Forecasting Exchange Rate Value at Risk using Deep Belief Network Ensemble based Approach

He, Kaijian; Ji, Lei; Tso, Geoffrey K.F.; Zhu, Bangzhu; Zou, Yingchao

Published in:
Procedia Computer Science

Published: 01/01/2018

Published version (DOI):
10.1016/j.procs.2018.10.213

Publication details:
He, K., Ji, L., Tso, G. K. F., Zhu, B., & Zou, Y. (2018). Forecasting Exchange Rate Value at Risk using Deep Belief Network Ensemble based Approach. *Procedia Computer Science*, *139*, 25-32. https://doi.org/10.1016/j.procs.2018.10.213

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 139 (2018) 25–32

Procedia Computer Science

www.elsevier.com/locate/procedia

The International Academy of Information Technology and Quantitative Management, the Peter Kiewit Institute, University of Nebraska

Forecasting Exchange Rate Value at Risk using Deep Belief Network Ensemble based Approach

Kaijian He[a,b,*], Lei Ji[b], Geoffrey K.F. Tso[c], Bangzhu Zhu[d], Yingchao Zou[e]

[a]*Hunan Engineering Research Center for Industrial Big Data and Intelligent Decision Making, Hunan University of Science and Technology, Xiangtan 411201, China*

[b]*School of Business, Hunan University of Science and Technology, Xiangtan, 411201, China*

[c]*Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong*

[d]*Business School, Nanjing University of Information Science & Technology, Nanjing 210044,China*

[e]*College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*

Abstract

In this paper, we propose a new Value at Risk estimate based on the Deep Belief Network ensemble model with Empirical Mode Decomposition (EMD) technique. It attempts to capture the multi-scale data features with the EMD-DBN ensemble model and predict the risk movement more accurately. Individual data components are extracted using EMD model while individual forecasts can be calculated at different scales using ARMA-GARCH model. The DBN model is introduced to search for the optimal nonlinear ensemble weights to combine the individual forecasts at different scales into the ensembled exchange rate VaR forecasts. Empirical studies using major exchange rates confirm that the proposed model demonstrates the superior performance compared to the benchmark models.

Keywords:
Exchange rate forecasting; Empirical Mode Decomposition; Deep Belief Network; Value at Risk
JEL: F31; C45; C53

1. Introduction

Exchange market is one of the most lucrative and actively traded financial markets. Exchange rate determined by the market supply and demand for currencies also reflects the influencing factors in the exchange market. It is one of the key macroeconomic factors linking the global financial system [1, 2]. Due to the global exposure of the

* Corresponding author. Tel.: +86-731-58290306 ; fax: +86-731-58290306.
E-mail address: kaijian.he@my.cityu.edu.hk

10.1016/j.procs.2018.10.213

exchange markets to wide range of investors, significant price fluctuations and risks have been observed in the market trading. However, the interaction among these factors are so complicated that it exceeds the computational boundary of the investors to conduct exhaustive investigations into the relationship between them and forecast the exchange rate movement accurately. The perceived market risk can be attributed to a diverse range of influencing factors, mostly concerning the macroeconomic indicator, etc. [1, 2]. The accurate risk measurement and modeling critically affect the theoretical research in international economics and finance such as the derivative pricing risk measurement.

In this paper, the downside risk in the exchange market is measured using the widely adopted Value at Risk (VaR) concept. VaR is defined as the particular quantile value of the exchange rate distributions over the given time horizon at the specified confidence level. VaR estimation models are usually classified into the parametric and non-parametric categories in the mainstream literature [3]. The parametric models include Exponential Weighted Moving Average model, GARCH model while the non-parametric models include historical simulation and Monte Carlo simulation models [3]. It summarizes the downside risk over a particular investment period into a single statistic number [3]. To estimate VaR more reliably, researches have been directed toward analyzing the fundamental and technical factors behind the exchange rate fluctuations, with the aims to exploit the key risk factors for more accurate risk modeling. Researches on analyzing and exploiting both factors for the risk measurement purposes have received continuous research attentions in the literature. But since the seminal work by [4], it has been shown that the major econometric models failed to predict the exchange rate changes better than the naive Random Walk (RW) model. So far there is no conclusive evidence on which exchange rate model is the best method to model the complex fluctuations in the exchange rate movement. Empirical studies show that the influence factors in the exchange markets are difficult to identify and they typically have nonlinear dynamic influences on the exchange rate movement over time. This gives rise to the necessity and renewed interests to model and control the significant risk in the exchange markets.

