A machine learning approach to forecasting carry trade returns

Xiao Wang*† Xiao Xie† Yihua Chen† Borui Zhao†

April 7, 2021

Abstract

Carry trade refers to a risky arbitrage in interest rate differentials between two currencies. Persistent excess carry trade returns pose a challenge to foreign exchange market efficiency. Using a data set of ten currencies between 1990 and 2017, we find: (i) a machine learning model, long short-term memory (LSTM) networks, forecast carry trade returns better than linear and threshold models and other machine learning models; and (ii) excess carry trade returns deteriorate after the 2007–2008 global financial crisis in all model forecasts, indicating that the uncovered interest rate parity may still hold in the long run.

JEL Classifications: C45, F21, F31, F37

Keywords: carry trade, uncovered interest rate parity, machine learning, long short-term memory networks

---

*Corresponding author.

†Xiao Wang: School of Management (International Institute of Finance), University of Science and Technology of China; email: iriswx@ustc.edu.cn. Xiao Wang acknowledges financial support from the National Natural Science Foundation of China (grant number: 72003181). Xiao Xie: Shanghai Advanced Institute of Finance, Shanghai Jiaotong University; email: xxie.20@saif.sjtu.edu.cn. Yihua Chen: School of Mathematical Sciences, University of Science and Technology of China; email: chenyihu@mail.ustc.edu.cn. Borui Zhao: School of Mathematical Sciences, University of Science and Technology of China; email: zbr1997@mail.ustc.edu.cn.

1 Introduction

Carry trade is a risky arbitrage based on interest rate differentials between two currencies. Carry trade returns depend on changes in the exchange rate—the relative price of currencies—and the interest rate differentials between two countries. While the persistent excess carry trade returns seem to challenge the efficient market hypothesis, there is also evidence that uncovered interest parity (UIP) may hold in the long run, and thus the excess carry trade returns will eventually reverse.[1]

We aim to improve carry trade return prediction and re-examine whether the UIP holds. Two strands of literature jointly suggest a machine learning approach to forecasting carry trade returns. In the first strand of the literature on the forecast of excess carry trade returns, Engel (2014) points out that a key improvement is to introduce the nonlinear relationship between explanatory factors for carry trade returns. For example, Jorda and Taylor (2012) unveil that augmenting linear models with thresholds of fundamentals strengthen the model predictability. Lustig and Verdelhan (2007) and Menkhoff et al. (2012) find that excess carry trade returns are related with the volatility of exchange rate or economic fundamentals. In summary, optimizing the use of higher moments and cross-moments of explanatory factors may help to capture the risk premium in carry trade returns and thus improve the return forecast. Moreover, as interest rates fluctuate near zero in US and many other countries after the 2007–2008 global financial crisis, the movement in the exchange rate becomes more dominant in carry trade returns. Rossi (2013) provides an overview that the predictability of exchange rate depends on the forecast horizon length besides the choice of model and data—the exchange rate is predictable at the ten-year horizon (Chinn, 2006) but not at shorter horizons. Overall, we may improve carry trade return prediction if the factor interactions and the horizon length of factors are more flexible in model specifications.

The second strand of literature is the fast growing application of machine learning in economic forecasting. Gu et al. (2020) find that machine learning allows nonlinear interactions between predictors and optimizes model specification. Therefore, machine learning opens up a new way for empirical asset pricing. Specifically, we focus on long short-term memory (LSTM) networks,

[1] See Burnside et al. (2008).

a machine learning model for time series forecasting. Hochreiter and Schmidhuber (1997) first develop LSTM networks that can "memorize" the long-term data pattern and determine how the input in every period can "enter" into the memory process. In a multi-period model structure, LSTM networks include input, hidden, and output layers in each period. The key mechanism is memory cells in the hidden layer that transmit information between periods. In a memory cell, a forget gate, an input gate, and an output gate function determine how the information from previous periods can be transmitted, how the input can be included, and how the information can be sent out respectively. Therefore LSTM networks optimize the use of explanatory factors and data length and thus can potential improve return forecasts.

