A Deep Coupled LSTM Approach for USD/CNY Exchange Rate Forecasting

Wei Cao
HeFei University of Technology

Weidong Zhu
HeFei University of Technology

Wenjun Wang
HeFei University of Technology

Yves Demazeau
CNRS-LIG, University Grenoble Alps

Chen Zhang
HeFei University of Technology

***Abstract*—Forecasting CNY exchange rate accurately is a challenging task due to its complex coupling nature, which includes market-level coupling from interactions with multiple financial markets, macrolevel coupling from interactions with economic fundamentals, and deep coupling from interactions of the two aforementioned kinds of couplings. This study develops a new deep coupled long short-term memory (LSTM) approach, namely, DC-LSTM, to capture the complex couplings for USD/CNY exchange rate forecasting. In this approach, a deep structure consisting of stacked LSTMs is built to model the complex couplings. The experimental results with 10 years data indicate that the proposed approach significantly outperforms seven other benchmarks. The DC-LSTM is verified to be a useful tool to make wise investment decisions through a profitability discussion. The purpose in this article is to clarify the importance of coupling learning for exchange rate forecasting, and the usefulness of deep coupled model to capture the couplings.**

CNY EXCHANGE RATE forecasting is of great importance due to the following reasons. First, the CNY exchange rate could highly affect Chinese economic development. In 2018, the overall exports and imports of goods and services stands at 39.78% of the Chinese GDP (http://data.stats.gov.cn/staticreq.htm), which means that exchange rate fluctuations have severe economic implications. Second, the promoted

Digital Object Identifier 10.1109/MIS.2020.2977283
Date of publication 28 February 2020; date of current version 12 May 2020.

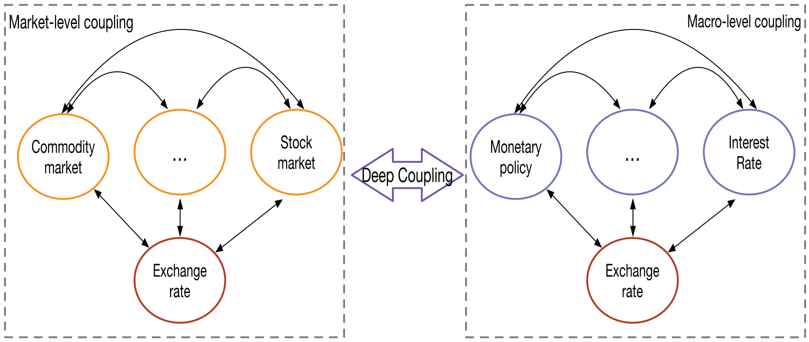


Figure 1. Overview of complex coupled nature of CNY exchange rate.

internationalization of CNY attracts attention from global investors, risk managers, and governments. On October 1, 2016, the official incorporation of CNY into the special drawing right (SDR) currency basket of the International Monetary Fund has been made. The ratio of CNY reserves to world total foreign exchange reserves increased from 1.23% in 2017 to 1.89% in 2018 (https://www.imf.org/en/Countries/CHN). Last but not least, the China–U.S. trade war, started in 2017, further increases the investment risk in the Chinese exchange rate market.

Currently, how to forecast CNY exchange rate accurately is still an open question with great challenges, the main reason behind this being the complex coupling nature of the CNY exchange rate (see Figure 1), where coupling refers to any relationship or interaction connecting two or more aspects.[1] As shown in the figure, there exist three types of couplings. 1) *Market-level coupling*, referring to the interactions between market influence factors (e.g., commodity market) and the CNY exchange rate, and also the interactions between the influence factors. 2) *Macrolevel coupling*, indicating the interactions between macroeconomic influence factors (e.g., interest rate) and the CNY exchange rate, and the interactions between the influence factors. Here, we discriminate the market-level coupling and macrolevel coupling since these two types of couplings take different influence mechanisms on the exchange rate. Specifically, the impact influence of market-level factors, such as commodities, is stronger at short horizons, and becomes weaker as the horizon increases, while the impact of macrofundamentals lasts for a long horizon.[2] 3) *Deep coupling* represents the interactions between market-level coupling and macrolevel coupling in a deep layer. This is easy to understand since the macrolevel factors could affect the market-level factors and vice-versa. Therefore, it is important to analyze and capture the complex couplings for CNY exchange rate forecasting, since coupling learning has been approved to be an effective way to tackle the economic and financial issues.[3]

Modeling and extracting such complex deep coupled relationships present a challenge for existing methods. Specifically, traditional statistical methods (e.g., ARIMA[4]) have difficulties when facing the multiple influence factors with nonlinear and interconnected relationships,[5] let alone the complex hidden couplings. Simple machine learning methods with shallow architectures (e.g., SVR[6]) cannot incorporate the three kinds of couplings with different working mechanisms listed above. Recently, with the development of deep learning and FinTech, deep neural networks (e.g., CNN[7] and LSTM[8]) which are appropriately structured to learn deep features, are utilized in financial time series forecasting with promising results.[9] Especially the LSTM model which can effectively learn long-term and short-term dependencies through memory cells and gates, has been verified to be a useful tool in financial time series analysis.[8] Interestingly, the capability of learning long-range and short-range time dependencies matching the working mechanisms of macrolevel coupling and market-level coupling, which means that the LSTM model is suitable for CNY exchange rate forecasting. However, how to build an advanced LSTM model to fully and systematically capture the complex deep couplings across the influence factors is still an issue.

