

A Machine Learning-based Recommendation System for Swaptions Strategies

Adriano Soares Koshiyama, Nick Firoozye and Philip Treleaven

Department of Computer Science

University College London

London, United Kingdom

adriano.koshiyama.15@ucl.ac.uk

October 15, 2021

Abstract

Derivative traders are usually required to scan through hundreds, even thousands of possible trades on a daily basis. Up to now, not a single solution is available to aid in their job. Hence, this work aims to develop a trading recommendation system, and apply this system to the so-called Mid-Curve Calendar Spread (MCCS), an exotic swaption-based derivatives package. In summary, our trading recommendation system follows this pipeline: (i) on a certain trade date, we compute metrics and sensitivities related to an MCCS; (ii) these metrics are feed in a model that can predict its expected return for a given holding period; and after repeating (i) and (ii) for all trades we (iii) rank the trades using some dominance criteria. To suggest that such approach is feasible, we used a list of 35 different types of MCCS; a total of 11 predictive models; and 4 benchmark models. Our results suggest that in general linear regression with lasso regularisation compared favourably to other approaches from a predictive and interpretability perspective.

1 Introduction

Derivative traders are usually required to scan through hundreds, even thousands of possible trades on a daily basis. A concrete case is the so-called Mid-Curve Calendar Spread (MCCS), a derivatives package that involves selling an option on a forward-starting swap and buying an option on a spot-starting swap with longer expiration [8, 26]. In such a package, traders look for the historical carry and the breakeven width levels, metrics that can be easily inferred from the terminal or aged payoff profile of the MCCS, shown in several heatmaps made by the research team. After that, they rank the most prominent ones to offer a client or to proceed in some proprietary trading. In general, the straightforwardness and swiftness that the decisions are made is the main upside of this framework.

However, one might notice that the main downsides of such approach are: (i) substantial information on the underlying like sensitivities, implied volatility, etc. are usually not taken into account; (ii) using the previous example, high

historical values for carry and breakeven widths are more necessary rather than sufficient conditions for a profitable MCCS trade, being such argument extensible to other trades as well; (iii) a trader can quickly judge if an individual trade is worthwhile to invest, but may take some time to find it; and (iv) after a given period, traders tends to only look at a small subset of possible trades (small area on the heatmap), rather than the all available selection. Hence, a systematic approach where more information at hand is crossed and aggregated to find good trading picks and undoubtedly increase the trader’s productivity.

Therefore, the objective of this work is to develop a trading recommendation system that can aid derivatives traders in their day-to-day routine. Being more specific, our solution is based on the following pipeline: (i) on a certain trade date, we compute metrics and sensitivities related to MCCS; (ii) these metrics are feed in a model that can predict its expected return for a given holding period; and after repeating (i) and (ii) for all trades we (iii) rank the trades using some dominance criteria. Our final solution is a model-based heatmap with the attractiveness scores for each trade, which can be offered to the traders and salespeople on a daily basis.

In this sense, we organised this work as follows: next section presents a literature review on existing approaches to return/price prediction/estimation in different areas and instruments, as well as a brief description on MCCS trades. The third section displays the dataset that comports the MCCS trades, showing how the information is computed and gathered, which variables are the input and outputs, and the main assumptions that are embedded in it. Then, we move to modelling strategy, highlighting the main models that are going to be used as candidates for the recommendation system, how they are tested and have their performance assessed. Finally, we exhibit the results and discussions, closing this work with some concluding remarks and future directions for research.

2 Background

2.1 Related Works

Literature provides a growing body of evidence that price changes can be predicted, that is, in particular circumstances and periods securities violate the Efficient Market Hypothesis [5, 24]. In this sense, researchers have employed different modelling approaches and information sets to predict price changes across a range of assets. When we scan the literature for cash instruments (equities, bonds, foreign exchange, etc.) focused only in using past returns as the main source for prediction, we can find works that tap into Bayesian forecasting [33], Nonparametric Predictive Inference [2], Forecasting Combination [12], Generalized Exponential Weighted Moving Average [25], Support Vector Machines (SVM) [22], Shallow and Deep Neural Networks architectures [7, 9, 18, 32], Random Forest and Gradient Boosting Trees [23], and so forth. The list of proposed methodologies keeps growing, in which equities or indices appears as the dominant asset class to apply these algorithms. Collectively, they provide evidence that some forecastability over returns can be achieved by putting in place complex models with a suitable training scheme.

Contrasting with the emphasis that researchers in cash instruments put on return predictability, when we devote our attention to research in derivatives

instruments (options, swaps, swaptions, etc.) it is clear that most of the effort is concentrated on pricing these contracts. In parallel to the traditional framework, alternative ways of pricing and trading started to emanate relying on fewer assumptions and more data-driven. We can pinpoint approaches that use Neural Networks for option pricing and hedging with daily S&P 500 index daily call options [17] as well as for real-time pricing and hedging options on currency futures of EUR/USD at tick level [31]. It is worth to mention other approaches in the derivatives realm, such that the prediction of pricing and hedging errors for equity-linked warrants with Gaussian Process models [19], building machine learning models for predicting option prices over KOSPI 200 Index options [27] and a general study on forecasting option price distributions using Bayesian kernel methods [28].

When we devote our attention to the asset type that this work is dedicated, interest rate swaptions, a similar pattern persists: most of the research is related to pricing and not to return prediction. Regarding pricing, the same tradition of relying on stochastic calculus techniques is followed [4, 29]. Regarding potential alternatives using more data-driven approaches as we saw with currency, indices and equities options, we can only mention the work of Souza et al. [30] which calibrates the Vasicek interest rate model under the risk neutral measure by learning the model parameters using Gaussian processes for regression. Considering trading strategies and return prediction, we can find even less academic research, being perhaps most of the research residing inside the counterparts that exchange such products (banks, hedge funds, etc.). This shortage of published research might be linked with the absence of ready to use and publicly available datasets, similar to the ones found in cash products since these instruments are traded off-exchange.

Based on this review of existing approaches to return/price prediction/estimation in different areas and instruments, to the best of our knowledge, our work is the first attempt to build a trading recommendation system in the context of derivatives. Our approach is not the only novel from a modelling perspective, but instead of trading the vanilla product (receiver/payer interest rate swap-tion), we prefer to focus on options strategies (calendar spreads, straddles, etc.) which in many cases is the package that is in practice traded. By thinking in terms of the package, in this case, a Mid-Curve Calendar Spread, rather than the individual constituents we unlock some features that can only be computed in this situation, like the carry at expiry, breakeven width and so on.

Therefore, we can train our models not only using past returns but also using sensitivities as well as information derived from the package payoff function. By portraying in this manner our investment strategy, we have a large information set that can substantially add information to aid forecasting returns. But as a counter-effect, this poses a new challenge on separating relevant features in a dynamic context. In this respect, the combination of temporal cross-validation, a diverse set of models and regularisation/feature selection can provide a robust framework for trading strategies backtesting and assessment. But before presenting such framework, next section gives a brief view on MCCSs trades.

2.2 Mid-Curve Calendar Spreads

Mid-Curve Calendar Spread (MCCS) is a package involving short selling an option on a forward-starting swap and going long a longer-expiry swaption on

the same underlying swap [8]. There is a counterpart with many similarities for equities – check in [26] for more information. Investors typically use MCCS to take a view on forwarding volatility. This comes from the fact that, conceptually, spot volatility can be decomposed into forward volatility and mid-curve volatility. Taking 10y10y¹ for example, Figure 1 illustrate the time periods covered by different interest rate volatilities and their instruments. The red lines indicate the time over which interest rate volatility exposure is taken, and the grey line indicates the underlying forward swap rate.

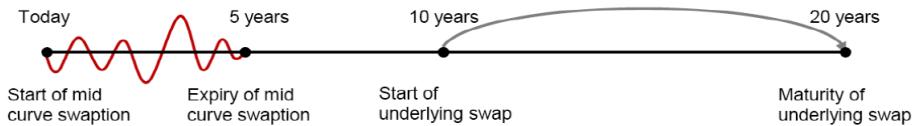


Figure 1: Mid-curve Swaption: 5y mid-curve on 5y10y swap rate – the volatility of a forward-starting swaption, called mid-curve, whose strike is set at inception and but the underlying swap starts several years following the option expiry date.

Figure 2 presents the payoff profile for an EUR 1m1y2y².

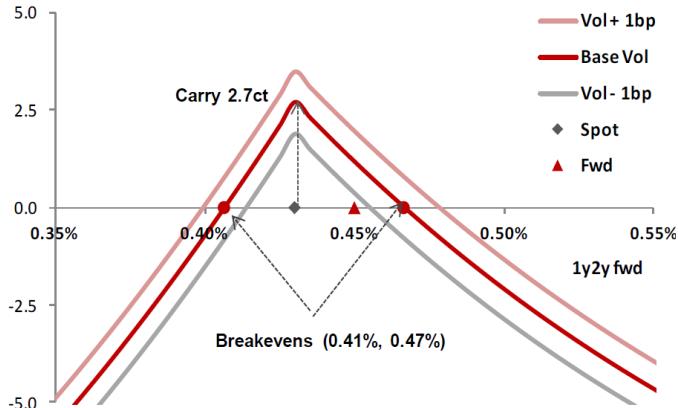


Figure 2: Payoff profile for an EUR 1m1y2y.

