A Multi-recurrent Network for Crude Oil Price Prediction

Oluwatamilore Orojo*, Jonathan Tepper*†, T.M. McGinnity*‡ and Mufti Mahmud*
*School of Science and Technology
Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom
Email: {oluwatamilore.orojo, mufti.mahmud}@ntu.ac.uk
†Perceptronix Ltd, Avon Way, Derby DE65 5AE, United Kingdom
Email: jtepper@perceptronix.net
‡Intelligent Systems Research Centre
University of Ulster, Magee Campus, Derry BT48 7JL, United Kingdom
Email: tm.mcginnity@ulster.ac.uk

***Abstract*—Crude oil is fundamental for global growth and stability. The factors influencing crude oil prices and more generally, the oil market, are well known to be dynamic, volatile and evolving. Subsequently, crude oil prediction is a complex and notoriously difficult task. In this paper, we evaluate the Multi-recurrent Network (MRN), a simple yet powerful recurrent neural network, for oil price forecasting at various forecast horizons. Although similar models, such as Long Short-Term Memory (LSTM) networks, have shown some success in this domain, the MRN is a comparatively simplified neural network model which exhibits complex state-based memories that are both flexible and rigid. We evaluate the MRN against the standard Feedforward Multilayered Perceptron (FFMLP) and the Simple Recurrent Network (SRN) in addition to the current state-of-the-art LSTM for specifically modelling the shocks in oil prices caused by the financial crisis. The in-sample data consists of key indicator variables sampled across the pre-financial crisis period (July 1969 to September 2003) and the out-sample data used to evaluate the models, is before, during and beyond the crisis (October 2003 to March 2015). We show that such simple sluggish state-based models are superior to the FFMLP, SRN and LSTM models. Furthermore, the MRN appears to have discovered important latent features embedded within the input signal five years prior to the 2008 financial crisis. This suggests that the indicator variables could provide Central Banks and governments with early warning indicators of impending financial perturbations which we consider an invaluable finding and worthy of further exploration.**

I. Introduction

Approximately, 99 million barrels of petroleum were consumed daily in 2018 [1]. Crude oil is therefore arguably the world's most important commodity, it is particularly key in ensuring nations are able to meet their energy demands [2]. Furthermore, oil prices have a massive effect on the price of other commodities and heavily influence macroeconomic projections for gross domestic product and inflation [3], [4]. Accordingly, it is not surprising that forecasting tools for crude oil prices are constantly being researched and developed [5].

These forecasting tools provide insight and foresight of crude oil price trajectories and possible shocks, which aid the attempts to reduce the associated risks to the economy from oil price uncertainty. Recent shocks and their impacts have further heightened interest in understanding oil price behaviour [6]. However, due to the number of factors that influence oil prices, the evolving oil market dynamics and the increasing oil price volatility, the oil price prediction task is not trivial [3].

Traditional statistical and econometric models have been applied to the oil price prediction task. The most common classes of models are linear and include Auto-Regressive Integrated Moving Average (ARIMA) and Generalised / Autoregressive Conditional Heteroskedasticity (G/ARCH). However, given the complexity of the task, the high non-linearity and irregularity of oil prices, these methods are not able to capture the underlying behaviour and dynamics sufficiently well [3].

Consequently, since 2001, researchers have applied Artificial Neural Networks (ANNs) to the task due to their non-linearity and universal function approximation abilities [3]. An evaluation of ANNs for the task highlights the need and importance of effective state-based memory in Recurrent Neural Networks (RNNs) [7].

The embedded memory mechanisms found within RNNs enable them to capture temporal dependencies in oil prices, which are crucial for identifying latent interactions amongst relevant feature variables thereby providing insights into their evolving behaviour dynamics [8]. Authors have applied a range of RNNs to the task; such as Simple Recurrent Network (SRN) and current state-of-the-art, Long-Short Term Memory (LSTM), and these models outperformed Feedforward Multi-layered Perceptrons (FFMLPs).

