Aalto University
School of Business

YIELD CURVE ARBITRAGE IN EUR SWAP RATES

A Hybrid Neural Network Approach

Master's Thesis
Lauri Stark
Aalto University School of Business
Finance
Spring 2020

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Lauri Stark	
Title of thesis Yield Curve Arbitrage in EUR Swap Rates: A Hybrid Neural Network Approach	
Degree Master of Science	
Degree programme Finance	
Thesis advisor(s) Matti Suominen	
Year of approval 2020	**Number of pages** 61 **Language** English

Abstract

In this thesis, I analyze the out-of-sample trading performance of a yield curve arbitrage strategy on EUR swap rates. For the implementation of the strategy, I introduce a novel hybrid neural network approach which uses the factors of an affine term structure model as inputs. I compare the results to the performance of a benchmark strategy that is based on the traditional two-factor Vasicek term structure model. The results imply that with reasonable transaction costs, the neural network model produces significant multifactor alpha, positively skewed returns with high kurtosis and a higher Sharpe ratio and higher absolute cumulative performance compared to the Vasicek model. However, the neural network strategy also has exposure to systematic risk factors and tail risk.

Keywords yield curve, machine learning, neural network, interest rate, arbitrage

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Lauri Stark

Työn nimi Hybridineuroverkon käyttö korkokäyrän arbitraasistrategiassa euroalueen vaihtokoroilla

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Matti Suominen

Hyväksymisvuosi 2020 **Sivumäärä** 61 **Kieli** Englanti

Tiivistelmä

Tässä tutkielmassa analysoin euroalueen vaihtokoroilla toteutettavaa, korkokäyräarbitraasiin pohjautuvaa sijoitustrategiaa jossa mallintaminen toteutetaan hybridineuroverkolla. Vertaan saamiani tuloksia perinteiseen kahden faktorin Vasicekin malliin pohjautuvaan strategiaan. Saamani tulokset osoittavat, että kohtuullisilla kaupankäyntikustannuksilla neuroverkkomalli tuottaa tilastollisesti merkittävää ylituottoa sekä korkeamman Sharpe-luvun ja kokonaistuoton verrattuna perinteiseen Vasicekin malliin. Neuroverkkomalliin pohjautuva strategia sisältää kuitenkin useaan systemaattiseen tuottofaktoriin liittyvää riskiä, sekä lisäksi häntäriskiä.

Avainsanat korkokäyrä, koneoppiminen, neuroverkko, sijoittaminen, arbitraasi

Yield Curve Arbitrage in EUR Swap Rates: A Hybrid Neural Network Approach

Lauri Stark

Master's Thesis

Department of Finance

Aalto University School of Business

Thesis advisor: Professor Matti Suominen

April 27, 2020

Abstract

In this thesis, I analyze the out-of-sample trading performance of a yield curve arbitrage strategy on EUR swap rates. For the implementation of the strategy, I introduce a novel hybrid neural network approach which uses the factors of an affine term structure model as inputs. I compare the results to the performance of a benchmark strategy that is based on the traditional two-factor Vasicek term structure model. The results imply that with reasonable transaction costs, the neural network model produces significant multifactor alpha, positively skewed returns with high kurtosis and a higher Sharpe ratio and higher absolute cumulative performance compared to the Vasicek model. However, the neural network strategy also has exposure to systematic risk factors and tail risk.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background and Motivation

Historically, fixed income arbitrage strategies such as yield curve arbitrage, swap spread arbitrage and volatility arbitrage have been associated as typical hedge fund strategies. The purpose of such arbitrage strategies is to recognize underpriced or overpriced fixed income instruments and trade the mispricings in a market-neutral way. In this paper, I analyze a yield curve arbitrage strategy on EUR swap curve. The idea is to use a neural network model that identifies the rich or cheap points on the swap curve and provides hedge measures for the mispriced points on the curve in order to make the total position market neutral. Especially, the inputs to the neural network model are the calibrated factor values of the two-factor Vasicek model. Using these factors as inputs for the neural network enables the calculation of the hedging measures so that the strategy is market neutral to the Vasicek risk factors. To my current knowledge, this kind of hybrid model which combines a term structure model with a machine learning framework has not been developed previously.

According to Duarte et al. (2006), fixed income arbitrage strategies have become very popular among the hedge fund industry after the beginning of the 21st century. This is despite the fact that the very same strategies also played a major role in the hedge fund crisis of 1998, during which the notorious collapse of the large hedge fund Long Term Capital took place. The aforementioned facts characterize fixed income arbitrage rather well: most of the time, these strategies may earn small positive returns but then as rare tail events occur, these returns are easily wiped out and the strategies suffer heavy losses. However, based on the results of Duarte et al. (2006), yield curve arbitrage has historically produced positively skewed returns with significant alpha. Despite the attractive nature of its returns, yield curve arbitrage has been explicitly studied by only a few previous papers, namely by Duarte et al. (2006) and Karsimus (2015).

Generally, a yield curve arbitrage strategy is a market-neutral strategy that seeks to profit by exploiting pricing inefficiencies between related fixed-income securities while neutralizing exposure to interest rate risk. In the previous studies of yield curve arbitrage, the modelling frameworks of the interest rate dynamics have been based on classical short-rate models.

Duarte et al. (2006) used the two-factor version of the Vasicek (1977) term-structure model and Karsimus (2015) used the two-factor version of Cox et al. (1985) and the Longstaff and Schwartz (1992) two-factor model. As with any model, also these models have limitations regarding the model assumptions, the dynamics of the model-implied interest rates and calibration of the model parameters. This is why a machine learning framework for modelling the interest rate is presented in this paper. Machine learning models have their limitations for example regarding to overfitting, but the assumptions behind the models are less stringent and the models learn the interest rate dynamics from the data. For a reader who is not acquainted with machine learning models, Goodfellow et al. (2016) provide a comprehensive, self-contained guide for the topic.

Because of the increased computational capacity and availability of data, machine learning methods have had increasing popularity in applications to finance-related problems during the recent years. Among the most recent ones related to interest rate modelling, Kirczenow et al. (2018) apply a machine learning model called Denoising Autoencoder to extract features of the yield curves of illiquid corporate bonds. In turn, Sambasivan and Das (2017) use Gaussian Process to forecast the US constant-maturity yield curve based on daily data. However, to my current knowledge no study has applied machine learning methods to modelling a swap curve.

The structure of the rest of the paper is as follows. First, I present the hypotheses. Second, I go through previous research on arbitrage, applications of term structure models and applications of machine learning models in asset pricing. Third, a separate chapter is devoted to the most essential term-structure models in chronological order. After this, I introduce the data set and go through the data cleaning procedures. In the following chapter, I present the specifications of the neural network model, the swap curve modelling methodology, the trading methodology and the valuation methodology for the swap contracts. The penultimate chapter presents and analyses the results and the final chapter concludes.

1.2 Hypotheses

In this paper I assess the empirical out-of-sample performance of a neural network model on a yield curve arbitrage strategy on EUR swap rates and compare the results to the performance of a benchmark model, namely the two-factor Vasicek model. The analysis is based on the following five hypotheses.

> **Hypothesis 1**: *The neural network model will produce a higher Sharpe ratio and a higher gain-loss ratio compared to the Vasicek model.*
>
> **Hypothesis 2**: *The neural network model produces positively skewed returns with high kurtosis.*
>
> **Hypothesis 3**: *The neural network model produces multifactor alpha with respect to systematic risk factors.*
>
> **Hypothesis 4**: *The neural network model has only minor exposures to systematic risk factors.*
>
> **Hypothesis 5**: *The neural network model has moderate to high tail risk with respect to some of the systematic risk factors.*

The first hypothesis is motivated by the following. First, machine learning models are rather data intensive and use state-of-the-art optimization algorithms which could make them able to learn the underlying dynamics of the interest rates better than the traditional short-rate models. Second, applications of machine learning models in options pricing have shown rather promising results in pricing accuracy: neural network models have generally been superior to traditional parametric options pricing models such as the Black-Scholes model (see e.g. Hutchinson et al. (1994), Anders et al. (1998), Amilon (2003), Bennell and Sutcliffe (2004), Stark (2017)). Thus, a neural network model could perform better also in yield curve modelling compared to traditional short-rate models.

The second hypothesis is motivated by the fact that fixed income arbitrage strategies are very typical hedge fund strategies which are supposed to be market neutral. Because of the assumption about market neutrality, the strategies are expected to produce positive returns regardless of the market environment. Thus, the second hypothesis states that the return distribution is assumed to be positively skewed with a fatter right tail. The third and the fourth hypothesis are based on the assumption that the yield curve arbitrage strategy produces abnormal positive returns and that the returns cannot be explained by the returns of systematic risk factors. The final hypothesis is related to the tail risk of the yield curve arbitrage strategy with respect to systematic risk factors. As stated before, based on historical events it seems that fixed income arbitrage strategies might contain significant amounts of tail risk.

2 Related Literature

This section presents and summarizes previous research on hedge fund arbitrage strategies, interest rate modelling and applications of machine learning models in finance. The section is divided as follows. First, I go through previous research on the characteristics of arbitrage strategies with a focus on hedge funds. Second, I present papers on practical applications of iterest rate models. Finally, I will present literature about applying machine learning models in the context of asset pricing. After the literature review chapter, a separate section is devoted for term structure models.

2.1 Risk and Return in Arbitrage Strategies

As mentioned before, fixed income arbitrage strategies, such as yield curve arbitrage, are considered to be very typical strategies for hedge funds. One of the most extensive studies about these kinds of strategies is done by Duarte et al. (2006). In their work, the authors provide an extensive analysis about the risk and return characteristics of the most widely used fixed income arbitrage strategies: swap spread arbitrage, yield curve arbitrage, mortgage arbitrage, volatility arbitrage and finally capital structure arbitrage. In the paper, the authors hypothesise whether fixed income arbitrage strategies are truly arbitrage, or if they are compensation for carrying tail risk as in "picking nickels in front of the steamroller". The results imply that all of the strategies produce positive excess returns on average and also that most of the returns are positively skewed. The authors also analyze the amount of market risk contained in the strategies and find that after adjusting for bond and equity factors, the strategies that contain most "intellectual capital", namely yield curve arbitrage, mortgage arbitrage and capital structure arbitrage, generate significant positive alphas. Especially, the strategies that generate significant alpha tend to be based on complex modelling frameworks which are used to detect mispricings and to make the strategies market neutral by using different hedging methodologies. Even after introducing hedge fund fees, some of the strategies produce significant positive alpha. However, the results also indicate that many of the strategies that are supposed to be market-neutral actually have exposure to systematic risk factors. Despite of this, the authors find very little evidence that the fixed income arbitrage strategies were merely compensation for carrying tail risk.

Focusing solely on capital structure arbitrage, Yu (2006) finds that capital structure arbitrage is indeed a rather attractive investment strategy. The idea of the strategy is to exploit the pricing difference between a company's debt and equity by recognizing cheap or rich credit default swap (CDS) spreads. In the work of Duarte et al. (2006), capital structure arbitrage is one of the more complex strategies which tend to produce significant positive alpha. In capital structure arbitrage, the complexity stems from the model that is used to detect mispriced credit default swaps. However, based on the results of Yu (2006), capital structure arbitrage seems to contain significant amount of risk. The risk is especially related to events when the CDS spread is shorted and the market spread skyrockets simultaneously. In these situations, the arbitrageur is forced to exit the position with large losses. Indeed, the maximum monthly losses of the strategy can be as large as -33% even though the strategy yields a very attractive Sharpe ratio of 1.54.

Focusing more on risk, Fung and Hsieh (2002) analyze the common sources of risk and return in fixed income hedge fund strategies. The results imply that the main common source of risk in such strategies is related to changes in interest rate spreads and also options on interest rate spreads. Especially, fixed income hedge fund strategies seem to be typically exposed to credit spreads. Okunev et al. (2006) develop nonlinear risk factors for analyzing the tail risk in fixed income hedge fund strategies. Based on the findings, it seems that the most significant risk factor in fixed income arbitrage strategies is a factor that is similar to being short on put options on high-yield bonds. More generally, using nonlinear risk factors seems to give increased estimates of the tail risk levels of hedge fund strategies. A similar framework that utilizes nonlinear risk factors is presented by Jawadi and Khanniche (2012). The motivation behind using nonlinear risk factors is to catch the asymmetric relationship between the risk and return in hedge fund strategies and also to be able to model the time varying nature of the exposure of hedge fund strategies to risk factors. The results indicate that the hedge fund risk factor exposures indeed vary over time and that hedge fund returns exhibit nonlinearity and asymmetry. Kelly and Jiang (2012) analyze the tail risk in hedge fund returns and conclude that hedge fund returns exhibit consistent exposures to extreme downside risk. Adrian et al. (2011) employ quantile regressions to assess the tail risk in different hedge fund strategies with respect to selected tail risk factors, such as a short-term liquidity spread, USD carry trade excess returns and a credit spread. Based on their results, it seems that correlations between different hedge fund styles increase in the tails which also increases the probability of simultaneous losses across hedge fund styles. Liu et al. (2002) study the effect of liquidity risk and default risk on the market price of interest rate swaps. Their results indicate that the credit premium in swaps is mostly compensation for liquidity risk, and that the liquidity premium in turn increases with maturity. On the other hand, the term structure of the default premium is basically flat, but both the liquidity premium and the default premium are time-varying. Also regarding market neutrality, Patton (2009) finds that about one quarter the hedge funds in their sample have significant exposures to market risk and thus are not market neutral in reality. Asness et al. (2001) show that hedge funds tend to load tail risk in their strategies in order to boost their alpha-creation.