Recently multiscale ensemble models have received significant research interests in the literature [5, 6]. Current approaches have used the wavelet analysis and Empirical Mode Decomposition to decompose the financial data into its constituent components and use the ensemble models to integrate forecasts in those constituent components [6, 7]. These models develop with the recent advancement of two recent researches development, i.e. EMD model and Artificial Intelligence (AI) ensemble model. On one hand, EMD was first proposed by Huang in 1998 and quickly found its applications in geophysics, vibration engineering, biomedicine, and other fields [8, 5]. [9] introduced EMD as a new tool for analyze the statistical characteristics of nonlinear and non-stationary data in the financial market. On the other hand, AI based ensemble approaches have served as a complementary approach to exchange rate modeling. Given the lack of definitive analytical tools for the market dynamics, AI can be used to model the dynamically changing weights in the ensemble models for risk estimate [10, 11, 12]. Till now limited positive performance have highlighted its significance. For example, empirical studies in the exchange rate and crude oil forecasting field have shown the higher level of forecasting accuracy [7].

As far as the AI model is concerned, the recent emergence of Deep Learning (DL) model and its success in physical disciplines compared to traditional neural network model showcase its potential to enhance the ensemble risk modeling accuracy in the exchange rate. For example, [13] used the deep belief restricted Boltzmann machine to learn and model the complex relationship between the real estate unit sales price and large number of economic variables. They found the improved accuracy and reduced computing time for their proposed approach. [14] used the deep belief network to model and forecast the hourly electricity load and found the improved forecasting accuracy. [15] used the deep neural networks to forecasting the S&P 500 stock index movement. The improved forecasting accuracy shows that there is exploitable forecasting opportunities in the stock market. Till now only limited exploratory attempts using the DBN model have been conducted in the exchange markets, mostly for the forecasting purpose. Positive performance improvements have already accumulated, etc. For example, [16] found that DBN model provided the superior exchange rate forecasting accuracy than the neural network model.

In this paper, we propose a EMD-DBN ensemble Value a Risk model that takes advantage of the DBN ensemble model and EMD model. It relies on the EMD model to recover the influencing factors for the exchange markets at different scales. It uses the DBN model to integrate the risk forecasts of a series of extracted constituent factors in the multi scale domain to produce the optimal risk estimate. We have conducted the empirical analysis applying the proposed model in the exchange markets. Experiment results show the superior performance of the proposed model and the validity of the extracted multi scale components. The major contribution of the paper is the proposition of DBN ensemble risk estimate model with EMD model. It introduces DBN model in the risk modeling to take full advantage of its nonlinear adaptive learning ability. As the DBN model is known for its capability to extract the nonlinear data pattern, it is used as the basis for the nonlinear ensemble model. As the DBN model is known to be less sensitive to the noise disruption, it contributes to the risk measurement literature to further performance improvement in the multi scale risk forecasts.

The rest of the paper is organized as follows. Section 2 provides a brief illustration of the DBN model. It further explains in detail the value at risk estimation model using the DBN ensemble model with EMD technique. Section 3 explains the results of the empirical studies useful using the comprehensive exchange rate dataset to evaluate the model performance. Section 4 summarizes the findings of the paper.

2. Methodology

2.1. Deep Belief Network (DBN)

Deep Learning (DL) has attracted significant research attentions from both academics and industry since it was originally proposed in 2006 [17]. Deep Belief Network (DBN) is a generative graphical model in the general deep learning framework. Like traditional shallow neutral network, the neurons in DBN have activation function and process the information. As the number of layers increases rapidly from the shallow neural network to the deep neural network, the number of the parameters involved in training deep neural network increases exponentially. To reduce the over fitting problem, the pre-training using a stack of a number of restricted Boltzmann machines(RBMs) is introduced in DBN. RBM is used as the pre-training model to learn the hidden data feature in an unsupervised learning process. A RBM is a neural network that consists of two layers called visible (input) layer and hidden layer. In RBM, neurons in the different levels of layers have mutual undirected connection while the neurons in the same level of layers are independent [18]. In the structure of RBM , v and h represent the state of the neurons of visible layer and hidden layer respectively. When a number of RBMs are stacked in DBN, the hidden layer of the previous RBM serves as the visible layer of the following RBM [19].

