UNIVERSITY OF TURKU

ABSTRACT

	Bachelor's thesis
x	Master's thesis
	Licentiate's thesis
	Doctor's thesis

Subject	Accounting and finance	Date	12.9.2019
Author	Eero Mäenpää	Student number	511126
		Number of pages	85
Title	Machine learning in FX trading with long short-term memory recurrent neural networks		
Supervisors	Prof. Mika Vaihekoski and M.Sc. Valtteri Peltonen		

Abstract

Advanced machine learning has outperformed previous benchmarks in numerous fields and applications, e.g. speech and image recognition, in an ever accelerating way during the past two decades. Furthermore, early evidence indicates its massive potential to improve over traditional techniques applied to financial market prediction tasks.

In this study, long short-term memory (LSTM) networks are applied to G10 currency market prediction task in a sample period from the beginning of 1999 to the end of 2018. The predictive factor set consists of past excess returns, forward premiums and seven factors capturing risk-aversion, price uncertainty, commodity returns and funding liquidity. The LSTM model reaches a total accuracy of more than 55% outperforming simple recurrent neural networks (RNN) which indicates that currency excess returns are partly driven by signals with more than a two-month temporal distance.

The model predictions are used in US dollar-neutral currency trading strategy in order to exploit the attractiveness of this approach in terms of performance metrics and to analyse its sources of profitability. The LSTM model is able to predict the profitability of carry and momentum strategies. In particular, the LSTM portfolio utilizes carry and long-term momentum signals during calm market conditions and short-term momentum signals in market turmoil. Consequently, the best performing LSTM portfolio delivers a Sharpe ratio of 0.32 with less tail risk than the carry portfolio that has a Sharpe ratio of 0.23. Furthermore, it is not exposed to the common risk factors driving currency carry and momentum trading strategies.

Attractive risk-return profile and low correlation with equity markets make the LSTM portfolio extremely suitable to the FX risk management of an international equity portfolio. When the LSTM portfolio is used in the modified portfolio mean-variance optimization routine that is introduced by Boudoukh et al. (2018), the Sharpe ratio of an unhedged international equity portfolio is almost doubled as the Sharpe ratio increases from 0.29 to 0.51.

Key words	Foreign exchange market, machine learning, forecasting
Further information	

TURUN YLIOPISTO

TIIVISTELMÄ

	Kandidaatintutkielma
x	Pro gradu -tutkielma
	Lisensiaatintutkielma
	Väitöskirja

Oppiaine	Laskentatoimi ja rahoitus	Päivämäärä	12.9.2019
Tekijä	Eero Mäenpää	Matrikkelinumero	511126
		Sivumäärä	85
Otsikko	Koneoppiminen ja kaupankäyntistrategiat valuuttamarkkinoilla		
Ohjaajat	Prof. Mika Vaihekoski ja KTM Valtteri Peltonen		

Tiivistelmä

Koneoppimisen sovellutukset ovat mahdollistaneet yhä uusien käytännön ongelmien, kuten puheen tai kuvien tunnistamisen, tehokkaan ratkaisemisen. Kahden viime vuosikymmenen aikaisten kehitysaskelien taustalla ovat muun muassa nopea koneoppimisen menetelmien kehitysvauhti ja laskentatehon kasvu. Koneoppimisen menetelmät ovat osoittautuneet monia perinteisiä menetelmiä tehokkaammiksi myös rahoitusmarkkinoiden ennustamisessa.

Tässä tutkielmassa tutkitaan LSTM-verkkojen soveltuvuutta G10-valuuttojen tuottojen ennustamiseen vuosien 1999-2018 aikana. LSTM-verkon ominaisuudet ja valikoidut syötteet mahdollistavat tuotonlähteiden löytämisen tuottotrendien, korkoeron, riskiaversion, hintaheilunnan, hyödykemarkkinan tuottojen ja likviditeetin ajallisista ja keskinäisistä suhteista. LSTM-verkolla saavutetaan yli 55 % ennustetarkkuus, joka ylittää selkeästi RNN-verkon ennustetarkkuuden. Tästä syystä osa valuuttamarkkinan tuottoja ennustavista signaaleista tunnistetaan yli kaksi kuukautta ennen tuoton realisoitumista.

LSTM-verkon ennusteiden pohjalta rakennetaan kaupankäyntistrategia niiden taloudellisen merkityksen ja tuotonlähteiden selvittämiseksi. Malli onnistuu maiden välistä korkoeroa hyödyntävän carry ja momentum -strategioiden voitollisuuden ennustamisessa. Rauhallisen nousumarkkinan aikana LSTM-verkon ennusteiden pohjalta luodun strategian tuottoja selittävät carry ja pitkän aikavälin momentum -strategiat, kun taas epävarmuuden kasvaessa tuottoja selittää lyhyen aikavälin momentum. Kaupankäyntistrategian Sharpen luku on 0.32, kun carry strategia yltää vain lukuun 0.23. Lisäksi LSTM-verkon ennusteiden pohjalta luotu strategia vähentää merkittävästi carry-strategialle ominaista suurta häntäriskiä ja altistusta carry ja momentum -strategioiden tuottoja selittäville riskitekijöille.

LSTM-verkon avulla saavutetut hyödyt eivät rajoitu vain spekulatiiviseen kaupankäyntiin. Strategian soveltaminen kansainvälisesti hajautetun osakeportfolion valuuttakurssiriskin optimaaliseen suojaamiseen lähes kaksinkertaistaa tuotto-riski-suhteen suojaamattomaan portfolioon verrattuna.

Asiasanat	Valuuttamarkkinat, koneoppiminen, ennustaminen
Muita tietoja	

UNIVERSITY OF TURKU
Turku School of Economics

MACHINE LEARNING IN FX TRADING WITH LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORKS

Master´s Thesis
in Accounting and Finance

Author:
Eero Mäenpää

Supervisors:
Prof. Mika Vaihekoski
M.Sc. Valtteri Peltonen

12.9.2019
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

CONTENTS

1 INTRODUCTION 9
1.1 Motivation 9
1.2 Objectives and structure 11

2 CURRENCY TRADING STRATEGIES 14
2.1 Carry 14
2.2 Momentum 19

3 LSTM NEURAL NETWORKS 24
3.1 Recurrent neural networks 24
3.2 Long short-term memory cells 26
3.3 Market prediction 28

4 METHODOLOGY 31
4.1 Data and definitions 31
4.2 Prediction routine 34
4.3 Performance evaluation 38
4.4 Portfolio construction 39
4.5 FX risk hedging 44

5 RESULTS 47
5.1 Sample description 47
5.2 LSTM and RNN predictions 56
5.3 LSTM portfolios 60
5.4 Application to FX risk management 70
5.5 Practical aspects of implementation 73

6 CONCLUSIONS 79

REFERENCES 81

FIGURES

Figure 1 Unfolded recurrent neural network (RNN) 25

Figure 2 Long short-term memory (LSTM) cell 27

Figure 3 LSTM model optimal parameter value selection 37

Figure 4 Cumulative excess currency returns 48

Figure 5 Currency forward premiums 49

Figure 6 Common factors 50

Figure 7 LSTMs and RNNs 26-week rolling accuracies 59

Figure 8 All minus RX 26-week rolling accuracies 59

Figure 9 Cumulative compounded LSTM portfolio returns 64

Figure 10 LSTM (All) optimization based and robust, carry and 3-month momentum cumulative compounded portfolio returns 64

Figure 11 LSTM (All) optimization based and robust, carry and 3-month momentum cumulative compounded portfolio returns during NBER recession 68

Figure 12 Cumulative compounded LSTM portfolio returns during NBER recession 69

Figure 13 Cumulative compounded international equity portfolio returns . 72

Figure 14 Cumulative compounded international equity portfolio returns during the financial crisis 73

TABLES

Table 1 Summary of the predictive factors of currency carry trade strategies 15

Table 2 Summary of the predictive factors of currency momentum strategies 20

Table 3 Summary statistics for the currency excess returns 47

Table 4 Summary statistics for the common factors 49

Table 5 In-sample OLS regression analysis with a one-week lag 51

Table 6 Correlation matrix of the common factors 52

Table 7 Latent factor loadings 53

Table 8 Univariate OLS regression R^2s of latent factor explanatory analysis 54

Table 9 Proportional and cumulative proportional variances of the latent factors 54

Table 10 In-sample OLS regression analysis with the latent factors and a one-week lag 55

Table 11 LSTM and RNN accuracies and probabilities that a model would have achieved the total accuracy by chance 57

Table 12 Diebold–Mariano test of model outperformance 58

Table 13 In-sample OLS regression analysis of LSTM accuracy differences with a lag of one 60

Table 14 Summary statistics of LSTM portfolio returns 61

Table 15 In-sample OLS regression analysis with the latent factors and a one-week lag 62

Table 16 Explanatory analysis of the LSTM portfolios 65

Table 17 In-sample OLS regression analysis with the latent factors during NBER recession 66

Table 18 Explanatory analysis of the LSTM portfolios during NBER recession 68

Table 19 Summary statistics of the international equity portfolio returns 70

Table 20 Correlation between equity and alpha seeking currency portfolio returns 72

Table 21 Currency portfolio transactions 74

Table 22 Summary statistics of international equity and currency portfolio returns after leverage constraint 76

Table 23 Currency portfolio transactions after the leverage constraint 78

1 INTRODUCTION

1.1 Motivation

An investor holding global equity portfolio has exposure to foreign currency (FX) risk. It is possible to fully hedge this risk, do nothing or actively manage FX exposures. Even though foreign currencies can be a remarkable source of volatility, the risk-return profile of FX and international equity portfolios can be enhanced through active management of currency exposures. In particular, this research investigates empirically if advanced machine learning can be used to exploit cutting-edge trading strategies in currency markets and further utilized in FX risk management framework of an international equity portfolio.

In general, actively managed foreign currency positions are held for two reasons: speculation and risk management. For speculation purposes, well researched sources of systematic returns are carry, momentum and value (Asness et al. 2013). In the carry trading strategy, one takes short positions in currencies with low interest rate and long positions in high interest rate currencies, whereas in the momentum strategy one goes long in currencies with the strongest recent performance and shorts currencies with the lowest performance in recent history. Finally, in the value strategy, long-term mean-reversion behaviour is utilized.

Regarding risk management, academics do not agree on an optimal approach for managing foreign currency risk. However, the majority of studies find evidence that support active currency management for equity investors (see e.g. Boudoukh et al. (2018), Barroso and Santa-Clara (2015), Topaloglou et al. (2011), Campbell et al. (2010), Jorion (1994), Glen and Jorion (1993) and Black (1989)). On the contrary, Froot (1993) asserts that long-term equity investors should not hedge their foreign currency exposures. His reasoning relies on long-term mean-reverting behaviour of real exchange rates which acts as a natural hedge. However, Schmittmann (2010) does not find enough evidence to support Froot's arguments.

Boudoukh et al. (2018) introduce a currency hedging procedure that almost doubles the Sharpe ratio of an unhedged global equity portfolio by combining speculative and risk minimizing currency portfolios into a dynamic currency allocation framework. They divide a global equity portfolio into the three following components: a fully hedged global equity portfolio, a foreign currency portfolio with minimum variance and a speculative foreign currency portfolio that combines carry, momentum and reversal signals. Boudoukh et al. (2018) suggest optimization for the construction of the minimum variance component but a heuristic approach for the speculative, alpha seeking component. However, they point out

that a method based on optimization might outperform their approach applied to the speculative portfolio if more advances techniques are utilized.

Boudoukh et al. (2018) rely on constant percentage allocation between the speculative currency portfolio exposures. However, Filippou and Taylor (2017), Barroso and Santa-Clara (2015) and Menkhoff et al. (2012a) argue that the profitability of currency carry and momentum trading strategies are time-varying and partly explained by financial and macroeconomic factors. Barroso and Santa-Clara (2015) develop a dynamic multi-style currency trading strategy based on momentum, carry and reversal. They state that the correct factors explaining the profitability of the strategy are not yet identified due to a high and statistically significant alpha. Consequently, Barroso and Santa-Clara (2015) argue that the most reasonable explanation for the returns is anomaly. However, their strategy to some extent is exposed to liquidity risk, equity volatility, foreign exchange volatility, excess returns on the equity market and carry trade risk. These findings are in line with Lustig et al. (2011) who argue that the returns on carry trade strategies are driven by global equity market volatility. Furthermore, Menkhoff et al. (2012a) demonstrate that carry and momentum strategies are exposed to foreign exchange volatility and liquidity risk.

The dynamic allocation routines developed by Filippou and Taylor (2017) and Barroso and Santa-Clara (2015) rely on models that are not able to capture potentially complex dependencies between the predictive factors and currency returns. In fact, advanced machine learning methods have been lacking broader popularity in financial market prediction despite of the tremendous amount of potential predictors presented by the financial market research (Krauss et al. 2017; Jacobs 2015). Exploiting a vast amount of predictive factors makes standard linear models useless due to, for example, collinearity and decreasing degrees of freedom (Gu et al. 2018). Furthermore, these techniques are not able to incorporate the complex and likely non-linear relationships in the data (Heaton et al. 2018).

Gu et al. (2018) investigate various machine learning models in time-series prediction of stock returns of almost 30,000 individual companies between 1957 and 2016. Furthermore, the data set used for prediction consist of more than 900 predictors. Gu et al. (2018) prove that non-linear models significantly improve the performance over linear models. In addition, they develop a long-short equity strategy with the S&P 500 index constituents based on the predictions of their 3-layer neural network architecture. They demonstrate that this method generates an exceptional Sharpe ratio of 2.35 in contrast to a Sharpe ratio of 0.89 generated by a benchmark linear method.

Krauss et al. (2017) compare three different statistical learning methods and their ensembles in stock market prediction. They show that an ensemble method

that combines three individual techniques: deep neural networks, gradient boosted trees and random forests outperforms each of the individual models. Furthermore, they construct a long-short strategy based on their best performing model which generates a Sharpe ratio of 1.81 after transaction costs. The Sharpe ratio is five times larger than the Sharpe ratio of the S&P 500 index during the sample period of 1992–2015. Fischer and Krauss (2018) show that a long short-term memory (LSTM) recurrent neural network (RNN) further improves the performance of the best performing model suggested in Krauss et al. (2017).

Filippou and Taylor (2017), Barroso and Santa-Clara (2015), Menkhoff et al. (2012a) and Lustig et al. (2011) provide evidence regarding the predictability of currencies and currency trading strategies. Moreover, the success of machine learning techniques exploiting non-linear and temporal dependencies in stock market prediction tasks raises the question, whether advanced statistical learning techniques could improve over previous benchmarks in currency markets or not (Gu et al. 2018; Fischer and Krauss 2018; Krauss et al. 2017).

1.2 Objectives and structure

The focus of this research is to test if the LSTM discovers profitable patterns in selected time-series data which could help to exploit trading opportunities in currency markets. Furthermore, if the model is able to extract these predictive signals, a currency trading strategy based on this information is constructed and applied to the FX risk management framework of a global equity portfolio. Other models or techniques are not considered, except the simple RNN, which performance is compared to the LSTM in order to find out if the complexity added by the latter is necessary.

The research objectives of this study are investigated through an empirical analysis as follows. First, the LSTM model is trained to predict the probability that a currency outperforms the cross-median of other currencies in the next time step in order to investigate the predictive power of the model. Next, two techniques are compared in the construction of the LSTM portfolios. The first portfolio is constructed as an equally weighted long-short portfolio based on the predicted outperformance probabilities. In the second technique, the output predictions of the LSTM model are ranked and historical scenarios are modified with entropy pooling introduced by Meucci (2008). The ranked inputs are handled as outperformance constraints as suggested by Meucci et al. (2014). Finally, the mean-variance optimal portfolio weights are computed based on estimated moments from the modified scenarios. The latter technique should solve the problem

in Fischer and Krauss (2018) as their portfolio construction technique overweights high volatility assets.

The factors used as inputs to the LSTM are chosen based on previous research on the predictability of currency markets (Londono and Zhou 2017; Menkhoff et al. 2017; Orlov 2016; Asness et al. 2013; Mancini et al. 2013; Bakshi and Panayotov 2013; Menkhoff et al. 2012a,b; Brunnermeier et al. 2009). Selected predictive factors are currency returns, forward premium, equity and foreign exchange market volatilities, funding liquidity, variance risk premium, currency variance risk premium, commodity price changes and credit risk. In fact, these factors cover the three most informative classes of predictors found in an analysis of over 900 different variables in equity market return prediction problem by Gu et al. (2018). Finally, aggregate variables, like 12-month momentum, are not used since the LSTM should be able to extract these patterns from the historical data. This is also empirically proven by Fischer and Krauss (2018).

The empirical analysis is conducted from the viewpoint of an investor whose home currency is the United States dollar (USD). Other base currencies are not taken into consideration. The currency pairs are formed from the G10 currencies: Euro (EUR), Pound sterling (GBP), Japanese yen (JPY), Australian dollar (AUD), New Zealand dollar (NZD), Canadian dollar (CAD), Swiss franc (CHF), Norwegian krone (NOK) and Swedish krona (SEK), which are denoted against the United States dollar (USD). The sample data covers the period from the 2nd of January 1999 to the 31st of December 2018. The prediction task and portfolio construction are conducted with a weekly observation frequency in order to solve the problems concerning different time-zones.

The FX risk management framework is applied to an international equity portfolio because currencies with negative correlation to global equity provide diversification and risk reduction opportunities. The MSCI World Index, that tracks the market capitalization weighted peformance of the equity market in 23 developed markets countries, is used as a proxy for the international equity portfolio (MSCI 2019). However, the framework investigated in this research is not applied to bond markets since Campbell et al. (2010) show that the majority of currencies are uncorrelated with international bond portfolios. Consequently, risk-minimizing bond investors should fully hedge their currency exposures.

Finally, practical aspects regarding leverage and turnover are covered as in Boudoukh et al. (2018). Overall, the approach developed in this study is applicable in practise. Consequently, the work could be implemented by asset managers and investors who want to pursue active foreign currency management either through hedging or speculation.

Combining portfolio construction and advanced predictive modelling with ma-

chine learning have not been widely studied. One explanation to this could be that it requires combining advanced techniques from multiple areas of expertise. Consequently, the key contributions of this study can be divided into four parts. First, the most suitable machine learning techniques and model architectures to time-series prediction tasks in financial markets are investigated and proposed. Second, an optimization based portfolio construction routine, that is designed to overcome flaws detected in the previous research regarding portfolio construction of rules based long-short trading strategies, is introduced. Third, the understanding of the sources of the profitability of the machine learning portfolios is increased through the explanatory analyses. Demystifying advanced machine learning methods should facilitate the deployment of these techniques among investment professionals. Fourth, it is showed that the currency trading strategies developed in this study are not only limited to currency speculation but are able to enhance the foreign exchange risk management of any investor with international equity exposures. Furthermore, in a broader academic context, this research contributes to theories on market efficiency and forward premium puzzle. The former theory has been under debate since it was first introduced by Fama (1965) and the latter has remained unsolved.

The rest of this research is organized as follows. Section 2 introduces the two most popular currency trading strategies: carry and momentum. Furthermore, the literature regarding the predictability of these strategies and currency markets in general is discussed in this section. Next, Section 3 covers feed-forward neural networks, recurrent neural networks, long short-term memory cells and discusses the literature covering machine learning in financial market prediction tasks.

The empirical part is divided into three parts. First, the methodology is introduced by going through practical aspects and choices made regarding the data and techniques. Second, Section 5 presents and discusses the results and, finally, Section 6 concludes and guides the future research.

2 CURRENCY TRADING STRATEGIES

2.1 Carry

In the carry trading strategy one goes long in currencies with high interest rates and shorts currencies with low interest rates. These currencies are called investment and funding currencies, respectively. The carry trade pay-off consists of two components which are ex-ante known interest rate differential and uncertain price movement off the currency pair. Filippou and Taylor (2017) demonstrate that the carry trade strategy has offered an annualized mean excess return w.r.t the USD of 4.24% and 2.79% for a sample of 48 currencies and a sample of the G11 currencies (G10 plus DKK), respectively, during the time period from 1985 to 2012. Furthermore, they report Sharpe ratios of 0.46 and 0.27 for the carry trade strategies formed with these two currency samples.

Currency forwards are priced based on no-arbitrage conditions. Thus, their prices should follow the covered interest rate parity (CIR)

$$\frac{f_t^{ij}}{s_t^{ij}} = \frac{1 + r_t^i}{1 + r_t^j}, \tag{1}$$

where f_t^{ij} and s_t^{ij} denote the forward and spot exchange rates for a unit of currency i per a unit of currency j at time t. In addition, r_t^i and r_t^j denote the nominal interest rates of zero coupon bonds denominated in currencies i and j, respectively. Equation 1 holds for an arbitrary maturity. However, it is assumed that the maturities of these default risk-free bonds and forward contracts coincide. (Fama 1984, 322).

If f_t^{ij} is replaced with $E_Q[s_{t+1}^{ij}]$, where Q is a risk-neutral measure, in Equation 1, we get the uncovered interest rate parity (UIP). The UIP implies that currencies with high interest rate should depreciate over time, whereas currencies with low interest rate should appreciate. According to this theory, expected carry trade pay-off should be zero, since the gain from the interest rate differential should be exactly offset by the appreciation off the funding currency against the investment currency. (Brunnermeier et al. 2009, 313). However, Fama (1984) finds that the forward premium

$$f_t^{ij} - s_t^{ij}$$

has actually a negative coefficient when it is used to explain spot prices in the future. Consequently, the funding currency is expected to depreciate against the investment currency. This unexplained phenomenon is known as the forward premium puzzle and it gives rise to the carry trade strategy. Since the empir-

ical findings are against the UIP, carry trades should have a positive pay-off on average. (Brunnermeier et al. 2009, 313).

In this research, all currency returns are calculated w.r.t. the USD unless stated otherwise. Consequently, an appreciation or depreciation of a currency means that it appreciates or depreciates against the USD. However, carry trade positions are not limited to trades relative to investors' home currencies. An investor with a home currency A could take advantage of a country B's high interest rate by funding the long position in a bond denominated in B's currency with a loan denominated in country C's currency. Alternatively, the exact same position could be composed using currency futures or OTC forwards.

Academics have tried to explain the sources of carry trade profits by introducing several explanatory factors and theories. These factors are summarized in Table 1. The coefficient column presents the sign of a factor coefficient linking the carry trade returns to the relevant factor. Furthermore, the time frames short, mid and long denote immediate or intra-day, one-day to one-month and one-month to 12-month lags, respectively.

Table 1: Summary of the predictive factors of currency carry trade strategies

Factor	Coefficient sign	Time frame	Evidence provided by
Equity market volatility	-	short	Lustig et al. (2011) , Brunnermeier et al. (2009)
Equity market volatility	+	mid and long	Brunnermeier et al. (2009)
Illiquidity	-	short	Bakshi and Panayotov (2013), Mancini et al. (2013), Menkhoff et al. (2012a), Brunnermeier et al. (2009)
Illiquidity	+	mid and long	Brunnermeier et al. (2009)
Currency market volatility	-	short and mid	Bakshi and Panayotov (2013) Menkhoff et al. (2012a)
Variance risk premium	+	mid and long	Londono and Zhou (2017)
Currency variance risk premium	-	mid and long	Londono and Zhou (2017)
Commodity prices	+	short and mid	Bakshi and Panayotov (2013)

Lustig et al. (2011) identify that ex post over 80% of the variance in foreign currencies relative to the USD is explained by two factors which they label as a level and a slope factor. Their findings are based on principal component analysis. However, they show that the latent level and slope factors, that account for 70% and 12% of the total variance, respectively, are followed by the appreciations and depreciations of the USD against a broad basket of foreign currencies and a carry trade factor that is proxied by the returns of a long-short carry trade strategy. Furthermore, Lustig et al. (2011) demonstrate that the results produced by the carry trade factor could be replicated with a factor describing global equity market volatility. They demonstrate that investment currencies tend to depreciate and funding currencies tend to appreciate as equity market volatility increases. Con-

sequently, by exploiting carry trades, investors increase their exposure to global risk.

