Forecasting Government Bond Spreads with Heuristic Models: Evidence from the Eurozone Periphery

Fernandes, FDS, Stasinakis, C & Zekaite, Z

Original citation & hyperlink:
Fernandes, FDS, Stasinakis, C & Zekaite, Z 2018, 'Forecasting Government Bond Spreads with Heuristic Models: Evidence from the Eurozone Periphery' *Annals of Operations Research*, vol (in press), pp. (in press)
https://dx.doi.org/10.1007/s10479-018-2808-0

DOI 10.1007/s10479-018-2808-0
ISSN 0254-5330
ESSN 1572-9338

Publisher: Springer

The final publication is available at Springer via http://dx.doi.org/10.1007/s10479-018-2808-0

Forecasting Government Bond Spreads with Heuristic Models: Evidence from the Eurozone Periphery

by

Filipa Da Silva Fernandes*

Charalampos Stasinakis **

Zivile Zekaite***

November 2017

Abstract

This study investigates the predictability of European long-term government bond spreads through the application of heuristic and metaheuristic Support Vector Regression (SVR) hybrid structures. Genetic, krill herd and sine-cosine algorithms are applied to the parameterization process of the SVR and Locally weighted SVR (LSVR) methods. The inputs of the SVR models are selected from a large pool of linear and non-linear individual predictors. The statistical performance of the main models is evaluated against a Random Walk (RW), an Autoregressive Moving Average (ARMA), the best individual prediction model and the traditional SVR and LSVR structures. All models are applied to forecast daily and weekly government bond spreads of Greece, Ireland, Italy, Portugal and Spain over the sample period 2000-2017. The results show that the sine-cosine LSVR is outperforming its counterparts in terms of statistical accuracy, while metaheuristic approaches seem to benefit the parameterization process more than the heuristic ones.

Keywords

Government bond spreads, Eurozone, Support Vector Regression, Krill Herd, Sine Cosine Algorithm

* School of Economics, Finance and Accounting, Coventry University, Priory Street, Coventry CV1 5FB, United Kingdom (filipa.dasilvafernandes@coventry.ac.uk)

**University of Glasgow Business School, University of Glasgow, Gilbert Scott Building, Glasgow G12 8QQ, United Kingdom (Charalampos.Stasinakis@glasgow.ac.uk)

*** University of Glasgow Business School, University of Glasgow, Gilbert Scott Building, Glasgow G12 8QQ, United Kingdom (zivile.zekaite@gmail.com)

1. Introduction

The introduction of the euro in January of 1999 has led to the elimination of the exchange rate risk premium and the beginning of common monetary policy with a clearly defined objective of the price stability that resulted in firmly anchored inflation expectations within the Economic and Monetary Union (EMU) (ECB, 2010). Consequently, it facilitated the financial integration process among the member countries, the degree of which could be measured by government bond yield differentials, i.e. yield spreads. For a given maturity, government yield spreads are typically calculated vis-a-vis yields of a country whose debt is viewed as very liquid and having low credit risk.[1] Hence, sovereign spreads are also good indicators of relative financing conditions of a specific country. Since the launch of the euro and up until the Global Financial Crisis (GFC) in 2007-2009, government yield spreads within the EMU stabilized at lower levels than before the EMU and also showed the tendency to co-move, indicating successful financial integration (Georgoutsos and Migiakis, 2012). In the absence of the exchange rate risk, yield spreads of individual countries were mainly driven by the credit risk and liquidity risk premia perceived to be relatively low (Aristei and Martelli, 2014; Dewachter et al., 2015).

Following the GFC and the subsequent Sovereign Debt Crisis in the Eurozone (SDC), government bond market developments in the currency bloc received unprecedented attention from policy makers, financial market participants, economists, academics and the public. Interest rates on public debt rose and diverged across the Eurozone with yield differentials widening sharply in some countries experiencing deteriorating fiscal positions and weak economic fundamentals. The sovereign bond market stress was likely amplified through the spill-over effects between periphery and core countries (Antonakakis and Vergos, 2013). For some countries, the exchange rate risk may have reappeared pushing yields upward due to speculations about the exit from the euro by those countries (Favero, 2013). Another related feature of this period was the divergence in retail bank lending rates across the Eurozone despite common accommodative monetary policy. As Darracq-Paries et al. (2014) show, the wedge between the private borrowing costs was to some extent driven by the fragmentation in bank funding costs across the currency bloc as a result of government bond market distress. Interest rates on public debt, which is perceived as relatively risk-free, are considered as the benchmark for the borrowing costs of the private sector. Consequently, higher yields on government bonds imply higher funding costs for households, financial and non-financial corporations. Thus, well-functioning bond markets are crucial for the transmission mechanism of monetary policy to real economy. On the other hand, heightened financial market stress may weaken this transmission if the link between the policy rate and market benchmark interest rates, upon which bank lending rates are typically based, breaks down (ECB, 2010; Darracq-Paries et al., 2014). Since the sovereign debt crisis, the European Central Bank has acknowledged the impact of sovereign bond market stress on the effectiveness of its policy and has

[1] In the case of the Eurozone, the German Bunds are commonly used as such a benchmark. For instance, see Aristei and Martelli (2014).

adopted a wide set of policies to enhance the policy transmission across the Eurozone in order to achieve its mandate (Lautenschläger, 2017).

It is clear from the above that government debt securities play an important role in the financial market. The Eurozone bond yields and yield spreads as well as their dynamics provide investors, policy makers, economists and academics with valuable information about the cost of public and private financing and credit risk within the EMU, the effectiveness of common monetary policy and the degree of financial integration among the countries in the Eurozone. Although several studies attempt to model government spreads in an effort to understand the underlying factors driving their erratic behaviour (Duffie et al. 2003; Gómez-Puig 2009; Manganelli and Wolswijk 2009; Paniagua et al., 2016; Leschinski and Bertram, 2017), not many researchers actually attempt to forecast the future levels of yield spreads. Such limited examples are the works of Diebold and Li (2006) and Favero (2013).

This gap in the literature is partially explained by the fact that European government bond spreads are characterized by non-linearities and high volatility. Long-term government bond spreads between euro-area countries and Germany are known to follow unstable patterns through time (Abad et al. 2010).Traditional linear and non-linear statistical models, despite their widespread use in macroeconomic forecasting applications, struggle to capture efficiently the nature of such data. As a consequence, practitioners, academic researchers and policy makers turn to more complex optimization techniques in their effort to overcome the ineffectiveness of easy-to-implement models. This explains why the forecasting literature is voluminous, when it comes to advanced prediction methodologies following the principles of machine learning, kernel-based optimizations and neural network structures (Kaastra and Boy 1996; Gilli and Schumann 2012; Patel et al. 2015).

The purpose of this study is to perform an empirical analysis that evaluates and compares the forecasting power of hybrid heuristic-SVR models applied to 10-year Government Bond Spreads (10YGBS) of the selected Eurozone countries. The main contribution of the paper is that we explore the utility of Krill Herd and Sine Cosine algorithms in the parameterization of the Support Vector Regression (SVRs) and Locally Weighted SVR counterparts. To the best of our knowledge, such a forecasting study has not been performed before. Additionally, we contribute in the financial literature of heuristics, as no other studies have investigated the robustness of the KH and SC SVR structures in financial and/or economic applications, while none has explored the potential benefits of combining SC with LSVR. Finally, this paper adds to the empirical literature on forecasting bond spreads (Diebold and Li 2006; Favero, 2013). The performance of KH and SC hybrid SVRs is benchmarked against the traditional SVR and LSVR models. All models are applied to forecast the 10YGBS of Greece, Ireland, Italy, Portugal and Spain (GIIPS) on a daily and weekly basis over the period 2000-2017. Therefore, the focus of our analysis is on the Eurozone periphery spreads that have presented a volatile and erratic behaviour during the SDC. Favero (2013) suggests that co-movements of yields in the Euro area imply either the impact of credit risk or a strong relationship between credit-risk and liquidity risk. For that reason, it is crucial to investigate the cases of GIIPS suffering more from the credit and liquidity tightening conditions observed especially after 2009. To make the statistical analysis more robust, the sample period is separated into two parts: 2000-2008 and 2008-2017. The first period covers the GFC, while the second the Eurozone SDC. In total, we carry out four forecasting exercises for each country. In terms of the results, the SC algorithm is found as the best prediction model of this study for all countries and periods considered. The SC-LSVR hybrid model outperforms the KH counterparts, while the evolutionary heuristics seems to be the less efficient in terms of forecasting accuracy. The forecasting results also show that yields appear more difficult to be forecasted within the SDC than the GFC, while models are performing generally worse when it comes to forecasting the 10YGBS of Greece, Ireland and Portugal. The rest of the paper is organized as follows. Section 2 is a brief literature review on heuristic and SVR techniques. The dataset and forecasting exercises are described in Section 3. The theoretical background of SVR and LSVR is outlined in Section 4. Section 5 includes the description of the heuristic models employed in this study along with the explanation of how their inputs and benchmarks are selected. The statistical performance of all models is analysed in Section 6. Section 7 includes some concluding remarks. Finally, all the information regarding the input set and technical characteristics of the model under study are summarized in the Appendix.

2. Brief Literature Review

Heuristic and metaheuristic optimization techniques have recently become a powerful tool in the hands of practitioners (Gilli et al. 2008). Several studies are focused on heuristics based on evolutionary principles (Fonseca and Fleming 1995; Alcaraz and Maroto 2001; Ahn and Kim 2008; Aguilar-Rivera et al. 2015). The main approach in such studies is the application of Genetic Algorithms (GAs) as proposed by Holland (1975). Although GAs apply the Darwinian principles of survival and reproduction of the fittest successfully, several GA based models suffer from local optima constraints during the search space. Metaheuristics are considered a superior class of heuristics because of their ability to avoid local optima trapping and over-fitting, while keeping the demands of computational time relatively low (Van Breedam 2001). Many researchers postulate that such algorithms are able to create the balance between the local and the global search space, by identifying potential suitable random variables that are not dependent on the problem under study (Talbi 2009). There are several trends in the metaheuristic modelling literature. The most prominent one is related to the nature-inspired metaheuristic approaches that are motivated by the evolution of species or their swarm movement behavior. Such examples are algorithms approximating ant and bee colonies movements (Dorigo et al. 2006; Karaboga and Akay 2009), firefly behavior (Yang 2010), bat flying (Yang and Gandomi 2012) and animal immigration patterns (Li et al 2014). Eventually, practitioners need to make a decision on screening for some robust heuristics in order to proceed with the parameterization and calibration of their forecasting models. In this study, apart from the traditional evolutionary GA heuristic approach, we focus on the performance of two other metaheuristics.

The first one is the KH, which is a successful metaheuristic method proposed by Gandomi and Alavi (2012). The algorithm is motivated by the behavior of ocean krill in herds. The intuition of KH is to incorporate the mean-reverse effect that predator attacks can inflict on the krill herd density into optimizing the future positions and eventually regrouping behavior of dispersed krill. With the KH approach, every krill's position is approximated through a time-dependent function that is based on three different motions, namely, the movement induced by other individuals, the foraging motion and random physical diffusion. According to Gandomi and Alavi (2012) and Wang et al. (2014), KH is found superior to previously mentioned metaheuristic models, due to the avoidance of a gradient search and expectation of fine-tuning of only one parameter. Unlike the first one, the second metaheuristic is not nature-based. Recently, Mirjalili (2016) proposed the Sine Cosine (SC) algorithm that is based on mathematical objective functions rather than bio-inspired ones. SC is a population-based optimization algorithm that is able to generate multiple random candidate solutions and orientate them towards the best solution through a mathematical approach of sine and cosine functions. These mathematical functions are infused with a set of random adaptive variables that optimize the balance between exploration and exploitation of the search space, unlike in the case of KH. This increases the probability of reaching the global optimum solution faster by exploring the most promising regions of the search space.