The typical structure of DBN is illustrated in figure 1 [17, 20, 21].

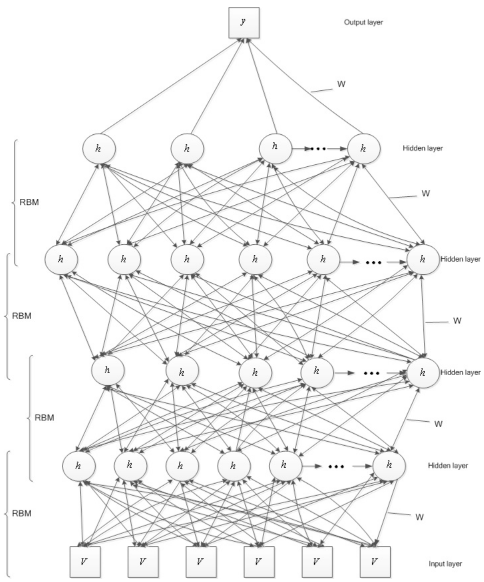

Fig. 1. The general architecture of DBN

The training process of DBN model can usually be divided into pre-training phase with an unsupervised learning process and fine tuning phase with a supervised learning process. In the pre-training phase ,the initial weights of the network are optimized so that the input data can be reconstructed. That way, the important data features in the input data are extracted in the RBM. The RBMs are sequantially trained with an algorithm called greedy layer-wise unsupervised learning algorithm. In the fine-tuning phase, some label units are added to the topmost layer from the pre-training phase. The initial weights from the pre-training phase are trained and adjusted by back-propagation algorithm in a supervised learning process. The global optimal weights of deep belief network are obtained [22, 23].

In the first pre-training step, the parameters of RBM is updated by the process of unsupervised learning. The weights of whole RBMs of DBN can be updated sequentially with a greedy layer-wise pre-training process. Given that v_i and h_j as the state of visible unit i and hidden unit j, the probability distribution of the unit is calculated by the entropy function since the RBM is an energy-based model [24]. The energy function is calculated as in Eq. (1) [20].

$$E(v,h;\theta) = -\sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij}^{R} v_i h_j - \sum_{i=1}^{M} a_i v_i - \sum_{j=1}^{N} b_j h_j \tag{1}$$

Where w_{ij}^{R} is the connecting weight that connects the neurons i in the visible layer and the neurons j in the hidden layer. The a_i is the bias of the visible neuron i, the b_j is the bias of the hidden neuron j. θ is the parameter set that $\theta = \{W, a, b\}$. The M and N are represent the number of neurons in the visible layer and hidden layer [25].

In RBM, the joint probability distribution of RBM with respect to units in the visible layer and hidden layer can be defined by the energy function as in Eq. (2) [20].

$$p(v,h;\theta) = \frac{e^{-E(v,h;\theta)}}{Z} \tag{2}$$

Where Z is a partition function that used for normalization $Z = \sum_{v,h} e^{-E(v,h;\theta)}$.

Because the joint probability distribution of RBM can be calculated as in Eq. (2), the marginal probability of visible neurons in RBM is calculated by summing the probability distribution of the all hidden units, as in Eq. (3) [20].

$$P(v;\theta) = \sum_{h} p(v,h;\theta) = \frac{\sum_h e^{-E(v,h;\theta)}}{Z} \tag{3}$$

To update the weight, the derivative of Eq. (4) is taken with respect to the weight. The weights of RBM is updated as in Eq. (4) [14].

$$\Delta w_{ij}^{R} = \mu\left(\frac{\partial \log p(v;\theta)}{\partial w_{ij}^{R}}\right) = \mu(E_{data}(v_i h_j) - E_{model}(v_i h_j)) \tag{4}$$

Where μ denotes the learning rate, $E_{data}(v_i h_j)$ represents the expected values of the observed data in the training data. $E_{model}(v_i h_j)$ represents the expected values of the sampling data in the model. In the literature, Gibbs sampling is usually used to calculate the approximate value for $E_{model}(v_i h_j)$ [20]. For example, Contrastive Divergence algorithm based on the Gibbs sampling has been shown to be an effective approach [22]. Through the above training process, the weights of the RBM can be update [22].