We plot the payoff profiles for current volatility and up and down volatility scenarios, noting that the long vega position means that the payoff profile shifts up in a rising volatility environment and correspondingly shifts down in a falling vol environment. We calculate the (volatility adjusted) breakevens as being 0.41% – 0.47%, giving little protection against selloffs. We note that forwards in a ± 1 volatility band leave them at 0.40% – 0.48%, a range just marginally larger than our breakeven range (i.e., the trade should pay off just slightly less

¹This notation is extensively used during this work. In this case, the first 10y means a spot swaption with 10 year of expiration, while the second 10y refers to the swap tenure.

²Short selling a 1m1y2y mid-curve swaption and going long a longer-expiry 13m2y spot swaption.

than 66% of the time). MCCS can result in what we think of as turbocharged carry, primarily because of the risks that they have (which fortunately can be balanced with the returns in a way which results in relatively attractive trades). Based on these characteristics of MCCS, next section presents how we elaborated the methodology to build this trading recommendation system.

3 Methodology

In summary, our solution develops the following roadmap (also schematically described in Figure 3):

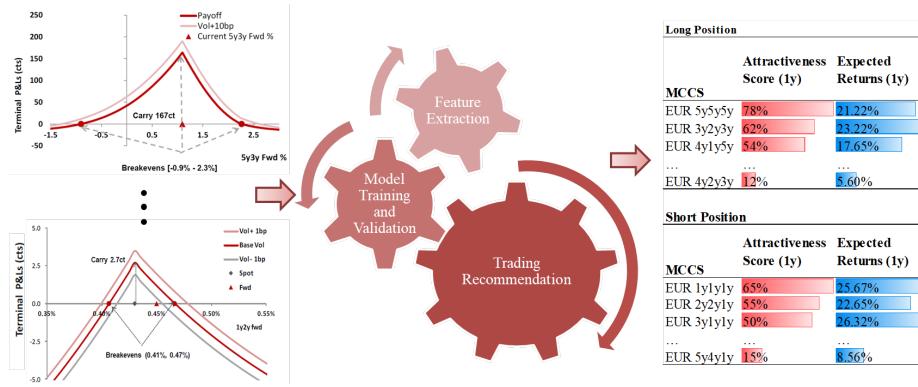


Figure 3: Flowchart describing the input-output schemes from the proposed trading recommendation system for MCCS trades.

1. **Data:** On a certain trade date, we **calculate metrics and sensitivities** related to an MCCS package;
2. **Modelling:** These metrics are **feed in a predictive model** that outputs its expected return for a given holding period (e.g., one year);
3. **Recommendation:** After repeating (i) and (ii) for all MCCS we (iii) rank them based on the expected returns using some criteria.

Following this outlined structure, the next three subsections describe in more details when and which MCCS trades were recorded (Dataset), which predictive models were trained and how they were assessed (Modelling) and how the long/short trading signal is computed for each MCCS (Recommendation). Finally, last subsection presents which metrics were used to evaluate the recommendation system performance when a certain predictive model candidate is underpinning it.

3.1 Dataset

During our experiments, we opted to use the trades displayed in Table 1. Although many other configurations are available in practice, these are the ones with longest historical data available, which is important when it is necessary

Table 1: Configuration of the MCCS trades used.

Currency	Expiry	Forward	Swap	Currency	Expiry	Forward	Swap
EUR	1y	1y	1y	EUR	3y	3y	2y
EUR	1y	1y	4y	EUR	3y	4y	1y
EUR	1y	2y	3y	EUR	3y	5y	5y
EUR	1y	2y	8y	EUR	4y	1y	1y
EUR	1y	3y	2y	EUR	4y	1y	4y
EUR	1y	4y	1y	EUR	4y	2y	3y
EUR	1y	5y	5y	EUR	4y	2y	8y
EUR	2y	1y	1y	EUR	4y	3y	2y
EUR	2y	1y	4y	EUR	4y	4y	1y
EUR	2y	2y	3y	EUR	4y	5y	5y
EUR	2y	2y	8y	EUR	5y	1y	1y
EUR	2y	3y	2y	EUR	5y	1y	4y
EUR	2y	4y	1y	EUR	5y	2y	3y
EUR	2y	5y	5y	EUR	5y	2y	8y
EUR	3y	1y	1y	EUR	5y	3y	2y
EUR	3y	1y	4y	EUR	5y	4y	1y
EUR	3y	2y	3y	EUR	5y	5y	5y
EUR	3y	2y	8y				

to fit a predictive model. As it can be seen, all trades are in Euro, ranging from different expiries (1y-5y), forwards (1y-5y) and swap tenures (1y-5y and 8y).

For each configuration, at time t we agree with a counterpart to trade this package using the At the Money Forward (ATMF) rate as the strike, paying or receiving the present value PV_t . The PV_t is computed via SABR model [29], using information and parameters (e.g., spot, forward rates and rate-rate correlation) calibrated using market data on a daily basis. From the same model that computed the PV_t , we can also obtain other metrics and sensitivities as those displayed in Table 2.

 Table 2: Metrics and sensitivities computed for each available package at time t .

Features	
<i>PV</i>	Strike
Carry at Expiry (Carry)	Breakeven Width (BE Width)
Aged 1y Carry	Theta
ATMF Implied Volatility (Implied Vol)	Gamma
Vega	Curve Carry (Aged 1y)
Time Carry (Aged 1y)	Volatility Carry (Aged 1y) (Vol Carry)

Carry and BE Width are those obtained looking at the payoff profile at expiry. The Aged 1y Carry is produced by ageing the trade by one year (moving closer to the expiration) and estimate the payoff profile computing the carry. Theta, Vega and Gamma are the sensitivities of the instruments by a change in time, volatility and a wider range of underlying rate movements, respectively. These and the ATMF Implied Vol are backed by the SABR model too. Curve, Time and Volatility Carry are the amount of Aged 1y Carry that can be attributed to the changes in certain sensitivities from spot to forward, such as the Delta (Curve), Theta (Time) and Vega (Volatility). These can also be used as tools to understand which factors most influence the instrument value over time.

After computing all these metrics at time t , we hold the trade until $t + h$ where h can be two weeks, one month, one year, and so on, as long as $t + h$ is before or at expiration. In time $t + h$ we compute the PV_{t+h} of the same trade

again, using the new economic scenario available (e.g. rates, change in model parameters). By agreeing on buying back or selling the current trade for PV_{t+h} we can compute the Holding k-period Return of the trade started at time t by:

$$R_t^{(h)} = \frac{PV_{t+h} - PV_t}{PV_t} \quad (1)$$

In summary, Table 3 presents an example of information in a wide format that is available when we combine the data from time t and $t + h$.

Table 3: Example of information available at time t and $t + h$ for the MCCS.

Instant (t)	PV_t	$R_t^{(h)}$	$R_{t-h-1}^{(h)}$	Strike	Features
1	300	0.3	-	3.4	...
2	320	0.2	-	3.2	...
...
$t + h - 1$	250	-0.1	-	1.8	...
$t + h$	260	-0.05	-	1.9	...
$t + h + 1$	270	-0.2	0.2	2.0	...
...
T	250	-	0.1	2.2	...

Note that in the most contemporaneous period (close to T) we do not have the PV_{T+h} and so, we cannot compute $R_T^{(h)}$. Conversely, if we want to use lagged returns $R_{t-h-1}^{(k)}$ as explanatory forces for $R_t^{(h)}$, then in the beginning this information is also not available. Therefore, our dataset is trimmed at the beginning and the end mainly by the value of h . If h is small, such as two weeks or one month, the trimming is imperceptible and, therefore, may not affect the model fitting and validation. However, if h is large such as two or three years, this might reduce the samples available substantially, decreasing the range of models and cross-validation schemes that might be employed for this task. Based on these procedures, metrics and observations, Table 4 express other details that we used during our experiments to generate the dataset.

Table 4: Details used to generate the MCCS trade dataset.