SVRs are regression-based models used to explore the non-linear and data-adaptive dynamics of financial time series (Vapnik 1995). They have been successfully applied in numerous financial forecasting applications. For example, Lu et al. (2009) suggest that SVRs' statistical performance is better than that of traditional random walks when forecasting daily stock prices. Wang and Zhu (2010) proposed a two-step kernel SVR for forecasting the S&P500 and NASDAQ indices with promising results. The LSVR is another class of SVRs that has proven in cases to be superior to the traditional SVR structures. The LSVR imposes penalties to past data and assumes that recent observations are more important. This is a concept suitable to the nature of financial time series. The applications of LSVR in the literature are not so extensive, however. Such examples are the works of Yang et al. (2009) and Wu and Akbarov (2011) who apply successfully the LSVR to forecast financial data and warranty claims.

The main disadvantage of the SVR is that its performance is sensitive to the parameterization process. This is where heuristic and metaheuristic models, such as GA, KH and SC, prove to be very useful (Gilli and Schumann 2015). Given that there is no formal theory on how to select optimal SVR parameters, researchers, instead of using the traditional cross-validation methods, resort in comparisons of hybrid heuristic-SVR techniques in different forecasting applications. Many of the above algorithms are successfully applied in the SVRs in order to obtain optimal SVR parameters and to maximize forecasting performance. Yuan (2012) claims that the combination of of GA and SVR provides better forecasts for sales volume than traditional SVR and NN models. Recently, Stasinakis et al. (2016) and Sermpinis et al. (2017) combine KH with SVR and LSVR, respectively, in a forecasting and trading application of exchange traded funds. Their results show that KH SVR optimization is superior to the traditional SVRs and GA-SVR models, while the advantages of LSVR are also validated. Finally, Li et al. (2018) propose a SC-SVR hybrid model that is applied in a financial forecasting competition between different heuristic-SVR techniques. Their findings suggest that SC optimization is superior to all other bio-inspired algorithms.

3. Dataset

In this study we focus on the 10-year Government Bond Spreads (10YGBS) of the five periphery Eurozone countries. More specifically, our empirical analysis is based on forecasting the daily and weekly 10YGBS of Greece, Ireland, Italy, Portugal and Spain over the period 2000-2017. For all the 10YGBS, the benchmark is the German Bund. The data series are the price indices of respective spreads and are summarized in Table 1. The exception is Spain, where we use bond yield spreads as the price index was not available.

****Insert Table 1****

Figure 1 plots the time series of the daily bond spreads for all countries under study over the sample period. The left-hand-side scale denotes the price of a spread index for Greece, Ireland, Italy and Portugal, while the right-hand-side of the figure represents the yield spread for Spain. Over the course of the GFC, the yield spreads in the GIIPS climbed up somewhat, especially so between the end of 2008 and the start of 2009. This increase was later partially reversed only to reappear again at the onset to the SDC that started in Greece in autumn of 2009. The crisis was quick to spread across the Eurozone between 2010 and 2012. During this period, the peripheral countries' government bond spreads went sharply upwards. In particular, bond spreads reached levels that had never been observed over the history of the EMU in 2011-2013. Based on this figure, it seems appropriate to forecast the country-specific 10YGBS for two distinct sample periods, namely, 2000-2008 and 2008-2017.

****Insert Figure 1****

The Jarque-Bera (1980) and Augmented Dickey–Fuller test performed on the 10YGBS series confirms their non-normality and non-stationarity at the 99% confidence interval respectively. For that reason, all series are transformed into daily and weekly series of the rate of returns using the following formula[2]:

$$R_t = \left(BS_t / BS_{t-1} \right) - 1 \quad (1)$$

[2] This is consistent with also the approach of Favero (2013). Note that in the case of Spain it is actually a percentage change in the yield spread rather than the rate of return on the price index as for other countries.

where R_t is the rate of return and BS_t is the 10YGBS value (daily or weekly) at time *t*. The descriptive statistics of the return series obtained for the two periods and frequencies are shown in the following Table 2:

****Insert Table 2****

The Jarque-Bera statistic confirms that the five return series are non-normal at the 99% confidence level, while the Augmented Dickey-Fuller (ADF) reports that the null hypothesis of a unit root is rejected at the 99% statistical level for every 10YGBS. The models under study, their benchmarks and individual predictors (pool of inputs) are applied to forecast the one-day-ahead rate of return ($E(R_t)$) of the five 10YGBS. Each model's performance is consequently evaluated based on four forecasting exercises presented in Table 3.

****Insert Table 3****

From the Table 3 above, it is clear that all models are optimized in-sample and their forecasts will be evaluated out-of-sample. F1 and F2 are spanning over the sample 2000-2008, while F3 and F4 refers to the period 2008-2017. In addition, separating F1 and F3 from F2 and F4, respectively, allows us to check the robustness of our results by employing different data frequency.

4. Theoretical Framework

SVR, as proposed by Vapnik (1995), is a class of Support Vector Machines (SVMs) applied to the principle of structural risk minimization. SVR distinguishes itself from other forecasting methods by combining its ability to project robust non-linear regression models with good generalization ability in previously unseen data. This is achieved consistently by relying only on a subset of the training observations known as the support vectors.

If $\{(x_1,y_1), (x_2,y_2)\ldots, (x_n, y_n)\}$, where $x_i \in X \subseteq R, y_i \in Y \subseteq R, i = 1 \ldots n$ are the training data and *n* the total number of training samples, the general SVR function can be specified as:

$$f(x) = w^T \varphi(x) + b \tag{2}$$

$\varphi(x)$ is the non-linear function that maps the input data vector *x* into a feature space where the training data exhibit linearity. In order to obtain *w* and *b*, the following regularized risk function must be minimized:

$$\left\{ R(C) = C\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(y_i, f(x_i)) + \frac{1}{2}\|w\|^2 ,\ L_\varepsilon(y_i, f(x_i)) = \begin{cases} 0 \ \ if\ |y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon \ \ if\ \ other \end{cases}, \varepsilon \ge 0 \right\} \tag{3}$$

Parameters C and ε are predefined by the practitioner, y_i is the actual value at time i, $f(x_i)$ is the predicted value at the same period and L_ε is ε-sensitive loss function The loss function identifies the predicted values that have at most ε deviations from the actual values y_i. The ε parameter defines the known 'tube' in the SVR literature. The problem is transformed into the following quadratic optimization problem, by introducing the slack variables ξ_i and ξ_i^*:

$$\text{Minimize } C\sum_{i=1}^{n}(\xi_i+\xi_i^*)+\frac{1}{2}\|w\|^2 \text{ subject to } \begin{Bmatrix} \xi_i \geq 0 \\ \xi_i^* \geq 0 \\ C>0 \end{Bmatrix} \text{ and } \begin{Bmatrix} y_i - w^T\varphi(x_i)-b \leq +\varepsilon+\xi_i \\ w^T\varphi(x_i)+b-y_i \leq +\varepsilon+\xi_i^* \end{Bmatrix} \quad (4)$$

Because of the unbounded nature of ε, Schölkopf *et al.* (1999) suggested the bounded SVR equivalent, where parameter $v \in (0,1)$ controls the ε allowing for a quicker and robust optimization solution. Equation (4) is transformed into the vSVR optimization problem that follows:

$$\text{Minimize } C\left(v\varepsilon+\frac{1}{n}\sum_{i=1}^{n}(\xi_i+\xi_i^*)\right)+\frac{1}{2}\|w\|^2 \text{ subject to } \begin{Bmatrix} \xi_i \geq 0 \\ \xi_i^* \geq 0 \\ C>0 \end{Bmatrix} \text{ and } \begin{Bmatrix} y_i - w^T\varphi(x_i)-b \leq +\varepsilon+\xi_i \\ w^T\varphi(x_i)+b-y_i \leq +\varepsilon+\xi_i^* \end{Bmatrix} \quad (5)$$

Equation (5) becomes a dual problem and its solution is based on the two Lagrange multipliers a_i, a_i^* and the kernel function $K(x_i, x)$:

$$f(x)=\sum_{\iota=1}^{n}(a_i-a_i^*)K(x_i,x)+b, \text{ where } 0 \leq a_i, a_i^* \leq \frac{C}{n} \quad (6)$$

In this study, the transformation of input space is achieved with the Gaussian Radial Basis Function (RBF) for all the SVR models applied. A RBF kernel is in general specified as:

$$K(x_i,x)=\exp(-\gamma\|x_i-x\|^2), \gamma>0 \quad (7)$$

where γ is the variance of the kernel function. RBFs require only one parameter to be optimized (γ) and provide good forecasting results in similar SVR applications (Lu et al. 2009; Yeh et al. 2011; Kao et al. 2013). From a theoretical perspective, parameter C satisfies the need to trade model complexity for a training error and vice versa (Cherkassky and Ma 2004). Additionally, the intuition of vSVR is that the parameter v is an approximation of the upper and lower bounds of the fraction of errors (Schölkopf et al. 1999). [3]

Although vSVR has been successfully applied in numerous relevant studies, it fundamentally does not account for the fact that recent information might be more relevant to interpret financial and economic series of data. The Locally Weighted Regression (LWR) is able to cope with that, since it is based on the assumption that the nearest to the predictor values are its best indicators. LWR is able to define an estimate *g(x)* for every value x in the dimensional space of the independent variables. Cleveland and Devlin (1988) suggest that the neighbourhood is defined by estimating the distances of q observations x_i from x, where $1 \leq q \leq n$. Closer points to x are assigned larger weights compared to those that are far, confirming the locality attribute of the method (Lee et al. 2005). The LWR is specified through the Euclidian distance ρ and a tricubic weight function *W(u)* for each training data (x_i, y_i) as below:

$$\left\{ W(u)=\begin{Bmatrix} (1-u^3)^3, 0 \leq u \leq 1 \\ 0, otherwise \end{Bmatrix}, \quad w_i(x_i)=W\left(\frac{\rho(x,x_i)}{d(x)}\right) \right\} \quad (8)$$

[3] For more details on the mathematical solutions and SVR modelling, the interested reader should refer to Vapnik (1995).

where $d(x)$ is the Euclidian distance specifically from the qth-nearest x_i to x. Weights are bounded as per $w_i \in [0,1]$, while the weight is maximized when x_i is closest to x and minimized for the q^{th}-nearest x_i to x. The work of Cleveland and Devlin (1988) provides a detailed explanation on the selection of weights and weight functions for the LWR purposes. Through the LWR, traditional vSVR can be transformed to a Locally Weighted vSVR (LSVR), where the parameter C is not constant, but locally adjusted as per $C_i = w_i * C$. Thus, the quadratic optimization problem of equation (5) is translated to the following:

$$\text{Minimize } C_i\left(v\varepsilon + \frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right) + \frac{1}{2}\|w\|^2 \text{subject to } \begin{Bmatrix} \xi_i \geq 0 \\ \xi_i^* \geq 0 \\ C_i > 0 \end{Bmatrix} \text{ and } \begin{Bmatrix} y_i - w^T\varphi(x_i) - b \leq +\varepsilon + \xi_i \\ w^T\varphi(x_i) + b - y_i \leq +\varepsilon + \xi_i^* \end{Bmatrix} \quad (9)$$

LSVR is theoretically advantageous over the traditional one, because the weighted C_i provides the better balance between the training error and model complexity. This improvement is derived by the fact that higher penalties (smaller weights) are imposed to bigger slack variables. Additionally, it is expected that the forecasting performance of LSVR increases gradually (Lee et al. 2005), unlike for the non-locally optimized vSVR.

5. Heuristic Models

This section summarizes the heuristic models applied to forecast the government bond spreads. Initially, the GA and KH SVR models are described. The SC counterpart is then explained. Finally, the SVR inputs and the benchmark model selection are explained.