While SRNs have provided better accuracy, they have a simple memory mechanism which favours the most recent state-based response over any historical responses. Consequently, gradients concerning historical yet important input observations typically begin to vanish quickly, thus limiting their predictive ability. Similarly, LSTMs have been shown to alleviate the vanishing gradient issue however, they use a complex gating mechanism. Ulbricht [9] applied a simple modification of the SRN, called the Multi-recurrent Network (MRN), and others such as Tepper [10] have found that minor modifications such as noise injection improved the learning ability of MRNs and also the quality of their underlying state dynamics. In this paper, we seek to identify whether the MRN has the ability to capture shocks in oil price trajectory and additionally identify if the MRN can provide early indications of the 2008 financial crisis; which could then be used by macroeconomic policymakers to mitigate against any subsequent risk to economic stability.

The structure of this paper is as follows. A summary of the existing approaches to modelling oil prices is given in Section 2, The methodology is given in Section 3. An empirical application to oil prices and results are given in Section 4 and Section 5 concludes.

II. Literature Review

A summary of the key techniques used to predict crude oil until 2013 is categorised by [2] into quantitative and qualitative models. They further classify quantitative models into econometric (e.g. time-series) and non-standard.

Traditional models such as ARIMA, ARCH, GARCH, Markov Switching (MS) model, Regime Switching (RS) model and the Random Walk (RW) model (which is a common benchmark), have been widely applied to crude oil price forecasting [11]. For example, in Sadorsky [12] the authors reported the application of univariate and multivariate statistical models, such as, RW, Historical mean model, Moving Average (MA) model, Exponential Smoothing, Least squares linear regression model, Auto-Regressive model, GARCH(1,1) model, GARCH(1,1) in mean model with variance, Threshold GARCH(1,1) model, State space model, Vector autoregression and Bivariate GARCH (BIGARCH). The predictive value of these models is typically evaluated with respect to the RW. The random walk refers to a model where the best forecast of the volatility for the next time step is the volatility of the current time step and it indicates the efficiency of the oil market[1]. These models were applied to daily closing future price returns on West Texas Intermediate (WTI) crude oil, heating oil, unleaded gasoline, and natural gas. The authors show that the best model is dependent on the chosen series and points out that most of these models outperform the random walk. However, a number of these traditional methods are linear methods and consequently, they are unable to predict the structural shifts found in non-linear time-series data such as oil prices properly. In addition, the need for iterative training and parameter re-estimation is time-consuming without guaranteed efficacy [13].

Due to the limitations of traditional statistical and econometric models, machine learning techniques such as FFMLPs, Support Vector Machine (SVMs) and Genetic Programming (GP) have been employed. We focus on the application of Artificial Neural Networks (ANNs) for crude oil prices forecasting. FFMLPs are considered a standard class of ANNs which are biologically inspired models of machine learning that can estimate powerful non-linear auto-regressive moving average models under appropriate conditions [2].

In particular, Hamdi [3] provide a review on research that has exploited FFMLPs for oil prices forecasting using data samples up until 2015. Moshiri [13] forecast crude oil futures prices for data samples between 1983 and 2003 using ARIMA, GARCH and FFMLP models where they show that the FFMLP outperformed the other models. However, Shabri and Samsudin [14] point out that they suffer from local minima, parameter selection sensitivity and over-fitting [15], [16], [17], [18]. Natarajan [7] highlight the need of advanced models such as RNNs for the oil price prediction task as FFMLPs can not properly identify interactions between variables across time thus they believe that the RNN's ability to utilise past observations recurrently is a more suitable approach for the task.

RNNs, are ANNs with recurrent connections to the same or preceding layers which creates a mechanism that is capable of modelling sequential data for sequence recognition and prediction. The hidden states in an RNN stores information from previous states thus creating a memory for the network [19]. This memory mechanism enables RNNs to handle sequential and time-series data properly thus enabling temporal processing [20].

Hu [21] compared the performance of three neural networks; FFMLP, Elman's SRN and Recurrent Fuzzy Neural Network (RFNN). In particular, they found that the RFNN had the best predictive power. Their results validate the appropriateness of RNNs for oil price prediction task as both RNNs outperformed the FFMLP. Some researchers, [22] used Multiple Wavelet Recurrent Neural Network (MWRNN) to predict the Brent and WTI crude oil prices. While, others researchers [23] have applied the Long Short-Term Memory (LSTM) to forecast crude oil prices between July 23, 2007 to February 24, 2017. LSTMs have presented a lead to combat the vanishing gradient problem by incorporating a gating mechanism that uses feedback connections and controls feedback weights [24]. However, LSTMs are ad-hoc techniques with high dimensionality and researchers such as [10] and [25] debate whether there are more optimal architectures for prediction tasks. Additionally, Chen, *et al* [23] argue that LSTMs are not adaptable to new changes that occur when predicting oil prices - we therefore maintain our focus on enhanced variants of the much simpler SRN.