Detail	Value
Period	September 2006 to September 2016
Holding Period (h)	1 year
Trade Frequency	Weekly (usually on Wednesday)
Strike	At the Money Forward (ATMF)
Lagged data ($h-p$)	$p = 1, 2$ and 3 lagged returns
Assumption	Characteristic
Bid-Ask	Middle Rate
Transaction Costs	Entry and Unwind = $0.75 \times Vega_t$
Funding Rate	Libor 3 month rate

Therefore, we gathered data from trades entered on a weekly basis from September 2006 to September 2016. These trades are struck ATMF, using the PV_t computed from the Middle Rate (in practice, some bid-ask spread would be imbued proportional to the Vega). After holding for one year ($h = 1y$) the trade, we compute the arithmetical returns that are, therefore by definition, automatically annualised. These returns are gross, and so we need to take into account the transaction costs (hedging costs and fixed fees charged by the

derivatives desk) as well as some future funding rate. These values are also outlined in Table 4, where the transaction costs of 0.75 as a fraction of Vega were chosen not only to take into account the transaction cost, but also some potential bid-ask spread on the start/unwind of the trade. The 3-month London Interbank Overnight Rate (LIBOR) was chosen as the funding cost/benchmark rate to compute excess returns.

Using these assumptions, next subsection presents the modelling strategy that taps into this dataset to create the recommendation system for the MCCS trades.

3.2 Modelling

In relation to modelling, our general model is a system of uncoupled equations:

$$R_{t,1}^{(1y)} = f_1(features_{t,1}) + \varepsilon_{t,1} = \hat{R}_{t,1}^{(1y)} + \varepsilon_{t,1} \quad (2)$$

$$R_{t,2}^{(1y)} = f_2(features_{t,2}) + \varepsilon_{t,2} = \hat{R}_{t,2}^{(1y)} + \varepsilon_{t,2} \quad (3)$$

...

$$R_{t,n}^{(1y)} = f_n(features_{t,n}) + \varepsilon_{t,n} = \hat{R}_{t,n}^{(1y)} + \varepsilon_{t,n} \quad (4)$$

where for each MCCS trade ($i = 1, \dots, n$) there is an i -th predictive model f_i that is feed with a set of pre-calculated features (BE Width, Carry, etc.) and returns an estimate of the holding 1y-period return $\hat{R}_{t,i}^{(1y)}$. As the model is an approximation, some noise/error is expected, and in the modelling aspect, this is expressed as the $\varepsilon_{t,i}$ component. After defining which variable is intended to be predicted, the remaining points are: which models are available to embody f_i and how the fitting, validation and selection of these models are going to be made.

About the first point, in the first rows of Table 5 we display the models that we used during our experiments, with their mathematical descriptions and usage found in the following references [3, 11, 16, 20].

Table 5: Parameters used to model the MCCS trade dataset.

Abbreviation	Model	Fixed Hyperparameters		Cross-Validated Hyperparameters
		None	Backward Selection	
Classic Regression	Classical Linear Regression			None
BackSel Regression	Stepwise Regression			None
Ridge Regression	Ridge Regression			$\lambda = \{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}$
Lasso Regression	Lasso Regression			$\lambda = \{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}$
KRR-RBF	Kernel Ridge Regression		Radial-Basis Function kernel	$\lambda = \{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}$
kNN	k-Nearest Neighbours			and $\gamma = \{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$
CART	Classification and Regression Tree		Euclidean Distance	$k = \{3, 5, 7, 9\}$
Random Forest	Random Forest		MSE Function	Max depth = {2, 3, 5, 7}
Grad Boost Reg	Gradient Boosting Tree			None
MLP	Multi-Layer Perception	Number of trees = 333 and Max depth = 5		Learning Rate: {0.1, 0.3, 0.5}
SVR-RBF	Support Vector Regression	Number of trees = 333 and Max depth = 5	Single hidden layer with hyperbolic tangent as transfer function	$\lambda = \{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}$
			Radial-Basis Function kernel	and $nn = \{5, 7, 12\}$
				$C = \{10^0, 10^1, 10^2, 10^3\}$
				and $\gamma = \{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$
Abbreviation	Baseline Model	Parameter		
Mean Pred	Average Prediction			
Naive	Naive Model			
Z-Score: BE-Width	BE Width feature			
Z-Score: CarryAtExpiry	Carry at Expiry feature			
	Other Parameters	Values		
	Warm-up Period (L)	$L_{outer} = 2$ years	$L_{inner} = 1$ year	
	k-rolling-cv	$k_{outer} = 1$ week for outer	$k_{inner} \approx (T_{train} - L_{inner}) / 5$ for inner	
	Outlier Treatment	Winsorizing		
	Winsorizing Quantiles	0.01 and 0.95		
	Missing Data Treatment	Remove		

In Table 5 Model column presents a plethora of models that this work has fitted for this prediction purpose: we started from simple predictive models such as Classical Linear Regression, k-Nearest Neighbours and Classification and Regression Tree, towards those that can seamlessly exhibit nonlinear behaviours, like Random Forest, Kernel Ridge Regression, Multi-Layer Perceptron and Support Vector Regression. Some of these methods had their hyperparameters held constant across all experiments (Fixed Hyperparameters column), or because we wanted to apply a particular form of a method (RBF kernel, single hidden layer, etc.) or because during a warm-up phase we noticed that they did not affect substantially the results (hyperbolic tangent, increasing number of trees, etc.).

For certain models, the Cross-Validated Parameters column shows which hyperparameters were optimised before the prediction step. For instance, suppose the case of Ridge Regression and the need to define the regularisation value (λ) appropriately. Consider that we have a set of training pairs ($features_t, R_t^{(1y)}\right)_{t=1}^L$ of size L , and for this sample we subset it in k-rolling-cross-validation (k-rolling-cv) folders (better explained later in this subsection). Then, we train and test using this scheme the Ridge Regression model with one of the predefined λ , say $\lambda = 10^0$. We compute some performance function on the test set (Mean Squared Error – MSE) and repeat this process for all λ values available. We use in the final model the λ that on average had the lowest MSE.

We fitted usual benchmarks found in the literature for regression and forecasting modelling: the Average and Naive models [16, 21]. We also implemented the benchmarks that traders use to assess whether a particular MCCS is worth to be pitched or traded: BE Width and Carry at Expiry. We replicated the way traders look to these features, by computing z-scores³ based on average and standard deviation on rolling window of size equal to 1 year. The signal for going long/short is done by a thumb rule with a simple rationale: if a certain metric has a z-score above or equal to ± 3 , the trader goes fully long (+)/short(-) in the trade, since it is a very extreme event. Otherwise, it reduces the leverage on it, until it below one standard deviation of distance from the rolling average.

We removed any missing data, and clipped extremes values, mainly in returns above the 95% percentiles (in our case it can be due to some numerical problems, or some extreme scenarios related to 2008-2009 financial crisis period). Next subsection presents the final component of our roadmap: recommendation system.

3.3 Recommendation

The recommendation of a certain trade can be made solely on some normalised version of the expected return for holding 1y-period the i-th trade ($\hat{R}_{t,i}^{(1y)}$). Given that each model will be providing individual forecasts for each MCCS and after that their performance will be assessed locally and globally, a more suitable manner to proceed would be to assign a credit based on the tracking record of a model to predict a particular MCCS trade. Hence, we will be weighted up or down a signal not only based on the magnitude of a model prediction but also by its quality. Then, consider as $\hat{R}_{t,i}^{(1y)}$ the expected return for holding 1y-period

³a z-score is defined by: $Z - score = \frac{X - \mu}{\sigma}$ where X represent the actual value of a certain variable, μ and σ the average and standard deviation of X in a period.

the i -th trade. Now, define the new signal function $S_{t,i}$ by:

$$S_{t,i} = \frac{\hat{R}_{t,i}^{(1y)} \times Rho_{\hat{R}_{t,i}^{(1y)}, R_{t,i}^{(1y)}}}{\max(|\hat{R}_{t,i}^{(1y)} \times Rho_{\hat{R}_{t,i}^{(1y)}, R_{t,i}^{(1y)}}|, \dots, |\hat{R}_{t-h,i}^{(1y)} \times Rho_{\hat{R}_{t-h,i}^{(1y)}, R_{t-h,i}^{(1y)}}|, 0)}$$

where the strength of the i -th long/short signal is given by its expected return, scaled by the maximum weighted return that a long/short position on the same trade (that is why the returns are in absolute terms) was expected to yield in the previous h -period (in this case 1 year). Therefore, the trade with the maximum weighted return in absolute terms will have $|S_{t,i}| = 1$ as well as those close to zero will yield $S_{t,i} \approx 0$. The weight/credit of a certain prediction is based on the historical Pearson correlation coefficient, that is, adherence between the actual and predicted values.