In this paper, we construct a monthly data set in G10 currencies (Australia, Canada, Germany, Japan, Norway, New Zealand, Sweden, Switzerland, the U.K., and the U.S.) between 1990 and 2017. We then train the LSTM networks model with data and make predictions. The performance of LSTM networks dominates other popular models for carry trade returns. The average return, the Sharpe ratio, and the gain/loss ratio are higher in the LSTM networks model in three forecast windows 2011-2015, 2012-2016, and 2013-2017. All model forecasting results show that the excess carry trade returns are lower after the 2007–2008 global crisis.

This paper contributes to the literature on carry trade returns in two dimensions. First, we introduce long short-term memory (LSTM) networks to predict carry trade returns and find that LSTM networks outperform commonly used models. To our knowledge, this paper is the first to explore the application of LSTM networks in a carry trade return forecast.[2] As a deep learning model, LSTM networks optmizes the use of long- and short-term information and the use of factor moments in carry trade return prediction.[3] Thus LSTM networks can improve the carry trade return forecast as the literature indicates that the UIP may hold in the long run and the risk premium in carry trade returns is related with higher moments of explanatory factors. We also find that the long short-term memory structure is the key to improve return predictions by comparing LSTM networks with otherwise identical machine learning models without such structure. Second, we find that carry trade returns deteriorate after the global financial crisis in all models as the interest

[2]See Davis et al. (2020) on machine learning applications for time series problems.

[3]Colombo et al. (2019) employs support vector machines to predict carry trade direction.

rate differentials shrink to near zero, consistent with Accominotti et al. (2019). Our findings provide supportive evidence that we may not be able to reject the UIP.

2 Data, carry trades, and model

Data. The monthly data set that spans from May 1990 to October 2017 includes the exchange rate and the one-month risk-free interest rate data from Datastream, and the consumer price index data from FRED Economic Data for G10 countries. Those countries' currency trading volume prevails in the foreign exchange market, and all countries take the floating exchange rate regime and allow free capital mobility.

Carry trade return. We follow Jordan and Taylor (2012) to construct a carry trade return estimation framework. Setting the U.S. as the home country in a currency pair, the ex-post nominal excess return for a carry trade s_{t+1} is

$$s_{t+1} = \Delta e_{t+1} + (i_t^* - i_t), \tag{1}$$

where Δe_{t+1} is the logged exchange rate (the home currency price of one unit of foreign currency) difference, and i_t and i_t^* are home and foreign one-period, risk-free interest rates. The UIP implies that $E_t(s_{t+1}) = 0$.

We further re-express the carry trade return in real terms. We define the real exchange rate as $q_{t+1} = \bar{q} + e_{t+1} + (p_{t+1}^* - p_{t+1})$, where $\bar{q}$ is the mean of q_t, and p_{t+1} and p_{t+1}^* are the logged aggregate prices home and abroad. Under purchasing power parity (PPP), real exchange rate q_t converges to $\bar{q}$ and thus $q_t - \bar{q}$ is stationary. If π_{t+1} is the inflation rate and the real interest rate is $r_t = i_t - \pi_{t+1}$, the carry trade return in equation (1) can be written as

$$s_{t+1} = \Delta q_{t+1} + (r_t^* - r_t). \tag{2}$$

Given that Δe_{t+1} and Δq_{t+1} are stationary, equations (1) and (2) can be viewed as an evolution from a stationary system:

$$\Delta v_{t+1} = \left[\Delta e_{t+1}, \pi^*_{t+1} - \pi_{t+1}, i^*_t - i_t\right]', \tag{3}$$

where the cointegration vector is $e_t + p^*_t - p_t = q_t - \bar{q}$, the deviation from PPP.

