Artificial Neural Networks in Fixed Income Markets for Yield Curve Forecasting

Manuel Nunes[a,1,*], Enrico Gerding[a,1], Frank McGroarty[b,1], Mahesan Niranjan[a,1]

[a]Electronics and Computer Science, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom
[b]Southampton Business School, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom

Abstract

The yield curve is the centrepiece in bond markets, a massive asset class with an overall size of USD100 trillion that remains relatively under-investigated using machine learning. This paper is the first comprehensive study using artificial neural networks in the context of yield curve forecasting. Specifically, two models were used for forecasting the European yield curve: multivariate linear regression and multilayer perceptron (MLP), at five forecasting horizons, from next day to 20 days ahead. Five variants of the MLP were analysed with different sets of features: target to predict (univariate); the most relevant features; all generated features; and the former two incorporating synthetic data generated by the linear regression model. Additionally, two different techniques of multitask learning were employed: simultaneous modelling and transformation into multiple single task learning. The results show that considering all forecasting horizons, the MLP using the most relevant features achieved the best results and the addition of synthetic data tends to improve accuracy. Furthermore, different targets and forecasting horizons resulted in different relevant features, reinforcing the importance of custom-built models. In the two multitask learning methodologies no clear differentiation could be demonstrated, and several explaining factors are identified. Overall, the outcome is very encouraging for the development of better forecasting systems for fixed income markets.

Keywords: machine learning, neural network, multitask learning, yield curve forecasting, yield forecasting, bond market

1. Introduction

The fixed income market is one of the most important sources of finance for governments, national and supranational institutions, banks, and private and public corporations that have access to this market. In fact, this is a massive asset class, and considering that the most significant part is represented by the bond market, the overall size is a staggering USD 102.0 trillion, as of 31-Dec-2016 (Bloomberg, 2017). This compares with a global equity market of USD 66.3 trillion. In addition, its importance also derives from two crucial sectors and top investors in fixed income: pension funds (USD 28.4 trillion) and insurance companies (USD 28.2 trillion). These two sectors are also the most important clients of the USD 43.2 trillion in investment funds (OECD, 2015a).

In our research we focus on predicting the yield curve, which is the centrepiece of bond markets. Taking government bonds as an example, the yield curve represents the annualised interest rates (or "yield") that a particular government has to pay to borrow funds from investors, as a function of the length of time in which the borrowing occurs (or "time to maturity"). The yield curve is also known as the term structure of interest rates.

The study of this asset class gains special relevance in the present moment for all the parties intervening in the financial industry and respective regulatory bodies. Indeed, fixed income markets are presently operating under very special circumstances in historic terms. First, we observe higher market risk due to potential inversion of the cycle, following declining yields in fixed income markets for more than three decades. Second, we observe higher levels of risk in investment portfolios, as a result of the new market conditions and the very low yield environment, leaving investors "searching for yield" (Becker and Ivashina, 2015; Kräussl et al., 2017; Mello, 2015; OECD, 2015a,b). Third and last, we observe higher levels of uncertainty and lower prediction ability of conventional models and tools used for policy making and asset management, following unprecedented actions of central banks around the world (Gogas et al., 2015; Morell, 2017).

From the literature it is evident that the vast majority of the academic works carried out on the use of machine learning in economics and finance are applicable to equities. See, for example, the works of Agrawal et al. (2013), Ballings et al. (2015), Booth et al. (2014a), Dunis et al. (2016), Eilers et al. (2014) and Vui et al. (2013), just to mention a few studies. Although the literature in this field is abundant, as well as in foreign exchange markets (Choudhry et al., 2009, 2012; Fletcher, 2012; Fletcher and Shawe-Taylor, 2013; Gradojevic and Yang, 2006; Huang et al., 2007), much less scientific work has been produced covering machine learning in fixed income markets (Castellani and Santos, 2006; Dunis and Morrison, 2007; Kanevski et al., 2008; Kanevski and Timonin, 2010; Sambasivan and Das, 2017). This is the case despite the paramount importance of this asset class for any economy. It

*Corresponding author. Tel: +44 (0)23 8059 2929.
Email addresses: m.nunes@soton.ac.uk (Manuel Nunes), eg@ecs.soton.ac.uk (Enrico Gerding), f.j.mcgroarty@soton.ac.uk (Frank McGroarty), mn@ecs.soton.ac.uk (Mahesan Niranjan)

[1]Declarations of interest: none. All authors have approved the final article.

clearly represents an opportunity, but also further stretches the challenges for our research. This will be further detailed in the literature review (Section 2).

Given this, the main innovative contribution of our research has been the extension of existing machine learning models to the fixed income asset class that has not been comprehensively and extensively covered using these techniques. To the best of the authors' knowledge this is the first comprehensive study on yield curve forecasting using artificial neural networks (ANN) and multitask learning techniques. Specifically, the contributions of the paper are:

- An assessment of selected machine learning techniques and evaluation of their adequacy for forecasting the European yield curve, at five forecasting horizons (0 or next day, 5, 10, 15 and 20 days into the future). Specifically, we consider multivariate linear regression models and multilayer perceptron (MLP) models.
- Consideration of a wide range of macroeconomic and financial time series (159), covering the period 1999-2017, and through feature selection determining the most relevant features.
- Estimation of the impact of additional financial market and macroeconomic information on forecasting accuracy.
- Evaluation of two different approaches to multitask learning, through the simultaneous modelling of all targets (MTL) and the transformation into multiple single task learning (STL) problems.
- Testing methodologies that could result in improved forecasting, such as the inclusion of synthetic data generated by other models.

The remainder of this paper is structured as follows. In Section 2 the relevant literature is presented. In Section 3 the theory behind the selected models used in our research and the multitask learning methodologies are described. Section 4 describes the dataset used and pre-modelling operations, while in Section 5 a global view of the empirical work performed is presented and explained. The results are presented and discussed in Section 6. Finally, in Section 7 the main conclusions are outlined together with potential future work.

2. Literature review

The literature review will cover a number of topics. First, classical financial models for time series and yield curve models are briefly presented (Section 2.1). Then, a review of the literature using machine learning specifically for fixed income markets is carried out (Section 2.2). Finally, given that this literature is limited, the review is extended to adjacent areas in financial markets (Section 2.3).

2.1. Classical financial modelling

Time series modelling has been a well established subject for many years (Hamilton, 1994; Enders, 2014; Box et al., 2015). The most popular models for time series analysis and forecasting are the autoregressive moving average (ARMA) and the autoregressive integrated moving average (ARIMA) models, a generalization of the former (Box and Jenkins, 1968). Autocorrelation is the basic assumption in these models. However, in a comprehensive literature review on the characterisation of financial time series (Sewell, 2011), the author clearly concludes that the autocorrelation of price changes, or returns, is largely insignificant.

Models for the complete yield curve pertain to two main groups: yields-only models, using only yield data to estimate the complete yield curve; and yields-macro models, which predict specified macroeconomic variables using the yield curve or vice-versa. In the former group, two of the most widely used models within this category are polynomial and Nelson-Siegel functions (Nelson and Siegel, 1987).

Most of these yields-macro models assume that the influence happens only in one-direction, macroeconomic variables affecting the yield curve or vice-versa, but without feedback effects. However, other models have been developed to study the possible feedback effects (Diebold and Li, 2006; Diebold et al., 2006; Diebold and Rudebusch, 2013). Departing from the Nelson and Siegel curve they developed a yield curve model incorporating both intrinsic yield factors (level, slope, and curvature) and macroeconomic factors (manufacturing capacity utilization, change of consumer price index over the past 12 months or annual price inflation and federal funds rate). This model is generally known as the Dynamic Nelson-Siegel model. These models are undoubtedly valuable and well established in the industry, but for new research using machine learning techniques, the limitation to factors considered in the model is undesirable. This is also emphasised by the fact that several studies in the literature note that additional domain-specific features could improve forecasting ability (Dunis and Morrison, 2007; Mettenheim and Breitner, 2010, 2011).