3.4 Evaluation Metrics

Below we outline two types of metrics: one that focuses on the predictive performance that the model provided, and other three that are based on the profit/loss that its application harvested during the backtest. Set by $R_t^{(S)} = R_t^{(1y)} \times S_t(\hat{R}_t^{(1y)})$ the strategy return (combination of the realized/observed excess returns and the signal – function of a model prediction)⁴, we can compute the following metrics:

- Pearson Correlation Coefficient (Rho): it is a dimensionless measure of the linear dependence between the actual and predicted values:

$$Rho = \frac{Cov[R_t^{(1y)}, \hat{R}_t^{(1y)}]}{\sqrt{Var[R_t^{(1y)}]Var[\hat{R}_t^{(1y)}]}} \quad (5)$$

where Cov and Var are the covariance and variance operators. It ranges from $[-1, +1]$, with -1 representing a perfect inverse linear association, and $+1$ the opposite. In our case, we benefit more when Rho is close to $+1$. In the context of linear models, a higher predictive power is a necessary condition for profitable trades (see [1]), hence by minimising the predictive error we are somewhat trailing a path for profits maximisation, albeit such causation is not very clear since this is not a sufficient condition.

- Average Return (Avg Return): is the arithmetic average of the strategy returns:

$$\bar{R}^{(S)} = \frac{\sum_{t=1}^T R_t^{(S)}}{T} \quad (6)$$

- Standard Deviation: is the estimator of the dispersion around the strategy average returns (a risk measure in certain sense):

$$\sigma_{R^{(S)}} = \sqrt{\frac{\sum_{t=1}^T (R_t^{(S)} - \bar{R}^{(S)})^2}{T}} \quad (7)$$

⁴For the sake of brevity we dropped the subscript that refers to a particular trade (i).

- Information Ratio: is the average annualized return of a strategy earned in excess of a particular benchmark per unit of risk (measured in terms of standard deviation):

$$IR = \frac{\bar{R}^{(S)} - \bar{B}}{\sigma_{R^{(S)}}} \quad (8)$$

where \bar{B} is the average return of the benchmark (e.g., treasury bond, equity index). In our case, it was already set to the 3-month LIBOR rate (Table 4). It should be mentioned that Information Ratio makes each strategy performance comparable: since we are adjusting average returns by the risk assumed for each strategy, it removes the leverage component that is magnifying/shrinking the returns provided by a certain strategy.

4 Results and Discussions

Figure 4 displays a heatmap with the results of all models for each trade regarding Average Return (%). Similarly, different remarks can be made over the global picture: (i) the Naive and Mean Pred models underperformed, but the traders benchmarks did perform reasonably well, surpassing the predictive models in many occasions; (ii) from the linear regression family, Lasso Regression followed by Ridge Regression are the ones that performed better; (iii) most nonlinear models failed to provide a decent average return; and (iv) MLP fared well for the trades in the EUR Xy1y1y range, but did not repeat a more stable across other trades.

When we take into account the variability seen in the stream of returns generated by the recommendation system, we may encounter a different picture. Figure 5 shows a heatmap with the Information Ratio for all the available combinations of models and trades.

In general, the models kept their positions unaltered in comparison to the Average Return (%) – linear models still fared better than the nonlinear ones –, but now all are standing on a similar scale. Based on these Information Ratio results, Table 6 presents a statistical analysis using the average ranks⁵, Friedman test and Holm posthoc procedure [10].

When we look at the average rank, Lasso Regression was the top positioned (3.23) while Mean Pred remained most of the time as the worst choice (12.86). The trader’s benchmarks performed pretty well, being placed in the third and fourth places. When we compare whether such result fared by Lasso Regression was substantially different from Ridge Regression (4.86), we arrive with a Z-score equal to 1.63 and a p-value of 0.0517. If we set our initial significance level as 0.05 and correct using the Holm procedure (last column) we can assert that Lasso did not perform significantly different from Ridge Regression, but way better than the other models. Therefore, Lasso Regression is capturing some information beyond that is being spanned by the trader’s benchmarks, as well as beating almost all other predictive models for this particular task. Our

⁵When we rank the models for a single MCCS, it means that we sort all them in such way that the best performer is in the first place (receive value equal to 1), the second best is positioned in the second rank (receive value equal to 2), and so on. We can repeat this process for all trades and compute metrics, like the average rank (e.g., 1.35 means that a particular model was placed mostly near to the first place).

	20.01	-4.23	23.07	7.70	4.95	12.92	2.76	4.19	5.04	11.82	-0.67	15.86	3.32	-39.38	19.71
EUR1y1y1y	-4.03	16.18	10.94	17.08	3.86	-4.22	15.27	5.63	10.29	-3.95	-6.92	-4.79	-6.64	-16.48	12.22
EUR1y1y4y	-0.65	5.37	6.14	12.51	-5.98	-7.60	-8.21	-3.52	-0.56	3.30	-9.50	-0.63	0.65	-31.01	-5.85
EUR1y2y3y	-5.50	14.47	-2.91	10.16	-1.73	-1.61	-0.22	0.95	-3.21	-2.95	-8.34	7.33	6.41	-15.91	-3.54
EUR1y2y8y	0.06	-9.86	3.51	7.00	4.02	-0.24	-12.81	-6.07	-2.76	-0.40	0.00	-4.07	-1.61	-27.69	-7.42
EUR1y3y2y	-1.41	-1.31	0.07	3.96	3.33	0.90	-8.37	-6.55	-3.00	-0.88	-1.29	-5.13	-1.97	-27.50	-7.39
EUR1y5y5y	1.94	7.89	10.17	8.42	-2.09	1.41	-0.74	0.03	1.19	-0.27	-5.01	5.26	5.41	-2.18	-4.23
EUR2y1y1y	31.31	11.11	23.29	8.19	3.87	5.72	6.85	8.56	10.62	27.62	-10.13	30.37	-5.16	-22.38	-14.24
EUR2y1y4y	8.95	-0.08	5.68	7.60	-5.69	-7.41	2.73	-1.44	4.00	-0.50	-5.54	2.72	2.41	-4.59	-6.25
EUR2y2y3y	13.29	8.14	10.84	8.33	0.85	-10.45	6.02	8.55	4.34	5.74	-1.24	4.80	0.17	-22.06	-12.30
EUR2y2y8y	-1.53	13.55	10.84	9.38	-2.27	-2.49	10.08	3.79	8.81	-0.24	-1.67	7.64	9.00	-9.95	-7.17
EUR2y3y2y	1.53	8.77	6.48	11.25	1.09	2.47	-8.38	-9.98	-2.65	2.19	-2.37	-1.76	0.04	-20.22	-9.53
EUR2y4y1y	-2.22	-2.10	5.90	5.00	4.80	2.94	-5.62	-6.24	-5.25	0.74	3.97	-1.72	0.90	-20.08	-7.06
EUR2y5y5y	-1.07	9.63	1.33	7.23	3.23	0.52	-1.55	2.43	3.79	-0.38	-4.06	6.98	7.91	-6.36	-1.57
EUR3y1y1y	24.38	10.39	16.02	13.59	3.46	-2.42	-1.84	-1.91	4.76	22.10	1.53	46.09	-21.89	4.21	-8.56
EUR3y1y4y	3.86	-0.38	3.11	6.81	-3.67	-8.95	-3.57	-3.12	-1.87	0.04	-7.78	0.38	9.04	-24.93	-15.89
EUR3y2y3y	7.89	7.75	8.54	9.66	0.30	-0.76	6.74	3.82	6.50	9.04	-2.15	8.26	0.42	-23.64	-19.93
EUR3y2y8y	-6.36	10.76	7.38	11.85	0.24	-0.67	-1.52	-0.71	10.42	3.71	-3.40	5.82	9.24	1.37	-0.64
EUR3y3y2y	4.78	11.17	4.36	8.66	2.87	0.52	-4.39	-5.88	-5.14	5.33	-3.37	-4.72	-2.81	-20.44	-14.08
EUR3y4y1y	3.88	2.59	2.46	4.13	6.20	6.05	-6.04	-5.09	-0.93	0.94	2.06	-2.49	0.56	-16.41	-8.43
EUR3y5y5y	3.37	6.95	7.29	7.49	3.42	3.28	-1.47	-1.02	1.01	-1.30	0.90	6.71	8.69	-4.05	2.83
EUR4y1y1y	-16.00	10.11	2.06	11.91	2.46	-4.53	4.17	-3.59	-6.99	15.63	3.55	38.77	-18.96	8.78	-26.05
EUR4y1y4y	-7.39	-8.22	-8.97	-5.68	-4.44	-3.38	-5.81	-8.27	-4.66	4.72	-6.02	2.43	15.83	-18.93	-8.09
EUR4y2y3y	-29.24	-3.85	14.90	-2.14	-7.53	-11.02	-4.99	-9.08	-6.66	-1.01	-10.09	15.32	23.54	-50.41	-17.16
EUR4y2y8y	3.01	1.66	6.54	11.85	0.06	0.62	-2.25	-2.76	0.21	-1.10	-1.69	1.65	3.84	2.99	-0.73
EUR4y3y2y	-7.90	-10.56	25.19	3.45	2.10	-6.48	-10.74	-13.12	-5.54	-9.83	-14.20	-14.58	14.28	-39.51	-28.92
EUR4y4y1y	-4.23	-7.23	-2.75	-0.38	5.08	6.38	-9.66	-7.74	-0.26	-1.91	0.14	-6.02	5.68	-31.13	-10.40
EUR4y5y5y	2.57	5.75	4.56	9.12	4.71	4.49	-0.57	1.90	0.07	0.02	0.19	5.75	8.04	0.23	0.73
EUR5y1y1y	-11.10	-9.81	-1.19	-1.10	-4.28	-2.18	0.33	0.71	-1.60	10.75	2.23	46.86	-26.03	1.13	-6.52
EUR5y1y4y	-3.97	-18.04	-20.98	-16.32	-3.34	-13.62	-14.79	-13.50	-15.53	1.26	-12.01	0.66	23.76	-41.99	-0.56
EUR5y2y3y	-6.59	-2.55	-0.01	7.21	-4.90	-3.26	-2.86	0.31	0.04	-2.57	-5.97	3.91	12.85	-11.12	-1.30
EUR5y2y8y	-9.08	4.24	0.83	6.67	-1.83	-1.82	-1.79	-0.92	1.52	-2.79	-2.41	-2.19	2.50	3.85	-1.32
EUR5y3y2y	-3.00	-0.65	5.70	11.76	-2.09	-2.60	-0.99	-1.18	-2.07	-1.06	-3.78	0.21	12.76	-14.76	-4.25
EUR5y4y1y	-3.71	-10.01	2.01	6.38	7.85	4.30	0.81	-0.63	2.57	-1.81	-1.81	6.67	13.69	-42.38	-10.78
EUR5y5y5y	-3.77	2.69	4.20	8.32	3.52	4.81	1.36	2.75	2.58	-0.72	0.92	1.83	9.36	3.94	1.24
Classic Regression															
BackSel Regression															
Ridge Regression															
Lasso Regression															
KRR-RBF															
KNN															
CART															
Random Forest															
Grad Boost Reg															
MLP															
SVR-RBF															
Z-Score: CarryAtExpiry															
Z-Score: BE-Width															
Mean Pred															
Naive															