RNNs have been shown to be generally effective for the oil prediction task. However, there are criticisms with the methods currently employed for example; the recurrency in SRNs enables the formation of temporal states. Nevertheless, due to its simple feedback loops, the historical knowledge (past inputs and states) decays rapidly [9]. Tepper [10] states this decay occurs as early as 5 to 10 discrete time steps. In this paper, we therefore seek to determine whether a slightly more sophisticated class of SRN, the MRN, is better able to capture the volatility and structural shocks naturally found in the oil price trajectory and therefore effecting higher predictive accuracy.

[1]Prices fully reflect all available information.

TABLE I
KEY INDICATOR VARIABLES OF OIL PRICES AND SOURCE FOR DATA.

Variable	Data source
Real WTI Crude Oil Price, Month-End Prices	Energy Information Administration, BLS MacroTrends Data Download
Gold Fixing Price 10:30 A.M. (London time) in London Bullion Market, based in U.S. Dollars Average daily price	DataStream
US External Trade: Goods, Deflator/Unit Value of Imports NADJ	DataStream
US Unemployment Rate SADJ	DataStream
US average weekly hours - Total Private Nonfarm VOLA	DataStream

III. METHODOLOGY

A. Data

We use five common indicator variables to predict monthly crude oil prices from July 1969 to March 2015, consisting of 549 observations. (see. Table. I for data source). Monthly data is preferred over daily data as there is less noise and the signal is not obscured which is essential for learning in the network.The change of direction for each variable was calculated, this is indicative of the direction between any given oil price and the month before and this was included as a feature for prediction. The data was divided into training and testing sets, training data accounted for 75% of the data and the remaining 25% of the data was out-of-sample testing. In particular, the in-sample (pre-financial crisis) was from July 1969 to September 2003 and the out-sample was from October 2003 to March 2015.

The data was transformed by standardization, where a variable is transformed into a normal variable using the mean and standard deviation. Muralidharan [26] shows that training the network with standardized data yields better results (note: the change of direction feature was not standardized). The transformed variables and the change of direction as described above are used as inputs for modelling thus for any given time, there are 10 input variables.

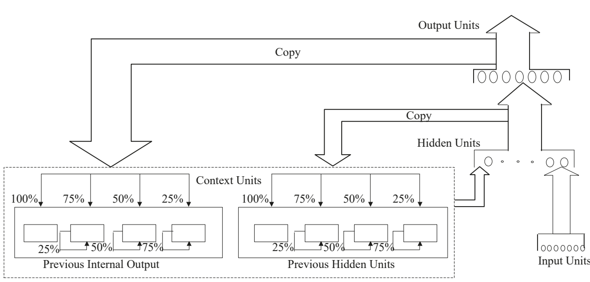

Fig. 1. The architecture of an MRN without noise injection (adapted from [10]).

B. Multi-recurrent Neural Network (MRN)

Multi-recurrent Neural Networks (MRNs) originally proposed by [9] are a class of RNNs with a combination of repeated memory banks with varying strengths. These memory banks comprise of feedback activations from the input, hidden or output layers and previous memory states.

As shown in (1) and (2), the composition is determined by a ratio of layer-level recurrency and self-recurrency within a set of memory banks, with the link recurrency ratio and number of memory banks being key hyper-parameters. The ratio between these links determines whether we have a sluggish or rapid memory and this memory in turn forms the context of the network. Such a combination of different degrees of recurrences enables rigid and flexible learning that effectively captures both variant and invariant properties of the time series [9].

The MRNs dynamic memory state enables it both to forget and retain knowledge which is the catalyst for enhanced performance [9]. In addition, the MRNs memory state enables the formation of a more comprehensive averaged history which is required when solving long term dependency problems [10], [27].