5.1 Genetic Algorithm Support Vector Regression

The previous section and the numerous applications of GAs in SVR parameterization tasks motivates us to implement a simple GA to allow us to search the feature space and encode the optimal SVR parameters into parameter genes. This is done by implementing the traditional one-point crossover and the mutation operator (Goldberg 1989). The purpose of the one-point crossover operators is to create two offsprings from every two parents. The associated probability of selecting an individual as a parent for the crossover operator is called crossover probability. In this study this probability is set at the value of 0.90. The crossover probability is not set to 1 in order to allow very good solutions of a population to pass through the next generation's population without being altered. The offspring created is then replacing its parents in the population in forthcoming iterations. The second applied operator, the mutation one, is applied in order to randomly select genes in the population based on the preset mutation probability. In this case, setting this probability to 0.1 allows us to avoid local optima and explore a larger portion of the search space. Following the principles of Holland (1975), the roulette wheel selection method is applied for the GA selection step. Based on that, the genes with the better fitness are more probable to be selected. The evolution process is continued based on the elitism principle. Elitism accelerates the evolution and ensures that the best solutions are made available to the new population at the end of every generation. [4]The population of chromosomes is initialized in the training sub-period, while, in order to achieve the optimal selection of the SVR parameters, the root mean squared error (RMSE) needs to be minimized in the test-sub period. Therefore, the following fitness function needs to be maximized:

$$Fitness = 1/(1 + RMSE) \tag{10}$$

This process is performed for the *v*SVR and LSVR structures (GA-*v*SVR, GA-LSVR) with the application of the RBF kernel (equation (7)). The optimized parameters of the best solution are used to train the *v*SVR and LSVR to produce the final optimized forecast, which is evaluated out-of-sample. The technical characteristics of the GA are presented in Appendix B, while the flowchart of GA-*v*SVR methodology is show in figure 3.

5.2 Krill Herd Support Vector Regression

The KH algorithm is a metaheuristic optimization technique inspired by the behavior of krill individuals. Krill move in herds in nature with an aim to reach food resources quicker. This behavior is characterized by a mean-reversion effect observed when sea predators attack the herd. In other words, an attack on the herd disperses the krill individuals and reduces its density. Once the attack is over, krill individuals orientate to return in closer positions to other krill in order to return the herd density to previous levels. This orientation is based on sensing nearby krill moving towards the optimal path to reach food. Gandomi and Alavi (2012) suggest that the position (*P*) of each krill in the search space is influenced by three motions, namely, the movement induced by other krill (*M*), the foraging action (*F*) and the random diffusion (*RD*). These three motions allow the practitioner to calculate the changes in the position of every krill *j* are captured by the Lagrangian formulation:

$$\frac{dP_j}{dt} = M_j + F_j + RD_j \tag{11}$$

The new movement motion M^{t+1} of each krill *j* is calculated as:

$$\left\{ \begin{array}{l} M_j^{t+1} = M^{\max} eff_j + k_M \mathrm{M}^t \\ eff_j = eff_j^{loc} + eff_j^{t\arg} \end{array} \right\} \tag{12}$$

where M^{max} the maximum induced speed, $k_M \in [0,1]$ the inertia weight of the motion, eff_j the direction of the motion and $eff_j^{loc}, eff_j^{t\arg}$ the local effect by a neighbor krill and the target direction effect by the best individual krill, respectively. Given the total number of individual krill N_k, the neighbor krill are identified through the calculation of a sensing distance from the j^{th} krill:

$$d_{s,j} = (1/\mathrm{N_k}) \sum_{j'=1}^{N_k} \left\| P_j - P_{j'} \right\| \tag{13}$$

[4] For more technical details on the use of GAs for the SVR parameterization refer to Sermpinis et al. (2015).

The updated foraging motion F^{t+1} of every krill j is based on the food location estimate and the krill's past experience in locating a correct food position:

$$\begin{Bmatrix} F_j^{t+1} = V_F \, floc_j + k_F F_j^t \\ floc_j = floc_j^{food} + floc_j^{best} \end{Bmatrix} \tag{14}$$

where V_F is the foraging speed, $k_F \in [0,1]$ is the inertia weight of the motion, $floc_j$ is the location of the food and $floc_j^{food}$, $floc_j^{best}$ is the food attraction and the effect from the best food-locating j^{th} krill so far, respectively. The *RD* motion is calculated as the maximum diffusion speed RD^{max} and a random directional vector δ with values between -1 and 1. In other words:

$$RD = RD^{\max}\delta \tag{15}$$

The KH algorithm approximates the global optima of the krill swarm based on the food attraction estimate. The above motions show that optimal krill positions are obtained through local and global search of the search space. Namely, krill initially are searching locally in order to increase the herd density. The more the density increases, the more krill start to orientate towards food than neighbor krill. The *RD* motion diffuses potential biased movements of krill towards food or nearby krill. [5]Finally, the position *Pj* of each krill at time $t+\Delta t$ is given as:

$$\begin{Bmatrix} P_j(t+\Delta t) = P_j(t) + \Delta t \dfrac{dP_j}{dt} \\ \Delta t = Z_{cr} \sum_{r=1}^{NP} \left(UpB_r - LowB_r \right) \end{Bmatrix} \tag{16}$$

where $Z_{cr} \in [0,2]$ is a constant number, NP is the number of parameters optimized (in our case the three SVR parameters) and $UpB_r, LowB_r$ is the upper and lower bounds of the parameters respectively. The Δt practically is the only parameter that needs fine-tuning.[6] This is the striking advantage of the method compared to other more complicated metaheuristics approaches. As the final step towards improvement of the krill positions, mutation and crossover operators are applied.

In our study, the KH is used to optimize the *v*SVR and LSVR parameters (KH-*v*SVR, KH-LSVR) with the application of the RBF kernel (equation (7)). The practitioner is required to predefine three parameters, while the potential three-dimensional space is identified through the range of the bounds of each SVR parameter. The KH algorithm is optimized based on the same fitness function as in the GA counterpart (equation (10)). The KH algorithm is trained in the training sub-period and its performance is evaluated in the test sub-period. The technical characteristics of KH are given in Appendix B also. The flowcharts of GA-*v*SVR and KH-LSVR are presented in Figure 2(i) and 2(ii) respectively below:

****Insert Figure 2****

[5] The reader interested in the exact mathematical details of the three motions should refer to Gandomi and Alavi (2012).
[6] The maximum induced speed and the foraging speed are set to 0.01 and 0.02 ms^{-1} respectively. The Z_{cr} is set at values lower than 1. k_M and k_F are initially set high (0.9) and then linearly decreased to 0.1 (Gandomi and Alavi 2012).

5.3 Sine Cosine Support Vector Regression

The SC algorithm is a population-based optimization technique as proposed by Mirjalili (2016). SC is able to search different areas of the given search space by combining an exploration and exploitation phase, a property that is also observed in the KH algorithm. This attribute allows such population-based optimization techniques to start with a set of random solutions and proceed to the global optima. The exploration property minimizes the probability of getting trapped in the local optima and the exploitation one suggests that the higher the number of random solutions, the higher the probability of obtaining the global optima. In the case of SC, the best global solution is found by updating the positions of the random candidate solutions towards the best solution with the use of sine and cosine functions as objective functions. The local search of different regions in the search space is achieved by allowing the sine and cosine functions to return to values greater than one or less than minus one. Following the work of Mirjalili (2016), the position updating equations are the following:

$$P_j'^{t+1} = \begin{cases} P_j'^{t} + r_1' \sin(r_2') \left| r_3' P_j''^{t} - P_j'^{t} \right|, r_4' < 0.5 \\ P_j'^{t} + r_1' \cos(r_2') \left| r_3' P_j''^{t} - P_j'^{t} \right|, r_4' \geq 0.5 \end{cases} \tag{17}$$

where $P_j'^{t}$ is the position of the current solution for the j^{th} dimension at t^{th} and $P_j''^{t}$ is the position of the destination point, r_1', r_2', r_3', r_4' are random variables.

Regarding the random variables, $r_1' = c'' - t(c''/T'')$ is a balancing random metric, where c'' is a constant, t is the current iteration and T'' is the maximum number of iterations. This random variable is crucial for balancing exploration and exploitation phases of the algorithm. Calibrating r_1' leads to an adaptive shift in the range of sine and cosine calculations and consequently dictates the next positions' region, allowing for higher exploration of the search space. This region would be either in the space between the current solution and the next destination or outside it (see Figure 3). This cyclical pattern is based on the properties of sine and cosine functions and enables the SC algorithm to reposition new solutions around other ones, infusing better exploitation of the search space. The random variable r_2' is bounded between $[0,2\pi]$ and indicates whether the random location will be within or outside the cyclical pattern. In other words, it defines the allowed movement towards or outwards of the next destination. The third random variable r_3' is a random weight defining the emphasis of the destination position in defining the distance. Finally, r_4' is between [0,1] and provides an equal switch between the sine and cosine functions in equation (17). Figure 3 illustrates the effects of sine and cosine functions on the next positon identification.

****Insert Figure 3****

The heuristic algorithm presents several advantages over other similar techniques. SC generates improved sets of random solutions and benefits from high exploration and local optima avoidance, compared with individual-based algorithms, such as GAs. The algorithm is able to separate the search space into different areas that are explored as more promising when sine and cosine values are between minus one and one. The adaptive range imposed on the two functions allows for a smooth transition between exploration and exploitation, unlike in the case of KH. Finally, the best approximated global optimum is stored as a destination point; therefore, this information is always exploited in the next steps of the optimization.

Applying the SC algorithm to the traditional SVR and LSVR models seems ideal to evaluate whether these advantages are significant compared to previously used heuristic models. In order to achieve this, we follow the guidelines of Li et al. (2018). These authors suggest that the SC algorithm is able to calibrate and provide the optimal parameters of the SVR method, but initially data normalization to the range of [-1.1] and search space reconstruction based on the C-C method (Kim et al. 1999) are required. The fitness function and kernel function of the SC based SVR models are as in equations (7) and (10). The pseudo code of the SC algorithm is presented in Appendix B.

5.4 Input Selection and benchmark models

For the purposes of applying SVR for forecast combinations, we employ a pool of potential inputs. This pool consists of potential linear and non-linear individual predictors of all 10YGBS, which are applied to the in-sample period. The linear pool includes Simple Moving Averages (SMA), Exponential Moving Averages (EMA), Autoregressive terms (AR) and Autoregressive Moving Average (ARMA) models. The non-linear individual predictors include Smooth Transition Autoregressive Models (STAR), Nearest Neighbors Algorithms (*k*-NN), a Multi-Layer Perceptron (MLP), a Recurrent Neural Network (RNN), a Higher Order Neural Network (HONN), a Psi-Sigma Neural Network (PSN), Adaptive RBF and PSO Neural Network (ARBF-PSO), Genetic Programming (GP) and Gene Expression Programming (GEP). These predictors create a pool of three hundred and twenty-nine individual predictors in total for each forecasting exercise. The heuristic models will combine the best predictors in order to generate superior out-of-sample statistical performance (Timmermann 2006). Those models are traditional linear and non-linear models in the forecasting literature, thus, it is out of the scope of this paper to describe these models in detail. [7]

In such forecasting tasks, the large dimension can impede the statistical performance of the SVR models and increase exponentially the computational costs. Therefore, we follow Jolliffe (2002) and perform Principal Component Analysis (PCA) to discard highly correlated variables while continuing to account

[7] More information regarding the set of inputs is provided further in Appendix A.

for the 95% variance. The final PCs are used as final input sets to the SVR models. These sets are presented in Table 4.

****Insert Table 4****

It is obvious that the PCA analysis vastly decreases the input set dimensions. From the selected final inputs, as expected, non-linear models are dominant in terms of number and performance compared to their linear counterparts. The non-linear nature of the 10YGBS makes this result not surprising. The input performing the best most of times is the ARBF-PSO, while the linear PC are in two cases selected as the best individual predictors. The efficiency of the heuristic models is benchmarked with three non-SVR models, namely the simple RW, ARMAs and the best individual predictor in each case. Additionally, as benchmarks we employ the *v*SVR and LSVR. The majority of the SVR studies suggest that *v*SVR models are superior to their "ε" counterparts, hence, our selection in this study. For the traditional *v*SVR, the grid search technique is selected for the parameter calibration.

6. Empirical Results

This section shows the statistical analysis performed for the models, forecasting exercises and country-specific 10YGBS. The statistically accuracy is evaluated through the Mean Absolute Error (MAE) and the RMSE. Their mathematical specifications are in Appendix C. The out-of-sample results for the forecasting exercises F1 and F2 spanning 2000-2008 period are displayed in Table 5 and 6[8].