A different alternative for yield curve modelling may be found in an emerging area in statistics called functional data analysis. This is a nonparametric statistical technique dealing with infinite-dimensional data as in the case of functions, curves, surfaces and images. This type of model, has been applied to forecast the US yield curve, where the yield curve is considered the functional variable (curve) that links maturities to yields (Caldeira and Torrent, 2017). The findings of this research produced mixed results, and did not demonstrate a systematic superiority of this approach. However, that was the case in several situations, in particular for forecasting short-maturity interest rates. Besides, the study did not include any macro or financial data apart from the yield curve itself, which could also be a limitation.

2.2. Machine learning models in fixed income applications

Studies applied specifically to fixed income markets are less common in the literature. In this market, we may be interested in modelling individual assets, such as individual bonds, bond indices, bond funds or bond futures. In this case the datasets for this purpose are time series and this is a single target regression problem. However, the main focus of our research is the yield curve. This is a more complex issue because the modelling target is a curve and not a single value. Specifically, one-dimension type problem in traditional financial modelling becomes a two-dimensional one: time and maturity (extra dimension).

Along this line, a study was conducted (Kanevski et al., 2008; Kanevski and Timonin, 2010) using spatial statistics, to map the yield curves into a two-dimensional space (maturity and time). Then, via interpolation using geostatistical or machine learning models, the authors reconstructed full yield curves from a specified number of points considered in the data as inputs. Promising results were obtained using artificial neural networks, although additional simulations are necessary under different market conditions and time horizons, to validate this methodology. Notwithstanding the potential of spatial / geostatistical models and possibly of hybrid approaches combining machine learning with those, our research is emphasising the use of machine learning techniques directly. What is more, those models base the entire mapping and forecasting processes on historic yield curve data only. For this reason, we perceive machine learning models with greater potential and flexibility. Furthermore, our work has the objective of going beyond the use of historic data from the target to predict. In particular, we will assess a wide range of potential explanatory features from a variety of fields.

A different approach is needed when modelling individual assets, since we do not have the additional maturity dimension. In this vein, Castellani and Santos (2006) used neural networks (a multilayer perceptron), among other models, to forecast monthly US 10-year Treasury bond yields using four economic indicators, namely: purchasing managers index (PMI), the consumer price index (CPI), the London interbank offered rate (Libor) and the volatility index (VIX). The results of this study were not very encouraging. In fact, as far as prediction accuracy is concerned, the best models were only marginally better than a basic one-step lag predictor, which predicts the yield of the US 10-year Treasury bond from the figure of the previous month. The study concludes by pointing out the difficulty in building reliable predictors for financial markets in general.

Another single-asset research was conducted by Dunis and Morrison (2007). By using state space modelling with a Kalman filter and neural network regression, they forecast the 10-year government bond yield of three countries: United Kingdom, United States and Germany. There are two interesting aspects in this study. First, they included a wide range of additional financial variables from main European countries, the United States and Japan, to work as features: bond yields, short-term interest rates, index stock prices, exchange rates and commodities. Second, the performance evaluation of the models was based on measures of accuracy, and also on results from a simulated trading strategy, with proper consideration of trading costs. The authors concluded that neural network regression models represent a promising alternative to more traditional techniques currently used in the industry.

From the studies covering directly fixed income, it was not possible to find a direct solution for modelling the yield curve using machine learning. This is a gap in terms of academic research and the aim of this paper is to fill this gap. A notable exception is the work carried out by Sambasivan and Das (2017) proposing a dynamic Gaussian process for modelling the yield curve. In this work, the authors compare the results of this machine learning model with multivariate time series forecasting (vector autoregressive model) and the dynamic Nelson-Siegel model. The results show that multivariate time series method performed best for yields with maturities up to 1 year, while the dynamic Gaussian process model was superior for the longer maturities (2 to 30 years). These results will be mentioned again in Section 6.2.5 for comparison purposes.

The studies described in this section were particularly fertile for potentially better models: from the use of ensembles to different types of hybrid models (Castellani and Santos, 2006; Kanevski et al., 2008; Kanevski and Timonin, 2010), with inclusion of broad information from several sources (Dunis and Morrison, 2007). Among those they should incorporate macroeconomic, financial and, whenever possible, practitioner type of information. Regarding machine learning models, mixed information and results were seen (Castellani and Santos, 2006; Dunis and Morrison, 2007) while using the same type of models, in this particular case, neural networks (multilayer perceptron).

2.3. Machine learning models in other financial applications

In foreign exchange markets, artificial neural networks have been used incorporating exogenous financial data for forecasting both the direction of the movement of the EUR-USD exchange rate and turning points in the prices of a basket of currencies (Fletcher, 2012; Fletcher and Shawe-Taylor, 2013). Other studies have been carried out using different nuances of ANN and high frequency market microstructure variables (Choudhry et al., 2012; Gradojevic and Yang, 2006; Huang et al., 2007). These studies have shown that they provide good results and can lead to profitable strategies, with proper consideration of transaction costs.

Studies covering equities are the most common applications of machine learning in financial markets. Starting with a published state-of-the-art review, for their broad scope, Vui et al. (2013) covered the application of artificial neural networks for stock market prediction, showing encouraging results with the use of this technique.

Moving into individual research studies, Arrieta-Ibarra and Lobato (2015) conducted a study using several machine learning techniques to forecast both stock market daily returns and squared returns. The results were not totally conclusive, but for predicting squared returns, neural networks and support vector machines (SVM) showed real potential for improving forecasting ability.

Since forecasting is frequently connected to the trading activity that could directly benefit from superior sources of information, some research studies aim to integrate the forecasting models within an automated trading system (Booth et al., 2014b, 2015). In this case, it was based on ensembles of random forests (RF), predicting price returns of the German DAX index. The results obtained were significantly better than simple averaging. Other studies have also emphasised the benefits of using ensembles for forecasting stock price direction with classifiers (Ballings et al., 2015) and for sentiment analysis in social applications (Araque et al., 2017).

Agrawal et al. (2013) and Arrieta-Ibarra and Lobato (2015) emphasise the challenges and difficulties of forecasting stock markets, a concern that is also applicable to fixed income markets. Nevertheless, it should be stressed that suitable techniques for market forecasting may be developed. In fact, there is evidence that they have been used in the industry (Burton, 2016; FRM, 2002; Kolanovic and Krishnamachari, 2017; Roux and Burton, 2017).

In the field of equity options, it emerges that radial basis functions (RBF) demonstrated the capacity to model the complex relationship between option price and the underlying stock price (Hutchinson et al., 1994; Niranjan, 1996), outperforming the parametric Black-Scholes model (Black and Scholes, 1973), most commonly used in the industry. Additionally, the inclusion of financial information leads to improved forecasting performance (Montesdeoca and Niranjan, 2016).

Other examples of the use of machine learning in financial applications include the prediction of recessions in the United States, applying support vector machine and using several interest rates from the yield curve to forecast the GDP cycle (Gogas et al., 2015). The results were promising but not completely satisfactory, with the out-of-sample overall accuracy of 66.7%, predicting correctly all recession periods one quarter ahead, but at the same time predicting as recession 60% of the growth periods, clearly undesirable.

Finally, a recurrent neural network (RNN) topology called shared layer perceptron was developed and used in financial applications (Mettenheim, 2010; Mettenheim and Breitner, 2010, 2011), which allows multi-asset and multi-step forecast. Positive results were obtained when compared to the benchmarks, in three forecasting applications: market value at risk, over the next 10 days; the economic indicator Baltic Dry Index, over the next 20 days, to identify a low entry point; and the sign of next day return of a portfolio.

In summary, the extended literature review on other asset classes and financial applications provided important input for our research. On type of model, although the results achieved were sometimes mixed, the following models were reported with positive results, in diverse markets: RBF, ANN, in particular MLP, SVM, RF, and RNN. Regarding methodologies that could result in improved forecasting and potentially better models, several studies mentioned the use of ensembles and different types of hybrid models (Ballings et al., 2015; Booth et al., 2014a; Vui et al., 2013). Regarding type of features, this is a common theme in the literature, reporting the benefits of including in the models additional information as features. These may include political, economic, financial and domain-specific factors (Agrawal et al., 2013; Choudhry et al., 2012; Fletcher, 2012; Fletcher and Shawe-Taylor, 2013; Montesdeoca and Niranjan, 2016; Vui et al., 2013). Again, as in the previous section, possible solutions for modelling the yield curve using machine learning techniques are not evident in these studies and respective applications.