The whole initial weights of DBN can be updated by sequential train RBMs using the above method. After the unsupervised pre-training step, the supervised fine-tuning step with back-propagation algorithm is used to adjust the weights slightly obtained from unsupervised learning process to obtain the optimal weights of DBN.

2.2. a Deep Belief Network Ensemble Value at Risk Estimate

We construct an ensemble Value at Risk model based on Empirical Mode Decomposition (EMD) and the Deep Belief Network (DBN) model. It consists of a series of steps to transform and model the data characteristics.

(1) Firstly, we assume that there is a multi scale risk structure in the exchange rate movement. Some of the risk factors have simultaneous impact across different markets while some of the risk factors have limited impact on single market. Given exchange rate time series data $x_i, i = 1, 2, ..., N$, EMD model is used to calculate the decomposed components IMF at different scales up to the maximal scale N and the residual e, as in Eq. (5).

$$x_t = \sum_{i=1}^{N} IMF_{i,t} + e_t \tag{5}$$

This is one of the key step in the model. It transforms the exchange rate data into a series of data components, upon which further ensemble modeling is performed. The multi scale transformation is done using the Empirical mode decomposition (EMD) model. The Empirical Mode Decomposition (EMD) method decomposes the original signal into a number of mutually independent Intrinsic Mode Functions (IMFs) through the sifting process [8]. Unlike wavelet analysis, no basis function needs to be set during the decomposition process. Compared with wavelet analysis, the empirical and adaptive nature of the EMD basis make it more suitable for accurately describing the physical characteristics of the signal when dealing with non-stationary nonlinear signals. It decomposes the data into several intrinsic mode functions (IMF) at different scales with unique characteristic. Although the empirical approach of EMD decomposition can't guarantee the mutual independence of IMFs mathematically, the correlation between IMFs are supposed to be very low in the sifting process. When applied to the exchange rate modeling, IMFs at different scales represent the influence of different investment factors on the market volatility and risk exposure.

(2) Secondly, we estimate the ARMA-GARCH model for individually extracted data components using the model training data. The conditional mean and conditional standard deviation are forecasted using ARMA-GARCH model, with normal assumption [26]. The model lags for both ARMA and GARCH models are determined using the model tuning data set by minimizing the AIC information criteria [26].

(3) Thirdly, we assume that there is nonlinear relationship between the individual risk factors and the total risk estimate. We use the DBN based nonlinear ensemble model to combine individual forecasts of both conditional mean and conditional standard deviation at different scales. This would produce the ensemble forecasts of conditional mean and conditional standard deviation.

(4) Fourthly, with the estimated ensemble conditional mean and conditional standard deviation, we forecast the future VaR using the standard parametric approach [3]. Assuming the normal distribution, VaR at confidence level cl is calculated as in Eq. (6) [3].

$$VaR_a = -\mu + \sigma Z_a \tag{6}$$

Where $a = 1 - cl$. Z_a is the normal variate at quantile a.μ and σ is the conditional mean and conditional standard deviation.

3. Empirical Studies

In this section, we conducted the empirical investigation of the performance of the proposed model against the benchmark model. ARMA-GARCH model is used as the benchmark model since it is one of the most robust models that provides the baseline performance in the literature [7]. During the empirical evaluation, we have used the extensive market dataset across major exchange markets. These exchange rates include Australian Dollar against Dollar (AUD), Dollar against Canadian Dollar (CAD), Dollar against Swiss Franc (CHF), and Euro against Dollar (EURO). All data were downloaded from the Quandl, which stores a large number of data from different sources and makes it easy for investors to access the data publicly [27].

