function $\Psi(\omega)$ with respect to the weight vector ω. Furthermore, the gradient specifies the direction that produces the steepest increase in Ψ. The negative of this vector thus provides us with the direction of the steepest decrease.

However, traditional gradient descent algorithms often fail to efficiently learn intricate patterns in the data due to the numerous possible initial settings. One of the efficient methods for learning the patterns in feed-forward neural networks, which we use, is the Levenberg–Marquardt back-propagation.

4.1. Focused time-delay neural network

To be able to fully explore the time dependence in the time series, we use a simple extension of the feed-forward framework, as dynamic neural networks are capable of more effectively learning the dynamics of time series relationships. The time-delay neural Network is a feed-forward network with a tapped delay line at the input. It is similar to a multilayer perceptron, as all connections feed forward. In addition, the inputs to any node consist of the outputs of earlier nodes from previous time steps. This is generally implemented using tap-delay lines.

The most straightforward general dynamic neural networks are of the class known as focused time-delay neural networks, which have delays only on the input units [47]. These models consist of a set of feed-forward networks with a tapped delay line capturing the autoregressive property of the inspected series. We propose the use of the focused time-delay neural network (FTDNN) for forecasting β_t loadings. The delay Δ is introduced into Eq. (7) as follows:

$$n_{k,t} = \omega_{k,0} + \sum_{i=0}^{m+1} \omega_{k,i} \widehat{\beta}_{t-(i-1)\Delta} \tag{9}$$

To forecast the three time series of β_t coefficients estimated by the Nelson–Siegel model, we will naturally use three separate networks. To prevent over-fitting, we use cross-validation over time with a fixed window. The best model is always chosen based on the cross-validation scheme. In-sample (training and validation) and out-of-sample (testing) datasets are chosen in usual ratio of 60%, 20%, and 20% for training, validation, and testing, respectively. In terms of the out-of-sample forecast period, we begin to forecast the futures prices in 2010. The same period is also used for forecasts from competing models defined in the next section.

The input layer consists of the m lags relevant for the forecast, where m can be determined by inspecting the respective sample autocorrelation function. To maintain the comparability of forecast results with AR(1) and VAR(1) models, we use one lag. A simple network with one hidden layer consisting of up to 20 hidden neurons is considered. An output neuron is the lth step-ahead forecast of a particular β_t coefficient: 1-month, 3-month-, 6-month- and 12-month-ahead forecasts have been examined. The final decision regarding the network structure is made based on the Hannan–Quinn information criterion,[8] as it punishes networks with an excessive number of parameters.

5. Out-of-sample forecasting performance

5.1. Competing models

The main interest of this work is to assess the out-of-sample forecasting performance of neural networks in forecasting the term structure of crude oil futures. Naturally, we assess their performance in comparison with competing models used in the literature. The first competing model we consider is a simple AR(1) process for all three $\widehat{\beta}_{i,t+h}$ coefficients $i = \{1,2,3\}$:

$$\widehat{\beta}_{i,t+h} = \widehat{c}_i + \widehat{\gamma}_i \widehat{\beta}_{it}, \tag{10}$$

where coefficients $\widehat{c}_i$ and $\widehat{\gamma}_i$ are obtained by regressing $\widehat{\beta}_{i,t}$ on $\widehat{\beta}_{i,t-h}$ and an intercept. Factor loadings $\widehat{\beta}_{i,t+h}$ may generally contain unit root, which will result in poor forecasts due to large possible biases in estimates. Nevertheless, the model is used in the literature modelling yield curves and term structures.

[8] $HQIC = \left[\ln\left(\sum_{t=1}^{N} \frac{(\beta_t - \widehat{\beta}_t)^2}{N}\right)\right] + \frac{k(\ln(\ln(N)))}{N}$.

The second benchmark model for forecasting the term structure that we consider is the VAR:

$$\widehat{\beta}_{t+h} = \widehat{c} + \widehat{\Gamma}\widehat{\beta}_t, \tag{11}$$

with $\widehat{c}$ and $\widehat{\Gamma}$ representing coefficients to be estimated. In autoregressive models, issues implied by the potential presence of a unit root in one of the series are not particularly severe. However, unrestricted VAR models perform quite poorly in forecasting tasks. This poor performance is primarily due by danger of over-parameterization because of the large number of parameters. Diebold and Li [13] also note that the factors do not share a cross-correlation structure, and hence we should not expect the VAR(1) model to produce superior forecasts. The situation is different for term structures, as the coefficients share interaction to be modelled.

As a final benchmark model, we consider a random walk, where the expected forecast is the previous lag:

$$\widehat{\beta}_{i,t+h} = \beta_{it}. \tag{12}$$

All four models are used to forecast the term structure of crude oil futures, both in one-step-ahead and multi-step-ahead predictions (we consider 1, 3, 6, and 12 months ahead).