****Insert Table 5 and 6****

From Tables 5 and 6, it is interesting to see that the models' statistical ranking is relatively consistent across all countries and periods under study. The SC algorithm proves to be successful in optimizing the SVR and obtaining the highest accuracy throughout our forecasting exercises. The SC-LSVR appears to be the best performing model, which suggests that LSVR combined with SC algorithm is providing in most cases superior forecasts compared to the simple SVR structure. It should be noted, though, that this property is observed also in the remaining models. The KH counterparts are consistently the second best model in terms of forecasting accuracy. The benchmarks GA-*v*SVR and GA-LSVR do not manage to overcome neither the bio- nor the sine-cosine-inspired metaheuristics. Nonetheless, it is always found superior to the traditional *v*SVR and LSVR models. Finally, we should note that regardless of the particular parameterization process applied, the SVRs with the suggested input set have higher predictive ability than the best individual predictor of each case. This confirms the academic belief that combining forecasts from individual models with simple or more complex methods decreases forecast errors. Finally, the RW is the worst model in terms of the statistical metrics computed. In terms of the periods and countries under study, we observe, as expected, that government bond spreads are more difficult to be forecasted in F3 and F4 covering the Eurozone debt crisis period. Additionally, the results imply that

[8] The in-sample results are consistent with the out-of-sample ones and are available upon request. They are not presented here to preserve space.

the forecasting performance is lower in the cases of Greece, Ireland and Portugal than in the cases of Italy and Spain. This is finding holds across different data frequencies (between F1 and F2) and different periods (F1-F3 and F2-F4). This could be explained by the fact that bond spreads might be driven by factors such as credit, liquidity and subprime loan default risks. These risks are found in to be greater in these three countries as compared to Italy and Spain as they seem to suffer more from spill-over and contagion effects (Favero, 2013).

In order to further validate the above performance of our forecasts, the Diebold Mariano (DM) (1995) test is calculated. Its null hypothesis is that two forecasts have equal predictive accuracy. For the purposes of this study, the test is applied to the pairs of out-of-sample forecasts. The first forecast is from the superior model based on the previously reported results, i.e. the SC-LSVR. The second forecast is every other model. For the calculation of the DM statistic we select the MAE as loss function. If the calculated DM statistic is negative, then the test indicates that the first model is more accurate than the second one. The lower the negative value, the more accurate are the SC-LSVR forecasts. In addition to the DM test, we also employ the Pesaran-Timmermann (PT) (1992) test. The PT tests the direction co-movements of the real and forecasted values. Its null hypothesis is that the model under study has no power in predicting the 10YGBS return series[9]. Table 7 below shows these results:

****Insert Table 7****

The two additional tests confirm the statistical ranking of the models utilized. The negative values of the DM statistic further support the statistical superiority of the SC-LSVR over all other models at 5% significance level. Only in four cases the DM statistic is not found significant at 5%, when the SC-LSVR is compared with the SC-*v*SVR. Nonetheless, in these cases the PT statistic is found more significant. Therefore, we conclude that the SC LSVR counterpart is the best performing model across all countries and forecasting exercises. The PT values also validate that metaheuristics are superior to the heuristic GA benchmark. Finally, in six and three cases RW and ARMAs are found to not be able to forecast the 10YGBS series respectively, while in only one case an individual predictor is failing to do so. This is not surprising, as the best individual predictors are non-linear adaptive models as shown in Table 5. These types of models are known to have a significant forecasting power in financial and economic series as our dataset. Overall, our findings support the strand of literature that suggests that the SVRs are suitable modelling tools for such forecasting exercises, while practitioners need to pay more attention to metaheuristic calibration of the SVR parameters in order to achieve increased predictive performance.

[9] The mathematical details of the tests can be found in Pesaran and Timmermann (1992) and Diebold and Mariano (1995).

7. Conclusions

This work aims to contribute to two strands of the literature. Firstly, we investigate the predictability of 10YGBS of the GIIPS during two sample periods, 2000-2008 and 2008-2017 (including in the out-of-sample exercises the GFS and SDC), and, consequently, we add to the existing studies investigating the puzzle of bond spreads co-movements. Secondly, this study explores the utility of heuristics and metaheuristics for SVR optimization purposes. For this reason, initially, a pool of individual forecasts is generated based on the traditional linear and non-linear models. These models are screened and the final input vector is selected for the SVR structures. The study applies the GA, KH and SC algorithm in *v*SVR and LSVR techniques. These models are benchmarked with a RW and the best individual predictor from the input set of each country and period under study. Apart from these two non-SVR benchmarks, the traditional *v*SVR and LSVR are also compared with the GA, KH, and SC counterparts in terms of forecasting accuracy.

The results provide several interesting findings. The SC algorithm is able to achieve high forecasting power for the 10YGBS and can be considered as the best prediction model of this study for all countries and periods considered. The SC-LSVR is found to be superior to the bio-inspired KH-LSVR, while both metaheuristic hybrids outperform the evolutionary heuristic GA-LSVR. This is in line with the literature that suggests that metaheuristics are able to achieve a beneficial trade-off between local and global search of the feature space. Our results are also consistent with a handful of studies that bring forward the superiority of LSVR over the traditional SVRs due to the higher penalties imposed to past observations (Yang et al., 2009; Wu and Akbarov, 2011; Sermpinis et al., 2017). In our case, all the LSVR structures regardless of the parameterization method are found to decrease forecast errors, the property that is also observed within the same class of models (e.g. SC-*v*SVR and SC-LSVR). Overall, this work provides insights to academics, researchers and policy makers on how heuristics and metaheuristics can benefit the calibration of sophisticated models, while the results highlight that the SVR parameterization is a crucial modelling aspect for the heuristics literature. This study could be further extended by incorporating bond spreads from the US and other developed countries and by investigating forecasting performance over different time horizons.

References

Abad, P., Chuliá, H., Gómez-Puig, M., et al. (2010). EMU and European government bond market integration. *Journal of Banking & Finance*, 34(12), 2851-2860.

Aguilar-Rivera, R., Valenzuela-Rendón, M., Rodríguez-Ortiz, J. J., et al. (2015). Genetic algorithms and Darwinian approaches in financial applications: A survey. *Expert Systems with Applications*, 42 (21), 7684-7697.

Ahn, H., Kim, K. J., et al. (2008). Using genetic algorithms to optimize nearest neighbors for data mining. *Annals of Operations Research*, 163(1), 5-18.

Alcaraz, J., Maroto, C., et al. (2001). A robust genetic algorithm for resource allocation in project scheduling. *Annals of Operations Research*, 102(1-4), 83-109.

Antonakakis, N., and Vergos, K., 2013. Sovereign bond yield spillovers in the Euro zone during the financial and debt crisis. *Journal of International Financial Markets, Institutions and Money*, 26, 258-272.

Aristei, D. and Martelli, D., 2014. Sovereign bond yield spreads and market sentiment and expectations: Empirical evidence from Euro area countries. *Journal of Economics and Business*, 76, 55-84.Chan, K. S., Tong, H., et al. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7 (3), 178–190.

Cherkassky, V., Ma, Y., et al. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17 (1), 113-126.

Cleveland, W.S., Devlin, S.J., et al. (1988). Locally Weighted Regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83 (403), 596-610.

Darracq-Paries, M., Moccero, D.N., Krylova, E., and Marchini, C., 2014. The retail bank interest rate pass-through the case of the euro area during the financial and sovereign debt crisis. European Central Bank, working paper No. 155.

Dewachter, H., Iania, L., Lyrio, M., and de Sola Perea, M., 2015. A macro-financial analysis of the euro area sovereign bond market. *Journal of Banking and Finance*, 50, 308-325.

Dorigo, M., Birattari, M., Stutzle, T., et al. (2006). Ant colony optimization. *IEEE computational intelligence magazine*, 1(4), 28-39.

Diebold, F.X., Mariano, R.S., et al. (1995). Comparing predictive ability. *Journal of Business & Economic Statistics*, 13 (3), 253-263.

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

Duffie, D., Pedersen, L.H., Singleton, K. J. et al. (2003). Modeling sovereign yield spreads: A case study of Russian debt. *The Journal of Finance*, 58(1), 119-159.

Dunis, C.L., Nathani, A., et al. (2007). Quantitative trading of gold and silver using non-linear models. *Neural Network World*, 16 (2), pp. 93-111.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14 (2)*, 179-211.

European Central Bank, 2010. Monetary policy transmission in the euro area, a decade after the introduction of the euro. *Monthly Bulletin*, May 2010, 85-98.

Favero, C.A. (2013). Modelling and forecasting government bond spreads in the euro area: a GVAR model. *Journal of Econometrics*, 177(2), 343-356.

Ferreira, C. (2001). Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. *Complex Systems*, 13(2), 87-129.

Fonseca, C.M., Fleming, P. J., et al. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary computation*, 3(1), 1-16.

Fix, E., Hodges, J. L., et al. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. *Technical Report 4*, USAF School of Aviation Medicine, Randolph Field, Texas.

Gandomi, A.H., Alavi, A. H., et al. (2012). Krill herd: A new bio-inspired optimization algorithm. *Communications in Nonlinear Science and Numerical Simulation*, 17 (12), 4831-4845.

Georgoutsos, D.A., and Migiakis, P.M., 2012. Heterogeneity of the determinants of euro-area sovereign bond spreads; what does it tell us about financial stability? Bank of Greece, working paper No. 143.

Gilli, M., Maringer, D, Winker, P. (2008). *Applications of heuristics in finance*. In, Seese, D., Weinhardt,C., & Schlottman, F. (Eds.), Handbook On Information Technology in Finance, International Handbooks on Information Systems (pp. 635-653). Germany, Springer.

Gilli, M., Schumann, E., et al. (2012). Heuristic optimisation in financial modelling. *Annals of operations research*, 193(1), 129-158.

Gómez-Puig, M. (2009). Systemic and Idiosyncratic Risk in EU-15 Sovereign Yield Spreads after Seven Years of Monetary Union. *European Financial Management*, 15(5), 971-1000.
Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley*.
Ghosh, J., Shin, Y., et al. (1991). The Pi-Sigma Network: An efficient higher-order neural networks for pattern classification and function *approximation. Proceedings of International Joint Conference of Neural Networks, 1, 13-18.*
Holland, J. (1975). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence. Cambridge, Mass: MIT Press.
Jarque, C.M., Bera, A.K., et al. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6 (3), 255-259.
Jolliffe, I. (2002). Principal component analysis and factor analysis. In Principal Component Analysis Springer Series in Statistics book series (pp. 150-166). New York, Springer-Verlag.
Kao, L. J., Chiu, C. C., Lu, C. J., Yang, J. L., et al. (2013) Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing*, 99(1), 534-542.
Karaboga, D., Akay, B., et al. (2009). A comparative study of artificial bee colony algorithm. *Applied mathematics and computation*, 214(1), 108-132.
Manganelli, S., Wolswijk, G., et al. (2009). What drives spreads in the euro area government bond market ?. *Economic Policy*, 24(58), 191-240.
Kaastra, I., Boyd, M., et al. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215-236.
Kim, H., Eykholt, R., Salas, J. D., et al. (1999). Nonlinear dynamics, delay times, and embedding windows. *Physica D: Nonlinear Phenomena*, 127(1), 48-60.
Koza, J., Poli, R., et al. (2005). *Genetic Programming*. In: Burke, E.K., Kendall, G. (eds.), Search Methodologies, Introductory Tutorials in Optimization and Decision Support Techniques (pp.127-164). New York: Springer.
Lautenschläger, S., 2017. Between low interest rates and bond purchases - has European monetary policy reached a dead end? Speech at Hohenheim University, Stuttgart, 9 October 2017.
Lee, D. E., Song, J. H., Song, S. O., Yoon, E. S., et al. (2005). Weighted support vector machine for quality estimation in the polymerization process. *Industrial & Engineering Chemistry Research*, 44(7), 2101-2105.