2.1.1 Limits to Arbitrage

Pure, textbook arbitrage requires no capital and carries no risk. Limits to arbitrage address the issues of implementing pure arbitrage in reality as such real-world arbitrage strategies require capital and are exposed to risks. Shleifer and Vishny (1997) state that arbitrage becomes ineffective in extreme market environments. The authors also state the possibility that many anomalies in the financial markets in fact stem from the volatility avoidance of arbitrageurs rather than hidden macroeconomic risks. When analyzing investment policies for theoretical arbitrage opportunities, Liu and Longstaff (2004) find that it is often actually most optimal for an investor to underinvest in an arbitrage opportunity rather than taking as big position as the margin requirements allow. Sometimes, it might even be so that it is optimal for an investor to walk entirely away from an arbitrage opportunity. Even if the optimal investment policy is followed, the simulation-based results of Liu and Longstaff (2004) imply that the arbitrage strategies often underperform the riskless asset and result in low Sharpe ratios. Also, the analyzed arbitrage strategies tend to experience losses before convergence. Regarding arbitrage trading of hedge funds, Siegmann and Stefanova (2009) analyze how liqudity affects the market neutrality of equity-based arbitrage strategies over time. The results indicate that during times of low liquidity, market neutrality decreases due to the difficulties of maintaining dynamic hedging strategies that are required for market neutrality. Mitchell and Pulvino (2012) analyze the effect of the 2008 financial crisis to hedge funds and arbitrage strategies. The outcome of the crisis was that leverage funding decreased, which forced hedge funds, i.e. the arbitrageurs, to liquidate their positions which in turn increased the mispricings between similar securities in the market. Fontaine and Nolin (2019) employ a relative value measure to analyze the limits to arbitrage in fixed income markets. The results are as expected: limits to arbitrage increase as funding decreases.

Generally, it seems that even though arbitrage opportunities can be recognized rather frequently, it is usually the case that market conditions prevent the exploitation of such arbitrage opportunities on a constant basis. Especially, limitations in liquidity and leverage funding seem to be the most essential barriers to exploiting the full potential of recognized arbitrage opportunities.

2.2 Applications of Term Structure Models

Applications of term structure models have been popular topics in the academic literature, especially regarding the analysis of swap curves and government yield curves such as the US Treasury yield curve. Duffie and Singleton (1997) use a two-factor affine term structure model, namely the two-factor Cox-Ingersoll-Ross model, for the purpose of modelling the US swap curve with counterparty credit risk and liquidity risk. Their results imply that both the liquidity factor and the credit factor were significant sources of variation in the swap rates. Liu et al. (2006) perform a similar study and find that the credit premium priced into swap rates is primarily compensation for liquidity risk. However, both liquidity and default premia vary significantly over time. Contrary to affine term structure models, Adrian et al. (2013) develop a linear regression model and apply the model for the US yield curve. A model that combines a term structure model with macroeconomic variables, namely inflation and economic growth rates, was developed by Ang and Piazzesi (2003). Diebold et al. (2006) performed a similar study in which the authors include real activity, inflation and monetary policy as the macroeconomic variables. In both studies, the authors find strong evidence that the macroeconomic variables have effects on the future evolution of the yield curve. Especially, the results of Ang and Piazzesi (2003) imply that 85% of the variation in the yield curve is explained by the macroeconomic factors. Mönch (2008) applies similar methodology for forecasting US government bond yields with a model that adds an affine term structure model and a factor-augmented vector autoregression. The model uses about 160 variables that contain data about industrial production, employment, price indices, monetary aggregates, survey data and stock indices, to mention a few. The results imply that the model outperforms several benchmark models for out-of-sample forecasts. Grinblatt (2001) develops a model that provides a closed form solution for swap spreads by employing one-factor term structure models for modelling the liquidity of government securities and short term borrowings.

Focusing more on the term structure models itself, Duffee (2002) introduces a class of term structure models that the authors label as "essentially affine models". The main idea in the paper is that standard affine term structure models are unable to reproduce the well-known failure of the expectations hypothesis which states that long-term interest rates are determined purely by current and future expected short-term rates. Dai and Singleton (2002) are able to model the key characteristics of the expectations puzzle with affine and quadratic term structure models. Dai and Singleton (2003) introduce an extensive survey in which the authors analyze the theoretical and empirical properties of several dynamic term structure models with empirical goodness-of-fit tests. Backus et al. (2001) develop two-currency versions of affine term structure models in order to characterise the so-called forward premium anomaly, which means that high-yielding currencies tend to appreciate over time. However, the authors state that the models have serious shortcomings in simultaneously producing the forward premium anomaly and reproducing the fundamental properties of currencies and interest rates. Collin-Dufresne and Solnik (2001) use a Vasicek-based framework to model the default risk in the swap term structure in order to explain the relation between corporate bond yields and the swap curve, i.e. the LIBOR-swap spread. Durham (2006) uses a three-factor Gaussian term structure model to estimate the inflation risk premium both in the nominal Treasury yield curve and the inflation-linked TIPS yield curve. Hördahl and Tristani (2012) perform similar analysis for Euro-area data by using a linear macro model combined with an affine term structure model.

2.3 Machine Learning Models in Asset Pricing

2.3.1 Machine Learning in Options Pricing

One of the most popular topics combining machine learning frameworks and financial instruments is option pricing with machine learning. Previous studies have focused on how machine learning models are capable of learning to price options and also on the hedging performance of the models. Usually, the studies use market prices of options as a target and then compare the performance of machine learning models to the Black-Scholes model.

In one of the first studies incorporating neural network methodology in options pricing, Malliaris and Salchenberger (1993) compare the pricing performance of a neural network model and the Black-Scholes model on S&P100 options. Another study was performed by Hutchinson et al. (1994) where the authors compare the pricing and also the hedging performance between the Black-Scholes model and a neural network model on S&P500 index options. Stark (2017), the author of this thesis, performs a similar study based on the methodology of Hutchinson et al. (1994) with more recent data on DAX30 index options. Similar studies have been also performed by Garcia and Gençay (2000) and Bennell and Sutcliffe (2004). At the highest level, the results of the aforementioned studies show evidence that the neural network models are superior to the Black-Scholes model. Amilon (2003) compares the pricing and hedging performances of a neural network model with European-style calls on OMX Stockholm index with implicit and historical volatilities. The results indicate that the neural network was superior in both cases. Andreou et al. (2008) formulate a hybrid framework in which parametric option pricing models (the Black-Scholes the and the Corrado-Su model) are combined with a neural network. The results indicate that these hybrid models outperform both the parametric models and the non-hybrid neural network models. Yao et al. (2000) focus solely on analyzing the pricing performance of a neural network model. The results indicate that neural networks are capable of pricing options more accurately during volatile times compared to the Black-Scholes model. Overall, the neural network models have shown to often outperform traditional options pricing models both in terms of pricing accuracy and hedging performance.

2.3.2 Machine Learning and Interest Rates

Only a few studies have studied interest rates with machine learning frameworks such as neural networks. Sambasivan and Das (2017) compare the performance of Gaussian Process regression, Nelson-Siegel model and a vector-autoregression model in forecasting US treasury yields. The findings indicate that the Gaussian Process is superior for short term maturities below one year and that the vector-autoregression model performs better for longer maturities starting from two years and up to 30 years. The performance of the Nelson-Siegel model is low especially for the longer tenors. Kirczenow et al. (2018) apply an autoencoder to price missing yields for illiquid corporate markets.

2.3.3 Other Machine Learning Applications in Asset Pricing

Apart from option pricing and interest rates, machine learning models have been used on a wide domain of asset pricing problems ranging from consumer credit risk analysis to stock return prediction. Gu et al. (2018) perform a wide comparison between different machine learning models in predicting returns of U.S. stocks. The conclusion from the results is that neural networks are the best performing machine learning method for this domain. In another study related to prediction, Malliaris and Salchenberger (1996) use neural networks to predict S&P100 implied volatility. Heaton et al. (2016) use a deep autoencoder for portfolio selection in Markowitz framework. Khandani et al. (2010) use generalized classification and regression trees to model consumer credit risk. In a similar study, Butaru et al. (2016) use decision trees, random forests and regularized logistic regression. Decision trees are also applied to analyze the drivers behind gold returns by Malliaris and Malliaris (2015). Commodities are also analyzed in Malliaris and Malliaris (2009) where the authors model the interdependence between oil, gold and the euro with a neural network model. Sirignano et al. (2016) and Kvamme et al. (2018) apply neural networks for modelling different risks in mortgage loans. Regarding traditional asset pricing and factor investing, Gu et al. (2019) formulate an asset pricing model based on an autoencoder with latent factors. A simple long-short strategy based on the model yields a Sharpe of 1.53 on monthly data, which is superior to the compared models including the Fama-French factor models. Moritz and Zimmermann (2016) use a tree-based method for portfolio sorting. Borovykh et al. (2017) use a convolutional neural network to predict S&P500 returns. Luss and d'Aspremont (2015) use text classification to predict abnormal returns from news.

As can be seen, machine learning models have been applied to a variety of problems related to asset pricing. In many of the aforementioned studies, the results imply that machine learning models perform rather well in several fields of asset pricing. It seems that in many cases, machine learning models can be considered as an alternative for traditional, well-established asset pricing models.

3 Overview of Term-Structure Models

In the following sections, I will briefly review the well-known traditional term-structure models. The mathematical notation and results in the sections are based on the book by Brigo and Mercurio (2007). The purpose of presenting the main properties of these models is to give the reader a perspective on the difference between the traditional term-structure models and machine learning models in yield curve modelling. Among the models presented in the following chapters, the benchmark studies for this thesis apply the two-factor Gaussian model (23) in Duarte et al. (2006) and the two-factor Cox-Ingersoll-Ross model (27) in Karsimus (2015).

3.1 Terminology in Term-Structure Models

The key idea behind term-structure models is to model the so-called *short rate*, usually denoted by r. The short-rate is a theoretical rate that represents the amount interest that is accumulated in an infinitesimally short amount of time. Under term-structure models, the short-rate follows a stochastic process specified by the model. A very closely related object to the short-rate is the zero-coupon bond price, usually denoted by P. The zero-coupon bond price at time t for tenor T represents the value at time t of one unit of currency that will be received at time T. When the process for the short-rate is specified, a zero-coupon bond price can be determined by the short rate. Based on no-arbitrage arguments, the zero-coupon bond is specified in terms of the short-rate under the risk-neutral measure as

$$P(t,T) = \mathbb{E}\Big[\exp\Big(-\int_t^T r_s ds\Big)\Big|\mathcal{F}_t\Big]. \tag{1}$$

Here $\mathcal{F}_t$ denotes the filtration of the short-rate process which can be thought as the information available up to time t. Based on the zero-coupon bond price, a corresponding continuously compounded spot rate (i.e. zero rate) R can be computed as

$$R(t,T) = -\frac{\ln P(t,T)}{\tau(t,T)}, \tag{2}$$

where $\tau(t,T)$ denotes the amount of time in years between times T and t. The zero-coupon bond price $P(t,T)$ can be thought as a discount factor for cash flows taking place at time T and the spot rate $R(t,T)$ is the corresponding discount rate.

Generally when it comes to different term-structure models, one is interested especially in properties such as the distribution of the short-rate r and the resulting zero-coupon bond price P. These aspects determine the analytical tractability of the models, for example if a model has a closed-form solution for the zero-coupon price. The following sections demonstrate these properties for several term-structure models.

3.2 One-factor term-structure models

3.2.1 Vasicek (1977)

The pioneering approach for interest rate modelling was proposed by Vasicek (1977). Under the risk neutral measure, the Vasicek model defines the dynamics of the short-rate r as

$$dr(t) = k[\theta - r(t)]dt + \sigma dW(t), r(0) = r_0 \tag{3}$$

where k, θ, σ and $r(0)$ are positive constants. This stochastic differential equation is an Ornstein-Uhlenbeck process, which is mean reverting. The parameter k denotes the speed of mean reversion, θ denotes the long-term mean of the process and σ denotes the volatility of the process. W denotes standard Brownian motion under the risk neutral measure. Because the parameters k, θ and σ are assumed to be constant, the model is time-homogeneous.

Historically, it has been a common view that the main drawback of the Vasicek model is that under its assumptions, we have a positive probability for negative short-rates. This was an unreasonable feature for an interest rate model before the global financial crisis of 2008. However, in the prevailing ultra-low interest rate environment where for example the short end of the EUR swap curve has been negative since year 2015, such model feature is not necessarily a disadvantage. Another drawback is that the one-factor Vasicek model cannot produce an inverted yield curve with any parameter values. In addition, the model is endogenous in a sense that its initial term structure might not match exactly the term structure observed in the market.

However, the model has also very attractive properties. First, the stochastic differential equation is linear and can be solved explicitly. Second, the distribution of the short rate is

Gaussian. Integrating the stochastic differential equation (3), we get for any $0 \leq s < t$

$$r(t) = r(s)e^{-k(t-s)} + \theta\left(1 - e^{-k(t-s)}\right) + \sigma \int_s^t e^{-k(t-u)} dW(u) \tag{4}$$

Based on (4), we can compute the zero-coupon bond price in closed form with (1):

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}, \tag{5}$$

where

$$A(t,T) = \exp\left\{\left(\theta - \frac{\sigma^2}{2k^2}\right)[B(t,T) - T + t] - \frac{\sigma^2}{4k}B(t,T)^2\right\}$$
$$B(t,T) = \frac{1}{k}\left[1 - e^{-k(T-t)}\right].$$

Thus, the Vasicek model has a rather straightforward formula for the zero-coupon price which makes the model transparent and analytically tractable.

3.2.2 Dothan (1978)

The dynamics of the short rate under the Dothan model (Dothan (1978)) are defined as

$$dr(t) = ar(t)dt + \sigma r(t)dW(t). \tag{6}$$

Integrating (6) yields

$$r(t) = r(s)\exp\left\{\left(a - \frac{1}{2}\sigma^2\right)(t-s) + \sigma(W(t) - W(s))\right\} \tag{7}$$

Thus, r is lognormally distributed. Because of the lognormal distribution, r is always positive for each t and thus the Dothan model overcomes the problem of negative rates of the Vasicek model (3). The Dothan model also has a closed form solution for zero-coupon bonds, but as noted by Brigo and Mercurio (2007), the solution is rather complex and includes a double integral which requires numerical integration and thus it is not computationally very efficient to evaluate.