Brunnermeier et al. (2009) argue that currency carry trades have high tail risk. They illustrate that investment currencies have high negative skewness when paired with funding currencies. In particular, these findings hold in the cross section of currency pairs despite using a physical skewness measure or risk-neutral option implied skewness. However, Menkhoff et al. (2012a) do not find that sensitivity to changes in carry trade portfolio's skewness would be statistically significant explanatory factor even though the sign of the factor price supports the explanation made by Brunnermeier et al. (2009).

Despite the high tail risk of carry trade positions, Jurek (2014) concludes that only one-third of the currency carry trade excess returns are explained by the tail risk premium. Consequently, carry trade returns are not explained by peso problems and remain high even after hedging against tail events. The evidence provided by Menkhoff et al. (2012a) further support this explanation as they show that foreign exchange volatility rules the premiums incorporated in currency prices even after excluding extreme events.

In addition, Brunnermeier et al. (2009) demonstrate that highly negative implied skewness of currency carry trade returns predicts less negative physical skewness for the future. This is in line with the global risk and funding liquidity phenomena. After a tail event, implied skewness becomes more negative as traders are willing to pay more for hedging. Furthermore, decreased positions in the investment currencies lower the future tail risk and, consequently, affect the physical skewness. In addition, Jurek (2014) argues that following a run of positive carry trade returns, physical skewness becomes more negative and implied skewness becomes more positive. Consequently, hedging tail risk is cheap before a tail event and expensive after the tail event has occurred.

Brunnermeier et al. (2009) demonstrate that cross section of currency carry trades are correlated and, thus, driven by the same systematic factors. They find co-movement amongst the investment currencies with high interest rates and co-movement amongst the funding currencies with low interest rates. Brunnermeier et al. (2009) provide broader risk based explanation to carry trade returns than Lustig et al. (2011) as they argue that movements of the carry trade components are driven by global liquidity and funding risks. However, in line with Lustig et al. (2011), Brunnermeier et al. (2009) show that carry trade strategies experience losses as global equity market risk increases. Furthermore, they conclude that increased global risk or risk aversion leads to tighter funding liquidity and the decapitalization of carry trade positions. Brunnermeier et al. (2009) explain that this implies losses in carry trades and increases hedging costs of tail risk. Furthermore,

tighter funding liquidity and carry trade losses increase the margin requirements which forces traders to unwind their carry trade positions even further.

Mancini et al. (2013) provide evidence from the period of 2007–2009 financial turmoil that support the importance of the liquidity factor in explaining carry trade return dynamics. They argue that the liquidities of foreign currencies co-move over time and, thus, are able to identify a market-wide currency liquidity factor. Moreover, Mancini et al. (2013) show that more liquid currency pairs such as EUR/USD and USD/JPY are less sensitive to the liquidity factor, whereas this is opposite for less liquid currency pairs such as AUD/USD and USD/CAD. They further examine that by exploiting carry trade strategies, investors load liquidity risk.

However, Brunnermeier et al. (2009) demonstrate that tighter funding liquidity and increased global risk predict positive carry trade returns in the future. Londono and Zhou (2017) report similar findings as they conclude that increases in variance risk premium predicts appreciation of investment currencies and depreciation of funding currencies with a one-month horizon. They approximate the variance risk premium as a difference between stock option-implied and realized variances.

In addition, Londono and Zhou (2017) identify that currency variance risk premium captures alternative dynamics of currency returns. They approximate the currency variance risk premium as a difference between currency option-implied and realized variances. Londono and Zhou (2017) find that an increase in global currency variance risk premium predicts a depreciation of the investment currencies relative to the USD, and that the predictive power is the largest with a four-month horizon. They depict this phenomenon as a flight-to-quality or a flight-to-liquidity effect. On the other hand, the predictive power of the variance risk premium is statistically significant only for the investment currencies.

Bakshi and Panayotov (2013) provide evidence regarding the time-series predictability of currency carry trades. They show that carry trade returns are driven by commodity returns, average currency volatility and global liquidity. Bakshi and Panayotov (2013) state that commodity returns are associated with a positive slope to returns from commodity currencies, like the AUD and the NZD. These currencies are usually used as investment currencies in carry trades. Furthermore, they demonstrate that increased average currency volatility or lower liquidity predicts lower carry trade returns with a one-month lag. Based on these findings, Bakshi and Panayotov (2013) construct a currency carry trade strategy that utilizes the trading signals generated by the commodity price, average currency volatility and global liquidity factors. This strategy enhances the Sharpe ratio and makes the skewness less negative compared to an unconditional currency carry trade strategy.

In addition, Bakshi and Panayotov (2013) argue that the predictive power of the VIX index or average equity market volatility, which are found to be significant explanatory factors of carry trade returns by Lustig et al. (2011) and Brunnermeier et al. (2009), diminishes after adding the commodity returns, average currency volatility and global liquidity factors. Bakshi and Panayotov (2013) further conclude that this holds for average long term rates and average term spreads after adding the three factors supported by their research.

Bakshi and Panayotov (2013) propose that the predictability of currency carry trade returns is due to the predictability of the investment currencies. This is in line with the findings regarding the currency variance risk premium investigated by Londono and Zhou (2017) as the predictive power of this factor is statistically significant only for the investment currencies.

In contrast to Bakshi and Panayotov (2013), Mancini et al. (2013) and Brunnermeier et al. (2009), Orlov (2016) finds that liquidity doesn't affect currency carry trade returns. Specifically, Orlov (2016) focuses on aggregate equity market liquidity, whereas Bakshi and Panayotov (2013) and Brunnermeier et al. (2009) use TED spread as a proxy for global liquidity. Orlov (2016) constructs the aggregate market liquidity measure as a value-weighted average of a monthly illiquidity measure developed by Amihud (2002). Amihud's illiquidity measure is an absolute value of a stock return divided by the trading volume of that stock. Consequently, it differs from the TED spread as this is more related to funding liquidity and credit risk as it captures changes in the creditworthiness of major banks. However, the findings provided by Orlov (2016) have some support from Bakshi and Panayotov (2013) and Menkhoff et al. (2012a), since they argue that the predictive power of various liquidity measures decreases significantly after currency volatility is taken into account. In addition, Mancini et al. (2013) focus only on the illiquidity effects during the market turmoil of 2007–2009. They admit that the importance of liquidity as an explanatory factor for carry trade returns might decrease during normal market conditions, since currency markets are usually characterized by high volumes and are the most liquid financial asset class.

Lustig et al. (2014) compare the high-minus-low and a dollar carry trade strategies. The dollar carry trade strategy consists of two components: USD and a basket of foreign currencies. In this strategy, one takes a short (long) position in USD whenever the level of the short-term U.S. treasury rate is lower (higher) than the average short-term interest rate of the basket of foreign countries and takes a long (short) position in the foreign currencies. Lustig et al. (2014) show that the dollar carry and high-minus-low carry trade strategies are uncorrelated and driven by different factors. They argue that the dollar carry trade returns are exposed to the U.S. economy related domestic factors, like industrial production

growth, whereas the high-minus-low currency carry trades are exposed to global risk factors. These findings are supported by Filippou and Taylor (2017). They derive latent factors from an extensive data set of macro economical variables. The latent factors identified as global variables show small R^2s that are not statistically significant for the dollar carry trade strategy. However, these factors are statistically significant when explaining the returns of the high-minus-low carry trade strategy. The opposite is true for factors identified as domestic variables as they are only valuable in explaining the returns of the dollar carry trade strategy. Finally, the conclusions made by Filippou and Taylor (2017) and Lustig et al. (2014) coincide with the findings summarized in Table 1, emphasizing the nature of the high-minus-low carry trade as a strategy that diversifies the country specific risks away.

2.2 Momentum

In a momentum trading strategy, one takes a long position in a basket of assets with strong recent excess returns, called winners, and a short position in a basket of assets with weak recent excess returns, called losers. Consequently, a momentum trade is a cross-sectional bet that the winner assets will continue to outperform the loser assets in the future. Alternatively, a strategy that exploits the time-series momentum of a single asset coincides with a trend following strategy. In the time-series momentum, one takes a long (short) position in an asset when it has had a strong (weak) recent performance. Jegadeesh and Titman (1993) identify the momentum strategy in the U.S. equity markets and show that it yields an average annualized excess return of 12.01%. Jegadeesh and Titman (1993) and Jegadeesh and Titman (2001) state that these returns are not explained by exposures to common risk factors but reflect investors' slow reactions to news to some extent.

In contrast to the high equity momentum returns presented by Jegadeesh and Titman (1993) and Jegadeesh and Titman (2001), Yang and Zhang (2019) and Bhattacharya et al. (2012) show that the post-2000 equity momentum returns have vanished as, for example, the same momentum strategy used by Jegadeesh and Titman (1993) realizes an average annual return of 0.72% during the years 2000 to 2015. However, Yang and Zhang (2019) argue that the momentum returns are retrieved after excluding the extreme winners from the portfolio. They show that this prevents the momentum portfolio from momentum crashes introduced by Daniel and Moskowitz (2016) and lifts the annualized average returns to 8.76%.

Momentum strategies have proven to be profitable in other asset classes as well. In the bond markets, Jostova et al. (2013) demonstrate that the momentum

strategy is profitable for non-investment grade bonds. However, this strategy does not perform well in the investment-grade or the government bond markets (Asness et al. 2013; Gebhardt et al. 2005). This link of credit risk and momentum profits is in line with Avramov et al. (2007) as they conclude that the profitability of equity momentum is significantly stronger amongst stocks with low credit rating. Furthermore, Asness et al. (2013) find strong support for momentum returns in commodity and G10 currency markets as the cross-sectional momentum strategy yields on average an annual 12.4% and 3.5% return, respectively, during the years from 1972 to 2011.

Despite the extensive academic research around the momentum phenomenon, it still predominantly remains an anomaly. However, the factors explaining the currency momentum returns are summarized in Table 2. The coefficient column

Table 2: Summary of the predictive factors of currency momentum strategies

Factor	Coefficient sign	Time frame	Evidence provided by
Past returns	+	short and mid	Daniel and Moskowitz (2016), Asness et al. (2013), Moskowitz et al. (2012)
Past returns	-	long	Menkhoff et al. (2017), Daniel and Moskowitz (2016), Asness et al. (2013), Moskowitz et al. (2012)
Equity market volatility	-	short and mid	Daniel and Moskowitz (2016), Orlov (2016)
Illiquidity	-	short	Orlov (2016), Asness et al. (2013)
Bear market	-	mid and long	Daniel and Moskowitz (2016)
Credit risk	-	short and mid	Asness et al. (2013), Menkhoff et al. (2012b)

presents the sign of a factor coefficient linking the carry trade returns to the relevant factor. Furthermore, the time frames short, mid and long denote immediate to one-month, one-month to 12-month and longer than 12-month lags, respectively.

Asness et al. (2013) provide evidence that momentum strategies are highly related to each other across different markets and asset classes and, thus, likely driven by the same factors. They report that an average equity momentum strategy has a correlation of 0.65 with an average equity momentum strategy exploited in other equity markets. Furthermore, Asness et al. (2013) show that the correlation between an average equity momentum strategy and an average non-equity momentum strategy is 0.37.

Filippou and Taylor (2017) show that the currency momentum strategy has offered an annualized mean of 5.17% and 1.57% for a sample of 48 currencies and a sample of G11 currencies, respectively, during the time period from 1985 to 2012. Furthermore, they report Sharpe ratios of 0.54 and 0.18 for the mo-

mentum strategies constructed with these two currency samples. Filippou and Taylor (2017) extract latent factors from an extensive data set of macro economical variables and use them to predict currency momentum returns. They argue that macro economical data performs poorly in predictive regressions of the currency momentum strategy. However, the latent factors that capture asset price indices, money and credit variables provide some value as the adjusted R^2's of their regressions are between 2% and 4%.

Daniel and Moskowitz (2016) argue that momentum returns show negative skewness and relatively long periods of negative returns. These findings coincide with the characteristics of the currency carry trades (Brunnermeier et al. 2009). However, Menkhoff et al. (2012b) conclude that currency momentum and carry trade strategies are unrelated to each other. Their findings are in line with Filippou and Taylor (2017) as they show that currency momentum returns are not explained by business cycle risk, liquidity risk or volatility risk that have proven to explain the profitability of currency carry and dollar-carry trade strategies (Filippou and Taylor 2017; Lustig et al. 2014; Menkhoff et al. 2012a; Brunnermeier et al. 2009). Furthermore, Menkhoff et al. (2012b) show that currency momentum returns are not exposed to a four-factor model proposed by Carhart (1997).

Menkhoff et al. (2012b) argue that the profitability of currency momentum strategy is linked to country risks, since momentum profits tend to be higher in countries with low credit ratings. The link between high momentum returns and poor credit ratings is supported by the profitability of momentum strategies in the equity and the bond markets (Jostova et al. 2013; Avramov et al. 2007; Asness et al. 2013; Gebhardt et al. 2005). However, Menkhoff et al. (2012b) show that currency momentum returns are exposed to minor currencies. This implies higher idiosyncratic and country specific risks as well as relatively high transaction costs that cut off approximately one half of the momentum returns. Furthermore, Menkhoff et al. (2012b) argue that the excess currency momentum returns are difficult to arbitrage away, since the momentum returns are unstable in the short-term which is a characteristic supported by Daniel and Moskowitz (2016).

Daniel and Moskowitz (2016) argue that the crashes of the cross-sectional momentum strategy take place during the recovery from a bear market. They show that during extensive financial market turmoil, momentum strategy doesn't perform poorly as the market declines but crashes when the market rebounds. This is because the betas of the loser assets increase significantly in periods of market recovery, whereas this is not the case for the winner assets. For example, after the market bottom of March 2009, the loser decile gained 163%, whilst the winner decile gained only 8% over a three-month period (Daniel and Moskowitz 2016). Consequently, momentum returns show dynamics that coincide with a short call

position which could be explained by the Merton's model for the equity market (Daniel and Moskowitz 2016; Merton 1974). However, for other asset classes this reasoning is not robust and, thus, the phenomenon could be explained by behavioural theories (Daniel and Moskowitz 2016).

Orlov (2016) and Asness et al. (2013) argue that various measures of illiquidity predict currency momentum returns with a negative slope. Orlov (2016) investigate liquidity effects focusing on Amihud's illiquidity measure computed from both equity and currency market observations (Amihud 2002). Orlov (2016) finds that equity illiquidity exhibits stronger predictive power than the according measure constructed from currency markets. On the other hand, Asness et al. (2013) argue that funding liquidity has a stronger explanatory power on momentum returns. Furthermore, Orlov (2016) provides some support for the relationship between lagged market volatility and currency momentum returns as presented by Daniel and Moskowitz (2016).

Moskowitz et al. (2012) investigate time-series momentum in the equity, currency, commodity and bond futures markets. They show that this strategy yields positive returns across all asset classes and are strongly correlated not only inside an asset class but also between different asset classes. Moreover, Moskowitz et al. (2012) find that the time-series momentum strategy provides the highest returns during extreme market events with both signs as, for example, market crashes are usually preceded by a run of more moderate losses. However, they argue that the time-series momentum returns are not explained by exposures to standard risk factors but rather show investors' initial under-reaction and delayed over-reaction. This explanation is shared by Menkhoff et al. (2012b) who provide evidence from this phenomenon in cross-sectional currency momentum returns. Furthermore, Moskowitz et al. (2012) argue that cross-sectional momentum is likely driven by the persistence of asset-wise trends. For example, cross-sectional and time-series momentum strategies deployed in currency market show a correlation of 0.75 during the years from 1965 to 2009.

Momentum returns have shown to reverse after holding periods longer than a year (Menkhoff et al. 2017; Asness et al. 2013; Moskowitz et al. 2012). Moskowitz et al. (2012) argue that time-series momentum signals last a year and after that the strategy incurs losses. This gives support to the theory that momentum returns are a result of investors' over-reaction. Furthermore, Asness et al. (2013) construct a profitable cross-sectional currency trading strategy that takes a long (short) position in a basket of currencies with the lowest (highest) 5-year real exchange rate return, thus, capturing long term price reversals. They show that this strategy has a strong negative correlation with the returns of the momentum strategy. In addition, Menkhoff et al. (2017) provide support for this contrary strategy

as they build a related currency price reversal strategy, however, utilizing more sophisticated metrics of currency value than Asness et al. (2013).

3 LSTM NEURAL NETWORKS

3.1 Recurrent neural networks

The following introduction of feedforward neural networks or multilayer perceptrons (MLP) follows the description by Goodfellow et al. (2016, 164–168). MLPs are used to approximate a function F^*. For example, the function F^* could be a binary classifier

$$y = F^*(\cdot),$$

where $y \in [0, 1]$. The idea of a MLP is to learn the parameters θ of a mapping

$$y = F(\cdot, \theta).$$

In fact, MLP is nothing but a chain of linear mappings and activation functions applied to the input data. For example, a two-layer MLP can be expressed by the following two equations

$$\boldsymbol{h} = \psi(\boldsymbol{W}_{xh}\boldsymbol{x} + \boldsymbol{b}_h)$$

$$\boldsymbol{y} = \boldsymbol{W}_{hy}\boldsymbol{h} + \boldsymbol{b}_y,$$

where $\boldsymbol{W}_{xh}$ is a weight matrix used in the input–hidden layer mapping, $\boldsymbol{W}_{hy}$ is a weight matrix used in the hidden–output layer mapping, $\boldsymbol{x}$ is a vector of the input data, $\boldsymbol{b}$ is a bias vector and ψ is an element-wise applied non-linear activation function.

With MLPs, it is possible to approximate any mapping given enough capacity. However, if recurrent connections are added, it is possible to approximate any algorithm. Recurrent neural networks (RNN) are used to process sequential data, for example time-series data. RNNs use the same weights for multiple time steps and they have recurrent connections. As a consequence from the former feature, called parameter sharing, RNNs are able to process data with various lengths. (Goodfellow et al. 2016, 367–368).

The following introduction of the RNN follows the description made by Graves et al. (2013). A standard RNN computes the hidden vector sequence $\boldsymbol{h} = (h_1, h_2, \ldots, h_T)$ and the output sequence $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)$. The computation is done iteratively through time $t \in [1, 2, \ldots, T]$ by the equations

$$\boldsymbol{h}_t = \psi(\boldsymbol{W}_{xh}\boldsymbol{x}_t + \boldsymbol{W}_{hh}\boldsymbol{h}_{t-1} + \boldsymbol{b}_h) \tag{2}$$

and

$$y_t = \boldsymbol{W}_{hy}\boldsymbol{h}_t + \boldsymbol{b}_y, \tag{3}$$

where $\boldsymbol{W}_{xh}$ is a weight matrix applied to input layer, $\boldsymbol{W}_{hh}$ and $\boldsymbol{W}_{hy}$ are weight

matrices applied to the output from the hidden layer at time $t-1$ and t respectively, $\boldsymbol{x}$ is input data, $\boldsymbol{b}$ denotes a bias vector and ψ is a function applied to the hidden layer. In a simple RNN, the hidden layer function ψ could be for example an element-wise application of the sigmoid function. The RNN is illustrated in Figure 1.

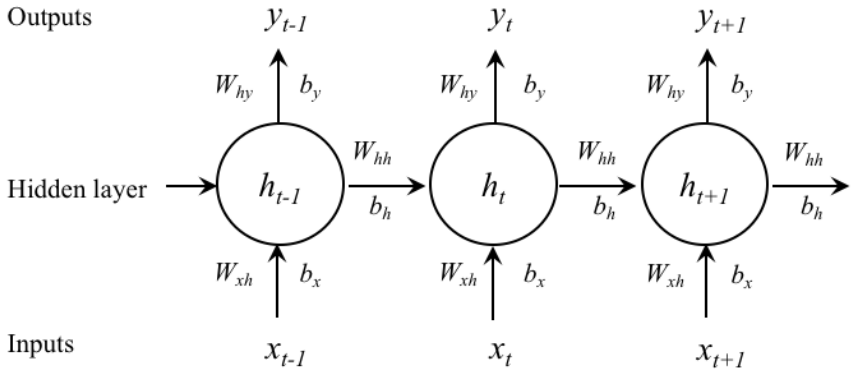

Figure 1: Unfolded recurrent neural network (RNN)

In order to optimize the model parameters in a RNN, gradient based back-propagation methods are usually applied through time. This method is called back-propagation through time (BPTT). The goal of the optimization is to minimize an objective function, $\mathcal{L}(y_t, \hat{y}_t)$ that determines the current error of the model. In the objective function y_t and $\hat{y}_t$ denote the target and the current output of the model. The choice of the objective function depends on the type of the predictive modelling problem. For example, a mean squared error could be used as the objective function in a predictive regression problem. The idea of the BPTT is to compute the gradients of the model parameters with respect to the error starting from the most recent time step and then moving step by step back in time. After the gradients are computed, the model parameters are updated with gradient descent. This procedure is repeated until the error is sufficiently small or the parameters are converged.

BPTT to the parameters in the Equations 2 and 3 is given as follows. First, gradients of the output layer are

$$\delta_t^y = \frac{\partial \mathcal{L}(y_t, \hat{y}_t)}{\partial y_t} \tag{4}$$

which are then propagated to the hidden layer according to

$$\delta_t^h = \psi'(\delta_t^y \boldsymbol{W}'_{hy} + \delta_{t+1}^h \boldsymbol{W}'_{hh}), \tag{5}$$

where ψ' denotes the element-wise applied derivative of the activation function and for the last time step T the term $\delta_{t+1}^h \boldsymbol{W}_{hh}$ is zero. Next, the gradient descent based update rules for the model parameters are

$$\boldsymbol{W}_{hy}^{i+1} = \boldsymbol{W}_{hy}^i + \eta \sum_{t=1}^{T} \delta_t^y \boldsymbol{h}_t',$$

$$\boldsymbol{b}_y^{i+1} = \boldsymbol{b}_y^i + \eta \sum_{t=1}^{T} \delta_t^y,$$

$$\boldsymbol{W}_{hh}^{i+1} = \boldsymbol{W}_{hy}^i + \eta \sum_{t=1}^{T} \delta_t^h \boldsymbol{h}_{t-1}',$$

$$\boldsymbol{W}_{xh}^{i+1} = \boldsymbol{W}_{xh}^i + \eta \sum_{t=1}^{T} \delta_t^h \boldsymbol{x}_{t-1}',$$

and

$$\boldsymbol{b}_h^{i+1} = \boldsymbol{b}_h^i + \eta \sum_{t=1}^{T} \delta_t^h,$$

where η is a predetermined learning rate and i denotes the number of the iteration.

3.2 Long short-term memory cells

RNN has a memory which means that it is able to use past states in predicting future targets. Bengio et al. (1994) show that BPTT becomes increasingly inefficient in training RNNs as the time spans of the dependencies increase. This is easily obtained through the Equations 4 and 5 as the effect of the predictive power of a signal to the gradient decreases exponentially as the number of time steps back increases. This phenomenon is called vanishing gradients.

Hochreiter and Schmidhuber (1997) developed a solution to the aforementioned problem by introducing long short-term memory cells (LSTM). LSTM cells can be applied to RNNs by replacing the activation function ψ in the Equation 2 by a composite function

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{ci}\boldsymbol{c}_{t-1} + \boldsymbol{b}_i)$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{cf}\boldsymbol{c}_{t-1} + \boldsymbol{b}_f)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t\boldsymbol{c}_{t-1} + \boldsymbol{i}_t \tanh(\boldsymbol{W}_{xc}\boldsymbol{x}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c) \tag{6}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{co}\boldsymbol{c}_{t-1} + \boldsymbol{b}_o) \tag{7}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \tanh(\boldsymbol{c}_t),$$

where

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \tag{8}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}},$$

$\boldsymbol{i}$, $\boldsymbol{f}$ and $\boldsymbol{o}$ are input, forget and output gate vectors, respectively. $\boldsymbol{c}$ is a cell vector and each weight matrix that is applied to the cell to gate mapping is diagonal. The LSTM cell is illustrated in the Figure 2.