3. Machine learning approaches

From the literature review, artificial neural networks, in particular the multilayer perceptron, stand out as a model with potential to be used as a forecasting tool in fixed income markets. This type of model possesses the necessary flexibility, taking into account the fact that we aim to incorporate a wide range of features. Further information on feed-forward neural networks and in particular the multilayer perceptron can be found elsewhere (see, for example, Bishop (2006), Hastie et al. (2013) and Rumelhart et al. (1986), the latter for the training process using the back-propagation algorithm). In this section, the multivariate linear regression model is briefly presented, with the main objective to describe the feature selection approach we use in this paper called LASSO. Additionally, the multitask learning methodology is described.

3.1. Linear regression

Linear regression models are still very popular nowadays, despite the advances in computer science. They are simple, for a large number of applications they provide adequate models, and the interpretability of those models is much higher. Unless more complex non-linear models offer a clear improvement versus the linear solution, one should favour the linear models due to its simplicity and advantages, in particular the easier optimisation process. It is also a good model to take as baseline to compare with more complex ones. The general equation for linear regression models can be written as follows (Bishop, 2006; Hastie et al., 2013; Niranjan, 2016):

$$\boldsymbol{f} = \boldsymbol{Y}\boldsymbol{a} \tag{1}$$

where $\boldsymbol{Y}$ is an $N \times (p+1)$ matrix, N being the number of observations and p the number of features; $\boldsymbol{a}$ is the $(p+1) \times 1$ vector of unknown parameters; and $\boldsymbol{f}$ is the $N \times 1$ vector of outputs. The unknown parameters are determined using the least squares method. The objective is to minimise the following error function:

$$E = ||\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{f}||^2 \tag{2}$$

Finally, the solution for the linear regression model can be obtained by equating the gradient to zero, or, alternatively, using a gradient descent algorithm to minimise Equation 2.

Feature selection can be performed using LASSO, which stands for Least Absolute Shrinkage and Selection Operator. The method can be expressed in the following form (Tibshirani, 1996):

$$E = ||\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{f}||_2^2 + \gamma||\boldsymbol{a}||_1^2 \tag{3}$$

where $||\ ||_1$ denotes the L^1-norm; $||\ ||_2$ the L^2-norm; and γ the regularisation parameter.

Hence, the LASSO method equates to performing a linear least squares regression on the variable, using an L^1-norm penalty for the weights (regularisation term). This constraint tends to lead to sparse solutions, enabling the identification of the most relevant features for the model. The reduction in features improves interpretability of models, helping also in cases of low bias / high variance, thus improving generalisation.

3.2. *Multitask learning*

In the machine learning domain, the standard methodology for regression problems is the modelling of one target variable (single task learning), using several inputs. For the yield curve, if we consider a reduced number of benchmarks, for example four, it would represent four different models to forecast the target bond yields. This is represented in Figure 1 using four neural networks. This method does not take into account the functional form of the yield curve. In other words, it does not consider that those interest rates in the yield curve tend to move together having some functional relationship, which could be beneficial for the model.

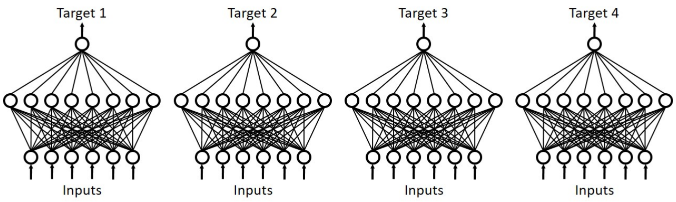


Figure 1: Single task learning for four different targets.

In contrast to single task learning, multitask learning enables the learning of several targets simultaneously. This is represented in Figure 2 and this methodology could be used to model the yield curve, through the modelling of its most relevant benchmarks.

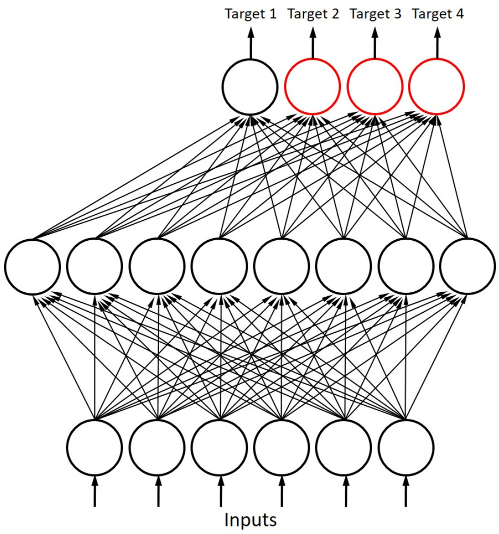


Figure 2: Multitask learning for four different targets.

From Figure 2 and the literature result some of the characteristics of multitask learning (Caruana, 1993, 1997; Borchani et al., 2015; Cai et al., 2014; Ruder, 2017): the hidden layer of the neural network is shared by all targets; the learning process occurs in parallel, simultaneously for all targets; some hidden units may specialise in specific targets, which can be useful for yield curve modelling where some features may be more important for short-term bonds and others for long-term bonds; the use of the domain specific information from additional targets functions as constraints to the overall model, improving generalisation accuracy.

A recent survey on multitask regression (Borchani et al., 2015) covered a varied number of applications of this methodology in different scientific fields, categorising the existing methods into two groups: problem transformation methods, in which the problem is transformed into independent single target problems; and algorithm adaptation methods, implying the modification of single output methods in order to handle multiple targets. Even though neural networks were not covered as a model in the survey, multitask learning using neural networks is not a new theme (Caruana, 1997; Ghosn and Bengio, 1997).

Taking into account the applications of this methodology, several uses strike as possible for financial time series prediction in the bond market, where targets would represent: different points in the yield curve at the same time t, representing a multi-asset process and enabling the forecast of the overall curve; the same yield of a particular bond, at different times, thus enabling the estimation of several time steps ahead in the future; a combination of previous items, enabling a multi-asset and multi-step forecast; additionally, when data is not available on time to be used as feature, it can still be used as target variable if it is relevant for the model.

In summary, multitask learning can be performed using two different methodologies: transforming the problem into multiple single target (STL) and using simultaneous modelling of all targets (MTL). Both techniques were used in our research.

4. Data

In this section we describe the dataset used, identifying features, targets and pre-modelling operations.

4.1. *Targets*

The focus of our work is the government bond asset class, which was selected for the following reasons. First, liquidity of this asset class is clearly higher than for the other bond classes. Second, the size of the market is also considerably higher. Third, this class encompasses a wide range of financial instruments available. Fourth and last, research on government bonds will attract the interest of entities such as national and supra-national institutions, in particular central banks, national government agencies managing the public debt, as well as asset management companies. Within this asset class, the Euro benchmark yield curve was selected and its modelling will be done through the modelling of its most relevant benchmarks. The benchmarks considered were: 3-month, 2, 5, 10 and 30-year bond yield, representing five targets to be predicted.

4.2. *Features*

Choosing relevant features is one of the most important factors to improve the performance of models. Given the interconnectedness and mutual influence of various asset classes in the markets, a large number of features from financial markets were considered. These were selected from government bond markets and from related classes and indicators: credit (corporate bonds), equities, currencies, commodities and volatility. Additional features were added, directly calculated from the previous features, mainly bond spreads, slope of the yield curve and simple technical analysis indicators. Furthermore, economic variables are also very important, as clearly exemplified by the well established yields-macro models presented in Section 2.1. Hence, a vast range of economic indicators is also included, from different geographic locations. The complete list includes 159 features and, due to its extension, is stored and made publicly available ([dataset] Nunes et al. (2018)).