The lognormality of r incorporates another shortcoming: theoretically, an arbitrary small time step can produce an infinite amount of money when modelling the evolution of a bank account with one unit of currency. This problem is related to all lognormal short-rate models.

3.2.3 Cox-Ingersoll-Ross (1985)

The Cox-Ingersoll-Ross (CIR) model proposed by Cox et al. (1985) uses a square-root diffusion process for the short-rate which prevents negative rates. The model is formulated as follows:

$$dr(t) = k[\theta - r(t)]dt + \sigma\sqrt{r(t)}dW(t), r(0) = r_0, \tag{8}$$

where k, θ, σ and r_0 are positive constants with the condition

$$2k\theta > \sigma^2 \tag{9}$$

With this formulation, r is always positive. (9) ensures that the process never hits the origin, and thus it is guaranteed that r is positive for all t. Also, instead of normal distribution the CIR process in (8) incorporates a noncentral χ^2-distribution. The CIR model has a closed form solution for zero-coupon bond price, given by

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}, \tag{10}$$

where

$$A(t,T) = \left[\frac{2h\exp\{\frac{(k+h)(T-t)}{2}\}}{2h + (k+h)(\exp\{(T-t)h\} - 1)}\right]^{\frac{2k\theta}{\sigma^2}},$$

$$B(t,T) = \frac{2(\exp\{(T-t)h\} - 1)}{2h + (k+h)(\exp\{(T-t)h\} - 1)}, \tag{11}$$

$$h = \sqrt{k^2 + 2\sigma^2}.$$

Thus, once the parameters of the model are known, the zero coupon bond price is straightforward to compute as with the Vasicek model (3).

3.2.4 Hull-White (1990)

The Hull-White model (Hull and White (1990)) is built on the Vasicek model (3) with the addition that the model parameters are now time-varying. This is why the Hull-White model is sometimes referred to as the Hull-White extended Vasicek model. Because of the time-varying parameters, the Hull-White model is able to fit the currently observed yield curve in the market. The dynamics of the model are formulated as follows:

$$dr(t) = [\vartheta(t) - a(t)r(t)]dt + \sigma(t)dW(t), \tag{12}$$

where ϑ, a and σ are deterministic functions of time. With this formulation, the model can be made to fit the initial term structure and the term structure of forward or spot-rate volatilities when pricing caps, floors or other interest rate derivatives with optionality. However, it is generally known that the future volatility structures implied by the model can be rather unrealistic compared to the ones observed in the market. This is why a simpler version of the model is usually used when focusing on the modelling of the term structure of short rates. The simpler model is essentially the same as (12) but with only one time-varying parameter:

$$dr(t) = [\vartheta(t) - ar(t)]dt + \sigma dW(t), \tag{13}$$

where ϑ is still a deterministic function of time but a and σ are now positive constants. From (13), we get

$$r(t) = r(s)e^{-a(t-s)} + \alpha(t) - \alpha(s)e^{-a(t-s)} + \sigma \int_s^t e^{-a(t-u)}dW(u), \tag{14}$$

where

$$\alpha(t) = f^M(0,t) + \frac{\sigma^2}{2a^2}(1 - e^{-at})^2. \tag{15}$$

Here $f^M(0,T)$ denotes the instantaneous market forward rate for maturity T at time 0. Based on (14), we know that r is normally distributed. As with the Vasicek model (3), also the Hull-White model has a positive probability for negative rates.

As with the models presented in earlier sections, the zero coupon bond price under the dynamics of the Hull-White model can also be solved explicitly since we know that r has a Gaussian distribution. Indeed, by using the definition of the zero-coupon bond price in (1), it can be shown that the integral $\int_t^T r(u)du$ also has a Gaussian distribution and that the zero-coupon bond price is given by

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}, \tag{16}$$

where

$$A(t,T) = \frac{P^M(0,T)}{P^M(0,t)} \exp\left\{B(t,T)f^M(0,t) - \frac{\sigma^2}{4a}(1-e^{-2at})B(t,T)^2\right\}.$$
$$B(t,T) = \frac{1}{a}\left[1 - e^{-a(T-t)}\right].$$

Here, $P^M(0,t)$ denotes the currently observed market price of a zero-coupon bond with maturity t.

3.2.5 Hull-White extension of the CIR model

The Hull-White methodology of time-varying parameters can also be applied to the CIR model. The dynamics of the short-rate are then

$$dr(t) = [\vartheta(t) - a(t)r(t)]dt + \sigma(t)\sqrt{r(t)}dW(t), \tag{17}$$

where ϑ, a and σ are again deterministic functions of time. However, with this specification, the zero-coupon bond price P does not have a closed-form solution and has to be evaluated numerically. Even when simplifying the model by using constant values for the parameters a and σ and allowing ϑ to be the only time-dependent parameter, a closed form solution for ϑ is not available. ϑ can be solved numerically, but this does not guarantee that r is positive which is required for the diffusion term $\sigma\sqrt{r(t)}dW(t)$ to be well defined. That is, there is a clear tradeoff: Gaussian, Vasicek-based models have nice properties when it comes to analytical tractability, but they also have the drawback of producing negative rates. The CIR model with time-varying parameters guarantees that the rates are positive, but analytical tractability is lost. To overcome this tradeoff, a Gaussian model that addresses the problem of negative rates was proposed by Black and Karasinski (1991).

3.2.6 Black-Karasinski (1991)

The Black-Karasinski model is a lognormal model with the dynamics of the short-rate defined as

$$d\ln(r(t)) = [\theta(t) - a(t)\ln(r(t))]dt + \sigma(t)dW(t), r(0) = r_0 \tag{18}$$

Because of the lognormality, the short-rate r is guaranteed to be positive. However, as generally with lognormal models, a closed form solution for the zero-coupon bond price

P is not available with this model. Also, the Black-Karasinski model has the problem of exploding bank account similar to the Dothan model (6) which is generally a problem with lognormal models. Also, another shortcoming stated by Brigo and Mercurio (2007) is that modelling forward rates with the Black-Karasinski model is rather heavy computationally since a trinomial tree with Monte-Carlo simulations is required for the process.

3.2.7 CIR with general deterministic shift extension

With shift extensions, the short-rate is generally defined as

$$r_t = x_t + \varphi(t),\ t \geq 0, \tag{19}$$

where x is a diffusion process and φ is a deterministic function of time. The Hull-White extended Vasicek model (12) is basically equivalent to using a deterministic shift for the Vasicek model (3). However, shift extension is especially useful with the CIR model (8). Remember that incorporating time-varying parameters in the CIR model (i.e. the Hull-White extension for CIR model (17)) resulted in a loss in analytical tractability. When using a deterministic shift for the CIR model, we get the desirable properties related to having time-varying parameters (i.e. the ability to match the initial term-structure exactly) and also analytical solutions for zero-coupon bond prices. Thus, we get a model that produces positive rates, matches the initial yield curve and is analytically tractable.

The dynamics of the shift-extended CIR model are defined as

$$\begin{aligned} dx(t) &= k[\theta - x(t)]dt + \sigma\sqrt{x(t)}dW(t), x(0) = x_0 \\ r(t) &= x(t) + \varphi(t), \end{aligned} \tag{20}$$

where x_0, k, θ and σ are positive constants with the condition $2k\theta > \sigma^2$ that ensures that the origin is not accessible. Under the model, the shift φ^{CIR} and instantaneous forward rate f^{CIR} become

$$\begin{aligned} \varphi^{CIR}(t;\alpha) &= f^M(0,t) - f^{CIR}(0,t;\alpha), \\ f^{CIR}(0,t;\alpha) &= \frac{2k\theta(\exp\{th\} - 1)}{2h + (k+h)(\exp\{th\} - 1)} + x_0 \frac{4h^2 \exp\{th\}}{[2h + (k+h)(\exp\{th\} - 1)]^2}, \end{aligned} \tag{21}$$

where $h = \sqrt{k^2 + 2\sigma^2}$ and f^M denotes the market-implied instantaneous forward rate. The price of a zero-coupon bond is then given by

$$P(t,T) = \bar{A}(t,T) \exp -B(t,T)r(t),$$

where

$$\bar{A}(t,T) = \frac{P^M(0,T)A(0,t)\exp\{-B(0,t)x_0\}}{P^M(0,t)A(0,T)\exp\{-B(0,T)x_0\}} A(t,T)e^{B(t,T)\varphi^{CIR}(t;\alpha)} \tag{22}$$

As before, P^M denotes the market-implied zero coupon bond price, i.e. the market discount factor.

3.3 Two-Factor Short-Rate Models

The main drawback of one-factor short-rate models is that under the one-factor dynamics, the short end and the long end of the yield curve (e.g. the 1-month rate and the 10-year rate) are perfectly correlated. This means that the yield curve moves in parallel shifts which is an unrealistic assumption in real world. Two-factor short-rate models try to overcome this problem by introducing more subtle correlation structures within the yield curve.

3.3.1 The Two-Factor Vasicek Model

The dynamics of the instantaneous short rate are defined as

$$r(t) = x(t) + y(t) + \varphi(t), r(0) = r_0, \tag{23}$$

where the dynamics of the factors x and y are in turn defined as

$$\begin{aligned} dx(t) &= -ax(t)dt + \sigma dW_1(t) \\ dy(t) &= -by(t)dt + \eta dW_2(t) \end{aligned} \tag{24}$$

with $x(0) = y(0) = 0$ Here, (W_1, W_2) is a two-dimensional Brownian motion with correlation ρ, that is

$$dW_1(t)dW_2(t) = \rho dt$$

and r_0, a, b, σ and η are positive constants. Integrating (23) yields

$$r(t) = x(s)e^{-a(t-s)} + y(s)e^{-b(t-s)} + \sigma \int_s^t e^{-a(t-u)} dW_1(u) + \eta \int_s^t e^{-b(t-u)} dW_2(u) + \varphi(t) \tag{25}$$

so that the short-rate r is again normally distributed. The zero-coupon price is then given by

$$
\begin{aligned}
P(t,T) &= \frac{P^M(0,T)}{P^M(0,t)}\exp\{\mathcal{A}(t,T)\}, \\
\mathcal{A}(t,T) &:= \frac{1}{2}\Big[V(t,T) - V(0,T) + V(0,t)\Big] - \frac{1-e^{-a(T-t)}}{a}x(t) - \frac{1-e^{-b(T-t)}}{b}y(t),
\end{aligned}
\tag{26}
$$

where $P^M(0,t)$ denotes the currently observed market price of a zero-coupon bond with maturity t. Note that usually we only observe the prices $P(0,t)$ for some fixed maturities such as 1 month, 3 months, 6 months, 1 year and so on, so in order to get a smooth mapping $T \mapsto P(0,T)$, $T > 0$ we must interpolate between the observed market prices. In the above formula, $V(t,T)$ denotes the variance of the Gaussian random variable I defined as

$$
I(t,T) = \int_t^T \Big[x(u) + y(u)\Big]du.
$$

3.3.2 The Two-Factor Cox-Ingersoll-Ross Model

The dynamics of the two-factor CIR model are analogous to the two-factor Vasicek model (23). The only difference is that now the processes x and y follow a square-root diffusion as with the one-factor CIR model (8). The two-factor CIR model is defined as

$$
\begin{aligned}
r(t) &= x(t) + y(t) + \varphi(t) \\
dx(t) &= k_1[\theta_1 - x(t)]dt + \sigma_1 dW_1(t) \\
dy(t) &= k_2[\theta_2 - x(t)]dt + \sigma_2 dW_2(t),
\end{aligned}
\tag{27}
$$

where k_1, θ_1, σ_1, k_2, θ_2, σ_2 are positive constants again with the conditions $2k_1\theta_1 > \sigma_1^2$ and $2k_2\theta_2 > \sigma_2^2$ and W_1 and W_2 are independent Brownian motions.

For analysing the model, it is convenient to first look at a two-factor CIR model without the shift $\varphi(t)$. This simplified model is labelled as CIR2. Without the shift, the zero-coupon price is

$$
P^{CIR2}(t,T;x(t),y(t),\alpha) = P^{CIR}(t,T;x(t),k_1,\theta_1,\sigma_1)P^{CIR}(t,T;y(t),k_2,\theta_2,\sigma_2), \tag{28}
$$

where $\alpha = (k_1, \theta_1, \sigma_1, k_2, \theta_2, \sigma_2)$ is the parameter vector and P^{CIR} denotes the zero-coupon bond price under the one-factor CIR model (8). Thus, the spot rate is given by

$$
R^{CIR2}(t,T;x(t),y(t),\alpha) = R^{CIR}(t,T;x(t),k_1,\theta_1,\sigma_1) + R^{CIR}(t,T;y(t),k_2,\theta_2,\sigma_2), \tag{29}
$$

where R^{CIR} denotes the spot rate for the one-factor CIR model. The zero-coupon bond price for the 2-factor CIR model with shift is then given as

$$P^{CIR2++}(t,T;x(t),y(t),\alpha) = \exp\left[-\int_t^T \varphi(s;\alpha)ds\right] P^{CIR2}(t,T;x(t),y(t),\alpha), \qquad (30)$$

where

$$\exp\left[-\int_t^T \varphi(s;\alpha)ds\right] = \exp\left\{\left[R^{CIR2}(0,T;\alpha) - R^M(0,T)\right]T - \left[R^{CIR2}(0,t;\alpha) - R^M(0,t)\right]t\right\}.$$

Thus, despite the additional factor, the calculations behind the two-factor models can generally be tracked back to the corresponding one-factor models.

Clearly, the advantage of traditional term-structure models presented in this chapter is the analytical tractability of the models and the known dynamics of the short rate process. However, one issue with term-structure models is the choice of the variables to which the model is calibrated[1]. For example, one could choose to calibrate a chosen term-structure model to interest rate caps, floors or swaptions. The performance of the model is naturally affected by this choice. Also, some of the calibration methods use tree structures such as binomial trees, which are rather intensive computationally.

[1]Calibration of term structure models is not covered in this thesis.