The dataset covers the period from 23, July 2007 to 3, August 2018. It contains 2772 daily observations. When constructing the dataset, the original data has been transformed using the log difference transformation. In this study, the first 70% of the dataset was used as the model training set, to estimate the ARMA-GARCH model. The rolling window was set to the length of the training set. The holding period was assumed to be 1 day. The rolling time window was moved one step ahead each time. The next 21% of the dataset is used as the model tuning dataset, to train and estimate the DBN model. The remaining 9% of the dataset, i.e. 250 observations, was reserved as model test dataset, to evaluate the model performance. The historical data is decomposed into 8 groups of mutually independent IMFs using the EMD method. The software implementations include tensorflow firm Google, the Python, R and the matlab software, as well as various packages including Keras, numpy,, etc [28]. The DBN is implemented using the software packages provided by [29].

Firstly we calculate the descriptive statistics and statistical test of data distribution characteristics. Results are listed in table 1.

Table 1. Descriptive statistics and statistical tests

Statistics	AUD	CAD	CHF	Euro
Mean(x10-5)	-7.2658	9.6327	-10.6395	-13.6789
Standard Deviation	0.0099	0.0068	0.0080	0.0067
Skewness	-0.6691	-0.0954	-1.7297	0.1900
Kurtosis	15.2254	8.7051	48.4565	6.0503
JB Test	0.001	0.001	0.001	0.001
BDS Test	0	0	0	0.0109

Results in Table 1 show that the exchange rate movement is characterized by non-normal distribution, with complex nonlinear behavior. The rejection of the null hypothesis for Jarque-Bera test confirms the non-normal distribution of the exchange rate movement. The rejection of the null hypothesis of BDS tests confirms that the exchange rate movement are not independently distributed. AUD market has the highest level of standard deviation while Euro market has the lowest level of standard deviation. Exchange rate has negative skewness in all exchange markets expect for Euro exchange market. Heavy tails with prevalence of extreme or transient events are observed in exchange rate distribution in all exchange markets as indicated by signfiicant kurtosis value. Among four exchange markets, the kurtosis value is especially large in the Swiss Frac market.

Then we conduct the experimenting to forecast the one step ahead VaR using different models. When estimating the DBN model, the empirical volatility is chosen as the proxy for the latent volatility level. It was calculated over a one year running time window, i.e. 250 trading days. The VaR model performance is evaluated using the back testing procedure [3]. The main performance measure is the number of exceedance calculated from the back testing procedure. We also calculate the p values of the test statics for the unconditional coverage test using the number of exceedances [3]. For benchmark ARMA-GARCH model, lag one is used and the model form is ARMA(1,1) - GARCH(1,1). As for DBN network, the network structure is 1 input layer, 3 hidden layers and 1 output layer, with 8-50-50-50-1 structure. For the initial tuning phase, the learning rate for is 0.005 and the maximal number of iterations is 40. For the fine tuning phase, the learning rate is 0.01 and the maximal number of iterations is 300. The activation function is Rectified Linear Unit (ReLU). The parameters for DBN model is trained on a interval of 126 trading days, i.e. the network is trained twice a year.

The forecasting accuracy in terms of different performance measures are listed in table 2.

Table 2. The model performance using the out-of-sample test data

Models	N_{AUD}	N_{CAD}	N_{CHF}	N_{Euro}	p_{AUD}	p_{CAD}	p_{CHF}	p_{Euro}
$AG_{95\%}$	13	16	7	8	0.8853	0.3294	0.0828	0.1632
$AG_{97.5\%}$	8	11	5	5	0.4965	0.0817	0.6000	0.6000
$AG_{99\%}$	4	4	3	1	0.3805	0.3805	0.7580	0.2781
$AG_{Average}$	8.3333	10.3333	5	4.6667	0.5875	0.2639	0.4803	0.3471
$DBN_{95\%}$	15	14	9	10	0.4812	0.6691	0.2860	0.4529
$DBN_{97.5\%}$	7	11	7	6	0.7656	0.0817	0.7656	0.9188
$DBN_{99\%}$	3	6	3	2	0.7580	0.0594	0.7580	0.7419
$DBN_{Average}$	8.3333	10.3333	6.3333	6	0.6683	0.2701	0.6032	0.7045

Where AG_{cl}, DBN_{cl},$cl = \{95\%, 97.5\%, 99\%, Average\}$ refers to VaR estimate using ARMA-GARCH model and Deep Belief Network model at cl confidence level. The $Average$ refers to the average value of VaR estimates at cl confidence levels.

