Aalto University
School of Business

Yield curve arbitrage in the EUR swap rates market
Replicating the strategies of quantitative arbitrageurs

Finance
Master's thesis
Lassi Karsimus
2015

Department of Finance
Aalto University
School of Business

Aalto University
School of Business

Yield curve arbitrage in the EUR swap rates market

Replicating the strategies of quantitative arbitrageurs

Master's Thesis
Lassi Karsimus
Spring 2015
Program

Approved in the Department of Finance __ / __ 2015 and awarded the grade

____

Author Lassi Karsimus

Title of thesis Yield curve arbitrage in the EUR swap rates market

Degree Master of Science

Degree programme Finance

Thesis advisor(s) Matti Suominen

Year of approval 2015 **Number of pages** 86 **Language** English

Abstract

OBJECTIVES OF THE STUDY:

In this thesis, I look into a hedge fund strategy known as a yield curve arbitrage, where arbitrageurs take relative value bets on interest rates. Earlier research has shown that the strategy produces favourable returns in the USD swap rates market in 1988-2004. My objective is to study whether the strategy yields attractive risk-adjusted returns and multifactor alpha in the recent period of 2002-2015 in the EUR swap rates space. I shall employ an enhanced modelling framework to implement the trading strategy. Moreover, I test the replicated strategy returns with respect to high-level and style-specific hedge fund index returns. Finally, I look into whether 'high-noise' periods in the markets coincide with large model-implied mispricing of rates. The empirical objectives of the thesis are linked to literature on yield curve formation and no-arbitrage.

DATA AND METHODOLOGY:

The dataset consists of monthly mid-market observations of constant maturity EUR swap rates for maturities of one to ten years. Also, Hedge Fund Research and Credit Suisse hedge fund index data for both high-level and style-specific indices is employed. Moreover, noise measure data by Jun Pan is employed to study the relationship of replicated returns to the level of noise. The methodology builds on Cox-Ingersoll-Ross and Longstaff-Schwartz two-factor stochastic short-rate models of interest rates. A calibration and trading algorithm is constructed based on these models to replicate the arbitrage strategy returns. Back tested trading is done explicitly out-of-the sample.

FINDINGS OF THE STUDY:

The yield curve arbitrage is found to produce attractive risk-adjusted returns and favourable return distributions. Moreover, the alpha of the strategy is statistically and economically significant when controlled by a number of commonly employed risk factors. Additionally, it is found that the replicated arbitrage strategy does not have statistically meaningful connection to neither high-level nor style-specific hedge fund indices. Finally, it is shown that high noise coincides with large model-implied mispricings when the measure of the mispricings is smoothed. No evidence is found in support of the idea that yield curve arbitrage alpha is compensation for carrying tail risk.

Keywords arbitrage, fixed income, trading strategy, hedge fund, alpha

Tekijä Lassi Karsimus

Työn nimi Korkokäyrän arbitraasi euroalueen koronvaihtosopimusmarkkinalla

Tutkinto Maisterin tutkinto

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Matti Suominen

Hyväksymisvuosi 2015 **Sivumäärä** 86 **Kieli** Englanti

Tiivistelmä

TUTKIMUKSEN TAVOITTEET:

Tässä lopputyössä tutkin vipurahastostrategiaa, joka tunnetaan korkokäyrän arbitraasina, missä arbitraasitoimijat tekevät suhteelliseen arvoon perustuvia sijoituksia koroissa. Aikaisempi tutkimus on näyttänyt, että strategia tuottaa hyvin dollarikorkomarkkinalla aikavälillä 1988–2004. Tavoitteenani on tutkia, tuottaako strategia mielenkiintoisia riskikorjattuja tuottoja ja monifaktori-alphaa tuoreella 2002–2015 aikavälillä eurokorkomarkkinalla. Käytän edistynyttä mallinnuskehikkoa strategian toteuttamiseen. Lisäksi vertaan mallinnetun strategian tuottoja erilaisiin vipurahastoindekseihin. Viimeisenä tutkin, havaitaanko markkinalla korkea hinnoitteluvirhe samaan aikaan kun mallini näyttää korkeaa hinnoitteluvirhettä. Lopputyön empiiriset tavoitteet sidotaan kirjallisuuten koskien korkokäyrän muodostusta ja arbitraasivapaata hinnoittelua.

DATA JA METODOLOGIA:

Data koostuu kuukausittaisista vakiopituisten euro-määräisten koronvaihtosopimusten korkotasoista, joiden maturiteetti on yhdestä kymmeneen vuotta. Tämän lisäksi käytän Hedge Fund Researchin ja Credit Suissen dataa koskien sekä yleisen tason että tyylikohtaisia vipurahastoindeksejä. Lisäksi käytän dataa liittyen Jun Panin kehittämään hinnoitteluvirhe-estimaattoriin tutkiakseni havaitun ja mallin näyttämän hinnoitteluvirheen yhteyttä. Metodologia perustuu Cox-Ingersoll-Rossin ja Longstaff-Schwartzin kahden faktorin stokastisiin lyhyenkoron malleihin. Rakennan kalibrointi- ja kaupankäynti-algoritmin perustuen näihin malleihin, jotta voin muodostaa arvion arbitraasituotoista. Tuotot muodostetaan erityisesti siten, että kaupankäynnin mallintamisessa ei käytetä tulevaisuuden dataa.

TUTKIMUKSEN TULOKSET:

Korkokäyräarbitraasin havaitaan tuottavan hyvää riskikorjattua tuottoa sekä suosiollisia tuottojakaumia. Lisäksi strategian alpha on tilastollisesti ja taloudellisesti merkittävä, kun sitä kontrolloidaan yleisesti tunnetuilla riskitekijöillä. Tämän lisäksi havaitsen, että arbitraasistrategialla ei ole merkittävää tilastollista yhteyttä korkean tason tai tyylikeskeisiin vipurahastoindekseihin nähden. Lopuksi näytän että korkea yleisesti havaittu hinnoitteluvirhe tapahtuu markkinalla samaan aikaan kun mallini näyttää korkeaa hintavirhettä, ottaen huomioon että hintavirhe on tasoitettu. En löydä tukea väitteelle, että korkokäyräarbitraasin alpha olisi kompensaatiota harvinaisista romahduksista.

Avainsanat arbitraasi, korkomarkkina, kaupankäyntistrategia, vipurahasto, alpha

Table of contents

List of figures

List of tables

1. Introduction

Hedge funds and other arbitrageurs engage in many trading and investment activities of highly varying nature. One of the most interesting categories of hedge fund styles is ‘arbitrage’ strategies, as they supposedly yield positive returns in all market environments, whether the general market is in a bull or a bear mode. As the term ‘arbitrage’ implies, these strategies are hedged in some way, and are thus meant to carry a relatively low risk in terms of e.g. volatility. Although some managers may be looking for the kind of textbook arbitrage, where an arbitrageur would generate genuine riskless profits with no initial capital, an ‘arbitrage’ in the in industry parlance usually refers to trades that are made neutral to changes in the key market variables. Such strategies do carry risk, yet the risks may be associated with unconventional premia with attractive characteristics.

Out of these arbitrage strategies, fixed income arbitrage styles are highly intriguing, as many of them are fairly complex, reducing competition in the area. Less competition generally leads to higher risk-adjusted returns for those involved, as implied by the weakened performance of a number of increasingly competed hedge fund strategies, such as convertible arbitrage, as illustrated by e.g. Agarwal et al. (2011). Moreover, earlier literature has indeed verified that many such strategies yield large multifactor alpha, high risk-adjusted returns by numerous measures, and are little correlated with the common risk factors and high-level hedge fund indices. Duarte, Longstaff and Yu (2007) show that yield curve arbitrage is one the most profitable strategies in the space of well-known fixed income arbitrage strategies. In essence, the yield curve arbitrage is a relative value trading strategy in the space of government debt or related interest rates. Hence, the strategy is about identifying overtly rich and cheap points on the yield curve with the assumption that these mispricings convergence in the near future, so that they can be traded profitably. Duarte et al. attribute the performance of this arbitrage mostly to the complexity of the analytical methodology necessary to implement a sophisticated strategy in this niche. Higher sophistication calls for additional human capital to employ a relevant modeling framework to both generate the trading signals and to hedge the bets.

This thesis builds on the theoretical framework of Vayanos and Vila (2009), who contemplate a preferred-habitat model of the term structure of interest rates. In their model, the yield curve is formed through the interaction of bond suppliers, investors and risk-averse arbitrageurs. Both demand and supply shocks can affect the term structure when arbitrageurs are risk-averse, yet it is the role of the arbitrageurs to render the yield curve arbitrage free. As the arbitrageurs shall impose the no-arbitrage condition through trading, their positions will in equilibrium carry risk associated with the bond risk premia. The arbitrageurs are explicitly assumed to observe an exogenous short-rate of interest, based on which they choose to dynamically take long and opposing short positions in the bonds and in the short-rate. This framework is a natural building block for this thesis, given that it explicitly calls for the arbitrageurs, and those arbitrageurs are assumed to incorporate expectations of the future short-rate into the bond prices. Similar to Duarte et al. (2007), as well as to this thesis, Vayanos and Vila let the short-rate follow a stochastic mean-reverting process. Building theoretically on the model by Vayanos and Vila, the agenda of this thesis is to model the strategies of those yield curve arbitrageurs who employ a quantitative approach to initiate and hedge the arbitrage trades. Vayanos and Vila also directly point out that the kind of risky yield curve arbitrage their arbitrageurs are thought to perform is similar to those strategies executed by hedge funds and proprietary-trading desks.

While the idea of generating excessive returns from trading liquid fixed income instruments certainly sounds lucrative, the only paper directly related to the replication of such strategies remains the one by Duarte et al. (2007). In their paper, the authors model the yield curve by employing Vasicek two-factor short-rate model, which they calibrate to the market curve. Based on the implied differences between the market and the model rates, they engage in making two-factor neutral long/short trades on the mispriced rates. Such trades are called the yield curve arbitrage, pertaining mostly to the fact of extensive market neutrality.

As to the contribution of this thesis, I set out to further develop the methodology outlined by Duarte et al. (2007) with a different and more recent data. For one, I am using the two-factor models by Cox, Ingersoll and Ross (CIR two-factor model, CIR2F) and Longstaff and Schwartz (Longstaff-Schwartz two-factor model, LS2F) to model the yield curve. I conduct the analysis of the arbitrage strategies with EUR swap rate data in the range of years 2002 and 2015, while Duarte et al. looked at the USD space during an earlier period of 1988 to 2004. Moreover, I study different trading strategies as logically implied by the models. Furthermore, throughout the thesis, I shall have an explicit out-of-the-sample focus in implementing the back test of the trades. Duarte et al.

acknowledge that their reported strategy is actually calibrated to the whole sample, including future information.

After replicating the arbitrage strategies, I carry out regressions to isolate the multifactor alpha of the strategies. I study the replicated returns with respect to the high-level hedge fund indices, as well as to the more specific style subindices. Finally, I employ the noise metric suggested by Hu et al. (2013) to study if high noise coincides with high model implied mispricings. Also, I look into whether information in the noise measure can be leveraged upon to produce better trading outcomes. The tests regarding the hedge fund subindices and the noise measure are included in the contribution of this paper.

I hypothesize that the yield curve arbitrage generates attractive risk-adjusted returns also in the EUR rates market in this more recent sample period. Additionally, I hypothesize that the return distributions have favorable characteristics. Secondly, I hypothesize that the strategies have a limited exposure to priced risk factors, i.e. that the multifactor alpha is positive and significant both economically and statistically. Also, I hypothesize that the arbitrage strategies I replicate explain to a significant extent the most relevant hedge fund subindex returns (e.g. the fixed income arbitrage style indices). Further, I make the hypothesis that the mispricings of the rates in the market are greater when there is a lot noise as quantified by the Hu et al. (2013) measure. Extrapolating this thought, I hypothesize that returns from the strategies are greater during the noisy periods.

I find evidence strongly in support of the hypotheses pertaining to the attractiveness of the returns. The yield curve arbitrage has continued to generate attractive risk-adjusted returns, and the phenomenon is replicated in the EUR swap space with the recent sample period from 2002 to 2015. The returns are basically pure multifactor alpha, and have high Sharpe and Gain-Loss ratios, as well as distributions that are heavily positively skewed with fat tails. Strong evidence is also found in support of the yield curve arbitrage yielding multifactor alpha. Moreover, the replicated returns are at best vaguely correlated with any of the considered hedge fund indices, implying that the returns are also ‘hedge fund alpha’. Finally, it is shown that the smoothed model implied mispricings correlate heavily with the Hu et al. (2013) noise measure, and that leveraging in the high-noise market environments does weaken the risk-adjusted trading performance.

As the volatility of the strategies is low, they can and need to be leveraged; Duarte et al. leverage their strategies so that the ex post volatility is 10% annually. As the strategies are implemented with swaps, scaling is straight forward and cost-effective, as one only needs to choose a higher amount of notional for a given trade. Thus, the leveraged annual returns are in the magnitude of 10%, which is economically very satisfactory for an annualized 10% volatility level. Thereby it seems that the risk-taking yet risk-averse arbitrageurs theoretically depicted in Vayanos and Vila (2009) do indeed enjoy attractive risk-adjusted returns, as suggested by their model.

The rest of the thesis is organized as follows. Section 2 presents the motivation and background for the ideas developed in this paper, including contribution and limitations of the study. Section 3 elaborates on the literature and theoretical framework of the ideas considered. Section 4 presents the hypotheses to be tested in the study, and Section 5 illustrates the data employed and goes through the methodology in detail. Section 6, 7, 8, and 9 present the results of the thesis, including conclusion regarding the hypotheses stated in Section 4. Section 10 concludes the thesis and its key findings.

2. Motivation and background

The idea behind studying the yield curve arbitrage as a trading strategy comes from the notion that some points of the term structure of interest rates may not at all times be in sync with each other. The yields for different maturities are not determined independently; they are all linked across the yield curve, as shown by e.g. Cox, Ingersoll, and Ross (1985). As pointed out in earlier research by Duarte, Longstaff and Yu (2007), trading of the rates that are out of the line with each other can result in highly attractive return profiles. No-arbitrage should guarantee that the yield curve is internally consistent, i.e. that all the forward rates are unique, and no 'textbook arbitrage' is possible. Nevertheless, the 'mispricings' on the yield curve arise from the nearby bond maturities trading at prices dissimilar enough. In other words, arbitrage opportunities related to the bond risk premia can exist in spite of the uniqueness of all forward rates. Such arbitrage opportunities are theoretically depicted in the preferred-habitat model of interest rates by Vayanos and Vila (2009), and the strategies were sought to be replicated in Duarte et al. (2007). Moreover, Vayanos and Vila

(2009) discuss that it is the kind of premia-driven arbitrage that the hedge funds and proprietary trading desks typically engage in practice.

As shown by Duarte et al. (2007), yield curve arbitrage in the USD space was highly profitable in the period from 1988 to 2004. Given the tide of material developments and events since 2004, it is certainly necessary to update the view on the attractiveness of the yield curve arbitrage strategy. This is logical given the pace of evolution in the hedge fund industry, as well other developments in finance from the rise of electronic trading to the increased regulation. Moreover, ever since 2004, the world has seen both the building and bursting of a housing and credit bubble, the advent of highly unorthodox central bank policies, ultra-low global rates, and the pre-crisis tightening cycle. In Europe in particular, we have witnessed escalation of the sovereign debt crisis, the near break-up of the Eurozone, the beginning of Outright Monetary Transaction, as well as an early ECB credit tightening, and finally the prospect of and advent of the ECB Quantitative Easing. In all, all the developments would warrant the studying of yield curve arbitrage in their own right. What is more, the strategies are yet to be tested in the EUR rates space, so I chose to combine studying the fresh time period with expanding the study to the euro space.

As a starting point in approaching the problem of arbitrage trading profits, the question is how exactly is the yield curve formed, and how does its construction allow for the risk premia related arbitrage that carries a risk. The theory on yield curve formation through no-arbitrage is built on Vayanos and Vila (2009) and Greenwood and Vayanos (2010). The empirical implementation is based on Duarte et al. (2007), and building on their work, I hypothesize in this thesis that the premia-related yield curve arbitrage generates attractive returns also in the EUR space. This is equivalent to saying that the term structure is not consistently priced at all times, and there is a role for arbitrageurs in enforcing the mispriced rates back into the line with the rest of the curve.

The specific purpose of the paper is, on the lines of Duarte et al., to model the behavior of the yield curve arbitrageurs – hedge funds, for instance – in order to get a picture of both what such arbitrageurs might be doing, and what are the characteristics of the returns their trading strategies yield. In essence, this thesis seeks to combine the theoretical models of yield curve arbitrage with the empirical replication methodology to study and explain the returns available for arbitrageurs in the EUR rates space.

Contributions to the literature include the expansion of the Duarte et al. (2007) study to the EUR swap space. Moreover, the time period is more recent, from 2002 to early 2015, as compared to the 1988 to 2004 period. Further, I employ somewhat more realistic two-factor models, namely the Cox-Ingersoll-Ross and Longstaff-Schwartz two-factor models. Additionally, I follow an explicit out-of-the-sample focus in trading, which is in contrast to the reported strategies in Duarte et al. Finally, the contribution includes testing of the strategy sensitivities to slight modifications, comparison of the replicated strategies to the most relevant hedge fund subindices, and application of the Hu et al. (2013) noise measure in explaining the mispricings, as well as in the implementation of the strategies.

As to the limitations of this paper, the focus is on the EUR constant maturity swap rate space, which differs from the definition of a vanilla interest rate swap. Therefore, strategy implementation with the vanilla swaps might differ from what is covered in this paper. The data employed consist of mid-swap rate quotes, which may be only partially executable in practice, given that a trader may have to trade at bid and ask prices occasionally, or even most of the time. In general, taking liquidity from the market by hitting existing best bids and offers would be detrimental to an arbitrageur, yet given the mature state of the swap market today, and hence the tight spreads, the effect to profit-and-loss should remain marginal.

The methodology with respect to the chosen short-rate models faces limitations in that it does not aim to cover all the relevant short-rate models, which are plentiful in the literature. Moreover, as the models are numerically calibrated to the data through analytical solutions for zero-coupon bond prices, a lot of the results may be sensitive to deliberate choices in calibration. At the end of the day, calibration is a blend of art and science, as the complexity of solving certain minimization problems, among other things, is an iterative process with no single best practice available.

Regarding the performance of the replicated strategies, two points are worth noting. First, no transaction costs are included, and second, the returns are before fees. When compared to the actual hedge fund returns, for instance, the reader should bear in mind that the reported returns by funds are usually after both performance and management fees, which can be substantial (often

20% and 2%, respectively). As a caveat regarding the ignorance of transaction costs, I note that the vanilla swap market is so mature and intensely competed that the associated bid-ask spreads should have an insignificant effect on the returns for a serious player in this space.

Considering the rest of this paper, I shall begin Section 3 by illustrating the above mentioned theory by Vayanos and Vila (2009) about term structure formation that calls explicitly for the arbitrageurs. This is important in order to understand the framework in which the work of the yield curve arbitrageurs happens in theory. After that, literature in general is review as relevant to the ideas in this thesis. Section 4 states the hypotheses, while Section 5 elaborates on the methodology and illustrates related key concepts. Sections 6 to 9 discuss findings of the thesis, as well as tests of the hypotheses. Finally, Section 10 concludes the agenda, context and findings of this paper.

3. Theory and literature

3.1. Formation and integration of the yield curve

As the dynamics of the yield curve are at the heart of this thesis, it is a key to have a solid economic framework for how the curve takes the shape and characteristics perceived in the market place. Vayanos and Vila (2009) offer a robust and tractable theoretical framework in this regard in their preferred habitat model of the term structure of interest rates. Their framework is more compatible with the recent empirical evidence from the market than the previous theories about the yield curve, and interestingly, their model explicitly assumes a role for the arbitrageurs. As the point of my research is to replicate the behavior and trading returns of such arbitrageurs, it is helpful to first take a glance at a high-level model illustrating the formation of term structure of interest rates through the workings of supply, demand and no-arbitrage.

In their work, Vayanos and Vila (2009) discuss the drivers for interest rates for different maturities. First, they point out that in standard economic theory interest rates for certain maturities can be explained by the willingness of agents to substitute consumption between now and the maturity of the bonds in question. After concluding that this approach would not explain e.g. the

U.S. Treasury's 2000-2002 bond buyback program's consequences, they look into the so-called preferred-habitat model for interest rates that was originally proposed by Culbertson (1957) and Modigliani and Sutch (1966). As the name implies, the idea in this view is that the interest rate for a particular maturity is driven by the supply and demand shocks local to the specific maturity. As pointed out by Cox, Ingersoll and Ross (1985), no-arbitrage conditions do not allow interest rates for different maturities to be independently determined, as they should be correlated with the nearby rates. Vayanos and Vila (2009) go on to develop a model where the yield curve is determined by the interaction between investor clienteles and risk-averse arbitrageurs. In the model, the investors have preferences for certain maturities (e.g., pension funds for longer and asset managers for shorter maturities), while the arbitrageurs' role is to render the yield curve arbitrage-free though trading.

No-arbitrage means that all extractable forward rates are unique, i.e. that the term structure is internally consistent, precluding a simple 'textbook arbitrage'. Given that the yield curve is swiftly rendered arbitrage free, in equilibrium the arbitrageurs' positions will carry risk related to the bond risk premia. In spite of the uniqueness of all forward rates, risky 'arbitrage' opportunities may arise, should the nearby bonds trade at prices (yields) dissimilar enough with respect to the nearby maturities. This kind of arbitrage situations could be labeled 'mispricings' to distinguish from the traditional textbook arbitrage. The arbitrageurs in Vayanos and Vila (2009) are able to locate such mispricings by observing the stochastic mean-reverting short-rate, similar to the ones employed in this thesis and in the paper by Duarte et al. (2007). By observing the short-rate, arbitrageurs are able to incorporate information about current and expected short-rates into the bond prices.

As the arbitrageurs are assumed to be risk-averse, the demand shocks inflicted by the investor clienteles (or the supply shocks caused by the issuers) affect the yield curve, given that the arbitrageurs cannot absorb the demand shocks perfectly. Thus, the yield curve is determined by both the supply-demand shocks as well as the exogenous stochastic short-rate, which can be affected by e.g. the Central Bank policy or macroeconomic conditions. As the investors are only interested in trading their preferred maturity, they are not interested in the short-rate; it is the arbitrageurs who can dynamically trade between the bonds and the short-rate. By doing this, the arbitrageurs integrate the maturity markets by making the mispricings to converge.

In the framework of Vayanos and Vila, arbitrageurs exercise carry trades based on their view on the evolution of the short-rate (the instantaneous interest rate for an infinitesimal time period). When the short-rate is seen rising, arbitrageurs will short bonds and invest at the short-rate (roll-up carry trade), as the bonds are expected to have a negative premia when the yield curve is inverting. Conversely, when they see the short-rate dropping, they shall go long bonds, borrowing at the short-rate (roll-down carry trade). This trading by the arbitrageurs provides the mechanism through which the yield curve is finally rendered consistent in terms of bond risk premia, as yields are adjusted to reflect changes in the current and expected future short-rates. In the absence of the arbitrageurs, the yield curve would be disconnected from the stochastic short-rate, and moreover, it would remain constant, should the investor clienteles' demands be constant over time.

Given that the carry trades executed by the arbitrageurs are not riskless, arbitrageurs will trade only when they assess that there are positive expected returns available from the arbitrage opportunities that compensate for the risks involved. Moreover, in the case where the bond prices are determined by both the short-rate and demand factors, the arbitrageurs will hedge their bets by taking offsetting long and short bond positions. For instance, a shock pushing the short-rate up will make the arbitrageurs go short the short-term bonds, and long long-term ones to hedge duration risk. This would have the effect of pushing short-term rates up and long-term rates down.