4 Data

The data I use in my study contain daily closing prices for three interest rate curves: the EUR swap curve for full-year tenors between 1 year and 10 years, the Euribor curve ranging between tenors of 1 week and 12 months, and the EUR overnight indexed swap (OIS) curve ranging from the overnight rate to 10-year rate. All data are acquired from Bloomberg Terminal. For the EUR swap curve, the data contain observations from January 3, 2000 until February 25, 2019. For the Euribor and OIS curves, the data contain observations for the out-of-sample testing period ranging from January 4, 2010 to February 25, 2019 where both of the curves are used in the valuation of the swap trades. The reason for having the longer data set for the EUR swap curve is that the observations before January 4, 2010 are used as a training dataset for the neural network model.

Data processing goes as follows. First, all the dates for which the EUR swap curve has missing values in any tenor are dropped. This results in dropping 47 dates from the EUR swap dataset. Second, the training dataset is constructed by selecting observations before the beginning of year 2010. After this, all the dates for which any of the endpoints of the Euribor curve or the OIS curve are missing, are dropped. For the OIS curve data, this results in dropping 53 dates. For the Euribor curve data, none of the dates are dropped. Finally, the test dataset is generated by merging the three cleaned datasets with an inner join such that the merged dataset has observations for dates that are present in all of the three separate datasets. Table 1 shows the summary statistics for the EUR swap data and Figure 1 shows the historical plots. An essential observation is that on October 22, 2015 the 1-year swap rate entered negative territory. After this, all tenors except the longer tenors of 8, 9 and 10 years have experienced negative values.

Table 1: Descriptive statistics for the EUR swap rates. The data contain observations ranging from January 3, 2000 until February 25, 2019. N denotes the number of observations. 25%, 50% and 75% are the corresponding percentiles. All values are in percentages.

	N	Mean	Std	Min	25%	50%	75%	Max
1-year swap rate	4947	1.95	1.75	-0.27	0.32	1.83	3.45	5.48
2-year swap rate	4947	2.08	1.76	-0.26	0.38	2.06	3.72	5.58
3-year swap rate	4947	2.24	1.76	-0.25	0.47	2.34	3.83	5.64
4-year swap rate	4947	2.40	1.76	-0.23	0.61	2.60	3.89	5.69
5-year swap rate	4947	2.55	1.74	-0.18	0.78	2.80	3.98	5.78
6-year swap rate	4947	2.68	1.72	-0.11	0.96	2.98	4.07	5.81
7-year swap rate	4947	2.81	1.71	-0.03	1.13	3.14	4.14	5.88
8-year swap rate	4947	2.92	1.69	0.06	1.30	3.26	4.24	5.94
9-year swap rate	4947	3.02	1.67	0.16	1.44	3.37	4.32	5.98
10-year swap rate	4947	3.11	1.65	0.24	1.58	3.46	4.40	6.02

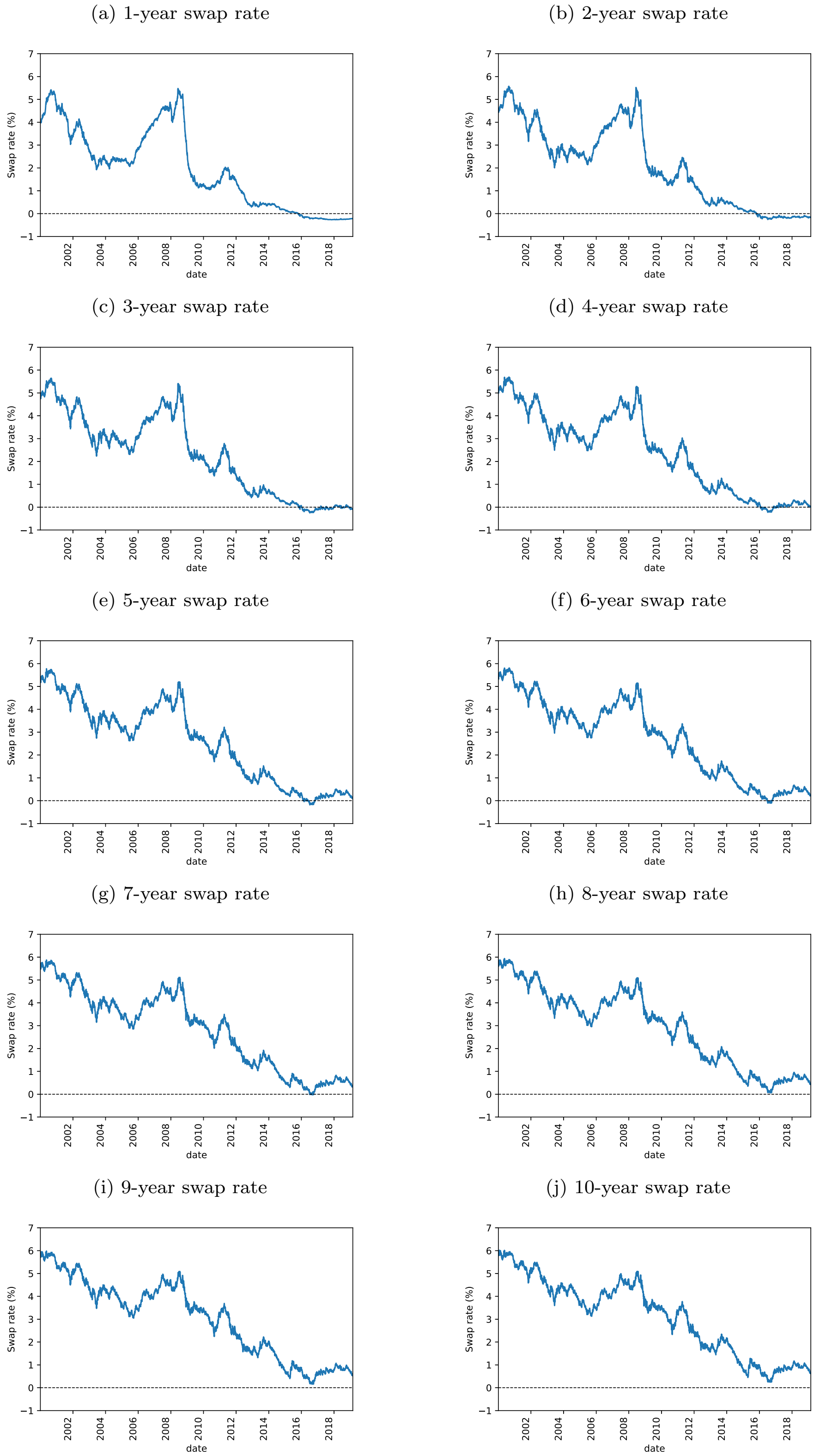


Figure 1: EUR swap rates ranging from January 2, 2000 to February 25, 2019.

5 Methodology

In this paper, I have chosen to use the two-factor Vasicek model for modeling the EUR swap curve. For the neural network model, the values of the two factors are used as inputs, and the benchmark strategy is based solely on the Vasicek model. The reasoning for choosing the two-factor Vasicek model is threefold. First, the model allows for negative rates. This is essential since the EUR swap rates entered the negative environment in the autumn of 2015. Second, the model is relatively easy and straightforward to calibrate. Finally, the model has already been used by Duarte et al. (2006) which makes it possible to compare the results.

5.1 Fitting the Vasicek factors

The neural network model uses the two factors of the Vasicek model as its inputs. Calculating the factor values requires calibrating the Vasicek model to the market swap curves. In the two-factor Vasicek model without shift extension, the processes for the factors are defined as

$$dx(t) = k_1[\theta_1 - x(t)]dt + \sigma_1 dW_1(t) \tag{31}$$

$$dy(t) = k_2[\theta_2 - y(t)]dt + \sigma_2 dW_2(t) \tag{32}$$

With these factor definitions, the zero-coupon bond price is

$$P(t,T) = A(t,T,k_1,\theta_1,\sigma_1)A(t,T,k_2,\theta_2,\sigma_2)\exp\{-B(t,T,k_1)x_t - B(t,T,k_2)y_t\}, \tag{33}$$

where

$$A(t,T,k,\theta,\sigma) = \exp\left\{\left(\theta - \frac{\sigma^2}{2k^2}\right)(B(t,T,k) - T + t) - \frac{\sigma^2}{4k}B(t,T,k)^2\right\} \tag{34}$$

and

$$B(t,T,k) = \frac{1}{k}\left[1 - e^{-k(T-t)}\right]. \tag{35}$$

The model-implied swap rate for tenor T is then

$$\hat{s}(T) = \frac{1 - P(0,T)}{\sum_{i=1}^{n} P(0,t_i)}, \tag{36}$$

where t_i are the times for the swap cash flows. For example, with a semiannual swap these would be apporiximately 0.5, 1.0, 1.5 and so on. Duarte et al. (2006) use the following parametrization

$$dx(t) = (\alpha - \beta x(t))dt + \sigma dW_1(t)$$
$$dy(t) = (\mu - \gamma y(t))dt + \eta dW_2(t)$$

and report the optimized parameters as

$$\alpha = 0.0009503$$
$$\beta = 0.0113727$$
$$\sigma = 0.0548290$$
$$\mu = 0.0240306$$
$$\gamma = 0.4628664$$
$$\eta = 0.0257381.$$

The optimized parameter values are used as an initial guess for the optimization process. The parametrization used by Duarte et al. (2006) can be transferred to the parametrization shown in (32) with the following relationships:

$$\kappa_1 = \beta$$
$$\theta_1 = \frac{\alpha}{\beta}$$
$$\sigma_1 = \sigma$$
$$\kappa_2 = \gamma$$
$$\theta_2 = \frac{\mu}{\gamma}$$
$$\sigma_2 = \eta$$

The goal is to calculate the values for x_0 and y_0 for each date in the data. This is done with the same process that is applied in both Duarte et al. (2006) and Karsimus (2015) with the modification that in this paper, the parameters are calibrated with a rolling window so that no forward-looking bias will occur in the calculated factor values. The process goes as follows. First, I pick the first half-year period (that is, the first 126 observations) of data from the beginning of the dataset. Second, by using the trial parameters, I solve the values of x_0 and y_0 for each date in the half-year window so that the model swap rates in tenors of 1 year and 10 years match exactly to the true market swap rates for the corresponding tenors. After this, the six parameters ($\kappa_1, \theta_1, \sigma_1, \kappa_2, \theta_2$ and σ_2) are optimized for the half-year window by minimizing the mean squared error for tenors ranging from 2 years to 9 years. This is a non-linear least-squares problem, which in this paper is solved with the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm[2]. Once the parameters are optimized, x_0 and y_0 are again solved so that the 1-year and 10-year model swap rates match the market rates exactly. The factor values for the last date in the half-year window are then saved, and the half-year window is shifted one day forward. The process continues similarly until the end of the dataset, with the exception that the most recent optimized values for the six parameters are used as trial values in the next half-year window. This process produces daily values for the factors, which are then used as an input to the neural network.

5.2 Swap Mechanics and Valuation Methodology

Interest rate swaps are derivatives instruments which exchange future cash flows based on two legs, the fixed leg and the floating leg. The owner of the swap receives the cash flows from one leg and pays the cash flows from the other leg to the counterparty of the contract. A *payer swap* pays the fixed leg and receives the floating leg and a *receiver swap* in turn receives the fixed leg and pays the floating leg. The cash flows of the fixed leg are determined by the specified fixed rate. The cash flows of the floating leg are usually determined by IBOR rates such as LIBOR or EURIBOR. In this paper, semi-annual EUR swaps are used which means that the floating leg is based on 6-month EURIBOR rates and cash flows are exchanged semi-annually.

The interest rate swap rates that are observed on the market are so-called *par swap rates*. A par swap rate is an annualized fixed rate that is determined in such way that the discounted cash flows of the fixed leg and the floating leg sum to zero at initialization of the swap. Thus, at initialization, the fixed leg cash flows are known for the whole lifetime of the swap since the fixed rate does not change. However, at initialization, only the first

[2] In this paper, I use the Python package `SciPy` which has an implementation of the algorithm.

floating cash flow is known: it is the 6-month EURIBOR rate that is observed at the market at the time. As time passes, the floating rate is updated on each semi-annual cash flow date to the 6-month EURIBOR rate of that date. Because the floating leg cash flows are not known in advance, a forward curve must be used to obtain the implied forward rates for each semiannual interval during the lifetime of the swap. These implied forward rates are treated as the future cash flows of the floating leg. The methodology for constructing the forward curve is explained in the next section.

5.2.1 OIS Discounting

After the financial crisis of 2008, using solely IBOR-based discounting has become infeasible because IBOR-rates do not incorporate credit risk and collateral payments. Thus, market participants have started to use the overnight-indexed-swap (OIS) curve for valuation of interest rate swaps. This is generally referred as OIS discounting.

For swap valuation, I follow the methodology presented in Smith (2013) where the author provides a practical, self-contained guide for valuing interest rate swaps with OIS discounting. The process goes as follows. First, the OIS zero rates are bootstrapped from the OIS rates. Cubic interpolation is then applied to obtain a continuous OIS zero curve. Second, the swap curve and the EURIBOR curve are interpolated with linear intepolation. After this, the OIS-consistent implied 6-month EURIBOR forward rates are bootstrapped by using the interpolated OIS zero curve, the interpolated EURIBOR curve and the interpolated swap curve. Finally, the zero rates at semiannual intervals are obtained from the forward curve and after linear interpolation, we have a continuous zero curve (P) that is used to discount the cash flows of the swap contracts. The value of a payer swap with tenor T at time t, $0 \leq t < T$ can then be expressed as

$$V = P(t, t_1)(\mathcal{L} - F) + \sum_{i=1}^{N} P(t, t_{i+1})(f(t, t_i, t_{i+1}) - F) \tag{37}$$

where $\mathcal{L}$ denotes the latest fixing of the floating rate, F denotes the fixed rate, N is the number of remaining semiannual payments at time t and t_i, $t_i < t_{i+1}$, $i = 1, ..., N$ are the remaining points of time of the semiannual payments with $t_N = T$. $f(t, T, S)$ denotes the forward curve at time t with tenor T and expiry S.