To address the above challenges, we propose a new deep coupled LSTM approach (DC-LSTM)

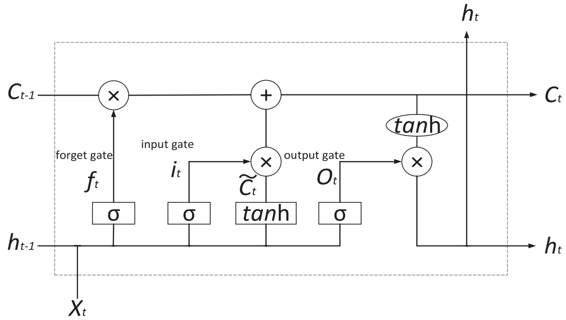


Figure 2. Structure of an LSTM memory block.

to encode the complex couplings for CNY exchange rate forecasting with a deep structure. At the first layer, we employ two LSTM models to learn the abstract features which represent market-level and macrolevel couplings. In the second layer, a coupled LSTM is built on the features learned from the first layer, so as to incorporate the short-term market-level coupling and long-term macrolevel coupling. Then, the learned deep coupled features are fed into a fully connected layer to conduct the final forecasting.

PROPOSED METHODOLOGY

Original LSTM Model

As depicted in Figure 2, the LSTM memory block is composed of a memory cell C_t and three gates with different purposes: a forget gate f_t specifies which information should be removed from the cell state C_{t-1}, an input gate i_t defines which information should be added to the cell state C_t, an output gate o_t defines which information from cell state C_t should be used as output. The detailed calculations and updating of the LSTM memory cells are performed with the following formulas.

$$f_t = \sigma\big(U_f X_t + W_f h_{t-1} + b_f\big) \tag{1}$$

$$i_t = \sigma(U_i X_t + W_i h_{t-1} + b_i) \tag{2}$$

$$\widetilde{C}_t = \tanh(U_c X_t + W_c h_{t-1} + b_c) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{4}$$

$$o_t = \sigma(U_o X_t + W_o h_{t-1} + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where X_t is the input variable at time t, U and W are weight matrices, b is the bias vector, $\widetilde{C}_t$ is the candidate of input to be stored, and h_t corresponds to the hidden state at time t. $\sigma(\cdot)$ is a sigmoid function, and symbol $*$ denotes element-wise multiplication. Then, the forward process can be described in three steps: 1) the first step is driven by forget gate f_t to determine how much information should be removed from the past cell state according to (1); 2) the second procedure comprises the three operations through input gate i_t according to (2) to (4): the input gate i_t determines how much new information should be stored, the candidate value $\widetilde{C}_t$ determines how much new information is received in the cell state, then the new cell state C_t is updated based on the previous two steps; and 3) the last step determines the hidden state h_t through output gate o_t according to (5) and (6).

Proposed DC-LSTM Model

The DC-LSTM, shown in Figure 3, owns a deep structure which consists of stacked LSTMs: there are two wings of classical LSTMs in the ground layer, where one wing represents the market-level coupling and the other captures the macrolevel coupling. Then, the two kinds of hidden couplings are fed into a new LSTM as input in the second layer, hence the second layer LSTM is used to model the deep interactions between market-level coupling and macrolevel coupling.

A. Representation of market-level coupling

Suppose there are I market indicators $\{\alpha_1, \alpha_2, \ldots, \alpha_I\}$ selected as market influence factors, and the element α_{it} denotes the observation of market indicator α_i at time t. As illustrated in the market-level coupling part in Figure 3, various market indicators α_{it} are served as inputs of one LSTM ($LSTM^{C_1}$) at the bottom layer at time t (i.e., $X_t^{C_1} = \{\alpha_{1t}, \alpha_{2t}, \ldots, \alpha_{It}, \mathrm{ER}_t\}$, where ER_t is the exchange rate value at time t), then, the hidden abstract patterns are learned through the gating mechanism. It is easy to find that the complex interactions between the indicators are encoded with the weight parameters U and W in (1) to (6). Therefore, the hidden features of market-level coupling at time t could be represented by $h_t^{C_1}$, where