As further discussed in Greenwood and Vayanos (2010), the arbitrageurs basically buy bonds with low clientele demand (high rates) and sell bonds with high clientele demand (low rates), ensuring thereby that nearby maturities trade at similar prices. The arbitrageurs face fundamental and non-fundamental risks in bridging the maturity markets; the fundamental one being the changes in the short-rate, and the non-fundamental being the shocks to the demand for bonds with particular maturities. Arbitrageurs receive compensation for taking these risks. As the arbitrageurs intermediate between bonds of different maturities, the bonds most sensitive to the demand shocks are the ones most sensitive to the market price of short-rate risk. When arbitrageurs are particularly risk-averse, multiple risk factors become relevant; if there are numerous such factors, arbitrageurs are less able to integrate the maturity markets, given the increasingly complex hedging requirements. As a consequence, the demand effects shall become more local.

The work of Vayanos and Vila (2009), and Greenwood and Vayanos (2010), offer an intriguing theoretical backdrop for the ideas developed in this thesis, as the point in my work is indeed to model the trading strategies of the arbitrageurs that these papers depict on a general level. If their theoretical assumptions are correct in that arbitrageurs are necessary for maintaining the consistency of the yield curve, and that arbitrageurs demand positive expected returns, it follows from this that the yield curve arbitrage strategies I employ should describe in some form the returns available from being such an arbitrageur. Moreover, Vayanos and Vila are very explicit in the risk-averse nature of the arbitrageurs. Linking this idea to the limits to arbitrage literature (more of which in the later section), one can logically move to the conclusion that the more risk-averse the arbitrageurs are, and the more they have limitations (e.g., arbitrageurs have limited capital), the less they will exercise the kind of yield curve integration trades described by Vayanos and Vila.

Greenwood and Vayanos (2014) find that low arbitrageur wealth makes the government bond supply and the yield curve slope stronger predictors of the future returns. This means that after the arbitrageurs have endured losses, they will be weaker in integrating the yield curve. Resulting from the increased limits to arbitrageurs' trading, mispricings on certain points of the yield curve should be larger and exist for longer, thereby possibly offering more attractive risk-adjusted returns to those arbitrageurs able to exercise their strategies. On the other hand, consistent with Vayanos and Vila (2009) and Greenwood and Vayanos (2014), it is also possible that low arbitrage capital will obstruct the convergence of the mispricings, given that shocks will be absorbed less efficiently.

3.2. Yield curve arbitrage methodology

Duarte, Longstaff, and Yu, in their paper "Risk and Return in Fixed Income Arbitrage: Nickels in Front of a Steamroller?" (2007, Review of Financial Studies), bring seminally light into the space of how hedge funds and bank proprietary trading desks presumably approach the modeling of fixed income arbitrage opportunities. Moreover, they study the characteristics of the resulting returns, which turn out to be highly favorable. Duarte et al. look at five strategies: swap spread, yield curve, mortgage, volatility and capital structure arbitrage. Out of these, the yield curve arbitrage is among the top three strategies that require more human capital to implement, which presumably explains the stronger risk-adjusted returns they generate.

Duarte et al. approach the yield curve arbitrage modeling the same way they think several large hedge funds view the problem: they employ a two-factor short-rate model to depict the yield curve. Namely, the model they use is Vasicek (1977) two-factor model. The model has two sources of uncertainty, meaning that it is built upon two mean-reverting stochastic processes. It has a well-known analytical solution for zero-coupon bond prices, i.e. for the discount factors. By turning swap rates into discount factors, the authors can match the model prices to the market prices. They use monthly USD swap rate data from 1988 to 2004 for swap maturities of one, two, three, four, five, seven and ten years. Section 5 in this thesis explores more carefully how discount factors, swaps, swap rates and the short-rate models play together in this context.

The authors make their short-rate model match exactly the 1-year and 10-year swap rates for each month, so that the two, three, four, five and seven year rates are fitted closely to the market rates but not necessarily exactly. The matching of model rates to market rates is achieved through a calibration procedure, essentially by minimizing the sum of squared differences between the market and the model rates. The calibration methodology is at the very core of the whole modeling exercise, so I will restate below how Duarte et al. iteratively calibrate the model.

To begin with, the Vasicek two-factor model has two factors, x_1 and x_2, as well as six parameters, three for each mean-reverting process; these can be described as mean-reversion speed of the short-rate (k), long-run mean of the short-rate (theta) and the instantaneous volatility of the short-rate (sigma). The model shall be introduced below, in Equation 1, where I present the analytical solution for the price of a zero-coupon bond in the model, as well as the underlying stochastic short-rate processes.

Equation 1 (Vasicek two-factor model zero-coupon bond price)

Vasicek two-factor model is built upon the two mean-reverting stochastic differential equations for x_1 and x_2, where dW_i terms are standard Brownian motions

$$dx_1(t) = k_1(\theta_1 - x_1(t))dt + \sigma_1 dW_1(t)$$

$$dx_2(t) = k_2(\theta_2 - x_2(t))dt + \sigma_2 dW_2(t)$$

A discount factor, or a zero-coupon bond, with a maturity T, as seen at time t, is priced by

$$DF(t,T) = A_1(t,T)A_2(t,T)e^{-x_1(t)B_1(t) - x_2(t)B_2(t)} \qquad (1)$$

Where,

$$B_i(t,T) = \frac{1 - e^{-k_i(T-t)}}{k_i}$$

$$A_i(t,T) = exp\left\{\left(\theta_i - \frac{\sigma_i^2}{2k_i^2}\right)(B_i(t,T) - T + t) - \frac{\sigma_i^2}{4k_i}B_i^2(t,T)\right\}$$

$i = 1, 2$ standing for the two factors.

Using the above analytical formulation, Duarte et al. calibrate the model to the market data by initially choosing trial values for the six parameters. The authors do not reveal what these values are, although it is often of relevance to the calibration process. After picking the trial values, the model is made to fit the 1-year and 10-year market rates exactly by choosing the factors x_1 and x_2 appropriately. After that, the sum of squared differences between the model and market rates is minimized within the whole sample period for the two, three, four, five, and seven year rates. Essentially, this means finding the six parameters that minimize the distance between the model and market rates jointly for the whole sample period for the ‘illiquid rates’.

The sample period for model calibration that Duarte et al. use is the whole 1988-2004 period for which they have monthly data. As they discuss, this would appear to create a look-ahead bias, which the authors, however, contest by saying the parameters are only used for computing the hedge ratios (which include the price of the zero-coupon bond). Moreover, they also calibrated the model to an in-the-sample period and tested it on a separate out-of-the-sample period, acquiring thereby similar results as with using the entire period as the sample for calibration.

After the six parameters are fixed after this calibration, the modeling continues so that for each month going forward, the values of the factors x_1 and x_2 are again chosen so that the respective 1-year and 10-year market rates are exactly fit to the model rates. Hence, the model shall imply market-to-model rate differences, or 'mispricings', for the rest of the points on the yield curve, namely for the 'illiquid rates' that are not fitted exactly. These mispricings are then considered as trading signals. Naturally, the arbitrageur goes short rich and long cheap market swap rates on the points of the curve where the mispricings are deemed large enough. After establishing positions, the trader waits for the market rates to converge to the model rates. Should this happen, positions will be unwound. The methodology shall be further discussed in Section 5.

3.3. Performance of yield curve arbitrage

Duarte et al. find strong evidence that yield curve arbitrage strategies generate monthly excess returns on the scale of 43.7 to 61.5 basis points after leveraging the positions so that they have an ex post annualized volatility of 10%. Leverage is hence around 16 times the equity in the respective swap trades. The excess returns are statistically significant as implied by the Newey-West (1987) autocorrelation robust test statistic, which attains a value of 3.42 for the equally weighted (EW) portfolio that trades all the rate maturities.

The authors look into strategies where only a single swap maturity is traded, as well as into an equally weighted portfolio made of all the maturities. The single-maturity strategies trade 2-year, 3-year, 5-year and 7-year swap rates. The monthly leveraged mean returns for these strategies are 54, 48.6, 61.5 and 43.7 bps, respectively, while the EW portfolio yields 51.9 bps. Sharpe ratios range from 0.524 to 0.738 in the single-swap trades, while the EW strategy has the highest Sharpe ratio at 0.785.

The returns from the strategy are highly positively skewed (0.995 for the EW strategy), which is in contrast to the common perception that fixed income arbitrage equals to collecting small returns most of the time, while suffering large losses occasionally (i.e., ‘picking nickels in front of a steamroller’). Kurtosis ranges for the EW portfolio is 3.269, meaning that the return distribution has fat tails, i.e. that extreme values for returns are more likely than is implied by the normal distribution. While positive skewness is certainly good for the arbitrageur, the favorability of heavy tails is more contestable, given that it may also increase exposure to extreme losses. Only 34.7% of the returns fall into the negative territory as reported by the authors.

Given the non-normality of the return distributions, Bernardo and Ledoit (2000) Gain-Loss ratios are also reported. These range from 1.643 to 2.355, the EW strategy having a ratio of 1.980, implying highly favorable performance. The ratio is basically computed as the ratio between the expected gain and the expected loss.

Yield curve arbitrage yields a highly statistically significant alpha of 59.8 bps per month before fees, the test statistic being 3.14, when the returns are controlled by priced risk factors, including Fama-French four factors, as well as rate and credit factors. R^2 of the regression is only 9.7%, implying that the beta factors explain little of the variation in the strategy’s returns. In contrast to the other fixed income arbitrage strategies discussed in Section 3.4., yield curve arbitrage has statistically significant alphas for all the portfolio modifications considered by Duarte et al. (2007). Finally, the yield curve arbitrage strategy appears to have no statistically meaningful connection to the actual high-level hedge fund indices, as Duarte et al. report a correlation coefficient close to zero between the two.

3.4. Performance of other arbitrage strategies

Mitchell and Pulvino (2001) study the risk arbitrage (merger arbitrage) strategy, where the arbitrageur usually takes a long position in the M&A target company while selling short the acquirer as a hedge. In the dataset spanning years from 1963 to 1998, the authors find that the strategy generated (annualized) Sharpe ratios from 0.44 to 1.0. The linear model alphas in the absence of transaction costs were 74 bps per month. Mitchell et al. liken the strategy to selling index put options, as the strategy does correlate positively with the general market during severe declines, yet earns stable returns most of the time. This characteristic also leads to the questioning of the Sharpe ratio as a prudent measure of risk-adjusted performance. This point shall be further discussed in Section 6 that delves deeper into the analysis of results.

Gatev et al. (2006) look into a classical Wall Street relative value, or statistical arbitrage strategy known as pairs trading. The strategy takes long/short positions in stocks that are historically cointegrated, i.e. have a long-run mean to which their spread convergences to. Stocks are matched to such pairs by minimizing the distances of normalized historical prices. The authors test a simple trading rule in a period 1962 to 2002, where the strategy yields up to 11% in annualized excess returns. The return distribution has a positive skew. The strategy generates Sharpe ratios that are four to six times higher than that of the market; the aggregate strategy has an annualized ratio of 1.56. As the excess returns are fairly robust, it seems that the strategy profits from the temporary mispricing of close substitutes. The returns are finally linked to a common factor that is uncorrelated with the conventional risk factors.

Yu (2006) studies capital structure arbitrage, where an arbitrageur takes off-setting positions in a company's debt and equity based on the trading signals generated by a Merton (1974) style structural model. Trades are entered into when a certain market-to-model spread threshold is exceeded, and similarly the trades are closed when the gap tightens, enough of losses accumulate, or if a time limit is reached. The trades are executed as credit default swap (CDS) positions versus stock positions. A core finding is that the capital structure arbitrage is very risky given a high drawdown potential. The most promising version of the strategy with a relatively high bound for trading yields an annualized Sharpe ratio of 1.54, yet the maximum loss in a month can be as high as 33%. Most of the losses occur when the arbitrageur is short the CDS and the spread soars abruptly, making the equity hedge ineffective. Strategy modifications that have less restrictive bounds trade more often, but also have much lower Sharpe ratios, ranging from 0.11 to 0.39. The returns have a weak positive statistical connection to the CSFB/Tremont Fixed Income Arbitrage index.

Duarte et al. (2007) also look into swap spread, volatility, capital structure and mortgage arbitrage. Equally weighted swap spread arbitrage strategy, where an arbitrageur trades swaps against Treasuries, has a monthly excess return of 41 bps and a Sharpe ratio of 0.597. Returns are negatively skewed and 39.4% of returns are negative. Gain-Loss ratio stands at 1.643 for the EW strategy. Mortgage arbitrage, where an arbitrageur isolates the prepayment risk, has an excess return of 40.8 bps and a Sharpe ratio of 0.514. Returns have a large positive skew of 6.369, yet 39.2% of the returns are negative. Gain-Loss ratio is reported at 1.489. Fixed income volatility arbitrage, where an arbitrageur basically sells rate volatility, has a monthly excess return of 58.4 bps for the EW strategy, with a Sharpe ratio of 0.72, and a Gain-Loss ratio of 1.709. The returns are not, however, statistically significant as the test statistic stands at only 1.79. Just 34.4% of returns are negative, yet the returns are negatively skewed, and the absolute minimum return exceeds the absolute maximum return. Again, capital structure arbitrage, where an arbitrageur forms a long/short portfolio of company's debt and equity, is reported to yield 70.5 bps per month with a Sharpe ratio of 1.203, and a Gain-Loss ratio of 4.117. Skewness is positive at 2.556, and the tails are very heavy with kurtosis at 8.607. 33% of the returns are negative. In spite of otherwise solid statistics, the returns are not statistically significantly different from zero, as the t-statistic is only 1.70 for the EW strategy. Out of these fixed income arbitrage strategies, only capital structure arbitrage has statistically significant alpha when controlled by the priced risk factors.

Avellaneda and Lee (2008) look into typical statistical arbitrage strategies in the U.S. equities market, and find that Principle Component Analysis (PCA) based strategies have an average annual Sharpe ratio of 1.44 in a period from 1997 to 2007. The performance is much stronger prior to 2003; after that the average ratio was 0.9. Exchange-traded fund (ETF) based strategies yielded a Sharpe ratio of 1.10 from 1997 to 2007, yet experienced a similarly degrading performance post-2002. Avellaneda et al. also find that taking into account daily trading volume information in the trading signals, the Sharpe ratio strengthens to 1.51 for the period 2003 to 2007 for the ETF-based strategies.

Agarwal et al. (2011) take a glance into common convertible arbitrage strategies, where an arbitrageur usually goes long convertible bonds and delta hedges the directional equity risk away. In terms of risk-adjusted performance, their synthetic hedge fund portfolios yield Sharpe ratios between 0.30 and 0.62. Basic OLS regression implies a significant monthly alpha of 40 bps, but controlling for the convertible bond supply effects, the alpha become negative, yet stays significant. This would suggest that the supply shock risk is priced in the convertibles market, as liquidity shocks (e.g. the LTCM event) are seen to hit the CB arbitrage funds rather hard.

3.5. Possible explanations for performance

Possible explanations as to why the yield curve arbitrage results in attractive returns include its analytical complexity, as suggested by Duarte et al. (2007), as well as challenges in implementation, especially the calibration. This explanation pertains to limited competition among the arbitrageurs; barriers to entry are relatively high due to extensive and costly human capital investments, resulting in a limited number of sophisticated hedge funds who resort to the strategies through the advanced analytical frameworks.

Other explanations are related to the efficiency of the markets and limits of arbitrage. Generally, in an efficient market, there are no limits to arbitrage, which itself makes the market efficient. Sophisticated arbitrageurs have unlimited firepower and frictions do not restrict trading, which will instantly correct any mispricings of securities. Today, there is an extensive pile of leading research on the limitations that the arbitrageurs face. These limitations are numerous and come in all flavors, yet one can think of them as being either exogenous to the arbitrageurs – like a cap on the leverage available – or choices made by the arbitrageurs themselves to maximize e.g. their expected risk-adjusted returns (or utility of wealth). In the ensuing subsections, I shall delve more in depth into these ideas that lend themselves as possible explanations for the superb performance of the yield curve arbitrage strategies. It should be noted that by and large, these possible explanations would be relevant for most fixed income or other arbitrage strategies; they are not unique to the yield curve arbitrage. On the footsteps of Duarte et al. (2007), I consider analytical complexity as a key differentiator between the yield curve arbitrage and most other strategies ranging from merger arbitrage to swap spread arbitrage.

3.5.1. Complexity of analytical methodology and model risk

Duarte et al. (2007) suggest that yield curve and capital structure arbitrage outperform other fixed income arbitrage strategies, as well as other less complex trading strategies due to their increased analytical complexity and thus the need to have additional human capital. As the yield curve arbitrage is a quantitative and systematic strategy employing little to no discretionary views of the arbitrageur, the quality of and insights in modeling are naturally key performance drivers. The complexity and innovativeness of the analytical methodology can act as a differentiator between hedge funds competing in this niche. Moreover, the number of hedge funds and other arbitrageurs competing in the space is limited by the supply of human capital. The arbitrageurs need to have the capacity to employ relevant multifactor interest rate models for pricing and hedging, as well as numerical methods necessary in calibrating the models.

Additionally, a trading strategy driven by mathematical and numerical methodologies is exposed to the risk that such methods end up losing their historical edge, or fail to hedge the risks appropriately. Thus, it appears logical that actual quantitative arbitrageurs would reserve dry powder for potentially large failures in terms of modeling. This would leave some of the perceived trading opportunities unharnessed and thus also alpha on the table.

In all, the necessary human capital investments and model risks associated with running the quantitative yield curve arbitrage strategy may act as constraints to driving the available alphas towards zero. Hence, those capable of running the respective strategies well enough may continue to be rewarded disproportionally in terms of risk-adjusted returns.

3.5.2. Limits of arbitrage

Shleifer and Vishny (1997) discuss the difference between textbook and real world arbitrage, noting the former requires no capital and carries no risk. They point out that in reality arbitrage does require capital, as well as entails risk. Moreover, such arbitrage is run by a limited number of highly specialized investors employing the capital from outside sources. Shleifer et al. show that especially in extreme circumstances, such arbitrage may not be fully effective in pushing price back to fundamental levels. In spite of the attractive returns available from mispricings, arbitrageurs would expose themselves to volatility and losses, and thereby to liquidations in the fund on the investors’ part. In conclusion, their work establishes that the avoidance of return volatility by arbitrageurs can hinder their ability to eliminate certain anomalies through trading.

Xiong (2001) studies convergence traders in particular, assuming logarithmic utility of wealth to them. In general, convergence traders reduce price volatility and provide liquidity to the market. However, when the arbitrageurs endure losses due to an unfavorable shock, they are forced to liquidate their positions as risk-bearing capacity shrinks, thereby amplifying the original shock. In other words, non-convergence of arbitrageurs' positions can lead to a vicious cycle of no convergences, as arbitrageurs' wealth is diminished, and their risk-aversion increases.

Mitchell and Pulvino (2010) discuss how arbitrageurs employing financial leverage are able to force even small pricing discrepancies to converge. They study the 2008 financial crisis, looking into the sudden and dramatic decrease in leverage available to hedge funds. From the arbitrageurs' point of view, seemingly long-term debt became short-term one, thus creating a mismatch between their assets and liabilities. Resulting from the withdrawal of financing, many hedge funds with relative-value strategies were unable to make assets with similar payoffs worth all but the same, i.e., to impose the no-arbitrage condition. The authors discuss how the magnitude and convergence time of the mispricings during the crisis can provide an indication of the arbitrageurs' role in enforcing no-arbitrage during the normal times.

Finally, Gromb and Vayanos (2010) summarize theoretical literature on the limits of arbitrage. They discuss how different mechanisms inflict costs to arbitrageurs, thereby preventing them from correcting mispricings and providing liquidity to other investors. First, Gromb et al. consider demand shocks that generate mispricings stemming from behavioral or institutional reasons. Secondly, they classify cost faced by arbitrageurs into categories: fundamental and non-fundamental risk, short-selling costs, leverage and margin constraints, and constraints on equity capital. Given the nature of this thesis, most promising single factor is probably risk. As Gromb et al. discuss, a number of papers explore dynamic multi-asset equilibrium settings with models assuming that the only cost for the arbitrageurs is risk. Greenwood (2005) and Hau (2009) portray arbitrageurs who absorb demand shocks of index investors following index redefinitions. In Gabaix et al. (2007), arbitrageurs have the unique ability to hold mortgage-backed securities; as they hedge interest-rate risk in the bond market, they continue to carry prepayment risk. Similarly, in Garleanu et al. (2009), arbitrageurs absorb shocks in the options market, delta hedging in the cash stock market, and thereby bearing the jump and volatility risk. On the same lines, as discussed earlier, are Vayanos and Vila (2009) and Greenwood and Vayanos (2010), where arbitrageurs absorb demand and supply shocks to specific maturities in the government bond market, hedging the bets by trading other maturities on the yield curve. While depicting different markets, the above literature comes together in that arbitrageurs transmit asset-specific shocks to other assets so that the effects are largest for the assets with the highest covariance with the original shock-hit asset. Jylhä and Suominen (2009), Plantin and Shin (2009), as well as Hau (2014), study similar effects in the FX market, while Naranjo (2009) looks into the futures space.

3.5.3. Noise in the markets

Hu, Pan and Wang (2013) propose a market-wide liquidity measure based on observed price deviations in the U.S. Treasury bonds, connecting this to the amount of arbitrage capital in the market. Similar to the reasoning of Vayanos and Vila (2009), Hu et al. consider that during normal times in the markets, there is abundant arbitrage capital that makes the Treasury yield curve smooth and keeps deviations small. However, during market stress periods, or crises, there will be shortages in the arbitrage capital, which can result in some of the yields moving out of the line with respect to the rest of the curve, meaning that there is more 'noise' in the bond prices.

Hu et al. choose the U.S. Treasury bond market as the space where they look for a measure that would capture market-wide liquidity short-falls. The logic behind the choice is that the Treasury market is known to be one of the largest and most liquid markets, and it is important also in collateral sense for funding purposes, as well as naturally as being a key asset class for investment. The point, however, is not to find a liquidity measure pertaining solely to the fixed income market, but for the broader financial markets across assets classes. The key is that the liquidity shocks that originate in other, potentially less liquid markets will be felt also in the very liquid Treasury bond market, especially when the liquidity shocks are significant enough. The measure is also designed to be robust in that it incorporates a number of points on the yield curve, and is therefore not concentrated on any single maturities, for instance.

The authors suggest that the noise as captured by their measure can be informative about the liquidity conditions of the broader market. They show that their noise measure is able to capture a number of liquidity crises across the financial markets in a way that is superior to earlier suggestions as proxies for liquidity. Moreover, Hu et al. show that employing their measure as a priced risk factor helps to explain the returns of hedge funds, which are understood to be sensitive to the broad liquidity conditions of the market.

The connection between the noise and yield curve arbitrage returns is thus based on the idea that high noise coincides with high mispricings, thereby offering arbitrageurs better trading opportunities. The stress periods presumably decrease the level of arbitrageur activity in the markets after the arbitrageurs endure losses, are unwilling to employ their capital to the full extent, or face tightened leverage and margin constraints. As a result, increased noise will be measured in the bond prices given that they are not smoothed aggressively enough by the arbitrageurs. This in turn may explain why those arbitrageurs willing to step in can reap attractive returns from relative value bets.

3.5.4. Compensation for tail risk

In general, hedge funds and certain strategies typical to hedge funds are often seen as having particular exposure to the negative tail of an asset price distribution, which is referred to as tail risk. The tail risk basically equals disaster, crash or rare-event risk. Having such exposure can naturally explain outperformance during an observation period when no tail risk materializes. In this subsection, I shall shortly present some literature discussing the tail risk, as it is a feasible explanation for the perceived and potential strength of the yield curve arbitrage strategy studied in this thesis.

Mitchell and Pulvino (2001) contemplate that merger arbitrage has similar characteristics to a strategy selling index put options. In other words, the strategy earns small profits most of the time, and has a limited correlation with the overall market, yet suffers heavily during the market downturns, when the correlation becomes significantly positive. Writing put options naturally equals selling disaster insurance and loading on the tail risk.

As Duarte, Longstaff and Yu (2007) discuss, fixed income arbitrage strategies' returns are often considered to be negatively skewed. This means that the strategies are thought to generate small returns on average, yet to suffer large losses occasionally. Should this be the case, then any alpha such strategies generate in a finite time period could be attributed as compensation for bearing crash risk. Taking long exposure to the crash risk has been shown to be characteristic to both hedge funds in general and to specific hedge fund styles in particular.