Figure 4: Heatmap with the historical Average Return (%) during test set.

throughout analysis suggests that Lasso Regression seems to provide in general the best results across a range of metrics and criteria. Figure 6 uses boxplots as a visualisation tool to decompose the aggregated results shown before, by informing per trade the returns obtained from using the Lasso Regression trading recommendation system.

Overall, some patterns can be spotted from the boxplots: (i) in general the medians are located above zero, meaning that more than half of the trades tended to yield positive returns; (ii) Lasso Regression is exploring long/short position regardless of the most frequent outcome for each MCCS tenure; (iii) it tended to perform well for EUR 3yXyXy trades and since these tended to be historically a challenging pick (medians are centred to zero in these trades), it means that the model is actually capturing some signal from the data and not naively guessing long/short positions; and (iv) the returns distribution, mostly with higher forward and swap tenure (second and third row), tended to be right-

	Classic Regression	BackSel Regression	Ridge Regression	Lasso Regression	KRR-RBF	kNN	CART	Random Forest	Grad Boost Reg	MLP	SvR-RBF	Z-Score: CarryAtExpiry	Z-Score: BE-Width	Mean Pred	Naive
EUR1y1y1y	0.39	-0.09	0.42	0.14	0.14	0.30	0.07	0.10	0.13	0.28	-0.02	0.34	0.06	-1.44	0.40
EUR1y1y4y	-0.31	0.48	0.36	0.57	0.19	-0.22	0.40	0.19	0.34	-0.20	-0.38	-0.18	-0.30	-0.72	0.51
EUR1y2y3y	-0.06	0.26	0.28	0.47	-0.56	-0.70	-0.43	-0.26	-0.03	0.20	-0.71	-0.04	0.04	-3.08	-0.57
EUR1y2y8y	-0.38	0.54	-0.21	0.39	-0.16	-0.14	-0.03	0.06	-0.22	-0.35	-0.61	0.32	0.28	-0.91	-0.39
EUR1y3y2y	0.01	-0.87	0.19	0.36	0.23	-0.01	-0.90	-0.46	-0.17	-0.04	0.00	-0.24	-0.10	-2.96	-0.71
EUR1y4y1y	-0.16	-0.11	0.01	0.23	0.25	0.06	-0.67	-0.56	-0.23	-0.08	-0.06	-0.40	-0.13	-3.12	-0.76
EUR1y5y5y	0.16	0.45	0.55	0.43	-0.32	0.12	-0.05	0.00	0.12	-0.04	-0.48	0.34	0.36	-0.20	-0.78
EUR2y1y1y	0.78	0.29	0.52	0.23	0.13	0.16	0.17	0.22	0.28	0.82	-0.32	0.73	-0.09	-0.60	-0.45
EUR2y1y4y	0.41	-0.01	0.26	0.37	-0.48	-0.70	0.14	-0.12	0.18	-0.07	-0.54	0.17	0.15	-0.15	-0.73
EUR2y2y3y	0.80	0.37	0.55	0.45	0.04	-0.76	0.28	0.37	0.18	0.43	-0.08	0.27	0.01	-1.34	-0.99
EUR2y2y8y	-0.19	0.60	0.44	0.39	-0.29	-0.29	0.42	0.22	0.40	-0.03	-0.19	0.34	0.37	-0.58	-0.94
EUR2y3y2y	0.10	0.37	0.43	0.62	0.07	0.13	-0.80	-0.92	-0.15	0.15	-0.16	-0.12	0.00	-1.36	-0.99
EUR2y4y1y	-0.24	-0.22	0.44	0.41	0.43	0.29	-0.55	-0.66	-0.52	0.10	0.23	-0.15	0.07	-1.50	-0.81
EUR2y5y5y	-0.16	0.61	0.17	0.46	0.27	0.06	-0.19	0.25	0.37	-0.06	-0.34	0.53	0.69	-0.49	-0.36
EUR3y1y1y	0.61	0.32	0.40	0.35	0.11	-0.07	-0.07	-0.06	0.17	0.62	0.04	1.21	-0.41	0.07	-0.22
EUR3y1y4y	0.14	-0.02	0.10	0.25	-0.18	-0.68	-0.19	-0.20	-0.09	0.00	-0.52	0.02	0.46	-1.78	-1.21
EUR3y2y3y	0.32	0.31	0.34	0.42	0.01	-0.04	0.26	0.15	0.27	0.43	-0.10	0.35	0.01	-0.80	-0.77
EUR3y2y8y	-0.60	0.48	0.31	0.55	0.03	-0.06	-0.14	-0.06	0.43	0.23	-0.32	0.28	0.43	0.08	-0.12
EUR3y3y2y	0.22	0.53	0.20	0.41	0.15	0.02	-0.31	-0.42	-0.36	0.26	-0.18	-0.19	-0.08	-0.75	-0.58
EUR3y4y1y	0.26	0.25	0.18	0.27	0.42	0.40	-0.63	-0.55	-0.10	0.09	0.16	-0.18	0.03	-1.01	-0.98
EUR3y5y5y	0.22	0.45	0.48	0.53	0.30	0.31	-0.17	-0.11	0.10	-0.19	0.07	0.44	0.68	-0.33	0.45
EUR4y1y1y	-0.31	0.27	0.07	0.36	0.14	-0.22	0.12	-0.11	-0.18	0.68	0.13	1.56	-0.38	0.24	-0.71
EUR4y1y4y	-0.33	-0.43	-0.46	-0.20	-0.32	-0.23	-0.44	-0.37	-0.17	0.18	-0.33	0.07	0.54	-1.18	-0.60
EUR4y2y3y	-0.85	-0.16	0.42	-0.06	-0.32	-0.46	-0.18	-0.30	-0.27	-0.04	-0.36	0.44	0.63	-1.47	-0.73
EUR4y2y8y	-0.40	0.15	0.30	0.67	0.01	0.06	-0.26	-0.25	0.02	-0.10	-0.19	0.09	0.21	0.25	-0.16
EUR4y3y2y	-0.19	-0.40	0.45	0.12	0.10	-0.24	-0.40	-0.49	-0.20	-0.61	-0.59	-0.29	0.40	-1.08	-0.58
EUR4y4y1y	-0.30	-0.40	-0.14	-0.02	0.26	0.34	-0.65	-0.53	-0.02	-0.18	0.01	-0.31	0.24	-1.41	-0.92
EUR4y5y5y	0.24	0.44	0.34	0.80	0.49	0.44	-0.06	0.25	0.01	0.00	0.02	0.35	0.57	0.02	0.19
EUR5y1y1y	-0.21	-0.33	-0.03	-0.03	-0.18	-0.07	0.01	0.02	-0.05	0.32	0.07	1.43	-0.42	0.03	-0.13
EUR5y1y4y	-0.13	-0.44	-0.59	-0.48	-0.11	-0.29	-0.37	-0.33	-0.41	0.03	-0.27	0.01	0.46	-1.15	-0.02
EUR5y2y3y	-0.57	-0.16	-0.00	0.41	-0.41	-0.24	-0.30	0.03	0.00	-0.21	-0.47	0.23	0.79	-0.72	-0.21
EUR5y2y8y	-0.53	0.28	0.06	0.53	-0.26	-0.21	-0.23	-0.17	0.14	-0.33	-0.30	-0.13	0.14	0.27	-0.16
EUR5y3y2y	-0.23	-0.05	0.29	0.40	-0.16	-0.14	-0.08	-0.07	-0.13	-0.11	-0.26	0.01	0.48	-0.55	-0.44
EUR5y4y1y	-0.20	-0.45	0.08	0.25	0.29	0.19	0.06	-0.05	0.18	-0.10	-0.08	-0.30	0.48	-1.65	-0.72
EUR5y5y5y	-0.30	0.21	0.26	0.63	0.34	0.41	0.13	0.30	0.29	-0.11	0.08	0.10	0.55	0.25	0.19