$$h_t^{C_1} = o_t^{\mathbb{C}_1} * \tanh\Big(C_t^{C_1}\Big). \tag{7}$$

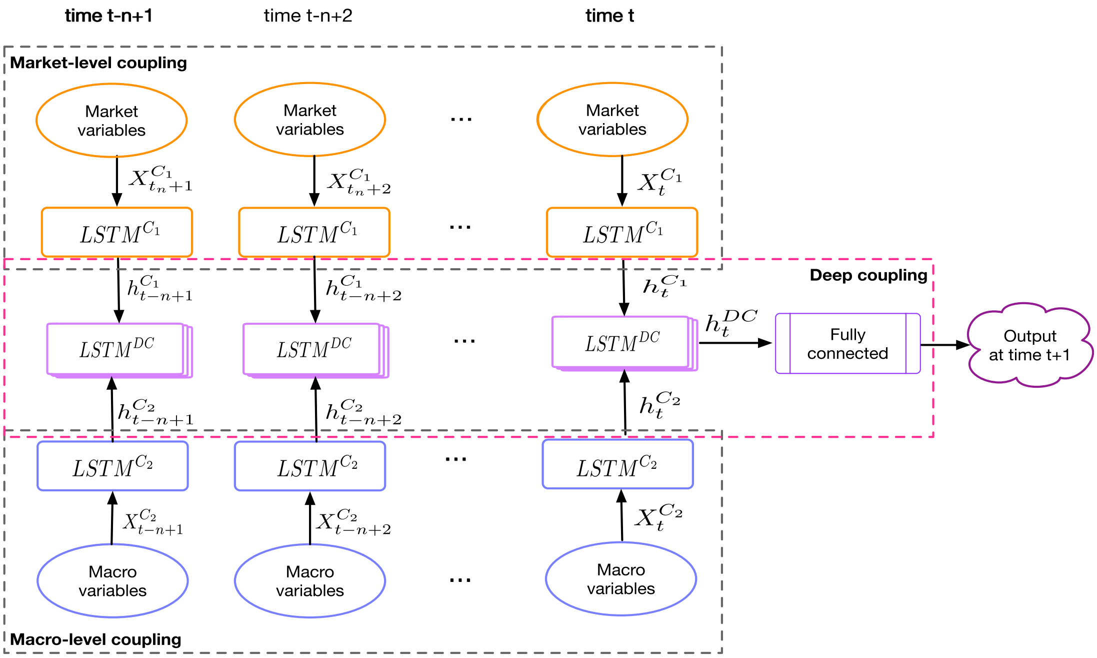


Figure 3. Architecture of the proposed DC-LSTM approach.

B. Representation of macrolevel coupling

Suppose there are J macroeconomic indexes $\{\beta_1, \beta_2, \ldots, \beta_J\}$, and β_{it} represents the observation of macro-based indicator β_i at time t. Similar to market-level coupling, at time t, the hidden coupling features $h_t^{C_2}$ are captured through an LSTM (LSTM^{C_2}) with the input $X_t^{C_2} = \{\beta_1 t, \beta_2 t, \ldots, \beta_J t, \text{ER}_t\}$ by

$$h_t^{C_2} = o_t^{C_2} * \tanh\left(C_t^{C_2}\right). \quad (8)$$

C. Forecasting based on DC-LSTM

As shown in Figure 3, suppose the window length is n (the data from time $t-n+1$ to time t are used to forecast the exchange rate at time $t+1$), then the output (hidden states) of LSTM^{C_1} ($\{h_{t-n+1}^{C_1}, h_{t-n+2}^{C_1}, \ldots, h_t^{C_1}\}$) and LSTM^{C_2} ($\{h_{t-n+1}^{C_2}, h_{t-n+2}^{C_2}, \ldots, h_t^{C_2}\}$) serves as the input of LSTM^{DC}, which means the DC-LSTM could jointly model the market-level and macrolevel couplings in a deep coupling layer. Then, at time t, the deep coupling features h_t^{DC} could be learned through the gate mechanisms of LSTM^{DC}

$$h_t^{\text{DC}} = o_t^{\text{DC}} * \tanh\left(C_t^{\text{DC}}\right) \quad (9)$$

where o_t and C_t are the output gate and cell state learned from X_t^{DC} and h_{t-1}^{DC}. Note that $X_t^{\text{DC}} = \{h_t^{C_1}, h_t^{C_2}\}$ here.

Accordingly, the learned deep coupling h_t^{DC} is passed through a fully connected layer to get the final prediction ER at time $t+1$ by:

$$\text{ER}_{t+1} = \delta(\Sigma h_t^{\text{DC}} W) \quad (10)$$

where δ is the activation function and W is the weight of connection between h_t^{DC} and ER_{t+1}.