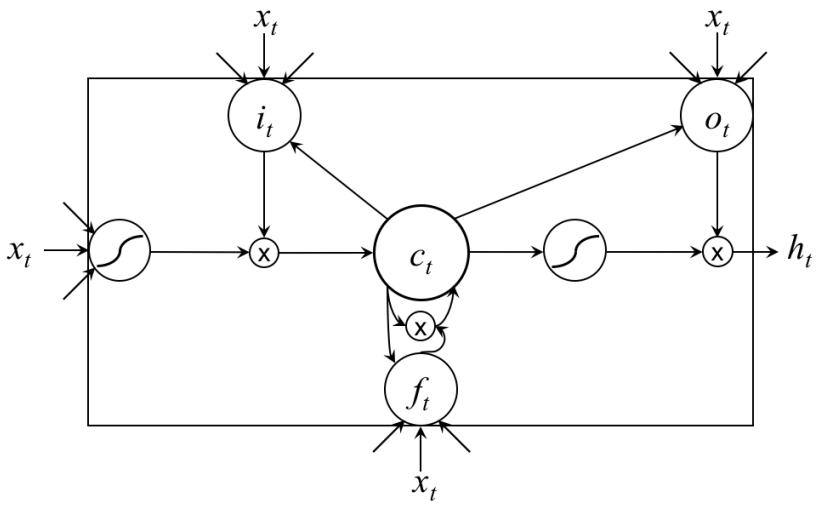

Figure 2: Long short-term memory (LSTM) cell (Graves et al. 2013)

The LSTM RNN is able to model dependencies with an arbitrary time span between the predictive signal and the target. The memory cell in the Equation 6 stores the temporal state values of the network and the gates, of which elements get values between 0 and 1, control the information flow. For example, if the values of the elements in the input gate vector are zero, new information is not added to the temporal state. Similarly, the output vector controls the output flow of the information that leaves the cell. The forget gate has been introduced by Gers et al. (2000) and it allows the LSTM cell to learn to reset its memory and, as a consequence, release internal resources. Finally, Gers et al. (2002) add peephole connections from the cell to the gates of an LSTM cell. These connections allow the model to learn precise timing as the gates are able to inspect the internal state of the cell.

3.3 Market prediction

During the past two decades, LSTM RNNs, LSTM henceforth, have been used to win multiple competitions and to break records of popular benchmark tasks that are based on large and complex data sets. This technique has been capable of solving previously unsolved problems that involve, for example, modelling the temporal distance and order of events, storing real numbers through long time intervals and recognizing patterns extended in time from noisy input data. In general, many of the machine learning methods, like neural networks, are proven to be valuable predictors of noisy and complex natural phenomena. (Schmidhuber 2015)

Despite the success in other applications, machine learning techniques have been lacking broader breakthroughs in financial market prediction, especially from the academic literature point of view. Jacobs (2015) summarizes 100 different market anomalies found in previous research. He divides these anomalies in 20 meta-strategy groups and demonstrates that each group has provided a positive alpha during investment periods with varying starting dates between 1926 and 1987 and ending in 2011. However, Krauss et al. (2017) point out that none of the strategies tested by Jacobs (2015) are utilizing advanced statistical learning techniques.

Gu et al. (2018) emphasize the fact that machine learning is suitable for prediction tasks, but these techniques cannot be used to explain the underlying economic mechanisms. The latter is possible with standard linear techniques. However, exploiting the vast amount of potential predictors presented by financial market research makes standard linear models inadequate in prediction tasks due to, for example, collinearity and decreasing degrees of freedom (Gu et al. 2018). Furthermore, these techniques are not able to incorporate the complex and likely non-linear relationships in the data (Heaton et al. 2018). Consequently, Heaton et al. (2018) list three main characteristics of machine learning which make it more suitable to financial market prediction compared to traditional methods used in this field. Firstly, all possibly relevant input data can be used in the prediction task. Secondly, machine learning techniques are capable of capturing non-linearities and complex interactions. Finally, over-fitting is easier to avoid in the learning process which increases the generalization and out-of-sample performance.

Gu et al. (2018) investigate different statistical learning models in time-series prediction of stock returns. They compare the performance of linear regression, generalized linear models with penalization, principal component regression, partial least squares, regression trees that include boosted trees and random forests and feed-forward neural networks relying on vast amount of data. Gu et al. (2018)

predict the excess returns of almost 30,000 individual stocks between 1957 and 2016. The data set used for prediction consists of 94 characteristics of each stock, eight aggregate time-series variables and 74 dummy variables related to industry sectors. Expectedly, in the class of linear models, penalization and dimension reduction techniques improve the out-of-sample performance due to the high dimensional data set of more than 900 predictors. Moreover, Gu et al. (2018) show that non-linear models further improve the performance over linear models.

Gu et al. (2018) develop a long-short equity strategy with the stocks in the S&P 500 index based on the predictions of their 3-layer neural network architecture. They show that this method generates a Sharpe ratio of 2.35. On the other hand, similar strategy that relies on the benchmark linear method annualizes a strikingly lower Sharpe ratio of 0.89.

By further analysis of the used methods and data sets, Gu et al. (2018) argue that the most important predictor classes are price trends, liquidity proxies and volatility measures. These findings hold regardless of the model used. 12-month momentum and both short-term and long-term reversals are amongst the most informative price trends.

Krauss et al. (2017) compare three different statistical learning methods and their ensembles in stock market prediction. These methods are deep neural networks, gradient boosted trees and random forests. Krauss et al. (2017) use only past return observations spanning three years back in time in their time-series prediction task of S&P 500 constituent stock returns. They show that the ensemble methods that combine the three individual techniques outperform each of the individual models. Furthermore, they construct a long-short strategy based on their best performing model which generates a Sharpe ratio of 1.81 after transaction costs. The Sharpe ratio is five times larger than the Sharpe ratio of S&P 500 index during the sample period of 1992–2015. In line with the most important price trends found by Gu et al. (2018), Krauss et al. (2017) show that their strategy utilizes predominantly the returns from the past 12-month and one-week periods. Furthermore, Krauss et al. (2017) and Gu et al. (2018) show consistent results considering the depth of the best performing models, since they both conclude that more shallow architectures of non-linear models outperform deep learners in financial market prediction.

Fischer and Krauss (2018) apply LSTMs to the same sample and setting investigated by Krauss et al. (2017). They conclude that LSTMs consistently outperform the prior benchmark techniques. More specifically, also after transaction costs, the LSTM alone provides a higher Sharpe ratio of 2.34 than the ensemble of the other three techniques. Finally, Fischer and Krauss (2018) show that the developed strategy utilizes mostly short-term reversals and picks high beta stocks.

Krauss et al. (2017) and Fischer and Krauss (2018) identify that their strategies have low exposures to well known risk factors. However, their strategies tend to pick high volatility stocks. This feature is most likely due to their heuristic portfolio construction technique that does not take volatilities into account. Consequently, an optimization based approach might enhance the risk-return profile of these strategies.

The usage of LSTMs comes with a higher computational cost, since in this model the number of parameters to be trained is large. Fischer and Krauss (2018) argue that the profitability of their strategy decreases significantly in the post 2010 period, fluctuating around zero after transaction costs. They explain that this is possibly a consequence of higher computational power and broader utilization of advanced machine learning techniques which has led the high excess returns to be arbitraged away. Increase in the computational power is likely to explain the success stories of deep learning techniques in general as, for example, today's computer has more than a million times the computational power of a desktop computer from thw early 1990s (Schmidhuber 2015, 94).

In general, models trained with historical data generalize well to unseen data if the dynamics of the system have not changed dramatically. This is usually the case with natural phenomena but not with financial markets. Consequently, simpler models that are built based on rigorous financial theories might provide greater confidence to investment decision making compared to the black box machine learning techniques. However, the attractive capabilities and proven success in terms of exceptional Sharpe ratios highlight the need to fill the research gap in advanced machine learning applications in financial markets prediction tasks. In fact, the popularity of these techniques can be increased securely through better understanding the dynamics of machine learning models and their implications.

4 METHODOLOGY

4.1 Data and definitions

The data span of the sample period is from the 2nd of January 1999 to the 31st of December 2018 and is chosen as the Euro has been traded only since 1999. The data are obtained from the Bloomberg database. The construction of the data, prediction task, portfolio formation and empirical analysis are conducted with a weekly frequency for two reasons. First, the weekly frequency helps to reduce the noise incorporated in higher frequencies. Second, weekly observations should decrease possible problems concerning different time-zones. However, daily closing observations are used to construct the factor proxies that require calculations of variances.

The currencies being traded are the G10 currencies: Euro (EUR), Pound sterling (GBP), Japanese yen (JPY), Australian dollar (AUD), New Zealand dollar (NZD), Canadian dollar (CAD), Swiss franc (CHF), Norwegian krone (NOK) and Swedish krona (SEK), which are denoted against the United States dollar (USD). These represent the most liquid currencies in the market and, consequently, are extremely suitable to trading strategies with higher frequencies. Price quotes from the spot and futures markets are used to calculate the currency returns.

The excess currency returns over one period, which are used in the prediction and portfolio construction phases, are calculated as follows

$$RX_{k,t+1} = i_t^k - i_t - \frac{s_{k,t+1} - s_{k,t}}{s_{k,t}} \approx \frac{f_{k,t,t+1} - s_{k,t+1}}{s_{k,t}},$$

where i^k and i denote the foreign and home country riskless interest rates whereas f_k and s_k denote the mid quote futures and spot prices of an unit of a foreign currency k per a unit of the home currency (Filippou and Taylor 2017). Moreover, the practical feasibility of the currency portfolios are evaluated based on the portfolio turnover and estimated transaction costs. The annualized portfolio turnover is computed as follows

$$Turnover = \frac{52}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} |w_{k,t+1} - (1 + RX_{k,t+1}) w_{k,t}|,$$

where w_k denotes the dollar amount of a currency k in the portfolio and $|(\cdot)|$ denote absolute value. Furthermore, K and T denote the total number of currency pairs and the total number of weeks, respectively. The estimated annual transaction

costs of a currency portfolio are computed as follows

$$Transaction\,costs = \sum_{k=1}^{K} Turnover_k \cdot BAS_k,$$

where $Turnover_k$ is an annual turnover of a currency k and the average bid–ask spread for a currency k

$$BAS_k = \frac{1}{T}\sum_{t=1}^{T} \frac{s^b_{k,t} - s^a_{k,t}}{(s^b_{k,t} + s^a_{k,t})/2},$$

where s^b and s^a denote bid and ask spot price quotes, respectively. Finally, transaction costs are estimated solely based on currency spot bid–ask spreads, since reliable data on the bid and ask quotes of the futures or forward contracts were not available.

The factors used in the predictions are currency returns, forward premium, equity and foreign exchange market volatilities, funding liquidity, variance risk premium, currency variance risk premium, commodity prices and credit risk. These factors are chosen based on the previous research findings regarding the predictability of currency markets and discussed more in depth in Section 2 (Londono and Zhou 2017; Menkhoff et al. 2017; Orlov 2016; Asness et al. 2013; Mancini et al. 2013; Bakshi and Panayotov 2013; Menkhoff et al. 2012a,b; Brunnermeier et al. 2009). Furthermore, the chosen factors cover the three most informative classes of predictors found in an analysis of over 900 different variables in an equity market return prediction problem by Gu et al. (2018) as they argue that various price trends, liquidity and volatility measures contribute the most to the prediction performance of their model.

For price trends, aggregate variables, like a 12-month return, are not used since LSTMs should be able to recognise these patterns from the historical data if they have predictive power. This is also empirically proved by Fischer and Krauss (2018). Consequently, RX is used as the first factor which proxies price trends. The second factor, forward premium, is calculated from the futures and spot prices as follows

$$FP_t = \frac{f_{k,t,t+1} - s_{k,t}}{s_{k,t}}.$$

Considering the risk proxies, the VIX index, that measures the option-implied volatility of the S&P 500 Index and is calculated by Cboe Options Exchange, is used as a proxy of equity market uncertainty as by Brunnermeier et al. (2009), whereas the approach by Menkhoff et al. (2012a, 692) is followed for the currency market volatility factor. Consequently, the proxy for a one-month currency market

volatility is

$$\sigma_t^{FX} = \frac{1}{K} \sum_{i=t-21}^{t} \left[\sum_{k=1}^{K} \mid \Delta s_i^k \mid \right],$$

where K is the number of currencies and Δs denotes the one-day change of a currency log spot price.

The construction of the variance risk premium and the currency variance risk premium follows the approach in Londono and Zhou (2017). Consequently, the proxy for the variance risk premium

$$VP_t = E_t^Q[\sigma_{r,t+1}^2] - E_t^P[\sigma_{r,t+1}^2], \tag{9}$$

where E^Q and E^P are the risk-neutral and physical expectations of the stock market variance σ_r^2. Londono and Zhou (2017) use a model-free option implied variance, similar to the method used to calculate the VIX Index, as a proxy for $E^Q[\sigma_r^2]$ but in this study, the VIX Index is used instead. Furthermore, one-month realized volatility of the S&P 500 index is used as a proxy for $E_t^P[\sigma_{r,t+1}^2]$ as by Londono and Zhou (2017). Consequently, the Equation 9 can be written as

$$VP_t = \frac{1}{12} VIX_t^2 - \sum_{i=t-21}^{t} r_i^2,$$

where r is a daily return of the S&P 500 Index. Furthermore, the currency variance risk premium

$$XVP_t = \frac{1}{K} \sum_{k=1}^{K} \left[E_t^Q[\sigma_{k,t+1}^2] - E_t^P[\sigma_{k,t+1}^2] \right], \tag{10}$$

where E^Q and E^P are the risk-neutral and physical expectations of the variance of a single currency k. Black-Scholes at-the-money option-implied volatility, which is calculated by Bloomberg, is used as a proxy for $E^Q[\sigma_k^2]$ and one-month realized volatility of a currency k is used as a proxy for $E_t^P[\sigma_{k,t+1}^2]$. Consequently, the Equation 10 can be written as

$$XVP_t = \frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{12} \sigma_{t,impl}^2 - \sum_{i=t-21}^{t} \Delta s_{k,i}^2 \right],$$

where σ_{impl} is the option-implied volatility of a currency pair.

TED spread is used as a proxy of the funding liquidity as in (Bakshi and Panayotov 2013; Brunnermeier et al. 2009). It is calculated as

$$TED_t = LIBOR_t^{USD} - TB_t,$$

where $LIBOR^{USD}$ is a three-month LIBOR (London Interbank Offer Rate) of interbank loans denominated in the USD in London and TB denotes the three-month U.S. treasury bill interest rate. Furthermore, the credit spread (CS) is

calculated as a difference between the yield of a Moody's Baa rated aggregate corporate bond index and the yield of an aggregate U.S. treasury note with 10-year maturities. Finally, the returns of a composite commodity index are used as a proxy of the commodity price factor (COM) as in Bakshi and Panayotov (2013). The commodity index used is the Rogers International Commodity Index that tracks the performance of a basket of 37 commodity futures contracts denominated in the USD.

4.2 Prediction routine

The goal of the prediction task is to learn K mappings

$$F^k : \boldsymbol{X}^k \rightarrow \boldsymbol{y}^k,$$

where $\boldsymbol{X}^k$ denotes the lagged input patterns for a currency k and $\boldsymbol{y}^k = [y_2^k, y_3^k, \ldots, y_{T+1}^k]$ denotes the targets

$$y_t^k = \begin{cases} 1 & if\, RX_t^k > m_t \\ 0 & if\, RX_t^k \leq m_t \end{cases},$$

where m_t is a cross-sectional median of the returns RX_t^k, where $k \in K$. Therefore, the objective is to learn to predict conditional probabilities that a currency k outperforms the cross-sectional median of all currency excess returns during the next time step. Two types of recurrent neural networks, LSTMs and simple RNNs are compared with two sets of input patterns. The input patterns at time t are a vector

$$\begin{bmatrix} RX_1^k & RX_2^k & \cdots & RX_{t-1}^k \end{bmatrix}'$$

of past excess currency returns or a matrix

$$\begin{bmatrix} RX_1^k & FP_1^k & VIX_1 & \sigma_1^{FX} & VP_1 & XVP_1 & TED_1 & CS_1 & COM_1 \\ RX_2^k & FP_2^k & VIX_2 & \sigma_2^{FX} & VP_2 & XVP_2 & TED_2 & CS_2 & COM_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ RX_{t-1}^k & FP_{t-1}^k & VIX_{t-1} & \sigma_{t-1}^{FX} & VP_{t-1} & XVP_{t-1} & TED_{t-1} & CS_{t-1} & COM_{t-1} \end{bmatrix}$$

of past excess currency returns, forward premiums and the seven factors that are common to all currencies. Since the model parameters are trained for each of the currencies separately, the total number of models investigated is $4 \times 9 = 36$. The performances of the LSTM models are compared to the performances of the RNN models in order to find out if the complexity added by the former is necessary. As discussed in Section 2, some predictability of currency excess returns have

been found with factor sets lagged up to even 12 months. If the LSTM models outperform RNN models, there exists predictive patterns with more than two months of temporal distance as a LSTM is capable to model longer temporal relationships than a simple RNN due to e.g. the vanishing gradient problem. Two sets of input patterns are investigated in order to analyse if the additional factors are valuable predictors of the currency return dynamics. Finally, the currency performances are predicted separately for two reasons. First, co-modelling the currency performances should be irrelevant since no co-movements have been found among the G10 currencies that would not otherwise be possible to capture from the predictive factor sets considered in this study. Second, smaller number of input patterns per model should significantly reduce the required computing power as the models are less complex.

A LSTM or RNN predictor is denoted as

$$\hat{\boldsymbol{y}^k} = F^k(\boldsymbol{X}^k; \theta^k),$$

where θ^k is a set of model parameters $\{\boldsymbol{W}, \boldsymbol{b}\}$. The true parameters θ^k are unknown and, thus, need to be estimated, i.e the model needs to be trained. The training is done by solving

$$\text{argmin}_\theta \frac{1}{T} \sum_{t=1}^{T} \mathcal{L}(y_t^k, \hat{y_t}^k), \tag{11}$$

where $\mathcal{L}(y_t, \hat{y}_t)$ is a loss function and $\hat{y_t}^k = F^k(\boldsymbol{X}_t^k; \theta)$.

Fischer and Krauss (2018) use a softmax activation function with two output neurons and a cross entropy loss function. In this research, however, a single output neuron with a sigmoid activation function in Equation 8 is used with a binary cross-entropy loss function for simplicity. The binary cross-entropy for a single target–output pair is

$$\mathcal{L}(y_t, \hat{y}_t) = -\big(y_t \cdot log(\hat{y}_t) + (1 - y_t) \cdot log(1 - \hat{y}_t)\big).$$

The training algorithm used is RMSprop as used by Fischer and Krauss (2018). The update rule of the parameters is similar to the gradient descent but instead of using fixed learning rate and gradients calculated from the full sample of training input patterns, RMSprop utilizes adaptive learning rates and mini-batches. The learning rates are divided by the square roots of average squared magnitudes of the gradients from the previous mini-batches and computed to each parameter separately. (Hinton 2012). In addition, similar to Fischer and Krauss (2018), dropout regularization is used as also adviced by Goodfellow et al. (2016, 420). The basic idea of the method, first introduced by Srivastava et al. (2014), is to drop out some fraction of randomly chosen units of the network during the training steps of the model. This decreases the probability of overfitting and, consequently,

should result in better out-of-sample performance. Finally, the input patterns of the training data are scaled so that the standard deviation is one in-sample before training. This speeds up the learning process and adds regularization without changing the signs of the factor observations. Furthermore, the standard deviations of the input patterns in the training data are used to scale the input patterns in the validation and prediction data as well.

The data is divided into training, validation and prediction sets. The training and validation data consists of the lagged input patterns and targets. The iterative optimization routine is started by first updating the model parameters in the training data based on the calculated training error that is the loss in the training data set. Next, the model and the updated parameter values are used to make predictions on the validation set, and these predictions are then used to calculate the validation error that is the loss in the validation data set. The validation error is used to control the model overfitting and, consequently, to improve generalization, i.e. out-of-sample performance, in the following way. After each parameter update, the validation error is calculated and the optimization is stopped when the validation error starts to increase or when it has been on the same level during the last fifteen epochs, that is a full presentation of the data set used in the learning process, as recommended by Goodfellow et al. (2016, 420). Furthermore, the maximum number of the epochs is capped to one thousand. Finally, the model parameters with the lowest validation error are retrieved. This routine is illustrated in Figure 3. After retrieving the optimal model parameters, the lagged input patterns of the prediction data set are used to make predictions. The targets of the prediction data set are used only in the performance evaluation when the performances of the models are compared afterwards.

The data splits are handled in the following way. The 20-year sample is divided into three parts. Three years (156 weeks) of data are used as a training set, whereas the next two and a half years (130 weeks) are used as a validation set. The rest of the years are used for out-of-sample predictions. However, after the model is optimized, it is used to generate the predictions for a year (52 weeks) following the end of the validation set. Single feature sets are 60 weeks long so that the model could recognize patterns with longer than 13 months temporal distances. They are constructed from the lagged observations so that a feature set used in the prediction of next week's performance probabilities is constructed from the observations obtained during the preceding 60 weeks. Consequently, the data set used in the prediction phase is rolled ahead for a week after each prediction. After the model that has been optimized with the training and validation sets is used to produce 52 predictions, the training set is extended by a year and the validation set is rolled ahead for a year. Then the model is re-trained with the new data

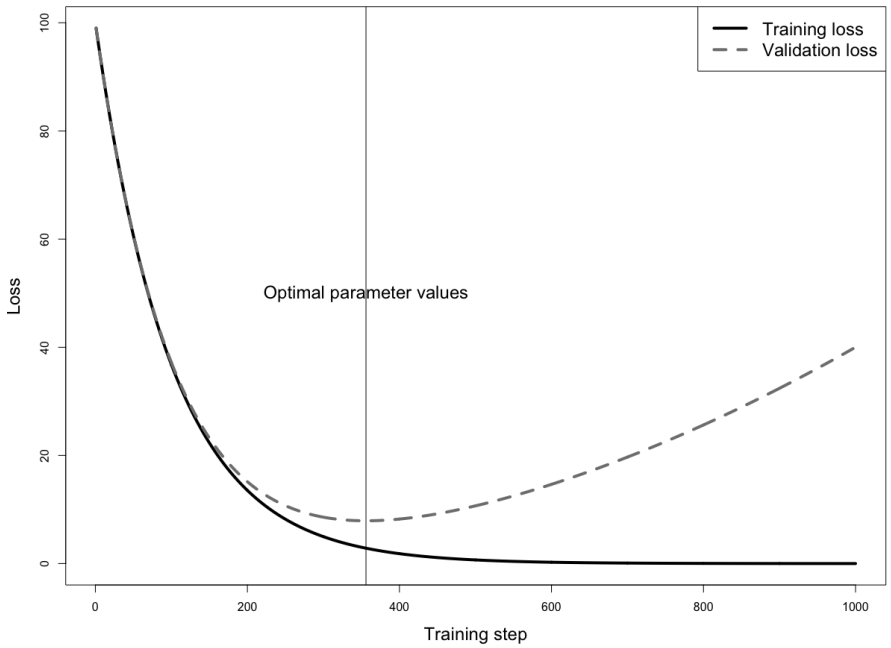

Figure 3: LSTM model optimal parameter value selection

splits and used to the out-of-sample predictions of the next 52 weeks. The model is re-trained in an annual frequency in order to save computational burden. As an illustration, to the model that utilizes nine factors the dimension of the first set of input patterns used in the parameter optimization is $97 \times 9 \times 60$ and the dimension of the output set produced is $97 \times 1 \times 1$. The number of outputs is smaller than the total number of time steps in the data set since one prediction requires data from the preceding 60 time steps. However, since the input patterns of the validation and training data sets overlap, the number of outputs produced in the validation phase is equal to the number of weeks in the split size. Consequently, the dimension of the input patterns of the validation set is $130 \times 9 \times 60$ and the dimension of the output set produced is $130 \times 1 \times 1$. Furthermore, the respective dimensions to the prediction data sets are $52 \times 9 \times 60$ and $52 \times 1 \times 1$. Note that after splitting the data the predictions are made with a rolling sets of lagged input patterns. Consequently, there is no look-ahead bias.