5.3 Modelling Methodology

A multi-layer perceptron, i.e. a vanilla neural network, is used to model the swap curve. The neural network is first trained with the training data ranging between January 10, 2000 and January 1, 2010. The specifications for the network are as follows.

- Parameter initialization method: LeCun normal (see Klambauer et al. (2017))
- Optimizer: Adam (see Kingma and Ba (2014))
- Number of hidden layers: 1
- Hidden layer size: 30
- Activation function: SELU (see Klambauer et al. (2017))
- Batch size: 32
- Number of training epochs: 1000
- Learning rate: 0.001

Number of hidden layers is set to one based on the universal approximator theorem by Hornik (1991). The chosen hidden layer size is rather arbitrary and based on experimentation. After the neural network has been trained with the training data, the trading strategy backtest is performed with the test data. This is done as follows. Each day in the backtest period, currently observed values for the model input variables, that is, the two fitted Vasicek factors, are fed to the network to produce the model-implied swap curve. After this, the network is updated by training the network with the currently observed input variable values for one epoch. The process continues similarly until the end of the backtest period.

5.4 Trading Methodology

In the trading methodology, I follow Duarte et al. (2006). Trades are initiated when the mispricing between some of the model-implied swap rates and the currently observed market swap rates is more than 5 basis points[3]. If the model-implied swap rate is lower than the current market swap rate, then a receiver swap is entered. Similarly, if the model-implied swap rate is higher than the current market swap rate, then a payer swap is entered. In

[3] Karsimus (2015) uses three different thresholds: 10, 15 and 20 basis points and Duarte et al. (2006) use 10 basis points.

case of multiple mispricings, only the largest mispricing is traded. However, there can be multiple trades that are open simultaneously because previously made trades do not need to be closed when a new mispricing occurs. A trade is considered converged if the current market swap rate deviates from the model-implied rate that was observed at the initiation of the trade by less than one basis point. A trade will be closed when convergence occurs. Also, the maximum length of each trade is 4 months, so if a trade has not converged in 4 months after initiation, the trade will be closed. A one basis point transaction cost will be applied for each closed trade which is similar as in Duarte et al. (2006). When closing trades, it is assumed that the swap position in question can be closed with its current market value. One caveat that should be noted here is that ongoing swaps are not generally quoted in the market. Thus, it is not possible to evaluate if in reality there would be enough liquidity to be able to close the ongoing swaps with their current market value at the desired point of time.

Each day during the backtest period, all open swap positions are valued with OIS discounting (see 5.2.1). Thus, the returns of the strategy are based on daily mark-to-market valuation of the total position.

5.4.1 Hedging methodology

The idea with hedging is to make the strategy market neutral by taking offsetting positions in 1-year and 10-year swaps. As the modelling is based on the two-factor Vasicek model, the only uncertainty stems from the two factors. Thus, the idea is to solve the weights for the 1-year and 10-year swaps so that the total position is neutral to the Vasicek factors. In formulas, this can be expressed as

$$\begin{bmatrix} \frac{\partial \hat{s}(t)}{\partial x} \\ \frac{\partial \hat{s}(t)}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial \hat{s}(1)}{\partial x} & \frac{\partial \hat{s}(10)}{\partial x} \\ \frac{\partial \hat{s}(1)}{\partial y} & \frac{\partial \hat{s}(10)}{\partial y} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_{10} \end{bmatrix} \tag{38}$$

where $\hat{s}(t)$ is the model-implied swap rate for tenor t and ω_1 and ω_{10} are the weights for the 1-year and 10-year swaps. For the Vasicek model, the partial derivatives of the swap rate with respect to the factors can be calculated in closed form as

$$\frac{\partial \hat{s}(T)}{\partial x} = \frac{B(0,T,k_1)P(0,T)\sum_{i=1}^{n} P(0,t_i) + (1 - P(0,T))\sum_{i=1}^{n} B(0,t_i,k_1)P(0,t_i)}{\left(\sum_{i=1}^{n} P(0,t_i)\right)^2}, \tag{39}$$

and respectively for y by just replacing k_1 with k_2. For the neural network model, the partial derivatives with respect to the factors are not available in closed form, but are instead computed with automatic differentiation in Tensorflow[4]. Once ω_1 and ω_{10} are solved, the nominals of the swaps are scaled in such way that the nominals of the hedge trades and the main trade sum up to 1.

[4]Computing the partial derivatives is done by computing the Jacobian matrix of the neural network output with Tensorflow's `GradientTape` object.

6 Results and Findings

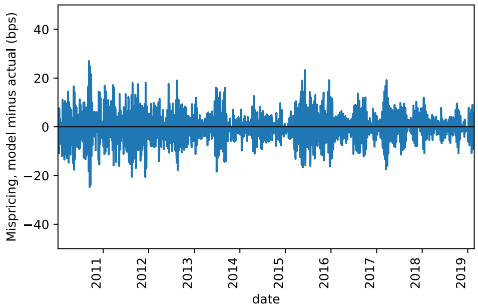
In this section, the results of the backtested strategies are presented and analyzed. This section is divided into three subsections. The first discusses the trading performance of the strategies, the second analyzes the risk factor exposures and the third analyzes the tail risk of the strategies.

6.1 Out-of-sample performance

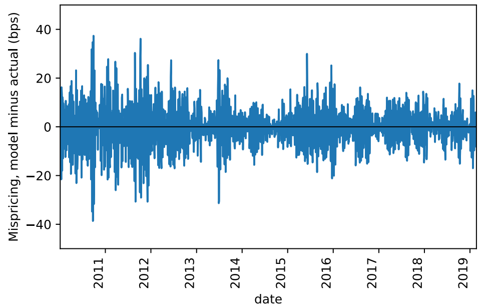
Table 2: Out-of-sample summary statistics of the leveraged daily returns for the yield curve arbitrage strategies. The training period for the neural network model is between January 10, 2000 and January 1, 2010 and the out-of-sample testing period is between January 2, 2010 and February 25, 2019 for both strategies. Trades are initiated when the mispricing between some of the model-implied swap rates and the currently observed market swap rates is more than 5 basis points. In case of multiple mispricings, only the largest mispricing is traded. However, there can be multiple trades that are open simultaneously (i.e. previously made trades do not need to be closed if a new mispricing occurs). Each initiated trade is hedged with 1-year and 10-year swaps so that the trade is neutral to the Vasicek factors. The notionals are scaled so that the total notional of each trade is 1. A trade is considered converged when the current market swap rate deviates from the model-implied rate that was observed at the initiation of the trade by less than one basis point. Upon convergence, a trade is closed with a one basis point transaction cost. New trades are not initiated if time to the end date of the backtest period is less than 4 months. N denotes the number of daily returns during the out-of-sample period. Leverage is the leverage ratio that generates an ex post annual standard deviation of 10%. G/L denotes the gain-loss ratio. MDD denotes the maximum drawdown in percentages and its length in days. The t-statistic, skewness and kurtosis are based on heteroscedasticity and autocorrelation robust standard errors using 1 lag. The Sharpe ratio is annualized. All return units are in basis points.

Model	N	Leverage	Mean	t-stat	Std	Min	Max	Skew	Kurt	Sharpe	G/L	MDD (%)	MDD (days)
NN	2330	32.30	3.22	2.54	62.98	-616.68	556.96	0.65	16.487	0.81	1.08	-16.48%	485
Vasicek	2236	8.33	2.68	2.05	62.98	-275.86	288.16	0.15	5.70	0.68	1.09	-28.70%	413

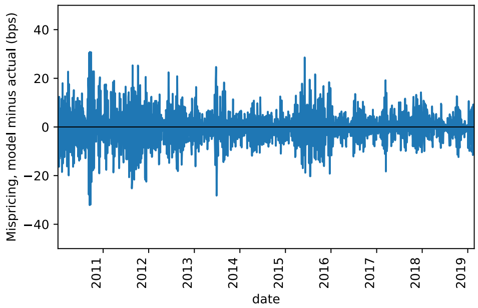
(a) Neural network, 2-year swap mispricings



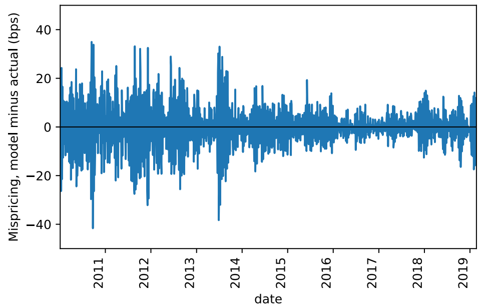
(b) Neural network, 3-year swap mispricings



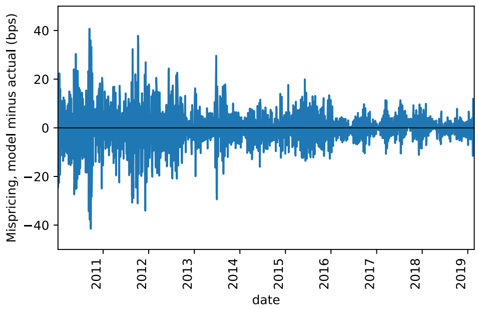
(c) Neural network, 4-year swap mispricings



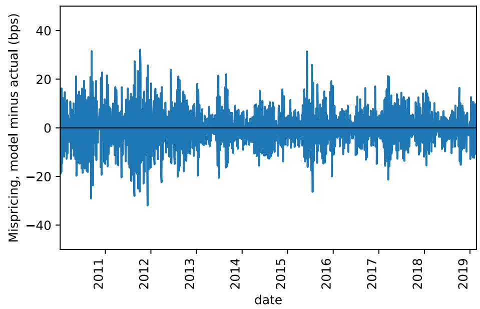
(d) Neural network, 5-year swap mispricings



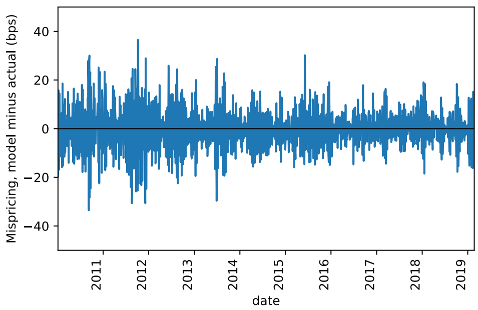
(e) Neural network, 6-year swap mispricings



(f) Neural network, 7-year swap mispricings



(g) Neural network, 8-year swap mispricings



(h) Neural network, 9-year swap mispricings

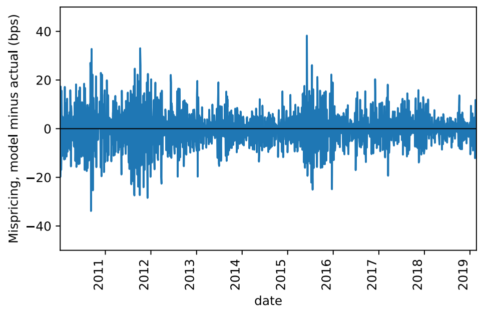
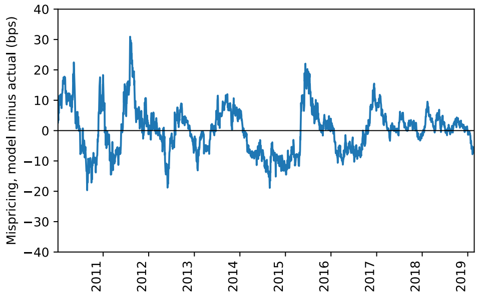
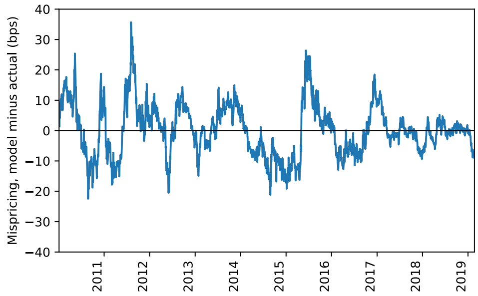


Figure 2: Neural network mispricings for swaps with tenors ranging from 2 years to 9 years.