Brunnermeier et al. (2009) provide evidence of a strong link between currency carry and currency crash risk. The strategy of being long high interest rate currencies and short low interest rate ones delivers returns with a negative skew. They show that speculators invest in high-carry currencies, arguing that currency crashes are linked to sudden unwinding of these positions. This may be explained by withdrawal of liquidity and lower speculator capital, as the currency crashes are positively correlated with the VIX index and the TED spread.

Jiang and Kelly (2012) document large and persistent hedge fund exposures to the downside tail risk. They show that funds with exposure to the tail risk earn annual returns of nearly 6% higher than their peers whom are tail risk-hedged, controlling for commonly employed hedge fund factors. They conclude that their results are consistent with the notion that a large part of hedge fund returns can be seen as compensation for selling disaster insurance.

4. Hypotheses

Based on the prior literature discussed in Section 3, as well as on general financial theory, I shall state a number of hypotheses to be tested in this thesis. Firstly, I hypothesize that the yield curve arbitrage in the EUR swap space generates attractive risk-adjusted returns within a sample period from 2002 to 2015, with a focus on out-of-the-sample trading from the late 2005 onward. Additionally, it is hypothesized that the core strategy's return distribution is tilted favorably to the arbitrageurs' benefit. Secondly, I hypothesize that these returns are mostly multifactor alpha, having thus limited exposure to known priced risk factors. Moreover, I make the hypothesis that the replicated yield curve arbitrage strategies explain to a statistically significant degree the most relevant hedge fund subindex returns, namely the HFR and Credit Suisse fixed income arbitrage, or relative value, indices. Further, I hypothesize that the mispricing of rates in the market is greater when there is a lot 'noise' as defined by Hu et al. (2013). Finally, I hypothesize that the returns from the strategies are greater during these noisy periods.

The Hypotheses 1 to 5 follow directly from the findings of Duarte et al. (2007). They are formulated so that the strategies are tested with the EUR swap dataset, with an enhanced methodology, as well as with modifications to the strategies.

Hypothesis 1: *Replicated yield curve arbitrage trading strategies generate attractive risk-adjusted returns in the EUR swap rates space.*

Hypothesis 2: *Replicated core yield curve arbitrage trading strategy's return distribution is non-normal, i.e. it is positively skewed with a high kurtosis.*

Hypothesis 3 pertains to the different implementations of the strategies, given that one can choose to trade the mispricings in a number of ways. The hypothesis states that the strategy directly following Duarte et al. (2007) will have the best risk-adjusted returns when compared to the other elementary modifications of the strategy.

Hypothesis 3: *A yield curve arbitrage trading strategy that trades multiple largest monthly mispricings at the same time will result in the most attractive risk-adjusted returns in the space of elementary strategy variations.*

Hypothesis 4: *Replicated yield curve arbitrage trading strategies' returns are mostly multifactor alpha, when controlled by known and potential risk factors across asset classes.*

Hypothesis 5: *Replicated yield curve arbitrage trading strategies' returns will not have a statistically significant explanatory power with respect to the high-level hedge fund indices.*

Hypothesis 6 is modification of Hypothesis 5 in that while the high-level hedge fund indices are expected to have no significant relation with the replicated strategy, the subindices closer to the EUR fixed income arbitrage space are conversely expected to have such a connection.

Hypothesis 6: *Replicated yield curve arbitrage trading strategies' returns will have a statistically significant explanatory power with respect to the most relevant hedge fund subindices.*

Hypotheses 7 and 8 are of a formulation that is yet to be tested anywhere. They intend to connect the yield curve arbitrage strategy to the research on the 'noise' in the (fixed income) markets, as well as to the limits of arbitrage literature as discussed in Section 3.

Hypothesis 7: *The extent of model implied mispricing of rates is highly positively correlated with the amount of 'noise' in the market.*

The Hypothesis 7 leads logically to the Hypothesis 8 on the grounds that noisy market environments are expected to coincide with the withdrawal of arbitrage capital [Hu et al. (2013), Greenwood and Vayanos (2014)], thereby possibly providing active arbitrageurs with more attractive opportunities to trade on.

Hypothesis 8: *Replicated yield curve arbitrage trading strategies' performance in terms of risk-adjusted returns is highly positively correlated with the amount of 'noise' in the market.*

Tests of the Hypotheses 1, 2, and 3 shall be discussed in Section 6. The Hypothesis 4 is covered in Section 7, while tests of the Hypotheses 5 and 6 are reported in Section 8. Finally, the Hypotheses 7 and 8 related to the noise measure are tested in Section 9.

5. Data and methodology

5.1. Data

The data consists of monthly observations of 1-year to 10-year constant maturity ('par' or 'zero') EUR swap rates. This swap data is available on the Bundesbank's website. The data ranges from the beginning of January 2002 until the end of January 2015. The rates are mid-swap rates, i.e. the average of the bid and offer quotes for a given rate maturity at the end of each month. The data is validated by cross-checking to a similar dataset extracted from a Bloomberg terminal. Figures 1 and 2 illustrate the swap rate data in question. Data regarding hedge fund indices and the noise measure by Hu et al. (2013) shall be illustrated in the related Sections 8 and 9, respectively.

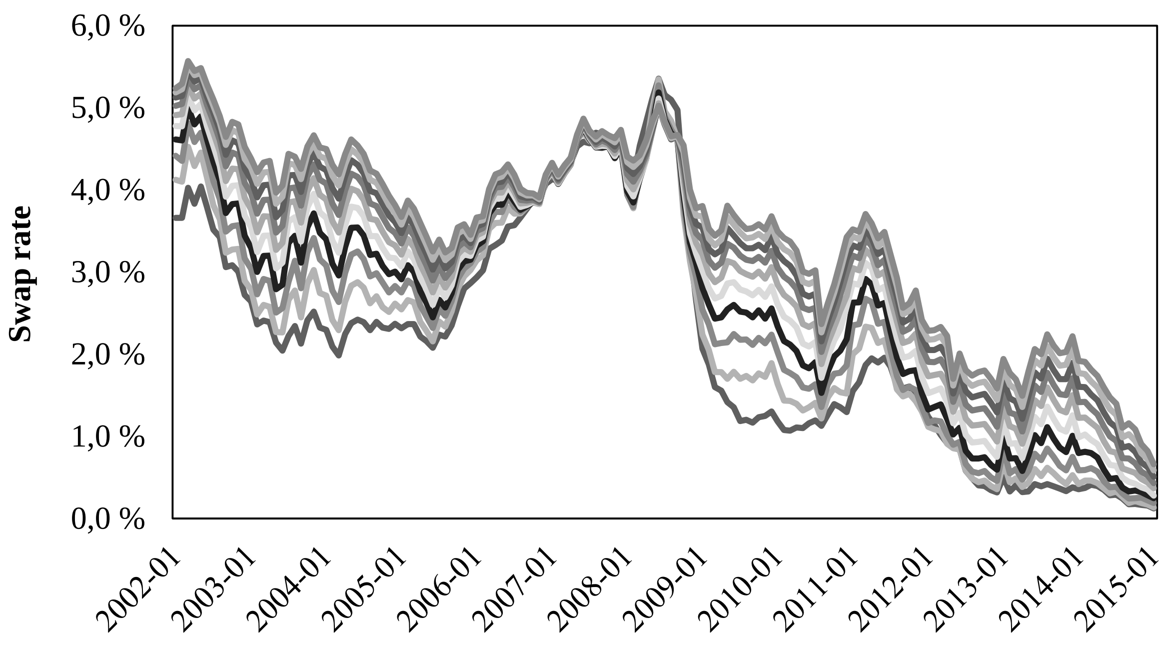

Figure 1. EUR swap rates evolution in time.

This figure illustrates the constant maturity EUR par swap rate data from the beginning of 2002 until the January of 2015. The swap rates are for maturities of one to ten years. The rates illustrated are mid-market quotes at the end of each calendar month, as available from the Bundesbank's website.

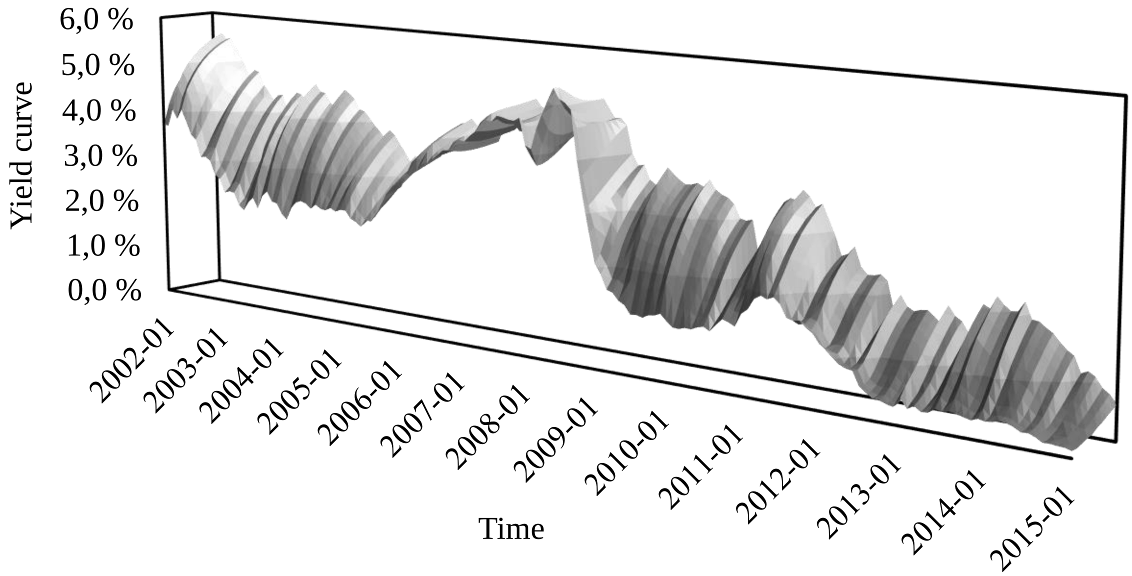

Figure 2. Yield curve's evolution in time.

This figure illustrates the constant maturity EUR par swap yield curve data from the beginning of 2002 until the January of 2015. The swap rate data is for maturities from one to ten years. The swap rate curve is depicted by the slope in the figure. The rates illustrated are mid-market quotes at the end of each calendar month, as available from the Bundesbank's website.

5.2. Swaps, swap rates and discount factors

A swap is a derivative agreement between two counterparties to exchange future cash flows related to a certain asset on pre-agreed dates. A vanilla fixed-to-floating interest rate swap is a contract where one party agrees to pay a fixed 'swap rate' and to receive a periodically determined 'floating rate' in exchange (payer swap). The counterparty would agree to have the opposite position (receiver swap). The floating rate is usually LIBOR or EURIBOR, and commonly itself has a maturity of three or six months. In the EUR space, the reference floating rate is generally the 3-month EURIBOR, in which case the payments are exchanged, or 'swapped', every 3 months. The fixed rate in question is chosen so that the value of the swap is zero at initiation. Therefore, the fixed rate's value depends on the expected future spot EURIBOR rate, namely its forward term-structure observed at the time.

The swap rate (or 'par swap rate') refers to the fixed rate at which at a given time a swap contract has a mark-to-market value of zero. As a swap is worth zero at initiation, this means that the counterparties entering into the contract would need capital only for collateral purposes, as the swap itself does not 'cost' anything to enter for neither of the parties. This is in contrast to dealing in the options market, for instance, where a premium would be paid or received when entering into the contract.

Below, Equation 2 presents the computation of a discount factor, or a zero-coupon bond, in the context of swap rates, which shall be employed throughout the thesis when moving from the swap rates to the discount factors for different maturities.

Equation 2 (Discount factor)

Making the observation at $t = 0$, the value of a unit notional zero-coupon bond, or a discount factor, is given by the below formula, where S_T refers to the yield, or swap rate, for that particular maturity T.

$$DF_{0,T} = \frac{1}{(1+S_T)^T} \qquad (2)$$

Equation 3 below illustrates the computation of the par swap rate. Such calculation is not necessary in this paper, given that the data itself is in the form of par swap rates. Hence, the equation is provided for the reader as a context for understanding how the market prices the swap rates.

Equation 3 (Par swap rate)

For $t \leq t_0$, the swap rate is given by the below formula, where t is the time of observation, t_0 is the initiation of the swap, and T is the maturity of the swap. In this thesis, it is necessary only to consider the case $t = t_0 = 0$ when computing the swap rates, i.e. to deal solely with the spot rates.

$$S_{t,T} = \frac{DF_{t,t_0} - DF_{t,t_0+T}}{\sum_{i=1}^{T} DF_{t,i}} \qquad (3)$$

5.3. Modeling methodology

Modeling is obviously at the very heart of replicating the returns available from arbitrage trading. The hypotheses and prior literature imply that the attractive returns from quantitative yield curve arbitrage strategies are at least partially explained by the relatively sophisticated methods employed. This is logical in the sense that required sophistication from arbitrageurs heightens barriers to entry to the arbitrage business.

In short, the modeling methodology chosen ought to be capable of replicating the behavior of arbitrageurs as discussed by Vayanos and Vila (2009) in their theory of yield curve formation and integration. By attempting to imitate the trading and hedging decisions of the depicted yield curve arbitrageurs, I hope to replicate the return characteristics of their strategies.

The modeling approach shall be similar across the short-rate models and the specific strategies considered. When studying the individual strategies, some modifications can be made, for instance the sample period for calibration can be made rolling instead of static. By and large, however, the core of the methodology described below will pertain to all of the strategies being considered.

First, in order to quantitatively locate potential mispricings in the market swap curve, one has to specify a certain model that is used as a framework for the analysis. Secondly, the model chosen has to match the market rates to a certain extent for the model to be relevant at any level. Thirdly, the model has to be able to point out tradable and economically significant mispricings to be of a practical benefit. This does not have to be true at all times; it is sufficient that the model points out mispricings frequently enough to offer economically acceptable (leveraged) total returns. Further, it has to be possible for the market rates to converge to the model rates to a decent degree, so that one can observe when a mispricing no longer exists. Finally, one needs to be able to extract some sort of sensitivity measures from the model. Namely, the model must be first-order differentiable with respect to the risk factors. This is necessary for an arbitrageur to be able to construct an arbitrage portfolio, which, by definition, is a portfolio that is neutral to (incremental) movements in the risk factors. If one were unable to find sensitivities to these factors, or the 'hedge ratios', then there would be little point in locating a mispricing in the market rates, as one would not be able to 'lock in' low-risk (or riskless) profits at any given time.

I chose Cox-Ingersoll-Ross (CIR2F) and Longstaff-Schwartz (LS2F) two-factor models as the tools for modeling the yield curve. This choice was motivated by the fact that both models are essentially considered state-of-the-art in the short-rate literature. In general, two factors are sufficient to explain around 99% of the variation in the bond returns, as shown by e.g. Litterman et al. (1991). Duarte et al. employed Vasicek two-factor model for this purpose. I hope that the choice of CIR2F and LS2F produces even better outcomes on the basis that the CIR2F model in particular adds sophistication to the Vasicek model by disallowing negative rates, for instance. Even in the present ultra-low rate environment, this seems to be a desirable property, given that the EUR swap rates are yet to go to the negative territory. Moreover, the Vasicek model implies normally distributed rates, whereas the CIR2F model has a non-central Chi-squared distribution, which seems like an enhancement, given that rates are certainly not normally distributed [see Cox et al. (1985) and Vasicek (1977) for further details of the models].

The modeling itself begins with the idea that the rates in the market have to fit the model rates to a reasonable degree, or more specifically, some rates have to match more or less exactly. Duarte et al. (2007) fit the model exactly to the 1-year and 10-year rates in the market data; as in my dataset, these are the end points of the yield curve. While Duarte et al. do not explicitly say why they choose specifically these rates, it is rather obvious: these are the most liquid rates in the 1-year to 10-year space. Amihud (2002) and Amihud et al. (1991), among others, show that liquidity affects the efficiency of pricing in the markets, and thus one can infer that other things being equal, the most liquid securities, or rates, convey the best information to a trader. In this regard, I will follow Duarte et al. in considering the 1-year and 10-years to be the ‘liquid rates’ and the 2-year to 9-year rates to be the ‘illiquid rates’.

To move towards fitting the models to the market rates, one can look into the analytical solutions of the models: both CIR2F and LS2F have an analytical solution for a zero-coupon bond price, or for a discount factor. This discount factor can then be easily turned into a corresponding yield, or equivalently, into a swap rate. So basically, one moves from discount factors to swap rates, and vice versa, when calibrating the models with respect to the rates observed in the market. Section 5.2 further elaborates on the swap rates and discount factors.

I follow the approach of Duarte et al. in fitting the models to the market rates. First I choose an in-the-sample period (my standard selection is four years) to which I calibrate the model, and where no trading is done. This ensures that I do not look in any way to the future data when choosing the model parameters. The historical calibration is done so that first, for each month in the sample, the models’ factors (x_1 and x_2 in CIR2F; r and V in LS2F) are chosen so that the 1-year and 10-year rates match exactly (or as close to as possible, at worst) the corresponding markets rates. Then, given these factor choices, the sum of squared differences between the market and the model is minimized for the ‘illiquid rates’ jointly across the sample. Illiquid rates mean all the rates from the 2-year rate to the 9-year rate. Calibrating ‘jointly across the sample’ means that the objective function in the minimization problem is the sum of squared differences across all the months in the sample, and for each month it contains the squared differences for rates from two to nine years. This step selects the 6 parameters that both models have. To illustrate, CIR2F’s parameters have a straight-forward economic interpretation: theta is the long-run mean of the short-rate, sigma is the volatility of the short-rate, and k stands for the speed of mean-reversion of the short rate. As there are two factors in the CIR2F, there are these three parameters regarding both of them, summing up to the six in the whole model.

Equation 4 presents the formula for the CIR2F zero-coupon bond price, while Equation 5 does the same for the LS2F model. Both the underlying stochastic short-rate processes and the resulting analytical formulations are presented. For the derivation of the bond price and further details, the reader is referred to Cox et al. (1985) and Vasicek (1977). The two-factor models follow straight-forwardly from the original one-factor models.

Equation 4 (CIR2F zero-coupon bond price)

CIR2F model has to factors that follow these stochastic processes

$$dx_1(t) = k_1(\theta_1 - x_1(t))dt + \sigma_1\sqrt{x_1(t)}dW_1(t)$$

$$dx_2(t) = k_2(\theta_2 - x_2(t))dt + \sigma_2\sqrt{x_2(t)}dW_2(t)$$

A discount factor, or a zero-coupon bond, with a maturity T, as seen at time t, is priced by

$$DF(t,T) = A_1(t,T)A_2(t,T)\exp(-x_1(t)B_1(t,T) - x_2(t)B_2(t,T)) \tag{4}$$

Where,

$$A_i(t,T) = \left(\frac{2h_i e^{(h_i+k_i)(T-t)/2}}{2h_i + (h_i + k_i)(e^{h_i(T-t)} - 1)}\right)^{2k_i\theta_i/\sigma_i^{\,2}}$$

$$B_i(t,T) = \left(\frac{2\left(e^{h_i(T-t)} - 1\right)}{2h + (h + k)(e^{h_i(T-t)} - 1)}\right)$$

$$h_i = \sqrt{{k_i}^2 + 2{\sigma_i}^2}$$

i = 1, 2 standing for the two factors.

Equation 5 (LS2F zero-coupon bond price)

A discount factor, or a zero-coupon bond, with a maturity T as seen at time t is priced by

$$DF(T) = A^{2\gamma}(T)B^{2\eta}(T)\exp(\kappa T + C(T)r + D(T)V) \qquad (5)$$

Where,

$$A(T) = \frac{2\phi}{(\delta + \phi)\exp\big((\phi T) - 1\big) + 2\phi}$$

$$B(T) = \frac{2\psi}{(\nu + \psi)\exp\big((\psi T) - 1\big) + 2\psi}$$

$$C(T) = \frac{\alpha\phi(\exp(\psi T) - 1)B(T) - \beta\psi(\exp(\phi T) - 1)A(T)}{\phi\psi(\beta - \alpha)}$$

$$D(T) = \frac{\psi(\exp(\phi T) - 1)A(T) - \phi(\exp(\psi T) - 1)B(T)}{\phi\psi(\beta - \alpha)}$$

And,

$$\nu = \xi + \lambda$$

$$\phi = \sqrt{2\alpha + \delta^2}$$

$$\psi = \sqrt{2\beta + \nu^2}$$

$$\kappa = \gamma(\delta + \phi) + \eta(\nu + \psi)$$

The next step is to employ these selected six parameters as we move forward in time, stepping into the out-of-the-sample space, which explicitly was left out of the calibration scheme thus far. Now, as the trading is to begin, month-by-month the model is fitted to the market data by selecting the two factors so that there is, again, a perfect (or as perfect as possible, at worst) match in the 1-year and 10-year rates between the market and the model. As these 'liquid rates' are made to fit the market exactly, the model will not fit the rest of the rates (2-year to 9-year rate) exactly most of the time, which is the purpose here. As there is a discrepancy between the market pricing and model pricing of the illiquid rates, it implies that either the market is wrong, or the model is wrong is determining the premia related no-arbitrage price of a given rate. The view here is obviously that the model is right, at least most of the time, in giving the arbitrage-free rate. The logic behind this view is that the model is internally arbitrage-free by construction, and if the 1-year and 10-year rates are more efficiently priced than the rest of the rates, then the model will imply arbitrage opportunities, should it differ from the observed market rates.

For a model to be relevant with respect to the market rates, it needs to have a proper specification of what is called the 'market price of risk' (interest rate risk in this case). This specification is attained through the calibration procedure. As the calibration is exact only for the 1-year and 10-year rates, the choice of the market price of risk quantity is also largely based on the assumption that these rates convey it appropriately, or in some sense better than the rest of the rates. Given that the choice of the parameters is based on the historical 'illiquid rates', part of the market price of risk calibration lies naturally in that data. Hence, one can say that the market price of risk is found as the combination of historical data and current liquid rate observations.

Now that the model is made to quantify the differences between the market and the model rates for the illiquid rates, one can consider if there is a mispricing in the markets. Again following the Duarte et al. (2007) approach, I consider there to be a mispricing, should the market rate deviate from the model rate by a certain level of basis points (bps, 1% is 100 bps). Duarte et al. set the limit to 10 bps; I shall also use the 10 bps as the standard choice, but will, in certain instances, also employ a five and a twenty basis point limit. When one concludes that there is a mispricing, it remains to decide how exactly to trade that observation.

Direction of the trade is clear: if the market rate is higher than the model rate, one expects the market rate to come down, and would thus receive the fixed rate and pay the floating leg, which is, of course, equal to being short the floating rate (or short-rate). If the opposite is true, and the market rate is lower than the model rate, then one would assume that the market rate would converge higher. In this case the trader would go long the floating rate and short the fixed rate, i.e. would enter into a payer swap. In either of these cases, one would enter into offsetting position in the liquid 1-year and 10-year rates to make the overall trade market neutral in the two factors. This is further discussed in Section 5.5.

Besides the direction and hedging of the trade, one would also need to decide which exact rate or rates to trade. It is often the case that the model implies multiple mispricings; in the extreme, the model would show that all the illiquid rates are mispriced (in practice, this does rarely if ever occur, however). One therefore has the choice to trade only the largest mispricing, some mispricings, or all of them. I will consider the strategies where only the largest mispricing is traded, as well as the strategy where all the mispricings are traded. Moreover, one can decide whether multiple trades initiated at different points in time are allowed. One can either decide that only a single trade can be on at any given time, and that until that trade is closed for any reason, no other trades are entered. I consider this strategy as the 'elementary version' of the possible choices. It can also be decided that for every month going forward, one chooses to trade the largest mispricing found in the market that month, but that as time goes by, one can have multiple trades on that originated in different months in history. This multi-trade specification is what Duarte et al. basically used, and it is the core strategy specification in this thesis. I do also consider a strategy where all the mispricings are traded. This means that for any given month, the arbitrageur opens positions in all the mispriced rates and hedges them. Going forward, any of these trades can be closed independently of each other if and when the mispricings disappear, for instance.