4.3. Datasets

The datasets were obtained from Bloomberg database and they cover the period from January 1999 to April 2017 [Bloomberg (2017)]. This is a longer period than covered in other studies (Dunis and Morrison, 2007; Sambasivan and Das, 2017). From the markets' point of view, this is an interesting period to study, spanning from the euro inception date on the 1st of January, 1999. This is also the starting date for most time series of the Euro benchmarks, in particular the yield curve data. Additionally, this period covers several temporary bull and bear markets and market moving events, such as: the dot-com bubble in 2000; the global financial crisis of 2008-2009, the Great Recession; the subsequent European debt crisis; the European recession in 2012-2013; and several phases of quantitative easing by the US Federal Reserve, European and UK central banks. Of note is the fact that, the principal overall trend in the bond market during this period has been of declining yields, although with significant and frequent temporary reversals. Regarding data frequency, the selection was daily closing values, which are easily available for financial assets in general.

4.4. Generation of additional features

In financial time series there is a natural temporal order that cannot be disrupted during modelling, since that ordering has in itself relevant information. Taking this into account, it is worth incorporating into the models past values of the time series. Hence, new features are generated from the original ones, corresponding to lagged values of the respective time series. In our research, six time steps were considered (5 past values plus 1 target), based on previous studies (Mahler, 2009). Consequently, we generated from the original 159 features a total of 795 features (159×5). These were filtered by the feature selection process, described in detail in Sections 3.1 and 5.2.

4.5. Train-test split and normalisation

As is common, we divided the data into two groups, for training and testing the models. In this case, a 70% / 30% split was considered. The training set is a moving window of historic data up to the time step being considered, corresponding to the last known data. In order to enable cross-validation, twenty different bootstrap samples were extracted from the above mentioned moving window of training data. Cross-validation is especially important for the neural network model, but in order to have a fair comparison, the same methodology was followed for the linear regression model. Furthermore, due to the large size of the testing set and the computing time necessary to retrain and forecast all points, only 50 random points were selected from the testing dataset for forecasting. The error calculations are based on the fifty predictions of testing data points, which are unknown data when training the models (out of sample error).

Finally, all data was normalised by subtracting the mean and dividing by the standard deviation of the training dataset. This is also essential, given the wide range of features we are considering, which have very different scales in some cases.

5. Methodology

In this section, the details of the methodology adopted are presented, including various analyses carried out in advance to the modelling process to justify the parameters adopted. Then, all models considered in this study are detailed. Finally, the concept of retraining of models is outlined. A global view of the empirical work carried out is summarised in Table 1 and explained below.

Table 1: Summary of empirical work.

Parameters	
Original features	159
Generated features	795
Targets	3M, 2Y, 5Y, 10Y, 30Y
Forecasting horizons	0 (next day), 5, 10, 15, 20 days
Analyses	
Regularisation param.	0 to 4, step 0.1
Selected	2 and 4
Moving window size	30, 100, 300, 500, 1000, 2000, 3000, 3290
Selected	3000
No. of hidden units	5, 10, 20, 50, 100, 150, 200
Selected	10
MTL mode	Yields as targets Forecasting horizon as targets
Models	
LR Linear Reg	Linear regression
1. NN GenFeat	MLP with all generated features
2. NN RelFeat	MLP with relevant features
3. NN TgtOnly	MLP with target data only
4. NN RelFeat+LRdata	NN RelFeat with synthetic data from Linear Regression model
5. NN TgtOnly+LRdata	NN TgtOnly with synthetic data from Linear Regression model

5.1. Forecasting horizon

Given a specific training dataset, forecasting the next value in the time series should be less complex than forecasting further into the future, when the time distance to the known data increases. Taking this into consideration, a forecasting horizon parameter was introduced in this study, equal to the number of days, or time steps, from the next value of the time series. In practice, a forecasting horizon equal to zero corresponds to forecasting the next value, i.e. one time step ahead, while a forecasting horizon equal to 20 corresponds to predicting the next value plus 20 days ahead. Our research was conducted using a range from 0 to 20, with 5 days increment (Table 1). The next day plus 20-day range (working days) was considered as it corresponds to one month, approximately. These limits have also been used in other studies (Arrieta-Ibarra and Lobato, 2015).

5.2. Feature selection

It should be emphasised that the most relevant features for each target yield are not known in advance and this is why this study included a wide range of original features (Table 1) to be submitted to feature selection. Linear regression using the LASSO method was performed to select the most relevant features (Equation 3). A range from 0 to 4 was considered for the regularisation parameter γ, with values 2 and 4 being selected. This selection is explained in detail when discussing the results in Section 6.1. Furthermore, as the impact on relevant features can change for different forecasting periods, we determine the relevant features separately per target and per forecasting horizon, resulting in a total of 25 combinations.

5.3. Retraining of models and size of moving window

One of the characteristics of financial time series is that they become available at a specified frequency, in this case on a daily basis. As the new information becomes available it can be incorporated in the models and proceed with their retraining using the new training window (moving window) that results from eliminating the oldest values and including the newly available values. The retraining of models is feasible in practice and the technique was used to take full advantage of the models being used.

The moving window size is another parameter that needs to be set and it was the object of a specific sensitivity analysis to study its impact on forecasting errors. For that purpose, the range shown in Table 1 was considered. The results showed that better predictions were obtained with larger windows, with a significant improvement until it reached 2000 observations and then the benefits were much smaller. A final moving window size of 3000 observations was considered.

5.4. Number of hidden units

An additional analysis was conducted to define the number of hidden units to use in the neural network model. The range of hidden units shown in Table 1 was tested for each target and for both single task and multitask learning. The main conclusion from the results is that 10 hidden units is a good compromise for the subsequent studies, with significant overfitting observed for neural networks with more than 100 units.

5.5. Single task and multitask learning

The modelling was carried out using the concepts of multitask learning described in Section 3.2, both in multitask learning mode, that is, considering simultaneously all targets in the same model, and through problem transformation into five single task learning models. For the multitask learning mode, two analyses were considered: with yields as targets (multi-asset forecasting) and forecasting horizon as targets (multi-step forecasting).

5.6. Models

All models studied in our research are listed in Table 1, which includes both the multivariate linear regression model and the three main models considered using the MLP architecture. Model 1 uses the complete list of generated features. Model 2 uses the relevant features determined during the feature selection process (features selected per target and per forecasting horizon). Finally, in Model 3 only past values of the target(s) to predict are taken into account, i.e. an univariate type of model.

In addition, we observed that results obtained with linear regression performed surprisingly well in some cases (further detail in Section 6.2). For this reason, it was decided to test the performance of hybrid models, incorporating the alternatives referred above with better performance (Models 2 and 3) and synthetic data generated by the linear regression model, used as additional feature (Table 1).

The main metric used for model comparison was the mean squared error (MSE), which is commonly used for this purpose. Nevertheless, other metrics were also calculated: mean absolute error (MAE) and root mean squared error (RMSE). Additionally, the statistical significance of differences was determined for all possible combinations.

6. Results and discussion

In this section the main results are presented and discussed, divided into two separate topics. First, we present the feature selection results to identify the most relevant features. Second, a thorough comparison of the models used and their variants is carried out, together with a comparison to results from the literature.

6.1. Feature Selection

A typical example of the feature selection results obtained, in this case for the 30-year bond yield, is shown in Figure 3. As can be seen, it is not necessary to examine a larger range for the regularisation parameter γ, because it starts stabilising very quickly with a small number of features. Note that the total number of generated features is 795 (Section 4.4).

However, since we are considering five different targets and most of the relevant features are not common to all targets, the total number of features to consider for the simultaneous modelling of all targets in multitask learning mode increases significantly, in relation to the number of features to consider in single task learning. For this reason, experiments with two selections of features were conducted: using linear regression with

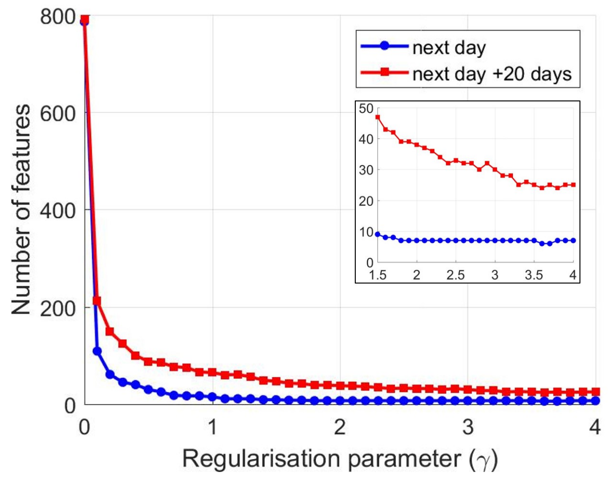


Figure 3: Change in the number of features as a function of the linear regression's regularisation parameter (γ), for target 30-year bond yield. Inside chart: zoom in range γ = 1.5 to 4.