An adapted MRN architecture without noise injection from [10] is illustrated in Fig. 1 employing the following feedbacks:

1) Hidden layer recurrency: the hidden layer is fed back to the input layer
2) Output layer recurrency: the output layer is fed back to the input layer
3) Self-recurrent links for the context units

The MRN functions are as follows, given the inputs at time t, I_t, the hidden units at time $t-1$, H_{t-1} and the hidden memory at time $t-1$, M_{t-1_h} as

$$M_{t_h} = (1/n_h \times H_{t-1}) + ((1 - 1/n_h) \times M_{t-1_h}) \tag{1}$$

where n_h is the total number of hidden memory banks, the output units at time $t-1$, O_{t-1} and the output memory at time $t-1$, M_{t-1_o} as:

$$M_{t_o} = (1/n_o \times O_{t-1}) + ((1 - 1/n_o) \times M_{t-1_o}) \tag{2}$$

where n_o is the total number of output memory banks, and their respective weights to the hidden layer W_{i_h}, $W_{M_{hh}}$ and $W_{M_{oh}}$, the net hidden units at time t are calculated as:

$$\hat{H}_t = \sum W_{i_h} I_t + \sum W_{M_{hh}} M_{t_h} + \sum W_{M_{oh}} M_{t_o} \tag{3}$$

The hidden units at time t are derived by calculating the sigmoid of the net input to the hidden units at time t using the formula below:

$$H_t = f(\hat{H}_t), f(x) = 1/(1 + e^{-x}) \tag{4}$$

and given the hidden units H_t at time t and the hidden to output weights W_{ho}, the output units at time t are calculated as:

$$O_t = \sum W_{h_o} H_t \tag{5}$$

To date, few researchers have successfully applied the MRN for prediction tasks, for example; [9] for weather forecasting, [28] for inflation forecasting, [29] for next word prediction task and [10] for grammar prediction task where it faired favourably with the LSTM. In addition, Tepper [10] also demonstrated how the underlying representations formed by the MRN were superior to those formed by the SRN, NARX and Echo-state Network for a complex grammar induction task, thus providing evidence that the MRN has the potential to model oil prices and improve prediction accuracy. This paper is the first to demonstrate conclusively the superiority of the MRN over other RNNs for the oil-price prediction task.

C. Forecasting Methodology

Sliding Window: A sliding window technique with a shift factor of 1 was used to generate the input sequences (see Fig. 2). These sequences also known as temporal input windows consist of a given number of previous input/output observations along with the current observation and are used as inputs at any given time-step to predict the output. The length of the input window is empirically established for each of the individual models evaluated.

Forecast Horizon: The desired output for any given time step is determined by the forecast horizon, that is how many months ahead the prediction is made.

For this task, the temporal input window size was between 60 and 300 monthly input observations to predict oil prices at forecast horizons of 1, 3, 6 and 12 months ahead. A number of experiments were undertaken to understand the impact on prediction with varying window sizes and horizons for the models.

$$\boxed{x_1, x_2, x_3,} x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, \ldots\ldots\ldots\ldots\ldots \; x_{48}, x_{49}, x_{50} \qquad t_1$$
$$x_1, \boxed{x_2, x_3, x_4,} x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, \ldots\ldots\ldots\ldots\ldots \; x_{48}, x_{49}, x_{50} \qquad t_2$$
$$x_1, x_2, \boxed{x_3, x_4, x_5,} x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, \ldots\ldots\ldots\ldots\ldots \; x_{48}, x_{49}, x_{50} \qquad t_3$$
$$\vdots$$
$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, \ldots\ldots\ldots\ldots\ldots \; \boxed{x_{48}, x_{49}, x_{50}} \qquad t_n$$

Fig. 2. Example of the sliding window technique with 50 input observations

D. Model Training

The MRN and SRN were trained with back-propagation through time [30] for a fixed number of epochs with a decaying learning rate that is, after each epoch, given n_e as the total number of epochs, the learning rate, L is recalculated as:

$$L = L - L/n_e \tag{6}$$

The models are treated as finite memory models and therefore the state memory is reset at the beginning of each input sequence (window). For each time step, a sequence is fed to the network as inputs and the network reads the inputs sequentially (i.e. one month at a time).