After initiating the trades in any possible strategy specification, one needs to naturally decide also when the trades are closed. The trades are closed when they achieve the purpose they were initiated for, or when they do not do so. Namely, if the market rate converges to the model rate, the mispricing is considered to have disappeared, and the trade and its hedges are closed. The convergence is not assumed to be perfect; I will consider it to have happened should 50% of the original difference have converged, or been eliminated from the market rates. If the trade does not converge in 12 months, it is closed due to a time limit, following the Duarte et al. (2007) approach.

The above described the modeling aspect itself. Below, I will go through the trading implementation in more detail to make sure the reader understands how the trading rules are structured. In the section after the trading decisions, I will go through the hedging procedure, which is also at the very core of the strategies.

5.4. Trading methodology

Once the modeling of the yield curve has been accomplished as described above, the arbitrageur needs to focus on how specifically he wants to trade the potential mispricings. Decision rules need to be in place for both initiating trades as well as closing them. The basic idea is to initiate trades when mispricings are found, and to close trades when the mispricings disappear, meaning that the market and model rates 'converge' enough. The model implied rates for the illiquid rates (2-year to 9-year maturities) generally deviate from the market rates by at least some basis points. One therefore has to decide how large a deviation is necessary in order to consider it a tradable mispricing, or an arbitrage opportunity. Following Duarte et al. (2007), I employ a 10 bps limit as the standard choice for the difference between the market and model rate that is considered to signal a trading opportunity (i.e., the difference has to be 10 bps or larger). Also 15 and 20 bps differences as the limit for trading are explored.

Overall, the model will imply one, several or no mispricings. If there is no mispricing, then the arbitrageur simply will not initiate a trade during that month. If there is only one mispricing, then the arbitrageur would set to arbitrage that mispricing. In the case of multiple mispricings – as is the most common case for the model – the arbitrageur can choose one or several trades. If he were to choose just one trade, it would be logical to choose the largest mispricing. If he were to choose multiple mispricings to be arbitraged, then basically any number can be chosen. I will run a strategy where only the largest mispricing is traded, as well as a strategy where all the mispricings are traded.

The other choice the arbitrageur can make is whether he wishes to have multiple trades going on in his portfolio. He can choose to trade just one mispricing and wait until it converges, or he can have multiple independent trades that are initiated and closed based on the traded rates' individual deviation and convergence with respect to the market.

The first and simplest trading methodology is to search for the largest single mispricing in the market yield curve by comparing it to the modeled curve ("One mispricing"). The point here is to trade only the largest difference between the curves, even if there were multiple differences exceeding a set limit. This approach looks at the returns available from being conservative in the trading decisions in that only the 'lowest hanging fruit' is picked up, with the expectation that it offers the best risk-adjusted return. The logic is that by committing capital only to the largest mispricing, one has the largest expected return, and presumably the smallest chance of an erroneous bet. Also, the largest mispricing can be assumed to get in line with the model the fastest, thereby contributing to annualized returns through increased turnover of capital from one profitable trade to another. With this trading methodology, the hedge fund that is modeled has at all times only one trade that it manages, i.e. the trade can be closed when some pre-defined conditions (e.g., convergence to the market rate) are met, and a new trading opportunity can then be looked for.

The second trading methodology attempts to trade all the mispricings exceeding a set limit at any given time ("All mispricings"). Therefore, if no trades are in place currently, and the model points to several mispricings, then this strategy would initiate trades in all the rates deemed enough mispriced. Then, going forward, the strategy checks if converge happens to any of the trades, and closes them based on that. At the same time, new trades are initiated on a rolling basis, given that the same maturity is not being traded already. In this strategy, the arbitrageur will enter to trades in all the mispricings he perceives every month going forward in the trading period.

Finally, there is the strategy replica of Duarte et al. (2007), i.e. the modeling of a 'hedge fund index' ("DLY") so that at any given time only one trade can be initiated, namely the one with the largest mispricing. The difference to the first strategy described is that multiple trades are allowed, and a new trade can start every month. Therefore this strategy can be thought of as a combination of the two above; it allows for multiple trades, yet it only trades the largest perceived mispricing each month.

All of these strategies work so that the market yield curve is compared to the fitted model, and then the differences between the 'illiquid' market and model rates are looked at. If the difference is great enough, usually set to the mentioned 10 bps, a trade is initiated as described above. Going forward, the trades are monitored for profit and loss (PNL) and for converge. If the convergence is deemed clear enough, the trade is closed as the mispricing is considered to have vanished. One could, of course, employ other trading rules such as a stop-loss or a stop-gain level. Following Duarte et al. (2007), I shall solely employ the convergence and time limit rules. This is a sound approach as the convergence rule, after all, is based on the assumption and theoretical framework that the mispricings, or arbitrage opportunities, will vanish in time as the arbitrageurs take action.

5.5. Hedging methodology

Theoretically broadly consistent with Vayanos and Vila (2009), Greenwood and Vayanos (2010), as well as the broader set of literature on arbitrageur behavior outlined in Section 3, the arbitrageurs in my work will hedge their bets. More specifically, I shall follow the practical ‘butterfly’ hedging approach outlined in Duarte et al. (2007). The theme in the literature is that arbitrageurs engage in hedging activities in order to isolate a less conventional source of return – such as bond, volatility or prepayment risk premia – they specialize in arbitraging. What is left unhedged by the arbitrageurs in this thesis is the convergence of the mispriced rates; the premia there might be best described as local bond risk premia. This section will shed light on how the trades are hedged in more detail.

To have an arbitrage trading portfolio, an arbitrageur needs to make his overall portfolio neutral to the (incremental) changes in the key risk factors affecting the mark-to-market value of the portfolio. Such a portfolio is comprised of the trade targeting the perceived mispricing, as well as the hedging instruments. I will call an ‘arbitrage portfolio’ a portfolio that is made of the arbitrage trade plus the hedges of that trade. In the two-factor short-rate modeling context, there are two sources of randomness, and thus two securities are necessary to fully hedge the trade. The arbitrage portfolio must be overall neutral to those two factors. Namely, the factors are x_1 and x_2 in the CIR2F model, and r and V in the LS2F model.

The hedges in the strategies are structured so that each arbitrage trade is hedged separately. One could alternatively choose to hedge the overall position, as the sensitivities, or ‘Greeks’, can simply be summed across all the positions in the portfolio. From the modeling perspective, I found it convenient to choose the hedges for each trade separately. To begin with, when one initiates a trade, the two-factor sensitivities for the rate (discount factor) to be traded are computed. The same sensitivities are computed also for the 1-year and 10-year rate, which are always the two rates employed for hedging. Then one simply finds the weights, or hedge ratios, for the hedging instruments (1-year and 10-year rates) that make the arbitrage portfolio's sensitivity zero to both of the factors. The formulas for these two-factor sensitivities are shown in Equations 7 to 10.

Equations 7 and 8 (CIR2F sensitivities)

DF, B_1 and B_2 are as defined in Section 5.3,

$$\frac{\partial DF}{\partial r} = \Delta = -B_1(t)DF(T) \qquad (7)$$

$$\frac{\partial DF}{\partial V} = \nu = -B_2(t)DF(T) \qquad (8)$$

Equations 9 and 10 (LS2F sensitivities)

DF, C(T) and D(T) are as defined in Section 5.3,

$$\frac{\partial DF}{\partial x_1} = \Delta = C(T)DF(T) \qquad (9)$$

$$\frac{\partial DF}{\partial x_2} = \nu = D(T)DF(T) \qquad (10)$$

One solves the below outlined linear algebra problem (Equation 6) for weights w_1 and w_2 to find the hedge ratios that make the arbitrage portfolio market neutral. Once the hedge ratios are known, one enters into a trade where the hedging instruments' weights have an opposite sign to the core trade itself. In other words, if the arbitrageur is short e.g. the 5-year floating rate, then he would be long both the 1-year and 10-year floating rate in the quantity specified by the hedge ratios. With respect to the notation, I will call sensitivities of the discount factors to the two factors delta and vega. This is not unlike in options trading, where delta and vega would be the first-order partial derivatives of the option price with respect to the price of the underlying instrument and implied volatility, respectively. Here, the derivatives are with respect to the factors in the models; for the CIR2F model, for instance, the economic interpretation of the differentials is less obvious than in the options' context. Nevertheless, the meaning and use of such sensitivities is by and large the same as it is in the case of hedging options with respect to the delta and vega, leading to the notation. Again, as there is the link between discount factors and swap rates, computing the sensitivities to the discount factors is equivalent to computing the same metric for the swaps in terms of how much the profit-and-loss changes when the factors move incrementally.

Equation 6 (Two-factor market-neutrality problem)

Δ and ν refer to the sensitivities of the discount factors to the two factors in the short-rate models. Subscripts indicate the part of the arbitrage portfolio to which the sensitivity pertains to; '1y' and '10y' mean the 1-year and 10-year rates (discount factors) employed as hedges, whereas the 'trade' refers to the mispriced rate targeted by the arbitrage.

$$\begin{bmatrix} \Delta_{1y} & \Delta_{10y} \\ \nu_{1y} & \nu_{10y} \end{bmatrix} \begin{bmatrix} w_{1y} \\ w_{10y} \end{bmatrix} = \begin{bmatrix} \Delta_{trade} \\ \nu_{trade} \end{bmatrix} \tag{6}$$

As to why the arbitrageur has these two opposing positions, and specifically in the 1-year and 10-year rates, follows from the two-factor model. In such a model, there are two sources of risk (or randomness) that need to be hedged; this is equal to having two equations to be solved, as in the above matrix specification. For the two equations to be solved, one needs to variables, which are the two rates used for the hedging. The 1-year and 10-year rates are considered to be the liquid rates that are most efficiently priced. Therefore, it is natural to use them to offset the factor risk in the mispriced rate, as the arbitrageur can assume that these two rates do not converge themselves to the model. By and large, such convergence for the liquid rates is not even possible, as they are made to fit the market curve exactly in the first place. With this logic in mind, the arbitrageur solves the system of the equations by choosing appropriate weights for the liquid rates.

6. Yield curve arbitrage returns

This section discusses the core findings of the thesis by presenting the returns of the backtested yield curve arbitrage strategies. The related Tables and Figures shall be presented in the subsections of 6.1 to 6.4. In general, the explored strategies produce economically and statistically significant leveraged monthly returns. As the Sharpe ratios for the strategies are high, leveraging them to an annualized ex post volatility of 10% results in returns of around 10% per annum. Return distributions of the strategies are positively skewed and have a high kurtosis. Summary statistics for the strategies are illustrated through Tables 1 to 6. The tables depict the trading outcomes of different trading strategies with different calibration choices. Tables 1 to 4 show the results for the Cox-Ingersoll-Ross two-factor model (CIR2F), while Tables 5 and 6 do the same for the Longstaff-Schwartz two-factor model (LS2F).

As the summary statistics in below subsection show, the replicated strategies generate economically significant leveraged monthly excess returns that are positively skewed with fat tails (positive kurtosis). Both the Sharpe ratios and Gain-Loss ratios are high, implying attractive risk-adjusted returns. At maximum a third of the returns are negative, and generally the highest return is larger than the lowest return. The returns differ significantly from zero, as implied by the Newey-West (1987) autocorrelation robust t-Statistics. A general conclusion is that the strategies that trade in-the-sample perform better than the ones trading out-of-the-sample, as could be assumed. Also, the strategies where the sample period for calibration is longer or rolling, instead of fixed, seem to fare better.

It also seems that trading each and every perceived mispricing (“All mispricings”) is not attractive in terms of risk-adjusted returns. This is probably because trading only the largest mispricings offers greater expected returns and more consistency in convergences. When lengthening the calibration sample from four years to eight, very attractive trading outcomes result. The returns are actually even better than those from calibrating the model in-the-sample, where the in-the-sample procedure incorporates the whole 14-year sample. This can suggest either a data mining issue, or that the trading environment is more favorable towards the end of the sample, which is where the strategy with the 8-year sample trades.

Trading solely a one mispricing at a time (“Single mispricing”) is not either the optimal solution. This is logical, as the strategy can always hold only one trade at a time, can thus miss superb opportunities, and is also less well diversified across mispricings. The strategy is mainly tested as a simplified benchmark for the core DLY strategy, yet its performance is highly correlated with the other strategies, implying robustness of the general methodology.

As suggested, among others, by Bernardo and Ledoit (2000), and Mitchell and Pulvino (2001), Sharpe ratio is generally not a good measure of the risk-adjusted performance of a strategy, should the returns be highly non-normally distributed. As this is the case in most of the arbitrage strategies, including the yield curve arbitrage, one must consider also other metrics in order to attain a more robust picture of the attractiveness of the risk-adjusted returns in question. Having this in mind, I employ the Bernardo and Ledoit Gain-Loss measure, which compares the expected gains to the expected losses in a given return sample. A fairly valued investment in the risk-neutral world should have a Gain-Loss ratio of one, yet both Duarte et al. (2007) and this thesis show that it is possible for the returns to exhibit much higher ratios, which is obviously an attractive characteristic for the strategies.

The CIR two-factor model consistently outperforms the LS two-factor model. This may be due that CIR does not allow negative rates, or because it is analytically more tractable and thus possibly also more stable regarding numerical calibration. See the Section 6.4 for more discussion on the performance of the LS2F model.

As to the Hypotheses, I find strong evidence in support of the Hypothesis 1, which states that the replicated strategies generate attractive risk-adjusted returns. This is indeed the case, as is shown by e.g. the Sharpe and Gain-Loss ratios in Tables 1 to 6. Strong evidence is found also in support of the Hypothesis 2, which states that the returns are far from normally distributed, and tilted to the arbitrageurs’ benefit. This is shown by the positive skewness and high kurtosis of the returns, as well as the plotted distribution of the returns in Figure 3. The Hypothesis 3 states that following the exact strategy implementation of Duarte et al. (2007) will yield the best risk-adjusted performance. This is found to be true across the Tables 1 to 4.

Overall, the results are consistent with each other, and it can be concluded that slight modifications to the strategies do not cause major differences, especially given the data mining considerations. This is promising in the sense that the methodology proves to be robust to secondary changes in implementation, giving credibility to the core of the analytical methodology. The Hypotheses regarding the attractiveness of returns are accepted on the basis on strong evidence.

6.1. CIR2F return statistics

This subsection collects the tables of the summary statistics for the returns regarding the studied yield curve arbitrage strategies. The covered strategies are “DLY”, “Single Mispricing”, and “All Mispricings”. Elaboration on the strategy specifications can be found in Section 5.4 on the trading methodology. Shortly put, the “DLY” strategy is the exact replica following Duarte et al. (2007), where the strategy trades the largest mispricing every month, should the mispricing exceed a given limit of basis points. The DLY strategy can carry multiple trades at the same time. “Single Mispricing” trades also the largest mispricing every month, but can carry only a single trade at a time, i.e. no new trades are enter before the earlier one is closed. “All Mispricings” strategy trades all the mispricing every month that exceed the given basis point limit. Trades are closed based on the individual performance (or lack thereof) of each trade in the portfolio.

“Out-of-the-sample” refers to a methodology where a model is calibrated solely to a dataset that is different from the one where the trading is done. In contrast, “In-the-sample” means that a model is calibrated also to the data where the trading is done, i.e. the calibration includes also ‘future’ data points, which would not be available in the reality. The sample period for model calibration (when trading out-of-the-sample) is either a fixed four or eight year period, or a rolling four year period, as indicated case by case.

Table 1. Summary statistics for yield curve arbitrage strategies. The below table reports the indicated summary statistics for the monthly percentage excess returns of different yield curve arbitrage strategies modeled by the Cox-Ingersoll-Ross two-factor framework. 'DLY' refers to the strategy employed by Duarte et al. (2007) that trades the largest mispricing every month. The DLY is modified with a rolling calibration period where trading is done out-of-the-sample, as well as with an in-the-sample version. Different trigger levels for trading are indicated in the parenthesis. 'One mispricing' refers to a strategy where a single trade is held in the portfolio at a time. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to January 2015.

CIR2F model
4-year sample period, 5 / 10 bps limit to initiate a trade

Strategy	N	Capital	Mean	*t*-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
DLY (5bps)	110	6.132	0.780	2.79	2.887	-7.175	10.599	0.615	5.125	0.327	0.040	2.314	0.939
DLY Rolling (5bps)	110	5.959	0.900	2.94	2.887	-6.713	10.740	0.489	5.387	0.236	0.242	2.658	1.082
DLY In-Sample	157	5.348	0.750	3.20	2.887	-8.040	12.527	1.242	7.135	0.198	0.042	2.670	0.903
DLY In-Sample (5bps)	157	5.937	1.050	4.22	2.887	-7.243	11.285	0.603	5.110	0.274	0.166	3.034	1.257
One Mispricing	110	6.445	0.600	2.28	2.887	-7.758	12.724	0.965	6.636	0.282	-0.075	2.079	0.726

Table 2. Summary statistics for the yield curve arbitrage strategies (20 bps trigger). The below table reports the indicated summary statistics for the monthly percentage excess returns of different yield curve arbitrage strategies modeled by the Cox-Ingersoll-Ross two-factor framework. 'DLY' refers to the strategy employed by Duarte et al. (2007) that trades the largest mispricing every month. The DLY is modified with a rolling calibration period where trading is done out-of-the-sample, as well as with an in-the-sample version. Different trigger levels for trading are indicated in the parenthesis. 'One mispricing' refers to a strategy where a single trade is held in the portfolio at a time. 'All mispricings' refers to a strategy where all the mispricings are traded each month. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to January 2015.

CIR2F model
4-year sample period, 20 bps limit to initiate a trade

Strategy	N	Capital	Mean	t-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
DLY	110	4.583	0.890	3.20	2.887	-6.109	12.655	1.659	7.483	0.136	0.016	4.165	1.064
All mispricings	110	5.356	0.570	2.03	2.887	-9.521	11.949	1.072	9.275	0.155	0.051	2.524	0.688
One mispricing	110	4.380	0.780	2.80	2.887	-6.393	12.102	1.755	8.074	0.136	0.015	3.652	0.931

Table 3. Summary statistics for the yield curve arbitrage strategies (8-year sample). The below table reports the indicated summary statistics for the monthly percentage excess returns of different yield curve arbitrage strategies modeled by the Cox-Ingersoll-Ross two-factor framework. 'DLY' refers to the strategy employed by Duarte et al. (2007) that trades the largest mispricing every month. The DLY is modified with a rolling calibration period where trading is done out-of-the-sample, as well as with an in-the-sample version. Different trigger levels for trading are indicated in the parenthesis. 'One mispricing' refers to a strategy where a single trade is held in the portfolio at a time. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to January 2015.

CIR2F model
8-year sample period, 10 bps limit to initiate a trade

Strategy	N	Capital	Mean	t-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
DLY	62	5.406	0.950	2.82	2.887	-5.365	12.024	1.835	7.989	0.242	-0.171	3.714	1.135
One mispricing	62	6.105	0.730	2.19	2.887	-6.388	13.432	1.816	9.643	0.258	-0.198	2.634	0.871

Table 4. Summary statistics for the yield curve arbitrage strategies (8-year sample / 15 bps trigger). The below table reports the indicated summary statistics for the monthly percentage excess returns of different yield curve arbitrage strategies modeled by the Cox-Ingersoll-Ross two-factor framework. 'DLY' refers to the strategy employed by Duarte et al. (2007) that trades the largest mispricing every month. The DLY is modified with a rolling calibration period where trading is done out-of-the-sample, as well as with an in-the-sample version. Different trigger levels for trading are indicated in the parenthesis. 'One mispricing' refers to a strategy where a single trade is held in the portfolio at a time. 'All mispricings' refers to a strategy where all the mispricings are traded each month. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to January 2015.

CIR2F model
8-year sample period, 15 bps limit to initiate a trade

Strategy	N	Capital	Mean	t-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
DLY	62	4.038	0.980	2.94	2.887	-3.963	15.355	2.672	12.667	0.161	-0.192	5.352	1.170
All mispricings	62	4.251	1.170	3.37	2.887	-3.293	13.408	2.122	8.253	0.161	-0.099	6.854	1.409
One mispricing	62	3.990	0.980	2.96	2.887	-4.010	15.540	2.672	12.883	0.161	-0.196	5.300	1.171

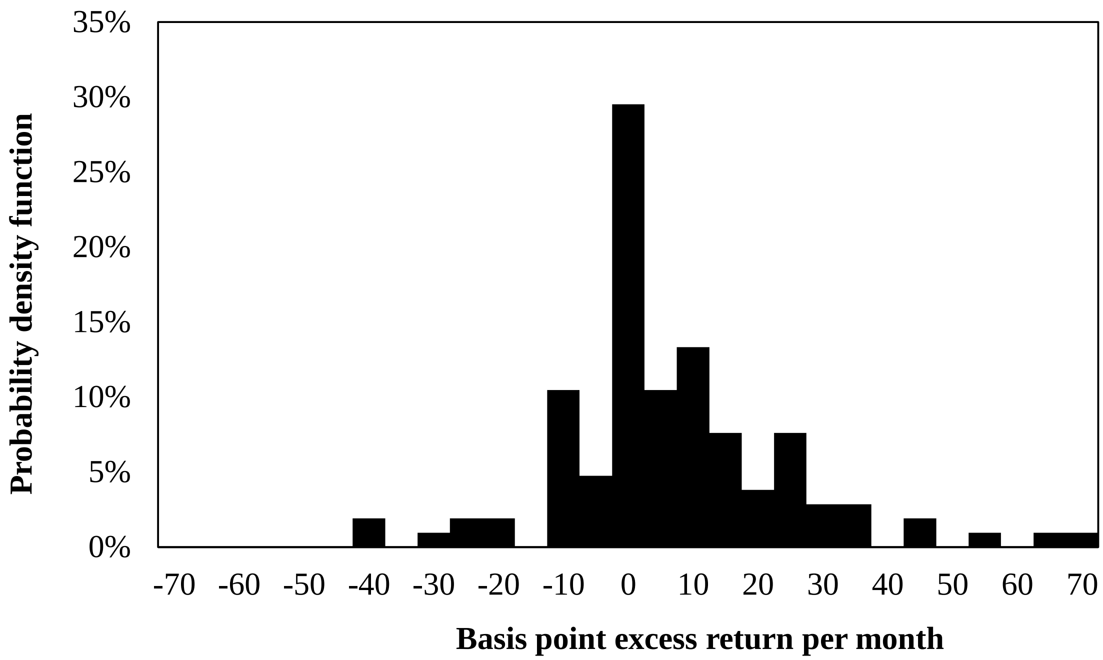

Figure 3. DLY strategy's return distribution.

Probability density function for the unleveraged core DLY strategy. Mean of the monthly excess returns is 4.8 bps and standard deviation 17.7 bps. The observed distribution is far from normal, i.e. it has fat tails and is positively skewed. Little evidence is found in support of frequent negative tail events.

6.2. CIR2F cumulative returns

This subsection illustrates graphically the cumulative returns of the explored yield curve arbitrage strategies. The covered strategies are "DLY", "Single Mispricing", and "All Mispricings". Elaboration on the strategy specifications can be found in Section 5.4 on the trading methodology. Shortly put, the "DLY" strategy is the exact replica following Duarte et al. (2007), where the strategy trades the largest mispricing every month, should the mispricing exceed a given limit of basis points. The DLY strategy can carry multiple trades at the same time. "Single Mispricing" trades also the largest mispricing every month, but can carry only a single trade at a time, i.e. no new trades are enter before the earlier one is closed. "All Mispricings" strategy trades all the mispricing every month that exceed the given basis point limit. Trades are closed based on the individual performance (or lack thereof) of each trade in the portfolio.

"Out-of-the-sample" refers to a methodology where a model is calibrated solely to a dataset that is different from the one where the trading is done. In contrast, "In-the-sample" means that a model is calibrated also to the data where the trading is done, i.e. the calibration includes also 'future' data points, which would not be available in the reality. The sample period for model calibration (when trading out-of-the-sample) is either a fixed four or eight year period, or a rolling four year period, as indicated case by case.