γ equal to 2 and 4. The latter value of the regularisation parameter further reduces the number of relevant features to consider in the models. In fact, despite the stabilisation trend shown in the plot (Figure 3), the number of relevant features continues to decrease with γ, in particular for predictions further away in the future. The results presented henceforth refer to the feature selection with $\gamma = 4$ (results with $\gamma = 2$ are not included in this paper because they do not provide any additional information to the main findings). Comparatively to those obtained with $\gamma = 2$, it leads to better results, with lower spread, mainly for the longer maturities considered (5, 10 and 30 years). The much higher number of features using $\gamma = 2$ tend to result in some overfitting of the neural network model.

An analysis of the feature selection results reveals that the relevant features depend on both the target yield to predict and the forecasting horizon. Table 2 shows the top relevant features per target, selected by weight above 0.01 and when they remain relevant in at least 4 of the 5 forecasting horizons studied.

For all targets, there is a dominant feature which is the last value of the target to predict. This is an expected result, since the last value should reflect all information available to the markets. This strong dominance is clear for the one step ahead forecasting but rapidly diminishes as the forecasting horizon increases and additional features are included in the model. Apart from this dominant feature, additional relevant features tend to come from assets with the same or adjacent type of maturity.

Now we will analyse the relevant features across target yields and across forecasting horizons. On the one hand, for a specific target, the number of features increases with the forecasting horizon (see Table 3). Most of the relevant features for one step ahead predictions remain relevant for forecasting more distant future values, but additional features are required for those more distant and more complex predictions. On the other hand, considering a particular forecasting horizon, each target yield tends to have a specific set of features. Adjacent targets may in some cases have some equal relevant features, but rarely does a specific feature remain relevant across the yield curve for all targets. The last row of Table 3 shows that the number of relevant features necessary to model all targets simultaneously (MTL) increases continuously with the forecasting horizon, from 31 (0 days) to 90 (20 days).

Table 2: Top relevant features per target, considering only those with weights above 0.01 and when they remain relevant in at least 4 of the 5 forecasting horizon studied. Dominant feature in bold.

ID	Feature Name [ticker] time step
3M	
4	Interest Rate Overnight [EUDR1T] t-1
5	Interest Rate Overnight [EUDR1T] t
772	Euro Generic Govt 3 Month Yield [GECU3M] t-3
773	Euro Generic Govt 3 Month Yield [GECU3M] t-2
775	**Euro Generic Govt 3 Month Yield [GECU3M] t**
780	Euro Generic Govt 2 Year Yield [GECU2YR] t
2Y	
45	Generic 2nd 3M Euribor Future [ER2] t
275	Equities Euro Stoxx 50 Index [SX5E] t
780	**Euro Generic Govt 2 Year Yield [GECU2YR] t**
5Y	
200	Bond Future Europe 2 Year Yield [DU1] t
291	Equities Tokyo Topix Index [TPX] t-4
785	**Euro Generic Govt 5 Year Yield [GECU5YR] t**
10Y	
210	Bond Future Europe 10 Year Yield [RX1] t
230	Swaps rate 10 Year [EUSA10] t
785	Euro Generic Govt 5 Year Yield [GECU5YR] t
790	**Euro Generic Govt 10 Year Yield [GECU10YR] t**
30Y	
210	Bond Future Europe 10 Year Yield [RX1] t
215	Bond Future Europe 30 Year Yield [UB1] t
235	Swaps rate 30 Year [EUSA30] t
356	Commodities Corn [C 1] t-4
795	**Euro Generic Govt 30 Year Yield [GECU30YR] t**

Table 3: Number of relevant features per yield, per forecasting horizon and in MTL mode (simultaneous modelling of all yields).

Yield	**Forecasting horizon (days)**				
	0	5	10	15	20
3M	11	22	23	29	36
2Y	5	18	19	18	23
5Y	8	11	17	17	25
10Y	5	13	17	22	25
30Y	7	11	20	21	25
MTL	31	58	71	76	90

6.2. *Comparison of models*

In this section, modelling results are presented, discussed in depth and finally compared with other results in the literature.

6.2.1. *Introduction*

Results from a direct comparison of the multilayer perceptron model using all generated features (Model 1, Table 1) versus MLP with relevant features (Model 2), demonstrated the clear advantage of performing an initial feature selection. The main advantages are twofold: better forecasting (lower errors and lower spread) and lighter models with lower number of features meaning less computing time. For this reason, the results presented in Figure 4 exclude Model 1.

The results are presented using the normalised metric since the non-normalised equivalents are scale dependent and, consequently, depend on the period we are analysing and the level of yield at that particular period. Hence, the normalised metric is used to facilitate the comparison of models in the literature. However, in order to enable a point of comparison between normalised and non-normalised results, an example is presented for the 10-year yield in Table 4. As can be seen the difference between normalised and non-normalised is not substantial given the range and level of 10-year yields analysed.

Table 4: Forecasting errors for 10Y yield (model: multilayer perceptron using relevant features; forecasting horizon: next day).

Error	**Normalised**		**Non-normalised**	
	Mean	Std Dev	Mean	Std Dev
MAE	0.03190	0.00194	0.03340	0.00185
MSE	0.00206	0.00021	0.00226	0.00018
RMSE	0.04529	0.00229	0.04754	0.00189

6.2.2. *Multilayer perceptron models*

The one step ahead forecasting, shown in Figure 4a, tends to produce results of the same magnitude for all models considered, with no significant difference between them. All results in this figure are presented in terms of normalised MSE. To visualise what this represents in real yields and give the reader an idea of the models' forecasting capability, a scatter plot of actual versus predicted yield is shown in Figure 5. This is presented as an example of results for 10-year yield and next day forecasting horizon.

Analysing the baseline linear regression model, the results for next day forecasting were surprisingly good (Figure 4a). However, we need to stress that linear regression and neural network models followed exactly the same procedure in what concerns: feature selection per target and per forecasting horizon and retraining of models at every time step (as discussed in Sections 5.2 and 5.3). As a result, this model is a more difficult benchmark to beat, especially for forecasting the next day.

When we move from forecasting one step ahead to forecasting further into the future, the error increases as expected, as can be seen in Figures 4b to 4e. More importantly, the outperformance of neural networks becomes more evident in relation to linear regression. Additionally, the model using relevant features (Model 2, Table 1), starts outperforming the model using only past values of the target yield to predict (Model 3), especially for higher forecasting horizons and longer maturities (10 and 30-year bonds). The latter results, demonstrate the importance of incorporating features from markets and economy in the models.

Globally, considering all models and forecasting horizons studied, the MLP using relevant features achieves the best overall results for yield forecasting. Given the moving window used and the continuous retraining of the model, it has also the advantage of being more flexible to potential changes in market regime in the future.

6.2.3. *Additional models with synthetic data*

Another important observation is that, despite the simplicity of linear regression, the neural network models including synthetic data generated by this model tend to improve results. Once again, this effect is more pronounced for higher forecasting horizons and longer maturities (5, 10 and 30 years). This is a promising result, showing the potential for the development of hybrid models using synthetic data from other models.

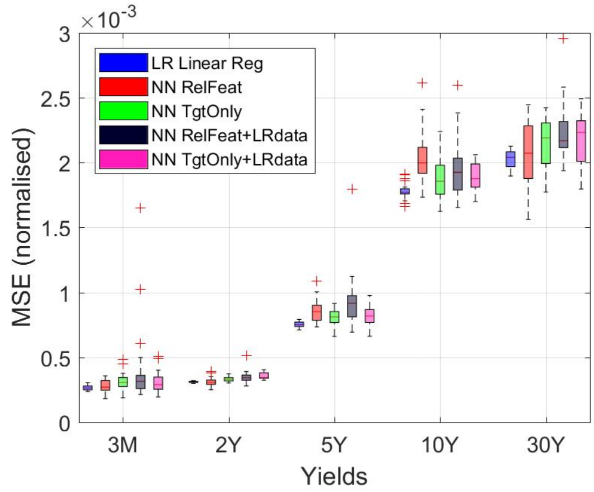
6.2.4. *Single task versus multitask learning*

Regarding the comparison of single task with multitask learning, using both yields as targets (multi-assets analysis) and forecasting horizon as targets (multi-step analysis), no clear differentiation among those two techniques could be demonstrated. As a result, we present only a few examples, shown in Figure 6.

