Inflation and Unemployment Forecasting with Genetic Support Vector Regression

Georgios Sermpinis

Charalampos Stasinakis

Konstantinos Theofilatos

Andreas Karathanasopoulos

February 2014

Abstract

In this paper a hybrid Genetic Algorithm – Support Vector Regression (GA-SVR) model in economic forecasting and macroeconomic variable selection is introduced. The proposed algorithm is applied to the task of forecasting the US inflation and unemployment. The GA-SVR genetically optimizes the SVR parameters and adapts to the optimal feature subset from a feature space of potential inputs. The feature space includes a wide pool of macroeconomic variables that might affect the two series under study. The forecasting performance of the GA-SVR is benchmarked with a random walk model, an Autoregressive Moving Average model, a Moving Average Convergence/Divergence model, a Multi-Layer Perceptron, a Recurrent Neural Network and a Genetic Programming algorithm. In terms of our results, GA-SVR outperforms all benchmark models and provides evidence on what macroeconomic variables can be relevant predictors of US inflation and unemployment in the specific period under study.

Keywords

Genetic Algorithms, Support Vector Regression, Forecasting, Inflation, Unemployment

*University of Glasgow Business School, University of Glasgow, Adam Smith Building, Glasgow, G12 8QQ, UK (georgios.sermpinis@glasgow.ac.uk)

**Bournemouth University, Business School, Executive Business Centre, Bournemouth, BH8 8EB, UK (cstasinakis@bournemouth.ac.uk)

***Pattern Recognition Laboratory, Dept. of Computer Engineering & Informatics, University of Patras, 26500, Patras, Greece (theofilk@ceid.upatras.gr)

**** University of East London, Royal Docks Business School, University Way, London, E16 2RD, UK (a.karathanasopoulos@uel.ac.uk)

INTRODUCTION

Developing highly accurate techniques for predicting the inflation and the unemployment is a crucial problem for economists and bankers. The forecasts for the unemployment and inflation play a crucial role in almost any monetary and policy decision process. As a result the empirical literature on forecasting macroeconomic variables is wealthy and extensive. Several statistical, technical and econometrical techniques have been applied to the problem with ambiguous results. Although researchers seem able to capture the pattern of the macroeconomic variables under normal market conditions, these models fail to provide accurate results during periods of recessions and economy shocks (Ager *et al.* (2009), Cogley *et al.* (2010) and Li (2012)). This can be explained by the fact that the relevant literature is dominated by linear models and/or models based on a fixed set of predictors. Inflation and unemployment though are affected by a number of different macroeconomic indicators and the underlying relation is likely to be changing depending on the state of the economy (D'Agostino *et al.*, 2013).

The novelty of our model lies in its ability to capture the asymmetries and nonlinearities in the given sample of predictors, select the optimal feature subsets and provide a single robust SVR forecast for the specific period and series under study. In that way, the hybrid Genetic Algorithm – Support Vector Regression (GA-SVR) sheds more light on the quest of nonlinear mapping of macroeconomic variables. From a technical point of view, our proposed model is superior to non-adaptive algorithms presented in the literature. It does not require analytic parameter calculation as Cherkassky and Ma (2004) propose, but also avoids time consuming optimization approaches (cross validation or grid search) that are used in similar applications (Lu *et al.*(2009) and Kim and Sohn (2010)). Additionally it is free from the data snooping bias, since all parameters of the GA-SVR model are optimized in a single optimization procedure.

This study introduces the GA-SVR and applies it to the task of forecasting the US inflation and unemployment. The GA-SVR model will attempt to capture the complex and non-linear behavior that dominates the two series. As potential inputs, the proposed algorithm will use a pool of 110 potential predictors. This will increase the model's flexibility and allow it to explore a large universe of potential relationships between the 110 predictors and the US inflation and unemployment. It is also worth noting that the proposed algorithm is fully adaptive something that excludes any bias from the practitioner. The

selection of the proposed model's inputs and parameters is based on a GA algorithm, while the pool of potential inputs is only limited by the data availability.

The proposed methodology intends to contribute to the growing literature that studies the utility of technical analysis and computational intelligence techniques in forecasting. The technical approach suggests that market action discounts all available information, prices move in trends and history tends to repeat itself (Murphy, 1999). Computational intelligent techniques, like Neural Networks (NNs), Genetic Algorithms (GAs) and Support Vector Machines (SVMs), attempt to model human cognitive abilities, such as reasoning, learning and decision making, which makes them useful in forecasting applications (Kim (2006), Kim and Sohn (2010) and Huang (2012)). In this paper, a random walk model (RW), an Autoregressive Moving Average model (ARMA) and a Moving Average Convergence/Divergence model (MACD), a Multi-Layer Perceptron (MLP), a Recurrent Neural Network (RNN) and a Genetic Programming (GP) algorithm will be utilized as benchmarks. This benchmark selection serves the purpose of this study. Namely, it allows GA-SVR to compete with three traditional models used in technical analysis (RW, ARMA, MACD) and three common computational intelligent methods (MLP, RNN, GP).