Leschinski, C., Bertram, P., et al. (2017). Time varying contagion in EMU government bond spreads. *Journal of Financial Stability*, 29, pp.72-91.
Li, S., Fang, H., Liu, X., et al. (2018). Parameter optimization of support vector regression based on sine cosine algorithm. *Expert Systems with Applications*, 91, 63-77.
Li, X., Zhang, J., Yin, M., et al. (2014). Animal migration optimization: an optimization algorithm inspired by animal migration behavior. *Neural Computing and Applications*, 24 (7-8), 1867-1877.
Liang, J.J., Qin, A.K., Suganthan, P.N., Baskar, S., et al. (2006). Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *IEEE Transactions on Evolutionary Computations*, 10 (3), 281–295.
Lin, C.J., Teräsvirta, T., et al. (1994). Testing the constancy of regression parameters against continuous structural changes, *Journal of Econometrics*, 62 (2), 211–228.
Lu, C.J., Lee, T.S., Chiu, C.C., et al. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 47 (2), 115-125.
Mirjalili, S., (2016). SCA: a sine cosine algorithm for solving optimization problems. *Knowledge-Based Systems*, 96, 120-133.

Paniagua, J., Sapena, J., and Tamarit, C., 2016. Sovereign debt spreads in EMU: The time-varying role of fundamentals and market distrust. *Journal of Financial Stability*, 10.1016/j.jfs.2016.06.004 (Available online 6 June 2016)

Patel, J., Shah, S., Thakkar, P., Kotecha, K., et al. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268.

Pesaran, M.H., Timmerman, A.G., et al. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, 10 (4), 461-465.

Schölkopf, B., Bartlett, P., Smola, A., Williamson, R., et al. (1999). *Shrinking the tube: a new support vector regression algorithm*, In: Kearns, M. J., (ed.), Advances in neural information processing systems 11 (pp. 330-336). Cambridge: MIT Press.

Shapiro, A. F. (2000). A Hitchhiker's guide to the techniques of adaptive non-linear models. *Insurance: Mathematics and Economics*, 26 (2-3), 119-132.

Sermpinis, G., Stasinakis, C., Hassanniakalager, A., et al. (2017). Reverse Adaptive Krill Herd: Application with Locally Weighted Support Vector Regression for forecasting and trading Exchange Traded Funds. *European Journal of Operational Research*, 263 (2), 540-558.

Sermpinis, G., Stasinakis, C., Theofilatos, K., Karathansopoulos, A., et al. (2015). Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms – support vector regression forecast combinations. *Journal of Operational Research*, 247(3), 831-846.

Sermpinis, G., Theofilatos, K.A, Karathanasopoulos, A.S., Georgopoulos, E.F., Dunis, C.L., et al. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization. *European Journal of Operational Research*, 225 (3), 528–540.

Stasinakis, C., Sermpinis, G., Psaradellis, I., Verousis, T., et al. (2016). Krill herd support vector regression and heterogeneous autoregressive leverage: evidence from forecasting and trading commodities. *Quantitative Finance*, 16(102), 1901-1915.

Talbi, E. G. (2009). *Metaheuristics: from design to implementation*. Hoboken, New Jersey, USA: John Wiley & Sons.

Tenti, P. (1996). Forecasting foreign exchange rates using recurrent neural networks, *Applied Artificial Intelligence*, 10 (6), 567-582.

Timmermann, A. (2006). *Forecast Combinations*, In: Elliott, G. Granger, C.W.J. &. Timmermann, A., (Eds), Handbook of Economic Forecasting (pp. 135-196). Amsterdam: Elsevier.

Van Breedam, A. (2001). Comparing descent heuristics and metaheuristics for the vehicle routing problem. *Computers & Operations Research*, 28 (4), 289-315.

Vapnik, V. (1995). *The nature of statistic learning theory*. New York: Springer-Verlag.

Wang, G. G., Guo, L., Gandomi, A. H., Hao, G. S., Wang, H., et al. (2014). Chaotic Krill Herd algorithm. *Information Sciences*, 274 (1), 17-34.

Wang, L., Zhu, J., et al. (2010). Financial market forecasting using a two-step kernel learning method for the support vector regression. *Annals of Operations Research*, 174(1), 103-120.

Wu, S., Akbarov, A., et al. (2011). Support vector regression for warranty claim forecasting. *European Journal of Operational Research*, 213 (1), 196-204.

Yang, H., Huang, K., King, I., Lyu, M.R., et al. (2009). Localized support vector regression for time series prediction, *Neurocomputing*, 72 (10–12), 2659-2669.

Yang, X. S. (2010). *Firefly algorithm nature-inspired metaheuristic algorithms*. United Kingdom: Luniver Press.

Yang, X. S., Gandomi, A. H., et al. (2012). Bat algorithm: a novel approach for global engineering optimization. *Engineering Computations*, 29 (5), 464–483.

Yeh, C. Y., Huang, CW., Lee, S. J., et al. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Systems with Applications*, (3), 2177-2186.
Yuan, F. C. (2012). Parameters optimization using genetic algorithms in support vector regression for sales volume forecasting. *Applied Mathematics*, 3 (1), 1480-1486.

Appendix

A. Predictors' summary

This appendix section provides some further information regarding the linear and non-linear models used to populate the individual forecast pools. The linear models used are SMA, EMA, AR and ARMA. Their specifications are provided in Table A.1. In total, the linear models' forecasts sum up to 290.

****Insert Table A.1****

Except from the above non-linear models, this study utilizes also Smooth Transition Autoregressive (STAR) models as proposed by Chan and Tong (1986) and Lin and Teräsvirta (1994). More specifically, we estimate the logistic and exponential specification, namely the LSTAR and ESTAR for orders 1 to 15. The STAR specifications combine two AR models with a function that defines the degree of non-linearity (smooth transition function). For more mathematical details, see Lin and Teräsvirta (1994).
The next non-linear model applied is the Nearest Neighbours (*k*-NN) inspired by the work of Fix and Hodges (1951). The intuition of the model is that past time series observations project patterns with resemblances to those of the future. Thus, the Euclidean distance is used to provide a sensitivity metric that transforms these patterns to nearest neighbours, which then are used to forecast the the immediate future. These neighbours are approximated by following the guidelines Dunis and Nathani (2007).

This paper also incorporates individual forecasts obtained by five NN architectures. The first is the traditional MLP. The MLP structure is three-layered and follows the training principles of back-propagation of errors and 'early stopping' explained by Shapiro (2000). For more information on MLPs refer to Stasinakis et al. (2016). The second NN is the RNN as proposed by Elman (1990). RNNs are extensions of the MLPs embodying an activation feedback offering short-term memory benefits (Tenti 1996). The third NN model also included in the input set is the HONN. HONNs can achieve superior simulations due to their ability to adapt to data with higher frequencies and orders. For more information on HONNs see Dunis *et al.* (2011). The fourth NN structure applied in this study is the PSN as introduced by Ghosh and Shin (1991). PSNs are a class of feed-forward fully connected HONNs and combine fast learning abilities with powerful and computationally quicker mapping properties. For more information regarding the properties of PSN, practitioner should refer to Sermpinis et al. (2015). The last NN used is the ARBF-PSO. Compared to other NN structures, the ARBF-PSO utilizes the Particle Swarm Optimization (PSO) algorithm to optimize the weights/parameters. The PSO algorithm is a nature inspired heuristic search algorithm based on the flock behaviour of birds (Liang et al, 2006), offering benefits in terms of over-fitting and data snooping bias. The complete description of the ARFB-PSO can be found in the work of Sermpinis *et al.* (2013). Table A.2 summarizes the learning algorithm, hidden and output node activation functions for all previous structures.

****Insert Table A.2****

In general, there is no formalized process for selecting the NN inputs and their characteristics, such as number of hidden neurons, learning rate, momentum and iterations. In this study, we follow studies suggesting that a sensitivity analysis on a pool of autoregressive terms of all 10YGBS series in the in-sample dataset (Tenti, 1996). We start our experiments with 500 number of iterations to 100000 (increasing the iterations by 500 at each experiment). Based on these experiments and the sensitivity analysis, the sets of variables selected are those that provide the higher statistical performance for each network in the in-sample period.

The final two non-linear models of the input set are inspired by the Darwinian principle of survival of the fittest, namely the GP and GEP. GP is considered a class of GAs as explained in section 4.1. The GP algorithm is designed with a focus on limiting and optimizing the computation time and the 'bloat effect'. Koza and Poli (2005) provide an extensive analysis of the GP structure. GEP is based on symbolic strings of fixed length that represent the genotype of an organism. Using GEP offers the safety to have lways valid expression trees generated, which is not always the case in GP. In general, GEP is considered superior to GP because fitness is established through the genotype and phenotype of an individual based on its chromosomes and expression trees respectively. The explanations of these terms and the GEP procedure can be found in Ferreira (2001).

B. Heuristic models' technical characteristics

In this appendix section we discuss some of the technical characteristics related to the heuristic models under study. Regarding the GA SVR models, we follow the guidelines of Sermpinis et al., (20015) where the parameter genes' encoding is done by using 50 bits as follows:

- 10 bits to represent the integer part of parameter C of SVRs (range [0-1024])
- 10 bits to represent the decimal part of parameter C of SVRs (~ 0.001 precision)
- 10 bits to represent the integer part of the γ parameter of RBF functions (range [0-1024])
- 10 bits to represent the decimal part of the γ parameter of RBF functions (~ 0.001 precision)
- 10 bits to represent the ν parameter of v-SVR (range [0-1] with ~ 0.001 precision)

In the initial step, all genes are randomly set with values 0 or 1 with equal probabilities for both of them. The training characteristics for the GA and KH SVR models per forecasting exercise and country under study are shown in Table B.1.

****Insert Table B.1****

Figures B.1 and B.2 illustrate the pseudo codes of the KH and SC applied within the SVR structure respectively.

****Insert Figure B.1 and Figure B.2****

C. Statistical performance measures.

The statistical performance measures are calculated as shown in Table C.1.

****Insert Table C.1****

Tables

Table 1: The government bond spreads under study

Countries	Description	Frequency	TICKER
Greece	Greece – Germany 10 year Government Bond Spread	Daily, Weekly	.GDRDEM10 G
Ireland	Ireland – Germany 10 year Government Bond Spread	Daily, Weekly	.IREDE10B G
Italy	Italy – Germany 10 year Government Bond Spread	Daily, Weekly	.ITLDEM10 G
Portugal	Portugal – Germany 10 year Government Bond Spread	Daily, Weekly	.PTEDEM10 G
Spain	Spain – Germany 10 year Government Bond Spread	Daily, Weekly	.SPGER10

Note: The source of the data is Bloomberg

Table 2: Summary statistics

Periods	Frequency	Statistic	Greece	Ireland	Italy	Portugal	Spain
2000-2008	Daily	Mean	0.0018	-0.0119	0.0019	0.0084	-0.0334
		Standard deviation	0.0605	2.1324	0.0484	0.7329	1.1547
		Skewness	8.5857	-8.9073	4.2741	29.1845	-31.1391
		Kurtosis	191.519	448.943	101.086	1427.389	1211.388
		Jarque-Bera (p value)	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		ADF (p value)	0.0000***	0.0001***	0.0001***	0.0001***	0.0001***
2000-2008	Weekly	Mean	0.0064	-0.0441	0.0102	-0.0271	-0.0263
		Standard deviation	0.1098	2.0018	0.1275	0.7899	1.1143
		Skewness	4.5745	-3.5363	6.2124	-18.2436	1.3031
		Kurtosis	46.776	115.972	84.303	373.639	62.530
		Jarque-Bera (p value)	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		ADF (p value)	0.0000***	0.0000***	0.0000***	0.0001***	0.0000***
2008-2017	Daily	Mean	0.0019	0.0011	0.0023	0.0089	0.0019
		Standard deviation	0.0379	0.0397	0.0398	0.0396	0.0440
		Skewness	-0.7905	-1.0698	0.5541	1.0387	0.3430
		Kurtosis	41.784	36.049	9.216	20.78450	12.732
		Jarque-Bera (p value)	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		ADF (p value)	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
2008-2017	Weekly	Mean	0.0103	0.0064	0.0122	0.0099	0.0098
		Standard deviation	0.0981	0.0992	0.0928	0.1041	0.1058
		Skewness	0.7691	0.1341	0.6510	1.0352	1.3124
		Kurtosis	11.041	9.193	5.859	8.874	9.845
		Jarque-Bera (p value)	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		ADF (p value)	0.0000***	0.0000***	0.0001***	0.0000***	0.0000***

Note: The table presents the descriptive statistics for each data frequency and sample period, along with the p-values of the Jarque-Bera and ADF test.