5.2. Evaluation of the forecasts

To statistically compare the accuracy of the forecasts from different models, we employ two common loss functions, namely the root mean square error (RMSE) and the mean absolute error (MAE). The measures are calculated for the $t = 1, \ldots, T$ forecasts as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{T}\left(\widehat{p}_{t+i} - p_{t+i}\right)^2} \tag{13}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{T}\left|\widehat{p}_{t+i} - p_{t+i}\right| \tag{14}$$

As Nomikos and Pouliasis [48] note, these metrics do not provide information on the asymmetry of the errors. While asymmetric errors are commonly found in the volatility literature, it may also be worthwhile to determine whether the models do not systematically over-, or under-predict the term structures. For example, Nomikos and Pouliasis [48], Wang and Wu [49], and Baruník and Křehlík [30] report that the majority of forecasting models over-predict the volatility on petroleum markets. This bias then translates into direct economic losses. Hence, as Nomikos and Pouliasis [48] suggest, we employ two additional mean mixed error (MME) loss functions [50] to assess the forecasts. These functions employ a mixture of positive and negative forecast errors with different weights, thereby allowing us to discover the cases in which the model tends to over- or under-predict:

$$\text{MME}(O) = \frac{1}{N}\left(\sum_{i\in U}\left|\widehat{p}_{t+i} - p_{t+i}\right| + \sum_{i\in O}\sqrt{\left|\widehat{p}_{t+i} - p_{t+i}\right|}\right), \tag{15}$$

$$\text{MME}(U) = \frac{1}{N}\left(\sum_{i\in U}\sqrt{\left|\widehat{p}_{t+i} - p_{t+i}\right|} + \sum_{i\in O}\left|\widehat{p}_{t+i} - p_{t+i}\right|\right), \tag{16}$$

where U is the set containing under-predictions and O is the set containing over-predictions.

Recently, considerable effort has been devoted to the development of testing procedures for situations in which several models are considered. While the influential work of Diebold and Mariano [51] allows us to statistically compare two models, the Reality Check of White [52], or Superior Predictive Ability (SPA) test of Hansen [53] propose multiple-testing procedures. Among the new developments, the Model Confidence Set (MCS) of Hansen et al. [54] consists of a sequence of statistic tests that permits one to construct a set of superior models, where the null hypothesis of equal predictive power is not rejected at a certain confidence level. To test significant differences in the loss functions of competing models, we use the MCS of Hansen et al. [54]. Given a set of forecasting models $\mathcal{M}_0$ we identify the model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha} \subset \mathcal{M}_0$, which is the set of models that contain the "best" forecasting model at a given level of confidence α. For a given model $i \in \mathcal{M}_0$, the p-value is the threshold confidence level. Model i belongs to the MCS only if $\widehat{p}_i \geqslant \alpha$. The MSC methodology repeatedly tests the null hypothesis of equal forecasting accuracy

$$H_{0,\mathcal{M}} : E[\mathcal{L}_{i,t} - \mathcal{L}_{j,t}] = 0, \quad \text{for all } i,j \in \mathcal{M}$$

with $\mathcal{L}_{i,t}$ being an appropriate loss function of the ith model. Starting with the full set of models $\mathcal{M} = \mathcal{M}_0$, this procedure sequentially eliminates the worst performing model from $\mathcal{M}$ when the null is rejected. The surviving set of models then belongs to the model confidence set $\widehat{\mathcal{M}}^*_{1-\alpha}$. Following Hansen et al. [54], we implement the MCS using a stationary bootstrap with an average block length of 20 days.[9]

Table 3
The average root mean square error (RMSE) across all constant maturities. The lowest RMSE is highlighted in bold.

Horizon	FTDNN	AR(1)	VAR(1)	RW
1 month	**4.398**	4.708	4.971	4.772
3 months	**6.077**	7.572	8.060	7.952
6 months	**6.425**	8.868	10.362	10.140
12 months	**7.881**	7.947	11.487	9.841

5.3. Discussion of the results

Four models – focused time-delay neural network (FTDNN), AR(1), VAR(1) and random walk (RW) – are used to forecast the term structure of crude oil futures, both in one-step-ahead and multi-step-ahead predictions. We begin with a discussion of the aggregate results. The average forecast error over all maturities, captured by RMSE loss function in Table 3, reveals that FTDNN produces forecasts with the lowest errors for all forecasting horizons considered. The second-best forecasting model is the AR(1) model, confirming the conclusions of Diebold and Li [13] regarding the yield curves data that the AR(1) model outperforms both VAR(1) and RW.

While the average results provide us with the first indication of how the models perform against one another, a rigorous statistical comparison is provided by Tables 4 and 5 in Appendix A, which present a summary of forecast performance results for individual maturities. For clarify the results, we report RMSE and MAE relative to the respective statistics from the random walk model as the benchmark. A simple ratio tells us, conveniently, how the model under evaluation compares with the benchmark random walk. Moreover, the model confidence set is found across all models for all times to maturity and multi-step-ahead forecasts.