Furthermore, a simple ensemble average of 6 models was used as model forecasts[2]. Finally, the models were evaluated using Root Mean Squared Error (RMSE) and Improvement over Random Walk (IORW).

[2]A simple ensemble approach consistently appeared to be more computationally effective than cross-validation for this type of time series task.

IV. Results and Discussion

In this section, the RW, FFMLP, SRN and MRN models will be used for the oil price prediction task and comparative results presented. The best of these models will then be compared with the currently accepted state-of-the-art LSTM model[3]. Various combinations of parameters and hyper-parameters were exploited for the task.

A. FFMLP, SRN and MRN

Random Walk (RW): For a RW model, the best prediction of volatility for any time in the future is the current volatility. The RW model is used as a benchmark.

Feedforward Multilayered Perceptrons (FFMLP): The FFMLP model accounts for the time factor by mapping time onto space, that is a fixed number (depending on the window size) of monthly observations for each feature variable is presented simultaneously as inputs to the FFMLP network. There were 100 hidden units for all the FFMLP models with an initial learning rate of 0.0007 and momentum of 0.0025.

Simple Recurrent Network (SRN): The SRN used is as described in [31], the previous hidden state along with the current observation is fed as inputs at any given time to the network. The SRN models had 20 hidden units, an initial learning rate of 0.01 and a momentum of 0.9999.

Multirecurrent Neural Network (MRN): The MRN combined both hidden and output layer recurrency and similar to the SRN, all the MRN models had 20 hidden units, with an initial learning rate of 0.01 and a momentum of 0.9999.

A descriptive summary of the prediction accuracy for all the models at 4 different horizons (1, 3, 6, 12) with 4 different window sizes (60, 120, 240, 300) is shown below in Fig. 3.

The tables in Fig. 3 show the RMSEs for all the models and IORWs for the neural networks (the best model for each horizon is highlighted in red). In general, the RMSEs for the MRN models are lower and they outperform both the FFMLP and SRN models thus supporting [9] and [10] claim of superior performance. This superior performance is attributed to the MRNs ability to exploit and latch onto past information more effectively which informs its prediction.

In particular, the FFMLP RMSE worsened as the window size increased indicating its poor ability to handle and process temporal data. On the other hand, the SRN in general had a better performance with larger window sizes, particularly, a window size of 240 had better prediction accuracy for both a horizon of 6 and 12 than a window size of 300. These results are indicative of the SRNs limited processing ability due to its simple feedback loops which cause a rapid decay of knowledge [9]. The MRN performed best with the largest window size due to its ability to combine both 'flexible' and 'rigid' memory thus forming stable representations that can latch onto important temporal invariant patterns.

The best models are visualised for each horizon, see Fig. 4 (note the window size varied for the best models). As expected

[3]The LSTM used a dropout regularization technique to reduce over-fitting and an adam optimiser.

	Random Walk			
	t+1	t+3	t+6	t+12
RMSE	6.85602266	15.1886427	23.20913236	26.97517002

Horizon		MLP t+1		t+3		t+6		t+12	
		RMSE	IORW	RMSE	IORW	RMSE	IORW	RMSE	IORW
Window size	60	1.1791	82.8025	1.6810	88.9327	2.2075	90.4886	2.3806	91.1750
	120	1.3955	79.6460	2.4261	84.0270	2.2208	90.4314	2.4131	91.0544
	240	1.3430	80.4110	3.0466	79.9416	2.3965	89.6741	2.3066	91.4491
	300	1.5382	77.5647	3.1619	79.1827	2.3844	89.7266	2.3838	91.1629
RMSE average		1.3639	80.1061	2.5789	83.0210	2.3023	90.0802	2.3710	91.2103

Horizon		SRN t+1		t+3		t+6		t+12	
		RMSE	IORW	RMSE	IORW	RMSE	IORW	RMSE	IORW
Window size	60	0.5591	91.8449	0.9132	93.9879	1.0890	95.3078	1.2670	95.3030
	120	0.5403	92.1187	0.8654	94.3020	1.1037	95.2447	1.2947	95.2003
	240	0.5033	92.6585	0.8824	94.1901	0.9654	95.8402	1.1127	95.8752
	300	0.3761	94.5143	0.8592	94.3431	0.9749	95.7994	1.1997	95.5527
RMSE average		0.4947	92.7841	0.8801	94.2058	1.0333	95.5480	1.2185	95.4828