Table 3: The total dataset and relevant forecasting exercises

FORECASTING EXERCISE	PERIODS	TRADING DAYS	START DATE	END DATE
F1 (Daily/ 2000-2008)	Total Dataset	2348	03/01/2000	31/12/2008
	Training Dataset	1305	03/01/2000	30/12/2004
	Test Dataset	520	03/01/2005	29/12/2006
	Out-of-sample Dataset	523	01/01/2007	31/12/2008
F2 (Weekly/ 2000-2008)	Total Dataset	469	07/01/2000	26/12/2008
	Training Dataset	261	07/01/2000	31/12/2004
	Test Dataset	104	07/01/2005	29/12/2006
	Out-of-sample Dataset	104	05/01/2007	26/12/2008
F3 (Daily/ 2008-2017)	Total Dataset	2523	01/01/2008	31/08/2017
	Training Dataset	1305	01/01/2008	31/12/2012

	Test Dataset	522	01/01/2013	31/12/2014
	Out-of-sample Dataset	696	01/01/2015	31/08/2017
F4 (Weekly/ 2008-2017)	Total Dataset	504	04/01/2008	25/08/2017
	Training Dataset	261	04/01/2008	28/12/2012
	Test Dataset	104	04/01/2013	26/12/2014
	Out-of-sample Dataset	139	02/01/2015	25/08/2017

Note: The in-sample periods are the sum of the training and test datasets.

Table 4: The SVR sets of inputs

	Greece	Ireland	Italy	Portugal	Spain
F1	AR(3), EMA(1), MLP, RNN, HONN, **PSN**	SMA(1), ARMA(1,3), MLP, HONN, GP, **GEP**	MLP, RNN, HONN, PSN, **ARBF-PSO**	MLP, RNN, HONN, **PSN**, ARBF-PSO, *k*-NN	LSTAR(2), ARMA(1,4), PSN, GP, GEP, **ARBF-PSO**
F2	ARMA(3,4), **MLP**, RNN, ARBF-PSO	EMA(3), **ESTAR(6)**, MLP, GP, GEP, ARBF-PSO	AR(3), SMA(3), MLP, RNN, HONN, PSN, **GP**	SMA(5), *k*-NN, MLP, RNN, PSN, GEP, **ARBF-PSO**	AR(5), ARMA(1,6), ARMA(3, 3), ESTAR(6), MLP HONN, **PSN**, GP, ARBF-PSO
F3	AR(2), SMA(4), EMA(3), MLP, **RNN**, PSN, GP, GEP	***k*-NN**, MLP, RNN, HONN, PSN	MLP, RNN, PSN, ARBF-PSO	**AR(3)**, ARMA(4,6), HONN, GP, GEP, ARBF-PSO, *k*-NN	MLP, **RNN**, HONN, PSN, GP, GEP, ARBF-PSO
F4	ARMA(3,6), PSN, GP, GEP, **ARBF-PSO**	MLP, RNN, PSN, **ARBF-PSO**	AR(1), ARMA(1,2), MLP, PSN, GP, GEP, **ARBF-PSO**	SMA(5), ARMA(4, 8), MLP, **PSN**, ARBF-PSO	SMA(4), ESTAR(5), MLP, RNN, PSN, **ARBF-PSO**

Note: The inputs in bold are the Best performing individual predictor in terms of statistical accuracy in the in-sample period of each forecasting exercise. These are used as benchmark models for the statistical analysis that follows.

Table 5: Out-of-sample Statistical Performance (2000-2008)

	Countries	Statistic	RW	ARMA	Best	*v*SVR	LSVR	GA-*v*SVR	GA-LSVR	KH-*v*SVR	KH-LSVR	SC-*v*SVR	SC-LSVR
F1	**Greece**	MAE	0.0112	0.0112	0.0102	0.0098	0.0097	0.0094	0.0091	0.0088	0.0081	0.0076	**0.0072**
		RMSE	0.0123	0.0119	0.0102	0.0096	0.0093	0.0090	0.0090	0.0086	0.0084	0.0081	**0.0079**
		THEIL-U	1.0155	1.0115	0.9994	0.9985	0.9884	0.9759	0.9764	0.9605	0.9518	0.9410	**0.9402**
	Ireland	MAE	0.0099	0.0095	0.0092	0.0088	0.0084	0.0080	0.0078	0.0074	0.0074	0.0070	**0.0069**
		RMSE	0.0096	0.0105	0.0097	0.0092	0.0090	0.0084	0.0082	0.0079	0.0076	**0.0073**	**0.0073**
		THEIL-U	1.0128	1.0185	1.0132	1.0081	0.9989	0.9945	0.9901	0.9845	0.9810	0.9775	**0.9708**
	Italy	MAE	0.0102	0.0101	0.0092	0.0082	0.0081	0.0075	0.0072	0.0066	0.0062	0.0058	**0.0054**
		RMSE	0.0099	0.0094	0.0088	0.0086	0.0084	0.0084	0.0080	0.0072	0.0071	0.0064	**0.0061**
		THEIL-U	1.0085	0.9985	0.9930	0.9815	0.9749	0.9684	0.9602	0.9448	0.9094	0.8847	**0.8648**
	Portugal	MAE	0.0125	0.0099	0.0096	0.0093	0.0093	0.0089	0.0085	0.0079	0.0078	0.0072	**0.0068**
		RMSE	0.00107	0.0098	0.0088	0.0081	0.0080	0.0075	0.0072	0.0069	0.0068	0.0064	**0.0060**
		THEIL-U	1.0280	1.0184	1.0165	1.0155	1.0085	0.9915	0.9945	0.9890	0.9884	0.9658	**0.9614**
	Spain	MAE	0.0115	0.0112	0.0087	0.0086	0.0083	0.0079	0.0076	0.0067	0.0065	0.0062	**0.0058**
		RMSE	0.0097	0.0092	0.0087	0.0085	0.0084	0.0081	0.0075	0.0070	0.0067	0.0062	**0.0059**
		THEIL-U	1.0081	1.0074	0.9986	0.9958	0.9886	0.9748	0.9658	0.9458	0.9335	0.9225	**0.9115**
F2	**Greece**	MAE	0.0110	0.0103	0.0099	0.0094	0.0092	0.0090	0.0091	0.0088	0.0085	0.0077	**0.0075**
		RMSE	0.0142	0.0140	0.0133	0.0105	0.0101	0.0098	0.0096	0.0095	0.0094	0.0090	**0.0088**
		THEIL-U	1.0165	1.0135	1.0122	1.0114	1.0105	1.0095	1.0025	0.9945	0.9801	0.9668	**0.9610**
	Ireland	MAE	0.0097	0.0098	0.0095	0.0086	0.0087	0.0084	0.0081	0.0076	0.0075	0.0070	**0.0067**
		RMSE	0.0106	0.0102	0.0096	0.0094	0.0094	0.0091	0.0089	0.0087	0.0084	**0.0081**	**0.0081**
		THEIL-U	1.0215	1.0188	1.0168	1.0100	1.0089	1.0080	1.0054	1.0021	0.9994	0.9951	**0.9749**
	Italy	MAE	0.0117	0.0119	0.0095	0.0088	0.0087	0.0081	0.0072	0.0069	0.0064	0.0060	**0.0055**
		RMSE	0.0119	0.0110	0.0102	0.0099	0.0097	0.0094	0.0092	0.0090	0.0085	0.0081	**0.0078**
		THEIL-U	1.0285	1.0185	1.0153	1.0102	1.0090	1.0070	1.0005	0.9965	0.9881	0.9746	0.9665

	Portugal	MAE	0.0125	0.0121	0.0092	0.0090	0.0087	0.0084	0.0085	0.0082	0.0077	0.0071	**0.0067**
		RMSE	0.0208	0.0200	0.0194	0.0105	0.0097	0.0096	0.0095	0.0092	0.0091	0.0087	**0.0084**
		THEIL-U	1.0265	1.0205	1.0102	1.0095	1.0050	1.0041	0.9950	0.9910	0.9842	0.9716	**0.9654**
	Spain	MAE	0.0113	0.0108	0.0089	0.0084	0.0081	0.0079	0.0074	0.0068	0.0064	0.0060	**0.0058**
		RMSE	0.0207	0.0209	0.0112	0.0095	0.0084	0.0084	0.0083	0.0081	0.0079	0.0074	**0.0070**
		THEIL-U	1.0119	1.0105	1.0085	1.0021	0.9980	0.9841	0.9749	0.9664	0.9620	0.9558	**0.9432**

Note: Best is the best individual predictors in terms of statistical performance respectively in the in-sample period (as outlined in table 5).

Table 6: Out-of-sample Statistical Performance (2008-2017)

	Countries	**Statistic**	**RW**	**ARMA**	**Best**	***v*SVR**	**LSVR**	**GA-*v*SVR**	**GA-LSVR**	**KH-*v*SVR**	**KH-LSVR**	**SC-*v*SVR**	**SC-LSVR**
F3	**Greece**	MAE	0.0185	0.0179	0.0177	0.0170	0.0164	0.0138	0.0120	0.0118	0.0109	0.0102	**0.0097**
		RMSE	0.0274	0.0258	0.0241	0.0217	0.0208	0.0198	0.0194	0.0189	0.0173	0.0149	**0.0125**
		THEIL-U	1.0123	1.0100	0.9981	0.9971	0.9804	0.9741	0.9730	0.9499	0.9418	0.9337	**0.9314**
	Ireland	MAE	0.0166	0.0185	0.0180	0.0155	0.0142	0.0128	0.0115	0.0109	0.0102	**0.0095**	**0.0095**
		RMSE	0.0214	0.0195	0.0186	0.0167	0.0156	0.0133	0.0129	0.0124	0.0118	0.0107	**0.0103**
		THEIL-U	1.0108	1.0089	1.0074	0.9948	0.9947	0.9853	0.9810	0.9740	0.9612	0.9441	**0.9226**
	Italy	MAE	0.0166	0.0136	0.0128	0.0099	0.0096	0.0095	0.0092	0.0090	0.0087	0.0086	**0.0084**
		RMSE	0.0245	0.0218	0.0207	0.0158	0.0117	0.0102	0.0097	0.0095	0.0093	0.0090	**0.0087**
		THEIL-U	1.0053	0.9990	0.9964	0.9758	0.9551	0.9564	0.9234	0.9015	0.8994	0.8842	**0.8445**
	Portugal	MAE	0.0140	0.0142	0.0128	0.0110	0.0108	0.0102	0.0100	0.0098	0.0096	0.0094	**0.0090**
		RMSE	0.0195	0.0144	0.0132	0.0151	0.0124	0.0119	0.0109	0.0108	0.0102	0.0097	**0.0095**
		THEIL-U	1.0245	1.0123	1.0099	0.9964	0.9954	0.9941	0.9807	0.9800	0.9718	**0.9515**	**0.9515**
	Spain	MAE	0.0144	0.0128	0.0120	0.0101	0.0098	0.0095	0.0094	0.0090	0.0087	0.0084	**0.0083**
		RMSE	0.0189	0.0180	0.0152	0.0140	0.0110	0.0105	0.0102	0.0097	0.0096	0.0095	**0.0090**
		THEIL-U	0.9940	0.9841	0.9784	0.9664	0.9752	0.9728	0.9514	0.9228	0.9207	0.9112	**0.9015**
F4	**Greece**	MAE	0.0188	0.0189	0.0184	0.0180	0.0176	0.0171	0.0164	0.0150	0.0144	0.0131	**0.0128**
		RMSE	0.0299	0.0266	0.0258	0.235	0.0225	0.0201	0.0195	0.0184	0.0172	0.0166	**0.0158**
		THEIL-U	1.0153	1.0105	0.9984	0.9980	0.9854	0.9801	0.9745	0.9551	0.9550	0.9447	**0.9401**
	Ireland	MAE	0.0160	0.0185	0.0183	0.0150	0.0142	0.0132	0.0118	0.0115	0.0105	0.0102	**0.0099**
		RMSE	0.0221	0.0201	0.0194	0.0171	0.0162	0.0155	0.0149	0.0130	0.0121	0.0118	**0.0107**
		THEIL-U	1.0180	1.0165	1.0140	1.0094	0.9988	0.9970	0.9815	0.9790	0.9715	0.9518	**0.9348**
	Italy	MAE	0.0188	0.0154	0.0108	0.0097	0.0094	0.0092	0.0089	0.0085	0.0086	0.0082	**0.0084**
		RMSE	0.0245	0.0233	0.0207	0.0158	0.0117	0.0102	0.0097	0.0095	0.0093	**0.0088**	**0.0088**
		THEIL-U	1.0253	1.0085	0.9842	0.9809	0.9779	0.9713	0.9664	0.9624	0.9557	0.9411	0.9335
	Portugal	MAE	0.0126	0.0119	0.0118	0.0113	0.0111	0.0097	0.0095	0.0091	0.0087	0.0085	**0.0085**
		RMSE	0.0195	0.0155	0.0132	0.0151	0.0124	0.0119	0.0109	0.0108	0.0102	0.0097	**0.0095**
		THEIL-U	1.0335	1.0254	1.0147	1.0106	1.0089	1.0074	0.9887	0.9748	0.9700	0.9674	**0.9531**
	Spain	MAE	0.0104	0.0121	0.0102	0.0098	0.0095	0.0094	0.0088	0.0085	0.0084	0.0080	**0.0078**
		RMSE	0.0149	0.0167	0.0134	0.0120	0.0113	0.0105	0.0099	0.0096	0.0093	0.0092	**0.0087**
		THEIL-U	1.0084	1.0080	0.9841	0.9745	0.9730	0.9505	0.9501	0.9488	0.9416	0.9338	**0.9228**

Note: Best is the best individual predictors in terms of statistical performance respectively in the in-sample period (as outlined in table 5).