In the case of the one-month-ahead forecast, FTDNN yields the lowest RMSE and MAE of the models. FTDNN is the only model in the MCS for maturities lower than 630 days according to RMSE. For longer maturities considered, specifically 660, 690, and 720, AR(1), and RW belong to the MCS, while VAR(1) is rejected in all cases. Regarding the MAE, the situation is very similar, with the only difference being that for maturities longer than 420, the FTDNN, AR(1), and RW models produce statistically

[9] We considered different block lengths, including ones depending on the forecasting horizons, to assess the robustness of the results, without any change in the final results. These results are available from the authors upon request.

Table 4
The table reports RMSE from model forecasts relative to the RMSE of the RW model. The competing model has lower error than RW if the ratio is lower than one, and vice versa. We consider multiple-step-ahead forecasts of 1, 3, 6, and 12 months for individual maturities using four competing models: FTDNN, AR(1), VAR(1), and RW. The model confidence set (MSC) is used to compare the errors across the four competing models. The ratio is bold in case of the model belonging to the $\widehat{\mathcal{M}}^{*}_{10\%}$.

RMSE	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																								
1 M	**0.81**	**0.81**	**0.80**	**0.80**	**0.80**	**0.79**	**0.79**	**0.79**	**0.78**	**0.78**	**0.77**	**0.77**	**0.77**	**0.77**	**0.77**	**0.77**	**0.77**	**0.77**	**0.77**	**0.77**	**0.78**	**0.79**	**0.80**	**0.83**
3 M	**0.73**	**0.73**	**0.73**	**0.72**	**0.72**	**0.72**	**0.72**	**0.72**	**0.72**	**0.72**	**0.72**	**0.72**	**0.72**	**0.73**	**0.73**	**0.73**	**0.74**	**0.74**	**0.74**	**0.75**	**0.75**	**0.76**	**0.77**	**0.78**
6 M	**0.59**	**0.59**	**0.58**	**0.58**	**0.58**	**0.58**	**0.58**	**0.59**	**0.59**	**0.60**	**0.61**	**0.61**	**0.62**	**0.62**	**0.63**	**0.64**	**0.64**	**0.65**	**0.66**	**0.66**	**0.67**	**0.67**	**0.68**	**0.69**
12 M	**0.74**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.74**	**0.74**	**0.74**	**0.74**	**0.75**	**0.75**	**0.75**	**0.75**
AR(1)																								
1 M	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	**0.99**	**1.00**	**1.02**
3 M	0.93	0.94	0.95	0.95	0.95	0.95	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.97	0.96	0.97	0.97	0.97	0.98
6 M	0.87	0.87	0.87	0.87	0.87	0.86	0.86	0.86	0.86	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.87**	**0.88**	**0.88**	**0.89**	**0.89**	**0.89**	**0.90**	**0.90**	**0.90**
12 M	0.87	0.87	0.85	0.84	**0.83**	**0.81**	**0.80**	**0.79**	**0.78**	**0.78**	**0.77**	**0.77**	**0.77**	**0.77**	**0.78**	**0.78**	**0.78**	**0.79**	**0.80**	**0.81**	**0.82**	**0.82**	**0.83**	**0.83**
VAR(1)																								
1 M	1.01	1.01	1.02	1.02	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.04	1.04	1.04	1.05	1.05	1.06	1.07	1.07	1.07	1.08	1.08	1.09	1.12
3 M	0.99	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04	1.05	1.05	1.05	1.06	1.08
6 M	1.08	1.07	1.06	1.05	1.03	1.02	1.02	1.01	1.01	**1.01**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**	**1.01**	**1.01**	**1.01**	**1.02**	**1.02**	**1.03**
12 M	1.21	1.20	1.19	1.18	1.16	1.15	1.14	1.13	1.13	1.13	1.13	1.13	1.13	1.14	1.14	1.15	1.16	1.17	1.18	1.20	1.21	1.22	1.22	1.23
RW																								
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

Table 5
The table reports MAE from model forecasts relative to the MAE of RW model. The competing model has lower error than RW if the ratio is lower than one, and vice versa. We consider multiple-step-ahead forecasts of 1, 3, 6, and 12 months for individual maturities using four competing models: FTDNN, AR(1), VAR(1), and RW. The model confidence set (MSC) is used to compare the errors across the four competing models. The ratio is bold if the model belongs to the $\widehat{\mathcal{M}}^{*}_{10\%}$.

