A machine learning approach to forecasting carry trade returns

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

August 23, 2020

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 regressions based on economic fundamentals; and (ii) excess carry trade returns deteriorate after the 2007–2008 global financial crisis in LSTM networks and other 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 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. 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]

This paper contributes to the literature on carry trade returns in two dimensions. First, we introduce long short-term memory (LSTM) networks—a machine learning model for time series forecasting—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 allow long- and short-term information to be optimally used in carry trade return prediction, as opposed to Colombo et al. (2019), who employ support vector machines to predict carry trade direction. Second, we find that carry trade returns deteriorate after the 2007–2008 global financial crisis in LSTM networks and in other models as the interest rate differentials shrink to near zero, consistent with the estimation in Accominotti et al. (2019). Our findings provide supportive evidence that we may not be able to reject the UIP.

Specifically 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. Hochreiter and Schmidhuber (1997) first developed LSTM networks that can "memorize" the long-term data pattern and determine how the input in every period can "enter" into the memory process. Therefore LSTM networks capture more data information than linear or threshold models and thus can potential improve return forecasts. 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

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

[2] See Gu et al. (2019) and Davis et al. (2020) on machine learning applications for time series problems.

the input can be included, and how the information can be sent out respectively. The performance of LSTM networks dominates other popular models for carry trade returns. All model forecasting results show that the excess carry trade returns are lower after the 2007–2008 global crisis.

2 Data, carry trades, and model

Data. The monthly data set that spans from January 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

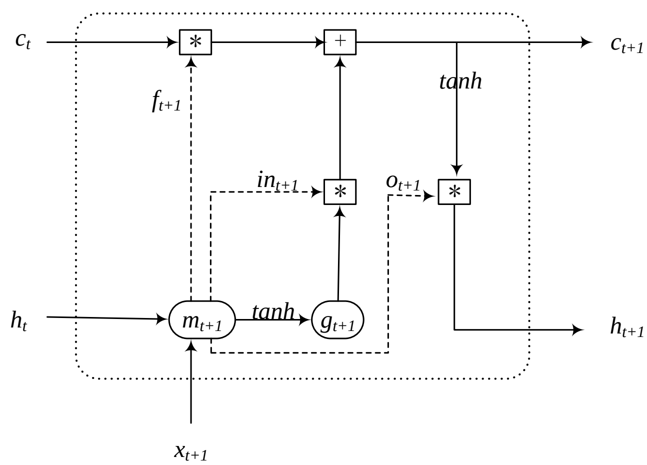

Fig. 1: A memory cell in LSTM networks

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.

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. 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, 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 g_{t+1} is the cell state candidate, 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} 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.

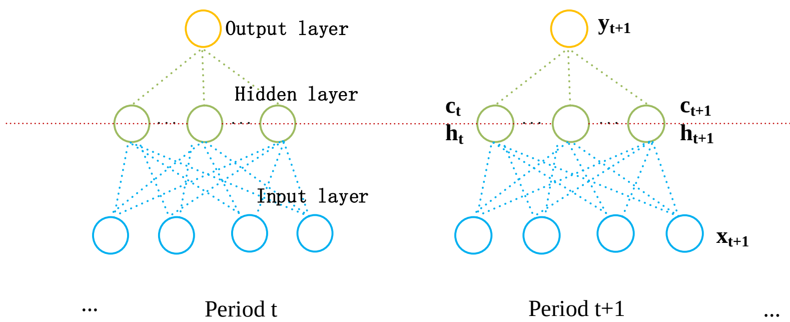

Fig. 2: LSTM networks in multiple periods

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 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. We choose the Adam algorithm (Kingma, et al., 2014) to minimize the loss function, because it is computationally efficient and performs well for a large number of parameters. Following Fischer and Krauss (2018), the number of parameters in LSTM layers is $4h(j + h) + 4h$, where h and j are the numbers of hidden units and input features respectively. The term $4h(j + h)$ is the number of parameters in four weight matrices, and the term $4h$ is the number of parameters in bias.

Predicting carry trade returns using LSTM networks. We divide the sample into the training set (in sample) from 1990 to 2011,[3] 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 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.[4] In the test data, we employ the rolling window strategy with one-step ahead forecast.

[3]A rule of thumb for splitting between training and test sets is 5:1. Results are robust with other splitting.
[4]Results are consistent under an alternative 80% training versus 20% validation rule.

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 rebalance the equal-weighted portfolio. Then we calculate the realized return from the exchange rate data.

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, while there are 2097 observations in the training sample. 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.

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.