The only difference between the LSTM and RNN model architectures is that instead of the LSTM cells, tanh activation function is used in the hidden layers of the simple RNN. The model architectures and hyper parameters are chosen based on experiments with the first training and validation data sets that cover the first five and a half years. Relatively shallow architectures perform better as models with a high number of hidden units tend to overfit. There is only minor changes in the validation error with hidden units ranging from 4 to 15. However, the chosen number of hidden units is 6 to the models utilizing all factors as input

patterns and 10 to the models utilizing only past excess currency returns as input patterns since with these the performances are the most consistent among the models. The number of weight and bias terms in a LSTM model is calculated as $4(u(i+1)+u^2)$, where i is the number of input features (factors) and u is the number of units in the hidden layer. Consequently, the numbers of weight and bias terms are $4(6(9+1)+6^2) = 384$ and $4(10(1+1)+10^2) = 480$ to the LSTM models with nine and one input features, respectively. Furthermore, the numbers of parameters are one-fourth of these to the simple RNNs. Finally, the choice to decrease the batch size, used in the parameter optimization, from one epoch to 52 feature sets decreases overfitting and improves validation set performance. Furthermore, tweaking the dropout ratio does not have significant effects and, consequently, the dropout ratio is set to 0.1 as in Fischer and Krauss (2018).

The empirical analysis is conducted with Python 3.6 by Python Software Foundation (2019). The LSTMs and simple RNNs are built with Keras by Chollet et al. (2015) which runs on top of Tensorflow by Abadi et al. (2015).

4.3 Performance evaluation

Classification accuracy is used to evaluate the prediction performances of the LSTM and RNN models. It measures the percentage of correct classifications and can be written as

$$\mathcal{A}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \frac{n - \sum_{i=1}^{n} |y_i - \hat{y}_i|}{n},$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of targets $\in (0, 1)$ and $\hat{\boldsymbol{y}}$ is a $n \times 1$ vector of class predictions so that

$$\hat{y}_i = \begin{cases} 1 & \text{if} \quad \mathcal{P}_i > 0.5 \\ 0 & \text{if} \quad \mathcal{P}_i \leq 0.5 \end{cases},$$

where $\mathcal{P}_i$ is the outperformance probability prediction of the model. Furthermore, in order to assess the accuracy of a model, a statistical estimate of the probability, that a model would have achieved the prediction accuracy by change, is computed as by Fischer and Krauss (2018). First, if the true model accuracy is 0.5, the number of correct predictions could be presented by a binomial distribution

$$\boldsymbol{X} \sim \text{Binom}(TK, 0.5),$$

where T is the number of time steps and K is the number of the assets. Consequently, for a large TK, the randomness of the model prediction performance

can be tested with a test statistic

$$BT = \frac{TK(\bar{a} - 0.5)}{\sqrt{0.5nk}} \sim \text{Norm}(0, 1),$$

where

$$\bar{a} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{A}(\boldsymbol{y}_k, \hat{\boldsymbol{y}}_k)$$

under a null hypothesis that the true model accuracy is 0.5.

The prediction performances of different models are compared as by Gu et al. (2018) and Fischer and Krauss (2018) with a test proposed by Diebold and Mariano (1995). In the Diebold–Mariano test, the null hypothesis states that the prediction accuracies of models 1 and 2 are the same, whereas the alternative hypothesis states that the prediction accuracy of the model 2 is less than the prediction accuracy of the model 1. The test statistic

$$DM_{12} = \sqrt{T-1}\frac{\bar{d}_{12}}{\sigma_d} \sim \text{t}_{T-1},$$

where

$$\bar{d}_{12} = \frac{1}{K \cdot T} \sum \boldsymbol{d}_{12},$$

where T is the number of time steps and K is the number of the assets and

$$\boldsymbol{d}_{12} = |\hat{\boldsymbol{e}}_1| - |\hat{\boldsymbol{e}}_2|,$$

where

$$\hat{\boldsymbol{e}}_i = \boldsymbol{y} - \hat{\boldsymbol{y}}_i.$$

Finally,

$$\sigma_d^2 = \frac{1}{K \cdot T} \sum (\boldsymbol{d}_{12} - \bar{d}_{12})^2.$$

The variable importance is exploited with linear models as by Gu et al. (2018) by reporting the R^2s of OLS regression models with different sets of explanatory variables. This exercise helps to explain the dynamics of the black box machine learning method being investigated.

4.4 Portfolio construction

At each time step t, a prediction model gives an output that represents the probability that the excess return on a given currency at time $t + 1$ is higher than the expected cross-median of all G10 currency excess returns at time $t + 1$. These predictions are done based on the information available at time t. The probabil-

ity predictions are used in portfolio construction in order to exploit whether the models provide alpha opportunities or not.

Two portfolio construction approaches are considered. Both techniques utilize the cross-sectional ranking of the models' output probabilities. Furthermore, in both cases, the resulting portfolio is a self-financed dollar-neutral allocation. Consequently, long positions are financed with equally large short positions.

First, a robust portfolio construction that relies on a simple heuristic approach is utilized. Starting from the cross-section of the outperformance probability predictions, these output probabilities are ranked in an descending order. Then the strategy takes long positions in the currencies with the highest rankings, thus, in currencies with rankings from six to nine. Similarly, short positions are taken in the currencies with the lowest rankings that are the ranks from one to four. Finally, the position in a currency with the ranking five is long if the predicted probability is higher than 0.5 and short otherwise.

After the signs of the currency positions are determined, the weights are scaled so that the short positions sum to minus one and the long positions sum to one. Furthermore, the long positions at time t are equally large and the short positions at time t are equally large.

The second portfolio construction technique is based on constrained mean–variance optimization of which inputs are processed by entropy pooling. This approach is motivated by the empirical evidence from Fischer and Krauss (2018), which states that the robust portfolio construction technique that utilizes the LSTM model output signals tends to overweight high volatility assets. However, Boudoukh et al. (2018) demonstrate that mean–variance optimization is sensitive to parameter uncertainty which results in extreme allocations when applied to the G10 currency markets. They argue that more advanced signal prosessing and parameter estimation techniques might lead to well-balanced portfolios. Consequently, restrictions are imposed to the portfolio volatility and holding sizes in order to avoid extreme allocations. Furthermore, the LSTM model predictions are processed by entropy pooling as this is an intuitive way to translate the outperformance probabilities into mean–variance optimization inputs so that the asset variance and correlation estimates are aligned with the scenario of the LSTM model estimates.

In order to optimize the portfolio allocation, second-order cone programming (SOCP) is utilized

$$\boldsymbol{z}^* = \text{argmin}_{\boldsymbol{z}}\, \boldsymbol{c}'\boldsymbol{z}$$

$$\text{subject to} \begin{cases} \boldsymbol{A}\boldsymbol{z} = \boldsymbol{a} \\ \boldsymbol{B}\boldsymbol{z} \geq \boldsymbol{b} \\ \|\boldsymbol{D}\boldsymbol{z} - \boldsymbol{q}\|_2 \leq \boldsymbol{k}'\boldsymbol{z} - g \end{cases},$$

where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ are conformable matrices, $\boldsymbol{z}$, $\boldsymbol{c}$, $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{q}$ are conformable vectors, $\boldsymbol{g}$ is a scalar and $\|(\cdot)\|_2$ denotes the L_2-norm (Meucci 2005, 313–314). The portfolio optimization problem reduces into a quadratically constrained linear programming (QCLP) problem, which is a subclass of the SOCP. Consequently, the problem formulation is presented according to the following form

$$\boldsymbol{u}_t^* = \text{argmax}_{\boldsymbol{u}}\, \boldsymbol{u}'\boldsymbol{p}_{t+1}$$

$$\text{subject to} \begin{cases} \boldsymbol{p}_t\boldsymbol{u} = 0 \\ \boldsymbol{B}\boldsymbol{u} \geq \boldsymbol{b} \\ \|\boldsymbol{C}\boldsymbol{u}\|_2 \leq \boldsymbol{\sigma_{max}} \end{cases},$$

where $\boldsymbol{u}$ is a vector with the amounts of the U.S. dollar units to be invested in the foreign currencies, $\boldsymbol{p}_{t+1}$ is a vector of the expected values of the foreign currency positions at time $t+1$ per a unit of the USD invested at time t. Values are used instead of returns as linear returns on wealth are not defined to a zero initial investment. Note that the maximization can be formulated back to a minimization by taking the negative sign of the objective function. Furthermore, the first linear equality constraint ensures that the short positions are as large as the long positions and, thus, the strategy is self-financed. $\boldsymbol{p}_t$ is a vector of the costs in the USD of the positions in the currencies at time t. The second linear inequality constraint sets the maximum and minimum asset-wise allocations. The matrix $\boldsymbol{B} = \big[\boldsymbol{I}_n, -\boldsymbol{I}_n\big]$, where $\boldsymbol{I}_n$ is an identity matrix and n is the number of the currencies. Furthermore, the vector $\boldsymbol{b} = \big[max_1, \ldots, max_n, min_1, \ldots, min_n\big]$ sets the asset-wise minimum and maximum exposure limits. Finally, the third constraint limits the volatility of the resulting portfolio to σ_{max}. The quadratic constraint

$$\boldsymbol{u}'\boldsymbol{\Sigma}\boldsymbol{u} \leq \sigma_{max}^2 \tag{12}$$

can be rewritten by decomposing the variance-covariance matrix $\boldsymbol{\Sigma}$ into two triangular Cholesky decomposition matrices $\boldsymbol{C}$ (Davidsson 2011, 176). Now Equation 12 becomes

$$\boldsymbol{u}'\boldsymbol{C}\boldsymbol{C}'\boldsymbol{u} \leq \sigma_{max}^2$$

$$(\boldsymbol{C}\boldsymbol{u})'\boldsymbol{C}\boldsymbol{u} \leq \sigma_{max}^2$$

$$\sqrt{(\boldsymbol{C}\boldsymbol{u})'\boldsymbol{C}\boldsymbol{u}} \leq \sqrt{\sigma_{max}^2},$$

which can be expressed as a second-order cone constraint

$$\|\boldsymbol{C}\boldsymbol{u}\|_2 \leq \sigma_{max}.$$

In order to perform the optimization routine, $\boldsymbol{p}_{t+1}$ and $\boldsymbol{\Sigma_{t+1}}$ need to be estimated.

$$\boldsymbol{p}_{t+1} = \boldsymbol{p}_t \odot e^{\hat{\boldsymbol{\mu}} + \frac{1}{2}\text{diag}(\hat{\boldsymbol{\Sigma}})}, \tag{13}$$

where $\odot$ denotes an element-wise multiplication, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the sample estimates for the first two moments of the posterior distribution $f_{\theta_{post}}$ derived with entropy pooling introduced by Meucci (2008).

The prior distribution determining the profit and loss distribution of the investment universe is denoted as

$$\boldsymbol{X} \sim f_{\boldsymbol{\theta}_{prior}},$$

where $\boldsymbol{X}$ is a set of risk factors, and $\boldsymbol{\theta}_{prior}$ is a set of parameters that fully describe the probability density function f of $\boldsymbol{X}$. In case the predictions or views $\boldsymbol{\mathcal{V}}_{t+1} \notin \boldsymbol{\theta}_{prior}$, which could be for example views on expected values or variances of $\boldsymbol{X}$, a probability distribution $f_{\boldsymbol{\theta}}$ so that $\boldsymbol{\theta} \in \boldsymbol{\mathcal{V}}$ should be searched. However, at the same time the prior should be changed as little as possible. Meucci (2008) shows that this posterior distribution could be expressed as $f_{\boldsymbol{\theta}_{post}}$, where

$$\boldsymbol{\theta}_{post} = \text{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\mathcal{V}}}\, \varepsilon(\boldsymbol{\theta} \| \boldsymbol{\theta} prior) \tag{14}$$

and

$$\varepsilon(\boldsymbol{\theta} \| \boldsymbol{\theta}_{prior}) \equiv \int_{\Theta} f_{\boldsymbol{\theta}}(\boldsymbol{x})\, ln\Big(\frac{f_{\boldsymbol{\theta}}(\boldsymbol{x})}{f_{\boldsymbol{\theta} prior}(\boldsymbol{x})}\Big)\, d\boldsymbol{x},$$

where $\Theta \in \mathbb{R}^n$ and n is the number of the risk factors.

The approach to handle ranking views proposed in Meucci et al. (2014) is modified to outperformance constraints such that

$$\boldsymbol{\mathcal{V}} = [\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_n],$$

where

$$\boldsymbol{\nu}_k = \mu_k - \mu_{k+1} + q \leq 0,$$

n is a number of constraints which is one less than the number of assets, mu denotes the currency mean in the reference model, k is a ranking of the currency outperformance probability and $q \in \mathbb{R}$.

After imposing the inequality views, entropy minimization has to be solved numerically. The numerical implementation follows Meucci (2008) with scenarios in discrete time. The reference model $\boldsymbol{X}$ is modelled with a $J \times K$ -dimensional matrix $\boldsymbol{\mathcal{X}}$, in which the rows have J historical scenarios for K assets. Consequently, a row j is a single joint scenario of the assets K and a column k describes the marginal distribution of that asset returns. As in Meucci (2008, 8), each scenario of the reference model is assigned with a probability $\omega_j = 1/J$. Finally, the same scenarios $\boldsymbol{\mathcal{X}}$ are used with a different set of probability weights $\tilde{\boldsymbol{\omega}}$ in order to represent the posterior distribution.

Meucci (2008, 8–9) shows that in discrete time, the numerical implementation

of the entropy minimization problem in Equation 14 becomes

$$\tilde{\boldsymbol{\omega}} = \text{argmin}_{\boldsymbol{A}\boldsymbol{\upsilon}+q\leq 0}\, \varepsilon(\boldsymbol{\upsilon}, \boldsymbol{\omega}),$$

where $\tilde{\boldsymbol{\omega}}$, $\boldsymbol{\omega}$ and $\boldsymbol{\upsilon}$ are $J \times 1$ dimensional vectors of probability weights used to represent the posterior and prior distributions, respectively, whereas $\boldsymbol{\upsilon}$ is used in search of $\tilde{\boldsymbol{\omega}}$. Furthermore,

$$\varepsilon(\boldsymbol{\upsilon}, \boldsymbol{\omega}) = \sum_{j=1}^{J} \upsilon_j (ln\, \upsilon_j - ln\, \omega_j)$$

and as by Meucci et al. (2014), the column k of the $J \times (K-1)$ matrix $\boldsymbol{A}$ is

$$\boldsymbol{A}_k = \boldsymbol{\mathcal{X}}_k - \boldsymbol{\mathcal{X}}_{k+1}, \qquad k = 1, \ldots, K-1.$$

Finally, q is set to a small real number. Meucci (2008) translates this into a dual formulation which can be efficiently solved as it is a linearly constrained quadratic programming problem (LCQP).

In order to implement the entropy pooling approach, 156 weeks of historical excess currency returns are used as the historical scenarios $\boldsymbol{\mathcal{X}}$ and q is set to 0.00001. After the optimal probability weights $\tilde{\boldsymbol{\omega}}$ are found, the posterior scenarios are represented by

$$\tilde{\boldsymbol{\mathcal{X}}} = \tilde{\boldsymbol{\omega}} \odot \boldsymbol{\mathcal{X}} J.$$

Lastly, the mean vector $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\boldsymbol{\Sigma}}$ are estimated from $\tilde{\boldsymbol{\mathcal{X}}}$.

The robust approach does not distinguish between different strengths of the predictive signals, whereas the optimization based approach gives credit to the currencies according to their ranked predictive signals. Furthermore, the robust approach does not take into account the riskiness of the positions or the correlation structure. On the other hand, this means that the robust approach will not result in an extreme allocation and is less exposed to estimation errors. However, by limiting the portfolio volatility and assigning caps and floors for the holdings, a well balanced allocation should be achieved with the optimized approach as well. The targeted level of risk is set based on the volatility of an equally weighted long-only currency basket as this is the simplest example of a currency portfolio. The annualized volatility of this portfolio is approximately 7% during the first 286 weeks of sample data that is used in the first training and validation data split. However, the portfolio risk is limited to 6% annual in-sample volatility as mean-variance optimal portfolios are likely to realize higher volatilities out-of-sample. Finally, the maximum absolute value of individual currency holdings is limited to 0.6 in order to prevent extreme concentration resulting from estimation errors.

Finally, the portfolio performances are compared to the carry and 3-month

momentum portfolio returns. These portfolios are constructed according to the robust approach. The performance metrics used in the comparisons are mean, standard deviation, Sharpe ratio, skewness, excess kurtosis, 95% Value-at-Risk and maximum drawdown.

4.5 FX risk hedging

An U.S. based investor holding euro area equities has a total exposure that is similar to the exposure of a local equity investor in the euro area plus the currency risk that comes from holding assets denominated in a foreign currency. The currency risk could be hedged by taking a short position in the Euro and immediately unhedged by taking an additional long position in the same currency. Consequently, an unhedged international equity portfolio EQU can be decomposed so that

$$EQU = (EQU - B) + B$$

$$EQU = EQ + B,$$

where EQ is a fully hedged international equity portfolio and B is a basket of foreign currencies with exposures in line with the sizes of the foreign equity investments. The resulting portfolio on the left side of the equation does not have to be an unhedged portfolio, but the investor could freely choose any combination of currency exposures to customize the overall exposure of the portfolio in addition to the fully hedged equity component. The return of such a portfolio can be expressed as follows

$$r_p = r_{EQ} + \boldsymbol{w}'\boldsymbol{r}_{FX},$$

where, r_{EQ} is the return on EQ, $\boldsymbol{r}_{FX}$ is a $K \times 1$ vector of currency returns w.r.t. the USD and $\boldsymbol{w}$ is a $K \times 1$ vector of currency weights. Naturally, an investor would either choose $\boldsymbol{w}$ to minimize the portfolio risk or to maximize risk-adjusted returns. Boudoukh et al. (2018) show that a hedging currency portfolio that minimizes the variance of an international equity portfolio has weights

$$\boldsymbol{w}_{hedge} = -\boldsymbol{\Sigma}_{FX}^{-1}\boldsymbol{\Sigma}_{EQFX},$$

where $\boldsymbol{\Sigma}_{FX}$ is a $K \times K$ covariance matrix of the currency excess returns and

$$\boldsymbol{\Sigma}_{EQFX} = \big[\mathrm{cov}(\boldsymbol{r}_{EQ}, \boldsymbol{R}_{FX,1}), \mathrm{cov}(\boldsymbol{r}_{EQ}, \boldsymbol{R}_{FX,2}), \ldots, \mathrm{cov}(\boldsymbol{r}_{EQ}, \boldsymbol{R}_{FX,K})\big]',$$

where $\boldsymbol{R}_{FX,k}$ is the kth column vector of a $T \times K$ currency excess return matrix $\boldsymbol{R}_{FX}$ and $\boldsymbol{r}_{EQ}$ is a $T \times 1$ vector of returns on EQ. Consequently, Boudoukh et al.

(2018) combine the minimum variance portfolio

$$EQ_{minvar} = EQ + FX_{hedge},$$

where FX_{hedge} is a currency basket with weights $\boldsymbol{w}_{hedge}$.

Boudoukh et al. (2018) argue that currency weights $\boldsymbol{w}^*$ that maximize the Sharpe ratio of the portfolio are

$$\boldsymbol{w}^* = \boldsymbol{w}_{hedge} + \kappa \boldsymbol{w}_{alpha},$$

where $\boldsymbol{w}_{alpha}$ is a $K \times 1$ vector of currency weights that maximize the Sharpe ratio of a currency portfolio regardless of other asset classes and

$$\kappa = \frac{\sigma^2_{EQ\,minvar}}{r_{EQ\,minvar}}. \tag{15}$$

They show that Equation 15 can be written alternatively as

$$\kappa = \sigma_{EQ}\Big(\frac{1 - (\mathrm{cor}(\boldsymbol{r}_{EQ}, \boldsymbol{r}_{FX\,hedge}))^2}{\mathcal{S}_{EQ} - \mathrm{cor}(\boldsymbol{r}_{EQ}, \boldsymbol{r}_{FX\,alpha})\mathcal{S}_{FX\,alpha}}\Big),$$

where $\mathcal{S}$ denotes Sharpe ratio and FX_{alpha} is a currency portfolio with weights $\boldsymbol{w}_{alpha}$. Finally, Boudoukh et al. (2018) combine the modified portfolio mean–variance optimal (MPMVO) international equity portfolio as

$$EQ_{MPMVO} = EQ + FX_{hedge} + \kappa \cdot FX_{alpha},$$

which return is

$$r_{MPMVO} = r_{EQ} + (\boldsymbol{w}_{hedge} + \kappa \boldsymbol{w}_{alpha})\boldsymbol{r}_{FX}.$$

In this research, two approaches to decide $\boldsymbol{w}_{alpha}$ are tested: robust and optimization based routines. These approaches are described in Section 4.4 and are utilizing the predictions done by the LSTM models. The LSTM alpha portfolios are then applied to the currency management of an international equity portfolio in the MPMVO framework. The MSCI World and the MSCI World 100% hedged to USD indices are used as a proxy for the EQU and EQ equity portfolios and are obtained from Bloomberg. Finally, the performances of the resulting portfolios are analysed according to the same set of metrics as described in Section 4.4.

The hedging currency portfolio weights $\boldsymbol{w}_{hedge}$ are computed as in Boudoukh et al. (2018). However, $\boldsymbol{\Sigma}_{FX}$ and $\boldsymbol{\Sigma}_{EQFX}$ are estimated from the past two years of weekly excess currency returns, instead of weekly returns overlapping daily, with exponential weights

$$\omega_t = (1 - \lambda)\lambda^q,$$

where $q \in [0, 1, 2, \ldots, T-1]$ and λ is set to 0.94 so that the most recent observation has the largest weight. Finally, the estimated correlations are shrunk so that the

final estimate is 0.75 of the initial estimate as done by Boudoukh et al. (2018). Furthermore, same Sharpe ratios of 0.5 and 0.3 to EQ and FX_{alpha}, respectively, are used in order to calculate the κ. The empirical test is done fully out-of-sample and the portfolios are rebalanced and combined weekly.

5 RESULTS

5.1 Sample description

Table 3 reports summary statistics of the weekly G10 currency excess returns with respect to the USD during the sample period from the beginning of 1999 to the end of 2018. All excess currency returns have significant excess kurtosis which highlights the high crash risk related to currency trading. However, regarding the investment currencies, crashes are mainly sharp depreciations against the USD, whereas the opposite is true for funding currencies. This can be interpreted from the skewness measures in Table 3 and from the spikes in the cumulative excess returns in Figure 4.

Table 3: Summary statistics for the currency excess returns

The table reports mean, average forward premium, standard deviation and Sharpe ratios of G10 excess currency returns w.r.t. the USD with a weekly observation frequency. These figures are annualized. Furthermore, kurtosis, skewness and autocorrelation (ACF) of the excess currency returns are reported. Kurtosis is excess to normal and ACF has a lag of one. The sample covers the period from the 1st of January 1999 to the 31st of December 2018. Finally, the excess currency returns do not include transaction cost adjustments.