All models will be applied to the task of forecasting the US inflation and unemployment from January 1974 to December 2012. The periods from January 1997 to December 2000, January 2001 to December 2004, January 2005 to December 2008 and January 2009 to December 2012 will act as out-of-samples by rolling forward the estimation by four years. The optimal parameters of GA-SVR are recalculated in every exercise. From an econometric perspective the rolling forward estimation will add validity to the results of the forecasting exercise. From an economic perspective, the unique architecture of the GA-SVR model will allow to study if variables that are significant in forecasting the inflation and the unemployment in the pro-crisis period (1997-2004) remain significant in crisis and after crisis period (2005-2012). The above selection of the out-of-sample periods allow us also to observe if the forecasting power of our models is reduced during the recession period, as it is repeatedly reported in the recent relevant literature.

During the last decades the dynamics of US inflation have changed considerably. Inflation forecasters have implemented a wide variety of multivariate models, such as Cogley and Sargent (2005) and Cogley *et al.* (2010). These models attempt to outperform simple univariate models like the Atkeson-Ohanian's (2001)

random walk model or time varying models with unobserved components as presented by Stock and Watson (2007). Their success though is inconsistent and their inflation forecasts are unstable. A concise survey of the instability of these models is given by Stock and Watson (2009). Stock and Watson (2003, 2004) also propose that the best predictive performance is attained through simple averaging of inflation forecasts derived from a very large number of models. McAdam and McNelis (2005) perform an inflation forecasting exercise in US, Japan and Euro area, where GAs are combined with NNs in order to achieve optimal non-linear Phillips curve specifications. Based on their results, the authors conclude that the payoff of the NN strategy comes in periods of structural change and uncertainty. Ang *et al.* (2007) compare and combine the forecasting power of several linear and non-linear models with survey-based measures. Their study shows that the use of surveys' information can lead to superior individual forecasts on the US infaltion. Inoue and Kilian (2008) apply the method of bootstrap forecast aggregation (bagging) to the task of forecasting the US CPI. The empirical evidence confirms the superiority of this method compared to the Bayesian model averaging or Bayesian shrinkage estimators used by other researchers (see amongst others Groen *et al.* (2010), and Stock and Watson (2012)).

Forecasting unemployment rates is also a well documented case study in the literature (Rothman (1998), Montgomery *et al.* (1998)). Swanson and White (1998) forecast several macroeconomic time series, including US unemployment, with linear models and Neural Networks (NNs). In their approach, NNs present promising empirical evidence against the linear VAR models. Moshiri and Brown (2004) apply a back-propagation model and a generalized regression NN model to estimate post-war aggregate unemployment rates in the USA, Canada, UK, France and Japan. The out-of-sample results confirm the forecasting superiority of the NN approaches against traditional linear and non-linear autoregressive models. Smooth transition vector error-correction models are also used to forecast the unemployment rates, as in the non-Euro G7 countries' study of Milas and Rothman (2008). The proposed model outperforms the linear autoregressive benchmark and improves significantly the forecasts of the US and UK unemployment rate during business cycle expansions. Wang (2010) combines several rival individual US unemployment forecasts with directed acyclical graphs. The results indicate that models that are not directly causally linked can be combined to project a more accurate composite forecast. Chua *et al.* (2012) present a latent variable approach to the same forecasting task. Their model exploits the time series properties of US

unemployment, while satisfying the economic relationships specified by Okun's law and the Phillips curve. The specification is advantageous since it provides an unemployment forecast consistent with both theories, but at the same time is less computational demanding than equivalent atheoretical models like VAR and BVAR. Finally, Olmedo (2013) performs a competition between non-linear models to forecast different European unemployment rate time series. The best results are provided by a vector autoregressive and baricentric predictor, but as the forecasting horizon lengthens the performance deteriorates.

The rest of the paper is organized as follows. Section 2 describes the dataset used for this study, while a brief description of the benchmark models is given in Section 3. Section 4 summarizes the theoretical background needed for the complete understanding of our proposed methodology. In Section 5 follows the complete description of the hybrid GA-SVR model. The empirical results are presented in Sections 6. Finally, some concluding remarks are provided in Section 7.