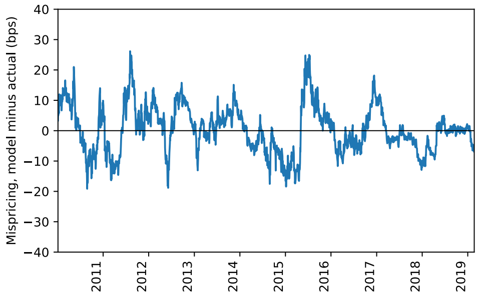
(a) Vasicek, 2-year swap mispricings



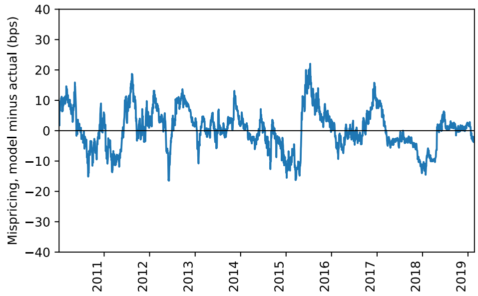
(b) Vasicek, 3-year swap mispricings



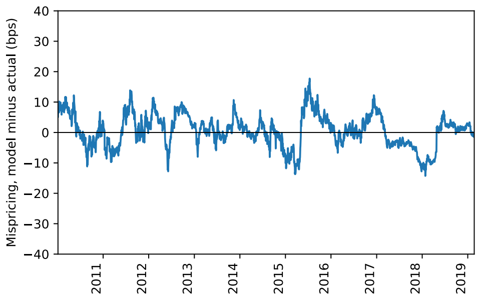
(c) Vasicek, 4-year swap mispricings



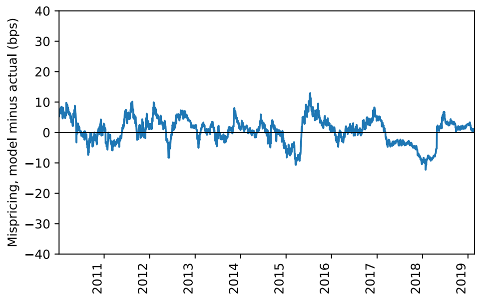
(d) Vasicek, 5-year swap mispricings



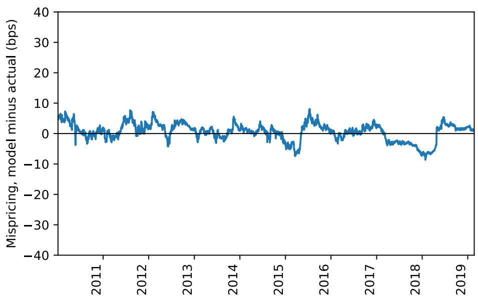
(e) Vasicek, 6-year swap mispricings



(f) Vasicek, 7-year swap mispricings



(g) Vasicek, 8-year swap mispricings



(h) Vasicek, 9-year swap mispricings

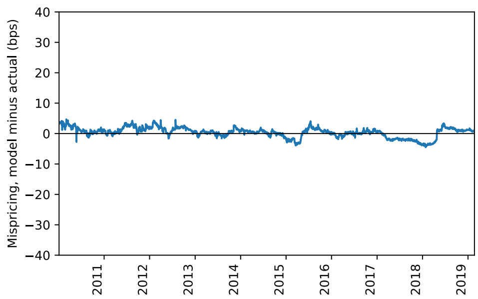
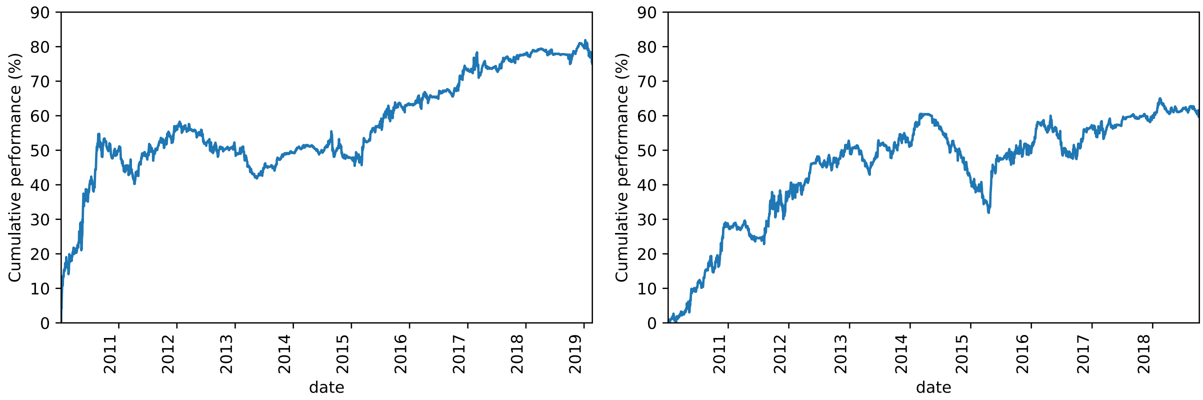


Figure 3: Vasicek model mispricings for swaps with tenors ranging from 2 years to 9 years.



(a) Neural network cumulative out-of-sample performance

(b) Vasicek model cumulative out-of-sample performance

Figure 4: Cumulative performances for the neural network model and the Vasicek model

Based on the summary statistics in Table 2 it seems that the neural network strategy earns attractive returns during the out-of-sample test period. The Sharpe ratio of the neural network strategy is 0.81 which is higher compared to the Sharpe ratio of 0.68 of the Vasicek model strategy. Duarte et al. (2006) achieve annualized Sharpe ratios ranging between 0.52 and 0.79 with different configurations of the strategy. In all of their configurations, Duarte et al. (2006) use a 10 basis point threshold for initiating trades. However, the authors declare that using a 5-basis-point threshold (which is used in this paper) yields similar results. Also, Karsimus (2015) achieves a Sharpe ratio of 1.26 by replicating the strategy of Duarte et al. (2006) with the two-factor Cox-Ingersoll-Ross model (27) with 5-basis-point-threshold for the period ranging from January 2002 to January 2015. However, this result is for in-sample backtest. For the out-of-sample results with 5-basis-point limits, Karsimus (2015) achieves a Sharpe ratio of 1.08 with rolling calibration and 0.94 with fixed calibration of the model.

The distribution of the returns of the neural network strategy is positively skewed and has high kurtosis which supports the second hypothesis. Also, the skewness and kurtosis values of the neural network are much higher compared to the Vasicek model. The t-value of 2.54 implies that the average return of the neural network differs from zero at the 1% significance level.

Looking at the maximum drawdowns of the strategies, it seems that the maximum drawdown of the neural network in percentages (-16.48%) is smaller in magnitude compared to the Vasicek model (-28.70%). One possible explanation for the large maximum drawdown of the Vasicek model might be the change in swap rate dynamics before entering the negative territory in the fall of year 2015. Looking at the cumulative performance of the Vasicek model in Figure 4 reveals that the maximum drawdown period between 2014 and 2017 contains indeed the period when the swap rates started entering negative territory. However, even though the maximum drawdown of the neural network is lower compared to the Vasicek model, the duration of the drawdown is quite long, 485 days.

Looking at the neural network mispricings in Figure 2 and the Vasicek model mispricings in Figure 3, it seems that the neural network is more sensitive and reacts quicker to changes in the swap rate dynamics. Also, the magnitude of the neural network mispricings seems to stay rather constant across tenors, while the Vasicek model mispricings seem to decay with longer tenors. This is a little surprising as the Vasicek factors are calibrated so that they match exactly the 1-year and 10-year swap rates. Thus, one might expect that the mispricings would be smaller in magnitude in the short and long end of the curve, and larger in the middle.

Generally, it seems that the neural network model is superior to the two-factor Vasicek model when performance is measured by Sharpe ratio. On the other hand, judging by gain-loss ratios, it seems that the performance of the models is almost equal, the neural network being slightly weaker than the Vasicek model. Also, it should be noted that even though the neural network model produces an out-of-sample cumulative performance of around 75%, it also uses a very high leverage ratio of 32.

6.2 Multifactor regressions

In this section, I analyze the exposure of the backtested strategies to systematic risk factors. As yield curve arbitrage strategies strive to be market neutral, they are assumed to generate returns that have somewhat minimal exposure to commonly known risk factors. In other words, the systematic risk factors are assumed to have only little explanatory power on the yield curve arbitrage returns. However, as the yield curve arbitrage strategies are not pure textbook arbitrage but more of market neutral relative value bets, it is reasonable to assume that the strategies have exposure to interest rate-related risk, as pointed by Vayanos and Vila (2009).

For the risk factor exposure analysis, I use two different factor models. In the first model, I combine the well-known Fama-French three-factor model (Fama and French (1993)) and the Fama-French momentum factor with hedge fund tail risk factors presented by Adrian et al. (2011). These hedge fund tail risk factors include factors for carry and short volatility strategy returns, which tend to generate steady, small returns most of the time but are also prone to experience large losses occasionally. Thus, these factors are well-suited for incorporating tail risk properties into the risk factors. The second factor model is the asset-based style factor model presented by Fung and Hsieh (2004), which is specifically hedge fund-oriented.

The first factor model is built as follows. For equity-related risk factors, I use the three factors by Fama and French (European versions): the size factor SMB (small-minus-big), the value factor HML (high-minus-low) and the momentum factor WML (winner-minus-loser).[5] In addition to the equity risk factors, I include the following risk factors. First, the Bloomberg Cumulative FX Carry Trade index (CTG10) is included for incorporating carry risk. Second, a Credit Suisse short VIX total return index is included for implied volatility-related risk (labelled as VIX) and a Credit Suisse short variance swap total return index is included for incorporating risk related to level shifts in volatility (labelled as VAR). Third, for credit risk, I include the Bloomberg Barclays US aggregated BAA total return index (labelled as BAA). Finally, I include the slope of the yield curve and the liquidity

[5]The data are from the website of Kenneth R. French: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html`.

spread as European bond risk factors. The slope factor is calculated as the daily change in the spread between the 10-year generic German government yield and the 3-month generic German government yield (10Y-3M). The liquidity spread (LIQ) is calculated similarly as the daily change in the spread between the Frankfurt 3-month Interbank Offered rate and the 3-month generic German government yield.[6] That is, the excess returns of the strategy are analyzed with the model

$$R = \alpha + \beta_0 R_M + \beta_1 R_{SMB} + \beta_2 R_{HML} + \beta_3 R_{WML} + \beta_4 R_{CTG10} + \beta_6 R_{VIX} + \beta_7 R_{VAR} + \beta_8 R_{BAA} + \beta_9 R_{10Y-3M} + \beta_{10} R_{LIQ}.$$

[6] All data except the Fama-French factors are acquired from Bloomberg Terminal.

Table 3: Multifactor regression results of the daily excess returns on systematic risk factors. Carry denotes the returns of a Bloomberg carry index on G10 currencies. Credit denotes the returns of Bloomberg Barclays US aggregated BAA total return index. For the Fama-French factors, HML denotes the value factor (high-minus-low) and SMB denotes the size factor (small-minus-big). The Fama-French factors factors are based on European data. Liquidity spread denotes the daily change in the spread between the Frankfurt 3-month Interbank Offered Rate and the 3-month generic German government yield. Slope denotes the daily change in the spread between the 10-year German generic government yield and the 3-month German generic government yield. Volatility (implied) denotes the returns of a Credit Suisse Short VIX index, and volatility (level) denotes the returns of a Credit Suisse Short Variance Swap index. Regressions are calculated with heteroscedasticity and autocorrelation robust standard errors with 1 lag. All values denote the t-statisics for the factors, and alpha is reported in basis points in parenthesis.

	Dependent variable:	
	Neural network	Vasicek
α	2.507**	2.284**
	(2.87)	(3.00)
Carry	-2.128**	-0.850
Credit	0.611	-2.862***
Fama-French HML	-0.701	-0.612
Fama-French Market	-0.124	-0.692
Fama-French Momentum	-2.208**	0.295
Fama-French SMB	1.527	2.055**
Liquidity spread	8.847***	2.374**
Slope	-10.494***	-2.457**
Volatility (implied)	-0.972	-1.058
Volatility (level shift)	-0.276	-0.318
R2	0.194	0.056
Note:	*p<0.1; **p<0.05; ***p<0.01	

The results in Table 3 imply that both models produce significant multifactor alpha at 5% significance level. The alpha of the neural network is 2.87 basis points per day, which is slightly less than the alpha of the Vasicek model, 3 basis points per day. In addition, the neural network strategy has significant exposures to carry, momentum, slope and liquidity spread. The significant exposures of the neural network strategy the momentum risk factor and the Vasicek model strategy to the size factor are somewhat surprising as one might think that yield curve arbitrage strategies should not have very much in common with equity markets. However, this kind of phenomenon where fixed income strategies contain stock market risk was also discovered by Campbell (1987) whose results imply that the risk premia on equity markets tend to move closely together with the risk premia of long-dated bonds.

Carry strategies are generally known to produce small, steady returns most of the time, but occasionally they tend to incur heavier losses. Thus, it seems that the neural network strategy has exposure to this kind of tail risk that is typical for carry strategies as the carry factor is significant at 5% level for the neural network strategy. Surprisingly, the neural network strategy does not seem to be exposed to significant credit risk as opposed to the Vasicek model, for which the credit factor is significant at 5% level. This is contrary to the results of Fung and Hsieh (2002), which indicate that fixed income arbitrage strategies are typically exposed to especially credit risk. Thus, it seems that the neural network model is able to mitigate the credit risk but in turn has exposure to carry risk.

To conclude, it seems that the strategy based on the Vasicek model has somewhat more desirable properties for a hedge fund-like arbitrage strategy mainly because the risk factors only explain about 5.6% of the variance in the Vasicek model returns as implied by the R^2 of the multifactor regression. However, both strategies generate multifactor alpha, which is in line with the results of both Duarte et al. (2006) and Karsimus (2015). Both strategies have significant exposures to interest rate-related risk factors, which is logical since both strategies take relative value bets on the interest rate swap curve. Also, both strategies have significant exposure to one equity factor which is in line with the results of Campbell (1987). The neural network model seems to contain carry risk premium, which is also logical due to the similar characteristics of carry strategies and fixed income arbitrage strategies. The Vasicek model has exposure to credit risk which is a similar result as in Fung and Hsieh (2002).

6.2.1 Asset-based style factor regressions

In addition to the regression model specified above, I also evaluate the exposure of the strategies to risk factors that are specific to hedge funds. This is done with the Fung-Hsieh asset-based style (ABS) factor model Fung and Hsieh (2004) that consists of eight factors: two equity risk factors, two interest rate-related risk factors, three trend-following risk factors and finally an emerging market risk factor. The ABS model is used to identify if the yield curve arbitrage strategies produce so-called hedge fund alpha. The equity factors consist of a market factor and a size spread factor. In this paper, I use the returns of STOXX Europe 600 index as the market factor. For small cap returns, I use STOXX Europe 200 Small index and the size spread factor is then defined as the difference between the small cap returns and the returns of the market portfolio. For the interest rate-related risk factors, I use the monthly change of the German 10-year constant maturity yield as a bond risk factor and Bloomberg Barclays US aggregated BAA total return index as the credit risk factor.[7] The three option portfolios which incorporate risk factors for currency trend following, bond trend following and commodity trend following are the original lookback straddle portfolios constructed by Fung and Hsieh (2004).[8] In this case, the regressions are run with monthly returns because the frequency of the Fung-Hsieh lookback straddle portfolios is on a monthly basis.

[7]Data for the equity factors, the interest rate-related risk factors and the emerging market risk factor are acquired from Bloomberg Terminal.

[8]The data for the lookback straddle portfolios are from the data library of David A. Hsieh: http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-FAC.xls.