Currency	Mean	Avg. FP	St. dev.	Sharpe	Skewness	Kurtosis	ACF
EUR	0.006	-0.006	0.098	0.061	-0.167	1.323	0.026
GBP	0.006	0.010	0.092	0.065	-0.821	6.250	-0.002
JPY	-0.019	-0.025	0.103	-0.184	0.348	3.200	-0.077
CHF	0.006	-0.021	0.108	0.056	0.884	12.996	-0.009
CAD	0.014	0.002	0.087	0.161	0.076	5.999	-0.028
NZD	0.045	0.018	0.134	0.336	-0.377	3.681	-0.043
AUD	0.039	0.019	0.128	0.305	-0.391	8.033	-0.067
SEK	0.003	-0.002	0.117	0.026	-0.043	1.575	-0.010
NOK	0.012	0.006	0.116	0.103	-0.227	2.069	-0.044

The cumulative time-series excess returns of the currencies in Figure 4 are split into two categories based on their ranked average forward premiums. The overall returns of the investment currencies with higher forward premiums are higher than the returns of the funding currencies. However, the outperformance is not consistent and it seems that the higher returns could be compensation from bearing additional risks. Furthermore, the mean excess returns are not solely explained by the forward premiums as the spot prices of the investment and the funding currencies appreciate against the USD almost exclusively. This can be interpreted by comparing the mean excess returns and the average forward premiums in Table

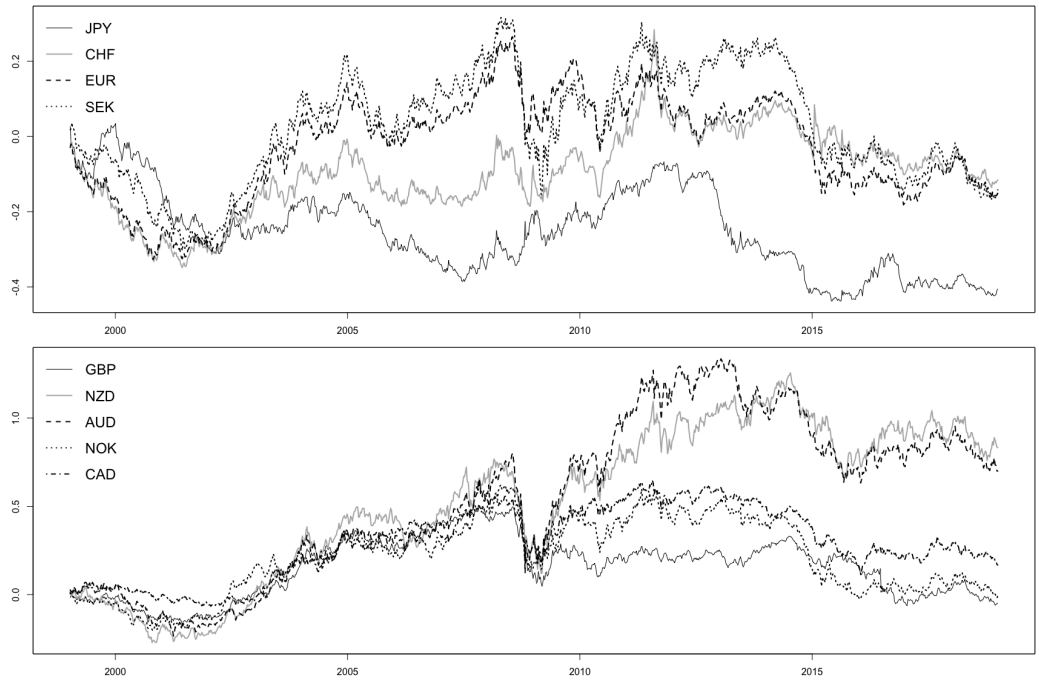

Figure 4: Cumulative excess currency returns

The figure displays time-series of cumulative compounded G10 excess currency returns w.r.t. the USD with a weekly observation frequency. The top (bottom) figure displays the returns of four (five) currencies with the lowest (highest) average forward premium. The sample covers the period from the 1st of January 1999 to the 31st of December 2018. Finally, the excess currency returns do not include transaction cost adjustments.

3. On the other hand, due to high volatilities and relatively low percentage returns, the Sharpe ratios are modest even for the best performing currencies.

The appreciation or depreciation trends, presented in Figure 4, of the currencies seem to be driven by the size of the interest rate differentials between the USD and other G10 currencies presented in Figure 5. The FED lowered the target interest rate before the great financial crisis as a result of the domestic housing market bubble burst. This increased the interest rate differentials and dragged the USD down. The opposite happened to the spot prices as other countries lowered the rates in the aftermath of the equity market crash in 2008. The interest rate differentials remained quite stable until the end of 2015 after which the interest rate differentials have declined due to the tightened monetary policy in the U.S.

Table 4 reports the summary statistics of the common factors used in the currency performance predictions. Each factor has high excess kurtosis which indicates high probability of extreme events. Furthermore, these extreme events are more often positive for VIX, σ^{FX}, TED and CS as interpreted from the positive skewness. They reach high values during market turmoil as visible from the spikes in Figure 6 that presents the time-series of the common factor values. However, VP and XVP exhibit negative skewness. Finally, the high autocorrelations in Table 4

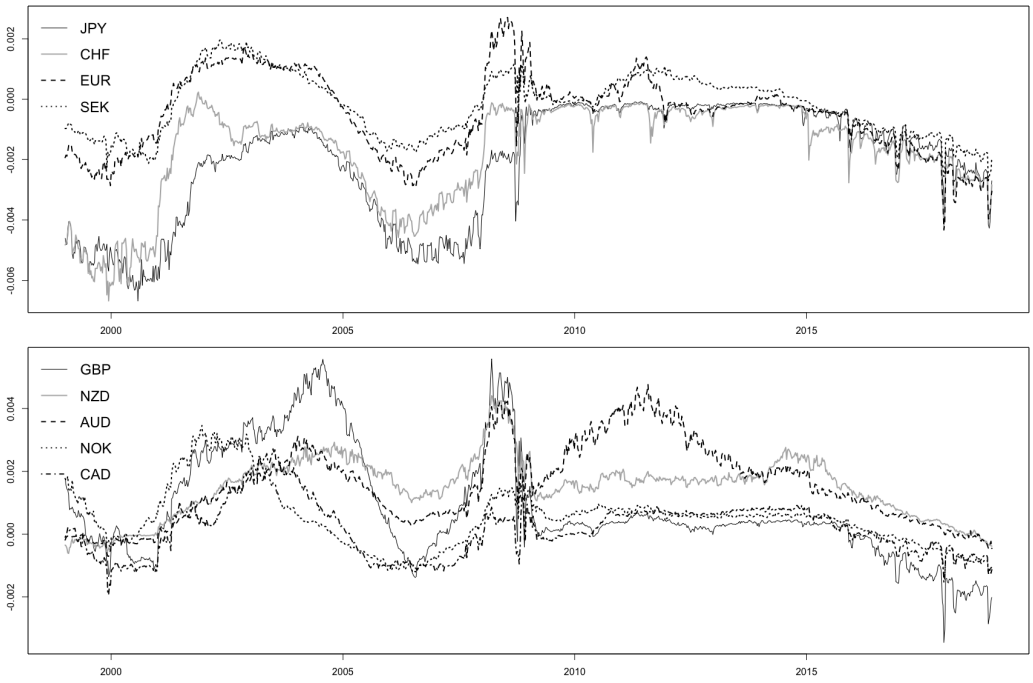

Figure 5: Currency forward premiums

The figure displays time-series of G10 currency 1-month forward premiums w.r.t. the USD with a weekly observation frequency. The top (bottom) figure displays the forward premiums of four (five) currencies with the lowest (highest) average forward premium. The sample covers the period from the 1st of January 1999 to the 31st of December 2018.

are partly explained by the temporal overlapping used in the construction of these factors.

Table 4: Summary statistics for the common factors

The table reports mean, maximum, minimum, standard deviation, skewness, kurtosis and autocorrelation (ACF) of the common factors with a weekly observation frequency. Kurtosis is excess to normal and ACF has a lag of one. The sample covers the period from the 8th of February 1999 to the 31st of December 2018.

Factor	Mean	Max	Min	St. dev.	Skewness	Kurtosis	ACF
VIX	0.058	0.231	0.027	0.025	2.006	6.632	0.923
σ^{FX}	0.105	0.371	0.044	0.036	2.698	12.534	0.961
VP	0.001	0.013	-0.032	0.003	-4.879	55.778	0.587
XVP	0.000	0.002	-0.005	0.000	-5.934	53.412	0.744
TED	0.004	0.046	0.000	0.004	4.034	23.897	0.952
CS	0.026	0.061	0.015	0.008	1.670	4.929	0.994
COM	0.001	0.217	-0.204	0.027	-0.242	8.203	-0.051

Next, in-sample predictions are conducted through OLS regression analysis. This is done in order to exploit the predictive power of the factors, and to conduct explanatory analysis before moving on to the black box predictions with recurrent neural networks. Excess currency returns are predicted with the past excess cur-

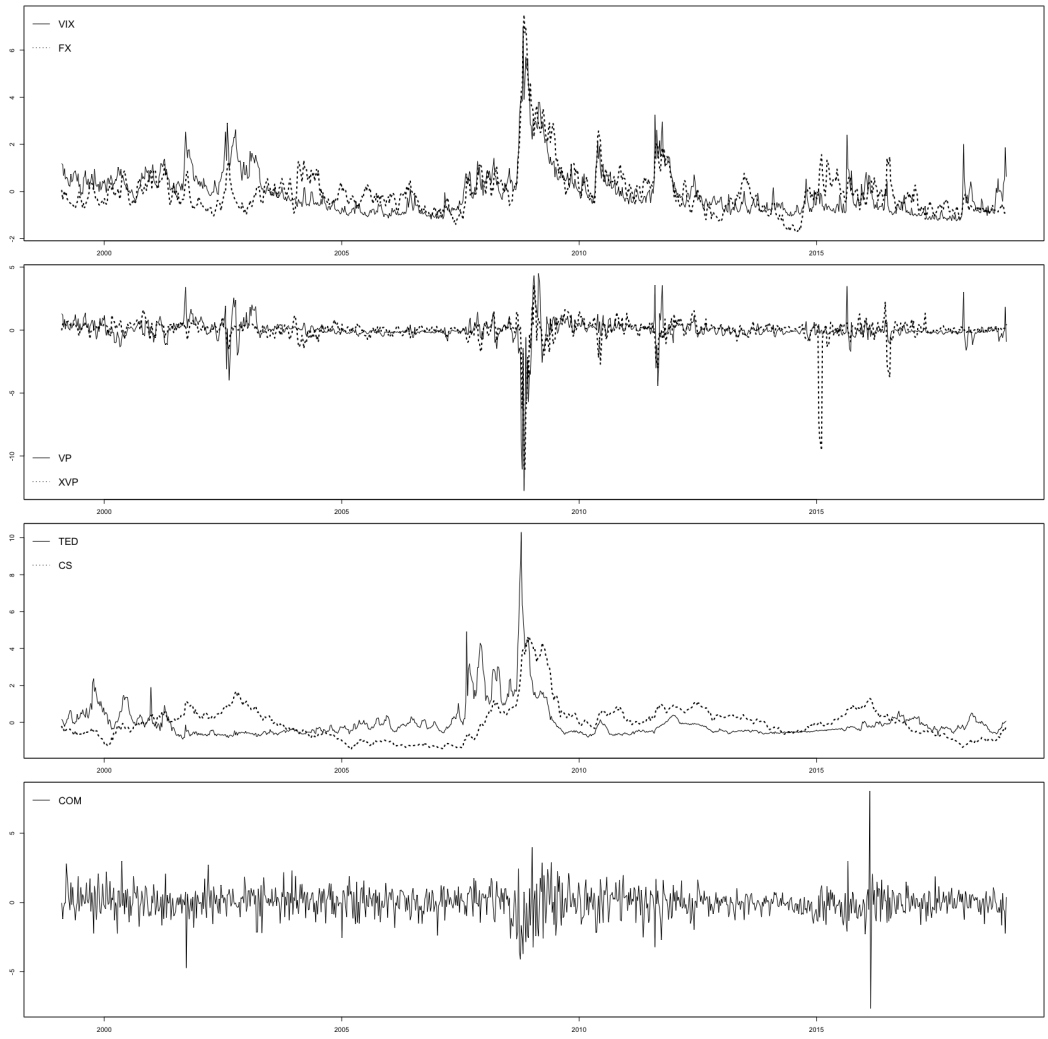

Figure 6: Common factors

The figure displays time-series of the common factors with a weekly observation frequency. All observations are standardized for the ease of comparison. The top figure displays the observations of the VIX index (VIX) and the currency volatility σ^{FX} (FX). The second figure displays the observations of the equity and currency market variance risk premiums (VP and XVP) whereas the third figure includes the TED and credit spreads (TED and CS). Finally, the bottom figure displays the returns of the commodity index (COM). The sample covers the period from the 8th of February 1999 to the 31st of December 2018.

rency returns (RX), the forward premiums (FP) and all seven common factors, which have one-week lags. Table 5 reports the signs of the β coefficients of each factor for every excess currency return series. Furthermore, the statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). The statistical significance is marked with ***, ** and * denoting 99%, 95% and 90% confidence levels, respectively. Finally, adjusted R^2s and p-values of F-statistics with a null hypothesis that none of the coefficients are different from zero are presented in the last two columns on the right hand side.

Table 5: In-sample OLS regression analysis with a one-week lag

The table reports the OLS regression coefficient signs, statistical significance of the coefficients, adjusted R^2s and p-values of the F-test. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variables are listed on the row label and the regressors are listed on the column label. The regressor observations are lagged with one week. Finally, the sample covers the period from the 8th of February 1999 to the 31st of December 2018.

Currency	RX	FP	VIX	σ^{FX}	VP	XVP	TED	CS	COM	Adj. R^2	p-value
EUR	+	+	-	+	+	-	-	+	-	0.002	0.247
GBP	-	+	-**	+	+**	+	-	+	-	0.025	0.000
JPY	-*	+	+	-	-***	-	+	+	+	0.019	0.001
CHF	-	+	-*	+	-	+	-	+	-	0.003	0.221
CAD	-	-	+	+	+	+	-	-	+	0.022	0.000
NZD	-	+	+	+	+	+	-	+	+	0.010	0.020
AUD	-	-	+	+	+	+	-	-	+	0.021	0.000
SEK	-	+	-	+	+	+	-	+	-	0.012	0.010
NOK	-	+	-	+	+**	-	-	+	+	0.014	0.005

The coefficients of the past excess currency returns or forward premiums are not statistically significant at 95% confidence level. Consequently, these coefficients do not differ statistically significantly from zero. However, the positive signs of the FP coefficients indicate that a relative increase in a foreign interest rate should increase the expected return of this currency position as the UIP does not hold. Furthermore, patterns in Figures 5 and 4 display this relationship supporting the profitability of the dollar carry trade strategy.

The coefficient signs of VIX are mixed and statistically significant only to GBP and CHF at 95% and 90% confidence levels, respectively. Furthermore, the coefficient signs of the σ^{FX} and XVP are not in line with the previous findings by Londono and Zhou (2017) and Bakshi and Panayotov (2013) nor are they statistically significant.

According to the results reported in Table 5, VP captures the most relevant information compared to the other factors individually. This factor is constructed exclusively of variables derived from the U.S. equity market and, consequently, is more linked to the uncertainty in the U.S. economy than truly global σ^{FX} and XVP factors. This might explain the depreciation of the USD against most other G10 currencies after an increase in risk–aversion as also argued by Londono and Zhou (2017). In contrast, the coefficients of TED support the safe haven phenomenon, which is in line with the evidence provided by Mancini et al. (2013) and Brunnermeier et al. (2009). In particular, during periods of poor liquidity, investors tend to decrease their foreign currency positions, thus, causing the appreciation of the USD.

Finally, the last two factors: CS and COM do not explain excess currency returns in the next time step as none of the coefficients are statistically significant at 90% confidence level. However, despite the mixed coefficient signs of the former factor, the coefficient signs of the latter factor are in line with the findings provided by Bakshi and Panayotov (2013). In particular, positive changes in commodity price index are followed by appreciations of the currencies of the main commodity exporters. Furthermore, Menkhoff et al. (2012b) find a negative link between the creditworthiness of a country and the momentum return of its currency. Consequently, the low explanatory power of the credit spread might be explained by the sample that does not include currencies of speculative countries.

Table 6 reports the correlation structure between the common factors. Due to the strong correlations between the uncertainty measures: VIX, σ^{FX} and CS and the risk-aversion measures: VP and XVP, the explanatory analysis, which results are reported in Table 5 suffer from collinearity. Consequently, the previous interpretations of single factor coefficients might not be reliable.

Table 6: Correlation matrix of the common factors

The table reports the pair-wise correlations of the common factors with a weekly observation frequency. The sample covers the period from the 8th of February 1999 to the 31st of December 2018.

	VIX	σ^{FX}	VP	XVP	TED	CS	COM
VIX	1						
σ^{FX}	0.71	1					
VP	-0.04	-0.31	1				
XVP	-0.20	-0.49	0.55	1			
TED	0.49	0.52	-0.36	-0.24	1		
CS	0.74	0.60	-0.12	-0.05	0.36	1	
COM	-0.15	-0.08	-0.01	0.01	-0.12	-0.08	1

In order to overcome the issue of collinearity, the factor set is decomposed to latent factors by first conducting a singular value decomposition (SVD)

$$\boldsymbol{X} = \boldsymbol{U}\Sigma\boldsymbol{W}',$$

where $\boldsymbol{X}$ is a $n \times m$ -dimensional matrix with the standardized original factors as column vectors, $\boldsymbol{U}$ is a $n \times n$ -dimensional matrix, Σ is a $m \times m$ -dimensional diagonal matrix with factor volatilities and $\boldsymbol{W}$ is a $m \times m$ -dimensional matrix of right singular vectors of $\boldsymbol{X}$ i.e. factor loadings. Then the orthogonal latent factors

$$\boldsymbol{L} = \boldsymbol{X}\boldsymbol{W}.$$

This technique is called principal component analysis (PCA).

Table 7 reports the factor loadings $\boldsymbol{W}$ of the common factors. The latent factors in the column of the table are constructed as linear combinations of the observed factors on the row of the table. The first latent factor L_1 has largest loadings on VIX, σ^{FX} and CS, whereas L_2 loads heavily on VP and XVP. Furthermore, L_3 loads almost exclusively on COM and L_4 loads on TED.

Table 7: Latent factor loadings

The table reports latent factor loadings of the common factors. The loadings are computed from the standardized time-series observations of the common factors by conducting singular value decomposition (SVD) to the observation sample that starts from the 8th of February 1999 and ends to the 31st of December 2018.

	L_1	L_2	L_3	L_4	L_5	L_6	L_7
VIX	0.464	0.368	0.086	0.051	-0.110	-0.609	-0.571
σ^{FX}	0.512	0.034	0.102	-0.192	-0.153	0.745	-0.332
VP	-0.257	0.640	0.091	-0.181	-0.625	0.090	0.291
XVP	-0.310	0.547	-0.040	0.493	0.435	0.253	-0.327
TED	0.398	-0.053	-0.113	0.796	-0.331	0.009	0.287
CS	0.439	0.340	0.183	-0.170	0.520	-0.013	0.599
COM	-0.095	-0.191	0.962	0.148	-0.071	-0.040	-0.013

The conclusions from the results in Table 7 are further supported by Table 8 that reports the R^2s of individual OLS regressions in which the latent factors in the column of the table are explained with the observed factors on the row of the table separately. VIX, σ^{FX} and CS each explain over 60% of the variation of the first latent factor L_1. Consequently, L_1 is named as a price risk factor PR. VP and XVP explain approximately one half of the variation of L_2 that is, thus, named as a risk-aversion factor RA. L_3 is named as a commodity price factor CP and L_4 is named as a funding liquidity factor FL since COM and TED are the only factors capable of explaining their variations meaningfully.

Table 8: Univariate OLS regression R^2s of latent factor explanatory analysis

The table reports R^2s of univariate OLS regressions, where the column response variables are explained with the row explanatory variables. The latent factors on the column label are principal components that are derived from the common factor observation sample by conducting singular value decomposition (SVD). The sample covers the period from the 8th of February 1999 to the 31st of December 2018.

	L_1	L_2	L_3	L_4	L_5	L_6	L_7
VIX	0.66	0.19	0.01	0.00	0.00	0.08	0.05
σ^{FX}	0.81	0.00	0.01	0.02	0.01	0.12	0.02
VP	0.20	0.57	0.01	0.02	0.17	0.00	0.02
XVP	0.30	0.42	0.00	0.16	0.08	0.01	0.02
TED	0.49	0.00	0.01	0.43	0.05	0.00	0.02
CS	0.60	0.16	0.03	0.02	0.12	0.00	0.07
COM	0.03	0.05	0.90	0.01	0.00	0.00	0.00

The last three latent factors are not intuitively linked to any observed factor. In addition, they capture only a small fraction of the overall variation in the factor set. Table 9 reports the proportional and cumulative proportional variances of the latent factors. The first four latent factors capture almost 90% of the variation in the standardized data. This means that the last three latent factors are rather irrelevant. Consequently, the usage of the four first latent factors is preferred over the usage of the full common factor set in the explanatory analysis henceforth.

Table 9: Proportional and cumulative proportional variances of the latent factors

The first row of the table includes the variances of the latent factors in proportion to the sum of all variances whereas the second row includes the cumulative sum of these proportional variances. The latent factors on the column label are principal components that are derived from the common factor observation sample by conducting singular value decomposition (SVD). The sample covers the period from the 8th of February 1999 to the 31st of December 2018.

	L_1	L_2	L_3	L_4	L_5	L_6	L_7
$\frac{\sigma^2(L_j)}{\sum_i \sigma^2(L_i)}$	0.418	0.211	0.143	0.100	0.065	0.035	0.029
Cumulative sum	0.418	0.628	0.771	0.871	0.936	0.971	1.000

Next, the results of the OLS regression analysis are reported in Table 5. The analysis is conducted with reduced dimensions of the factors. Excess currency returns are regressed in-sample against the four first latent factors. Table 10 reports the coefficient signs of the latent factors and their statistical significance is marked with ***, ** and * denoting 99%, 95% and 90% confidence levels, respectively. In addition, adjusted R^2s and p-values of F-statistics with a null hypothesis that none of the coefficients are different from zero are reported in the last two columns on the right hand side.

Table 10: In-sample OLS regression analysis with the latent factors and a one-week lag

The table reports the OLS regression coefficient signs, statistical significance of the coefficients, adjusted R^2s and p-values of the F-test. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variables are listed on the row label and the regressors are listed on the column label. The regressor variables are the first four principal components of the common factor observation sample. Furthermore, the regressors are lagged with one week. Finally, the sample covers the period from the 8th of February 1999 to the 31st of December 2018.

Currency	PR	RA	CP	FL	Adj. R^2	p-value
EUR	-	+	-	-	0.003	0.138
GBP	-***	+*	+	-**	0.020	0.000
JPY	+**	-**	+	+	0.013	0.002
CHF	-	+	+	-	-0.002	0.788
CAD	-	+**	+	-	0.022	0.000
NZD	-	+*	+	-	0.013	0.002
AUD	-	+**	+	-	0.019	0.000
SEK	-	+**	+	-	0.016	0.000
NOK	-	+***	+	-	0.014	0.001

The coefficient signs of PR and FL in Table 10 are in line with each other, whereas RA seems to have the opposite effect on the excess currency returns. In line with the results reported in Table 5 and with Mancini et al. (2013) and Brunnermeier et al. (2009), increased price uncertainty or illiquidity cause appreciation of the safe haven assets, the USD and the JPY. In contrast, increased risk aversion tend to weaken the USD and JPY against other currencies. This coincides with the previous interpretation of the variance risk premium but not with the currency variance risk premium. However, Londono and Zhou (2017) show that the dependency of these two factors and the excess currency returns is at its strongest after one to four months and, thus, might differ from the analysis based on a one-week lag between the observations. Finally, as in Table 5, commodity prices CP are not statistically significant.