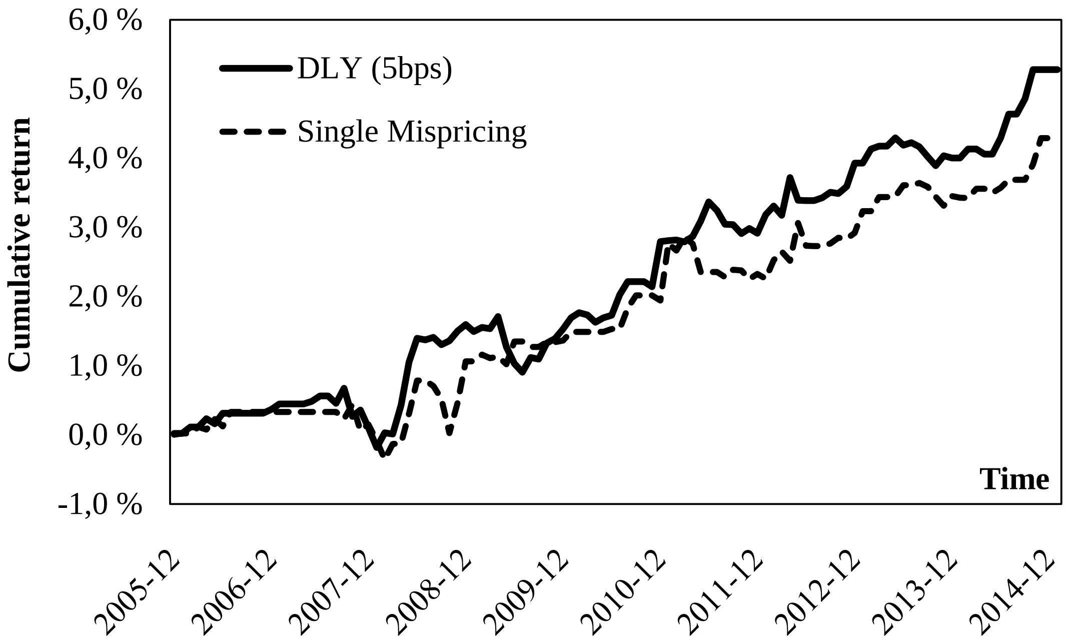

Figure 4. Cumulative returns for the DLY and Single Mispricings strategies.

The chart describes the evolution of the cumulative unleveraged returns of the out-of-the-sample 'DLY' (five bps trigger level) and 'Single Mispricing' (10 bps trigger level) strategies with a four year sample period for model calibration. The DLY results are nearly identical for the 10 bps trigger level, but in that case the strategy does not begin trading immediately. See the beginning of Section 6.2, or alternatively Section 5.4 for more detail on the strategy specifications.

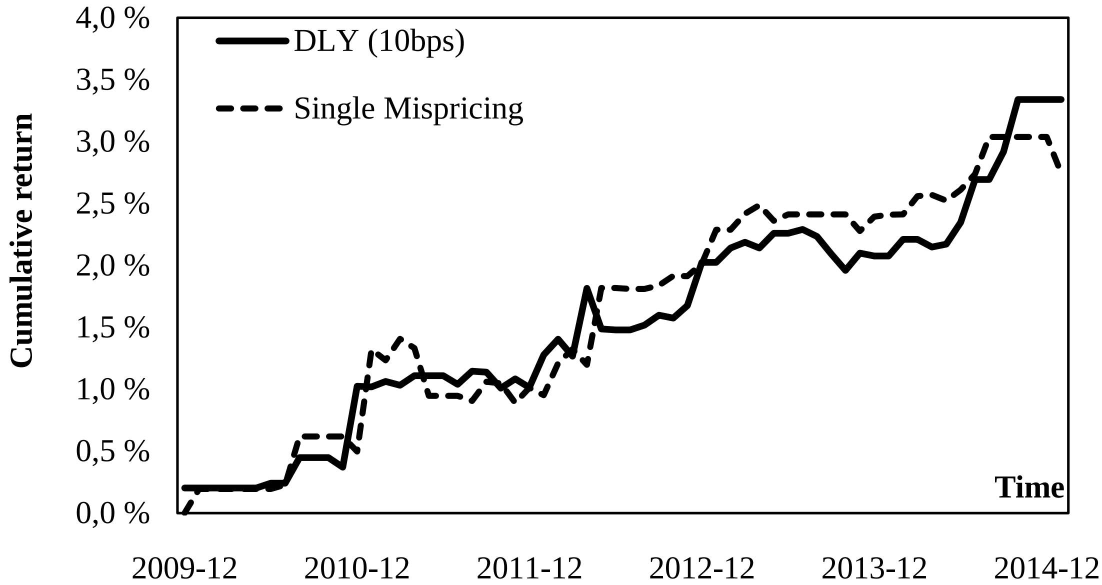

Figure 5. Cumulative returns for the DLY strategy with an 8-year sample.

The chart describes the evolution of the cumulative unleveraged returns for the in-the-sample 'DLY' and 'Single Mispricing' (10 bps trigger level) strategies with an eight year sample period for model calibration. See the beginning of Section 6.2, or alternatively Section 5.4 for more detail on the strategy specifications.

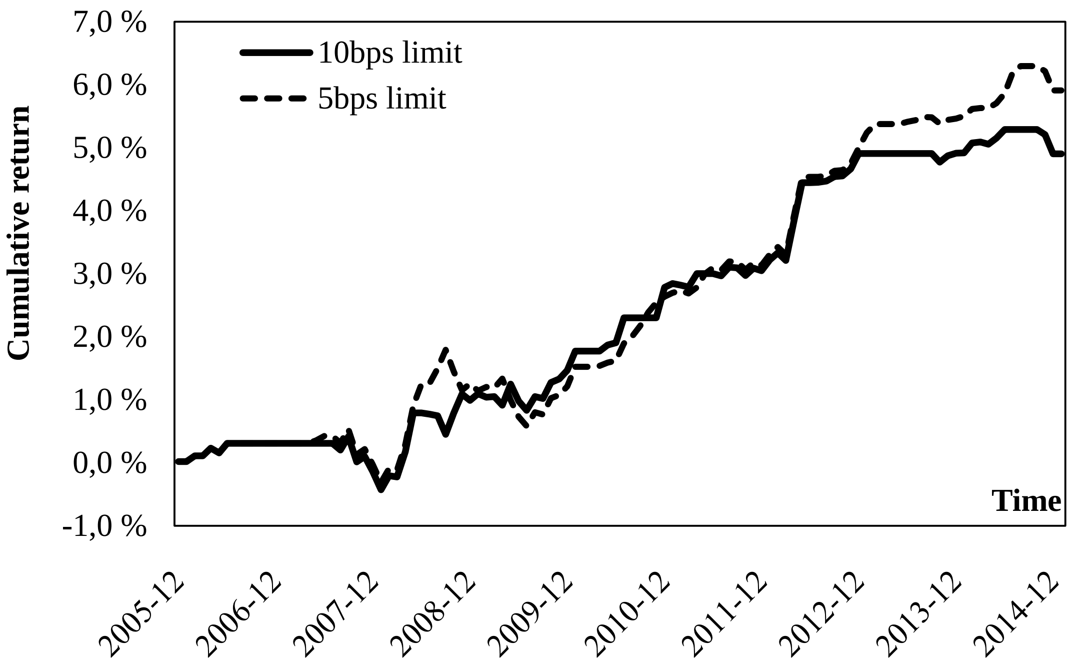

Figure 6. Cumulative returns for the DLY strategy with a rolling sample.

The chart describes the evolution of the cumulative unleveraged returns for the out-of-the-sample 'DLY' strategy (five and ten bps trigger levels), whose calibration period is always four years from the latest trading observation (trades only out-of-the-sample). See the beginning of Section 6.2, or alternatively Section 5.4 for more detail on the strategy specifications.

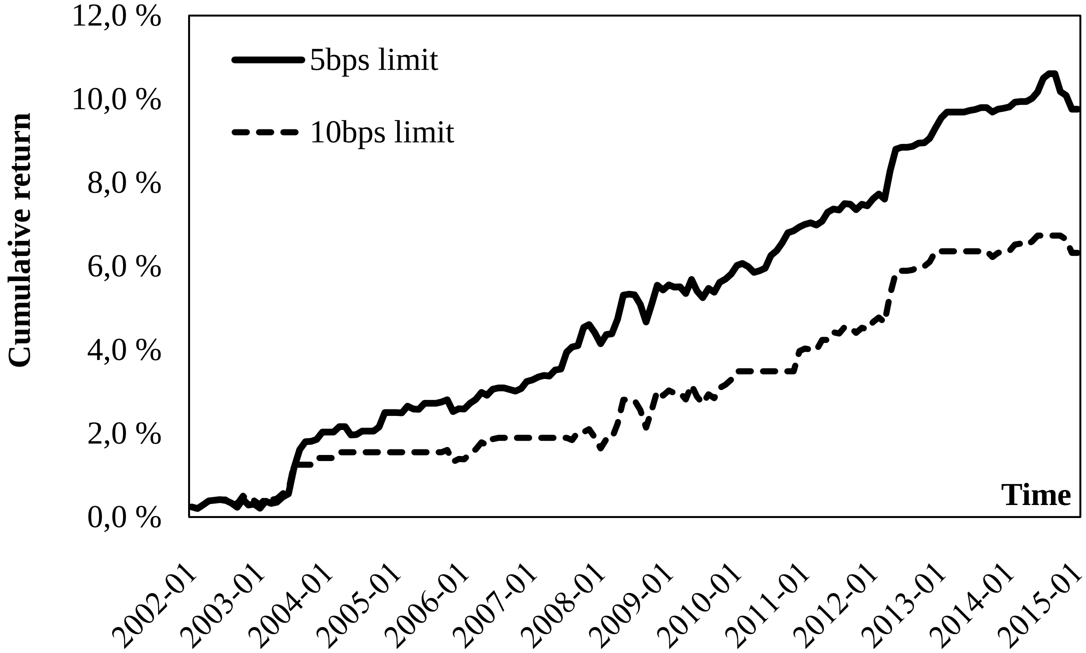

Figure 7. Cumulative returns for the in-the-sample DLY strategy.

The chart describes the evolution of the cumulative unleveraged returns for the in-the-sample 'DLY' strategy. The model is calibrated to the whole sample of rates, i.e. also to 'future' information. See the beginning of Section 6.2, or alternatively Section 5.4 for more detail on the strategy specifications.

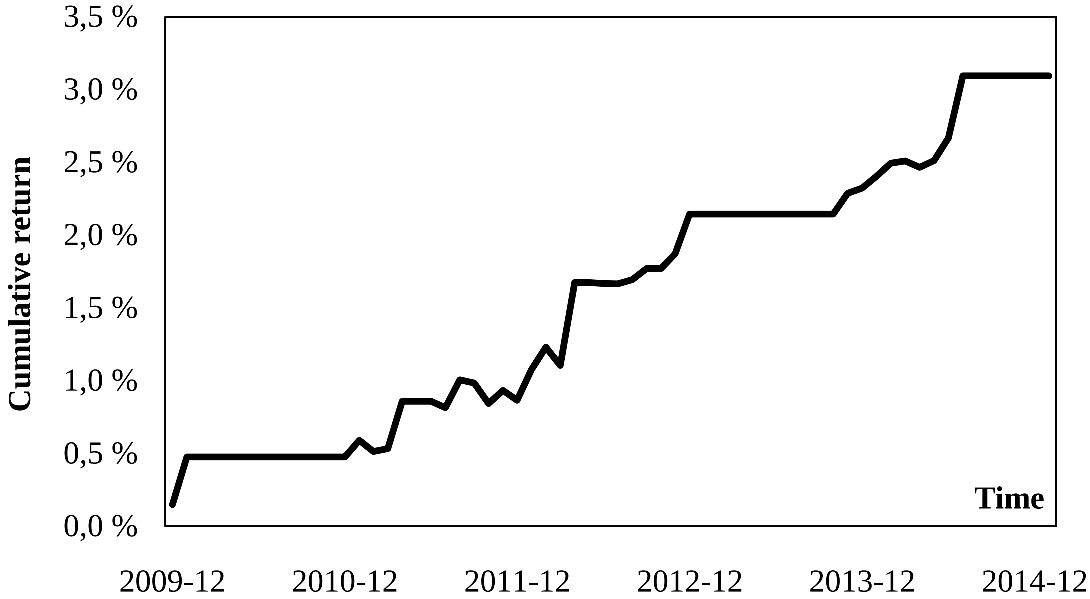

Figure 8. Cumulative returns for the out-of-the-sample ‘All Mispricings’ strategy (8-year sample).

The chart describes the evolution of the cumulative unleveraged returns of the out-of-the-sample strategy that trades all the mispricings every month, if the mispricing exceed the limit of 15 basis points. Trades are closed based on the individual performance. The model is calibrated to the first eight years in the overall sample, and it trades only out-of-the-sample. Performance of a strategy with a four year sample period for calibration yields similar returns, but trades less often in the beginning of the trading period. See the beginning of Section 6.2, or alternatively Section 5.4 for more detail on the strategy specifications.

6.3. CIR2F market-to-model differences

'Mispricing' is considered as the market-to-model difference in the respective swap rates. Figures 9 and 10 show the extent of the mispricings, as well as their evolution in time. Overall, the mispricings are small enough to make the model credible, yet large enough to generate economically significant trading signals. The mispricings of the rates seem to move in sync so that usually the sign of the difference is equal to all or most of the rates.

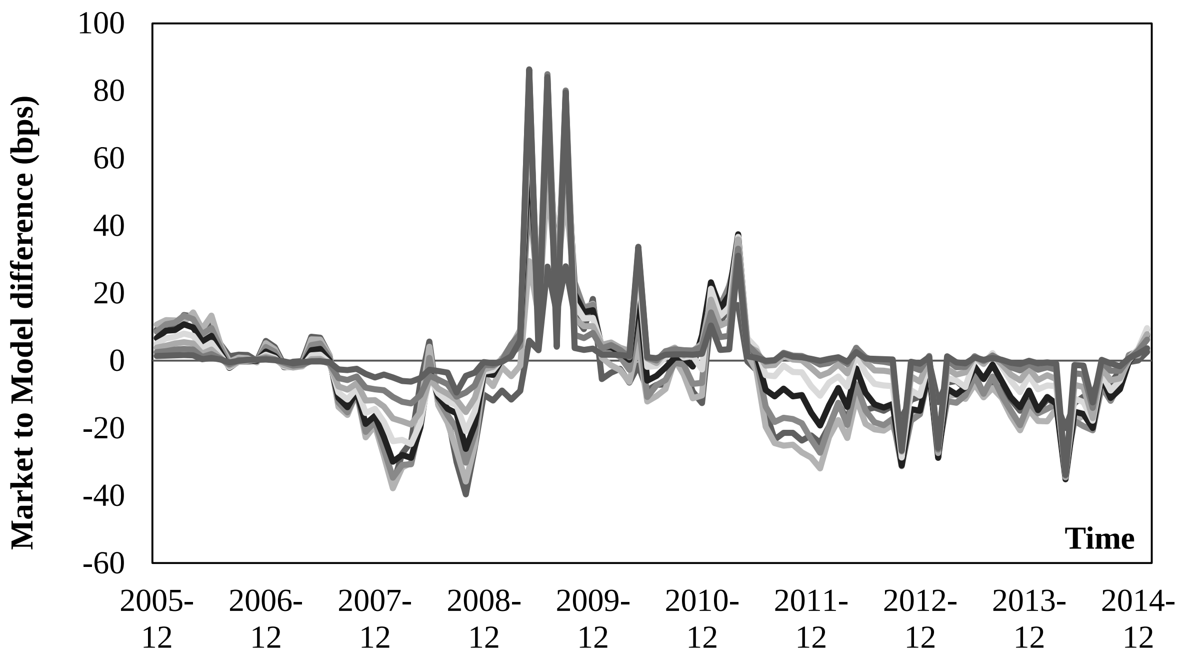

Figure 9. Market-to-model differences in the out-of-the-sample strategy.

The figure shows the basis point difference between the market and model implied swap rates for different maturities (2 to 9 years) in time as produced by the core DLY strategy. See the Appendix for the differences of individual rates.

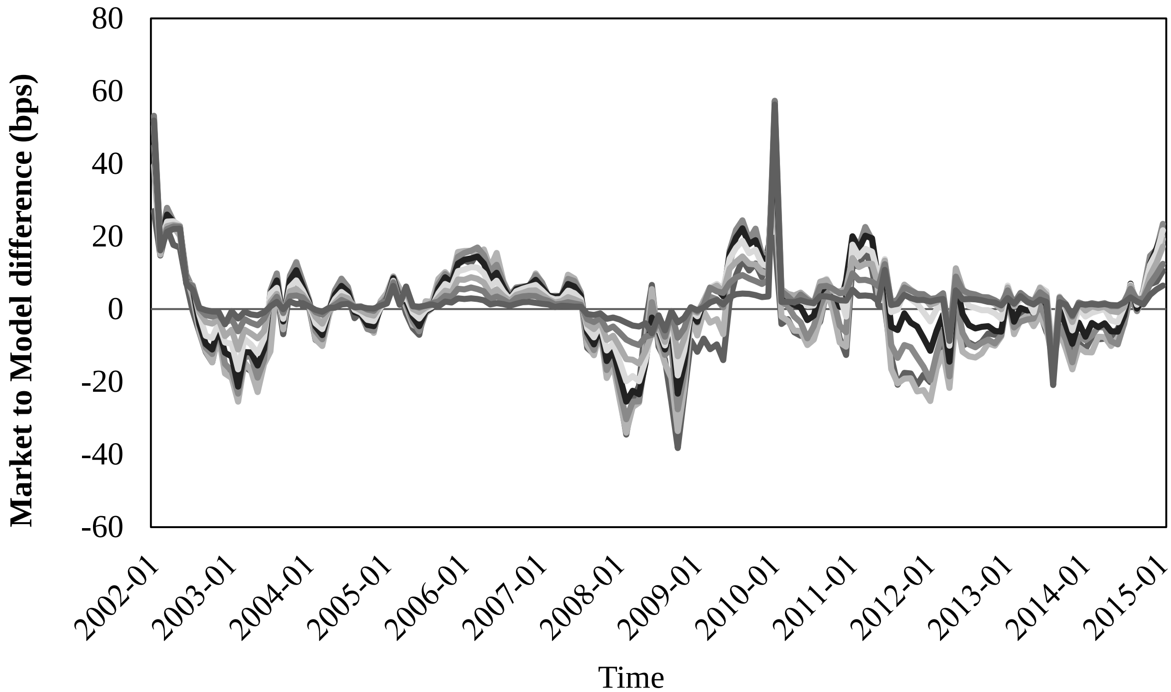

Figure 10. Market-to-model differences regarding the in-the-sample strategy.

The figure shows the basis point difference between the market and model implied swap rates for different maturities (2 to 9 years) in time as produced by the in-sample DLY strategy.

6.4. LS2F return statistics

Although the Longstaff-Schwartz two-factor model can be shown to equal the Cox-Ingersoll-Ross two-factor model under certain assumptions, it is analytically less tractable than the CIR model. It also allows for negative rates, which is yet to be true for the EUR swaps. Possibly because of these features, or due to instability with respect to calibration, the LS2F model by and far yields inferior trading results compared to the CIR2F model. The Return statistics section below shall describe the outcomes of a number of strategy variations. I will not analyze more in depth the results from this model, given that the CIR2F has shown more promise, and thereby constitutes the core model of this paper in analyzing the results.

Table 5. Summary statistics for the yield curve arbitrage strategies (LS2F). The below table reports the indicated summary statistics for the monthly percentage excess returns of different yield curve arbitrage strategies modeled by the Longstaff-Schwartz two-factor framework. 'DLY' refers to the strategy employed by Duarte et al. (2007) that trades the largest mispricing every month. The DLY is modified with a rolling calibration period where trading is done out-of-the-sample, as well as with an in-the-sample version. Different trigger levels for trading are indicated in the parenthesis. 'One mispricing' refers to a strategy where a single trade is held in the portfolio at a time. 'All mispricings' refers to a strategy where all the mispricings are traded each month. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to January 2015.

LS2F model
4-year sample period, 5 /10 bps limit to initiate a trade

Strategy	N	Capital	Mean	t-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
DLY	110	7.184	0.480	1.83	2.887	-9.466	11.554	0.676	5.928	0.327	-0.105	1.727	0.571
One mispricing	110	7.318	0.410	1.57	2.887	-7.516	11.343	0.750	5.235	0.327	-0.076	1.593	0.498
DLY In-Sample	157	9.504	0.520	2.24	2.887	-8.734	11.469	0.575	5.651	0.312	0.028	1.741	0.627
DLY In-Sample (5bps)	157	9.622	0.470	2.08	2.887	-8.626	13.199	0.683	6.303	0.369	-0.055	1.623	0.560

Table 6. Summary statistics for yield curve arbitrage strategies with an 8-year sample (LS2F). The below table reports the indicated summary statistics for the monthly percentage excess returns of different yield curve arbitrage strategies modeled by the Longstaff-Schwartz two-factor framework. 'DLY' refers to the strategy employed by Duarte et al. (2007) that trades the largest mispricing every month. The DLY is modified with a rolling calibration period where trading is done out-of-the-sample, as well as with an in-the-sample version. Different trigger levels for trading are indicated in the parenthesis. 'One mispricing' refers to a strategy where a single trade is held in the portfolio at a time. 'All mispricings' refers to a strategy where all the mispricings are traded each month. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to January 2015.

LS2F model
4-year sample period, 20 bps limit to initiate a trade

Strategy	N	Capital	Mean	t-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
DLY	110	6.416	0.540	2.10	2.887	-6.546	12.936	1.385	7.190	0.236	-0.140	2.010	0.644
All mispricings	110	7.037	0.540	2.20	2.887	-10.231	11.794	1.091	7.997	0.246	-0.200	2.057	0.649
One mispricing	110	6.664	0.520	2.05	2.887	-8.103	12.454	1.157	6.837	0.236	-0.160	1.960	0.622

7. Multifactor regression of the returns

Hedge funds' existence is largely based on the idea that they can generate alpha, or uncorrelated returns that cannot be explained by well-known risk factors (betas). If the returns are not statistically explained by the betas, it can be deduced that they do not stem from a priced risk exposure, which is a valuable property for an investment. As suggested by Vayanos and Vila (2009), the returns available for yield curve arbitrageurs should be related to the bond risk premia. Therefore, instead of being a genuine arbitrage, the trading of mispriced rates is assumed to carry risk. The risk should be highly limited, though, given the fairly sophisticated two-factor butterfly hedging strategy employed, where both the parallel shifts as well as twists of the curve are hedged extensively. In other words, the yield curve arbitrage strategy is a market neutral one, just as many relative value bets characteristically are. Market neutrality, however, does not mean that the strategy is neutral to all risks; in fact, the neutrality can be seen as a way to isolate more elusive risk premia, in this case the local bond risk premia that is assumed to converge.

As discussed by Fung and Shieh (1997), factor analysis of the market neutral funds is extraordinarily challenging, as exposure to the market betas is by construction made very close to zero. Even the analysis of funds that are *not* market neutral is demanding enough, due to the highly dynamic asset allocation strategies. For one, a certain hedge fund might never be genuinely market neutral, but if it alters long and short positions in an asset frequently enough and in a balanced way, its return correlation with the asset will approach zero in a sufficiently long time period.

In spite of the complexities regarding the market neutral strategies, I shall follow Duarte et al. (2007) in building the multifactor regression model for explaining the replicated arbitrage returns. To gain insight into whether the yield curve arbitrage as a trading strategy generates alpha, it is necessary to run such a regression analysis, where one controls the returns of the replicated strategy by the risk factors that have been shown to systematically explain returns elsewhere.

Ever since Fama and French (1992), equity-related returns are often controlled by at least the original three Fama-French factors, where the CAPM by Sharpe (1964) and Lintner (1965) is augmented by the small-minus-big (SMB) and high-minus-low (HML) factors that control for size and value premiums, respectively. One could add any number of equity-related long/short portfolios or indices as controls to the regression model, and I shall follow Duarte et al. (2007) by going with a four-factor model that incorporates the momentum (WML) factor to the classical Fama-French model. Additionally, given the banking context of fixed income trading, bank stock portfolios (S&P 500 Banks and MSCI European Financials) are used here as additional equity controls for the strategies.