2. DATA DESCRIPTION

This paper implements two forecasting exercises with monthly data over the period of January 1974 to December 2012. The first exercise attempts to forecast the percentage change in the US inflation. As a proxy for the US inflation, we use the US Consumer Price Index (CPI). The second one focuses on predicting the percentage change of the US unemployment (UNEMP). Figure 1 below presents the two series under study in levels.

[Insert Figure 1]

Following similar studies (Wright (2009) and Koop and Korobilis (2012)), we select eleven predictors that can explain the economic premises of inflation and unemployment or are found to be useful in forecasting them. The pool of our potential inputs includes the first ten autoregressive terms of these predictors. Thus, the feature space consists of hundred ten series of monthly percentage changes. All series are seasonally adjusted, where applicable. The sources of our data are the Federal Reserve Bank of St. Louis (FRED) and Bloomberg (BLOOM). Table 1 below summarizes the list of variables used in this application[1].

[1] In this application we use monthly frequency data. This constrains us from including other relevant predictors that are available quarterly. Nonetheless, the GA-SVR uses the majority of the important macroeconomic indicators, as explored by landmark similar studies (i.e. Stock and Watson (2012))

[Insert Table 1]

3. BENCHMARK FORECASTING MODELS

The proposed GA-SVR model is benchmarked with a Random Walk model (RW), an Autoregressive Moving Average model (ARMA), a Moving Average Convergence/Divergence Model (MACD), a Multi-Layer Perceptron (MLP), a Recurrent Neural Network (RNN) and a Genetic Programming (GP) algorithm. This section provides a brief description of these models.

3.1 Random Walk Model (RW)

The random walk model (RW) is a process where the current value of a variable is calculated from the past value plus an error term. The error term follows the standard normal distribution. The specification of the model is:

$$\hat{Y}_t = Y_{t-1} + e_t, \quad e_t \sim N(0,1) \tag{1}$$

Where $\hat{Y}_t$ is the forecasted inflation/unemployment for period t and Y_{t-1} is the actual inflation/unemployment of period *t-1*.

The RW is a non-stationary process with a constant mean, but not a constant variance.

3.2 Auto-Regressive Moving Average Model (ARMA)

An ARMA model embodies autoregressive and moving average components and can be specified as below:

$$\hat{Y}_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t - w_1 \varepsilon_{t-1} - w_2 \varepsilon_{t-2} - \ldots - w_q \varepsilon_{t-q} \tag{2}$$

Where:

- $\hat{Y}_t$ is the forecasted inflation/unemployment at time t
- $Y_{t-1}, Y_{t-2}, \ldots Y_{t-p}$ are the lagged actual inflation/unemployment values

- $\varphi_0,\varphi_1,\ldots,\varphi_p$ are the regression coefficients
- ε_t is the error term
- $\varepsilon_{t-1},\varepsilon_{t-2},\ldots,\varepsilon_{t-q}$ are the previous values of the error terms
- $w_1, w_2,\ldots,w_q$ are the error weights

The ARMA models are selected using the correlogram and the information criteria in the in-sample period as a guide. The back-casting technique is used to obtain pre-sample estimates of the error terms (Box and Jenkins, 1976). The null hypotheses that all coefficients (except the constant) are not significantly different from zero and that the error terms are normally distributed are rejected at the 95% confidence interval.

3.3 Moving Average Convergence/Divergence Model (MACD)

A moving average model is defined as:

$$M_t = \left(Y_{t-1} + Y_{t-2} + \ldots + Y_{t-n+1}\right) / n \tag{3}$$

Where:

- M_t: moving average at time t
- n: the number of terms in the moving average
- $Y_{t-1,\ldots,} Y_{t-n+1}$: the actual inflation/unemployment at periods $t-1,\ldots,t-n+1$

The MACD line derived by two moving average series with different lengths (short and long) is used to forecast the two series under study. The short and long terms used in the estimation of the moving averages are commonly determined based on the forecaster's judgement and practical previous knowledge. In our case, the combinations that perform best over the in-sample sub-period are retained for out-of-sample evaluation.

3.4 Neural Network Architectures (NNs)

The Multi-Layer Perceptron (MLP) and the Recurrent Neural Network (RNN) are the two traditional NNs used as benchmarks for this forecasting application. Both these architectures have at least three layers. The first layer is called the input layer (the number of its nodes corresponds to the number of explanatory variables). The last layer is called the output layer (the number of its nodes corresponds to the number of response variables). An intermediary layer of nodes, the hidden layer, separates the input from the output

layer. Its number of nodes defines the amount of complexity the model is capable of fitting. In addition, the input and hidden layer contain an extra node called the bias node. This node has a fixed value of one and has the same function as the intercept in traditional regression models. Normally, each node of one layer has connections to all the other nodes of the next layer. The training of the network (which is the adjustment of its weights in the way that the network maps the input value of the training data to the corresponding output value) starts with randomly chosen weights and proceeds by applying a learning algorithm called backpropagation of errors (Shapiro (2000)). The iteration length is optimised by maximising a fitness function in the test dataset.