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																								
1 M	**0.79**	**0.80**	**0.81**	**0.81**	**0.83**	**0.83**	**0.83**	**0.83**	**0.83**	**0.84**	**0.83**	**0.83**	**0.83**	**0.83**	**0.84**	**0.84**	**0.84**	**0.83**	**0.84**	**0.83**	**0.84**	**0.85**	**0.85**	**0.88**
3 M	**0.73**	**0.71**	**0.70**	**0.69**	**0.68**	**0.68**	**0.68**	**0.68**	**0.69**	**0.68**	**0.69**	**0.68**	**0.68**	**0.68**	**0.68**	**0.70**	**0.71**	**0.72**	**0.72**	**0.73**	**0.74**	**0.75**	**0.77**	**0.77**
6 M	**0.61**	**0.62**	**0.63**	**0.63**	**0.63**	**0.64**	**0.65**	**0.65**	**0.66**	**0.66**	**0.67**	**0.67**	**0.67**	**0.68**	**0.68**	**0.69**	**0.69**	**0.70**	**0.70**	**0.70**	**0.70**	**0.71**	**0.72**	**0.74**
12 M	**0.68**	**0.69**	**0.70**	**0.71**	**0.72**	**0.72**	**0.72**	**0.72**	**0.71**	**0.71**	**0.70**	**0.70**	**0.69**	**0.69**	**0.68**	**0.68**	**0.67**	**0.67**	**0.66**	**0.66**	**0.66**	**0.65**	**0.65**	**0.64**
AR(1)																								
1 M	0.95	0.96	0.97	0.98	0.99	0.99	0.99	1.00	1.00	1.01	0.99	1.00	0.99	**1.00**	**1.01**	**1.01**	**1.01**	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**	**1.02**
3 M	0.98	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.96	0.96	0.96	0.96	0.97
6 M	0.92	0.93	0.94	0.95	0.95	0.94	0.93	0.92	**0.92**	**0.92**	**0.92**	**0.91**	**0.91**	**0.91**	**0.90**	**0.90**	**0.89**	**0.89**	**0.89**	**0.88**	**0.88**	**0.87**	**0.87**	**0.88**
12 M	0.83	0.84	0.85	0.85	0.84	0.84	**0.82**	**0.81**	**0.78**	**0.77**	**0.74**	**0.73**	**0.73**	**0.72**	**0.72**	**0.71**	**0.71**	**0.71**	**0.72**	**0.73**	**0.75**	**0.76**	**0.77**	**0.77**
VAR(1)																								
1 M	1.02	1.02	1.03	1.03	1.04	1.04	1.05	1.05	1.06	1.07	1.07	1.08	1.09	1.09	1.11	1.12	1.12	1.12	1.12	1.12	1.13	1.14	1.14	1.17
3 M	1.02	1.03	1.03	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.01	1.02	1.02	1.02	1.03	1.04	1.04	1.05	1.05	1.06	1.07	1.08	1.09	1.12
6 M	1.14	1.14	1.13	1.12	1.10	1.07	1.05	1.03	**1.02**	**1.00**	**0.98**	**0.97**	**0.96**	**0.96**	**0.96**	**0.97**	**0.97**	**0.98**	**0.98**	**0.98**	**0.99**	**1.00**	**1.01**	**1.03**
12 M	1.17	1.16	1.16	1.16	1.16	1.16	1.15	1.14	1.14	1.14	1.14	1.14	1.16	1.17	1.19	1.20	1.21	1.22	1.25	1.26	1.27	1.28	1.28	1.28
RW																								
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

Table 6
The table reports the average number of cases in which the error in the forecasts of four competing models, FTDNN, AR(1), VAR(1), and RW, are negative. The higher the number than 0.5, the more under-predictions the model yields. We consider multiple-step-ahead forecasts of 1, 3, 6, and 12 months for individual maturities. The model confidence set (MSC) is used to compare the MME(U) errors across the four competing models. The ratio is bold when the model belongs to the $\widehat{\mathcal{M}}_{10\%}$.