In the fixed income space, one would generally seek to attain control for systematic rate and credit factors. On the lines of Duarte et al. (2007), I choose to employ industrial and bank bond (rated A/BBB, or investment grade) index excess returns as controls for the credit spread premia. Moreover, Duarte et al. used Fama-French Treasury portfolios (two, five, and ten year maturities) to control for rates. I control for the rates by employing EUR constant maturity swap rates directly, so that the duration adjusted returns from receiving the fixed constant maturity swap rates are used as independent variables in the regression (all of the rates from one to ten years). Both the Treasury and swap portfolios should capture the roll-down effect of the yield curve, i.e. the term premia of interest rates.

The regression equation (Equation 11) is formulated below, and summary statistics for the regression output are reported in Table 7 for the core DLY strategy (5 bps trigger level), its in-the-sample version, and for the relevant hedge fund indices. R_{S1} is the European bank stock index (MSCI Financials Europe), and R_{S2} is the U.S. equivalent (S&P 500 Banks). R_I is an industrial bond portfolio, and R_B a bank bond portfolio, where the subscript states whether it is A or BBB rated iBoxx total return index. R_j refers to the returns from receiving the fixed leg in a constant maturity swap of maturity j. The returns are in excess of the risk-free rate, except for the swaps.

$$R_{i,t} = \alpha + \beta_1 R_{M,t} + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 WML_t + \beta_5 R_{S1,t} + \beta_6 R_{S2,t} + \beta_7 R_{I_A,t} + \beta_8 R_{I_{BBB},t} + \beta_9 R_{B_A,t} + \beta_{10} R_{B_{BBB},t} + \sum_{j=1}^{10} \beta_{10+j} R_{j,t} + \varepsilon_t \qquad (11)$$

Table 7. Multifactor regression summary statistics. This table reports the indicated summary statistics for the regression of the replicated DLY strategies' monthly excess returns on the excess returns of the indicated equity and bond portfolios. DLY refers to the yield curve arbitrage strategy modeled directly based on Duarte et al. (2007), but which trades out-of-the-sample. 'In-sample' refers to the strategy that trades in-the-sample. Results for the CS and HFR hedge fund indices are also reported. Descriptions of the hedge fund indices are given in Section 8. RM is the excess return on a value weighted index, SMB, HML and WML are the Fama-French small-minus-big, high-minus-low, and winner-minus-loser market factors, respectively. The Fama-French factors are computed with European data. R_S is the excess return of an index of bank stocks both in Europe and in the United States. R_I and R_B are the excess returns of iBoxx A/BBB-rated industrial and bank bond total return indices, respectively. R_2, R_5, and R_{10} are portfolios that receive the fixed rate in an indicated constant maturity EUR swap. The sample period for the indicated strategies is December 2005 to January 2015.

Strategy	α	*t*-Stat.	R_M	SMB	HML	WML	$R_{S,\,USA}$	$R_{S,\,EUR}$	$R_{I,\,A}$	$R_{I,\,BBB}$	$R_{B,\,A}$	$R_{B,\,BBB}$	R_2	R_5	R_{10}	R^2
DLY (5bps)	1.22	3.89	-0.32	0.90	-1.22	-1.36	0.16	0.89	-0.41	0.20	1.23	-2.46	0.60	-1.16	-1.56	0.20
DLY in-sample	1.24	3.93	1.17	1.52	-0.30	-0.22	-1.13	-0.73	-1.25	-0.21	2.38	-1.84	-0.44	-1.46	-1.28	0.26
CS HF Index	0.28	3.26	7.04	4.80	-0.39	3.19	-2.60	-0.22	-1.70	1.81	2.83	-0.62	0.26	0.54	2.05	0.83
CS FI Arb	0.23	1.85	1.62	0.11	-0.19	-0.85	-1.44	-4.16	7.72	5.23	-3.09	-2.16	-0.49	-0.80	0.00	0.76
CS Global Macro	0.47	3.72	5.84	2.34	1.10	-0.39	-4.03	-2.98	-2.74	2.42	4.21	-3.20	0.61	0.70	1.33	0.57
CS Multi Strategy	0.41	4.53	3.22	5.44	-1.83	3.20	-2.78	2.14	-0.73	1.93	2.65	-0.30	0.52	1.06	1.33	0.81
HFRI Fund Weighted Comp.	0.21	2.73	8.91	5.17	-1.41	2.26	-1.25	0.18	-1.14	1.33	1.71	0.08	-1.06	0.36	1.56	0.89
HFRI RV Total	0.41	5.81	3.54	4.01	-0.62	0.82	-2.89	0.54	-2.32	5.32	3.43	-0.88	-0.32	0.36	1.74	0.85
HFRX RV FI Sov.	0.11	0.71	3.90	1.08	-0.61	-1.21	-2.62	-0.87	-2.76	5.52	4.64	-4.83	-0.33	2.24	0.03	0.70
HFRI RV FI Corp.	0.22	2.32	1.16	2.74	-0.72	1.46	-0.66	1.03	-2.53	5.51	2.27	1.20	-0.45	0.68	1.13	0.84
HFRI Macro Systematic	0.25	1.13	4.02	1.62	0.00	2.01	-1.48	-0.68	-0.55	-2.43	-0.13	0.96	-0.92	-0.83	1.14	0.41
HFRX Western/Pan Europe	0.46	3.32	1.95	6.34	-1.80	4.50	-0.45	4.34	0.29	-1.44	1.72	-2.11	-0.98	1.02	1.11	0.69

As can be seen from the summary statistics in Table 7, the multifactor regression model explains very little of the replicated arbitrage strategy's ("DLY") returns compared to the hedge fund indices. Around 80% of the return variation in the hedge fund indices is accounted for by the model, as implied by the R^2. The regression shows a considerably higher monthly alpha of 1.2% for the replicated strategy than for any of the hedge fund indices, out of which the Global Macro style fares the best with its 0.47% alpha. The alpha is also easily statistically significant, whereas, for example, the CS Fixed Income Arbitrage index struggles to attain the 5% significance level for the returns. Out of the control variables, only the BBB-rated bank bond index has statistically significant explanatory power on the strategy's returns, coefficient for which is actually -0.48, meaning that the replicated strategy is short the BBB bank bonds. To elaborate on this, Fung and Hsieh (2002) find that by and large, fixed income hedge funds have static exposure to fixed income spreads, with weak evidence that these funds employ convergence and market timing strategies. This regression in contrast implies little to no static exposure to such spreads regarding the yield curve arbitrage's returns.

My regression model explains the arbitrage strategy's return variation better than reported previously, as the R^2 is now 20%, whereas Duarte et al. report a figure of 10% or less. This can be seen as an evidence of a weakening relation between the indices and the strategy, as my model explains starkly better the indices than the replicated returns, while Duarte et al.'s model does not explain either well. Discrepancy in explanatory power is interesting, given that it implies a genuinely uncorrelated alpha for the replicated strategy particularly in the more recent observation period.

Turning attention to the hedge fund indices in the regressions, Duarte et al. (2007) report mostly insignificant loadings to the risk factors when explaining the indices, as well as a low R^2. My regression model explains the return variation of the hedge fund indices rather robustly in contrast to Duarte et al. This may be due to the dilution of the actual arbitrageurs' alphas in time, as Duarte et al. studied an earlier period of 1988-2004, after which some alphas have shown a tendency to decline, as is discussed in Section 3.4.

In all, I find strong support for the Hypothesis 4 stating that the replicated core strategy's returns are multifactor alpha. As the Table 7 illustrates, the priced risk factors explain relatively little of the return variation of the replicated DLY strategy. Hence, also the alpha is large and statistically significant. Moreover, there is a stark contrast between the alphas of the hedge fund indices and the replicated strategy, even after noting that the hedge fund returns are after fees. This further lends support for the idiosyncratic nature of the returns generated by the DLY strategy.

To conclude, it seems that while most of the hedge fund and fixed income arbitrage indices have significant positive alpha, the replicated yield curve arbitrage strategy has this effect even more pronounced. It is hard to explain the replicated returns of the yield curve arbitrageurs by the known and priced risk factors. This implies that the yield curve arbitrageurs who follow the here outlined quantitative methodology to pick relative value market neutral bets do enjoy economically and statistically significant multifactor alpha. This is broadly consistent with Vayanos and Vila (2009), as their model predicts that the risk-averse arbitrageurs would trade only when compensated sufficiently for the bond risk premia related uncertainty in their positions. Vayanos and Vila do not imply whether their arbitrageurs should be able to gain extraordinary risk-adjusted returns. It is known that their arbitrageurs do hedge in a similar fashion than in this thesis and in Duarte et al. (2007), indicating that low market exposure is expected also from the theoretical strategies.

As a final point to the literature discussing tail risk compensation in hedge fund returns, I find no evidence (in here or in Section 6) that the considerable tail events such as the 2007 quant meltdown, the 2007-2008 global financial crisis, the 2012 euro sovereign crisis nor the highly unconventional central bank policies of the Federal Reserve, Bank of Japan, as well of the ECB, would have constituted such a realized negative outcome in terms of large losses (diminished alphas) to the strategy. Hence, the yield curve arbitrage strategy’s alpha is difficult to see solely as a compensation for the tail risk. On the other hand, a positive tail event could be identified in the form of the extraordinary monetary policies, as Vayanos and Vila (2009) show that the central bank policies cater to the arbitrageurs. This phenomenon may thus have been in motion ever since the 2008 meltdown, after which the Fed begun the first round of Quantitative Easing.

8. Hedge fund indices and replicated returns

At this point we have replicating the returns available for arbitrageurs involved in the integration of the yield curve as described in theory by Vayanos and Vila (2009), and further discussed in Section 3. Now, a question naturally arises whether the actual fixed income arbitrageurs have had similar results. If the replicated returns resemble those of the related hedge fund indices (from HFR and Credit Suisse), one can say that the replicated arbitrage strategy characterizes what many of the actual funds do. If there is no connection, it is likely that the strategy or methodology introduced in this thesis differ enough from the average strategies captured by the hedge fund indices so as to have explanatory power. In other words, it is possible that even hedge fund subindices contain rather heterogeneous strategies either by nature or methodology.

Whether the replicated returns do or do not explain the hedge fund index returns is interesting for different reasons. Firstly, if there is a connection between the replicated and actual returns, then one can say that leaning on the methodology in this paper, we know to a certain degree what the hedge funds in this space are doing. Secondly, if there is no connection, then we can say that the methodology developed here differs enough from what managers on average employ so as to offer a differentiated source of returns, and alpha. Differentiated alphas would actually be even more precious than those copied from the existing funds, as they would be less prone to competition, and thereby would also be less likely to suffer from liquidity or similar shocks to crowded positions among the arbitrageurs.

I shall give a brief introduction to the hedge fund indices here, further illustrated by the Figures 11 and 12. CS Hedge Fund Index is a Credit Suisse's aggregate hedge fund index, the other indices being specific strategy subindices. Credit Suisse Fixed Income Arbitrage Index (CS FI Arb.) is a subindex made of funds that have said they focus on making relative value bets in the fixed income space. CS Global Macro is an index made of funds focusing on trading rates, currencies and equities globally, either on a directional or relative value basis. CS Multi Strategy index is comprised of funds that have indicated that they are inclined to run a set of different strategies given market conditions. HFRI Fund Weighted Composite Index is HFR's aggregate hedge fund index, the other HFR indices being specific strategy subindices. HFRI RV Total is an aggregate relative value index, under which all the relative value funds belong regardless of an asset class focus. HFRI Macro Total is a subindex made of funds focusing on trading rates, currencies and equities globally, either on a directional or relative value basis. HFRI Macro Systematic is a global macro subindex that comprises of funds focusing on trading e.g. rates, currencies and equities globally, either on a directional or relative value basis, so that the trading signals are generated through quantitative, systematics methods. HFRX RV FI Sovereign is a fixed income subindex whose participant funds focus on trading government debt securities on a relative value basis. HFRX Western/Pan Europe is a cross-strategy index of funds that focus on trading exposures in the geographical area of Europe.

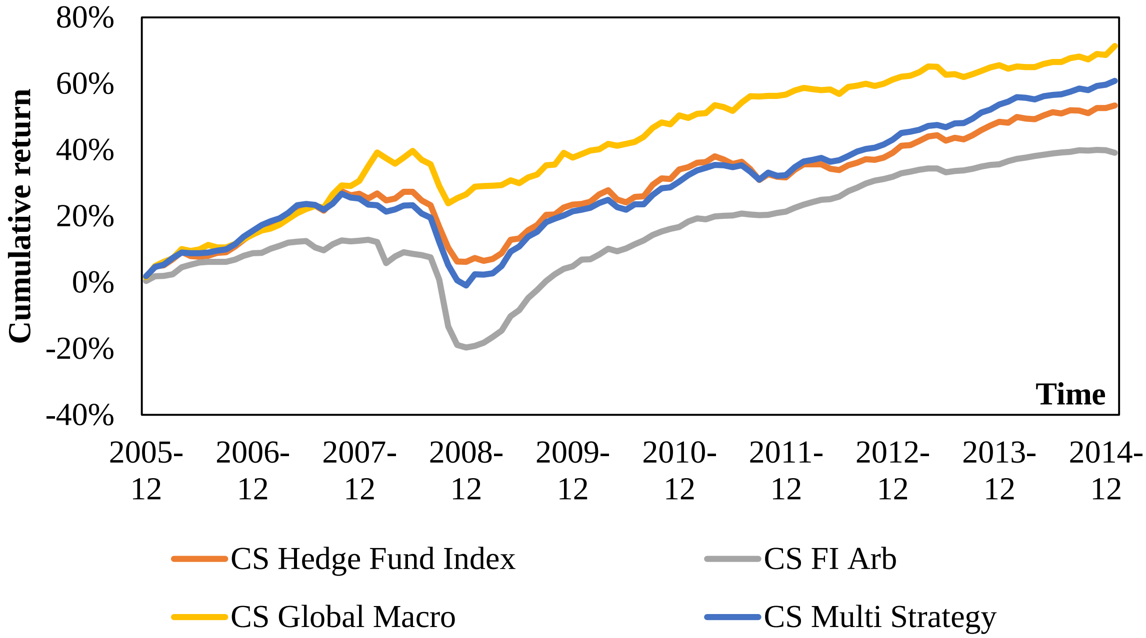

Figure 11. Credit Suisse hedge fund index data.

The figure illustrates the Credit Suisse cumulative hedge fund index and subindex returns that are considered relevant regarding the yield curve arbitrage strategy. CS Hedge Fund Index is a Credit Suisse's aggregate hedge fund index, the other indices being specific strategy subindices. CS FI Arb is a fixed income arbitrage subindex, i.e. an index made of funds that have said that they focus on making relative value bets in the fixed income space. CS Global Macro is an index made of funds focusing on trading rates, currencies and equities globally, either on a directional or relative value basis. CS Multi Strategy index is made of funds that have indicated that they are inclined to run opportunistically a set of different strategies given market conditions.

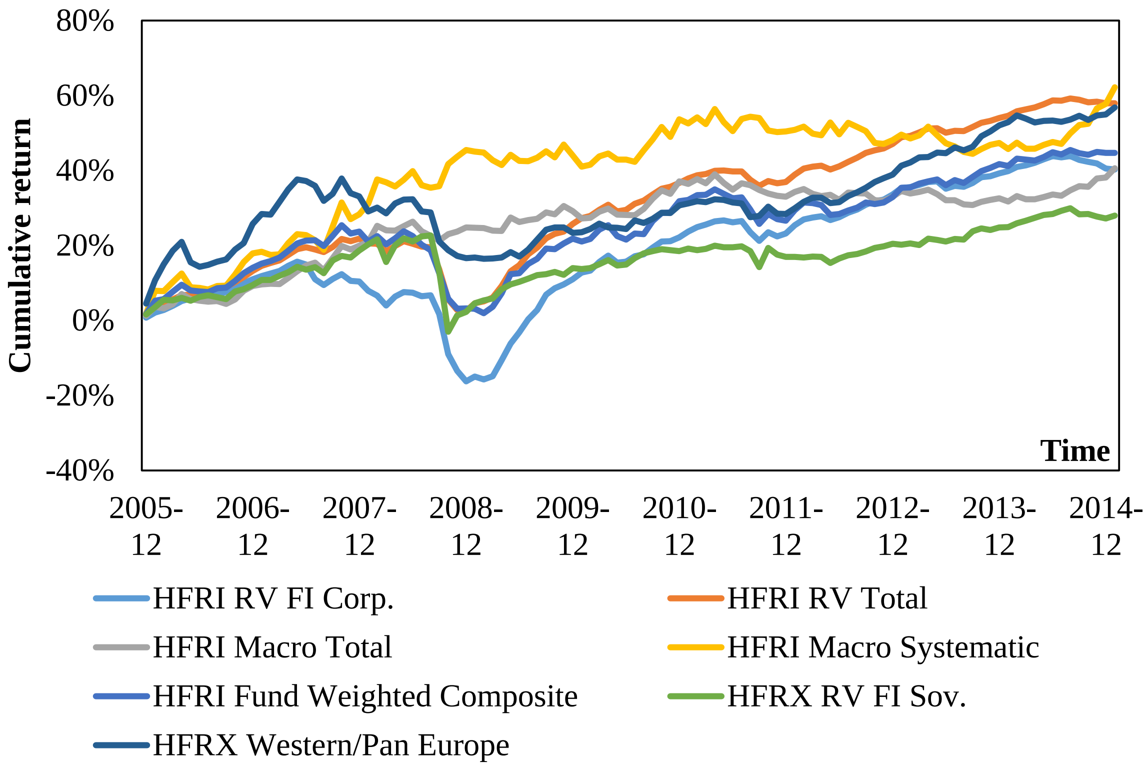

Figure 12. Hedge Fund Research hedge fund index data.

The figure illustrates the Hedge Fund Research cumulative hedge fund index and subindex returns that are considered relevant regarding the yield curve arbitrage strategy. HFRI Fund Weighted Composite Index is HFR's aggregate hedge fund index, the other indices being specific strategy subindices. HFRI RV Total is an aggregate relative value index, under which all relevant value strategies funds belong regardless of an asset class focus. HFRI Macro Total is a subindex made of funds focusing on trading rates, currencies and equities globally, either on a directional or relative value basis. HFRI Macro Systematic is a global macro subindex made of funds focusing on trading rates, currencies and equities globally, either on a directional or relative value basis, so that the trading signals are generated through quantitative, systematics methods. HFRX RV FI Sovereign is a fixed income subindex whose participant funds focus on trading government debt securities on a relative value basis. HFRX Western/Pan Europe is an aggregate index of funds who focus on trading exposures in the geographical area of Europe.

As illustrated by the Tables 8 to 11 below, the replicated returns have rather little statistically significant connection to the related hedge fund indices. Correlation between the indices and the strategy is at most 12.8% for the out-of-the-sample strategy (Table 8). At best, the relation is spurious and subject to data mining bias, as the correlations, for instance, are highly dependent on the time period of observation. It would be difficult to show in a robust manner that there existed a systematic relation between the indices and the replicated returns. This is consistent with the Duarte et al. (2007), where they find that actual-to-replicated hedge fund correlations range from -10% to 30% for all the fixed income arbitrage strategies, and from -2% to 2% for the yield curve arbitrage strategies in particular. The most likely explanation for this is that the indices are made of tons of somewhat varied strategies, many of which are completely different in nature and some of which have a different modeling methodology even if the basic idea was the same. Especially the market neutrality of the replicated strategy can be seen as a hindrance to finding a statistically significant connection.

In contrast, Mitchell et al. (2001) find that their replicated merger arbitrage strategy has a 36% correlation with the general HFR index, and a somewhat varying correlation with the selection of individual hedge funds, with a correlation coefficient up to 41% within the whole sample. In this regard, the merger arbitrage replication seems to better explain the hedge funds' behavior, yet the sample period and market focus is rather different from my work. Regarding other hedge fund strategies, Jylhä and Suominen (2009) find that their risk-adjusted carry trade strategy explains 16% of the overall hedge fund index returns, and 33% of the fixed income arbitrage subindex returns.

The Hypothesis 5 stating that the high-level hedge fund indices do not have a statistically significant relation with the replicated strategy is accepted. This is evident given the regression outputs (Tables 10 and 11) and the levels of correlation (Tables 8 and 9). Interestingly enough, the Hypothesis 6 stating that the most relevant hedge fund subindices have a statistically significant connection with the replicated returns is discarded. Thus, even if the most relevant style indices comprised yield curve arbitrage funds, this does not come across from the data. A number of issues could explain this, ranging from the diversity of the indices to the differences in strategy implementations. Of course, it also possible that the relative value or fixed income arbitrage indices do not actually comprise of funds making this kind of yield curve arbitrage bets, at least not in the EUR space.

Finally, I make the inference that even the hedge fund subindices – such as Fixed Income Arbitrage or Global Macro – are diversified enough not to explain the replicated yield curve arbitrage in the EUR swap rates space. Moreover, I would suggest that the U.S. market focus of the hedge fund indices in any case limits the exposure to the EUR fixed income arbitrage even further. Evidence to this direction comes from the fact that it is the European focused hedge fund index HFRX Western/Pan Europe that has the highest correlation and strongest sensible regression outcome with respect to the replicated returns. The conclusion is that the yield curve arbitrage, besides yielding multifactor alpha, also produces 'hedge fund alpha' in the sense that no combination of the relevant hedge fund indices can explain the returns well.

Table 8. Out-of-the-sample strategy's correlation with hedge fund indices. The table illustrates the correlations between the different hedge fund indices and subindices and the replicated, leveraged yield curve arbitrage strategy referred to as the DLY. The DLY is the strategy implemented following the methodology of Duarte et al. (2007), further discussed in Section 5.4. Out-of-the-sample means that the model is calibrated to a different dataset where the trading is done. The time period for computing the correlations is December 2005 to January 2015. Descriptions of the hedge fund indices are given earlier in this Section 8.

	CS Hedge Fund Index	CS FI Arb	CS Global Macro	CS Multi Strategy	HFRI RV FI Corp.	HFRI RV Total	HFRI Macro Total	HFRI Macro Systematic	HFRI Fund Weighted Composite	HFRX RV FI Sov.	HFRX Western/Pan Europe
Leveraged DLY	5.4 %	0.4 %	0.8 %	0.0 %	2.5 %	1.8 %	2.7 %	-1.4 %	6.9 %	6.3 %	12.8 %

Table 9. In-the-sample strategy's correlation with hedge fund indices. The table illustrates the correlations between the different hedge fund indices and subindices and the replicated, leveraged yield curve arbitrage strategy referred to as the DLY. The DLY is the strategy implemented following the methodology of Duarte et al. (2007), further discussed in Section 5.4. In-the-sample means that the model is calibrated to the same dataset where the trading is done. The time period for computing the correlations is January 2002 to January 2015. Descriptions of the hedge fund indices are given earlier in this Section 8.

	CS Hedge Fund Index	CS FI Arb	CS Global Macro	CS Multi Strategy	HFRI RV FI Corp.	HFRI RV Total	HFRI Macro Total	HFRI Macro Systematic	HFRI Fund Weighted Composite	HFRX RV FI Sov.	HFRX Western/Pan Europe
Leveraged DLY	9.0 %	-5.5 %	12.5 %	6.7 %	-6.8 %	3.1 %	19.3 %	15.1 %	7.1 %	2.0 %	17.4 %

Table 10. Explaining the out-of-the-sample replicated returns with hedge fund indices. The table shows the output of a regression where the monthly excess returns of the core DLY strategy trading out-of-the-sample are explained with the indicated hedge fund indices' returns. Descriptions of the hedge fund indices are given earlier in this Section 8. DLY refers to the yield curve arbitrage strategy that is modeled directly after Duarte et al. (2007). Stars indicate standard levels of statistical significance.