Table 7: DM and PT statistics

	Countries	**RW**	**ARMA**	**Best**	***v*SVR**	**LSVR**	**GA-*v*SVR**	**GA-LSVR**	**KH-*v*SVR**	**KH-LSVR**	**SC-*v*SVR**	**SC-LSVR**
F1	**Greece**	-10.22	-9.44	-8.33	-8.87	-7.54	-6.99	-6.15	-6.05	-5.48	**-1.17**	-
		1.25	**1.35**	**1.88**	3.15	4.28	5.36	7.48	8.12	9.15	9.99	10.22

	Ireland	-9.45	-8.18	-8.08	-7.13	-7.02	-6.77	-6.28	-6.01	-5.19	-4.25	-
		1.90	**1.93**	3.14	5.55	6.328	7.19	8.20	9.05	10.18	11.56	12.15
	Italy	-12.22	-10.48	-10.36	-10.02	-8.99	-7.14	-6.29	-5.84	-5.37	-5.08	-
		3.45	3.68	4.23	5.10	6.22	7.19	8.13	8.57	9.08	9.66	11.28
	Portugal	-10.33	-10.48	-9.99	-9.16	-8.19	-7.45	-6.22	-5.41	-4.27	**-1.57**	-
		2.15	3.08	3.99	4.28	5.66	6.08	7.14	7.79	8.19	8.54	9.63
	Spain	-8.86	-8.99	-8.12	-7.49	-6.80	-6.54	-5.84	-5.13	-4.08	-4.01	-
		2.99	3.02	3.25	4.20	5.64	5.69	6.36	7.12	7.86	8.59	8.97
F2	**Greece**	-11.42	-11.01	-10.25	-9.85	-8.79	-8.42	-7.15	-6.88	-5.32	-4.12	-
		1.15	2.18	4.42	4.99	6.16	7.22	7.36	8.54	8.93	9.68	10.35
	Ireland	-13.45	-12.32	-12.48	-10.48	-9.84	-8.47	-5.18	-5.07	-4.87	-3.22	-
		3.63	3.63	4.09	5.28	5.97	6.50	7.53	7.86	8.84	9.08	10.55
	Italy	-11.28	-10.99	-10.68	-9.45	-7.59	-6.60	-5.49	-3.28	-2.98	-2.78	-
		1.08	3.14	5.03	5.55	.64	6.99	7.05	7.66	8.45	9.55	11.02
	Portugal	-15.12	-14.44	-13.37	-10.78	-9.67	-8.87	-8.06	-7.34	-5.61	-3.37	-
		5.53	5.17	5.68	6.61	6.88	7.15	7.28	8.44	8.97	10.18	11.22
	Spain	-13.66	-12.99	-12.54	-11.77	-10.20	-9.98	-9.16	-9.04	-8.14	-7.74	-
		2.89	4.19	5.23	6.84	7.19	7.66	7.67	7.89	7.95	8.19	9.65
F3	**Greece**	-18.45	-17.60	-17.25	-15.23	-14.10	-12.08	-9.88	-8.47	-8.13	-6.55	-
		1.94	2.15	3.02	4.48	6.08	7.15	9.45	8.88	10.28	11.19	11.39
	Ireland	-16.22	-14.55	-13.48	-10.47	-10.08	-8.84	-8.09	-7.41	-5.64	**-1.22**	-
		3.56	3.07	3.48	5.15	6.48	6.89	9.08	9.85	9.05	10.36	11.53
	Italy	-15.05	-12.86	-17.10	-14.17	-12.33	-10.42	-9.41	-9.03	-8.84	-8.07	-
		4.45	4.87	5.12	5.66	6.38	6.87	7.89	7.95	8.15	9.45	10.35
	Portugal	-10.10	-10.30	-10.38	-9.91	-9.73	-9.06	-8.45	-8.17	-7.14	-4.22	-
		3.33	-3.49	3.85	4.42	5.25	6.03	6.88	7.45	7.96	8.45	9.60
	Spain	-9.45	-9.79	-9.40	-8.87	-8.05	-7.33	-7.18	-6.28	-5.57	-3.36	-
		2.96	3.47	3.45	3.86	3.49	5.58	7.28	8.05	8.49	9.72	9.86
F4	**Greece**	-10.18	-10.35	-9.99	-9.55	-9.31	-8.08	-7.15	-6.35	-4.28	-3.01	-
		0.85	**1.56**	2.86	3.44	4.25	4.48	5.06	6.18	6.66	7.45	8.93
	Ireland	-13.28	-12.47	-11.27	-10.29	-9.97	-9.49	-9.30	-8.14	-7.01	-4.14	-
		4.28	4.10	4.59	5.61	5.97	6.08	6.79	7.08	7.19	7.90	8.45
	Italy	-8.87	-9.07	-8.42	-7.56	-6.17	-5.55	-5.02	-4.84	-4.51	-2.95	-
		3.06	3.14	3.28	3.66	3.97	4.55	4.29	6.64	6.67	7.42	8.01
	Portugal	-11.03	-11.59	-11.08	-10.47	-10.22	-8.47	-7.55	-7.10	-6.33	-5.15	-
		4.44	4.90	5.18	6.04	6.87	7.79	7.86	8.84	8.94	9.08	10.25
	Spain	-10.14	-9.87	-9.25	-8.83	-8.02	-7.14	-5.37	-4.12	-3.84	**-1.09**	-
		3.56	4.10	4.22	5.36	6.77	7.05	7.14	8.99	9.62	10.12	12.24

Note: For every country, the first and second raw reports the DM and PT statistics respectively. The top left value corresponds to the DM statistics between the SC-LSVR and RW forecast. Every value and absolute value higher than 1.96 indicates significance of 5% for the PT and DM test respectively. The bold values are found insignificant at 5%.

Table A.1: The specification of the linear models

LINEAR MODELS	DESCRIPTION	TOTAL INDIVIDUAL FORECASTS

SMA (q)	$$E(R_t) = (R_{t-1} + ... + R_{t-q}) / q$$ *Where:* • *q=3...30*	28
EMA (q')	$$E(R_t) = \frac{R_{t-1} + (1-a')R_{t-2} + ... + (1-a')^{q'-1} R_{t-q'}}{a' + (1-a') + ... + (1-a')^{q'-1}}$$ *Where:* • *q'=3...30* • *a'*=2/(1+N_{days}), N_{days} *is the number trading days*	28
AR (q'')	$$E(R_t) = \beta_0 + \sum_{i'=1}^{q''} \beta_{i'} R_{t-i'}$$ *Where:* • *q''=1,…,24* • $\beta_0, \beta_{i'}$ *the regression coefficients*	24
ARMA (m', n')	$$E(R_t) = \bar{\varphi}_0 + \sum_{j'=1}^{m'} \bar{\varphi}_{j'} R_{t-j'} + \bar{a}_0 + \sum_{k'=1}^{n'} \bar{w}_{k'} \bar{a}_{t-k'}$$ *Where:* • *m', n'=1,..,15* • $\bar{\varphi}_0, \bar{\varphi}_{j'}$ *the regression coefficients* • $\bar{a}_0, \bar{a}_{t-k'}$ *the residual terms* • $\bar{w}_{k'}$ *the weights of the residual terms*	210

Note: The total number of individual inputs calculated is 329

Table A.2: Neural Network Design and Training Characteristics

PARAMETERS	MLP	RNN	HONN	PSN	ARBF-PSO
Learning algorithm	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Particle Swarm Optimization
Initialisation of weights	N(0,1)	N(0,1)	N(0,1)	N(0,1)	-
Hidden node activation function	$F(z_\psi) = 1/(1+e^{-z_\psi})$	$F(z_\psi) = 1/(1+e^{-z_\psi})$	$F(z_\psi) = 1/(1+e^{-z_\psi})$	$F(z_\psi) = \sum_{\psi=1}^{n''} z_\psi$	$F(z_\psi) = \exp\left(\frac{\lVert z_\psi - C' \rVert^2}{2\sigma'^2}\right)$
Output node activation function	$F(z_\psi) = \sum_{\psi=1}^{n''} z_\psi$	$F(z_\psi) = \sum_{\psi=1}^{n''} z_\psi$	$F(z_\psi) = \sum_{\psi=1}^{n''} z_\psi$	$F(z_\psi) = 1/(1+e^{-z_\psi})$	$F(z_\psi) = \sum_{\psi=1}^{n''} z_\psi$

Note: The input of every node is z_ψ, where ψ = 1… n'' and n'' is the number of nodes of the previous layer. The vector indicating the center of the Gaussian function is C' and σ' is the value indicating its width.

Table B.1: GA and KH training characteristics

Countries		GA				KH				
		F1	F2	F3	F4	Forecasting Exercise	F1	F2	F3	F4
Greece	**Population Size**	60	60	60	60	**Population Size**	60	60	60	60
	Maximum Generations	800	800	800	800	Δt **,** Z_{cr}	15.25, 0.61	20.38, 0.47	14.25, 0.94	20.32, 0.66
Ireland	**Population Size**	80	80	80	80	**Population Size**	80	80	80	80
	Maximum Generations	1000	1000	1000	1000	Δt **,** Z_{cr}	10.44, 0.45	32.04, 0.85	18.01, 0.85	20.38, 0.68
Italy	**Population Size**	50	50	50	50	**Population Size**	50	50	50	50
	Maximum Generations	1000	1000	1000	1000	Δt **,** Z_{cr}	21.33, 0.49	11.48, 0.79	21.14, 0.44	20.10, 0.62
Portugal	**Population Size**	80	80	80	80	**Population Size**	80	80	80	80
	Maximum Generations	500	500	500	500	Δt **,** Z_{cr}	18.47, 0.77	10.14, 0.42	21.96, 0.36	19.57, 0.81
Spain	**Population Size**	75	75	75	75	**Population Size**	75	75	75	75
	Maximum Generations	900	900	900	900	Δt **,** Z_{cr}	14.53, 0.87	22.15, 0.86	14.58, 0.72	21.44, 0.47
	Selection Type	Roulette Wheel Selection				**Foraging Speed**	0.02 ms^{-1}			
	Elitism	Best individual is kept in the next generation.				**Maximum Motion Speed**	0.01 ms^{-1}			

All countries	**Crossover Probability**	0.9	**Maximum Diffusion Speed**	[0.002, 0.010] ms^{-1}
	Mutation Probability	0.1	**Inertia Weights**	[0,1]