Table 1 displays summary statistics, in particular, carry trade return on average is above zero.[4]

Table 1: Summary Statistics

	Currencies	Months	Mean	s.d.	Skewness	Kurtosis
s	9	330	0.0005	0.0305	-0.3098	4.9616
Δe	9	330	-0.0001	0.0304	-0.3506	5.1727
$\pi^* - \pi$	9	330	-0.0005	0.0041	0.6771	7.3543
$i^* - i$	9	330	0.0005	0.0020	0.5453	7.2803

Predicting carry trade returns using LSTM networks. We utilize LSTM networks to estimate the carry trade return. The LSTM networks employ the long-term data pattern and determine how the new information in every period can enter the model. In LSTM networks, time period t spans from 0 to T. In every period, the model consists of an input layer, one or more hidden layers, and an output layer. The input layer includes the explanatory variables for the carry trade return: Δe_t, $\pi^*_t - \pi_t$, $i^*_t - i_t$ and $q_t - \bar{q}$, $t = 0, 1, \cdots, T-1$. The output layer generates the prediction for exchange rate $\hat{e}_{t+1}$. The hidden layer is the key structure of LSTM networks as it consists of memory cells that determine how information can be injected and carried to the next period. In each memory cell, three types of gates—forget gate, input gate, and output gate—are status activation functions and jointly determine how information is transmitted in the cell state. We explain the information transmission process and show the full model structure in detail in the online appendix. We also show in the appendix that if shutting down the memory cell gate structure, LSTM networks may degenerate to its parent model, the recurrent neural network (RNN).

We divide the sample into the training set (in sample) from 1990 to 2011,[5] and three test windows (out of sample): 2012–2015, 2013–2016, and 2014–2017. As the training set includes the 2007–2008 global financial crisis, LSTM networks can adapt the abnormal carry trade returns

[4] We also confirm that $q_t - \bar{q}$ is stationary by the panel cointegration tests.

[5] A rule of thumb for splitting between training and test sets is 5:1. Results are robust with other splitting.

during the turmoil periods into model training. We also split out the validation set (between 2010 and 2011) from the training set to avoid the potential over-fitting problem.[6] In the test data, we employ the rolling window strategy with one-step ahead forecast.

Following Jorda and Taylor (2012), we build an equally weighted portfolio with $1/N(N = 9)$ portion of investment on each foreign currency, because the equally weighted strategy makes comparing portfolio performance in different models possible. Given the model forecast $\hat{e}_{t+1}$, the investors can long or short each currency to re-balance the equal-weighted portfolio. Then we calculate the realized return from the exchange rate data.

3 Results

In this section, we first compare the LSTM performance with widely used models for carry trade returns and then investigate the resource of performance difference.

LSTM networks specify the interactions between forecasting factors and the lag terms of factors more flexibly. In order to show the resource of forecasting improvement in it, we compare LSTM networks with other popular models for carry trade return forecast and an otherwise identical machine learning model with no memory structure. We choose random walk, vector autoregression (VAR), the threshold vector error correction model (TECM), and RNN as the benchmark models. In linear models, the random walk model assumes that the exchange rate movement Δe_{t+1} is independently and identically distributed, while the VAR model with the optimal lag as one period assumes that economic fundamentals are capable of forecasting future exchange rate movement. As a nonlinear model, the TECM (one period as the optimal lag order and $q_t - \bar{q}$ as the cointegration variable) assumes parsimonious thresholds—whether the interest rate differential and the real exchange rate are above their median values. The RNN model is a degeneration of LSTM networks by shutting down the memory cell gate structure.

Table 2 displays the model performance comparison in forecasting carry trade returns. Firstly, in all three rolling windows, LSTM networks dominate the other four models. The Diebold and

[6] Results are consistent under an alternative 80% training versus 20% validation rule.