Finally, the null hypothesis that none of the factors are capable of predicting the excess currency returns in the next time step are reported in the last column of Table 10. The null hypothesis is rejected for all currencies, except EUR and CHF. Furthermore, the R^2s are low for all G10 currencies, CAD having the highest figure of 2.2%.

A concern regarding the results of the OLS regressions is the dominance of the period around the great financial crisis since during that period, the absolute values of the observations are significantly larger compared to calmer market conditions.

Consequently, weighted least squares regressions [1] are conducted, as a robustness check. However, this analysis results only in minor changes in the coefficient signs which do not change the bigger picture. On the other hand, almost all coefficients become more uncertain and the adjusted R^2s decrease. This result is expected, since the market is more likely driven by price uncertainty, risk-aversion and funding liquidity during market turmoil than during calm periods as also pointed out by Mancini et al. (2013).

To conclude, excess G10 currency returns are characterized by high probability of extreme events and low Sharpe ratios even for the best performing currencies. Furthermore, forward premiums or past returns do not predict excess currency returns with one-week lags. However, currencies with relatively high interest rates tend to appreciate, whereas currencies with relatively low interest rates tend to depreciate over long time periods, thus, neglecting the UIP. In addition, price uncertainty, risk-aversion and funding liquidity are driving the excess currency returns especially during market turmoil. On the other hand, these factors explain only a small fraction of the overall variance in the excess currency returns.

5.2 LSTM and RNN predictions

In this section, the predictive accuracies of LSTM and RNN models are reported. Furthermore, two statistical tests are conducted in order to analyse the predictive power of the models as well as to evaluate if the additional complexity from the LSTM cells is necessary. Finally, the difference of prediction accuracies of the two best performing models are analysed through an explanatory factor analysis.

Throughout this section, two different input data sets are considered. Notations RX and All refer to past excess returns and all factors (RX, FP, VIX, σ^{FX}, VP, XVP, TED, CS and COM), respectively. Moreover, the sample covers the period from the 8th of February 1999 to the 31st of December 2018. However, the predictions start from the 3rd of October 2005.

Table 11 reports out-of-sample prediction accuracies of the LSTM and simple RNN models with two factor sets: RX and All. Accuracies are reported to each currency separately as well as to the total accuracy of a model. Finally, probabilities that the models would have achieved the total accuracies by chance are listed on the bottom line.

Each model performance differs statistically significantly from a random guess,

[1] Each time step is weighted based on the inverse of the conditional variance of the currency excess returns in hand. The conditional variances are computed with a Garch(1,1)-model and the residuals are assumed to be normally distributed.

Table 11: LSTM and RNN accuracies and probabilities that a model would have achieved the total accuracy by chance

The table reports the out-of-sample prediction accuracies of the models on the column label. RX denotes that only the excess currency returns are used as predictive factors, whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. Furthermore, the table reports the prediction accuracies for the currencies on the row label separately, as well as the total prediction accuracies of each model choice. Finally, probabilities that the models would have achieved the total accuracies by chance are reported in the bottom row of the table. The sample covers the period from the 8th of February 1999 to the 31st of December 2018. However, the predictions start from the 3rd of October 2005.

Currency	LSTM (RX)	LSTM (All)	RNN (RX)	RNN (All)
EUR	0.614	0.601	0.553	0.568
GBP	0.600	0.591	0.512	0.564
JPY	0.556	0.561	0.510	0.527
CHF	0.581	0.581	0.530	0.549
CAD	0.539	0.523	0.512	0.493
NZD	0.519	0.527	0.484	0.506
AUD	0.483	0.481	0.488	0.504
SEK	0.569	0.561	0.539	0.555
NOK	0.536	0.525	0.533	0.520
Total	0.555	0.550	0.518	0.532
BT	(0.000)	(0.000)	(0.000)	(0.000)

since the probabilities that the prediction accuracies would be 50% are practically zero. Both LSTMs are more accurate than RNNs with AUD and NOK being the only exceptions. However, regarding the NOK, LSTMs provide higher accuracies than RNNs if the methods are compared along the same input data, whereas none of the models provide valuable insight regarding the AUD. Finally, LSTMs deliver the highest prediction accuracies for the EUR, GBP and CHF.

A statistical test proposed by Diebold and Mariano (1995) is used in order to compare the goodness of the models. Table 12 reports the p-values of the Diebold–Mariano test. p-value smaller than 0.05 means that the prediction accuracy of the model in the row label is superior w.r.t. the model in the column label at 95% confidence level. Both LSTM models deliver higher prediction accuracies than the RNNs at confidence levels above 99%. However, the Diebold–Mariano test doesn't indicate statistical difference between the prediction accuracies of the LSTMs.

The difference between prediction accuracies is higher among the models using only historical return data as a predictive factor. As a result, the model complexity might be more valuable when only price trends are explored, whereas other factors show also short-term dependencies which were also concluded by running the in-sample OLS regressions in Section 5.1. Consequently, the LSTM most likely captures long-term trends from the past excess return data, since a simple RNN

Table 12: Diebold–Mariano test of model outperformance

The table reports the p-values of the Diebold–Mariano test. p-value smaller than 0.05 means that the prediction accuracy of the model in the row label is superior w.r.t. the model in the column label at 95% confidence level. RX denotes that only the excess currency returns are used as predictive factors, whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. The sample covers the period from the 8th of February 1999 to the 31st of December 2018. However, the predictions start from the 3rd of October 2005.

	LSTM (RX)	LSTM (All)	RNN (RX)	RNN (All)
LSTM (RX)	–	0.163	0.000	0.000
LSTM (All)	0.837	–	0.001	0.002
RNN (RX)	0.999	0.999	–	0.947
RNN (All)	0.999	0.998	0.053	–

is proven to miss these patterns due to the vanishing gradient problem. However, the models utilizing all factors, might capture predictive signals from shorter temporal distances, because the difference in prediction accuracies is lower between LSTM(All) and RNN(All) than between the RX models. Even though, Fischer and Krauss (2018) are not investigating the differences between simple RNNs and LSTMs, they prove that LSTM outperforms less complex random forests and deep neural networks in price trend recognition. Consequently, their findings are in harmony with the results reported in Tables 11 and 12.

In order to exploit the potential time variation and differences of the model performances, 26-week rolling accuracies are presented in Figure 7. The time-series of the RNN model accuracies are placed at the top panel, whereas the bottom panel displays the time-series accuracies of the LSTM models. The RNN accuracies do not follow clear pattern. Furthermore, their performances stay below the 50% threshold several times for approximately one-year periods. On the other hand, LSTMs show common cyclical performance with approximately 3-year cycles and the rolling prediction accuracies fluctuate between 50% and 60%. Consequently, the LSTM models provide not only higher but also consistently valuable prediction performance.

Even though there is not a statistically significant difference between the LSTM model prediction accuracies, there seems to be periods of systematic outperformance between the models. Figure 8 displays the difference of the 26-week rolling accuracies of the LSTM models. The difference is calculated as the rolling prediction accuracy of the LSTM model utilizing all factors minus the rolling prediction accuracy of the LSTM model that use only the past returns as a predictive factor. The outperformance of LSTM(All) lasts from the end of 2005 until the mid 2010. After this period LSTM(RX) delivers higher prediction accuracy during the next year and a half. However, the difference fluctuates around zero until the end of

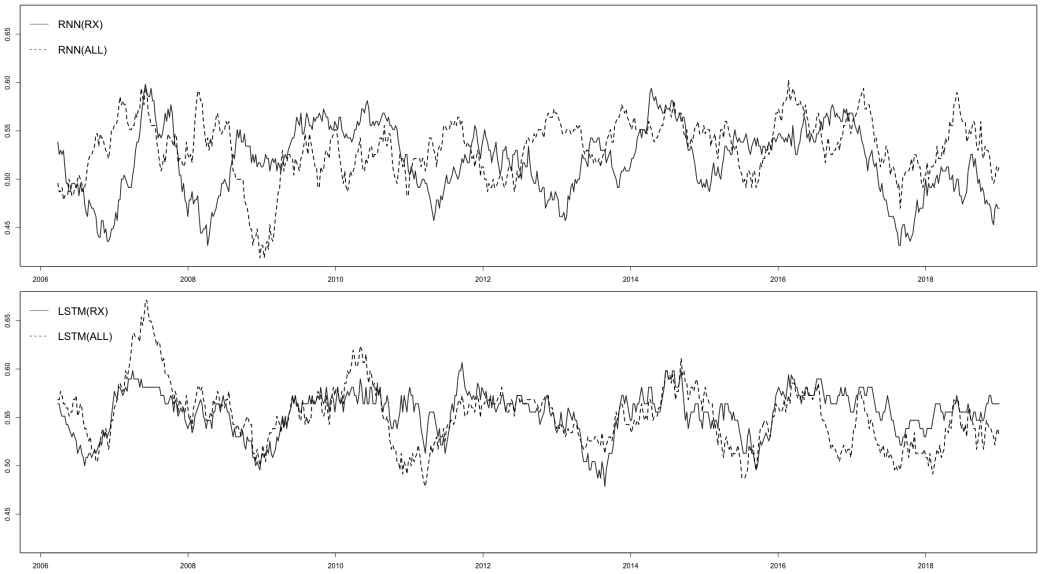

Figure 7: LSTMs and RNNs 26-week rolling accuracies

The figure displays the time-series of a 26-week rolling total prediction accuracies of the prediction models. The top (bottom) figure displays the time-series of the simple RNN (LSTM) model rolling prediction accuracies. The sample of the prediction accuracy observations covers the period from the 3rd of October 2005 to the 31st of December 2018.

2015. Finally, LSTM(RX) outperforms LSTM(All) during the rest of the sample period.

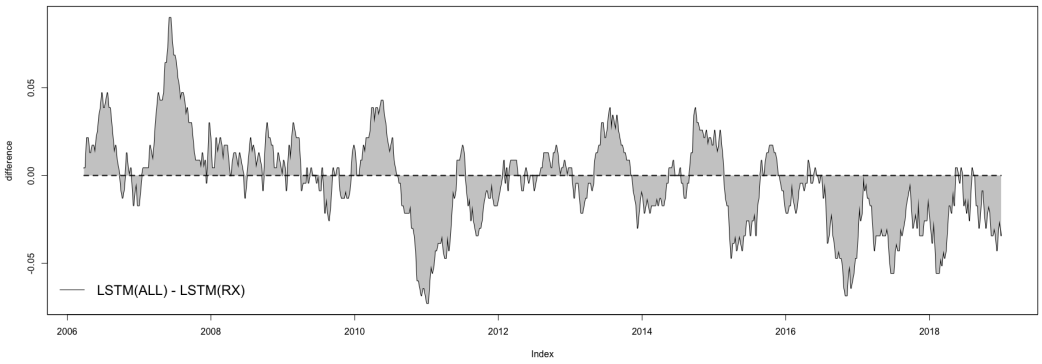

Figure 8: All minus RX 26-week rolling accuracies

The figure displays the time-series of the difference between 26-week rolling total prediction accuracies of the LSTM(All) and the LSTM(RX) models. The sample of the prediction accuracy observations covers the period from the 3rd of October 2005 to the 31st of December 2018.

An in-sample OLS regression analysis is conducted to provide a statistical explanation on the time-variation of the prediction accuracy differences. Table 13 reports the coefficient signs, statistical significance of the coefficients, an adjusted R^2 and a p-value of the F-test. ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The regressors are the first

four principal components derived from the common factor observation sample by conducting a singular value composition plus a dummy variable FC_{dummy} which has a value of one between December 2007 and June 2009 and a value of zero otherwise.

Table 13: In-sample OLS regression analysis of LSTM accuracy differences with a lag of one

The table reports the OLS regression coefficient signs, statistical significance of the coefficients, an adjusted R^2 and a p-value of the F-test. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variable on the row label is computed as a difference between 26-week rolling total prediction accuracies of the LSTM(All) and the LSTM(RX) models. The regressors are the first four principal components derived from the common factor observation sample by conducting singular value composition plus a dummy variable FC_{dummy} which has a value of one between December 2007 and June 2009 and a value of zero otherwise. This period is classified as a recession by NBER (2010). Furthermore, the regressors are not lagged. Finally, the sample covers the period from the 8th of February 1999 to the 31st of December 2018. However, the first predictions start from the 3rd of October 2005.

Acc. Diff.	PR	RA	CP	FL	FC_{dummy}	Adj. R^2	p-value
LSTM: ALL - RX	+	-	-	+	+	0.041	0.000

In Table 13, none of the factors are statistically significant at 90% confidence level. The coefficient signs of price risk, commodity prices, funding liquidity and the financial crisis dummy factors indicate that market turmoil leads to the out-performance of the LSTM(All) model, whereas the risk-aversion factor seems to have the opposite effect. However, in addition to the low statistical significance of the factor coefficients, the R^2 of 4.1% means that the majority of the variation remains unexplained. This variation might be better analysed in the form of the currency portfolios that are built based on the LSTM model predictions.

5.3 LSTM portfolios

This section provides summary statistics and an explanatory factor analysis of the currency portfolios, which are constructed based on the predictions of the LSTM models analysed in the previous section. Two portfolio construction methods are applied to both LSTM model predictions, which utilize different predictive input factor sets. Notations RX and All refer to past excess returns and all factors (RX, FP, VIX, σ^{FX}, VP, XVP, TED, CS and COM), respectively. Furthermore, the two portfolio construction methods are the robust and the optimized approach. In addition to comparing the performances of the LSTM portfolios, their performances are compared against the carry and the 3-month momentum portfolio returns. The

two latter portfolios are constructed with the robust approach which is a common choice in the academic literature. Finally, 3-month momentum is chosen since it is the only momentum strategy that was profitable during the sample period. The portfolio returns are reported and analysed in the sample period from the 3rd of October 2005 to the 31st of December 2018.

Table 14 reports the descriptive statistics of the currency portfolio returns. Among the LSTM portfolios, the mean returns of the robust portfolios are higher than the mean returns of the optimized portfolios. However, portfolios based on the LSTM(All) model predictions present higher returns than the portfolios that use the predictions of the LSTM(RX) model, even though, the prediction accuracy of the latter model was slightly higher.

Table 14: Summary statistics of LSTM portfolio returns

The table reports mean, standard deviation and Sharpe ratios of currency portfolio returns with a weekly observation frequency. These figures are annualized. Furthermore, excess kurtosis, skewness, sample 95% value-at-risk and maximum drawdown of the currency portfolio returns are reported. The first four portfolios on the row label are built based on the LSTM models predictions. RX denotes that only the excess currency returns are used as predictive factors, whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. Robust (Opt) denotes that the robust (optimization based) portfolio construction technique is used. Furthermore, the carry and 3-month momentum (Mom3) portfolios are constructed with the robust portfolio construction technique. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018. Finally, the currency portfolio returns do not include transaction cost adjustments.

	Mean	St.dev.	Sharpe	Skewness	Kurtosis	VaR_{95}	Max D.d.
Robust (RX)	0.018	0.083	0.217	-0.017	9.926	-0.017	-0.206
Robust (ALL)	0.022	0.068	0.324	-0.462	2.185	-0.015	-0.151
Opt (RX)	0.006	0.071	0.085	-0.501	2.199	-0.016	-0.225
Opt (ALL)	0.016	0.078	0.205	-1.232	8.680	-0.016	-0.231
Carry	0.018	0.080	0.225	-0.733	6.474	-0.017	-0.197
Mom3	0.002	0.068	0.029	0.234	7.226	-0.015	-0.212

The risk-adjusted returns reported in Table 14 in form of Sharpe ratios set a clear difference between the portfolios utilizing all predictive factors and the portfolios using only past returns. The carry portfolio delivers the second highest Sharpe ratio after the Robust(RX) portfolio. However, further analysis of higher moments and alternative performance measures show clear outperformance of the Robust(All) portoflio relative to the other portfolios. It has less negative skewness than the carry portfolio and the lowest excess kurtosis of all portfolios. Furthermore, the robust approach that uses all factors has less tail risk according to the VaR and maximum drawdown measures.

The performance metrics do not reveal exposures to common risk factors in currency trading strategies. These exposures are analysed through in-sample OLS

regressions and the results are reported in Table 15. The table reports the OLS regression coefficient signs, statistical significance of the coefficients, adjusted R^2s and p-values of the F-test. The regressors are the first four principal components derived from the common factor observation sample by conducting a singular value composition plus a dummy variable FC_{dummy} which has a value of one between December 2007 and June 2009 and a value of zero otherwise.

Table 15: In-sample OLS regression analysis with the latent factors and a one-week lag

The table reports the OLS regression coefficient signs, statistical significance of the coefficients, adjusted R^2s and p-values of the F-test. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variables are listed on the row label and the regressors are listed on the column label. The first four portfolios on the row label are built based on the LSTM RNN model predictions. RX denotes that only the excess currency returns are used as predictive factors whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. Robust (Opt) denotes that the robust (optimization based) portfolio construction technique is used. Furthermore, the carry and 3-month momentum (Mom3) portfolios are constructed with the robust portfolio construction technique. The regressors are the first four principal components derived from the common factor observation sample by conducting a singular value composition plus a dummy variable FC_{dummy} which has a value of one between December 2007 and June 2009 and a value of zero otherwise. This period is classified as a recession by NBER (2010). The regressor observations are not lagged. Finally, the sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

	α	PR	RA	CP	FL	FC_{dummy}	R^2	p-value
Robust (RX)	-	$-^{**}$	$-^{***}$	$+^{***}$	-	+	0.223	0.000
Robust (ALL)	+	+	-	$+^{***}$	+	-	0.040	0.000
Opt (RX)	+	-	$-^{**}$	-	-	+	0.015	0.008
Opt (ALL)	+	-	-	+	-	+	0.010	0.039
Carry	-	$-^{**}$	$-^{***}$	$+^{***}$	-	+	0.208	0.000
Mom3	+	+	$+^{**}$	-	+	-	0.040	0.000

None of the models have a statistically significant α. However, RA is a statistically significant risk factor for Robust(RX), Robust(All), carry and 3-month momentum portfolios and its coefficients have negative signs, except the 3-month momentum portfolio that has a positive coefficient sign. Furthermore, CP has positive and statistically significant coefficients for both of the robust LSTM portfolios and the carry portfolio. Finally, the price risk factor PR has statistically significant and positive coefficients for the carry and Robust(RX) portfolios. In fact, these portfolios share similar risk factor exposures overall.

Robust(RX) and the carry portfolios have a significant amount of variance explained by the five factors, which can be interpreted from the R^2s of above 0.2. By interpretation of the coefficient signs of the statistically significant factors for these two strategies, Robust(RX) and carry portfolios tend to face losses during

market turmoil, since the volatility and risk-aversion are likely to pick-up, while the commodity prices are depreciating. These dynamics are in line with the findings from Londono and Zhou (2017), Bakshi and Panayotov (2013) and Lustig et al. (2011). In contrast, the coefficient signs of the FC_{dummy} factor indicate that these portfolios would perform better during the period of severe financial distress after taking into account other factors. However, this dummy variable is not statistically significant for any of the portfolios which also applies to the coefficients of the funding liquidity factor FL.

Unlike with the carry and Robust(RX) portfolios, the five factors capture only a small fraction of the variation in the returns of the best performing model, Robust(All), as well as the returns of the optimized approaches. The robust approach that uses all factors has a positive and statistically significant coefficient sign of the commodity price factor, whereas the Opt(RX) has a negative and statistically significant coefficient sign only for RA which is opposing to the risk exposure of the 3-month momentum strategy. Furthermore, Opt(All) does not have any statistically significant factor coefficients.

The time-series of the cumulative compounded returns of the LSTM portfolios are plotted in Figure 9. The Robust(RX) and Robust(All) portfolios end up higher than the respective optimized portfolios. However, the outperformance of the cumulative returns is due to a massive shift in return trends, since after the long bull run from 2009 to 2016, the optimized portfolios suffer from severe losses during the following two years, whereas the pattern of the robust portfolio returns is almost the opposite. The robust portfolio returns start to rebound rapidly at the end of 2015 after a three year decline. Finally, the optimized approach utilizing all factors outperforms the respective RX strategy consistently in terms of returns. In addition, the risk measures in Table 15 do not differ much between the optimized approaches and, thus, including all factors in the prediction task seems to add value.

Despite the more attractive performance metrics of the Robust(All) portfolio in Table 14 and lower risk exposures in Table 15, the outperformance periods of the robust portfolios differ in Figure 9 as the Robust(All) portfolio delivers higher returns during the financial crisis and between 2016 and the end of the sample period. On the other hand, the Robust(RX) portfolio outperforms the Robust(All) portfolio in the beginning of the sample period as well as during the turning point of 2013.

Figure 10 demonstrates the time-series of the cumulative compounded returns of the Robust(All), Opt(All), carry and 3-month momentum portfolios. The LSTM portfolios using only past excess currency returns are excluded due to their less attractive performances. The optimized portfolio returns are not clearly linked to

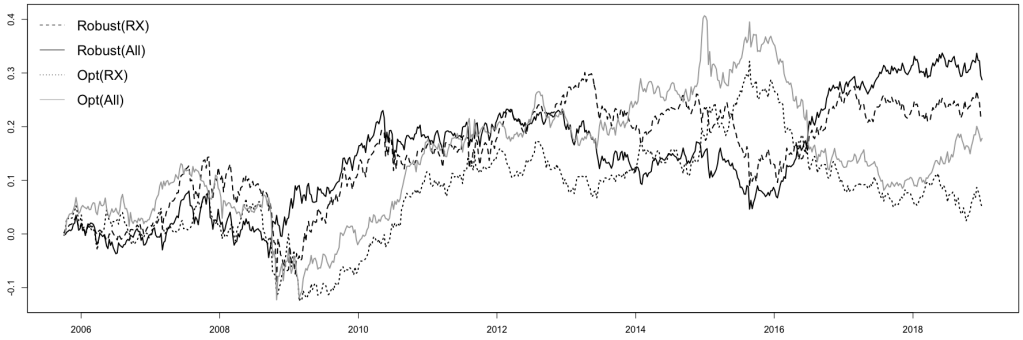

Figure 9: Cumulative compounded LSTM portfolio returns

The figure demonstrates time-series of the cumulative compounded LSTM portfolio returns with a weekly observation frequency. The top (bottom) figure displays the returns of the LSTM portfolios using the RX factor (all factors) in the outperformance probability predictions. Furthermore, two portfolio construction techniques are utilized, from which opt (robust) denotes the optimization based (robust) approach. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

the carry or momentum portfolios. However, the robust portfolio returns seem to follow the momentum portfolio patterns more closely until the end of 2012. After this period, the robust portfolio returns are more closely in line with the carry trade returns, however, they first decline slightly more but then step into a positive trend along with the carry portfolio.

Figure 10: LSTM (All) optimization based and robust, carry and 3-month momentum cumulative compounded portfolio returns

The figure displays time-series of cumulative compounded LSTM (All) optimization based and robust, carry and 3-month momentum cumulative compounded portfolio returns with a weekly observation frequency. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

The findings reported in Table 15 indicate that the LSTM portfolios utilizing all factors are not exposed to the same risk factors than the carry trade returns. Furthermore, only a small fraction of the variation is explained by the set of risk factors proposed in the literature. However, the Robust(All) portfolio manages

to capture the best performing periods of both carry and momentum portfolios displayed in Figure 10. Consequently, the sources of the profitability of the LSTM portfolios are analysed through another explanatory analysis in which the returns on carry and five different momentum strategies are used as regressors. The univariate OLS regression coefficient signs, their statistical significance and R^2s are reported in Table 16.