Unlike MLPs, RNNs have an activation feedback which embodies short-term memory. In other words, the RNN architecture can provide more accurate outputs because the inputs are (potentially) taken from all previous values. Tenti (1996) notes that RNNs need more connections and memory than standard back-propagation networks. However, RNNs can yield better results in comparison with simple MLPs due to the additional memory inputs. For more information on MLPs and RNNs see Sermpinis *et al.* (2013).

There is no formal theory behind the selection of the inputs of NNs. For that reason, we conduct neural networks experiments and a sensitivity analysis on a pool of autoregressive terms of all available series in the in-sample dataset. The aim is to select as inputs those sets of variables that provide the best statistical performance for each network in the in-sample period. Based on the guidelines (Lisboa and Vellido (2000) and Zhang (2009)) we experiment with the first fifteen autoregressive terms of each forecasted series in all in-sample periods. More details about the design and training characteristics our NNs are included in Appendix A.

3.5 Genetic Programming Algorithm (GP)

GP algorithms are a class of Genetic Algorithms (GAs) and the intuition behind this technique is the Darwinian principle of reproduction and survival of the fittest. GP applies the Darwinian theory of evolution to a population of computer programs of varying sizes and shapes, which run in various environments in order to produce forecasts at a high level of accuracy (Chen, 2002). Dissimilar to NNs, GP

creates an initial population of models and evolves it using genetic operators, in order to calculate the mathematical expression which best fits the specified data input in the system.

Our GP application evolves tree-based structures that present models (sub-trees) of input – output. It utilizes formulas to evolve algebraic expressions that enable the analysis and optimization of results in a genetic tree structure. This structure consists of nodes, which are essentially functions that perform actions within this structure. The maximum tree depth is the maximum length of each model (of each tree structure) and it depends on the functions and terminals of each individual model. The design phase of our GP application focuses primarily on execution time optimization and then on limiting the 'bloat effect', a similar issue as overfitting in NNs.

GP reproduces newer models replacing the weaker ones in the population according to their fitness. In our case, the fitness is measured by Mean Squared Error (MSE) of the forecasted value and the actual value of each month. Obviously the lowest MSE is considered as a criterion of better fitness. Then, the best models (tournament winners) are exposed to two genetic operators, known as mutation and crossover. This genetic procedure produces superior offsprings that will replace the worst models (tournament losers) and rearrange the initial population for the next iteration. This is constrained by the size of the models, namely the tournament size, and their goodness of fit. The iterations stop and the final forecast results are obtained, when our model reaches the critical value of the termination criterion. The termination criterion is in general arbitrarily chosen, but our choice is based on optimizing the statistical performance with the least possible bloat effect in the in-sample period. For more details on the functionality aspects of GP and the genetic operators see Koza and Andre (1996).

The figure below describes the structure of a typical GP algorithm.

[Insert Figure 2]

The parameters of our GP application are defined based on which model presents the best statistical performance in the in-sample sub-period and are presented in Appendix A.

4. THEORETICAL BACKGROUND

In this section follows a short theoretical background on Support Vector Regression (SVR), Genetic Algorithms (GAs) and the issues of parameter and feature subset selection.

4.1 The ε-SVR and v-SVR

Support Vector Machines (SVMs) are non-linear algorithms used in supervised learning frameworks in order to solve classification problems. SVM processes belong to the general category of kernel methods (Scholkopf and Smola (2002)). Their main advantage is that they can generate non-linear decision boundaries through linear classifiers. Another advantage is that the practitioner can apply kernel functions to data that their vector space is not fixed in terms of dimensions. The SVMs can be used in regression problems by implementing the ε-sensitive loss function by Vapnik (1995). This function established SVRs as a robust technique for constructing data-driven and non-linear empirical regression models.

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

$$f(x) = w^T \varphi(x) + b \tag{4}$$

where w and b are the regression parameter vectors of the function and $\varphi(x)$ is the non-linear function that maps the input data vector x into a feature space where the training data exhibit linearity (see figure 3).