MME(U)	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																								
1 M	**0.53**	**0.56**	**0.60**	**0.61**	**0.61**	**0.58**	**0.58**	**0.58**	**0.56**	**0.60**	**0.56**	**0.56**	**0.58**	**0.60**	**0.60**	**0.60**	**0.58**	**0.54**	**0.58**	**0.56**	**0.56**	**0.58**	**0.54**	**0.54**
3 M	**0.55**	**0.53**	**0.55**	**0.56**	**0.56**	**0.56**	**0.53**	**0.53**	**0.51**	**0.51**	**0.51**	**0.49**	**0.49**	**0.47**	**0.47**	**0.45**	**0.44**	**0.44**	**0.42**	**0.40**	**0.40**	**0.40**	**0.40**	**0.38**
6 M	**0.40**	**0.44**	**0.42**	**0.44**	**0.46**	**0.46**	**0.46**	**0.42**	**0.40**	**0.40**	**0.40**	**0.40**	**0.40**	**0.40**	**0.38**	**0.37**	**0.37**	**0.37**	**0.37**	**0.33**	**0.33**	**0.33**	**0.31**	**0.31**
12 M	**0.74**	**0.72**	**0.72**	**0.70**	**0.72**	**0.74**	**0.72**	**0.74**	**0.72**	**0.72**	**0.67**	**0.63**	**0.61**	**0.61**	**0.61**	**0.59**	**0.59**	**0.61**	**0.61**	**0.61**	**0.61**	**0.61**	**0.61**	**0.59**
AR(1)																								
1 M	0.51	0.51	0.53	0.53	0.54	0.54	0.54	0.56	0.56	0.61	0.60	0.60	**0.60**	**0.63**	**0.61**	**0.63**	**0.61**	**0.60**	**0.61**	**0.60**	**0.60**	**0.60**	**0.58**	**0.60**
3 M	**0.60**	**0.60**	**0.60**	**0.62**	**0.62**	**0.64**	0.62	**0.60**	**0.60**	0.60	**0.60**	**0.58**	**0.60**	**0.58**	**0.58**	**0.58**	**0.58**	**0.55**	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**
6 M	**0.69**	**0.67**	**0.65**	**0.67**	**0.67**	**0.67**	**0.67**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.58**	**0.58**	**0.58**	**0.56**	**0.56**	**0.54**
12 M	0.83	0.83	0.83	0.83	**0.78**	**0.78**	**0.78**	**0.78**	**0.78**	**0.74**	**0.72**	**0.65**	**0.61**	**0.59**	**0.59**	**0.59**	**0.52**	**0.48**	**0.50**	**0.43**	**0.39**	**0.37**	**0.33**	**0.33**
VAR(1)																								
1 M	0.53	0.53	0.54	0.54	0.54	0.53	0.54	0.54	0.51	0.47	0.47	0.47	0.47	0.47	0.47	0.46	0.46	0.46	0.44	0.44	0.44	0.44	0.44	0.44
3 M	**0.62**	**0.65**	**0.65**	**0.60**	**0.58**	**0.58**	0.58	**0.58**	**0.56**	0.55	**0.55**	0.53	0.49	0.49	0.47	0.45	0.45	0.44	0.44	0.42	0.42	0.38	0.36	0.38
6 M	**0.75**	**0.75**	**0.75**	**0.75**	**0.73**	**0.73**	**0.69**	**0.67**	**0.67**	**0.67**	**0.63**	**0.60**	**0.60**	**0.56**	**0.52**	**0.54**	**0.56**	**0.54**	**0.52**	**0.50**	**0.48**	**0.46**	0.46	0.42
12 M	0.78	0.78	0.76	0.76	0.74	0.72	0.72	0.65	0.61	0.59	0.57	0.50	0.50	0.48	0.46	0.46	0.43	0.43	0.43	0.41	0.39	0.35	0.35	0.35
RW																								
1 M	0.53	0.54	0.54	0.51	0.51	0.51	0.53	0.53	0.53	0.51	0.53	0.53	**0.51**	**0.51**	**0.47**	**0.47**	**0.51**	**0.49**	**0.51**	**0.51**	**0.51**	**0.49**	**0.49**	**0.47**
3 M	0.53	0.51	0.51	0.53	0.53	0.53	0.53	**0.51**	**0.51**	0.51	**0.51**	0.49	**0.51**	0.55	0.55	**0.53**	**0.51**	**0.53**	**0.53**	**0.51**	**0.51**	**0.51**	**0.51**	**0.51**
6 M	0.60	0.60	**0.58**	**0.54**	**0.50**	0.50	0.50	0.48	0.48	0.48	0.48	0.48	0.46	0.46	0.42	0.42	0.42	0.42	0.42	0.40	0.40	0.40	0.40	0.40
12 M	0.59	0.59	0.59	0.59	0.57	0.57	0.50	0.50	0.46	0.43	0.41	0.37	0.37	0.39	0.39	0.39	0.37	0.37	0.37	0.37	0.33	0.30	0.30	0.30

Table 7
The table reports the average number of cases in which the error in the forecasts of four competing models, FTDNN, AR(1), VAR(1), and RW, are positive. The higher the number is than 0.5, the more over-predictions the model yields. We consider multiple-step-ahead forecasts of 1, 3, 6, and 12 months for individual maturities. The model confidence set (MSC) is used to compare the MME(O) errors across the four competing models. The ratio is bold when the model belongs to the $\widehat{\mathcal{M}}_{10\%}$.