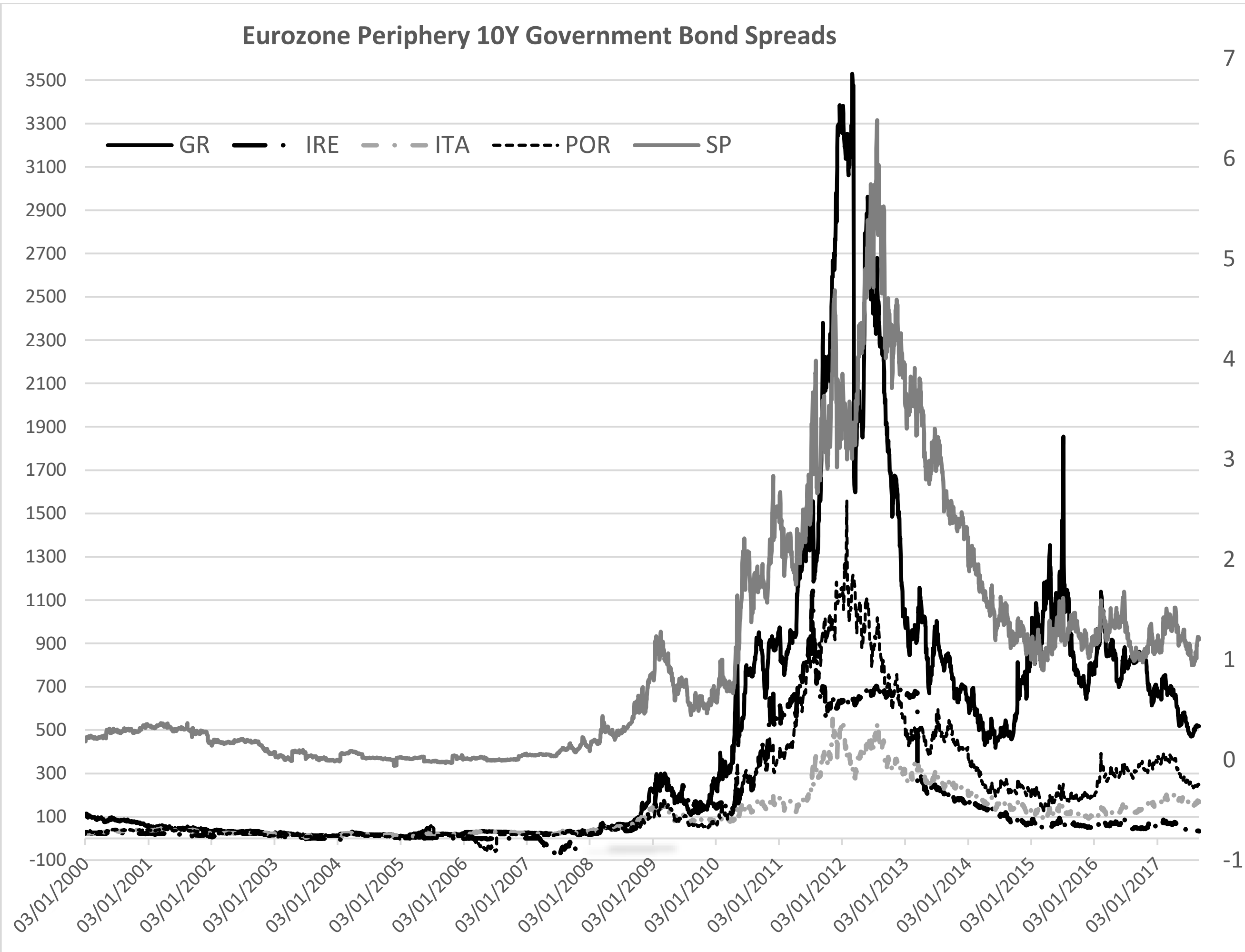
Note: The same population size and generations per forecasting exercise and country are applied to the SC training

Table C.1: Statistical performance measures

STATISTICAL PERFOMANCE MEASURES	DESCRIPTION
Mean Absolute Error	$MAE = \frac{1}{N'}\sum_{\tau=t+1}^{t+N'}\left\|E(R_\tau)-Y_\tau\right\|$ *with* Y_τ *being the actual value and* $E(R_\tau)$ *the forecasted value and* N' *the number of forecasts*
Root Mean Squared Error	$RMSE = \sqrt{\frac{1}{N'}\sum_{\tau=t+1}^{t+N'}(E(R_\tau)-Y_\tau)^2}$
Theil-U Statistic	$Theil-U = \frac{\sqrt{\frac{1}{N'}\sum_{\tau=t+1}^{t+N'}(E(R_\tau)-Y_\tau)^2}}{\sqrt{\frac{1}{N'}\sum_{\tau=t+1}^{t+N'}E(R_\tau)^2}+\sqrt{\frac{1}{N'}\sum_{\tau=t+1}^{t+N'}Y_\tau^2}}$

Figures

Figure 1: Eurozone periphery government bond spreads to German Bund



Note: Out-of-sample forecasting accuracy is calculated within the periods 2007-2008 and 2015- end of August 2017.

Figure 2: The flowcharts of i) GA-*v*SVR and ii) KH-LSVR

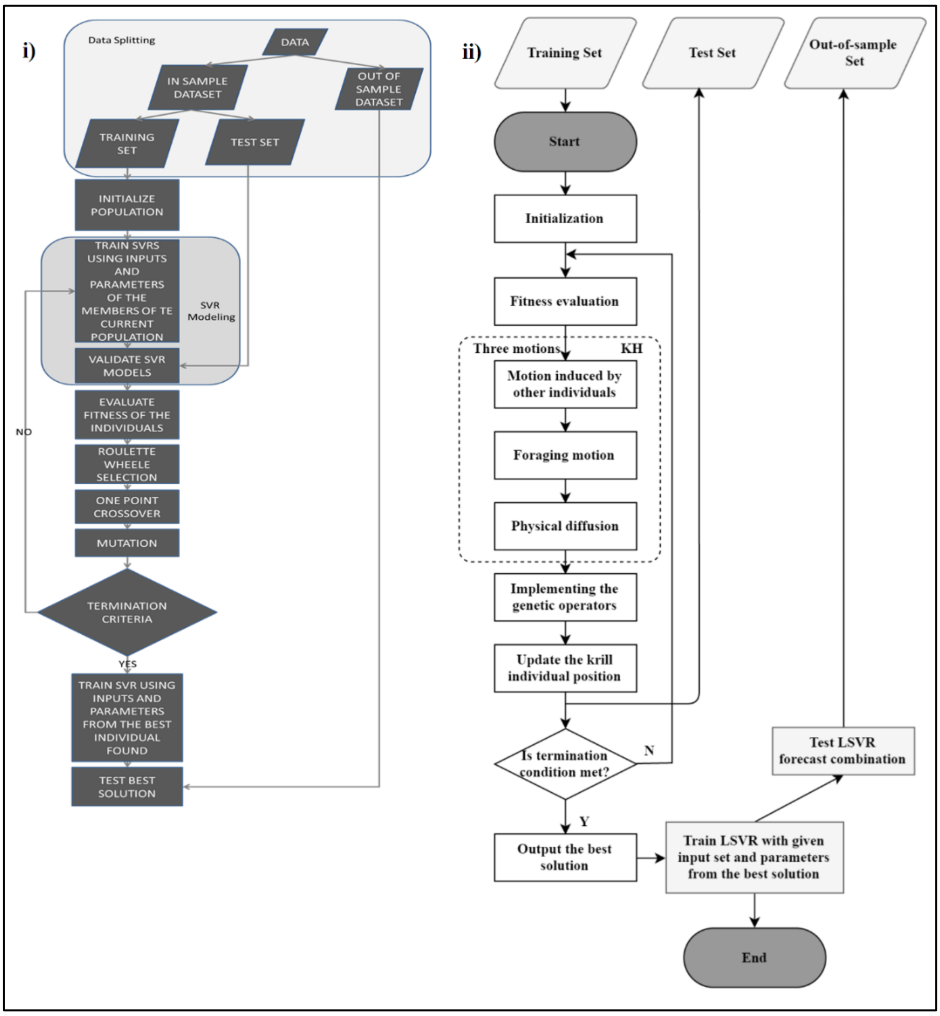
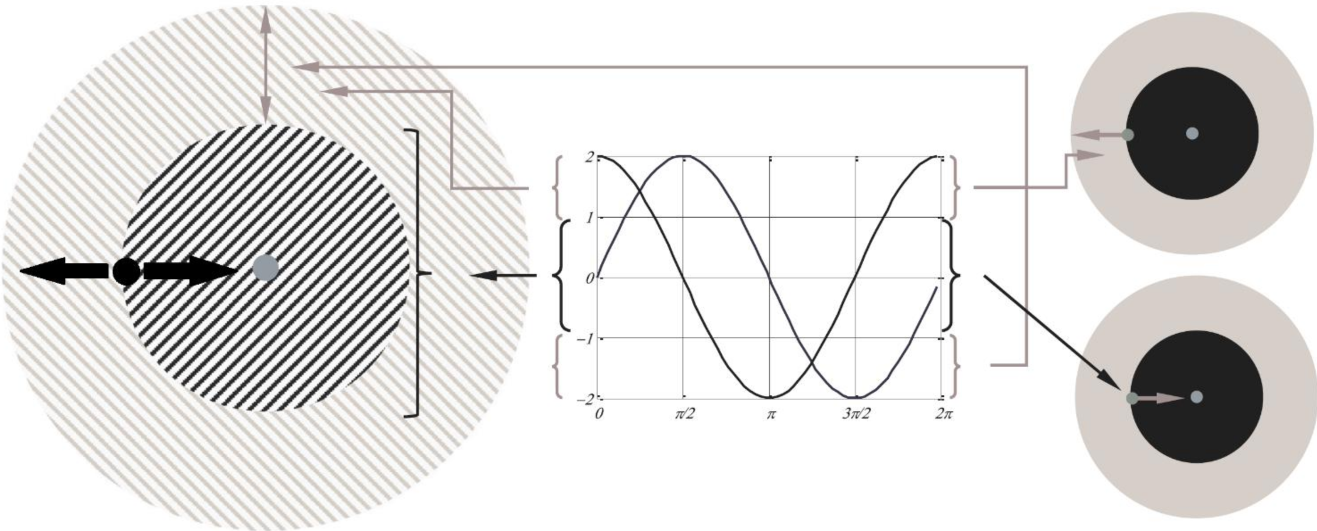


Figure 3: Sine Cosine algorithm in defining positions in the search space



Note: On the left hand side, the black and grey dot is the current solution and the destination position, respectively. The next position will be the inner (outer) circle for $r_1' < 1 (> 1)$. *The sine and cosine within the range [−2,2] allow a solution to go around (inside the space between them) or beyond (outside the space between them) the destination. It is illustrated this equation decreases the range of sine and cosine functions over the course of iterations, as the SC algorithm explores the search space when the ranges of sine and cosine functions are in (1,2] and [-2, -1). The search space is exploited when the ranges are in the interval of [-1,1] (Mirjalili, 2018).*

Figure B.1: KH algorithm pseudo-code

BEGIN

Step 1: Initialization

Initialize the generation counter GEN, the population of N_k krill randomly, V_F, D^{max} and RD^{max}.

Step 2: Fitness calculation

Calculate fitness for each krill according to its initial position P_j.

Step 3: While GEN < maximum generation criterion

Sort the population according to their fitness.

For j=1: N_k (all krill) **do**

Perform the following motion calculations:

Motion induced by other individuals (M_j)

Foraging motion (F_j)

Random Diffusion motion (RD_j)

Implement the genetic operators.

Update the krill position in the search space.

Calculate fitness for each krill according to its new position.

End for *j*

GEN=GEN+1

Step 4: End While

END

Figure B.2: SC algorithm pseudo-code

<u>**Sine Cosine Algorithm**</u>
Initialize
Set the random solutions of the search agents
Get the objective values of the initial random solutions
Get the position of the destination point
Get the best objective value so far
Iteration
While $t <= T''$ **do**
Update the range of sine and cosine
Update the position of solutions
Update the objective values
Update the position of the destination point
Update the best objective value so far
End while
Return *the best objective values is obtained as the global optimum along with the best solutions*