Horizon		MRN t+1		t+3		t+6		t+12	
		RMSE	IORW	RMSE	IORW	RMSE	IORW	RMSE	IORW
Window size	60	0.5330	92.2262	0.8036	94.7093	1.0160	95.6223	1.2046	95.5344
	120	0.4912	92.8358	0.8726	94.2552	1.0632	95.4193	1.1597	95.7010
	240	0.4904	92.8467	1.0193	93.2888	1.0416	95.5121	1.3474	95.0049
	300	**0.3679**	**94.6346**	**0.7455**	**95.0916**	**0.9585**	**95.8702**	**1.0152**	**96.2365**
RMSE average		0.4706	93.1358	0.8603	94.3362	1.0198	95.6060	1.1817	95.6192

Fig. 3. Descriptive Summary of Results for the test dataset

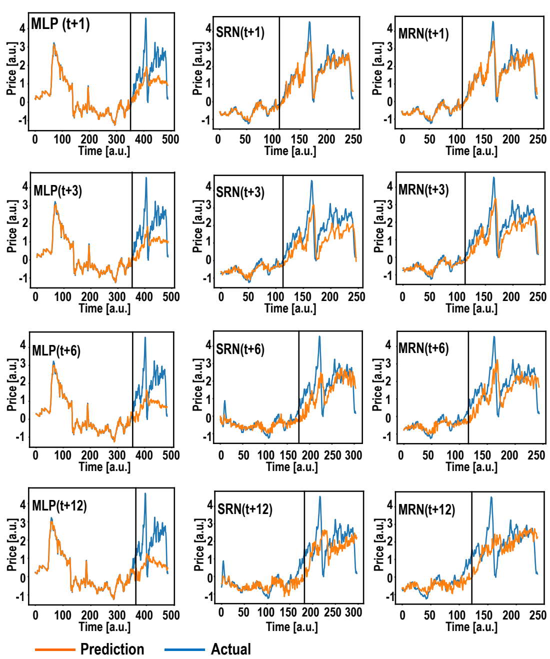

Fig. 4. Crude oil price forecast for the best models for each horizon (the in-sample is on the left of the vertical line and the out-sample on the right.)

the models performed best for the shortest horizon, $t+1$, predicting a month ahead which has the least volatility than the other horizons which are further ahead in time.[4]

The main impact and benefit of a model is seen in its ability to estimate and model unseen data. The graphs in Fig. 4 shows that the FFMLP overfits on the training data (fitting too closely to the training set) but generalizes poorly (deviating when predicting the test set) which is indicative of the lack of robustness in the FFMLP. The SRN and MRN appear to have a good trade-off between over-fitting and generalizing. In particular the MRN, outperformed the SRN as it consistently had both a lower training RMSE and test RMSE thus showing not only its appropriateness for the task but its ability to deal with noise and recognise patterns in the signal better. The MRN's ability to capture temporal dependencies in the evolving oil market and superior performance is consistent with the improved learning, attributed to the varying degrees of its embedded memory.

In addition, we see that the MRN's prediction of crude oil prices provides early indication of the 2008 financial crisis thus showing the ability of the MRN to act as an Early Warning System. This finding in and of itself is significant as the MRN demonstrates that there is important latent signal within the indicator variables to predict the perturbations in oil prices 12 months ahead that would have been invaluable to macroeconomic policy makers such as governments and Central Banks. (Note that the MRNs weights are only tuned up to September 2003, with the outsample data starting from October 2003, showing the robustness of the results.)

B. MRN and LSTM

The LSTM has been widely accepted as the current state-of-the-art. In particular, it has memory blocks and similar to the MRN these blocks utilise self-connections to store the temporal state. The LSTM has gates that control the flow and storage of information and these gates are widely believed to mitigate a common weakness of RNNs; gradient descent problem which results in loss of information. Particularly, the forget gates control what information is forgotten when new information is obtained which is believed to ensure the network retains valuable information which aids learning and better prediction accuracy is achieved [32].