MME(O)	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																								
1 M	**0.47**	**0.44**	**0.40**	**0.39**	**0.39**	**0.42**	**0.42**	**0.42**	**0.44**	**0.40**	**0.44**	**0.44**	**0.42**	**0.40**	**0.40**	**0.40**	**0.42**	**0.46**	**0.42**	**0.44**	**0.44**	**0.42**	**0.46**	**0.46**
3 M	**0.45**	**0.47**	**0.45**	**0.44**	**0.44**	**0.44**	**0.47**	**0.47**	**0.49**	**0.49**	**0.49**	**0.51**	**0.51**	**0.53**	**0.53**	**0.55**	**0.56**	**0.56**	**0.58**	**0.60**	**0.60**	**0.60**	**0.60**	**0.62**
6 M	**0.60**	**0.56**	**0.58**	**0.56**	**0.54**	**0.54**	**0.54**	**0.58**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.60**	**0.62**	**0.63**	**0.63**	**0.63**	**0.63**	**0.67**	**0.67**	**0.67**	**0.69**	**0.69**
12 M	**0.26**	**0.28**	**0.28**	**0.30**	**0.28**	**0.26**	**0.28**	**0.26**	**0.28**	**0.28**	**0.33**	**0.37**	**0.39**	**0.39**	**0.39**	**0.41**	**0.41**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.39**	**0.41**
AR(1)																								
1 M	**0.49**	0.49	0.47	0.47	0.46	**0.46**	**0.46**	**0.44**	**0.44**	**0.39**	**0.40**	**0.40**	**0.40**	**0.37**	**0.39**	**0.37**	**0.39**	**0.40**	**0.39**	**0.40**	**0.40**	**0.40**	**0.42**	**0.40**
3 M	0.40	0.40	0.40	0.38	0.38	0.36	0.38	0.40	0.40	0.40	0.40	0.42	0.40	0.42	0.42	0.42	0.42	0.45	0.47	0.47	0.47	0.47	0.47	0.47
6 M	0.31	0.33	0.35	0.33	0.33	0.33	0.33	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.42	0.42	**0.42**	**0.44**	**0.44**	**0.46**
12 M	**0.17**	**0.17**	**0.17**	**0.17**	**0.22**	**0.22**	**0.22**	**0.22**	**0.22**	**0.26**	**0.28**	**0.35**	**0.39**	**0.41**	**0.41**	**0.41**	**0.48**	**0.52**	**0.50**	**0.57**	**0.61**	**0.63**	**0.67**	**0.67**
VAR(1)																								
1 M	0.47	0.47	0.46	**0.46**	**0.46**	**0.47**	**0.46**	**0.46**	**0.49**	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**	**0.54**	**0.54**	**0.54**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**	**0.56**
3 M	0.38	0.35	0.35	0.40	0.42	0.42	0.42	0.42	0.44	0.45	0.45	0.47	0.51	0.51	0.53	0.55	0.55	0.56	0.56	0.58	0.58	0.62	0.64	0.62
6 M	0.25	0.25	0.25	0.25	0.27	0.27	0.31	0.33	0.33	0.33	0.37	0.40	0.40	0.44	0.48	0.46	0.44	0.46	0.48	0.50	**0.52**	**0.54**	**0.54**	**0.58**
12 M	0.22	0.22	0.24	0.24	0.26	0.28	0.28	0.35	0.39	0.41	0.43	0.50	0.50	0.52	0.54	0.54	0.57	0.57	0.57	0.59	0.61	0.65	0.65	0.65
RW																								
1 M	**0.47**	0.46	**0.46**	**0.49**	**0.49**	**0.49**	**0.47**	**0.47**	**0.47**	**0.49**	**0.47**	**0.47**	**0.49**	**0.49**	**0.53**	**0.53**	**0.49**	**0.51**	**0.49**	**0.49**	**0.49**	**0.51**	**0.51**	**0.53**
3 M	0.47	0.49	0.49	0.47	0.47	0.47	0.47	0.49	0.49	0.49	0.49	0.51	0.49	0.45	0.45	0.47	0.49	0.47	0.47	0.49	0.49	0.49	0.49	0.49
6 M	0.40	0.40	0.42	0.46	0.50	0.50	0.50	0.52	0.52	0.52	0.52	0.52	0.54	0.54	0.58	0.58	0.58	0.58	0.58	0.60	**0.60**	**0.60**	**0.60**	**0.60**
12 M	**0.41**	**0.41**	**0.41**	**0.41**	**0.43**	**0.43**	**0.50**	**0.50**	**0.54**	**0.57**	**0.59**	**0.63**	**0.63**	**0.61**	**0.61**	**0.61**	**0.63**	**0.63**	**0.63**	**0.63**	**0.67**	0.70	0.70	0.70

indistinguishable forecasts, while FTDNN produces the lowest average statistics.

The difference between FTDNN and all other models is even more pronounced when forecasting 3 months ahead, where the forecasts from FTDNN are the only ones that are included in the set of best forecasts using MCS for all times to maturity. The FTDNN hence decisively produces significantly better forecasts than all other models at all maturities.

Longer forecasts for 6 months ahead show that FTDNN produces even greater improvements in terms of RMSE and MAE than at shorter horizons, where it is the only model belonging to the MCS. The longer the horizon, the lower the gains from the FTDNN against all other models are. While FTDNN produces the lowest average RMSE and MAE, none of the models can be rejected from the MCS for maturities longer than 300 days. This means that all models produce statistically similar 6-month-ahead forecasts for longer horizons.

The longest horizon forecasts of one year show similar results to the 6-month forecast, with VAR(1) and RW being rejected from MCS for all maturities. For the short maturities, the FTDNN produces the best forecasts, while for longer maturities, AR(1) is also included in the MCS.