Table 4: Regression results of the monthly excess out-of-sample returns of the yield curve arbitrage strategies on Fung-Hsieh asset-based style factors. Bond denotes the monthly change of German 10-year constant maturity yield. Bond lookback straddle denotes the returns of a portfolio of lookback straddles on bond futures. Credit denotes the returns on Bloomberg Barclays US aggregated BAA total return index. Commodity lookback straddle denotes the returns of a portfolio of lookback straddles on commodity futures. Emerging market is the return on MSCI Emerging Markets index. FX lookback straddle denotes the returns of a portfolio that consists of lookback straddles on currency futures. Market returns are the returns of STOXX Europe 600 index. Size spread denotes the difference between returns on STOXX Europe 200 Small and STOXX Europe 600. Regressions are calculated with heteroscedasticity and autocorrelation robust standard errors with 1 lag. All values denote the t-statisics for the factors, and alpha is reported in parenthesis in basis points.

	Dependent variable:	
	Neural network	Vasicek
α	2.024**	2.677***
	(38.049)	(63.151)
Bond	-3.884***	-0.819
Bond lookback straddle	-2.244**	-0.695
Commodity lookback straddle	-1.521	0.119
Credit	-2.690***	-1.999**
Emerging market	1.777*	-0.447
FX lookback straddle	1.110	-0.228
Market	-0.612	-0.563
Size spread	-0.715	0.348
R2	0.277	0.114
Note:	*p<0.1; **p<0.05; ***p<0.01	

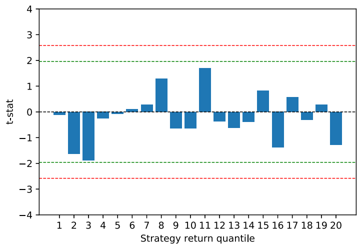
Based on the results in Table 4, it seems that both strategies generate multifactor hedge fund alpha. The alpha of the neural network strategy is significant at 5% level, and respectively the alpha of the Vasicek model is significant at 1% level. Also, the Vasicek model alpha (63 basis points per month) is almost double compared to the neural network alpha (38 basis points per month). The neural network seems to have significant exposure to bond and credit factors, and also to the bond trend following factor. Especially the significant bond and credit exposures are somewhat expected based on previous literature since as pointed by Vayanos and Vila (2009), yield curve arbitrage strategies are expected to have some exposure to interest rate-related risks. Thus, it is surprising that neither the bond risk factor nor the bond trend following risk factor are significant for the Vasicek model. The only significant factor for the Vasicek model is the credit factor. Thus, it seems that also with the asset-based style factors, the Vasicek model has less exposure to the factors compared to the neural network strategy. The neural network strategy also has exposure to emerging market risk factor, but this result is significant only at 10% level. Most probably, this emerging market exposure is mostly due to noise since the yield curve arbitrage strategies operate on the developed market interest rates of the Eurozone.

Generally, based on the results of the Fung-Hsieh ABS regressions, it seems that the Vasicek model is more attractive from hedge fund perspective compared to the neural network model since the Vasicek model produces clearly higher multifactor alpha and has exposure to only one of the ABS factors. In addition, the R^2 of 0.114 is a lot smaller compared to the neural network (0.277).

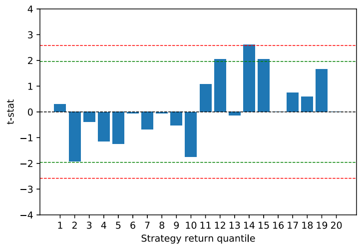
6.3 Quantile Regressions

This section analyzes the tail risk of the strategies in more detail. The yield curve arbitrage strategy returns are divided into 20 quantiles, and then the strategy returns are regressed with the previously presented Fama-French/tail risk factor regression model inside each quantile. Quantile regressions are used in a similar manner by Adrian et al. (2011) to evaluate the tail risk exposures of hedge fund strategies with the difference that Adrian et al. (2011) regress the strategy returns pairwise with each risk factor but here multifactor regressions are employed for each quantile in order to avoid omitted variable bias in the quantile regressions.

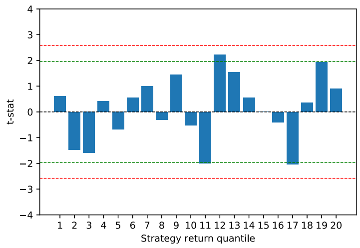
(a) Neural network, carry



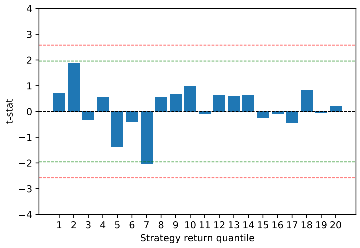
(b) Neural network, credit



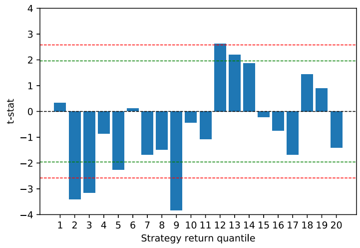
(c) Neural network, high-minus-low



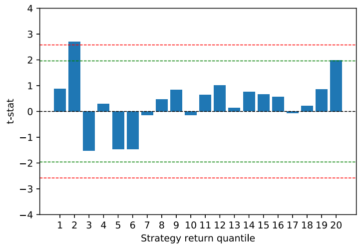
(d) Neural network, market



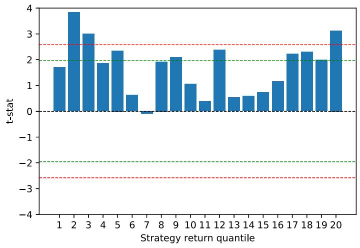
(e) Neural network, momentum



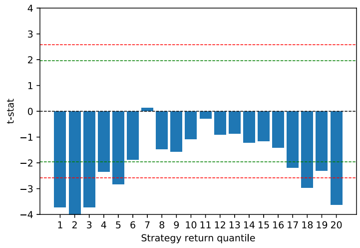
(f) Neural network, small-minus-big



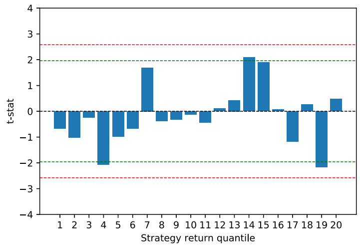
(g) Neural network, liquidity



(h) Neural network, slope



(i) Neural network, volatility (implied)



(j) Neural network, volatility (level shift)

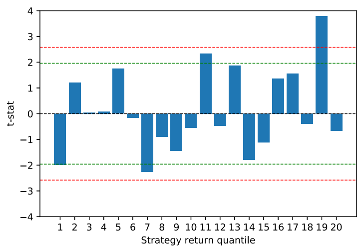
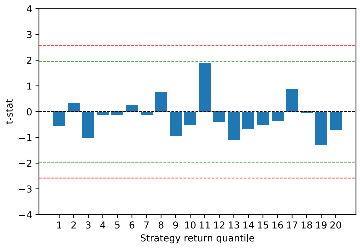
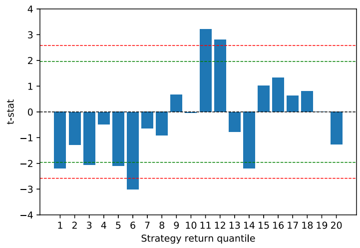


Figure 5: Neural network strategy, multifactor quantile regression t-statistics. The dashed lines denote the 5% (green) and 1% (red) significance levels.

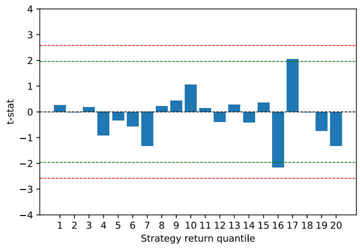
(a) Vasicek, carry



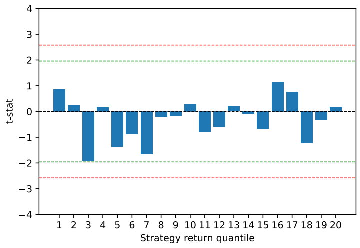
(b) Vasicek, credit



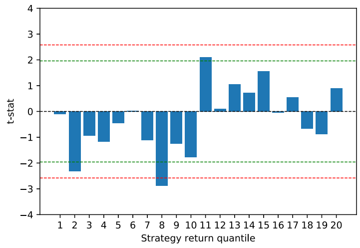
(c) Vasicek, high-minus-low



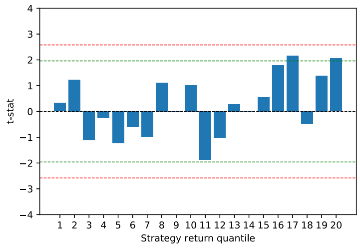
(d) Vasicek, market



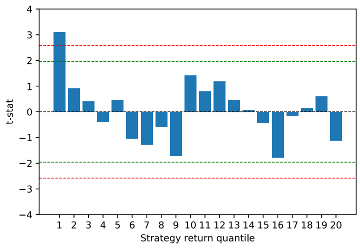
(e) Vasicek, momentum



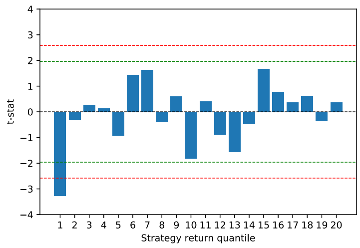
(f) Vasicek, small-minus-big



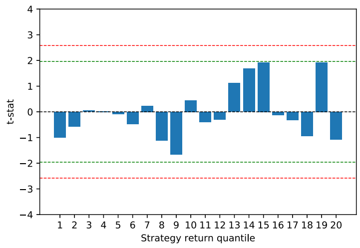
(g) Vasicek, liquidity



(h) Vasicek, slope



(i) Vasicek, volatility (implied)



(j) Vasicek, volatility (level shift)

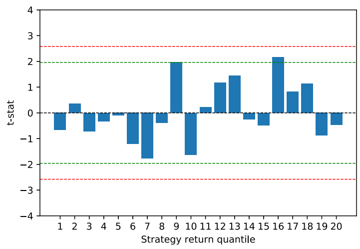


Figure 6: Vasicek model strategy, multifactor quantile regression t-statistics. The dashed lines denote the 5% (green) and 1% (red) significance levels.

Table 5: Multifactor quantile regression results of daily excess returns, left and right tails. The number of quantiles is 20. Left tail denotes the first quantile of the yield curve arbitrage returns, and right tail denotes the twentieth quantile. Carry denotes the returns of a Bloomberg carry index on G10 currencies. Credit denotes the returns on Bloomberg Barclays US aggregated BAA total return index. For the Fama-French factors, HML denotes the value factor (high-minus-low) and SMB denotes the size factor (small-minus-big). The Fama-French factors factors are based on European data. Liquidity spread denotes the daily change in the spread between Frankfurt 3-month Interbank Offered Rate and the 3-month generic German government yield. Slope denotes the daily change in the spread between the 10-year German generic government yield and the 3-month German generic government yield. Volatility (implied) denotes the returns of a Credit Suisse Short VIX index, and volatility (level) denotes the returns of a Credit Suisse Short Variance Swap index. Regressions are calculated with heteroscedasticity and autocorrelation robust standard errors with 1 lag. All values denote the t-statisics for the factors.

	Dependent variable:			
	Neural left tail	Vasicek left tail	Neural right tail	Vasicek right tail
Carry	-0.115	-0.544	-1.271	-0.717
Credit	0.301	-2.202**	0.011	-1.263
Fama-French HML	0.620	0.259	0.912	-1.315
Fama-French Market	0.721	0.860	0.230	0.164
Fama-French Momentum	0.330	-0.103	-1.411	0.903
Fama-French SMB	0.880	0.329	1.997**	2.058**
Liquidity spread	1.707*	3.101***	3.128***	-1.110
Slope	-3.719***	-3.282***	-3.615***	0.356
Volatility (implied)	-0.671	-0.992	0.471	-1.088
Volatility (level shift)	-1.988**	-0.663	-0.677	-0.468
R2	0.433	0.2	0.432	0.349

Note: *p<0.1; **p<0.05; ***p<0.01

The quantile regression graphs for the neural network in Figure 5 imply that the neural network strategy is exposed to tail risk with respect to the interest rate-related risk factors slope and liquidity spread. The graphs for the quantile t-statistics of liquidity spread and slope are both U-shaped, which means that the risk factor significance increases in the tails of the strategy. For the Vasicek model, the slope and liquidity spread factors are also significant in the left tail of the strategy, but the graphs in Figure 6 do not show as clear U-shaped pattern as with the neural network.

Table 5 presents the multifactor regression results in the left and right tails of both of the yield curve arbitrage strategies. The results imply the following. Regarding the credit factor, the strategies behave rather differently in the tails. For the left tail of the Vasicek model strategy, the credit factor is significant at 5% level whereas it is insignificant for the neural network strategy in both tails. This result is in line with the previous regression results of the Fama-French/hedge fund tail risk factor regression model. Also, reflecting on the previous regression results, it is a bit surprising that the carry risk factor is insignificant for the neural network model in both tails. Thus, it seems that the neural network strategy has exposure to carry risk, but this risk is not purely tail risk. Instead of carry tail risk, the neural network strategy seems to pick up level-shift volatility risk in the left tail. However, the coefficient of the volatility level shift is negative, which means that in the left tail of the neural network strategy, the effect of negative returns for the short variance swap portfolio is positive for the neural network strategy. This could be beneficial, since during market tail events, volatility tends to spike. Thus, as the neural network strategy has negative coefficient for a short volatility portfolio in its left tail, during market turmoil the volatility spiking can have a positive effect on the neural network returns, assuming that the neural network strategy is also simultaneously in its left tail.

With equity factors, both strategies have significant coefficients for the size factor in the right tails. However, based on previous literature, the exposure of the strategies to equity factors is not as surprising as it first might sound. As mentioned previously, Campbell (1987) finds that fixed income strategies can indeed contain significant amounts of stock market risk.

Considering the interest rate-related risk factors, both the neural network strategy and the Vasicek model strategy have significant exposure to the interest rate risk in the left tail, as implied by the significant t-statistics for the slope risk factor. Regarding the liquidity spread, both strategies have 1% significance in both left and right tails. Again, this is somehow expected as the liquidity spread mainly depicts overall liquidity and credit risk in the markets and fixed income arbitrage strategies tend to be exposed such risks. Also, the swap curve incorporates a credit premium in itself, as noted by Liu et al. (2002).