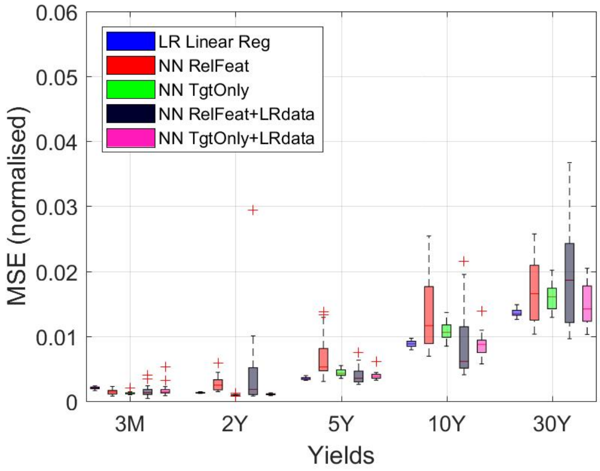
Given that the literature highlighted numerous benefits of simultaneous modelling of targets in multitask learning mode on a wide range of applications, this lack of differentiation was somehow unexpected. In fact, in our research study we have also compared these two techniques when models are trained with a fixed training dataset and no retraining (see Section 5.3) is carried out. Some benefits of multitask learning were observed in this case, which could not be reproduced once retraining was used (Figure 6).

Thus, it is worth reflecting on possible reasons justifying the results obtained. In the neural network model with relevant features, going from a single task learning method to a multitask mode implies the incorporation of all relevant features for each target (in MTL with yields as targets) or all relevant features for each forecasting horizon (in MTL with forecasting horizon as targets). This corresponds to a very significant increase in the number of features the model has to deal with, which may not be best for generalisation, i.e. performance outside the known training data.

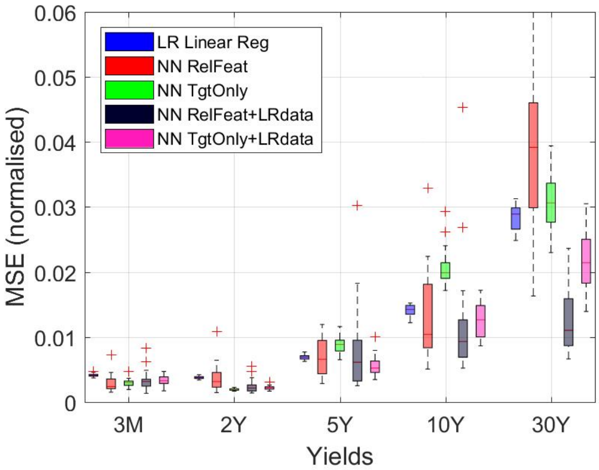
Other possible reasons for the lack of performance of MTL may be obtained from a recent study by Ciliberto et al. (2015). The authors reported the benefits of using multitask learning, but also concluded that its advantage decreases as the amount of training examples increases. From both mean and standard deviation of errors presented in the same research study, it can be concluded that the performance improvement from MTL is not substantial. Also, in the cases where the benefit of multitask learning is higher, the overall performance of models is



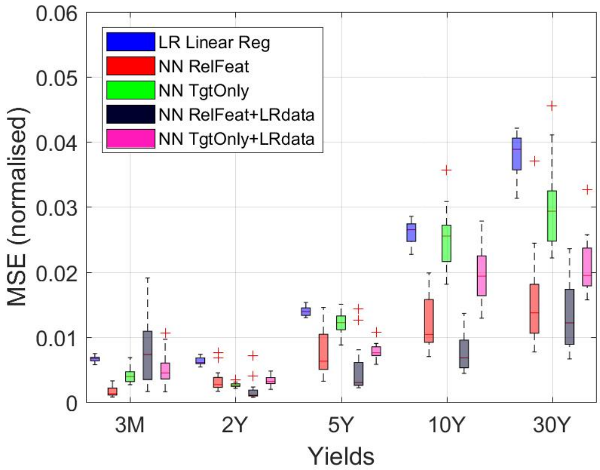
(a) Forecasting horizon = 0 days (next day).



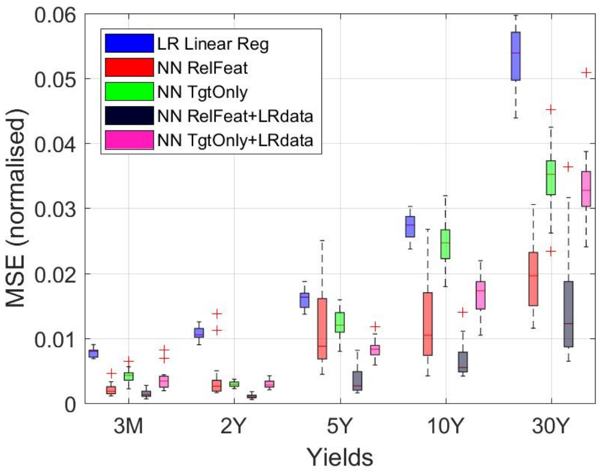
(b) Forecasting horizon = 5 days.



(c) Forecasting horizon = 10 days.



(d) Forecasting horizon = 15 days.



(e) Forecasting horizon = 20 days.

Figure 4: Comparison of models: linear regression (LR Linear Reg); multilayer perceptron using relevant features per target and per forecasting horizon (NN RelFeat); multilayer perceptron using only past values of the target(s) to predict (NN TgtOnly); and the last two models with synthetic data from the linear regression model as additional feature (NN RelFeat+LRdata and NN TgtOnly+LRdata, respectively). In all cases: neural network (NN) models with 10 hidden units and feature selection with regularisation parameter γ equal to 4.

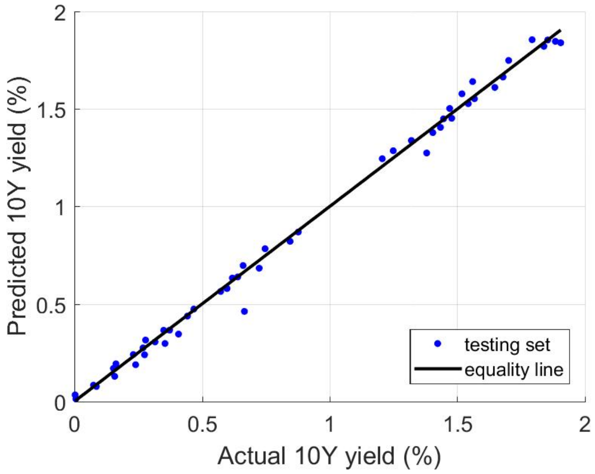
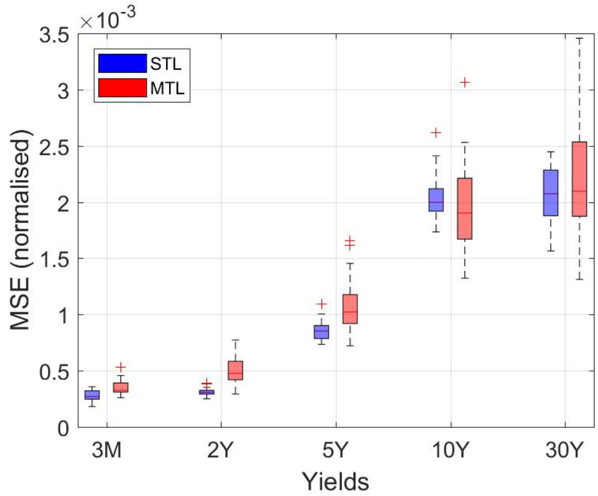
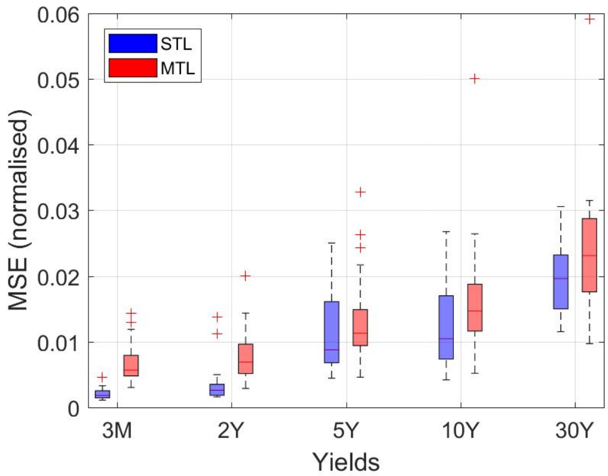


Figure 5: Forecasting results for 10Y yield (model: multilayer perceptron using relevant features; forecasting horizon: next day).