In terms of the average p value, both the proposed model and ARMA-GARCH model have passed the unconditional coverage test as no p value is less than 0.05 at all confidence levels. As for ARMA-GARCH model, the risk forecasts are more aggressive in AUD and CAD markets since the number of exceedances are less than the theoretical values. They are more conservative in CHF and Euro markets since the number of exceedances are larger than the theoretical values. Compared with the benchmark model ARMA-GARCH model, risk forecasts produced by DBN based model are more conservative in AUD and CAD markets at 95% and 97.5% confidence levels and more aggressive in CHF and Euro markets at all confidence levels. Thus the DBN based model produces generally more optimal forecasts. In terms of the average p value, the DBN based model performs better than ARMA-GARCH model. We found that using the DBN model leads to the best out-of-sample forecasting performance. The use of nonlinear ensemble integrating forecasts at 8 different scales shows that forecasts at different scales do not contribute the same weight in the nonlinear ensemble algorithm. In the meantime, the DBN model is trained on a less regular basis. The model parameters remain the same between two different trainings. The resulting improved risk estimation accuracy implies that multiscale structure for the contribution of forecasts at different scales are stable over a half year period.

4. Conclusions

In this paper, we have proposed a DBN based nonlinear ensemble model with EMD technique, for estimating value at risk. DBN has been used to aggregate the ensemble multiscale risk forecasts in the foreign exchange market. Positive performance improvement in risk estimates have been observed. Results in this paper have demonstrated that the use of DBN model could identify more optimal ensemble weights and better integrate the partial information from extracted risk estimates. This approach leads to the forecasting accuracy improvement of the multiscale VaR estimate models. Work in this paper implies that new innovative deep learning model can take into account the domain knowledge in the risk estimate field to gain better insights in the risk estimate field and achieve the improved forecasting accuracy.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC No. 71671013), and Humanities and Social Sciences Youth foundation of Ministry of Education of China (No. 16YJC790026).

References

[1] A. Abhyankar, L. Sarno, G. Valente, Exchange rates and fundamentals: Evidence on the economic value of predictability, Journal of International Economics 66 (2) (2005) 325–348.

[2] K. Chen, S. Zhang, What's news in exchange rate dynamics: A {DSGE} approach, Economics Letters 134 (2015) 133 – 137.

[3] K. Dowd, Measuring Market Risk, John Wiley & Sons Inc., Chichester, England; Hoboken, NJ, 2005, 9780470013038; Kevin Dowd.; :ill. ;25 cm. +1 CD-ROM (4 3/4 in.; Includes bibliographical references (p. [365]-377) and index.
URL http://lib.cityu.edu.hk/record=b1977867;http://www.loc.gov/catdir/toc/ecip0511/2005010796.html

[4] R. A. Meese, K. Rogoff, Empirical exchange-rate models of the seventies - do they fit out of sample, Journal of International Economics 14 (1-2) (1983) 3–24.

[5] L. Yu, S. Wang, K. K. Lai, Forecasting crude oil price with an emd-based neural network ensemble learning paradigm, Energy Economics 30 (5) (2008) 2623–2635. doi:10.1016/j.eneco.2008.05.003.

[6] K. HE, Ensemble forecasting of value at risk via multi resolution analysis based methodology in metals markets, Working paper, City University of Hong Kong (2010).

[7] K. He, C. Xie, S. Chen, K. K. Lai, Estimating var in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network, Neurocomputing 72 (16-18, Sp. Iss. SI) (2009) 3428–3438. doi:10.1016/j.neucom.2008.09.026.

[8] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454 (1971) (1998) 903–995.

[9] N. E. Huang, M.-L. C. Wu, S. R. Long, S. S. P. Shen, W. Qu, P. Gloersen, K. L. Fan, A confidence limit for the empirical mode decomposition and hilbert spectral analysis, Proceedings: Mathematical, Physical and Engineering Sciences 459 (2037) (2003) 2317–2345.