We choose random walk, vector autoregression (VAR), and the threshold vector error correction model (TECM) as the benchmark models to assess LSTM networks' forecasting ability. The random walk model assumes that the exchange rate movement Δe_{t+1} is independently and identically distributed. In contrast, the VAR model with the optimal lag as one period assumes that economic fundamentals are capable of forecasting future exchange rate movement. The TECM (one period as the optimal lag order) with the demeaned real exchange rate $q_t - \bar{q}$ as the cointegration variable assumes parsimonious thresholds—whether the interest rate differential and the real exchange rate are above their median values.

Table 1 displays the model performance comparison in forecasting carry trade returns. In all three rolling windows, LSTM networks dominate the three other models. The Diebold and Mariano (1995) test results show that LSTM networks are more accurate in return predictions.[5]

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

Table 1: Carry trade return

	Random Walk	VAR	TECM	LSTM
Realized profits 2012–2015				
Mean	-0.0009	-0.0008	-0.0006	0.0014
Std. dev.	0.0177	0.0116	0.0118	0.0100
Skewness	0.1053	-0.4649	-0.5411	0.2139
Sharpe ratio (annual)	-0.1863	-0.2750	-0.1806	0.4636
Gain/loss	0.8550	0.8698	0.8720	1.4427
D-M test *p*-value	0.0000	0.0000	0.0001	-
Realized profits 2013–2016				
Mean	-0.0004	-0.0003	-0.0010	0.0028
Std. dev.	0.0181	0.0106	0.0121	0.0077
Skewness	0.0328	-0.3385	-0.4850	0.5110
Sharpe ratio (annual)	-0.0807	-0.1262	-0.3181	1.2413
Gain/loss	0.9505	0.9189	0.7817	2.4960
D-M test *p*-value	0.0000	0.0000	0.0001	-
Realized profits 2014–2017				
Mean	0.0009	-0.0003	-0.0011	0.0024
Std. dev.	0.0183	0.0104	0.0134	0.0073
Skewness	-0.1544	-0.2916	-0.3871	0.9213
Sharpe ratio (annual)	0.1702	-0.1091	-0.3004	1.1308
Gain/loss	1.1022	0.9141	0.7891	2.5344
D-M test *p*-value	0.0000	0.0000	0.0000	-

The LSTM networks in 2012–2015 generate an average monthly return of 14 basis points with low standard deviation and positive skewness, which is better than the negative returns from other models. The average return, Sharpe ratio, and 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. However, the highest annual returns in the three windows 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.

In Table 2, we further investigate when the LSTM networks outperform other models. The near-zero interest rate differentials between 2012 and 2017 are not the main profit contribution component, and thus the accuracy of exchange rate prediction largely determines the realized carry trade returns. 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. Utilizing TECM as the comparison model, LSTM networks perform better when two 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. Moreover, carry trade returns through LSTM forecasts exhibit positive skewness, while those under TECM have negative skewness. Consequently, the LSTM networks help to avoid "bad carries" with low returns and negative skewness and thus improve model performance as pointed out by Bekaert and Panayotov (2020).

Table 2: Comparison of LSTM and TECM

	Where models agree		Where models disagree	
	LSTM	TECM	LSTM	TECM
Realized profits 2012–2015				
Number of observations	206	206	217	217
Mean	0.0001	0.0001	0.0030	-0.0018
Std. dev.	0.0267	0.0267	0.0274	0.0274
Skewness	-0.1538	-0.1538	0.1930	-0.1652
Max	0.0709	0.0709	0.0794	0.0793
Min	-0.0871	-0.0871	-0.0784	-0.0772
Realized profits 2013–2016				
Number of observations	202	202	221	221
Mean	0.0001	0.0001	0.0037	-0.0027
Std. dev.	0.0278	0.0278	0.0276	0.0276
Skewness	-0.0967	-0.0967	0.1191	-0.0993
Max	0.0893	0.0893	0.0774	0.0793
Min	-0.0871	-0.0871	-0.0784	-0.0770
Realized profits 2014–2017				
Number of observations	197	197	199	199
Mean	0.0003	0.0003	0.0045	-0.0033
Std. dev.	0.0282	0.0282	0.0270	0.0271
Skewness	-0.0700	-0.0700	0.1788	-0.1731
Max	0.0893	0.0893	0.0773	0.0776
Min	-0.0871	-0.0871	-0.0784	-0.0770

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.

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., Panayotov, G. (2020). Good Carry, Bad Carry. Journal of Financial and Quantitative Analysis, 55(4), 1063-1094.

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

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

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

[7] 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.

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

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

[10] 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.

[11] Kingma, D. P., Ba, J. (2014). Adam: A Method for Stochastic Optimization.

[12] 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.