Overall based on the quantile regression analysis, it seems that both the neural network and the Vasicek model have exposure to mainly interest-rate related tail risks. Based on the shapes of the graphs in Table 5, the risk factor exposures experience somewhat random behaviour across quantiles for most of the risk factors. This is a desirable property for the yield curve strategies as such strategies are expected to have very little co-movement with the risk factors despite of the market environment.

7 Conclusions

In this thesis, I analyze the out-of-sample trading performance of a yield curve arbitrage strategy on EUR swap curve where the modelling is based on a novel hybrid neural network approach in which a neural network uses the fitted factors of the two-factor Vasicek model as its inputs. I compare the results to an identical benchmark strategy where the modelling is based solely on the two-factor Vasicek model. Evaluation of the performance is done by comparing well-known investment statistics such as the Sharpe ratio, gain-loss ratio and multifactor alpha. The first hypothesis is that the neural network-based strategy outperforms the benchmark strategy when performance is evaluated by Sharpe ratio and gain-loss ratio.

The results regarding the first hypothesis are two-fold. Based on Sharpe ratio, the neural network model performs clearly better compared to the benchmark strategy: the out-of-sample Sharpe ratio of the strategy 0.81 including transaction costs is generally higher compared to the benchmark strategy, which has Sharpe ratio of 0.68. The gain/loss ratios of the strategies are practically equal.

The second hypothesis is that the neural network-based strategy produces positively skewed returns with high kurtosis. The results support this hypothesis. Also, the skewness and kurtosis values for the neural network strategy are quite much higher compared to the Vasicek model.

The third hypothesis is that the neural network model produces positive, significant multifactor alpha. The results support this hypothesis as after controlling for well-known systematic risk factors, both the neural network strategy and the Vasicek model strategy produce significant alpha. This result is also in line with previous literature on yield curve arbitrage strategies.

The fourth hypothesis is that the neural network model has low exposure to well-known risk factors. The results of the multifactor regression analysis do not fully support this hypothesis as the neural network strategy has significant coefficients for especially interest rate-related risk factors. In this sense, the Vasicek model benchmark strategy seems to have more suitable properties for a hedge fund arbitrage strategy compared to the neural network model.

The final hypothesis states that the neural network strategy has notable levels tail risk with respect to systematic risk factors. The results of the quantile regression analysis imply that both of the strategies have significant tail risk coefficients for interest rate-related risk factors. In addition, the neural network model has significant left tail exposure for level shifts in volatility, and the Vasicek model has significant left tail exposure to a credit risk factor.

Even though the Vasicek model benchmark strategy has less exposure to risk factors compared to the neural network strategy, the Vasicek model strategy has a larger maximum drawdown of around -28% compared to -16% of the neural network strategy. One possible explanation for this is that the maximum drawdown takes place on a time period when the interest rates entered negative territory in the Eurozone, and that the Vasicek model is not able to adapt to the change in the interest rate dynamics as quickly as the neural network model.

To conclude, it seems that the Vasicek model benchmark strategy has more of the features of a market neutral hedge fund arbitrage strategy as it has less exposure to risk factors and produces higher multifactor alpha. On the contrary, the neural network strategy is more desirable from an investment point of view as it has rather high Sharpe ratio, the magnitude the maximum drawdown is smaller and the absolute cumulative performance is higher during the out-of-sample period. The promising results of the neural network strategy show the potential of applying machine learning models in the context of interest rates, and further research could focus on applying machine learning models for interest rates in a more general setting.

References

Adrian, T., Brunnermeier, M. K., and Nguyen, H.-L. Q. (2011). Hedge Fund Tail Risk. In *Quantifying Systemic Risk*, pages 155–172. University of Chicago Press.

Adrian, T., Crump, R. K., and Moench, E. (2013). Pricing the Term Structure with Linear Regressions. *Journal of Financial Economics*, 110(1):110–138.

Amilon, H. (2003). A Neural Network versus Black-Scholes: A Comparison of Pricing and Hedging Performances. *Journal of Forecasting*, 22(4):317–335.

Anders, U., Korn, O., and Schmitt, C. (1998). Improving the Pricing of Options: A Neural Network Approach. *Journal of Forecasting*, 17(5-6):369–388.

Andreou, P. C., Charalambous, C., and Martzoukos, S. H. (2008). Pricing and Trading European Options by Combining Artificial Neural Networks and Parametric Models with Implied Parameters. *European Journal of Operational Research*, 185(3):1415–1433.

Ang, A. and Piazzesi, M. (2003). A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics*, 50(4):745–787.

Asness, C. S., Krail, R. J., and Liew, J. M. (2001). Do Hedge Funds Hedge? *The Journal of Portfolio Management*, 28(1):6–19.

Backus, D. K., Foresi, S., and Telmer, C. I. (2001). Affine Term Structure Models and the Forward Premium Anomaly. *The Journal of Finance*, 56(1):279–304.

Bennell, J. and Sutcliffe, C. (2004). Black-Scholes Versus Artificial Neural Networks in Pricing FTSE 100 Options. *Intelligent Systems in Accounting, Finance & Management: International Journal*, 12(4):243–260.

Black, F. and Karasinski, P. (1991). Bond and Option Pricing when Short Rates are Lognormal. *Financial Analysts Journal*, 47(4):52–59.

Borovykh, A., Bohte, S., and Oosterlee, C. W. (2017). Conditional Time Series Forecasting with Convolutional Neural Networks. *arXiv preprint arXiv:1703.04691.*

Brigo, D. and Mercurio, F. (2007). *Interest Rate Models-Theory and Practice: with Smile, Inflation and Credit.* Springer Science & Business Media.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., and Siddique, A. (2016). Risk and Risk Management in the Credit Card Industry. *Journal of Banking & Finance*, 72:218–239.

Campbell, J. Y. (1987). Stock Returns and the Term Structure. *Journal of Financial Economics*, 18(2):373–399.

Collin-Dufresne, P. and Solnik, B. (2001). On the Term Structure of Default Premia in the Swap and LIBOR Markets. *The Journal of Finance*, 56(3):1095–1115.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). An Intertemporal General Equilibrium Model of Asset Prices. *Econometrica: Journal of the Econometric Society*, pages 363–384.

Dai, Q. and Singleton, K. (2003). Term Structure Dynamics in Theory and Reality. *The Review of Financial Studies*, 16(3):631–678.

Dai, Q. and Singleton, K. J. (2002). Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure. *Journal of financial Economics*, 63(3):415–441.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach. *Journal of Econometrics*, 131(1-2):309–338.

Dothan, L. U. (1978). On the Term Structure of Interest Rates. *Journal of Financial Economics*, 6(1):59–69.

Duarte, J., Longstaff, F. A., and Yu, F. (2006). Risk and Return in Fixed-Income Arbitrage: Nickels in Front of a Steamroller? *The Review of Financial Studies*, 20(3):769–811.

Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *The Journal of Finance*, 57(1):405–443.

Duffie, D. and Singleton, K. J. (1997). An Econometric Model of the Term Structure of Interest-rate Swap Yields. *The Journal of Finance*, 52(4):1287–1321.

Durham, J. B. (2006). An Estimate of the Inflation Risk Premium Using a Three-Factor Affine Term Structure Model. *FEDS Working Paper, No. 2006-42. Available at SSRN: https://ssrn.com/abstract=951413.*

Fama, E. F. and French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3–56.

Fontaine, J.-S. and Nolin, G. (2019). Measuring Limits of Arbitrage in Fixed-Income Markets. *Journal of Financial Research*, 42(3):525–552.

Fung, W. and Hsieh, D. A. (2002). Risk in Fixed-Income Hedge Fund Styles. *The Journal of Fixed Income*, 12(2):6–27.

Fung, W. and Hsieh, D. A. (2004). Hedge Fund Benchmarks: A Risk-Based Approach. *Financial Analysts Journal*, 60(5):65–80.

Garcia, R. and Gençay, R. (2000). Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint. *Journal of Econometrics*, 94(1-2):93–115.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT press.

Grinblatt, M. (2001). An Analytic Solution for Interest Rate Swap Spreads. *International Review of Finance*, 2(3):113–149.

Gu, S., Kelly, B., and Xiu, D. (2018). Empirical Asset Pricing via Machine Learning. Technical report, National Bureau of Economic Research.

Gu, S., Kelly, B. T., and Xiu, D. (2019). Autoencoder Asset Pricing Models. *Available at SSRN*.

Heaton, J., Polson, N. G., and Witte, J. (2016). Deep Portfolio Theory. *arXiv preprint arXiv:1605.07230*.

Hördahl, P. and Tristani, O. (2012). Inflation Risk Premia in the Term Structure of Interest Rates. *Journal of the European Economic Association*, 10(3):634–657.

Hornik, K. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural networks*, 4(2):251–257.

Hull, J. and White, A. (1990). Pricing Interest-Rate-Derivative Securities. *The Review of Financial Studies*, 3(4):573–592.

Hutchinson, J. M., Lo, A. W., and Poggio, T. (1994). A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks. *The Journal of Finance*, 49(3):851–889.

Jawadi, F. and Khanniche, S. (2012). Modeling Hedge Fund Exposure to Risk Factors. *Economic Modelling*, 29(4):1003–1018.

Karsimus, L. (2015). Yield Curve Arbitrage in the EUR Swap Rates Market. Master's thesis, Aalto University School of Business.

Kelly, B. T. and Jiang, H. (2012). Tail Risk and Hedge Fund Returns. *Chicago Booth Research Paper*, (12-44).

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer Credit-Risk Models via Machine-learning Algorithms. *Journal of Banking & Finance*, 34(11):2767–2787.

Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

Kirczenow, G., Fathi, A., and Davison, M. (2018). Machine Learning for Yield Curve Feature Extraction: Application to Illiquid Corporate Bonds (Preliminary Draft). *arXiv preprint arXiv:1806.01731*.

Klambauer, G., Unterthiner, T., Mayr, A., and Hochreiter, S. (2017). Self-Normalizing Neural Networks. In *Advances in neural information processing systems*, pages 971–980.

Kvamme, H., Sellereite, N., Aas, K., and Sjursen, S. (2018). Predicting Mortgage Default Using Convolutional Neural Networks. *Expert Systems with Applications*, 102:207–217.

Liu, J. and Longstaff, F. A. (2004). Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. *The Review of Financial Studies*, 17(3):611–641.

Liu, J., Longstaff, F. A., and Mandell, R. E. (2002). The Market Price of Credit Risk: An Empirical Analysis of Interest Rate Swap Spreads. Technical report, National Bureau of Economic Research.

Liu, J., Longstaff, F. A., and Mandell, R. E. (2006). The Market Price of Risk in Interest Rate Swaps: The Roles of Default and Liquidity Risks. *The Journal of Business*, 79(5):2337–2359.

Longstaff, F. A. and Schwartz, E. S. (1992). Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model. *The Journal of Finance*, 47(4):1259–1282.

Luss, R. and d'Aspremont, A. (2015). Predicting Abnormal Returns from News Using Text Classification. *Quantitative Finance*, 15(6):999–1012.

Malliaris, A. and Malliaris, M. (2009). Time Series and Neural Networks Comparison on Gold, Oil and the Euro. In *International Joint Conference on Neural Networks (IJCNN), 2009.*, pages 1961–1967. IEEE.

Malliaris, A. G. and Malliaris, M. (2015). What Drives Gold Returns? A Decision Tree Analysis. *Finance Research Letters*, 13:45–53.

Malliaris, M. and Salchenberger, L. (1993). A Neural Network Model for Estimating Option Prices. *Applied Intelligence*, 3(3):193–206.

Malliaris, M. and Salchenberger, L. (1996). Using Neural Networks to Forecast the S&P 100 Implied Volatility. *Neurocomputing*, 10(2):183–195.

Mitchell, M. and Pulvino, T. (2012). Arbitrage Crashes and the Speed of Capital. *Journal of Financial Economics*, 104(3):469–490.

Mönch, E. (2008). Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-augmented VAR Approach. *Journal of Econometrics*, 146(1):26–43.

Moritz, B. and Zimmermann, T. (2016). Tree-Based Conditional Portfolio Sorts: The Relation Between Past and Future Stock Returns. *Available at SSRN 2740751.*

Okunev, J., White, D., et al. (2006). Hedge Fund Risk Factors and the Value at Risk of Fixed Income Trading Strategies. *The Journal of Fixed Income*, 16(2):46–61.

Patton, A. J. (2009). Are Market Neutral Hedge Funds Really Market Neutral? *The Review of Financial Studies*, 22(7):2495–2530.

Sambasivan, R. and Das, S. (2017). A Statistical Machine Learning Approach to Yield Curve Forecasting. In *International Conference on Computational Intelligence in Data Science (ICCIDS), 2017*, pages 1–6. IEEE.

Shleifer, A. and Vishny, R. W. (1997). The Limits of Arbitrage. *The Journal of Finance*, 52(1):35–55.

Siegmann, A. and Stefanova, D. (2009). Limits to Arbitrage: Time-Varying Market Neutrality of Hedge Funds. *Available at SSRN 1471322.*

Sirignano, J., Sadhwani, A., and Giesecke, K. (2016). Deep Learning for Mortgage Risk. *arXiv preprint arXiv:1607.02470*.

Smith, D. J. (2013). Valuing Interest Rate Swaps Using Overnight Indexed Swap (OIS) Discounting. *The Journal of Derivatives*, 20(4):49–59.

Stark, L. (2017). Machine Learning and Options Pricing: A Comparison of Black-Scholes and a Deep Neural Network in Pricing and Hedging DAX 30 Index Options. Bachelor's Thesis, Aalto University School of Business.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics*, 5(2):177–188.

Vayanos, D. and Vila, J.-L. (2009). A Preferred-Habitat Model of the Term Structure of Interest Rates. Technical report, National Bureau of Economic Research.

Yao, J., Li, Y., and Tan, C. L. (2000). Option Price Forecasting Using Neural Networks. *Omega*, 28(4):455–466.

Yu, F. (2006). How Profitable is Capital Structure Arbitrage? *Financial Analysts Journal*, 62(5):47–62.