Summarizing the results from the RMSE and MAE, FTDNN produces the forecasts with the significantly lowest errors in relative to the competing models for short maturities and short forecasting horizons. For maturities longer than 300 days and for and longer forecasting horizons, the other models are relevant. Often, forecasts from the AR(1) model cannot be statistically distinguished from the forecasts of FTDNN. Note here that FTDNN includes only one delayed input to make the model comparable to the AR(1) and VAR(1) strategies used in the literature, and the FTDNN forecasts would improve if an increasing number were introduced. While we have experimented with the number of lags, and obtained even lower errors, the sample size does not allow us to rigorously study these models, and we leave this task for future research.

To determine whether the models over- or under-predict the term structures, we employ the MME(U) and MME(O) statistics.

Table 6 shows the average number of cases in which the error from the model is negative for all models across forecasting horizons and maturities. Table 7 reports the average number of cases in which the error from the model is positive. In addition, asymmetric errors are tested using MME(U), and MME(O) in the MCS framework. In brief, Table 6 indicates whether the models tend to under-predict the term structures, while Table 7 tends to over-predict the term structures.

The important observation from the asymmetric loss functions is that the models, in general, produce symmetric short-term forecasts and for short times to maturity. With longer times to maturity, the FTDNN tends to under-predict for the 1-month- and 12-month-ahead forecasts, while it over-predicts for the 3-month-, and 6-month-ahead forecasts. AR(1) tends to under-predict at all forecasting horizons. For the longest forecasting horizon of 1 year, and longer times to maturities, AR(1) substantially over-predicts futures prices. The results reveal a similar pattern in terms of forecast comparisons. FTDNN is never rejected from the MCS.

6. Conclusion

This paper investigates the properties of the term structure of crude oil markets and proposes the use of dynamic neural networks for forecasting in this context.

The term structure of crude oil futures prices exhibits very similar behaviour to the yield curve on government bonds, and the three-factor dynamic Nelson–Siegel model [13] used in the literature to model yield curves captures the shapes of the term structure very well. We further forecast the factors using a dynamic neural network.

Our proposed framework yields significant improvements in the futures price forecasts when compared with the benchmark models. We report the performance of the models on the 1-month, 3-month, 6-month and 12-month forecasting horizons. Moreover, the forecasting errors from our approach have traceable patterns. For a fixed forecasting horizon, the deviation between the forecast and observed futures price decreases as the time to maturity increases. Furthermore, for more distant forecast horizons, the deviation, on average, increases as expected.

In summary, this work has shown that the crude oil term structure can be successfully modelled and predicted using the parsimonious Nelson–Siegel model that was primarily developed for interest rates when coupled with a generalized regression framework of neural networks. Future research will determine whether our results also hold for other commodities. An interesting and important approach would be to use the framework to study the commonalities between factors across various commodities.

Appendix A

See Tables 4–7.

References

[1] Hamilton JD. Causes and consequences of the oil shock of 2007-08. Technical report, National Bureau of Economic Research; 2009.
[2] Pan H, Haidar I, Kulkarni S. Daily prediction of short-term trends of crude oil prices using neural networks exploiting multimarket dynamics. Front Comput Sci China 2009;3(2):177–91.
[3] Büyüksahin B, Harris JH. Do speculators drive crude oil futures prices? Energy J 2011;32(2):167–202.
[4] Lautier D. Term structure models of commodity prices: a review. Technical report, CEREG; 2005.
[5] Zhang Y-J. Speculative trading and WTI crude oil futures price movement: an empirical analysis. Appl Energy 2013;107:394–402.
[6] Narayan PK, Narayan S, Zheng X. Gold and oil futures markets: are markets efficient? Appl Energy 2010;87(10):3299–303.
[7] Narayan PK, Narayan S, Popp S. Investigating price clustering in the oil futures market. Appl Energy 2011;88(1):397–402.
[8] Tang B-J, Shen C, Gao C. The efficiency analysis of the european CO_2 futures market. Appl Energy 2013;112:1544–7.
[9] Brennan MJ, Schwartz ES. Evaluating natural resource investments. J Bus 1985;58(2):135–57.
[10] Schwartz ES. The stochastic behavior of commodity prices: implications for valuation and hedging. J Finance 1997;52(3):923–73.
[11] Gabillon J. The term structures of oil futures prices. Oxford Institute for Energy Studies. Working paper; 1991.
[12] Cortazar G, Schwartz ES. Implementing a stochastic model for oil futures prices. Energy Econ. 2003;25(3):215–38.
[13] Diebold FX, Li C. Forecasting the term structure of government bond yields. J Economet 2006;130(2):337–64.
[14] Karstanje D, van der Wel M, van Dijk D. Common factors in commodity futures curves; 2015.
[15] Nomikos NK, Pouliasis PK. Petroleum term structure dynamics and the role of regimes. J Futures Markets 2015;35(2):163–85.
[16] Almansour A. Convenience yield in commodity price modeling: a regime switching approach. Energy Econ 2014.
[17] Heidorn T, Mokinski F, Rühl C, Schmaltz C. The impact of fundamental and financial traders on the term structure of oil. Energy Econ 2015;48:276–87.
[18] Grønborg NS, Lunde A. Analyzing oil futures with a dynamic Nelson–Siegel model. J Futures Markets 2015.
[19] Haykin S. Neural networks: a comprehensive foundation. Englewood Cliffs, NJ: Prentice Hall; 2007.
[20] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2(5):359–66.
[21] Siami-Irdemoosa E, Dindarloo SR. Prediction of fuel consumption of mining dump trucks: a neural networks approach. Appl Energy 2015;151:77–84.
[22] Kelesa D, Scelleb J, Paraschivc F, Fichtnera W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks (ann). Appl Energy 2016;162:218–30.
[23] Kialashaki A, Reisel JR. Modeling of the energy demand of the residential sector in the united states using regression models and artificial neural networks. Appl Energy 2013;108:271–80.