Factor	Coefficient	t-Stat.
Constant	0.01**	3.10
CS Hedge Fund Index	2.06*	1.68
CS FI Arb	-0.06	-0.16
CS Global Macro	-0.28	-0.68
CS Multi Strategy	-1.53**	-2.14
HFRI RV FI Corp.	0.44	0.87
HFRI RV Total	-0.73	-0.73
HFRI Macro Total	0.91	1.21
HFRI Macro Systematic	-0.70*	-1.87
HFRI Fund Weighted Comp.	-0.84	-0.97
HFRX RV FI Sov.	0.23	0.93
HFRX Western/Pan Europe	0.51**	2.16
R^2		0.11

Table 11. Explaining the in-the-sample replicated returns with hedge fund indices. The table shows the output of a regression where the monthly excess returns of the core DLY strategy trading in-the-sample are explained with the indicated hedge fund indices' returns. Descriptions of the hedge fund indices are given earlier in this Section 8. DLY refers to the yield curve arbitrage strategy that is modeled directly after Duarte et al. (2007). Stars indicate standard levels of statistical significance.

Factor	Coefficient	t-Stat.
Constant	0.01	1.61
CS Hedge Fund Index	0.14	0.11
CS FI Arb	-0.27	-0.70
CS Global Macro	-0.17	-0.39
CS Multi Strategy	-0.07	-0.10
HFRI RV FI Corp.	-0.90*	-1.70
HFRI RV Total	1.75*	1.70
HFRI Macro Total	1.31*	1.69
HFRI Macro Systematic	-0.48	-1.23
HFRI Fund Weighted Comp.	-0.86	-0.97
HFRX RV FI Sov.	0.04	0.14
HFRX Western/Pan Europe	0.41*	1.70
R^2		0.14

9. Arbitrage returns and noise in the markets

As is discussed in Section 3, building on Vayanos and Vila (2009), and Greenwood and Vayanos (2014), one can draw the assumption that high arbitrageur risk-aversion, as well as diminished arbitrageur capital (endured losses), will limit the arbitrageurs' capability to integrate maturity markets. Decreased arbitrageur participation would make other factors more pronounced in explaining the future bond returns. Namely, bond supply and yield curve slope would become stronger attributes, as shown by Greenwood and Vayanos (2014). Given the decreased arbitrageur participation, local demand-supply shocks would affect the term structure with more impact, as the arbitrageurs would not have the firepower to connect the curve to the information in the short-rate they perceive and trade, as discussed in Section 3.1. In all, one can connect the high arbitrageur risk-aversion to the amount of noise in the market prices. In other words, it can be assumed that high noise coincides with the arbitrageurs withdrawing bets and enduring losses. In this thesis, 'noise' refers to the measure developed by Hu, Pan and Wang (2013), which incorporates U.S. Treasury bond data. The measure is discussed more in depth in Section 3.5.3. Figure 13 below depicts the measure's evolution in time.

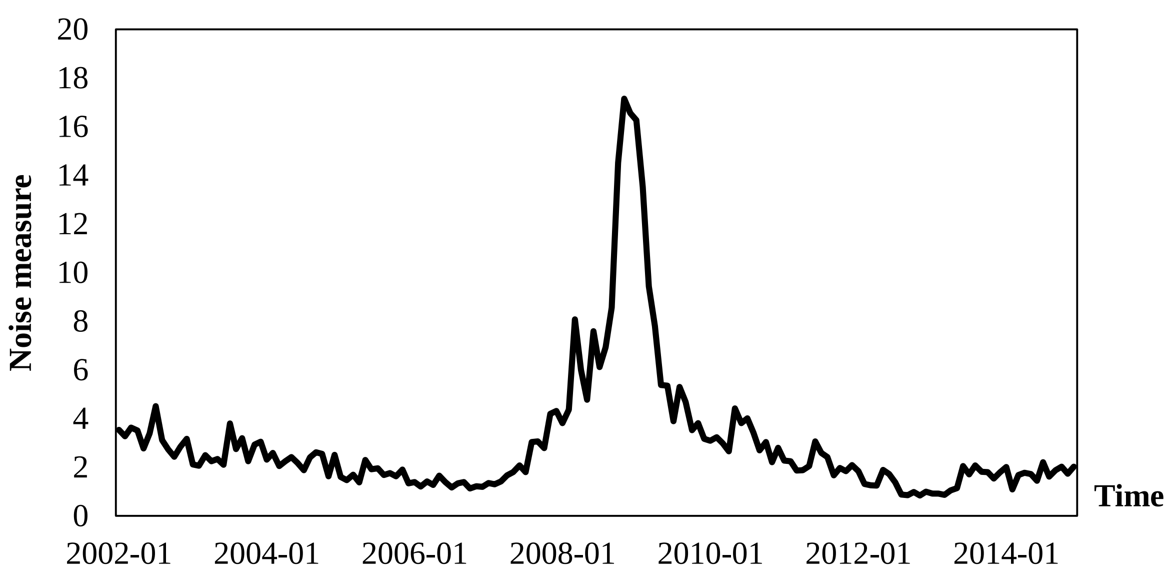

Figure 13. Noise measure data.

The figure shows how the noise measure by Hu et al. (2013) evolves in time. Data for the measure is available until the December 2014. The noise measure is observed at the end of each calendar month. The data is available on the website of Jun Pan. See Section 3.5.3 for more details on the measure's construction.

Following the theoretical connection between the noise measure and the arbitrageur risk-aversion, the Hypothesis 7 stated in Section 4 says that the mispricings on the yield curve are larger when there is a lot of noise in the markets as measured by the Hu et al. (2013) metric. As a measure of 'mispricing' I look at the differences between the market and the model implied rates, i.e. at by how many basis points (bps) the model implied rates deviate from those observed in the market. I consider both the core DLY strategy and its in-the-sample version in connecting the mispricings to the noise. The core DLY strategy explicitly trades out-of-the-sample, whereas the in-the-sample version is calibrated to and trades within the whole sample.

As is obvious from Figures 9 and 10 in Section 6.3, the market-to-model differences, or mispricings, are fairly unstable quantities themselves. To make them smoother, I take a 12-month moving average (MA) of both the noise measure and the average absolute mispricing to attain a more robust picture of their relation. Figures 14 and 15 below illustrate the considerable positive correlation that the smoothed mispricings have with the noise measure. To be specific, the correlations with respect to the noise measure are 60% and 48% for the out-of-the-sample and in-the-sample strategies' implied mispricings, respectively.

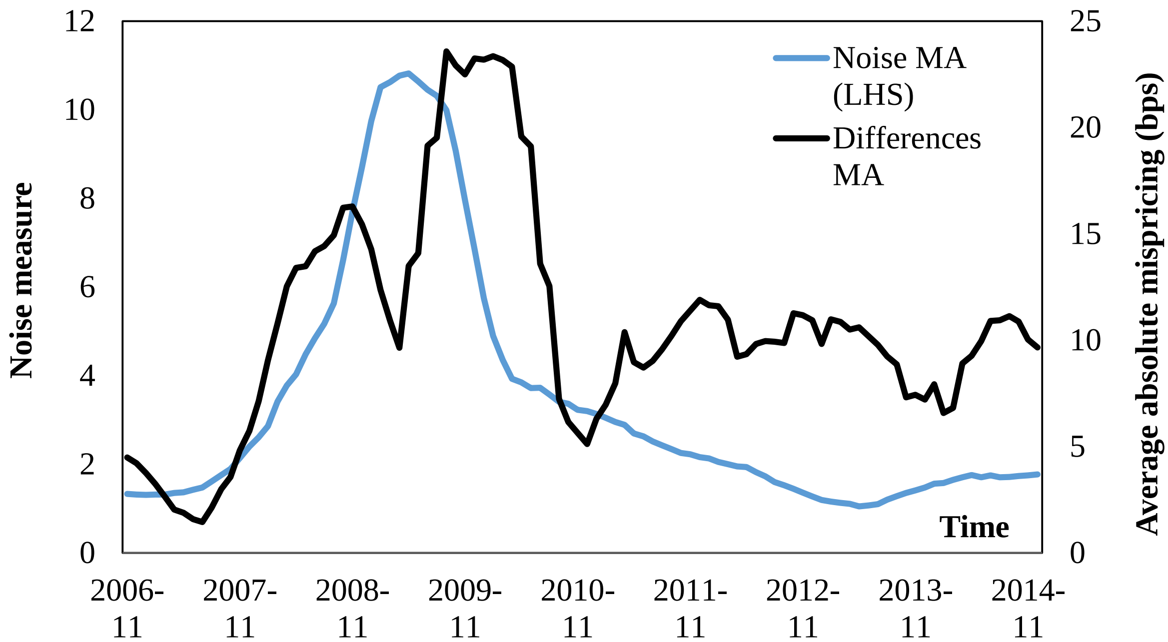

Figure 14. Noise measure and out-of-the-sample strategy mispricings.

The figure plots 12-month moving averages of the Hu et al. (2013) noise measure and the model implied mispricing in the case of trading done out-of-the-sample. The mispricing refers to the mean of the absolute market-to-model differences of the two to nine year rates ('illiquid rates'). Correlation between the two series is 60%. The time period for the observations is November 2006 to December 2014.

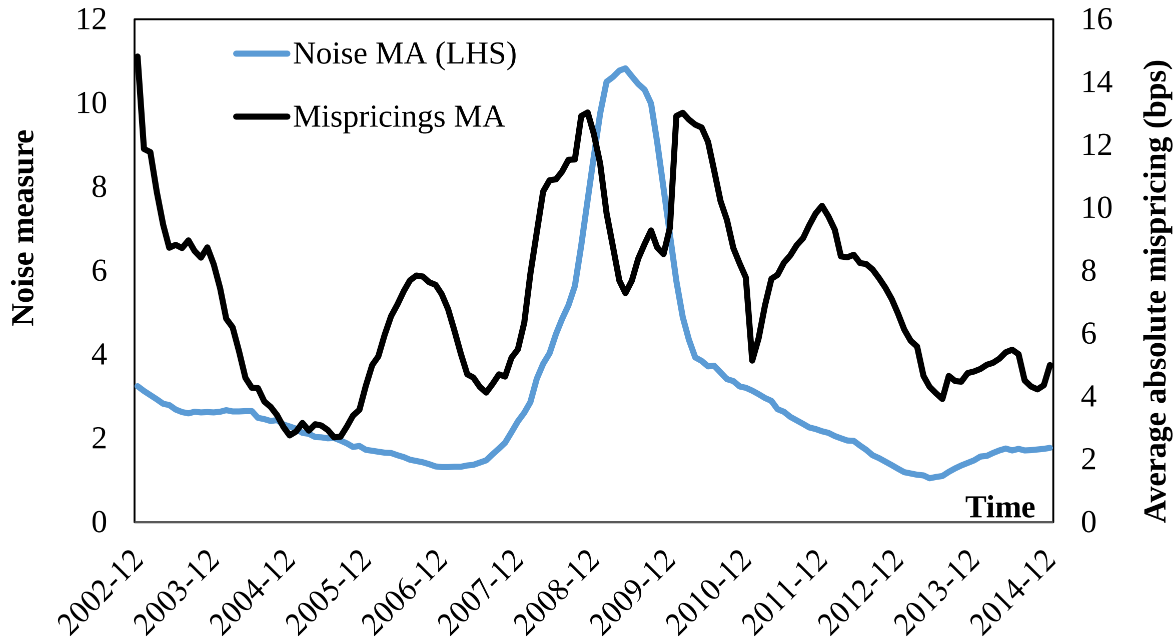

Figure 15. Noise measure and in-the-sample strategy mispricings.

The figure plots 12-month moving averages of the Hu et al. (2013) noise measure and the model implied mispricing in the case of trading done in-the-sample. The mispricing refers to the mean of the absolute market-to-model differences of the two to nine year rates ('illiquid rates'). The mispricing refers to the mean of the absolute market-to-model differences for the two to nine year rates. Correlation between the two series is 48%. The time period for the observations is December 2002 to December 2014.

Given the high correlations between the moving averages, regressions where the mispricings are explained by the noise measure naturally yield highly significant betas of 1.14 and 0.48 with the test statistics of 7.4 and 5.1 for the out-of-the-sample and in-the-sample strategies, respectively. The noise measure explains 36% of the variation in the out-of-the-sample strategy's, and 18% of the variation in the in-the-sample strategy's implied mispricings. This lends support for the Hypothesis 7, which states that high noise coincides with high model implied mispricings. After employing the moving averages to smoothen the measures of the implied mispricings, one indeed finds a high positive correlation to the noise measure. The caveat is that the relation would significantly weaken, should one employ the non-smoothed measure of the mispricings, which is fairly unstable in itself.

The other noise related hypothesis, the Hypothesis 8, states that a high noise environment would lead to attractive trading outcomes. I run a strategy that leverages the basic DLY strategy two times (2x) when the noise measure is above its 70th percentile, and leverages 0.5x (i.e., under leverages) when it is below the 30th percentile. I refer to this strategy as ‘leveraging high noise’. Moreover, I do the opposite, i.e. leverage more when the noise measure is low; this strategy is referred to as ’leveraging low noise’. Both of the leveraging strategies trade also the average noise conditions, yet in that case employ the unaltered leverage of 1x.

In contrast to the Hypothesis 8, I find evidence that favorable trading outcomes coincide with low noise, as shown in Table 12 and Figure 16, which illustrate the attractive risk-adjusted performance of (further) leveraging when there is low noise. Therefore, The Hypothesis 8 is discarded on decent evidence. This is defensible in the sense that convergences can actually be expected to happen in the low-noise environment. Given the inferior performance of leveraging in a noisy market, I suggest that while noisy market environments (high risk-aversion, low liquidity) may offer extraordinary buying opportunities in general, it does not facilitate the work of arbitrageurs making their living on convergence, or relative value, based bets. This view is further discussed in the convergence context by e.g. Xiong (2001).

Table 12. Leveraging high and low noise periods. The table presents the performance of three strategies. First is shown the basic version of the strategy as in Section 6, directly according to Duarte et al. (2007), then the same strategy which leverages 2x when the noise measure exceeds the 70th percentile, and leverages 0.5x when the measure is below the 30th percentile. Lastly is depicted the strategy which leverages 2x when the noise measure is below the 30th percentile, and leverages 0.5x when the measure is above the 70th percentile. Percentiles for the noise measure are computed in-the-sample. N denotes the number of monthly excess returns. Capital is the initial amount of capital required per €100 notional of the arbitrage strategy to give a ten-percent annualized standard deviation of the excess returns. Mean is the leveraged monthly excess return. Test statistics are computed with the Newey-West (1987) autocorrelation robust measure. Min/max are the minimum and maximum of the leveraged monthly excess returns. Gain/Loss is the Bernardo and Ledoit (2000) gain/loss ratio for the strategy. Sharpe ratios are annualized. The overall sample period for the strategies is January 2002 to December 2014. Trading is replicated out-of-the-sample.

CIR2F model
4-year sample period, 5 bps limit to initiate a trade

Strategy	N	Capital	Mean	t-Stat	Std. Dev.	Min.	Max.	Skew.	Kurt.	Ratio Neg.	Serial Corr.	Gain/ Loss	Sharpe Ratio
Basic DLY	109	6.159	0.787	2.79	2.887	-7.139	10.564	0.605	5.079	0.330	0.040	2.314	0.944
Leverage high noise	109	9.069	0.620	2.01	2.887	-9.695	13.825	0.764	8.196	0.330	0.251	2.100	0.744
Leverage low noise	109	6.769	0.909	3.60	2.887	-5.986	16.112	1.859	10.159	0.330	-0.168	2.999	1.091

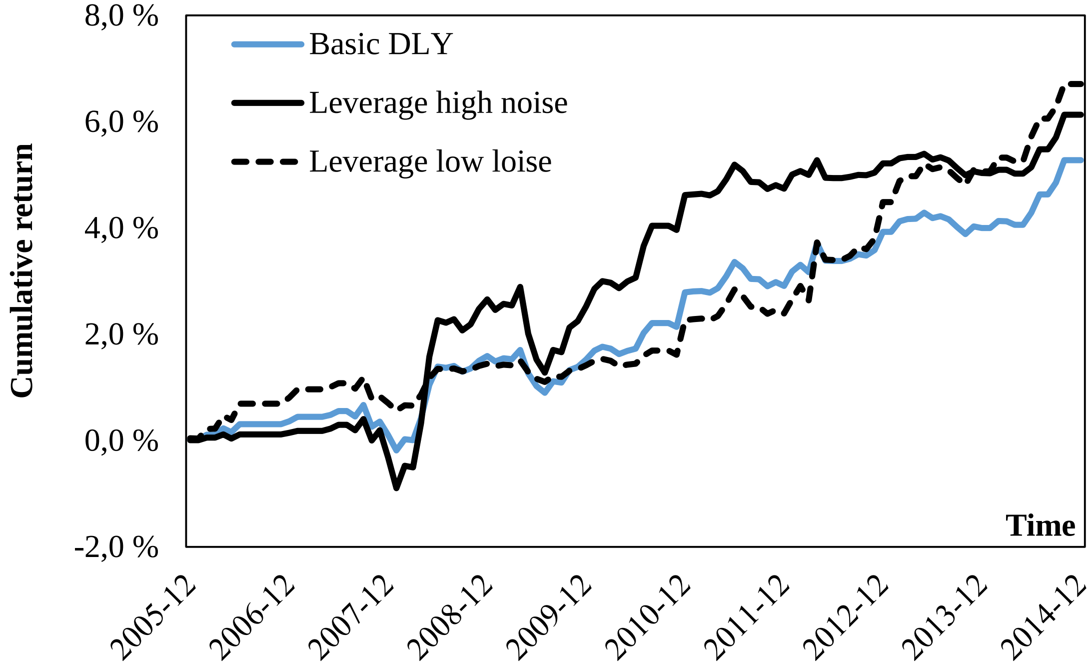

Figure 16. Cumulative returns from leveraging noise.

The figure plots the cumulative unleveraged returns of the basic version of the strategy ('DLY') as in Section 6, directly according to Duarte et al. (2007), then the same strategy but leveraging 2x when the noise measure is above its 70th percentile, and leveraging 0.5x when the measure is below its 30th percentile ('Leverage high noise'). Also plotted is the strategy that leverages 2x when the noise measure is below its 30th percentile, and leverages 0.5x when the measure is above its 70th percentile ('Leverage low noise'). Percentiles for the noise measure are computed in-the-sample.

10. Conclusion

Hedge funds have been shown to generate highly favorable returns in the space of fixed income arbitrage, as illustrated in e.g. Duarte et al. (2007). As the term 'arbitrage' implies, these strategies are in a way or another hedged, and thus carry a low risk in terms of volatility. Although some may look for the kind of textbook arbitrage, where one would generate riskless profits with no capital, 'arbitrage' in the in industry parlance usually refers to portfolios that are made neutral to changes in priced high-level risk factors. In essence, yield curve arbitrage is a relative value trading strategy where an arbitrageur trades overtly high or low interest rates against the more efficiently priced ones, assuming that the perceived mispricings shall be eliminated in time. Duarte et al.

(2007) attribute the yield curve and capital structure arbitrage strategies' performance mostly to the increased complexity and thus additional human capital required to employ the relevant models in locating the mispricings and implementing the trades.

The theoretical foundation of the yield curve formation and no-arbitrage, as discussed in this paper, is based chiefly on Vayanos and Vila (2009), complemented by Greenwood and Vayanos (2010 and 2014). The practical methodology and its execution is either directly or with modifications after Duarte et al. (2007). In general, the driving idea in my work is to combine the theoretical view of yield curve arbitrage with the empirical arbitrageur trading methodology to paint a picture of the attractiveness of the yield curve arbitrage in the EUR rates space, as done by hedge funds or proprietary trading desks.

I set out to further develop the methodology outlined in Duarte et al. (2007) with a different and more recent dataset. For one, Duarte et al. employ a two-factor Vasicek model, while I am using two-factor models suggested by Longstaff and Schwartz (LS two-factor model) and Cox, Ingersoll and Ross (CIR two-factor model). I conduct the analysis of the strategies with recent (2004-2015) data on the EUR swap rates, while the earlier paper looks at the USD space in 1988-2004. Moreover, I test different trading implementations as logically implied by the models. Finally, I study the strategies with respect to hedge fund indices and subindices, as well as with respect to the noise metric by Hu et al. (2013). Throughout the thesis, I make explicitly sure that no look-ahead bias emerges, unless when done on purpose for comparison. Duarte et al. imply that they calibrate their model to the whole sample (while conducting as a diagnostic also an explicit out-of-the-sample trading test), which includes also future data in choosing the parameters.

I hypothesize that the yield curve arbitrage strategy in the EUR space generates attractive risk-adjusted returns in terms of common ratios and regarding the return distributions. Strategies' exposure to priced risk factors (betas) is hypothesized to be low, meaning that returns would be driven by multifactor alpha. Additionally, I hypothesize that the arbitrage strategy explains to a significant degree the most relevant hedge fund subindex returns. Further, it is hypothesized that mispricing of rates in the market is greater when there is a lot of 'noise' as implied by the Hu et al. (2013) measure, and that returns from the strategies are greater during these noisy periods.

I find evidence strongly in support of the hypotheses regarding the attractiveness of the replicated returns. The yield curve arbitrage strategies have generated attractive risk-adjusted returns in the EUR swap space, having high Sharpe and Gain-Loss ratios. Moreover, the returns are statistically significant as measured by the autocorrelation robust Newey-West (1987) statistic. The return distributions are far from normal, as they are highly positively skewed with heavy tails. The positive skew is certainly favorable to the arbitrageurs, yet the attractiveness of the fat tails in the distributions can be contested.

The returns are basically pure multifactor alpha, as priced risk factors lack explanatory power regarding the strategies. As the volatility of the strategies is low, they can and need to be leveraged. Duarte et al. (2007) leverage their strategies so that the ex post volatility is 10% annually. With the same specifications, the yield curve arbitrage strategies are found to be capable of achieving leveraged returns up to 10% annually in the EUR swap rates market. This is actually superior to the earlier results by Duarte et al. both in the yield curve arbitrage space and when compared to other fixed income arbitrage styles (e.g. capital structure and MBS arbitrage).

When the replicated yield curve arbitrage returns are compared to the actual hedge fund indices, both high-level and style-specific, little connection between the returns is found. The yield curve arbitrage is essentially ‘hedge fund alpha’ in the sense that no combination of the hedge fund indices explains the replicated returns to any significant extent. Similar to Duarte et al. (2007), high-level hedge fund indices are not explained by the strategy. Against my hypothesis, I find that neither the most relevant subindices are not well explained by the arbitrage strategy. This suggests that either the strategy itself or its deliberate implementation diverge enough from the average strategies in the indices so as to offer little explanatory power in the cross-section. This can be seen as an attractive characteristic for the strategy, as it is differentiated from the mainstream relative value trades, thereby leading to less competition and less crowded positioning.

As hypothesized, high model implied mispricings coincide with high noise in the markets as measured by the Hu et al. (2013) metric. As the model implied mispricing is a fairly unstable quantity in itself, this result is statistically robust only when the time series is smoothed by taking a moving average. Moreover, high noise in the markets does not lead to stronger trading outcomes, which is in a sense intuitive given the nature of convergence trading. Evidence is found that further leveraging positions when there is low noise leads to better risk-adjusted returns.

This thesis also sheds further light into the subtleties of modeling, as well as describes more in-depth the sensitivities of the strategies to certain variations in implementation. In all, yield curve arbitrage produces highly favorable risk-adjusted returns with little sensitivity to the exact choices made through modeling, calibration and implementation, yet with a data mining angle one can certainly find parameters that maximize the out-of-the-sample performance of the strategies. The robustness of the results regarding the choices in the modeling highlights the attractiveness of the core methodology, as one can say that with all likelihood the findings cannot be solely due to data mining.

As Duarte, Longstaff and Yu (2007) concluded, there seems to be little evidence to label fixed income arbitrage as ‘picking nickels in front of a steamroller’, which says that one would have decent returns most of the time and large losses every now and then. More to the contrary, the return distributions have a mean, skew and kurtosis of the caliber that the yield curve arbitrage can be considered to be a highly attractive trading strategy in terms of gains to losses and excess returns to volatility, as well as in yielding uncorrelated returns.

The quantitative yield curve arbitrage seems to produce both economically and statistically significant alpha that seriously outperforms the alphas of the related hedge fund indices. This is consistent with the evidence that the strategies modeled here have little correlation with the hedge fund indices. If some of the hedge fund indices comprise similar yield curve arbitrage strategies, their weight is small enough to get neutralized in aggregate. In a sense, this makes the returns from the replicated strategies even more attractive, as the evidence points to the direction that they are sourced from a relatively rare origin that the bulk of sophisticated investors do not in practice seem to harness.