EXPERIMENTAL SETTINGS AND RESULTS

Data Gathering and Preparation

The weekly exchange rate USD/CNY is selected as the target since it is the most important type of the CNY exchange rate.[10] Moreover, the selected market-level and macrolevel factors are listed in Table 1.

The data are collected from Wind Database (http://www.wind.com.cn/), covering the period from June 2009 to April 2019. The dataset is divided into an in-sample subset (June 2009 to March 2016) and an out-of-sample subset (April 2016 to April 2019). In addition, all data is normalized to [0,1] by $I_t' = \frac{I_t - I_{\min}}{I_{\max} - I_{\min}}$,where I_t is the original value of the index at time t.

Evaluation Methodology

A. Technical perspective

- *Directional forecasting accuracy (Acc)* is a measure of the performance in predicting

Table 1. Selected factors.

Category		Factors
Market-level data	Commodity market	WTI crude oil price
		Gold price
	Stock market	Shanghai Stock Exchange Composite Index
		Dow Jones Index
Macrolevel data	Monetary policy	Chinese money supply index (M1 and M2)
		Chinese consumer index
		Producer pricer index
		Industrial production index
	Interest rate	Shibor 1-week rate (Chinese)
		Federal fund rate (US)
	Inflation	Chinese inflation rate
	Trade balance	Trade balance index
		Payment balance index
	Policy uncertainty	Chinese economic policy uncertainty index (CEPU)
		US economic policy uncertainty index (UEPU)
		Global economic policy uncertainty index (GEPU)

the direction of value changes, ranges between 0 and 1 and a higher value indicates a better forecasting performance

$$Acc = \frac{1}{N}\sum_{t=1}^{N} d(t) \times 100\%, \text{where}\, d(t) = \begin{cases} 1 \text{ if } & [y(t+1)-y(t)][\widehat{y}(t+1) \\ -y(t)] \geq 0 & \\ 0 \text{ otherwise} & \end{cases} \tag{11}$$

where $\widehat{y}(t)$ and $y(t)$ denotes the forecasting value and the actual value at time t, respectively, and N is the size of the forecasting interval.

- Mean absolute error (MAE) is a measure of average of the differences between the actual and forecasting values, and a smaller value represents a higher forecasting accuracy

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t|. \tag{12}$$

- Root mean square error (RMSE) is the square root of the mean of the square of all of the errors, and a smaller value denotes a better forecasting performance

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t|}. \tag{13}$$

B. Statistical perspective

The Diebold–Mariano (DM) test[11] and Pesaran–Timmermann (PT) test[12] as statistical tests for multiple forecast comparison, are employed to compare the proposed approach with other benchmarks. The mean square error (MSE) and Acc are selected as the target functions for these two statistical tests, respectively.

- DM test

At first, the forecast error of model i is given by $u_{i,t} = \hat{y}_{i,t} - y_t$, where $\hat{y}_(t)$ and $y(t)$ denote the forecasting value and the actual value at time t, respectively. A loss differential between two forecasts i_1 and i_2 is built by $d_t = g(u_{i_1,t}) - g(u_{i_2,t})$, where $g(\cdot)$ is a given loss function (here, we use MSE). Then, the DM statistic could be computed by

$$\text{DM} = \frac{\overline{d}}{\sqrt{2\pi\hat{f}_d(0)/T}} \sim N(0,1) \tag{14}$$

Table 2. Input variables and parameters of various approaches.

Approaches	Factors			Parameters
	Market-level	Macrolevel	History value (USD/CNY)	
ARIMA	X	X	O	Order=(2,1,0)
SVR	O	O	O	see the work by Swami and Jain[13]
CNN	X	X	O	Seven layers; Optimizer: Adam; Activate function: RELU
LSTM-single	X	X	O	Hidden units: 30; Learning rate:0.0001; Epoch:200
LSTM-all	O	O	O	Hidden units: 30; Learning rate:0.0001; Epoch:200
LSTM-market	O	X	O	Hidden units: 25; Learning rate:0.0001; Epoch:200
LSTM-macro	X	O	O	Hidden units: 15; Learning rate:0.0001; Epoch:200
DC-LSTM	O	O	O	Hidden units: 25(market-level LSTM), 15(macrolevel), 30(deep layer); Learning rate:0.0001; Epoch:200 Optimizer: Adam; Activate function: RELU

O denotes employment of the factors while X represents exclusion of the factors.

where $\overline{d} = \frac{1}{T}\sum_{t=1}^{T}\left(g\left(u_{1,t}\right) - g\left(u_{2,t}\right)\right)$, and $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$, which represents the spectral density of the d_t at frequency 0. Then, the corresponding p value is obtained; if the p value is less than the threshold, the null hypothesis that the forecasting abilities of the two models i_1 and i_2 are equivalent is rejected.