Figure 5: Heatmap with the Information Ratio during the test set.

skewed – because the third quartile is far from the median, whereas the first is squeezed towards the median. This last fact denounces that Lasso Regression frequently generates small negative outcomes, and dangerous scenarios are not as likely, which tends to be a desired property for quantitative strategies in general. Given that, we now look at the aggregated returns harvested by Lasso Regression during its test phase. These results are consolidated in Figure 7, where: (i) the top plot shows the average return with standard deviations obtained across all MCCS per trading week; (ii) the middle reveal histograms, where the left one represents the returns obtained across all MCCS regardless of the trading date, while the right ones displayed the same data but conditioned per position; and (iii) finally the bottom image presents the trading success rate for each long/short position suggested by the model (left), with the break down by long/short position displayed as well (right). To clarify, trading success in this context means being long/short when the returns of trade were

Table 6: Average ranks, Friedman and Holm post-hoc statistical tests and analysis for Information Ratio.

Model	Avg Rank	Z-score	p-value	Holm Correction
Mean Pred	12.86	9.63	<0.0001	0.0036
Naive	11.89	8.66	<0.0001	0.0038
SVR-RBF	10.60	7.37	<0.0001	0.0042
CART	10.11	6.89	<0.0001	0.0045
Random Forest	9.31	6.09	<0.0001	0.0050
kNN	8.23	5.00	<0.0001	0.0056
Classic Regression	8.17	4.94	<0.0001	0.0063
Grad Boost Reg	7.69	4.46	<0.0001	0.0071
KRR-RBF	7.49	4.26	<0.0001	0.0083
MLP	7.20	3.97	<0.0001	0.0100
BackSel Regression	6.37	3.14	0.0008	0.0125
Z-Score: CarryAtExpiry	6.09	2.86	0.0021	0.0167
Z-Score: BE-Width	5.91	2.69	0.0036	0.0250
Ridge Regression	4.86	1.63	0.0517	0.0500
Lasso Regression	3.23	-	-	-
Friedman Chi-Square	117.22	<0.0001		

positive/negative regardless of its magnitude. In respect to the top image, we can see that Lasso Regression started well during the first year but suffered a drawdown in the second to third year. This period was marked by higher volatility, mainly due to the final developments of the Euro Crisis period (2010-2012). However, from the third year onwards the average returns always scored positive values, usually ranging from 10% to 20% in average. Such performance can be seen stamped on the middle left histogram, where the bulk of returns lies above zero, and not only that but concentrated close to 15%. This performance was largely generated by Lasso Regression suggesting short positions (middle right), while the long positions were not so successful. Such pattern can be better seen in the histogram located at the bottom, where a bimodal distribution for trading recommendation success rate is depicted. Probably the verified outperformance coming from taking short positions in the MCCS is linked with betting against the volatility/variance risk-premium trade [6]. Roughly this strategy harvest the premium paid by a counterpart for the insurance on large swings in the market (almost the same as selling a put for equities options). Since in general, the market tends to remain range-bounded, the investor shorting the trade can repurchase it later for a smaller premium, profiting from this differential. Lasso Regression did dynamically the opposite and profited from it, largely because in this last 5-6 years was populated of higher volatility periods and tail events.

Figure 8 help us to analyse which features are being most significant by Lasso Regression for each particular trade.

Each cell corresponds to a normalised t-stats⁶ from the model coefficients built in the last step from k-rolling-cv. Implied Vol was the most significant feature pointed out by Lasso Regression, is negatively related with the MCCS returns. Other important features were the BE Width – slightly positively correlated with returns – and the Carry at Expiry – negatively related, but probably due to the depressed levels of carry that has been seen in the last batch of data. Lasso Regression promoted in general very sparse models, being

⁶By normalised t-stats we mean dividing each coefficient t-stat by the sum of the absolute values of all t-stats in the model. The result is a number between -1 and +1, indicating the significant magnitude in comparison to other variables, as well as the direction in which it affects the model predictions. We multiplied it by a one hundred just to work on a more convenient scale of -100% and 100%.

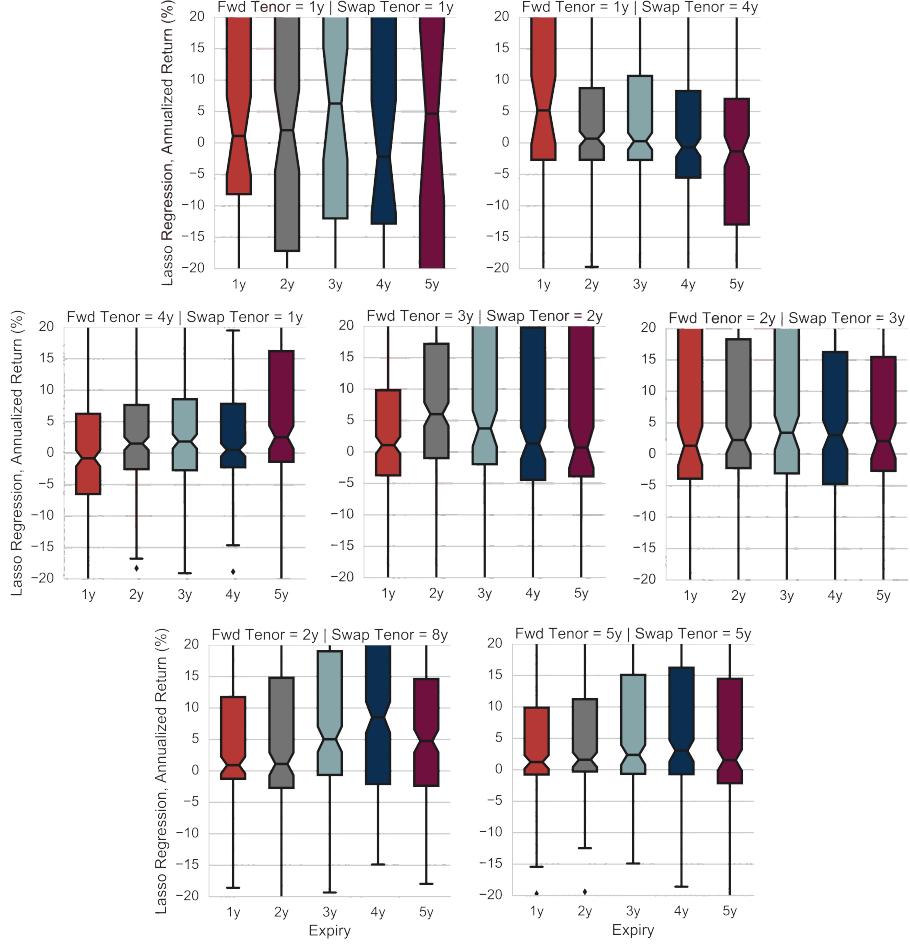


Figure 6: Boxplot with the returns obtained from Lasso Regression for each instrument. All y-axes were fixed between 0.3 and -0.3 (30% and -30% annualised return) to facilitate their visualisations and comparisons.

the other features playing specific roles for some trades like Time Carry for short-dated trades. Finally, the lagged returns were just relevant for few trades, and perhaps could be omitted for certain trades in the future to guarantee a broader dataset. We close this section showing results of Lasso Regression for a specific trade: EUR 4y5y5y – Figure 9.

Starting from the top, it can be seen that the Lasso Regression was able to predict reasonably well the observed returns after 1 year holding this trade. It is not perceived any over/underestimation of values, with the mirror-shape format indicating a good fit. This observation is reflected in the middle image, where the long/short positions track well the observed returns of the EUR 4y5y5y. The main mistakes are due to events that were not incorporated into the model and perhaps were also unforeseeable: (i) March to April 2012, possibly due to the Greek Debt Restructuring Agreement; (ii) May to November 2013, linked with the Taper Tantrum event in the US and its effect on Europe rates.