[10] A. Das, P. Das, Chaotic analysis of the foreign exchange rates, Applied Mathematics and Computation 185 (1) (2007) 388 – 396.

[11] P. Caraiani, E. Haven, Evidence of multifractality from {CEE} exchange rates against euro, Physica A: Statistical Mechanics and its Applications 419 (2015) 395 – 407.

[12] L. Yu, S. Y. Wang, K. K. Lai, A novel nonlinear ensemble forecasting model incorporating glar and ann for foreign exchange rates, Computers & Operations Research 32 (10) (2005) 2523–2541.

[13] R. Hrasko, A. G. Pacheco, R. A. Krohling, Time series prediction using restricted boltzmann machines and backpropagation, Procedia Computer Science 55 (2015) 990 – 999, 3rd International Conference on Information Technology and Quantitative Management, ITQM 2015. doi:https://doi.org/10.1016/j.procs.2015.07.104.

[14] A. Dedinec, S. Filiposka, A. Dedinec, L. Kocarev, Deep belief network based electricity load forecasting: An analysis of macedonian case, Energy 115, Part 3 (2016) 1688 – 1700, sustainable Development of Energy, Water and Environment Systems.

[15] C. Krauss, X. A. Do, N. Huck, Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500, European Journal of Operational Research 259 (2) (2017) 689 – 702. doi:https://doi.org/10.1016/j.ejor.2016.10.031.

[16] F. Shen, J. Chao, J. Zhao, Forecasting exchange rate using deep belief networks and conjugate gradient method, Neurocomputing 167 (2015) 243 – 253.

[17] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504.

[18] Z. Zhang, Q. He, J. Gao, M. Ni, A deep learning approach for detecting traffic accidents from social media data, Transportation Research Part C: Emerging Technologies 86 (2018) 580 – 596. doi:https://doi.org/10.1016/j.trc.2017.11.027.

[19] D. Y. Kim, H. Y. Song, Method of predicting human mobility patterns using deep learning, Neurocomputing 280 (2018) 56 – 64, applications of Neural Modeling in the new era for data and IT. doi:https://doi.org/10.1016/j.neucom.2017.07.069.

[20] J. Qiao, G. Wang, W. Li, X. Li, A deep belief network with plsr for nonlinear system modeling, Neural Networks 104 (2018) 68 – 79. doi:https://doi.org/10.1016/j.neunet.2017.10.006.

[21] C. Luo, D. Wu, D. Wu, A deep learning approach for credit scoring using credit default swaps, Engineering Applications of Artificial Intelligence 65 (2017) 465 – 470. doi:https://doi.org/10.1016/j.engappai.2016.12.002.

[22] G. Fu, Deep belief network based ensemble approach for cooling load forecasting of air-conditioning system, Energy 148 (2018) 269 – 282. doi:https://doi.org/10.1016/j.energy.2018.01.180.

[23] L. Yu, R. Zhou, L. Tang, R. Chen, A dbn-based resampling svm ensemble learning paradigm for credit classification with imbalanced data, Applied Soft Computing 69 (2018) 192 – 202. doi:https://doi.org/10.1016/j.asoc.2018.04.049.

[24] M. H. Rafiei, H. Adeli, A novel machine learning model for estimation of sale prices of real estate units, Journal of Construction Engineering and Management 142 (2) (2016) 04015066.

[25] Q. Zhang, L. T. Yang, Z. Chen, P. Li, A survey on deep learning for big data, Information Fusion 42 (2018) 146 – 157. doi:https://doi.org/10.1016/j.inffus.2017.10.006.

[26] C. Brooks, Introductory Econometrics for Finance, 2nd Edition, Cambridge [England] ; New York : Cambridge University Press, 2008.

[27] Quandl, EIA and OPEC various end-of-day data (2018).
URL https://www.quandl.com/data/EIA-U-S-Energy-Information-Administration-Data,https://www.quandl.com/data/OPEC/ORB-OPEC-Crude-Oil-Price

[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org (2015).
URL https://www.tensorflow.org/

[29] albertbup, A python implementation of deep belief networks built upon numpy and tensorflow with scikit-learn compatibility (2017).
URL https://github.com/albertbup/deep-belief-network