- PT test

The objective of the PT test is to determine whether a forecast does a good job in predicting the change in the direction of a time series. Suppose $\hat{y}_(t)$ and $y(t)$ denote the forecasting value and the actual value at time t, and the number of observations is T, then, the PT statistic could be obtained through

$$PT = \frac{p_{y\hat{y}} - p}{\sqrt{v - w}} \sim N(0, 1) \tag{15}$$

where $p = p_y p_{\hat{y}} + (1 - p_y)(1 - p_{\hat{y}})$,

$v = \frac{p(1-p)}{T}$,

$w = (2p_y - 1)^2 q_{\hat{y}} + (2p_{\hat{y}} - 1)^2 q_y + 4q_y q_{\hat{y}}$,

$I_t(i) = \begin{cases} 1; & t_i > 0 \\ 0; & t_i \leq 0 \end{cases}$, $p_t = \frac{1}{T}\sum_{i=1}^{T} I_t(i)$, and $q_t = \frac{p_t(1-p_t)}{T}$. Similar to the DM test, the null hypothesis that $\hat{y}$ is not a good forecast of y could be rejected if the corresponding p value is less than the threshold.

Baseline Algorithms

We compare the performance of the proposed method with the following seven benchmarks, and the input variables and parameters for the benchmarks and the proposed DC-LSTM are listed in Table 2.

- ARIMA[4]: This is a statistical method for analyzing and building a forecasting model which best represents a time series by modeling the correlations in the data. We use it as a baseline method.
- SVR[6]: As a standard machine learning method, we can derive the incremental value-add of the proposed DC-LSTM model through comparison.
- CNN[7]: This is a typical deep learning model, the effect of long-term and short-term memories could be observed through comparing the performance of CNN and LSTM.
- LSTM-single: This is a single LSTM model which only considers the history values of the USD/CNY exchange rate as input.
- LSTM-all: The different performance of this model with DC-LSTM could reflect the effect of deep coupling analysis.
- LSTM-market: This is a submodel of DC-LSTM, which simply models the market-level coupling without considering the macrolevel coupling.

Table 3. Performance comparison of different models.

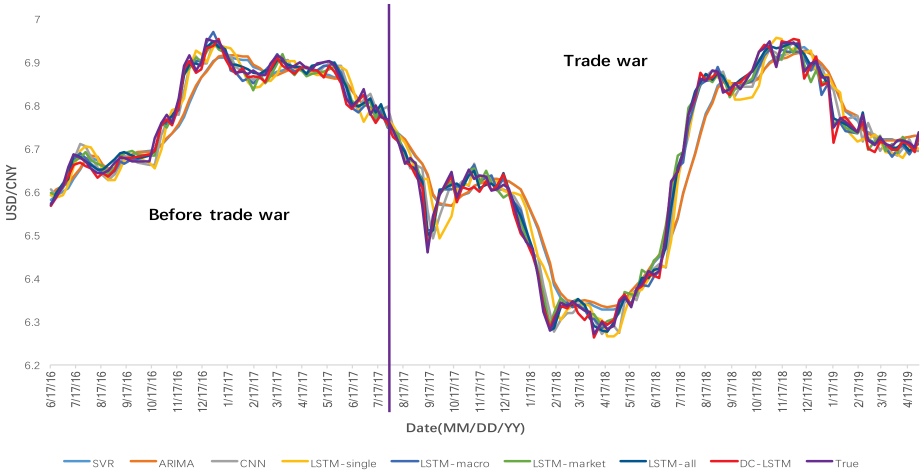
(a) One-month window length						
Models	MAE	Rank	RMSE	Rank	Acc(%)	Rank
ARIMA	0.0257	8	0.0350	8	53.85	8
SVR	0.0231	5	0.0326	6	55.48	5
CNN	0.0243	7	0.0336	7	54.84	7
LSTM-single	0.0238	6	0.0323	5	55.48	5
LSTM-market	0.0191	4	0.0252	3	66.45	3
LSTM-macro	0.0230	3	0.0287	4	64.51	4
LSTM-all	0.0151	2	0.0202	2	69.68	2
DC-LSTM	0.0145	1	0.0189	1	71.61	1
(b) Three-months window length						
Models	MAE	Rank	RMSE	Rank	Acc	Rank
ARIMA	0.0261	8	0.0356	8	54.42	8
SVR	0.0255	7	0.0304	7	57.53	6
CNN	0.0226	6	0.0302	6	57.53	6
LSTM-single	0.0210	5	0.0275	5	61.64	5
LSTM-market	0.0160	4	0.0196	4	68.49	3
LSTM-macro	0.0153	3	0.0193	3	67.12	4
LSTM-all	0.0134	2	0.0179	2	71.92	2
DC-LSTM	0.0127	1	0.0168	1	75.34	1
(c) Six-month window length						
Models	MAE	Rank	RMSE	Rank	Acc	Rank
ARIMA	0.0272	8	0.0369	8	50.75	8
SVR	0.0247	7	0.0336	7	55.64	7
CNN	0.0236	6	0.0331	6	57.89	6
LSTM-single	0.0229	5	0.0303	5	60.15	5
LSTM-market	0.0191	4	0.0241	4	65.41	3
LSTM-macro	0.0188	3	0.0233	3	63.16	4
LSTM-all	0.0169	2	0.0208	2	72.93	2
DC-LSTM	0.0168	1	0.0202	1	75.93	1