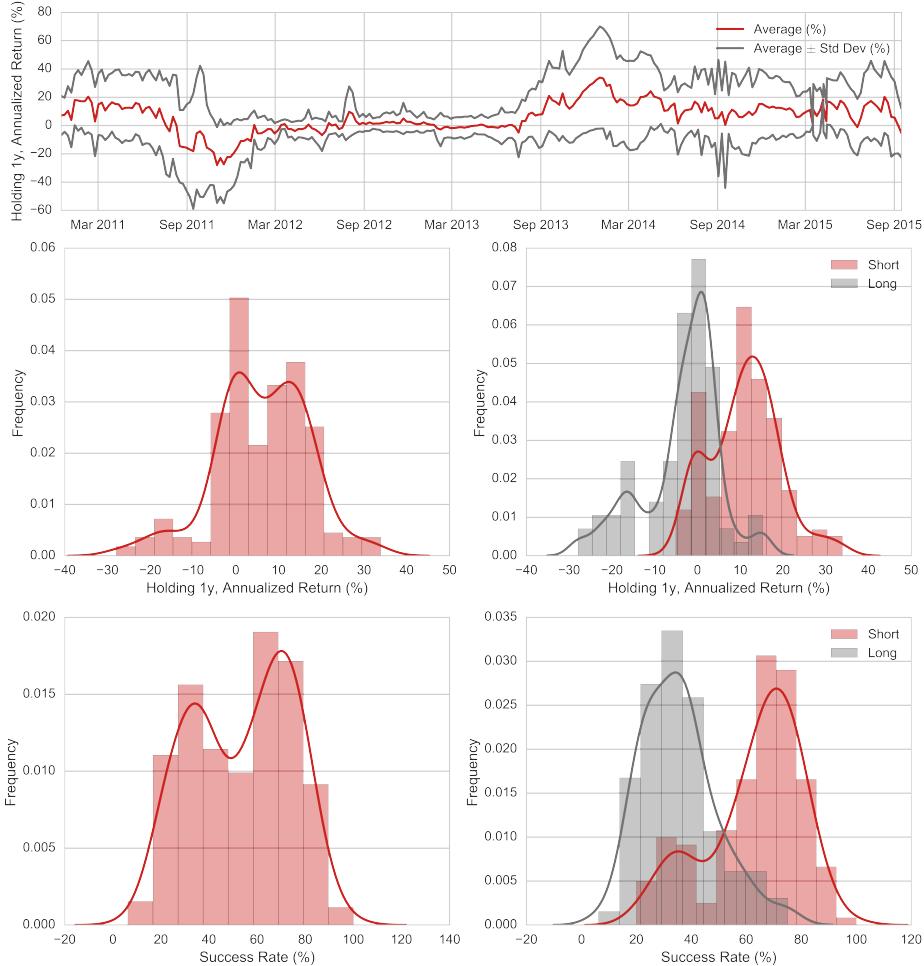


Figure 7: Aggregated returns over the period and histogram of aggregated returns and success rate from trades suggested using Lasso Regression.

Although such events have influenced in the strategies return, last plot shows that the trading success of Lasso Regression has attained a historical Area Under the Curve (AUC) of 0.85, with the capacity to control the false alarms to 5% and still recommend trades accurately 40% of the time. This is a good indicator since in the onset of the recommendation system it is better to reduce the chances of suggesting a bad trade than missing a good pick.

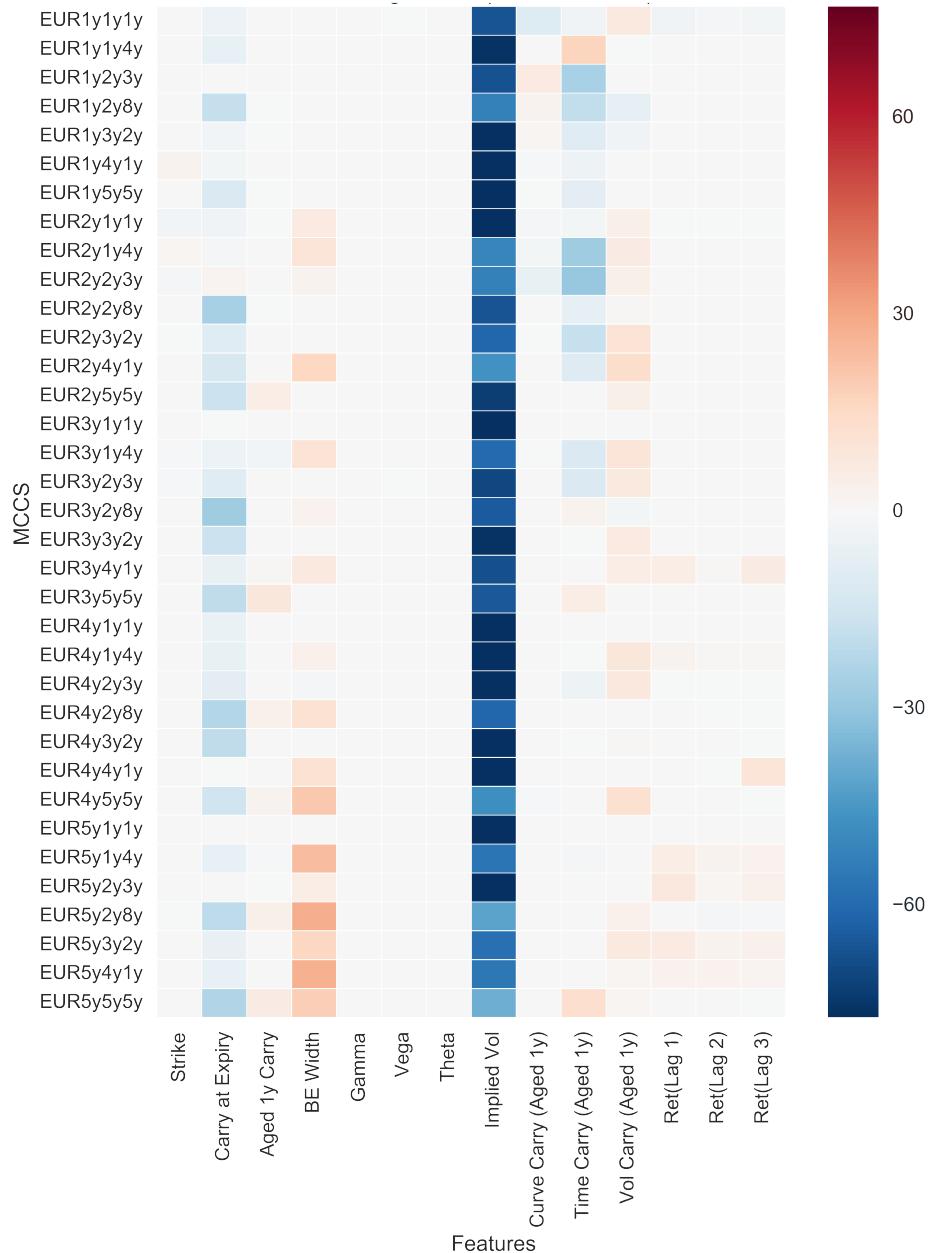


Figure 8: Heatmap with the Feature Significance (%) obtained from normalised t-stats of Lasso Regression coefficients.