[24] An H, Gao X, Fang W, Ding Y, Zhong W. Research on patterns in the fluctuation of the co-movement between crude oil futures and spot prices: a complex network approach. Appl Energy 2014;136(5):1067–75.
[25] Fan Y, Liang Q, Wei Y-M. A generalized pattern matching approach for multi-step prediction of crude oil price. Energy Econ 2008;30(3):889–904.
[26] Yu L, Wang S, Lai KK. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Econ 2008;30(5):2623–35.
[27] Xiong T, Bao Y, Hu Z. Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. Energy Econ 2013;40:405–15.
[28] Jammazi R, Aloui C. Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modeling. Energy Econ 2012;34 (3):828–41.
[29] Papadimitriou T, Gogas P, Stathakis E. Forecasting energy markets using support vector machines. Energy Econ. 2014.
[30] Baruník J, Křehlík T. Coupling high-frequency data with nonlinear models in multiple-step-ahead forecasting of energy markets' volatility. Available at SSRN 2429487; 2014.
[31] Chiroma H, Abdulkareem S, Herawan T. Evolutionary neural network model for west texas intermediate crude oil price prediction. Appl Energy 2015;142:266–73.
[32] Wang L, An H, Liu X, Huang X. Selecting dynamic moving average trading rules in the crude oil futures market using a genetic approach. Appl Energy 2015.
[33] Baumeister C, Guérin P, Kilian L. Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. Int J Forecast 2015;31(2):238–52.
[34] Holton GA. Value-at-risk: theory and practice. Academic Press; 2003.
[35] Hagan PS, West G. Interpolation methods for curve construction. Appl Math Finance 2006;13(2):89–129.
[36] Kilian L, Murphy DP. The role of inventories and speculative trading in the global market for crude oil. J Appl Economet 2014;29(3):454–78.
[37] Mahadeva BF, Kilian L, et al. The role of speculation in oil markets: what have we learned so far? Energy J 2013;34(3).
[38] Hotelling H. The economics of exhaustible resources. J Polit Econ 1931:137–75.
[39] Litzenberger RH, Rabinowitz N. Backwardation in oil futures markets: theory and empirical evidence. J Finance 1995;50(5):1517–45.
[40] Haubrich JG, Higgins P, Miller J. Oil prices: backward to the future. Federal Reserve Bank of Cleveland; 2004. December.
[41] Nelson CR, Siegel AF. Parsimonious modeling of yield curves. J Bus 1987;60 (4):473–89.
[42] Steeley JM. Testing term structure estimation methods: evidence from the uk strips market. J Money Credit Bank 2008;40(7):1489–512.
[43] Sarker RA, Kamruzzaman J, Begg R. Artificial neural networks in finance and manufacturing. IGI Global; 2006.
[44] Bliss RR. Testing term structure estimation methods. Technical report, Federal Reserve Bank of Atlanta; 1996.
[45] Duffie D, Kan R. A yield-factor model of interest rates. Math Finance 1996;6 (4):379–406.
[46] Vela D. Forecasting latin-american yield curves: an artificial neural network approach. Borradores de Economia 2013;761. Banco de la Republica de Colombia.
[47] Clouse DS, Giles CL, Horne BG, Cottrell GW. Time-delay neural networks: representation and induction of finite-state machines. IEEE Trans Neural Netw 1997;8(5):1065–70.
[48] Nomikos NK, Pouliasis PK. Forecasting petroleum futures markets volatility: the role of regimes and market conditions. Energy Econ 2011;33(2):321–37.
[49] Wang Y, Wu C. Forecasting energy market volatility using Garch models: can multivariate models beat univariate models? Energy Econ 2012;34 (6):2167–81.
[50] Brailsford TJ, Faff RW. An evaluation of volatility forecasting techniques. J Bank Finance 1996;20(3):419–38.
[51] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econ Stat 2002;20(1):134–44.
[52] White H. A reality check for data snooping. Econometrica 2000:1097–126.
[53] Hansen PR. A test for superior predictive ability. J Bus Econ Stat 2005;23(4).
[54] Hansen PR, Lunde A, Nason JM. The model confidence set. Econometrica 2011;79(2):453–97.