- LSTM-macro: This is a submodel of DC-LSTM, which only considers the macrolevel coupling while overlooking the macromarket coupling.

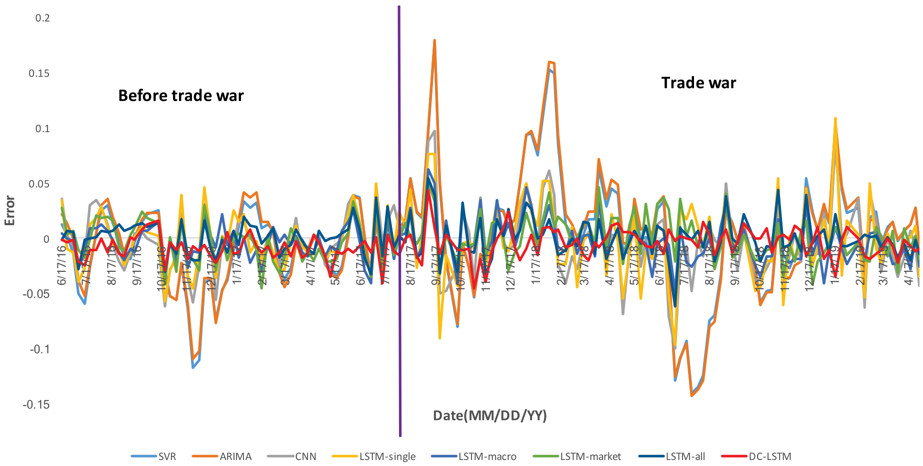
Results

A. Results of forecasting performance

Table 3 compares the forecasting performances of the eight approaches in terms of forecasting errors and directional accuracy via MAE, RMSE, and Acc, with three different window lengths. The proposed DC-LSTM approach outperforms other benchmarks under all evaluation criteria and with all window lengths. For example, the DC-LSTM exhibits 3.5%, 8.2%, and 7% improvements compared to LSTM-all, LSTM-macro, and LSTM-market in terms of Acc (see Table 3), which



(a) The forecasting values



(b) The forecasting errors

Figure 4. Forecasting results using different models: three-month window length. (a) Forecasting values. (b) Forecasting errors.

verifies the effect of deep coupling analysis, when compared with shallow coupling and simple coupling. Interestingly, the performance of LSTM-macro and LSTM-market are conflicting with each other. Specifically, LSTM-macro performs better than the LSTM-market with longer window length in terms of MAE and RMSE. And this is consistent with former research[2] that macroeconomic factors have long-term influence on exchange rate. In addition, it is worth noting that the LSTM-based models obviously outperform the simple SVR, ARIMA, and CNN, which further proves the superiority of the LSTM model which is capable of capturing the long-term and short-term dependencies in time series.

Figure 4 reports the forecasting results of the eight approaches with three-months window length. It can be seen in Figure 4(a) that the fittings and predictions look reasonably accurate, and the results obtained by the models are similar. But, based on Figure 4(b) which characterizes forecasting errors, we can observe that the proposed DC-LSTM approach has the smallest fluctuations (errors) through the whole testing period when compared with other benchmarks. Especially in the China–U.S. trade war period (started in August, 2017), full of uncertainty, the DC-LSTM still achieves low forecasting errors while the performance of other models fluctuate sharply.

B. Results of the statistical test

The DM and PT tests are used to statistically evaluate forecasting errors and accuracy. The results are depicted in Table 4.

Table 4. Statistical results of different models.

(a) DM results			
	Window length		
	One-month	Three-month	Six-month
ARIMA	−5.5756***	−5.7469***	−5.2406***
SVR	−4.1386***	−5.6312***	−5.0017***
CNN	−4.5436***	−5.5542***	−4.9291***
LSTM-single	−4.4943***	−5.1367***	−3.4557***
LSTM-market	−3.0475***	−4.5983***	−2.1484**
LSTM-macro	−5.4956***	−3.9168***	−2.3851 **
LSTM-all	−2.1129**	−2.1477**	−2.0490**

(b) PT results			
	Window length		
	One-month	Three-month	Six-month
ARIMA	1.0221	1.3292*	0.6064
SVR	1.2018	1.8277**	1.3041*
CNN	1.0455	1.6634**	1.6517**
LSTM-single	1.3292*	2.6736**	2.3488**
LSTM-market	3.9481***	4.3330***	3.3974***
LSTM-macro	3.6086***	4.1585***	3.0593***
LSTM-all	4.7432***	5.3274***	5.3090***
DC-LSTM	5.3901***	6.9817***	6.0337***

*denotes a rejection of null hypothesis at the 10% significance level.
**denotes a rejection of null hypothesis at the 5% significance level.
***denotes a rejection of null hypothesis at the 1% significance level.