(a) Forecasting horizon = 0 days (next day).



(b) Forecasting horizon = 20 days.

Figure 6: Example of single task versus multitask learning for the multilayer perceptron model using relevant features per target and per forecasting horizon. In both cases: neural network models with 10 hidden units and feature selection with regularisation parameter γ equal to 4.

comparatively poor, with the amount of data probably also having an important effect. The benefits of MTL with limited data have also been reported by Benton et al. (2017). In the research reported in this paper, the amount of data collected was large considering both the overall period and the amount of observations available.

In summary, there are several factors that may have contributed to the lack of differentiation between single task and multitask learning: the large amount of data used for training the models; the optimisation of models by using the relevant features per target and per forecasting horizon together with full retraining of models at every time step; the large increase in the number of features as we consider MTL; and the fact that the relevant features tend to be different for all targets. There is an advantage with multitask learning, which is the running of only one model for all targets instead of five. However, this is not as relevant nowadays, given the modern computing capabilities of super-computers and GPU computing.

6.2.5. Comparison with results in the literature

A comparison of results with the ones available in the literature is difficult given the scarce number of studies having some type of overlapping with this study. Notwithstanding, some common ground can be found in Castellani and Santos (2006), Dunis and Morrison (2007) and Sambasivan and Das (2017), covered in Section 2.2. A direct comparison of results can only be achieved by using exactly the same data for the models being compared, which falls outside the objectives of our research. Bearing in mind the limitations, having an indicative comparison of the magnitude of errors would be important in this type of empirical work.

In more detail, in Castellani and Santos (2006), monthly data was used for forecasting the US 10-year yield, and the best models achieved levels of accuracy only marginally better than forecasting using the last available value. The closest situation in this study would be the comparison with results obtained for a forecasting horizon of 20 (working) days, approximately a calendar month. However, given the different type of data frequency used (monthly versus daily), the comparison is done on a qualitative basis only, emphasising the fact that in this study all models led to results significantly better than using the last available value.

A closer comparison can be attempted with Dunis and Morrison (2007). The authors used daily data for next-day forecasting 10-year yields (German, UK and US), using a feedforward neural network with one hidden layer and five hidden units, among other models. This may be compared with our results shown in Table 4. The results presented in this paper compare favourably in all cases, being of the same magnitude as the best results obtained in that study, achieved in the case of the UK yields. Main limitations of this comparison are due to the different dataset used, both in terms of period analysed and features considered.

Finally, a comparison of results with the work carried out by Sambasivan and Das (2017). The dataset used in our research (January 1999 to April 2017) fully includes the period considered in the above mentioned study (February 2006 to February 2017). In this case, we have to take into account that only one step ahead forecasting was implemented. Taking all this into consideration, the results presented in this paper are significantly better for all target yields considered. In our research additional information from macroeconomic and market features was included, as well as a more extended period for the datasets, totalling over 18 years of data.

7. Conclusions and future work

The literature review revealed two main gaps and the objective of our research is to fill those gaps. The first concerns the limited amount of publications on the use of machine learning techniques applied specifically to fixed income markets. Other financial areas have been the target of much higher levels of interest from the research community, such as equities and foreign exchange markets.

The second gap is more specific and relates to the lack of direct solutions in the literature for modelling the yield curve as a whole, specifically using machine learning models. For this purpose, multitask learning has been identified as a possible solution to be evaluated in our research and it represents a novel application. It can be performed either via the consideration of multiple targets used simultaneously in the same model with parallel learning process, or via the transformation of the problem into multiple STL models.

In our research, a number of models were used for forecasting the main benchmarks of the European yield curve. Using the most relevant features for each target and for each forecasting horizon, the multilayer perceptron achieved overall higher levels of accuracy, when compared to the linear regression and MLP using only past values of the target variable. In addition, the model considers a moving window of training data to incorporate most recent information as it becomes available, and the retraining of models at every time step. Consequently, this model is more flexible to changing market conditions and its outperformance becomes more evident for forecasting further ahead into the future. The results presented also compare positively with the scarce error data found in the literature for the same specific purpose of yield forecasting using machine learning techniques.

On the linear regression model, despite its simplicity, it is already a tough benchmark to beat for one step ahead forecasting, given the dominant characteristic of the last available value of the variable to predict. Additionally, by including synthetic data generated by this model, the neural network models tend to improve results. This gives excellent indications about the possibility of using hybrid models incorporating data generated by industry-established models as additional features.

Furthermore, the importance of an adequate feature selection, through fine tuning the regularisation parameter γ, was emphasised from the results, leading to better results with lower standard deviations, thus reflecting less overfitting of the models. Additional findings concerning feature selection, in particular that different targets and forecasting horizons lead to different relevant features, demonstrate that the "one set of features fits all" methodology does not work for forecasting bond yields. It also reinforces the importance of custom-built models, taking into account the specific targets and conditions of the problem.

On the comparison of the two multitask learning techniques used for yield curve forecasting, no clear differentiation could be demonstrated. Several factors were pointed out that could justify these results, which are supported by previous studies found in the available literature, namely: large amount of training data; full retraining of models at every time step; large increase in the number of features as we consider MTL; and the fact that relevant features tend to be different for all targets.

With respect to future work, additional research on multitask learning is needed to identify the conditions under which this methodology can be used with improved performance. It is also necessary to develop a theory behind multitask learning and techniques to actually select extra tasks or targets that would benefit from a simultaneous modelling multitask learning.

Acknowledgements

This work was supported by the UK Engineering and Physical Sciences Research Council, Doctoral Training Partnership with the University of Southampton.

References

Agrawal, J., Chourasia, V., Mittra, A., 2013. State-of-the-art in stock prediction techniques. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2 (4), 1360–1366.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., Iglesias, C. A., 2017. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications 77, 236–246.

Arrieta-Ibarra, I., Lobato, I. N., 2015. Testing for predictability in financial returns using statistical learning procedures. Journal of Time Series Analysis 36 (5), 672–686.

Ballings, M., Van den Poel, D., Hespeels, N., Gryp, R., 2015. Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications 42 (20), 7046–7056.

Becker, B., Ivashina, V., 2015. Reaching for yield in the bond market. The Journal of Finance 70 (5), 1863–1902.

Benton, A., Mitchell, M., Hovy, D., 2017. Multitask learning for mental health conditions with limited social media data. In: European Chapter of the Association for Computational Linguistics, EACL. Vol. 1. pp. 152–162.

Bishop, C. M., 2006. Pattern recognition and machine learning (Information Science and Statistics). Springer-Verlag New York.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of political economy 81 (3), 637–654.

Bloomberg, 2017. Bloomberg Professional Database | Subscription Service.

Booth, A., Gerding, E., Mcgroarty, F., 2014a. Automated trading with performance weighted random forests and seasonality. Expert Systems with Applications 41 (8), 3651–3661.

Booth, A., Gerding, E., McGroarty, F., 2014b. Predicting equity market price impact with performance weighted ensembles of random forests. In: IEEE Conference on Computational Intelligence for Financial Engineering & Economics, CIFEr. IEEE, pp. 286–293.