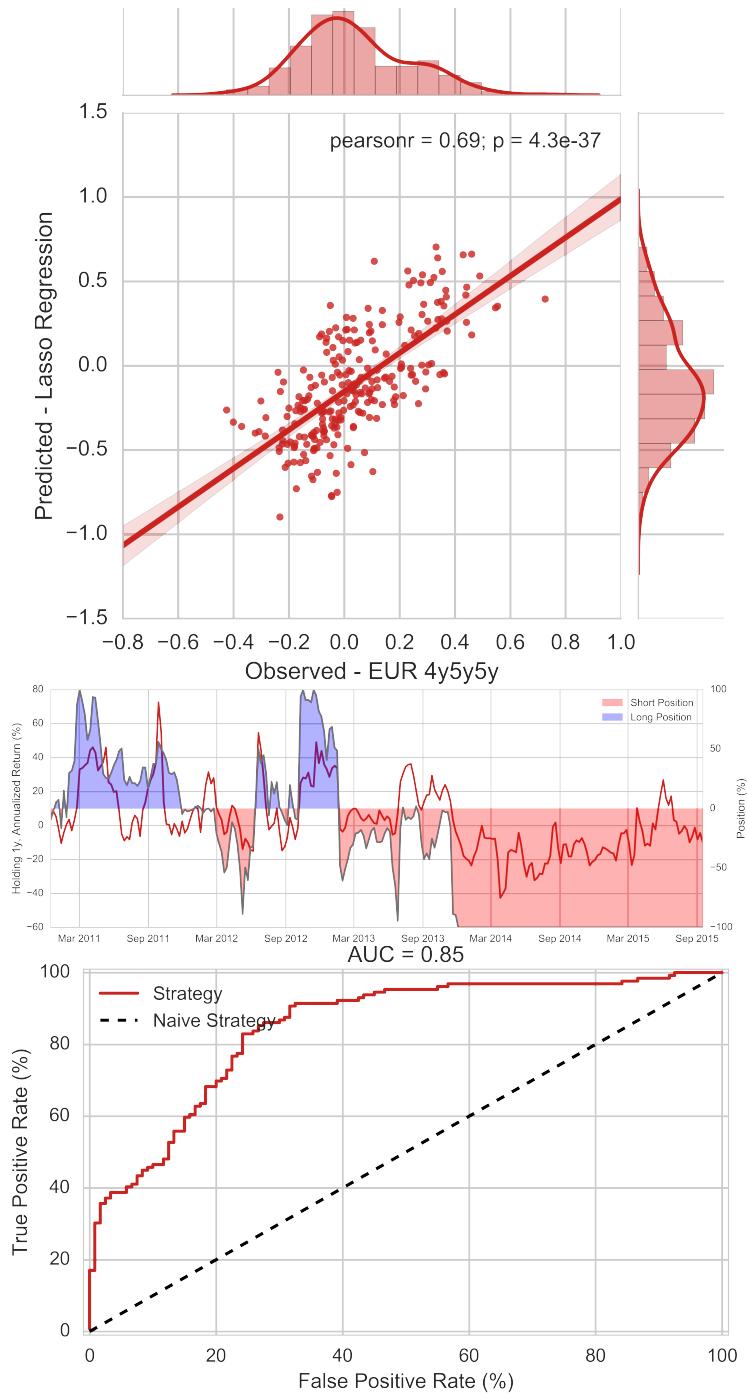


Figure 9: Lasso Regression results for EUR 4y5y5y: predicted versus observed values, returns and long/short signals over the period and receiving operating characteristic curve based on the success rate.

5 Conclusions

This work proposed a trading recommendation system for Mid-Curve Calendar Spread Trades (MCCS). We proposed a recommendation system that could analyse and rank a set of fixed income derivatives trades. Our first experiment is designing and applying this method for Mid-Curve Calendar Spread trades. Therefore, we started the methodology by showing the dataset: it comprised of 35 MCCS trades, ranging from September 2006 to September 2016, with different expirations, forward and swap tenures. For each particular trade, we described how the sampling of inputs (metrics, sensitivities and lagged returns) and outputs (returns from unwinding the trade after one year of its start) were computed on a weekly basis. Then, we displayed the modelling strategy by highlighting the models that were trained as well as which hyperparameters were investigated during the nested resampling step. Before entering the results section, we presented the backtesting setting with the performance measures used to compare different methodologies.

Most models provided results better than the modelling benchmarks (Mean and Naive), yet very few were able to outperform the trader’s benchmarks. Our results suggested that linear models with shrinkage procedures (e.g., Ridge and Lasso) tended to perform better than their nonlinear counterparts (like Kernel Ridge Regression, SVR and MLP). Also, regarding interpretability, they tend to be easier to convey to the traders, since most are versed in linear models. When we delved into Lasso Regression results, we found out that this model wielded some interesting features like: (i) it learned a type of volatility buying/selling strategy without being programmed to do so; (ii) its returns distribution across all MCCS tended to be right-skewed, meaning that we are more hedged towards dangerous scenarios with greater chances of upsides; (iii) it matched traders view on selecting good trades, but adding some dynamic view on it since Carry at Expiry is now negatively linked with returns, rather than the original view from the traders. We believe that Lasso Regression will be our choice for a first version of the trading recommendation system, with future developments giving space to different models and mixed approaches.

Acknowledgment

Adriano Soares Koshiyama wants to acknowledge the funding for its PhD studies provided by the Brazilian Research Council (CNPq) through the Science Without Borders program. Also, the authors would like to thanks, Guillaume Andrieux, Tomoya Horiuchi, Gerald Rushton, Tam Rajendran, and Anthony Morris for all the comments and support during this research.

References

- [1] Emmanuel Acar and Stephen Satchell. *Advanced trading rules*. Butterworth-Heinemann, 2002.
- [2] Rebecca M. Baker, Tahani Coolen-Maturi, and Frank P. A. Coolen. Nonparametric predictive inference for stock returns. *Journal of Applied Statistics*, 44(8):1333–1349, 2017.

- [3] C Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2007.
- [4] Damiano Brigo and Fabio Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.
- [5] John Y Campbell, Andrew Wen-Chuan Lo, and Archie Craig MacKinlay. *The econometrics of financial markets*. Princeton University press, 1997.
- [6] Hoyong Choi, Philippe Mueller, and Andrea Vedolin. Bond variance risk premiums. *Review of Finance*, 21(3):987–1022, 2017.
- [7] Eunsuk Chong, Chulwoo Han, and Frank C. Park. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187 – 205, 2017.
- [8] Howard Corb. *Interest Rate Swaps and Other Derivatives*. Columbia University Press, 2012.
- [9] Y. Deng, F. Bao, Y. Kong, Z. Ren, and Q. Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems*, 28(3):653–664, 2017.
- [10] Joaquín Derrac, Salvador García, Daniel Molina, and Francisco Herrera. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1):3–18, 2011.
- [11] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.
- [12] Graham Elliott, Antonio Gargano, and Allan Timmermann. Complete subset regressions. *Journal of Econometrics*, 177(2):357 – 373, 2013.
- [13] Nick Firoozye and Qilong Zhang. Turbo carry zooms ahead: A performance update of the turbo-carry trades. *Nomura International plc, Nomura Research*, 2014.
- [14] Nick Firoozye and Qilong Zhang. Usd short-term front-end turbo carry usd 1m1y2y trades: Short horizon for sizeable carry. *Nomura International plc, Nomura Research*, 2014.
- [15] Nick Firoozye and Xiaowei Zheng. Market update: Forward vol and mid-curve calendar spreads in usd and eur recent levels and carry and trades of note. *Nomura International plc, Nomura Research*, 2016.
- [16] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.
- [17] R. Gencay and Min Qi. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12(4):726–734, 2001.

- [18] Eduardo A. Gerlein, Martin McGinnity, Ammar Belatreche, and Sonya Coleman. Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*, 54:193 – 207, 2016.
- [19] Gyu-Sik Han and Jaewook Lee. Prediction of pricing and hedging errors for equity linked warrants with gaussian process models. *Expert Systems with Applications*, 35(12):515 – 523, 2008.
- [20] Simon S Haykin. *Neural networks and learning machines*, volume 3. Pearson, 2009.
- [21] Rob Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.
- [22] Andreas Karathanasopoulos, Konstantinos Athanasios Theofilatos, Georgios Sermpinis, Christian Dunis, Sovan Mitra, and Charalampos Stasinakis. Stock market prediction using evolutionary support vector machines: an application to the ase20 index. *The European Journal of Finance*, 22(12):1145–1163, 2016.
- [23] Christopher Krauss, Xuan Anh Do, and Nicolas Huck. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the s&p 500. *European Journal of Operational Research*, 259(2):689 – 702, 2017.
- [24] Burton G Malkiel. The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, 17(1):59–82, 2003.
- [25] Masafumi Nakano, Akihiko Takahashi, and Soichiro Takahashi. Generalized exponential moving average (ema) model with particle filtering and anomaly detection. *Expert Systems with Applications*, 73:187 – 200, 2017.
- [26] Sheldon Natenberg. *Option volatility and pricing: advanced trading strategies and techniques*. McGraw Hill Professional, 2014.
- [27] Hyejin Park, Namhyoung Kim, and Jaewook Lee. Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over {KOSPI} 200 index options. *Expert Systems with Applications*, 41(11):5227 – 5237, 2014.
- [28] Hyejin Park and Jaewook Lee. Forecasting nonnegative option price distributions using bayesian kernel methods. *Expert Systems with Applications*, 39(18):13243 – 13252, 2012.
- [29] Riccardo Rebonato, Kenneth McKay, and Richard White. *The SABR/LIBOR Market Model: Pricing, calibration and hedging for complex interest-rate derivatives*. John Wiley & Sons, 2011.
- [30] J. Beleza Sousa, M. L. Esquvel, and R. M. Gaspar. Machine learning vasicek model calibration with gaussian processes. *Communications in Statistics - Simulation and Computation*, 41(6):776–786, 2012.

- [31] Christian von Spreckelsen, Hans-Jrg von Mettenheim, and Michael H. Breitner. Real-time pricing and hedging of options on currency futures with artificial neural networks. *Journal of Forecasting*, 33(6):419–432, 2014.
- [32] Tianle Zhou, Shangce Gao, Jiahai Wang, Chaoyi Chu, Yuki Todo, and Zheng Tang. Financial time series prediction using a dendritic neuron model. *Knowledge-Based Systems*, 105:214 – 224, 2016.
- [33] Xiaocong Zhou, Jouchi Nakajima, and Mike West. Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *International Journal of Forecasting*, 30(4):963–980, 2014.