Table 2: Carry trade return

	Random Walk	VAR	TECM	RNN	LSTM
Profits 2012–2015					
Mean	-0.0009	-0.0008	-0.0006	-0.0036	0.0014
Std. dev.	0.0177	0.0116	0.0118	0.0172	0.0100
Skewness	0.1053	-0.4649	-0.5411	-0.2375	0.2139
Sharpe ratio (annual)	-0.1863	-0.2750	-0.1806	-0.7482	0.4636
Gain/loss	0.8550	0.8698	0.8720	0.5541	1.4427
D-M test p-value	0.0000	0.0000	0.0001	0.0000	-
OOS R^2	0.0045	0.0264	0.0090	0.0400	0.0474
Profits 2013–2016					
Mean	-0.0004	-0.0003	-0.0010	-0.0041	0.0028
Std. dev.	0.0181	0.0106	0.0121	0.0174	0.0077
Skewness	0.0328	-0.3385	-0.4850	0.3379	0.5110
Sharpe ratio (annual)	-0.0807	-0.1262	-0.3181	-0.8372	1.2413
Gain/loss	0.9505	0.9189	0.7817	0.5348	2.4960
D-M test p-value	0.0000	0.0000	0.0001	0.0000	-
OOS R^2	0.0649	0.0400	0.0357	0.0370	0.0609
Profits 2014–2017					
Mean	0.0009	-0.0003	-0.0011	-0.0022	0.0024
Std. dev.	0.0183	0.0104	0.0134	0.0182	0.0073
Skewness	-0.1544	-0.2916	-0.3871	0.2296	0.9213
Sharpe ratio (annual)	0.1702	-0.1091	-0.3004	-0.4199	1.1308
Gain/loss	1.1022	0.9141	0.7891	0.7382	2.5344
D-M test p-value	0.0000	0.0000	0.0000	0.0000	-
OOS R^2	0.0286	0.0512	0.0239	0.0417	0.0769

Mariano (1995) test[7] and the out-of-sample R^2 show that LSTM networks are more accurate in return predictions. The LSTM networks in 2012–2015 generate an average monthly return of 14 basis points, better than the negative returns from other models. The low standard deviation and positive skewness of carry trade returns in LSTM networks suggest that the carry trade portfolio has low volatility and is less likely to generate large loss. Therefore, its Sharpe ratio and gain/loss ratio are also higher.[8] The average return, the Sharpe ratio, and the gain/loss ratio are higher in 2013–2016 and 2014–2017 than those in 2012–2015, indicating that carry trade returns are low during the near-zero interest rate periods due to unprecedented expansionary monetary policies but improve with the step-down of quantitative easing. In summary, the flexible modeling of the higher moments and interactions among explanatory factors and the optimal use of data length help to strengthen carry trade predictability. As the literature indicates that excess carry trade return movement is related with the volatility of fundamentals and that the long run changes in interest rate differentials tend to explain the exchange rate movement better, LSTM networks generate higher carry trade returns.

Secondly, the highest annual returns in the three windows in Table 2 are 1.7% ($1.0014^{12} - 1$), 3.4%, and 2.9% respectively, much lower than the 7.1% in Jorda and Taylor (2012) and the 7.4% in Accominotti et al. (2019) before the global financial crisis. Overall, the lower returns after the crisis provide supporting evidence that the UIP may hold in the long run. Facing mixed evidence on whether the UIP holds, Bekaert and Hodrick (2001) and West (2012) conclude that if the econometrician's sample is small and biased, we may reject UIP. The persistent excess carry trade returns up to the 2000s may simply be a result of small sample. With more data after the global financial crisis, we find that the UIP holds. Another possible reason is that the UIP tends to hold in the long run, as proposed in Chinn (2006). LSTM networks optimizes the information use including the long-term "memory" and improves the exchange rate forecast, and thus the UIP may hold.

[7] The null versus alternative hypotheses are that the model (random walk, VAR, TECM, or RNN) has the same forecast accuracy versus that it has less forecast accuracy with LSTM networks.

[8] The Sharpe ratio is measured as the difference between the portforlio return and the risk-free one-month bond rate, divided by the standard deviation of the portfolio. The gain/loss ratio is measured as the ratio of the probability of gains over that of losses.