Table 4 shows the DM test results in terms of DM statistics, here the MSE is considered as the loss function and the proposed DC-LSTM approach is compared against other seven models. From the Table 4, it can be clearly seen that: 1) compared with the LSTM-all model, all DM statistics are less than -2 corresponding to P-values are less than 0.05, which indicates that the proposed DC-LSTM is better than LSTM-all under 95% confidence level with all window lengths; 2) compared with LSTM-market and LSTM-macro approaches, the DC-LSTM approach is significantly superior to the two approaches under 99% confidence level with one-month and three-month window lengths, while under 95% confidence level with six-month window length; and 3) compared with the LSTM-single, CNN, SVR, and ARIMA approaches, DC-LSTM is better with under 99% confidence level with all window lengths.

Table 4 illustrates the PT test results in terms of PT statistics, and the results check how well the predicted directions follow the real directions. As shown in the Table: 1) our proposed DC-LSTM displays the best statistical performance with the highest PT statistics with all window lengths; 2) the coupling-based approaches (LSTM-all, LSTM-macro, and LSTM-market) are significantly valid forecasting approaches under 99% confidence level, which further demonstrates the importance of coupling analysis in exchange rate forecasting; and 3) simple models like ARIMA, SVR, and CNN are almost ineffective since their p-values are greater than 0.1 with one-month window length.

Backtesting on Profitability

Since having a high prediction accuracy does not imply that the strategy generates profit, here, we compare the performance of our approach against the seven other approaches with three-months window length in the testing period from June 2016 to April 2019, together with two other trading strategies: 1) buy and hold strategy: a long-term investment strategy based on the view that in the long run financial markets give a good rate of return despite periods of volatility or decline. According to the trading strategy, an investor buys an asset and holds it until the end; and 2) risk free strategy: it is a look-ahead strategy which means that an investor has perfect foresight of the sign of the trend change from today to tomorrow. It is clear that this strategy can never be achieved, and here we treat it as an interesting upper bound of the performance. The wealth evolutions are shown in Figure 5.

We have the following settings for the investment: 1) the initial capital investment is USD100;

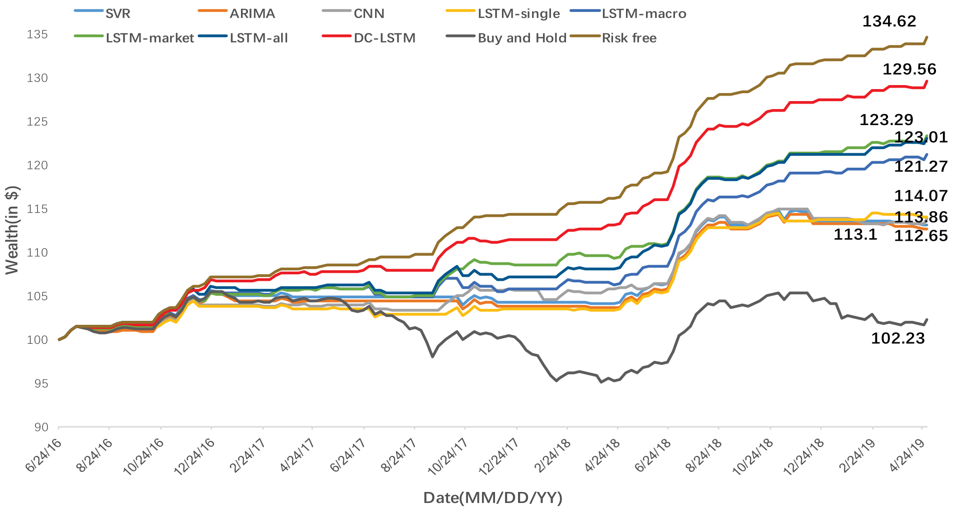


Figure 5. Investor's wealth evolution.

2) no new capital will be added thereafter; 3) the investor buys and sells the index according to the trends forecasted by each approach (buy when there is an upward forecasting and sell while downward); and 4) there are no transition fees. The following conclusions can be drawn from the figure: a. the DC-LSTM approach performs best, that is, an investor taking recommendations from the approach can make profit at USD29.56, which represents a return of 34.62% in around three years during the trade war period; and b. LSTM-all, LSTM-market, and LSTM-macro achieve almost similar results, which is much better than LSTM-single, CNN, SVR, and ARIMA. We believe these results further demonstrate the importance of coupling analysis in exchange rate forecasting, and to verify the proposed DC-LSTM as a useful tool to help investors make wiser trading decisions.