Booth, A., Gerding, E., McGroarty, F., 2015. Performance-weighted ensembles of random forests for predicting price impact. Quantitative Finance 15 (11), 1823–1835.

Borchani, H., Varando, G., Bielza, C., Larrañaga, P., 2015. A survey on multi-output regression. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 5 (5), 216–233.

Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., 2015. Time series analysis: forecasting and control, 5th edition. Wiley Series in Probability and Statistics. Wiley.

Box, G. E. P., Jenkins, G. M., 1968. Some recent advances in forecasting and control. Journal of the Royal Statistical Society. Series C (Applied Statistics) 17 (2), 91–109.

Burton, K., 2016. Inside a moneymaking machine like no other. Bloomberg.

Cai, H., Huang, Z., Zhu, X., Zhang, Q., Li, X., 2014. Multi-output regression with tag correlation analysis for effective image tagging. In: International Conference on Database Systems for Advanced Applications, DASFAA. Springer, pp. 31–46.

Caldeira, J., Torrent, H., 2017. Forecasting the US term structure of interest rates using nonparametric functional data analysis. Journal of Forecasting 36 (1), 56–73.

Caruana, R. A., 1993. Multitask learning: A knowledge-based source of inductive bias. In: International Conference on Machine Learning. pp. 41–48.

Caruana, R. A., 1997. Multitask learning. Machine Learning 28 (1), 41–75.

Castellani, M., Santos, E. A. d., 2006. Forecasting long-term government bond yields: an application of statistical and AI models. ISEG, Departamento de Economia.

Choudhry, T., McGroarty, F., Peng, K., Wang, S., 2009. Artificial neural network and high frequency exchange rate prediction. Forecasting Financial Markets Conference, FFM.

Choudhry, T., McGroarty, F., Peng, K., Wang, S., 2012. High-frequency exchange-rate prediction with an artificial neural network. Intelligent Systems in Accounting, Finance and Management 19 (3), 170–178.

Ciliberto, C., Mroueh, Y., Poggio, T., Rosasco, L., 2015. Convex learning of multiple tasks and their structure. In: International Conference on Machine Learning. Vol. 37. pp. 1548–1557.

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of econometrics 130 (2), 337–364.

Diebold, F. X., Rudebusch, G. D., 2013. Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach. Econometric and Tinbergen Institutes lectures. Princeton University Press.

Diebold, F. X., Rudebusch, G. D., Aruoba, S. B., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of econometrics 131 (1), 309–338.

Dunis, C. L., Middleton, P. W., Karathanasopolous, A., Theofilatos, K., 2016. Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics. New Developments in Quantitative Trading and Investment. Palgrave Macmillan UK.

Dunis, C. L., Morrison, V., 2007. The economic value of advanced time series methods for modelling and trading 10-year government bonds. European Journal of Finance 13 (4), 333–352.

Eilers, D., Dunis, C. L., Mettenheim, H.-J. v., Breitner, M. H., 2014. Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. Decision support systems 64, 100–108.

Enders, W., 2014. Applied Econometric Time Series, 4th Edition. Wiley Series in Probability and Statistics. Wiley.

Fletcher, T., 2012. Machine learning for financial market prediction. Ph.D. thesis, University College London.

Fletcher, T., Shawe-Taylor, J., 2013. Multiple kernel learning with fisher kernels for high frequency currency prediction. Computational Economics 42 (2), 217–240.

FRM, 2002. Medallion International Ltd. investment analysis. Tech. rep., Financial Risk Management.

Ghosn, J., Bengio, Y., 1997. Multi-task learning for stock selection. Advances in Neural Information Processing Systems, 946–952.

Gogas, P., Papadimitriou, T., Matthaiou, M., Chrysanthidou, E., 2015. Yield curve and recession forecasting in a machine learning framework. Computational Economics 45 (4), 635–645.

Gradojevic, N., Yang, J., 2006. Non-linear, non-parametric, non-fundamental exchange rate forecasting. Journal of Forecasting 25 (4), 227–245.

Hamilton, J. D., 1994. Time series analysis. Princeton University Press, Princeton, New Jersey.

Hastie, T., Tibshirani, R., Friedman, J., 2013. The elements of statistical learning. Data mining, inference, and prediction, 2nd Edition. Springer Series in Statistics.

Huang, W., Lai, K. K., Nakamori, Y., Wang, S., Yu, L., 2007. Neural networks in finance and economics forecasting. International Journal of Information Technology & Decision Making 6 (01), 113–140.

Hutchinson, J. M., Lo, A. W., Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. The Journal of Finance 49 (3), 851–889.

Kanevski, M., Maignan, M., Pozdnoukhov, A., Timonin, V., 2008. Interest rates mapping. Physica A: Statistical Mechanics and its Applications 387 (15), 3897–3903.

Kanevski, M., Timonin, V., 2010. Machine learning analysis and modeling of interest rate curves. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN.

Kolanovic, M., Krishnamachari, R. T., 2017. Big data and AI strategies: machine learning and alternative data approach to investing. J.P. Morgan Quantitative and Derivatives Strategy Report.

Kräussl, R., Lehnert, T., Rinne, K., 2017. The search for yield: Implications to alternative investments. Journal of Empirical Finance 44, 227–236.

Mahler, N., 2009. Modeling the S&P 500 index using the Kalman filter and the LagLasso. In: IEEE International Workshop on Machine Learning for Signal Processing, MLSP. IEEE, pp. 1–6.

Mello, M. A. d. C. d. S. e., 2015. Search-for-yield in Portuguese fixed-income mutual funds and monetary policy. Master's thesis, Nova School of Business and Economics.

Mettenheim, H.-J. H. v., 2010. Advanced neural networks: finance, forecast and other applications. Ph.D. thesis, Gottfried Wilhelm Leibniz Universität Hannover.

Mettenheim, H.-J. H. v., Breitner, M. H., 2010. Robust decision support systems with matrix forecasts and shared layer perceptrons for finance and other applications. In: International Conference on Information Systems. ICIS, p. 83.

Mettenheim, H.-J. H. v., Breitner, M. H., 2011. Forecasting complex systems with shared layer perceptrons. In: Operations Research Proceedings 2010. Springer, pp. 15–20.

Montesdeoca, L., Niranjan, M., 2016. Extending the feature set of a data-driven artificial neural network model of pricing financial option. In: IEEE Symposium Series on Computational Intelligence, SSCI. IEEE, pp. 1–6.

Morell, J. V. T., 2017. The decline in the predictive power of the US term spread: a structural interpretation. Journal of Macroeconomics 55, 314–331.

Nelson, C. R., Siegel, A. F., 1987. Parsimonious modeling of yield curves. Journal of business, 473–489.

Niranjan, M., 1996. Sequential tracking in pricing financial options using model based and neural network approaches. In: Neural Information Processing Systems, NIPS. pp. 960–966.

Niranjan, M., 2016. Machine learning. Electronics and Computer Science, University of Southampton.

Nunes, M., Gerding, E., McGroarty, F., Niranjan, M., 2018. Dataset information for "Artificial neural networks in fixed income markets for yield curve forecasting". University of Southampton repository. DOI to be registered: https://doi.org/10.5258/SOTON/D0450 [Dataset].

OECD, 2015a. OECD business and finance outlook. OECD Publishing, Paris.

OECD, 2015b. Pension markets in focus. OECD Publishing, Paris.

Roux, P., Burton, K., 2017. This hedge fund may be poised to create the most billionaires. Bloomberg.

Ruder, S., 2017. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098v1.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning internal representations by error propagation. In: D. E. Rumelhart & J. L. McClelland (Eds.), Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1. MIT Press, pp. 318–362.

Sambasivan, R., Das, S., 2017. A statistical machine learning approach to yield curve forecasting. arXiv preprint arXiv:1703.01536v1.

Sewell, M., 2011. Characterization of financial time series. UCL Department of Computer Science, Research Note RN/11/01.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.

Vui, C. S., Soon, G. K., On, C. K., Alfred, R., Anthony, P., 2013. A review of stock market prediction with artificial neural network (ANN). In: IEEE International Conference on Control System, Computing and Engineering, ICCSCE. IEEE, pp. 477–482.