Divergence from the positioning of other arbitrageurs can be extremely beneficial for a hedge fund, as it will hold less crowded positions. This is because crowded trading positions will suffer more when competitors are forced to liquidate, as was shown by the 2007 quant meltdown, for instance. As for the rare event risks, evidence does not suggest that the tail event of 2008 Lehman Brothers bankruptcy, for example, would have caused the strategy to blow up, or even to cause a major drawdown to the returns. In conclusion, a strategy that appears little competed and does not amount to selling disaster insurance would be a valuable add to basically any portfolio as a source of alpha.

References

Agarwal, V., Fung, W. H., Loon, Y. C., Naik, N. Y., 2011. Risk and Return in Convertible Arbitrage: Evidence from the Convertible Bond Market. Journal of Empirical Finance 18, 175–194.

Amihud, Y. 2002. Illiquidity and Stock Returns: Cross Section and Time Series Effects. Journal of Financial Markets 5, 31–56.

Amihud, Y., Mendelson, H., 1991. Liquidity, Maturity, and the Yields on U.S. Treasury Securities. The Journal of Finance 46, 1411–1425.

Avellaneda, M., Lee, J.-H., 2010. Statistical Arbitrage in the U.S. Equities Market. Quantitative Finance 8, 761-782.

Brunnermeier, M., Nagel, S., Pedersen, L., 2009. Carry Trades and Currency Crashes. NBER Macroeconomics Annual 2008, 313–347.

Cox, J., Ingersoll, J., and Ross, S., 1985. A Theory of the Term Structure of Interest Rates. Econometrica 53, 385-408.

Culbertson, J. M., 1957. The Term Structure of Interest Rates. Quarterly Journal of Economics, 71, 485–517.

Duarte, J., Longstaff, F.A., Yu, F., 2007. Risk and Return in Fixed-Income arbitrage: Nickels in Front of a Steamroller? Review of Financial Studies 20, 769–811.

Fama, E. F., Bliss, R. R., 1987. The Information in Long-Maturity Forward Rates. The American Economic Review 77, 680-692.

Fama, E. F., French, K. R., 1992. The Cross-Section of Expected Stock Returns. The Journal of Finance 47, 427–465.

Fung, W., Hsieh, D.A., 2002. The Risk in Fixed-Income Hedge Fund Styles. The Journal of Fixed Income 12, 6-27.

Fung, W., Hsieh, D.A., 1997. Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds. The Review of Financial Studies 10, 275-302.

Gabaix, X., Krishnamurthy, A., Vigneron, O., 2007. Limits of Arbitrage: Theory and Evidence from the Mortgage-backed Securities Market. Journal of Finance. 62, 557-95.

Garleanu, N., Pedersen, L., Poteshman A., 2009. Demand-Based Option Pricing. Review of Financial Studies 22, 4259-4299.

Gatev, E., Goetzmann, W., Rouwenhorst, K., 2006. Pairs Trading: Performance of a Relative-Value Arbitrage Rule. Review of Financial Studies 19, 797-827.

Greenwood, R., 2005. Short- and Long-term Demand Curves for Stocks: Theory and Evidence on the Dynamics of Arbitrage. Journal of Financial Economics 75, 607-649.

Greenwood, R., Vayanos D., 2010. Price Pressure in the Government Bond Market. American Economic Review 100, 585-590.

Greenwood, R., Vayanos, D., 2014. Bond Supply and Excess Bond Returns. Review of Financial Studies 27, 663-713.

Gromb, D., Vayanos, D., 2010. Limits of Arbitrage: The State of the Theory. Annual Review of Financial Economics 2, 251-275.

Hau, H., 2009. Global Versus Local Asset Pricing: Evidence from Arbitrage of the MSCI Index Change. Working Paper, INSEAD.

Hau, H., 2014. The Exchange Rate Effect of Multi-Currency Risk Arbitrage. Journal of International Money and Finance 47, 304-331.

Hu, G.X., Pan, J., and Wang, J., 2013. Noise as Information for Illiquidity. Journal of Finance, 68, 2341–2382.

Jiang, H., Kelly, B., 2012. Tail Risk and Hedge Fund Returns. Working Paper, NBER.

Jylhä, P., Suominen, M., 2009. Speculative Capital and Currency Carry Trades. Journal of Financial Economics 99, 60-75.

Khandani, A. E., Lo, A. W., 2007. What Happened To The Quants In August 2007? Journal of Investment Management 5, 5-54.

Litterman, R. B., Scheinkman, J., 1991. Common Factors Affecting Bond Returns. The Journal of Fixed Income 1, 54-61.

Liu, J., Longstaff, F. A., 2004. Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. The Review of Financial Studies 17.

Merton, R. C., 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. Journal of Finance 29, 449-470.

Mitchell, M., Pulvino, T., 2001. Characteristics of Risk in Risk Arbitrage. Journal of Finance 56, 2135–2175.

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

Modigliani, F., Sutch R., 1966. Innovations in Interest-Rate Policy. American Economic Review, 56, 178–97.

Naranjo, L., 2009. Implied Interest Rates in a Market with Frictions. Working Paper, ESSEC.

Newey, W., West, K., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

Plantin, G., Shin, H., 2009. Carry Trades and Speculative Dynamics. Working Paper, Toulouse School of Economics.

Shleifer, A., Vishny, R. W., 1997. The Limits of Arbitrage. The Journal of Finance 52, 35-55.

Vasicek, O., 1977. An Equilibrium Characterization of the Term Structure. Journal of Financial Economics 5, 177-188.

Vayanos, D., Vila, J.-L., 2009. A Preferred-Habitat Model of the Term Structure of Interest Rates. Working Paper, NBER.

Xiong, W., 2001. Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets. Journal of Financial Economics 62, 247–292.

Yu, F., 2006. How Profitable Is Capital Structure Arbitrage? Financial Analysts Journal 62, 47-62.

Appendix A (Model implied mispricings per rate maturities)

This Appendix shows the basis point differences between the market and model implied swap rates for individual rate maturities from two to nine years in time as produced by the core DLY strategy trading out-of-the-sample. The interpretation for all the following Figures is that they plot the ‘mispricing’, or basis point spread between the market-observed and model implied constant maturity swap rate. The observation period for the differences is from December 2005 to January 2015. Details on how the model implied rate is computed can be found in Section 5.3.

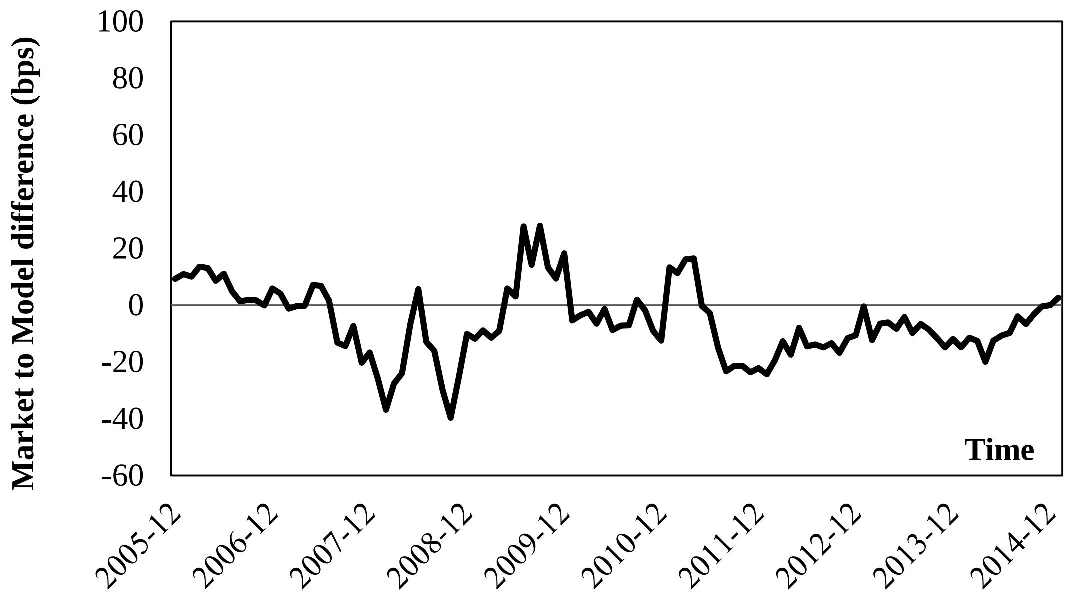

Figure 16. 2-year rate market-to-model differences in the out-of-the-sample DLY strategy.

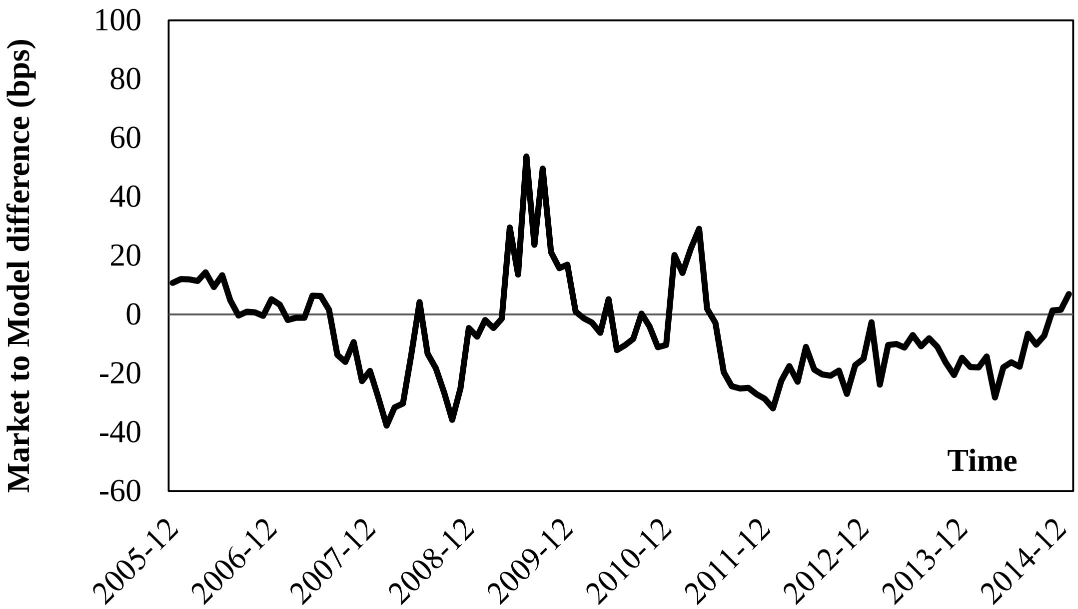

Figure 17. 3-year rate market-to-model differences in the out-of-the-sample DLY strategy.

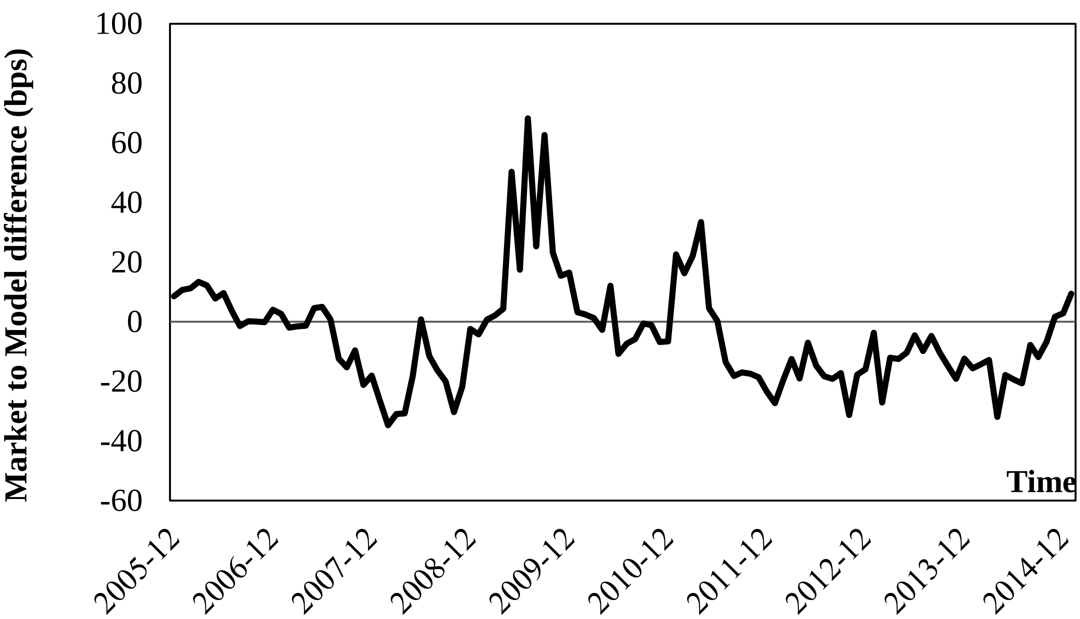

Figure 18. 4-year rate market-to-model differences in the out-of-the-sample DLY strategy.

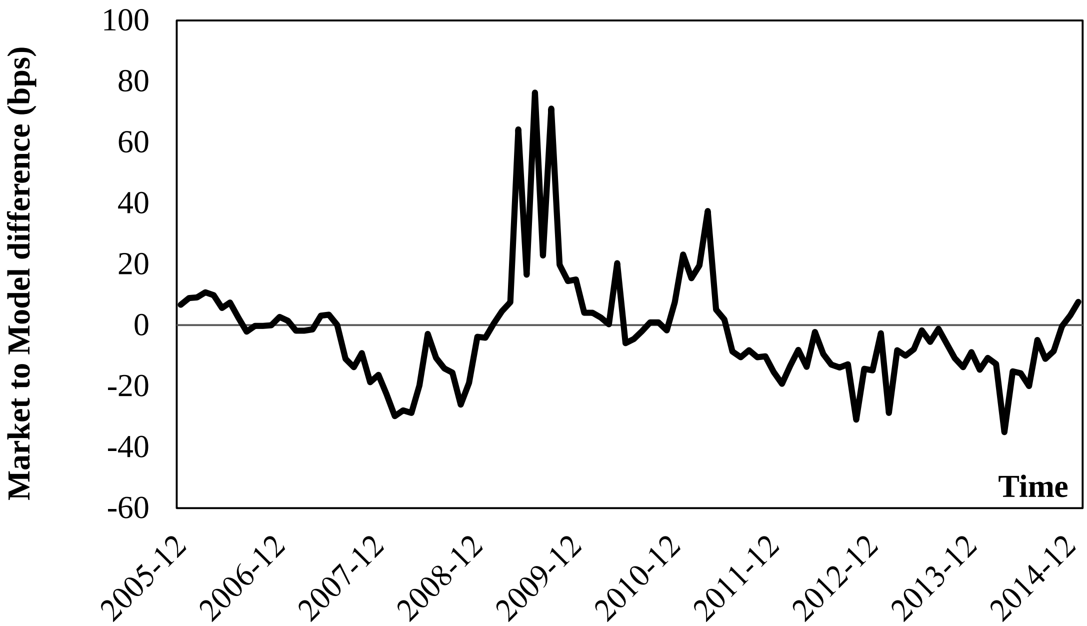

Figure 19. 5-year rate market-to-model differences in the out-of-the-sample DLY strategy.

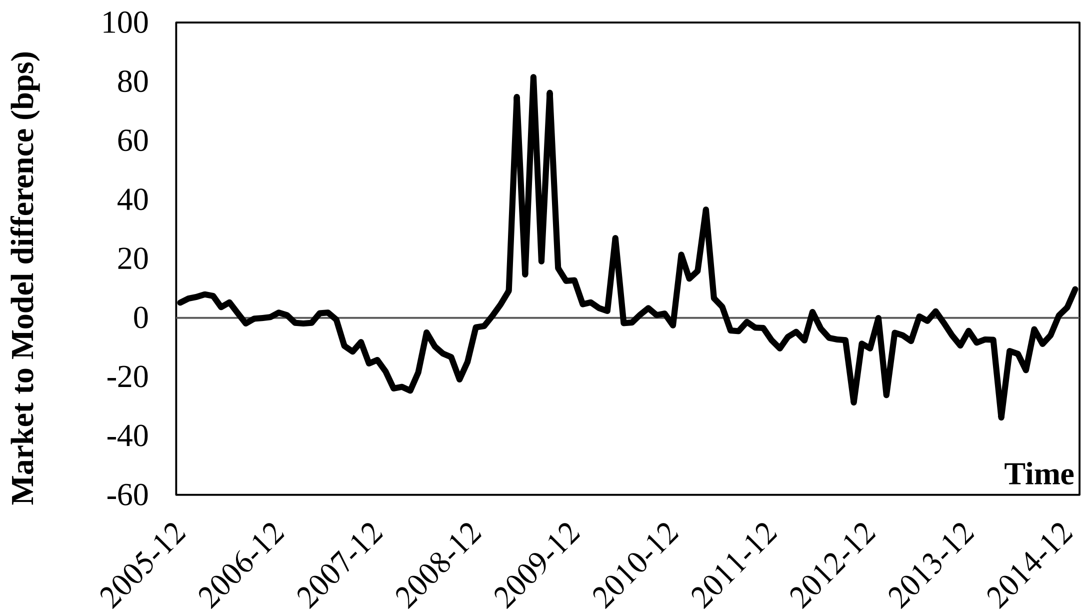

Figure 20. 6-year rate market-to-model differences in the out-of-the-sample DLY strategy.

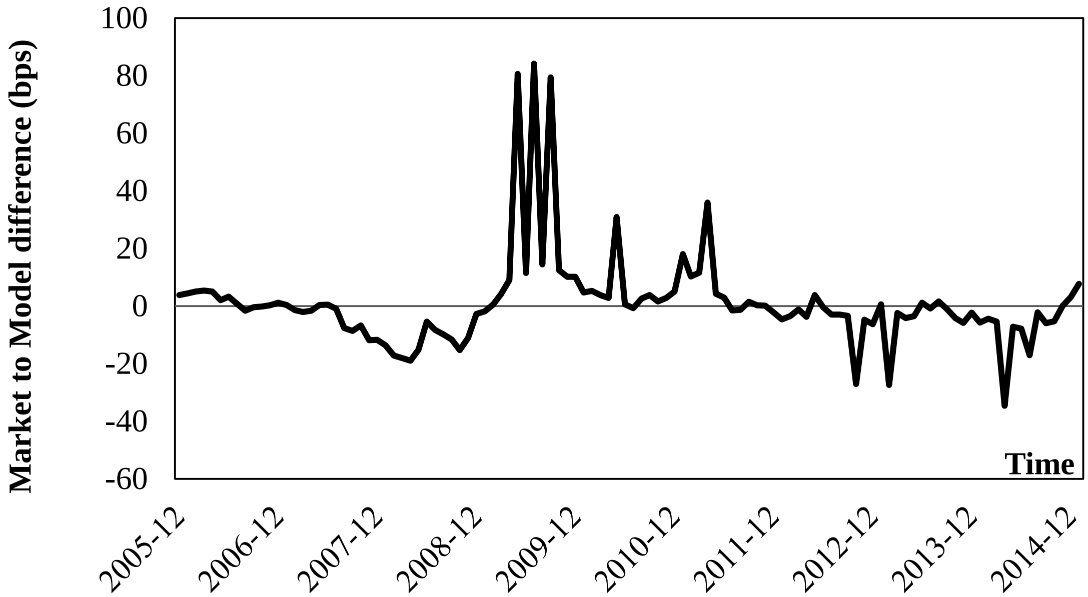

Figure 21. 7-year rate market-to-model differences in the out-of-the-sample DLY strategy.

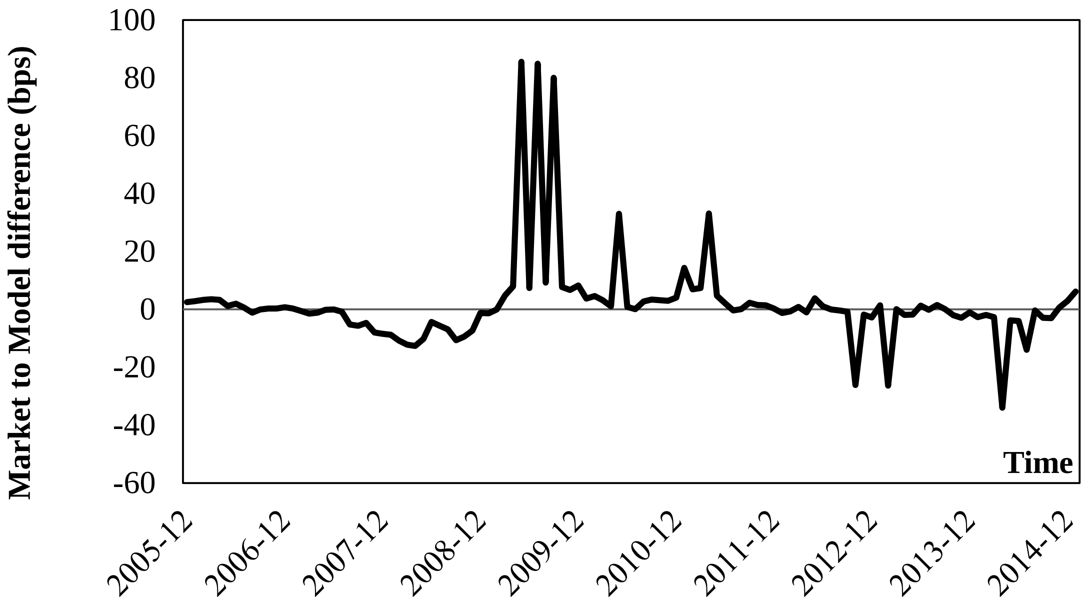

Figure 22. 8-year rate market-to-model differences in the out-of-the-sample DLY strategy.

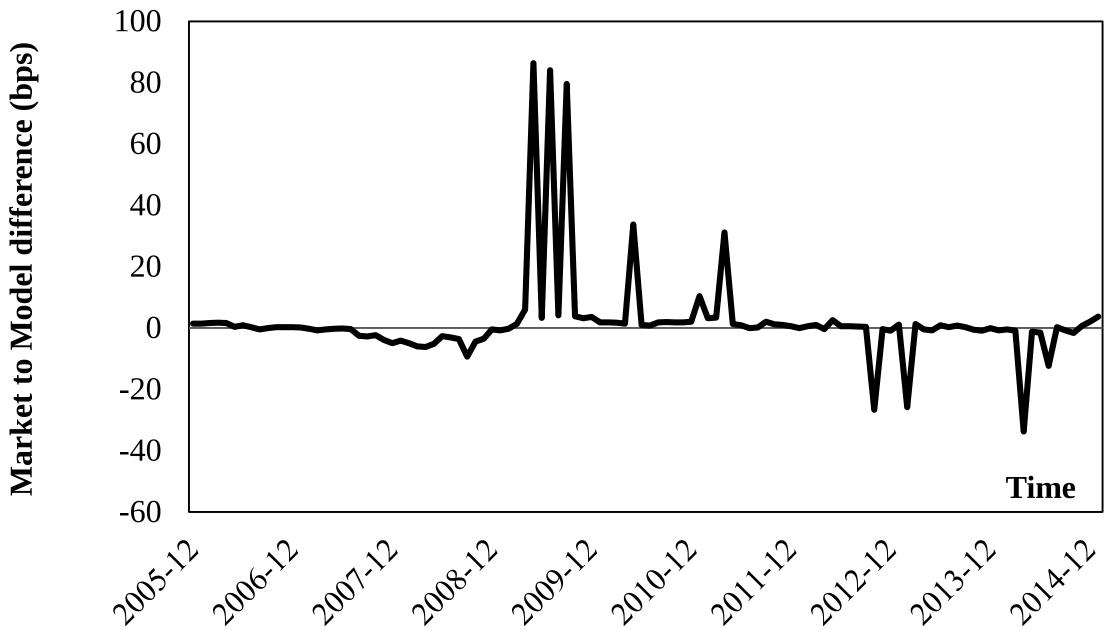

Figure 23. 9-year rate market-to-model differences in the out-of-the-sample DLY strategy.