CONCLUSION

Forecasting the CNY exchange rate is very challenging especially under the complex couplings and fluctuations due to CNY internationalization and the unstable economic environment. In this study, the DC-LSTM model is proposed to learn the complex deep couplings between multiple influence factors of the USD/CNY exchange rate. We believe the experimental results with 10 years data can not only report the importance of coupling learning in exchange rate forecasting, but also the superiority of the proposed DC-LSTM model in capturing the couplings when compared with various benchmarks.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Project Nos. 71801072, 71774047, and 71971071.

REFERENCES

1. L. Cao, "Coupling learning of complex interactions," *Inf. Process. Manage.*, vol. 51, no. 2, pp. 167–186, 2015.
2. C. Amat, T. Michalski, and G. Stoltz, "Fundamentals and exchange rate forecastability with simple machine learning methods," *J. Int. Money Finance*, vol. 88, pp. 1–24, 2018.
3. W. Cao, L. Hu, and L. Cao, "Deep modeling complex couplings within financial markets," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2518–2524.
4. G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
5. Y. Wei, S. Sun, J. Ma, S. Wang, and K. K. Lai, "A decomposition clustering ensemble learning approach for forecasting foreign exchange rates," *J. Manage. Sci. Eng.*, vol. 4, pp. 45–54, 2019.
6. S. Fu, Y. Li, S. Sun, and H. Li, "Evolutionary support vector machine for RMB exchange rate forecasting," *Phys. A, Statist. Mech. Appl.*, vol. 521, pp. 692–704, 2019.

7. E. Hoseinzade and S. Haratizadeh, “CNNpred: CNN-based stock market prediction using a diverse set of variables,” *Expert Syst. Appl.*, vol. 129, pp. 273–285, 2019.
8. Y. Baek and H. Y. Kim, “Modaugnet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module,” *Expert Syst. Appl.*, vol. 113, pp. 457–480, 2018.
9. Y. Liu, “Novel volatility forecasting using deep learning long short term memory recurrent neural networks,” *Expert Syst. Appl.*, vol. 132, pp. 99–109, 2019.
10. G. Cao, L. Xu, and J. Cao, “Multifractal detrended cross-correlations between the Chinese exchange market and stock market,” *Phys. A, Statist. Mech. Appl.*, vol. 391, no. 20, pp. 4855–4866, 2012.
11. F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” *J. Bus. Econ. Statist.*, vol. 13, no. 1, pp. 134–144, 1995.
12. M. H. Pesaran and A. Timmermann, “A simple nonparametric test of predictive performance,” *J. Bus. Econ. Statist.*, vol. 10, no. 4, pp. 461–465, 1992.
13. A. Swami and R. Jain, “Scikit-learn: Machine learning in python,” *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, 2013.

Wei Cao is currently an Associate Professor with the School of Economics, HeFei University of Technology, Hefei, China. Her research interests include data mining, machine learning, deep learning, and coupling analysis in finance. She received the Ph.D. degree in information technology from the University of Technology, Sydney, NSW, Australia. Contact her at weicao@hfut.edu.cn.

Weidong Zhu is currently a Professor with the School of Economics, HeFei University of Technology, Hefei, China. His research interests include evidence theory, decision theory, and data mining, with applications in: business management and finance. He received the Ph.D. degree in management science from the HeFei University of Technology. Contact him at zhuwd@hfut.edu.cn.

Wenjun Wang is currently an Associate Professor with the School of Economics, Hefei University of Technology, Hefei, China. Her research interests include data mining, price and risk modeling on financial markets. She received the Ph.D. degree in management science and engineering from the University of Science and Technology of China, Hefei, China. Contact her at wjwang@hfut.edu.cn

Yves Demazeau is currently an ECCAI Fellow and Doctor Honoris Causa with SDU Odense, Director of Research with CNRS and Invited Professor with SDU, President of the French Association for AI (AFIA) since 2011. His current research in AI include: user centration, ethics by design, organizations dynamics, observation and evaluation, with applications in: energy management, bio-inspired UAVs, social network analysis, and financial modeling. Contact him at yves.demazeau@imag.fr.

Chen Zhang is currently a Professor with the School of Management, HeFei University of Technology, Hefei, China. Her research interests include data mining, deep learning, and QCA analysis in financial risk measurement and financial asset pricing. She received the Ph.D. degree in management science from the HeFei University of Technology. Contact her at sm.zhangchen@hfut.edu.cn.