Table 16: Explanatory analysis of the LSTM portfolios

The table reports the univariate OLS regression coefficient signs, statistical significance of the coefficients and R^2s. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variables are listed on the row label and the explanatory variables of the regressions are listed in the column label. The portfolios on the row label are built based on the LSTM RNN model predictions. RX denotes that only the excess currency returns are used as predictive factors, whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. Robust (Opt) denotes that the robust (optimization based) portfolio construction technique is used. Furthermore, the carry and momentum portfolios are constructed with the robust portfolio construction technique and the numbers after Mom indicate how many months are included in the computation of a momentum signal. The regressor observations are not lagged. Finally, the sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

	Carry	Mom1	Mom3	Mom6	Mom12	Mom13
Robust (RX)	$+^{***}$	-	$-^{*}$	-	-	-
R^2	0.758	0.026	0.065	0.024	0.019	0.011
Robust (ALL)	$+^{***}$	+	+	$+^{**}$	$+^{***}$	$+^{***}$
R^2	0.330	-0.001	-0.001	0.024	0.043	0.047
Opt (RX)	-	-	+	+	+	+
R^2	0.001	0.001	-0.001	0.000	0.012	0.019
Opt (ALL)	+	-	-	-	+	+
R^2	0.007	0.017	0.001	0.005	-0.001	-0.001

The carry trade returns explain over 75% of the return variation of the Robust(RX) portfolio according to the high value of the R^2 in Table 16. This is in line with the similar risk factor exposures reported in Table 15 and in the analysis thereafter. Furthermore, the performance measures of these strategies reported in Table 14 were similar. Consequently, it seems that the LSTM model recognizes the carry trade signals even without having any information on the forward premiums except the implicit premiums in the historical excess currency returns.

The carry trade returns capture 33% of the return variation of the Robust(All) strategy. Furthermore, long-term momentum returns explain some of the return variation as well. However, 3-month momentum is not capturing any variation in Robust(All) portfolio returns even though it delivered a similar performance during the financial crisis as visible in Figure 10. Finally, the return variations of the optimized portfolios remain almost completely unexplained after the explanatory

analyses with the common risk factors and with the carry and the five momentum portfolios.

The two best performing portfolios: Robust(All) and carry are utilizing similar trading patterns based on the interpretation of the cumulative returns in Figure 10 and the factor analysis that reports the R^2 of 33% in Table 15. However, their performances differ substantially during the financial crisis. In fact, it seems that the difference is due to the use of all factors in the LSTM model predictions, since the LSTM model using only the historical excess currency returns is suffering from the financial crisis alike the carry portfolio. Consequently, the factor set that drives the profitability of the carry trade, might enable the LSTM model to exploit profitable patterns during the financial crisis as the market is more sensitive to these risk factors as interpreted earlier in the sample descriptions and as argued by Mancini et al. (2013) and Brunnermeier et al. (2009). In order to confirm this hypothesis, the same explanatory analyses as reported in Tables 15 and 16 are conducted with a sample period starting from December 2007 until the end of June 2009 as this period is classified as a recession by NBER (2010).

Table 17: In-sample OLS regression analysis with the latent factors during NBER recession

The table reports the OLS regression coefficient signs, statistical significance of the coefficients, adjusted R^2s and p-values of the F-test. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variables are listed on the row label and the regressors are listed in the column label. The first four portfolios on the row label are built based on the LSTM RNN model predictions. RX denotes that only the excess currency returns are used as predictive factors whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. Robust (Opt) denotes that the robust (optimization based) portfolio construction technique is used. Furthermore, the carry and 3-month momentum (Mom3) portfolios are constructed with the robust portfolio construction technique. The regressors are the first four principal components derived from the common factor observation sample by conducting a singular value composition plus a dummy variable FC_{dummy} which has a value of one between December 2007 and June 2009 and a value of zero otherwise. This period is classified as a recession by NBER (2010). The regressor observations are not lagged. Finally, the sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

	α	PR	RA	CP	FL	R^2	p-value
Robust (RX)	$+^{*}$	-	$-^{***}$	$+^{***}$	-	0.421	0.000
Robust (ALL)	$-^{**}$	$+^{***}$	+	+	$+^{**}$	-0.001	0.424
Opt (RX)	+	-	$-^{***}$	+	$-^{*}$	0.081	0.031
Opt (ALL)	$+^{**}$	-	$-^{**}$	+	$-^{***}$	0.144	0.003
Carry	+	-	$-^{***}$	$+^{***}$	-	0.360	0.000
Mom3	$-^{**}$	$+^{**}$	$+^{***}$	$-^{*}$	+	0.230	0.000

Table 17 reports the OLS regression coefficient signs, statistical significance of the coefficients, adjusted R^2s and p-values of the F-test. The regressors are

the first four principal components derived from the common factor observation sample by conducting a singular value composition. The sample spans the NBER recession period between December 2007 and June 2009. The table reports higher R^2s to carry, Robust(RX), Opt(All) and 3-month momentum than in Table 15. Consequently, the returns of the common systematic sources of currency trading are more driven by the set of risk factors during market turmoil as concluded by Mancini et al. (2013) and Brunnermeier et al. (2009). However, liquidity does not explain the return variation of the carry trade which contradicts with the evidence provided by Mancini et al. (2013) and Brunnermeier et al. (2009). Bakshi and Panayotov (2013), on the other hand, argue that the explanatory power of liquidity proxies decreases significantly if commodity price returns and currency market volatility are included into the analysis of the carry trade returns. In fact, commodity price returns are a highly statistically significant explanatory factor in Table 17 for the carry and Robust(RX) portfolios along with the risk aversion factor that is statistically significant for all except the Robust(All) portfolio returns. Finally, price uncertainty factor is statistically significant only for the 3-month momentum and Robust(All) portfolio returns.

The R^2s differ considerably between Tables 15 and 17. During the NBER recession, carry and momentum portfolio returns are more driven by the latent factors since the R^2s increase from 21% and 4% during the whole sample period to 36% and 23% during the recession. However, the risk factors are not capturing any variation of the Robust(All) portfolio returns during the recession even though they captured 4% of the variation during the whole sample period. Consequently, the Robust(All) portfolio is actually less exposed to the risk factors during the market turmoil, which contradicts with the return dynamics of the carry and 3-month momentum returns.

The findings reported in Table 17 indicate that the Robust(All) portfolio is not exposed to the risk factors during the market turmoil, whereas the carry and 3-month momentum portfolios are more exposed to the risk factors during this period. Consequently, the sources of the profitability of the LSTM portfolios are analysed through an explanatory analysis in which the returns on carry and five different momentum strategies are used as regressors during the NBER recession period. The univariate OLS regression coefficient signs, their statistical significance and R^2s are reported in Table 18.

According to the results reported in Table 18, Robust(RX) portfolio returns are not only explained by the carry trade returns but also by the momentum returns. However, their coefficients have opposite signs. This might be due to a strong negative correlation of the carry and momentum returns during this period which is visible in Figure 11 that displays the cumulative excess returns of the

Table 18: Explanatory analysis of the LSTM portfolios during NBER recession

The table reports the univariate OLS regression coefficient signs, statistical significance of the coefficients and R^2s. The statistical significance is tested with a t-test that is based on the HAC standard error estimates by Newey and West (1987) with lags chosen based on Andrews (1993). ***, ** and * denote statistical significance at 99%, 95% and 90% confidence levels, respectively. The response variables are listed on the row label and the explanatory variables of the regressions are listed in the column label. The portfolios on the row label are built based on the LSTM RNN model predictions. RX denotes that only the excess currency returns are used as predictive factors whereas All denotes that the excess currency returns, forward premiums, VIX, σ^{FX}, VP, XVP, TED, CS and COM are used as predictive factors. Robust (Opt) denotes that the robust (optimization based) portfolio construction technique is used. Furthermore, the carry and momentum portfolios are constructed with the robust portfolio construction technique and the numbers after Mom indicate how many months are included in the computation of a momentum signal. The regressor observations are not lagged. Finally, the sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

	Carry	Mom1	Mom3	Mom6	Mom12	Mom13
Robust (RX)	$+^{***}$	-	$-^{***}$	$-^{***}$	$-^{***}$	$-^{***}$
R^2	0.694	0.127	0.451	0.612	0.431	0.388
Robust (ALL)	+	$+^{*}$	+	+	+	+
R^2	0.002	0.045	-0.002	-0.012	0.010	0.007
Opt (RX)	+	$-^{*}$	-	-	-	-
R^2	0.006	0.033	0.026	0.003	-0.009	-0.012
Opt (ALL)	+	$-^{**}$	$-^{**}$	$-^{**}$	$-^{*}$	-
R^2	0.109	0.173	0.237	0.191	0.126	0.103

LSTM(All), carry and 3-month momentum portfolios during the NBER recession. The same interpretation applies to the returns of the Opt(All) portfolio that has similar but lower exposures to the carry and momentum portfolios during this period.

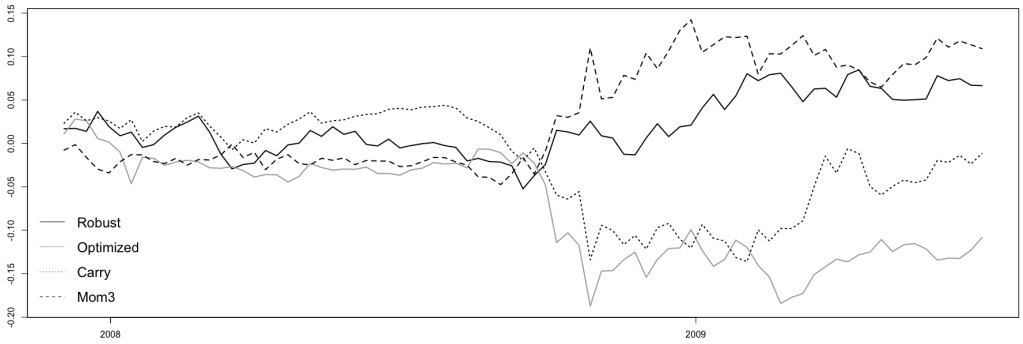

Figure 11: LSTM (All) optimization based and robust, carry and 3-month momentum cumulative compounded portfolio returns during NBER recession

The figure displays time-series of the cumulative compounded LSTM(All) optimization based and robust, carry and 3-month momentum cumulative compounded portfolio returns with a weekly observation frequency. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

The Robust(All) portfolio is driven by the 1-month momentum signals during the recession and is the only profitable LSTM portfolio during this period as visible in Figure 12 that demonstrates the cumulative excess returns of the LSTM portfolios during the NBER recession. However, the Robust(All) strategy utilizes the carry and long-term momentum signals during normal market conditions. The LSTM model that uses all factors seems to be able to recognize patterns that predict the performances of carry and momentum strategies.

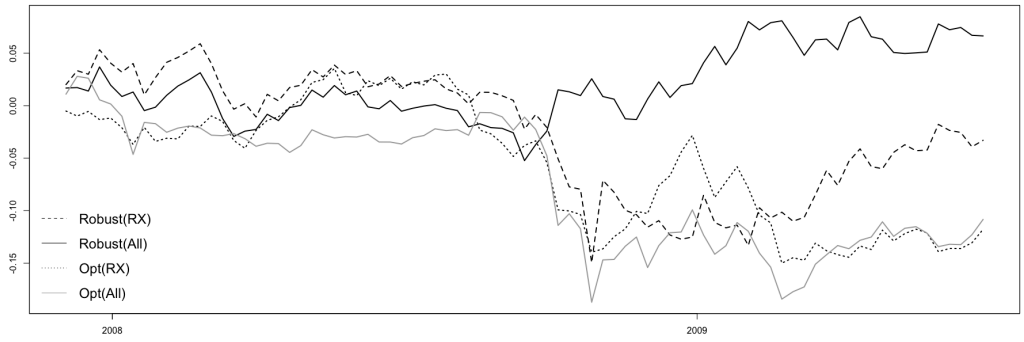

Figure 12: Cumulative compounded LSTM portfolio returns during NBER recession

The figure displays time-series of the cumulative compounded LSTM portfolio returns with a weekly observation frequency. The top (bottom) figure displays the returns of the LSTM portfolios using the RX factor (all factors) in the outperformance probability predictions. Furthermore, two portfolio construction techniques are utilized, from which opt (robust) denotes the optimization based (robust) approach. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

The LSTM model which predictions are based on all factors, recognizes the most profitable trading patterns which realizes in higher returns and lower exposures to common risk factors in currency trading strategies. Furthermore, the LSTM model utilizing historical excess currency returns is able to recognize the carry trade signals even without explicitly offered forward premiums. However, after all predictive factors are added, the LSTM model recognizes signals that predict not only carry trade and momentum signals but also their profitability. In particular, based on these predictions, the robust strategy utilizes carry and long-term momentum signals during calm market conditions, whereas it shifts to the short-term momentum signals as uncertainty picks-up.

The choice of the portfolio construction technique seems to make a bigger impact on the allocation choices than the difference between the predictive signals from the LSTM(All) and LSTM(RX) models. The optimization based portfolio construction technique offers high returns during calmer periods. However, the optimized portfolios have a high crash risk which is measured by steep drawdowns. This might be due to the high number of parameters that need to be estimated and

their estimation errors which are most likely higher after regime shifts. Entropy pooling does not offer stable enough estimates, and as Boudoukh et al. (2018) state, more robust estimates are needed. The estimation routine of the portfolio optimization inputs should be more adaptive, which could be done through introducing regime shift models. Overall, utilizing more robust optimization based portfolio construction technique seems attractive since a 50-50 combination portfolio of the Robust(All) and Opt(All) allocations delivers a Sharpe ratio of 0.4, cutting the volatility to half with a similar return when compared to the characteristics of the carry portfolio.

5.4 Application to FX risk management

In this section, the results of the foreign currency management of an international equity portfolio are reported. Only the LSTM portfolios based on all factors and the carry trade portfolio are included since the rest of the strategies were less attractive from the mean–variance point of view. All strategies are run out-of-sample and the sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

Table 19 reports descriptive statistics of the international equity portfolio returns. The annualized mean return of the unhedged equity portfolio (EQU) is improved by fully hedging the currency exposures as the fully hedged portfolio (EQ) provides slightly higher returns. Furthermore, the risk measures of EQ are more attractive than the risk measures of EQU since the former outperforms the latter in terms of lower volatility and tail risk.

Table 19: Summary statistics of the international equity portfolio returns

The table reports mean, standard deviation and Sharpe ratios of international equity portfolio returns with a weekly observation frequency. These figures are annualized. Furthermore, excess kurtosis, skewness, sample 95% value-at-risk and maximum drawdown of the international equity portfolio returns are reported. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018. Finally, the currency portfolio returns do not include transaction cost adjustments.

	Mean	St.dev.	Sharpe	Skewness	Kurtosis	VaR_{95}	Max D.d.
EQU	0.054	0.184	0.293	-0.700	6.392	-0.042	-0.607
EQ	0.059	0.172	0.343	-0.705	5.573	-0.038	-0.573
Minvar	0.035	0.110	0.318	-1.046	3.887	-0.026	-0.479
MPMVO(Carry)	0.074	0.175	0.423	0.000	7.341	-0.036	-0.535
MPMVO(Robust)	0.070	0.138	0.507	-0.840	3.048	-0.033	-0.437
MPMVO(Opt)	0.046	0.136	0.338	-0.765	9.874	-0.031	-0.552

The results reported in Table 19 indicate that the performance of an international equity portfolio can be improved by actively managing the currency exposures. First, the Minvar portfolio that aims to minimize the overall variance, reduces the risk significantly as the annualized standard deviation and weekly VaR_{95} of 11% and -2.6%, respectively, are significantly lower when compared to the EQ and EQU portfolios. However, the reduction in the risk metrics comes with a price of lower returns which is, after all, quite moderate as the Minvar portfolio delivers a Sharpe ratio similar to the Sharpe ratios of EQ and EQU.

Second, the MPMVO approach further improves the performances of the international equity portfolios from the usage of only the hedge portfolio in Minvar approach as in Boudoukh et al. (2018). In comparison to the Minvar portfolio, the MPMVO also adds a dynamic allocation to an alpha seeking currency portfolio. In the MPMVO framework, the carry portfolio delivers the highest returns with a volatility similar to the EQ and EQU portfolios. However, the MPMVO portfolio utilizing the LSTM based Robust(All) currency portfolio outperforms all other approaches when the risk measures are taken into account. The MPMVO(Robust) portfolio delivers a Sharpe ratio of 0.507 which is almost twice the size of the Sharpe ratio of the EQU portfolio. This approach has only a slight increase in volatility compared to the Minvar but it doubles the returns. The MPMVO(Robust) portfolio also has a considerably lower tail risk than the other approaches.

Even though the Robust(All) portfolio delivers higher returns than the carry portfolio, the MPMVO(Robust) portfolio has a lower return than the MPMVO(Carry) portfolio. This is due to a lower exposure to the alpha seeking currency component in the currency basket of the MPMVO(Robust) portfolio. The dynamic weight κ realizes lower values if the correlation between EQ and the alpha seeking currency portfolio decreases. Table 20 reports the correlations of the alpha seeking currency portfolios and the EQ equity portfolio. The carry portfolio has the highest correlation with the equity portfolio and, thus, has the highest exposure in the MPMVO portfolio. This translates into higher realized returns of the the MPMVO(Carry) portfolio during the sample period. On the other hand, the high exposure and equity correlation of the carry portfolio results in more volatile returns than the returns of the other MPMVO portfolios. Due to the same dynamics, the MPMVO(Opt) has low exposure to the alpha seeking currency component and, therefore, has lower returns.

Figure 13 demonstrates the cumulative compounded returns of the six international equity portfolios. The two MPMVO portfolios with the carry and the robust all factor currency allocations display close relation in return patterns. However, Figure 14, that demonstrates the cumulative compounded portfolio returns during

Table 20: Correlation between equity and alpha seeking currency portfolio returns

The table reports correlation between the international fully-hedged equity portfolio EQ and the three alpha seeking currency portfolios: Robust(All), Opt(All) and Carry. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

	Robust(All)	Opt(All)	Carry
EQ	0.406	0.095	0.68

the NBER recession period between December 2007 and June 2009, brings out the strong outperformance of the LSTM approach over the carry trade strategy during a period of market turmoil as the MPMVO(Robust) portfolio delivers even better returns than the Minvar portfolio.

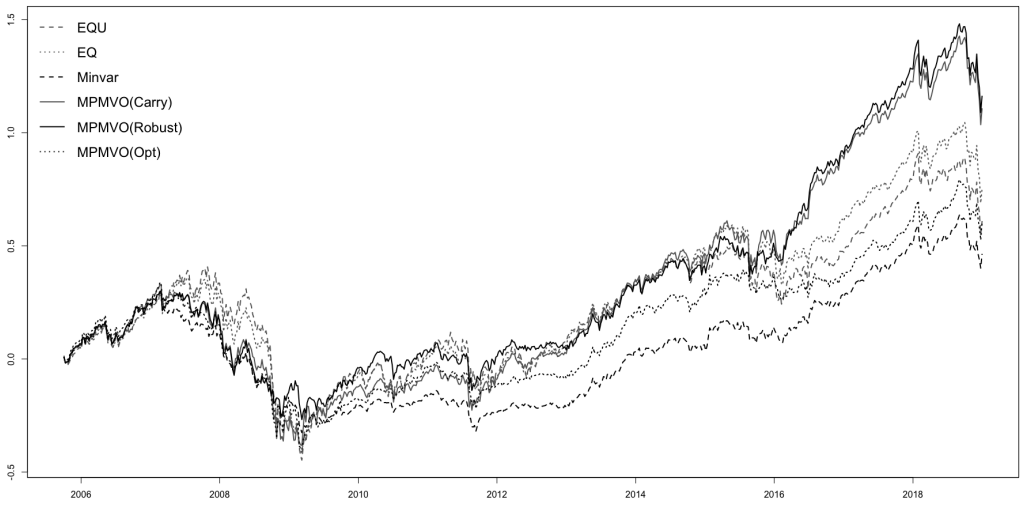

Figure 13: Cumulative compounded international equity portfolio returns

The figure displays time-series of cumulative compounded unhedged (EQU), fully-hedged (EQ), Minvar, MPMVO(Robust), MPMVO(Carry) and MPMVO(Opt) portfolio returns with a weekly observation frequency. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

By active currency management, the risk-return profile of an international equity portfolio can be improved significantly. Utilizing LSTM model predictions in the construction of the alpha seeking component of the currency basket can further improve the performance over using carry trade portfolios especially by reducing risk. The carry trade strategy is highly correlated with an international equity portfolio and driven by similar risk factors especially during market turmoil. The LSTM model, however, is capable of recognizing profitable patterns in the currency carry and momentum strategies which performances vary significantly in different market regimes. Furthermore, it is less exposed to the common

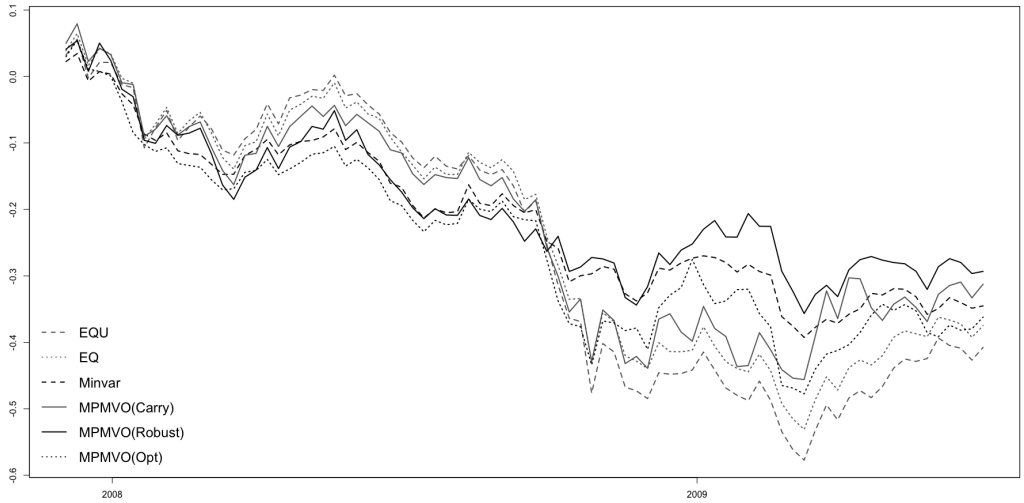

Figure 14: Cumulative compounded international equity portfolio returns during the financial crisis

The figure displays time-series of cumulative compounded unhedged (EQU), fully-hedged (EQ), Minvar, MPMVO(Robust), MPMVO(Carry) and MPMVO(Opt) portfolio returns with a weekly observation frequency. The sample covers the period between December 2007 and June 2009.

risk factors driving the returns of the most popular currency trading strategies. Consequently, the LSTM model is in great value when applied to the management of currency exposures of an international equity portfolio.

5.5 Practical aspects of implementation

This section concentrates on investigating whether the implementation of the LSTM portfolios is practical or not. First, the turnover and transaction costs of the currency portfolios are analysed. Second, the performance metrics and summary statistics of the currency portfolios are reported after imposing a leverage constraint that limits the long and short positions to one. Finally, the turnover and transaction costs of these constrained portfolios are reported.

Table 21 reports the portfolio turnover and estimated transaction costs of the currency portfolios. These figures are annualized averages computed from the sample period starting from the 3rd of October 2005 to the 31st of December 2018. The results are reported on the currency baskets only and do not include equity transactions. The carry portfolio has by far the lowest turnover whereas the Opt(All) portfolio has the highest turnover. The relatively low turnover of the carry portfolio is most likely due to the high persistence of short-term interest rates. Consequently, the ranking order of the forward premiums is changing slowly.

In contrast, the 3-month momentum portfolio has more than six times higher turnover than the carry portfolio. The high turnover is a characteristic of the momentum strategies as concluded by Menkhoff et al. (2012b). In particular, short-term momentum strategies have higher turnovers, since the momentum signals are based on rolling estimation windows and, thus, the long-term momentum signals are more persistent.