The MRN and LSTM differ mainly in information processing, the MRN utilises and stores all the information in varying capacity that is different strengths of memory thus providing an averaged informed overview whereas the LSTM forgets some information and retains others. Thus the LSTM has a limited overview. In addition, the LSTM suffers from uncontrolled and uncoordinated writing (of information) leading to chaos and overflow [23]. The best MRNs are compared to an LSTM (see. Table. II), the MRNs are substantially simpler in terms of architecture and number of free parameters and achieve a better prediction accuracy.

V. Conclusion

In this paper, four neural network models, the FFMLP, SRN, LSTM and MRN were applied for the crude oil price prediction task. The performance of these models was benchmarked against the random walk of crude oil prices. The results obtained indicate that the MRN is consistently better able to model crude oil prices more robustly than the other models. We further demonstrate that the MRN is an effective

[4]The point where the vertical line appears in the graph is dependent on the window size and horizon which determine the number of observations.

TABLE II
HYPER-PARAMETERS FOR EACH NETWORK AND COMPARATIVE NETWORK PERFORMANCE OF THE MRN AND LSTM.

Horizon	1		3		6		12	
	MRN	LSTM	MRN	LSTM	MRN	LSTM	MRN	LSTM
Layer	3	4	3	4	3	4	3	4
Memory bank	9	0	5	0	8	0	7	0
Total trainable parameters	2624	205,301	1664	52,651	2224	13,826	2144	810,601
Learning Rate	0.01	-	0.01	-	0.01	-	0.01	-
Momentum	0.9999	-	0.9999	-	0.9999	-	0.9999	-
Dropout	-	0.7	-	0.9	-	0.9	-	1
RMSE test set	0.36785	0.4244	0.74552	0.70107	0.9585	0.99568	1.01521	1.15299

and comparatively simple modelling technique, compared to the more complex state-of-the-art LSTM. In addition, it is a strong candidate for an early warning model of impending financial crises. The results show that there was important information in the indicator variables 4-5 years before the crisis. Additionally, varying window sizes and horizons were exploited to identify the impact on the models performance accuracy. Given the positive outcome of our experiments, future work will involve further optimizing the gradient descent learning algorithm to better utilize the architecture of the MRN to determine the sluggish nature of the representations formed to identify temporal patterns and provide early warning signals for medical applications. Furthermore, the MRN will be rigorously compared to other state-of-the-art techniques for such early warning systems with an overriding evaluation principles of parsimony, simplicity and interpretability.

ACKNOWLEDGMENT

Oluwatamilore Orojo was supported by Nottingham Trent University Vice-Chancellor's Doctoral Scholarship.