In order to open the black box in machine learning, we further investigate when the LSTM networks outperform other models. Given that the portfolio is equally weighted, we examine the realized return at the currency pair level in order to dissect the portfolio return dominance under LSTM networks. In Table 3, we first compare LSTM networks with TECM, the representative of regression models, and find that LSTM networks perform better when models suggest contradicting strategies of carry trade: returns based on LSTM and TECM forecasts are 0.0030 versus -0.0018, 0.0037 versus -0.0027, and 0.0045 versus -0.0033 in three rolling windows. We then compare LSTM networks with RNN to investigate whether the memory cell gate structure in LSTM networks helps predicting returns. Again returns from LSTM are higher when two models suggest different exchange rate movement. Overall, LSTM networks improve the prediction accuracy, avoid large trading losses, and thus reduce the carry trade return risk and achieve a higher portfolio return.

Table 3: Comparison of LSTM, TECM, and RNN

	Where models agree		Where models disagree		Where models agree		Where models disagree	
	LSTM	TECM	LSTM	TECM	LSTM	RNN	LSTM	RNN
Profits 2012–2015								
Number of obs.	206	206	217	217	177	177	246	246
Mean	0.0001	0.0001	0.0030	-0.0018	-0.0064	-0.0064	0.0039	-0.0026
Std. dev.	0.0267	0.0267	0.0274	0.0274	0.0260	0.0260	0.0270	0.0270
Skewness	-0.1538	-0.1538	0.1930	-0.1652	-0.8650	-0.8650	0.5015	-0.3353
Max	0.0709	0.0709	0.0794	0.0793	0.0651	0.0651	0.0879	0.0793
Min	-0.0871	-0.0871	-0.0784	-0.0772	-0.0726	-0.0726	-0.0784	-0.0871
Profits 2013–2016								
Number of obs.	202	202	221	221	177	177	246	246
Mean	0.0001	0.0001	0.0037	-0.0027	-0.0050	-0.0050	0.0063	-0.0051
Std. dev.	0.0278	0.0278	0.0276	0.0276	0.0250	0.0250	0.0288	0.0288
Skewness	-0.0967	-0.0967	0.1191	-0.0993	-0.7049	-0.7049	0.7635	-0.6160
Max	0.0893	0.0893	0.0774	0.0793	0.0820	0.0820	0.0893	0.0793
Min	-0.0871	-0.0871	-0.0784	-0.0770	-0.0716	-0.0716	-0.0784	-0.0882
Profits 2014–2017								
Number of obs.	197	197	199	199	163	163	233	233
Mean	0.0003	0.0003	0.0045	-0.0033	-0.0039	-0.0039	0.0039	-0.0024
Std. dev.	0.0282	0.0282	0.0270	0.0271	0.0253	0.0253	0.0288	0.0289
Skewness	-0.0700	-0.0700	0.1788	-0.1731	-0.5401	-0.5401	0.4627	-0.2941
Max	0.0893	0.0893	0.0773	0.0776	0.0651	0.0651	0.0879	0.0820
Min	-0.0871	-0.0871	-0.0784	-0.0770	-0.0882	-0.0882	-0.0815	-0.0871

4 Conclusion

While the persistent positive carry trade returns of the early 2000s have posed a challenge to the UIP, the literature is silent on the new features of carry trade after the global financial crisis. We introduce a novel machine learning model, LSTM networks, to forecast carry trade returns and find that LSTM networks improve return forecasts. Excess returns remain low in the post-crisis periods, suggesting that we may not reject the UIP in the long run. The success of LSTM networks implies that we can improve carry trade return forecast by a better use of information embodies in economic fundamentals. LSTM networks may help to construct more profitable carry trade portfolios. Given that it can explore the flexible interactions between explanatory factors, we expect that LSTM networks may improve forecast in other economic settings in future research.

References

[1] Accominotti, O., Cen, J., Chambers, D., Marsh, I. (2019). Currency Regimes and the Carry Trade. Journal of Financial and Quantitative Analysis, 54(5), 2233-2260.

[2] Bekaert, G., Hodrick, R. J. (2001). Expectations Hypotheses Tests. The Journal of Finance, 56(4), 1357-1394.

[3] Burnside, C., Eichenbaum, M., Rebelo, S. (2008). Carry Trade: The Gains of Diversification. Journal of the European Economic Association, 6(2-3), 581-588.