Table 21: Currency portfolio transactions

The table reports annualized turnover and transaction cost estimates of the currency portfolios that are rebalanced weekly. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018.

	Turnover	Transaction costs
Robust(All)	25.18	0.0177
Opt(All)	92.25	0.0664
Carry	5.65	0.0037
Mom3	36.52	0.0250
FX(Hedge)	12.27	0.0078
Robust(All)κ+FX(Hedge)	33.31	0.0226
Opt(All)κ+FX(Hedge)	61.35	0.0437
Carryκ+FX(Hedge)	17.51	0.0110

The turnover of the Robust(All) portfolio is between the turnover figures of the carry and 3-month momentum portfolios. This is in line with the previous conclusions as the Robust(All) utilizes both carry and momentum signals. However, the optimization based approach of the portfolio construction increases turnover significantly as demonstrated by the figures in Table 21. In fact, the turnover of the Opt(All) portfolio is approximately 3.7 times higher than the turnover of the Robust(All) portfolio. The robust portfolio construction technique limits the single currency exposures to 0.25, whereas the optimized approach limits the single currency exposures to 0.6 and the overall portfolio volatility to 6% prior to the allocation decision. Furthermore, the absolute size of the currency positions is always two for the robust portfolios, whereas it practically is not capped or floored for the optimized portfolio. Consequently, the higher turnover of the optimized portfolio is not likely to be explained by this feature since the absolute size of the currency positions can be smaller or higher. However, the optimized approach requires a vast amount of parameters to be estimated in order to pursue the allocation decision. In fact, there is not only uncertainty in the signal predictions but also in the covariance matrix and reference model estimation. Therefore, the optimized weights are more likely to fluctuate strongly and even change signs when compared to the robust portfolio approach. Moreover, the optimized portfolio construction

routine allows higher single currency exposures and, therefore, more extreme estimations have more extreme currency exposures. On the other hand, transaction costs could be included in the portfolio optimization routine and when combined with more robust estimates, this approach could lower the turnover significantly.

High turnover translates into high transaction costs. In particular, deducting transaction costs from the currency portfolio returns changes the mean excess returns of the Opt(All) and 3-month momentum portfolios from positive to negative. On the other hand, the Robust(All) portfolio delivers positive excess returns after transaction cost adjustment despite a turnover of over 25. However, deducting the transaction costs reduces the returns of the Robust(All) portfolio by 80% whereas the carry trade returns are reduced only 20% after the transaction cost adjustment. Consequently, transaction costs seem to have an important role in the profitability of the LSTM portfolios. However, rebalancing the portfolios less frequently should decrease turnover and, thus, bring down the transaction costs. In order to proceed with this hypothesis, an analysis if the LSTM model is able to recognize profitable patterns also during a longer time frame than a week should be conducted.

The FX(Hedge) currency portfolio that aims to minimize the volatility of an international equity portfolio has a relatively low turnover compared to the other optimization based approach, Opt(All). However, in the FX(Hedge) portfolio construction, only the covariance matrix and equity correlations are estimated and, thus, it has less uncertain parameters than the Opt(All). Furthermore, these parameters are most likely more persistent than the LSTM model output signals which lowers the turnover. Finally, due to the optimization based portfolio construction, the turnover of the FX(Hedge) portfolio remains higher than, for example, the turnover of the robust carry portfolio.

The currency components of the MPMVO portfolios have mixed turnover figures when compared to the Robust(All), Opt(All) and carry portfolios. The turnover figures of the carry and Robust(All) portfolios increase once they are combined with the FX(Hedge) portfolio, whereas the turnover of the Opt(All) and FX(Hedge) combination portfolio is significantly lower than the turnover of the Opt(All) portfolio. This might be explained by two reasons. First, the dynamic allocation weights, which are measured by the κ parameters, should be smaller for the LSTM portfolios due to their lower correlation with the international equity market than the highly correlated carry portfolio. In particular, Opt(All) is less correlated with the equity market than the Robust(All) portfolio which means that it has a lower average κ. Consequently, the turnover is reduced significantly, since the Opt(All) has a significantly smaller exposure in the MPMVO portfolio than the carry and Robust(All) portfolios. Second, negative correlation between the changes in single currency exposures of the FX(Hedge) and the alpha portfolios

should decrease the turnover, whereas positive correlation has an opposite effect. Consequently, the currency exposures of the carry portfolio might change in line with the changes of the FX(Hedge) portfolio or more likely be uncorrelated.

The transaction costs of the currency component of the MPMVO Opt(All) portfolio are extremely high despite the lower turnover when compared to the Opt(All) portfolio alone. However, the transaction costs of the MPMVO Carry and Robust(All) currency components are closer to each other and more modest. In particular, the MPMVO carry and Robust(All) portfolio returns are decreased by approximately 15% and 30% after deducting the transaction costs, respectively. This reduction leads to Sharpe ratios of 0.36 for the carry and 0.35 for the Robust(All) MPMVO portfolios. These Sharpe ratios are still higher than the Sharpe ratios of EQU and EQ, even though the difference is small.

Another practical consideration is related to the portfolio leverage that is measured by the sum of the absolute portfolio weights divided by two. The leverage restriction is imposed as by Boudoukh et al. (2018) and restricts the leverage of the currency portfolios to 100% or less. Table 22 reports the summary statistics of the currency portfolios after imposing the leverage constraint. Currency exposures of the robust portfolios: carry, 3-month momentum and Robust(All) are unchanged, since their leverage is always 100%. However, the leverage of the optimization based approaches: Opt(All), Minvar and MPMVO portfolios might exceed 100% and, thus, the results in the table can differ from the results in Table 19 that reports the results without the leverage restriction.

Table 22: Summary statistics of international equity and currency portfolio returns after leverage constraint

The table reports mean, standard deviation and Sharpe ratios of international equity and currency portfolio returns after leverage constraint with a weekly observation frequency. These figures are annualized. Furthermore, excess kurtosis, skewness, sample 95% value-at-risk and maximum drawdown of the international equity portfolio returns are reported. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018. Finally, the currency portfolio returns do not include transaction cost adjustments.

	Mean	St.dev.	Sharpe	Skewness	Kurtosis	VaR_{95}	Max D.d.
Robust(All)	0.022	0.068	0.324	-0.462	2.185	-0.015	-0.151
Opt(All)	0.012	0.055	0.218	-1.260	9.779	-0.011	-0.171
Carry	0.018	0.080	0.225	-0.733	6.474	-0.017	-0.197
Mom3	0.002	0.068	0.029	0.234	7.226	-0.015	-0.212
Minvar	0.044	0.116	0.379	-1.074	3.811	-0.028	-0.490
MPMVO(Carry)	0.058	0.147	0.395	-0.705	3.484	-0.033	-0.541
MPMVO(Robust)	0.067	0.141	0.475	-0.833	3.293	-0.033	-0.480
MPMVO(Opt)	0.057	0.139	0.410	-0.850	5.190	-0.033	-0.563

Table 22 reports higher Sharpe ratios for the Opt(All) and Minvar portfolios

than Table 19. After imposing the leverage restrictions, these portfolios have less extreme allocations. In fact, this brings down the volatility and reduces the tail risk of the Opt(All) portfolio significantly. In contrast, the Minvar portfolio has slightly more tail risk but higher returns. This is intuitive, since the FX(Hedge) portfolio, that is used with the international equity portfolio in the Minvar portfolio, has a negative return in the sample period and is constructed to decrease the riskiness of the equity portfolio. Consequently, smaller exposure to FX(Hedge) enhances the return but increases the riskiness of the Minvar portfolio.

The Sharpe ratios of the MPMVO portfolios are higher for the optimized approach and lower for the carry and robust approaches after imposing the leverage restriction. The higher Sharpe ratio of the MPMVO(Opt) portfolio is explained by the decreased exposure to the FX(Hedge) currency component and the reasoning for that follows the previous discussion. In contrast, MPMVO(Robust) and MPMVO(Carry) have lower Sharpe ratios. Their currency components have high exposures to the alpha portfolios and, thus, decreases the overall leverage of the currency basket bringing down the returns. However, due to a higher correlation of the carry portfolio with the equity market and, hence, higher exposure to the alpha component in the currency basket than with the MPMVO(Robust) portfolio, volatility of the MPMVO(Carry) portfolio decreases significantly as the leverage is capped to 100%. Despite the volatility reduction of the MPMVO(Carry), its mean return declines even more resulting in a lower Sharpe ratio. In fact, both LSTM based MPMVO portfolios have higher Sharpe ratios than the MPMVO(Carry) portfolio after imposing the leverage restriction.

Table 23 reports the portfolio turnover and estimated transaction costs for currency portfolios after the leverage restriction of 100%. All portfolios affected by the restriction have smaller annual turnover. The transaction costs of the Opt(All) and the combination of the Opt(All) and the FX(Hedge) portfolios are yet extremely high compared to their returns. Next, the transaction costs of the FX(Hedge) and the combinations of the Robust(All) and the carry portfolios with the FX(Hedge) portfolio are slightly lower than before applying the leverage restriction. However, due to the reductions in the returns of the two latter currency portfolios, the outperformances of the MPMVO portfolios over the EQ and EQU equity portfolios are not as impressive as prior to the transaction cost reductions and the leverage restriction.

After all, the transaction costs of the LSTM and MPMVO portfolios are so high that they can not be ignored. In particular, weak persistence of the LSTM model prediction signals and the parameter uncertainty of the optimized portfolio construction technique, resulting in high turnover and transaction costs, impose a challenge to the implementation of the LSTM portfolios in practice. However,

Table 23: Currency portfolio transactions after the leverage constraint

The table reports mean, standard deviation and Sharpe ratios of the international equity portfolio returns with a weekly observation frequency. These figures are annualized. Furthermore, excess kurtosis, skewness, sample 95% value-at-risk and maximum drawdown of the international equity portfolio returns are reported. The sample covers the period from the 3rd of October 2005 to the 31st of December 2018. Finally, the currency portfolio returns do not include transaction cost adjustments.

	Turnover	Transaction costs
Robust(All)	25.18	0.0177
Opt(All)	64.49	0.0457
Carry	5.65	0.0037
Mom3	36.52	0.0250
FX(Hedge)	10.09	0.0065
Robust(All)κ+FX(Hedge)	26.45	0.0181
Opt(All)κ+FX(Hedge)	42.72	0.0302
Carryκ+FX(Hedge)	14.03	0.0089

changing the rebalancing frequency of the portfolio from weekly to monthly or quarterly should decrease the turnover significantly. For example, Boudoukh et al. (2018) show that the turnover of the MPMVO portfolio that combines carry, momentum and reversal strategies decreases from 12.4 to 6.6 when the rebalancing frequency is changed from monthly to quarterly. Furthermore, this change has hardly any impact on the Sharpe ratio of the portfolio. This finding should be in line with the LSTM approach due to the similarities of their sources of returns.

6 CONCLUSIONS

This research investigated whether a long short-term memory (LSTM) recurrent neural network (RNN) is able to recognize profitable patterns in the G10 currency market from the 4th of January 1999 to the 31st of December 2018 or not. The model inputs are historical excess currency returns w.r.t. the USD, forward premiums and seven factors that capture price uncertainty, risk-aversion, commodity price changes and funding liquidity. Furthermore, two alternative portfolio construction techniques, robust and optimized, are applied to the LSTM model predictions and the resulting portfolios are compared to the well-researched currency carry and momentum strategies. Finally, the best performing portfolios are applied to a FX risk management framework of an international equity portfolio.

First, LSTM and simple RNN model predictions are compared. The LSTM is proven to be significantly more accurate than the simple RNN by reaching total accuracies of more than 55%, which indicates that currency excess returns are partly driven by signals with more than two months temporal distance. Second, robust and optimized portfolio construction techniques are utilized in order to exploit if the predictions of two alternative LSTM models can be used to build successful trading strategies. The LSTM model recognizes carry trade signals from the historical excess currency returns. However, after adding all predictive factors, the LSTM model is able to predict the profitability of the carry and momentum strategies. In particular, the LSTM portfolio utilizes carry and long-term momentum signals during calm market conditions and short-term momentum signals during market turmoil. The best performing LSTM portfolio delivers a Sharpe ratio of 0.32 with less tail risk than the carry portfolio that has a Sharpe ratio of 0.23.

Attractive risk-return profile and low correlation with equity markets make the LSTM portfolio extremely suitable to the FX risk management of an international equity portfolio. In fact, when the LSTM portfolio is used as an alpha seeking currency component in the modified portfolio mean-variance optimization (MPMVO) routine introduced by Boudoukh et al. (2018), the Sharpe ratio of an unhedged international equity portfolio is almost doubled as the Sharpe ratio increases from 0.29 to 0.51. In comparison, the carry portfolio manages to increase the Sharpe ratio only to 0.42.

The LSTM portfolios, that utilize all predictive factors, are not exposed to the common risk factors that are driving the currency carry and momentum returns, especially during the financial crisis. However, the choice of the portfolio construction technique has a considerable impact on the performances in market regime shifts. The optimization based approach, which translates the LSTM model predic-

tions into mean-variance optimization inputs through entropy pooling, is sensitive to parameter uncertainty, whereas the robust technique relying on heuristics is more stable. However, their combination portfolio delivers an impressive Sharpe ratio of 0.4. This suggests that more robust optimization based approaches might be well-suited for this task.

After all, the substantial benefits might not fully materialize to investors trying to implement the LSTM strategies, investigated in this research, in practise. The restriction that limits the portfolio leverage to 100% has only limited impact on the results, whereas high turnover and transaction costs impose a challenge to the LSTM portfolios as a major part of the excess returns are cut off by the trading costs. However, less frequent rebalancing frequency should decrease the transaction costs significantly and result in a feasible trading strategy in practise. Exploiting lower prediction and rebalancing frequencies is left for future research.

REFERENCES

Abadi, M. – Agarwal, A. – Barham, P. – Brevdo, E. – Chen, Z. – Citro, C. – Corrado, G. S. – Davis, A. – Dean, J. – Devin, M. – Ghemawat, S. – Goodfellow, I. – Harp, A. – Irving, G. – Isard, M. – Jia, Y. – Jozefowicz, R. – Kaiser, L. – Kudlur, M. – Levenberg, J. – Mané, D. – Monga, R. – Moore, S. – Murray, D. – Olah, C. – Schuster, M. – Shlens, J. – Steiner, B. – Sutskever, I. – Talwar, K. – Tucker, P. – Vanhoucke, V. – Vasudevan, V. – Viégas, F. – Vinyals, O. – Warden, P. – Wattenberg, M. – Wicke, M. – Yu, Y. – Zheng, X. (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. <http://tensorflow.org/>, retrieved 16.6.2019.

Amihud, Y. (2002) Illiquidity and stock returns: cross-section and time-series effects. *The Journal of Financial Markets*, Vol. 5 (1), 31–56.

Andrews, D. W. K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica*, Vol. 61 (4), 821–856.

Asness, C. S. – Moskowitz, T. J. – Pedersen, L. H. (2013) Value and momentum everywhere. *The Journal of Finance*, Vol. 68 (3), 929–985.

Avramov, D. – Chordia, T. – Jostova, G. – Philipov, A. (2007) Momentum and credit rating. *The Journal of Finance*, Vol. 62 (5), 2503–2520.

Bakshi, G. – Panayotov, G. (2013) Predictability of currency carry trades and asset pricing implications. *Journal of Financial Economics*, Vol. 110 (1), 139–163.

Barroso, P. – Santa-Clara, P. (2015) Beyond the carry trade: Optimal currency portfolios. *Journal of Financial and Quantitative Analysis*, Vol. 50 (5), 1037–1056.

Bengio, Y. – Simard, P. – Frasconi, P. (1994) Learning long-term dependencies with gradient descent is difficult. *Neural Networks*, Vol. 5 (2), 157–166.

Bhattacharya, D. – Kumar, R. – Sonaer, G. (2012) Momentum loses its momentum: Implications for market efficiency. SSRN Electronic Journal. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1762264>, retrieved 16.6.2019.

Black, F. (1989) Universal hedging: Optimizing currency risk and reward in international equity portfolios. *Financial Analysts Journal*, Vol. 45 (4), 16–22.

Boudoukh, J. – Richardson, M. P. – Thapar, A. K. – Wang, F. (2018) Optimal currency hedging for international equity portfolios. SSRN Electronic Journal, <https://ssrn.com/abstract=3265403>, retrieved 16.6.2019.

Brunnermeier, M. K. – Nagel, S. – Pedersen, L. H. (2009) Carry trades and currency crashes. *NBER Macroeconomics Annual*, Vol. 23 (1), 313–347.

Campbell, J. Y. – Medeiros, K. S.-D. – Viceira, L. M. (2010) Global currency hedging. *The Journal of Finance*, Vol. 65 (1), 87–121.

Carhart, M. M. (1997) On persistence in mutual fund performance. *The Journal of Finance*, Vol. 52 (1), 57–82.

Chollet, F. et al. (2015) Keras. <https://keras.io/>, retrieved 16.6.2019.

Daniel, K. – Moskowitz, T. J. (2016) Momentum crashes. *Journal of Financial Economics*, Vol. 122 (2), 221–247.

Davidsson, M. (2011) Portfolio theory and cone programming. *Journal of Applied Finance & Banking*, Vol. 1 (2), 173–187.

Diebold, F. X. – Mariano, R. S. (1995) Comparing predictive accuracy. *Journal of Business & Economic Statistics*, Vol. 13 (3), 253–263.

Fama, E. (1965) The behavior of stock-market prices. *The Journal of Business*, Vol. 38 (1), 34–105.

Fama, E. (1984) Forward and spot exchance rates. *Journal of Monetary Economics*, Vol. 14, 319–338.

Filippou, I. – Taylor, M. P. (2017) Common macro factors and currency premia. *Journal of Financial and Quantitative Analysis*, Vol. 52 (4), 1731–1763.

Fischer, T. – Krauss, C. (2018) Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, Vol. 270 (2), 654–669.

Froot, K. A. (1993) Currency hedging over long horizons. *NBER Working Papers 4355*, <http://www.nber.org/papers/w4355>, retrieved 16.6.2019.

Gebhardt, W. R. – Hvidkjaer, S. – Swaminathan, B. (2005) Stock and bond market interaction: Does momentum spill over? *Journal of Financial Economics*, Vol. 75 (3), 651–690.

Gers, F. A. – Schmidhuber, J. – Cummins, F. (2000) Learning to forget: Continual prediction with lstm. *Neural Computation*, Vol. 12, 2451–2471.

Gers, F. A. – Schraudolph, N. N. – Schmidhuber, J. (2002) Learning to forget: Continual prediction with lstm. *Journal of Machine Learning Research*, Vol. 3, 115–143.

Glen, J. – Jorion, P. (1993) Currency hedging for international portfolios. *The Journal of Finance*, Vol. 48 (5), 1865–1886.

Goodfellow, I. – Bengio, Y. – Courville, A. (2016) *Deep Learning*. MIT Press, <http://www.deeplearningbook.org/>, retrieved 16.6.2019.

Graves, A. – Mohamed, A. – Hinton, G. E. (2013) Speech recognition with deep recurrent neural networks. Cornell University Library, arXiv.org, Ithaca.

Gu, S. – Kelly, B. – Xiu, D. (2018) Empirical asset pricing via machine learning. NBER Working Papers 25398, <https://ideas.repec.org/p/nbr/nberwo/25398.html>, retrieved 16.6.2019.

Heaton, J. B. – Polson, N. G. – Witte, J. H. (2018) Deep learning in finance. Cornell University Library, arXiv.org, Ithaca.

Hinton, G. (2012) Neural networks for machine learning: Lecture 6e.

Hochreiter, S. – Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, Vol. 9 (8), 1735–1780.

Jacobs, H. (2015) What explains the dynamics of 100 anomalies? *Journal of Banking & Finance*, Vol. 57, 65–85.

Jegadeesh, N. – Titman, S. (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, Vol. 48 (1), 65–91.

Jegadeesh, N. – Titman, S. (2001) Profitability of momentum strategies: an evaluation of alternative explanations. *The Journal of Finance*, Vol. 56 (2), 699–720.

Jorion, P. (1994) Mean/variance analysis of currency overlays. *Financial Analysts Journal*, Vol. 50 (3), 48–56.

Jostova, G. – Nikolova, S. – Philipov, A. – Stahel, C. W. (2013) Momentum in corporate bond returns. *The Review of Financial Studies*, Vol. 26 (7), 1649–1693.

Jurek, J. W. (2014) Crash-neutral currency carry trades. *Journal of Financial Economics*, Vol. 113 (3), 325–347.

Krauss, C. – Do, X. A. – Huck, N. (2017) Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, Vol. 259 (2), 689–702.

Londono, J. M. – Zhou, H. (2017) Variance risk premiums and the forward premium puzzle. *Journal of Financial Economics*, Vol. 124 (2), 415–440.

Lustig, H. – Roussanov, N. – Verdelhan, A. (2011) Common risk factors in currency markets. *Review of Financial Studies*, Vol. 24 (11), 3731–3777.

Lustig, H. – Roussanov, N. – Verdelhan, A. (2014) Countercyclical currency risk premia. *Journal of Financial Economics*, Vol. 111 (3), 527–553.

Mancini, L. – Ranaldo, A. – Wrampelmeyer, J. (2013) Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance*, Vol. 68 (5), 1805–1841.

Menkhoff, L. – Sarno, L. – Schmeling, M. – Schrimpf, A. (2012a) Carry trades and global foreign exchange volatility. *The Journal of Finance*, Vol. 67 (2), 681–718.

Menkhoff, L. – Sarno, L. – Schmeling, M. – Schrimpf, A. (2012b) Currency momentum strategies. *Journal of Financial Economics*, Vol. 106 (3), 660–684.

Menkhoff, L. – Sarno, L. – Schmeling, M. – Schrimpf, A. (2017) Currency value. *The Review of Financial Studies*, Vol. 30 (2), 416–441.

Merton, R. C. (1974) On the pricing of corporate debt: the risk structure of interest rates. *The Journal of Finance*, Vol. 29 (2), 449–470.

Meucci, A. (2005) *Risk and Asset Allocation*. Springer, Berlin.

Meucci, A. (2008) Fully flexible views: Theory and practice. *Risk*, Vol. 21 (10), 97–102.

Meucci, A. – Ardia, D. – Colasante, M. (2014) Portfolio construction and systematic trading with factor entropy pooling. *Risk*, 56–61.

Moskowitz, T. J. – Ooi, Y. H. – Pedersen, L. H. (2012) Time series momentum. *Journal of Financial Economics*, Vol. 104 (2), 228–250.

MSCI (2019) Msci world index. <https://www.msci.com/documents/10199/178e6643-6ae6-47b9-82be-e1fc565ededb>, retrieved 16.6.2019.

NBER (2010) Us business cycle expansions and contractions. <https://www.nber.org/cycles.html>, retrieved 16.6.2019.

Newey, W. K. – West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, Vol. 55 (3), 703–708.

Orlov, V. (2016) On persistence in mutual fund performance. *The Journal of Banking and Finance*, Vol. 67, 1–11.

Python Software Foundation (2019) Python 3.6.8 documentation. <https://docs.python.org/3.6/>, retrieved 16.6.2019.

Schmidhuber, J. (2015) Deep learning in neural networks: An overview. *Neural Networks*, Vol. 61, 85–117.

Schmittmann, J. M. (2010) Currency hedging for international portfolios. *IMF Working Paper*, Vol. 10 (151), 1–44.

Srivastava, N. – Hinton, G. – Krizhevsky, A. – Sutskever, I. – Salakhutdinov, R. (2014) Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol. 15, 1929–1958.

Topaloglou, N. – Vladimirou, H. – Stavros, Z. A. (2011) Optimizing international portfolios with options and forwards. *Journal of Banking and Finance*, Vol. 35, 3188–3201.

Yang, X. – Zhang, H. (2019) Extreme absolute strength of stocks and performance of momentum strategies. *Journal of Financial Markets*, Vol. In Press.