REFERENCES

[1] U.S. Energy Information Administration, "Short-term energy outlook." [Online]. Available: https://www.eia.gov/outlooks/steo/report/global_oil.php
[2] N. Bashiri Behmiri and J. R. Pires Manso, "Crude Oil Price Forecasting Techniques: A Comprehensive Review of Literature," *SSRN Electron. J.*, 2013.
[3] M. Hamdi and C. Aloui, "Forecasting crude oil price using artificial neural networks: A literature survey," *Econ. Bull.*, vol. 35, no. 2, pp. 1339–1359, 2015.
[4] L. A. Gabralla, R. Jammazi, and A. Abraham, "Oil price prediction using ensemble machine learning," in *Proc. ICCEEE 2013*, Aug. 2013, pp. 674–679.
[5] S. Deng and A. Sakurai, "Crude Oil Spot Price Forecasting Based on Multiple Crude Oil Markets and Timeframes," *Energies*, vol. 7, no. 5, pp. 2761–2779, Apr. 2014.
[6] M. M. Mostafa and A. A. El-Masry, "Oil price forecasting using gene expression programming and artificial neural networks," *Econ. Model.*, vol. 54, pp. 40–53, Apr. 2016.
[7] G. S. Natarajan and A. Ashok, "Multivariate Forecasting of Crude Oil Spot Prices using Neural Networks," Nov. 2018, arXiv: 1811.08963.
[8] U. Güçlü and M. A. J. van Gerven, "Modeling the dynamics of human brain activity with recurrent neural networks," *Front. Comput. Neurosci.*, vol. 11, Feb. 2017, arXiv: 1606.03071.
[9] C. Ulbricht, "Multi-recurrent Networks for Traffic Forecasting," *Proceedings of the National Conference on Artificial Intelligence*, vol. 2, pp. 883–888, 1994.
[10] J. A. Tepper, M. S. Shertil, and H. M. Powell, "On the importance of sluggish state memory for learning long term dependency," *Knowl. Based Syst.*, vol. 96, pp. 104–114, Mar. 2016.
[11] H. Mohammadi and L. Su, "International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models," *Energy Econ.*, vol. 32, no. 5, pp. 1001–1008, Sep. 2010.
[12] P. Sadorsky, "Modeling and forecasting petroleum futures volatility," *Energy Econ.*, vol. 28, no. 4, pp. 467–488, Jul. 2006.
[13] S. Moshiri and F. Foroutan, "Forecasting Nonlinear Crude Oil Futures Prices," *Energy J.*, vol. 27, no. 4, pp. 81–95, Oct. 2006.
[14] A. Shabri and R. Samsudin, "Daily Crude Oil Price Forecasting Using Hybridizing Wavelet and Artificial Neural Network Model," *Mathematical Problems in Engineering*, vol. 2014, pp. 1–10, 2014.
[15] A. Lee, Z. Geem, and K.-D. Suh, "Determination of Optimal Initial Weights of an Artificial Neural Network by Using the Harmony Search Algorithm: Application to Breakwater Armor Stones," *Appl. Sci.*, vol. 6, no. 6, p. 164, May 2016.
[16] M. A. Wani, "Comparative Study of Back Propagation Learning Algorithms for Neural Networks," *J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 12, p. 6, 2013.
[17] T. Xiong, Y. Bao, and Z. Hu, "Beyond one-step-ahead forecasting: Evaluation of alternative multi-step-ahead forecasting models for crude oil prices," *Energy Econ.*, vol. 40, pp. 405–415, Nov. 2013.
[18] L. Yu, S. Wang, and K. K. Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," *Energy Econ.*, vol. 30, no. 5, pp. 2623–2635, Sep. 2008.
[19] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent Advances in Recurrent Neural Networks," Dec. 2017, arXiv: 1801.01078.
[20] J. A. Bullinaria, "Recurrent Neural Networks,," p. 20, 2015. [Online]. Available: http://www.cs.bham.ac.uk/ jxb/INC/ l12.pdf
[21] J. W.-S. Hu, Y.-C. Hu, and R. R.-W. Lin, "Applying Neural Networks to Prices Prediction of Crude Oil Futures," *Math. Probl. Eng.*, vol. 2012, pp. 1–12, 2012.
[22] T. Mingming and Z. Jinliang, "A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices," *J. Econ. Bus.*, vol. 64, no. 4, pp. 275–286, Jul. 2012.
[23] Y. Chen, K. He, and G. K. Tso, "Forecasting Crude Oil Prices: a Deep Learning based Model," *Procedia Comput. Sci.*, vol. 122, pp. 300–307, 2017.
[24] C. Tallec and Y. Ollivier, "Can recurrent neural networks warp time?" Mar. 2018, arXiv: 1804.11188.
[25] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. ICML 2015*, vol. 37, 2015, pp. 2342–2350.
[26] K. Muralidharan, "A note on transformation, standardization and normalization," *Int. J. Oper. Quant. Manage.*, vol. IX, no. 1 & 2, pp. 116–122, 2010.
[27] G. Dorffner, "Neural Networks for Time Series Processing," 1996.
[28] J. Binner, P. Tino, J. Tepper, R. Anderson, B. Jones, and G. Kendall, "Does money matter in inflation forecasting?" *Physica A*, vol. 389, no. 21, pp. 4793–4808, Nov. 2010.
[29] M. S. Shertil, "On the Induction of Temporal Structure by Recurrent Neural Networks," Ph.D. dissertation, Nottingham Trent University, 2014.
[30] P. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.
[31] J. L. Elman, "Finding Structure in Time," *Cognitive Science*, vol. 14, pp. 179–211, 1990.
[32] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling," p. 5.