[4] Chinn, M. D. (2006). The (Partial) Rehabilitation of Interest Rate Parity in the Floating Rate Era: Longer Horizons, Alternative Expectations, and Emerging Markets. Journal of International Money and Finance, 25(1), 7-21.

[5] Colombo, E., Forte, G., Rossignoli, R. (2019). Carry Trade Returns with Support Vector Machines. International Review of Finance, 19: 483-504.

[6] Davis, S.J., Liu, D., Sheng, X. (2020). Economic Policy Uncertainty in China Since 1949: The View from Mainland Newspapers. Working paper.

[7] Diebold, F. X., Mariano, R.S. (1995) Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, 253-263.

[8] Engel, C. (2014). Exchange Rates and Interest Parity. In G. Gopinath, E. Helpman & K. Rogoff (Eds.), Handbook of International Economics (Vol. 4, pp. 453-522): Elsevier.

[9] Fischer, T., Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. European Journal of Operational Research, 270(2), 654-669.

[10] Gu, S., Kelly, B., Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. Review of Financial Studies, 33(5), 2223-2273.

[11] Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

[12] Jordà, Ò., Taylor, A. M. (2012). The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself. Journal of International Economics, 88(1), 74-90.

[13] Lustig, H., Verdelhan, A. (2007). The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk. American Economic Review, 97(1), 89-117.

[14] Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A. (2012). Carry Trades and Global Foreign Exchange Volatility. The Journal of Finance, 67(2), 681-718.

[15] Rossi, B. (2013). Exchange Rate Predictability. Journal of Economic Literature, 51(4), 1063-1119.

[16] West, K. D. (2012). Econometric Analysis of Present Value Models When the Discount Factor is Near One. Journal of Econometrics, 171(1), 86-97.

Online Appendix (Not for Publication)

LSTM networks. Below we first explain the information transmission within a memory cell, then show the full model structure.

We elaborate the mechanism in a memory cell in Figure 1. There are three types of gates that control how information flows in a memory cell: the input gate, the forget gate, and the output gate. Denoting the hidden state as h_t and the input as x_{t+1}, we define the combined input as $m_{t+1} = [h_t, x_{t+1}]$. First, the forget gate activation f_{t+1} determines how information in the combined input can be transmitted:

$$f_{t+1} = \sigma\left(W_f m_{t+1} + b_f\right), \tag{4}$$

where W_f is the forget gate weight, b_f is its associated bias, and the sigmoid function $\sigma(s) = \frac{1}{1+e^{-s}}$ ranges between 0 and 1. The forget gate passes little information to period $t+1$ if f_{t+1} is close to 0. Second, the input gate activation in_{t+1} is also a sigmoid function:

$$in_{t+1} = \sigma\left(W_{in} m_{t+1} + b_{in}\right), \tag{5}$$

where W_{in} is the input gate weight and b_{in} is the bias vector. The input gate allows more information to "enter" if in_{t+1} is close to 1. Third, denoting the cell state as c and the cell state candidate as g, its evolution process follows

$$g_{t+1} = tanh\left(W_g m_{t+1} + b_g\right), \tag{6}$$

$$c_{t+1} = f_{t+1} \cdot c_t + in_{t+1} \cdot g_{t+1}, \tag{7}$$

where W_g and b_g are the weight and bias respectively, and the state activation function $tanh(s) = \frac{e^{2s}-1}{e^{2s}+1}$ with range [-1,1] guarantees a reasonable cell state value. Overall, the forget gate f_{t+1}

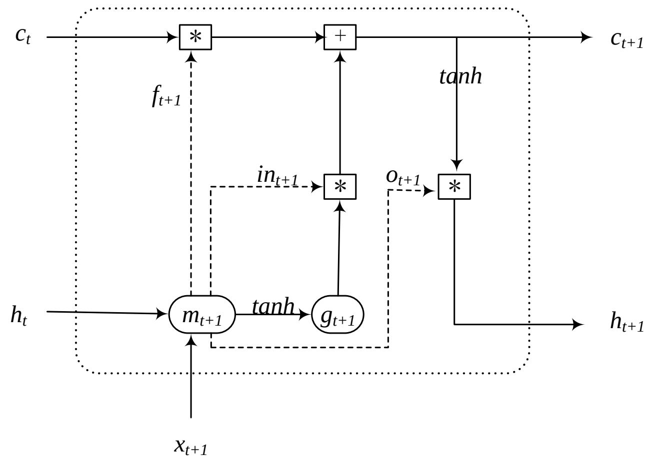

Fig. 1: A memory cell in LSTM networks

controls how the cell state c_t passes into c_{t+1}, and the input gate in_{t+1} controls how the combined inputs are incorporated into c_{t+1}. Finally, the output gate o_{t+1} and the hidden state h_{t+1} are

$$o_{t+1} = \sigma\left(W_o m_{t+1} + b_o\right), \tag{8}$$

$$h_{t+1} = o_{t+1} \cdot tanh\left(c_{t+1}\right), \tag{9}$$

where W_o is the output gate weight and b_o is its bias vector. The output gate determines how the information in the cell state c_{t+1} can be transmitted into the new hidden state. Following Sherstinsky (2020), let $\mathcal{Q}$ be an invertible map that transforms the hidden state to the output, then the output is

$$y_{t+1} = \mathcal{Q}(h_{t+1}). \tag{10}$$

Figure 2 illustrates how LSTM networks evolve over periods. LSTM networks update the cell state and the hidden state from c_t and h_t to c_{t+1} and h_{t+1} with the input x_{t+1}, and generate the

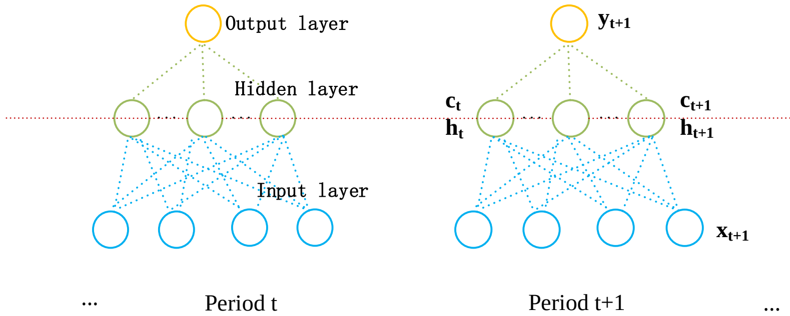

Fig. 2: LSTM networks in multiple periods

output y_{t+1}. This process iterates until the last period $t = T$.

During the evolving process, the model minimizes the loss function in the data. Given the output is the exchange rate movement $\hat{e}_{t+1}$, the mean squared error between the actual and predicted values is a natural loss function. Following Fischer and Krauss (2018), the number of parameters in LSTM layers is $4d(j+d)+4d$, where d and j are the numbers of hidden units and input features respectively. The term $4d(j+d)$ is the number of parameters in four weight matrices, and the term $4d$ is the number of parameters in bias.

In summary, the LSTM networks have four input features Δe_t, $\pi_t^* - \pi_t$, $i_t^* - i_t$, and $q_t - \bar{q}$ (all are stationary after unit root tests), and we choose time step 2 as the best fit from time step candidates 1 to 5. The hidden layer has eight hidden neurons, corresponding to 416 parameters. Zhang et al. (2017) show that neutral networks allow a relatively high ratio of the number of parameters over that of observations. The output layer contains two neurons as a standard configuration.

Note. LSTM networks may degenerate to its parent model, the recurrent neural network (RNN), by shutting down gates in (7) and (9):

$$c_{t+1} = W_c c_t + W_h h_t + W_x x_{t+1} + b_c,$$

$$h_{t+1} = tanh(c_{t+1}),$$

where W_c, W_h, W_x, and b_c are weight matrices and bias to be estimated. The RNN model also fits time series prediction but cannot optimize long and short term data transmission without gates.

References

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O. (2017) Understanding Deep Learning Requires Rethinking Generalization. International Conference on Learning Representations (ICLR). arXiv:1611.03530.