[Insert Figure 3]

The ε-sensitive loss L_ε function (see figure 3b) finds the predicted points that lie within the tube created by two slack variables ξ_i, ξ_i^*:

$$L_\varepsilon(x_i) = \begin{cases} 0 \; if \; |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon \; if \; other \end{cases}, \varepsilon \geq 0 \tag{5}$$

The goal is to solve the following argument:

$$\text{Minimize} \quad C\sum_{i=1}^{n}(\xi_i + \xi_i^*) + \frac{1}{2}\|w\|^2 \text{ subject to } \begin{Bmatrix} \xi_i \geq 0 \\ \xi_i^* \geq 0 \\ C > 0 \end{Bmatrix} \text{and } \begin{Bmatrix} y_i - w^T\varphi(x_i) - b \leq +\varepsilon + \xi_i \\ w^T\varphi(x_i) + b - y_i \leq +\varepsilon + \xi_i^* \end{Bmatrix} \tag{6}$$

The above quadratic optimization problem is transformed in a dual problem and its solution is based on the introduction of two Lagrange multipliers a_i, a_i^* and mapping with a kernel function $K(x_i, x)$:

$$f(x) = \sum_{i=1}^{n} (a_i - a_i^*) K(x_i, x) + b \text{ where } 0 \le a_i, a_i^* \le C \tag{7}$$

The application of the kernel function transforms the original input space into one with more dimensions, where a linear decision border can be identified. Factor *b* is computed following the Karush-Kuhn-Tucker conditions. A detailed mathematical explanation of the above solution can be found in Vapnik (1995). Support Vectors (SVs) are called all the x_i that contribute to equation (7), thus they lie outside the ε-tube, whereas non-SVs lie within the ε-tube[2]. Increasing ε leads to less SVs' selection, whereas decreasing it results to more 'flat' estimates. The norm term $\|w\|^2$ characterizes the complexity (flatness) of the model and the term $\left\{ \sum_{i=1}^{n} (\xi_i + \xi_i^*) \right\}$ is the training error, as specified by the slack variables. Consequently the introduction of the parameter *C* satisfies the need to trade model complexity for training error and vice versa (Cherkassky and Ma, 2004). In general, both terms cannot be minimal or close to zero at the same time. The SVR algorithm estimates the *w* and *b* of the linear function of equation (4) with the predefined ε and *C*, in order the resulting regression function to achieve good generalization ability. This result should not be too complex and at the same time avoid many training errors. If this balance is achieved, then the SVR offers a solution to the over-fitting problem. The parameter ε, though, takes nonnegative unconstrained values, which makes its optimal setting very challenging.

The *v*-SVR algorithm can be used to make this task easier. This alternative SVR approach encompasses the ε parameter in the optimization process and controls it with a new parameter $v \in (0,1)$. In *v*-SVR the optimization problem transforms to:

$$\text{Minimize } C\left(v\varepsilon + \frac{1}{n} \sum_{i=1}^{n} (\xi_i + \xi_i^*) \right) + \frac{1}{2} \|w\|^2 \text{ subject to } \begin{Bmatrix} \xi_i \ge 0 \\ \xi_i^* \ge 0 \\ C > 0 \end{Bmatrix} \text{and } \begin{Bmatrix} y_i - w^T \varphi(x_i) - b \le +\varepsilon + \xi_i \\ w^T \varphi(x_i) + b - y_i \le +\varepsilon + \xi_i^* \end{Bmatrix} \tag{8}$$

The methodology remains the same as in ε-SVR and the solution takes a similar form:

[2] A SV is either a boundary vector ($(a_i - a_i^*) \in [-C/n, C/n], \xi_i = \xi_i^* = 0$) or an error vector ($a_i, a_i^* = C/n \text{ and } \xi_i, \xi_i^* > 0$).

$$f(x) = \sum_{i=1}^{n} (a_i - a_i^*) K(x_i, x) + b \text{ where } 0 \le a_i, a_i^* \le \frac{C}{n} \quad (9)$$

Based on the '*v*-trick', as presented by Scholkopf *et al.* (1999), increasing ε leads to the proportional increase of the first term of $\left\{ v\varepsilon + \frac{1}{n} \sum_{i=1}^{n} (\xi_i + \xi_i^*) \right\}$, while its second term decreases proportionally to the fraction of points outside the ε-tube. So *v* can be considered as the upper bound on the fraction of errors. On the other hand, decreasing ε leads again to a proportional change of the first term, but also the second term's change is proportional to the fraction of SVs. That means that ε will shrink as long as the fraction of SVs is smaller than *v*, therefore *v* is also the lower band in the fraction of SVs. For a more detailed mathematical analysis of the above solutions see Vapnik (1995).

Although SVR has emerged as a highly effective technique for solving non-linear regression problems, designing such a model can be impeded by the complexity and sensitivity of selecting its parameters (C, ε or *v* and kernel function parameter). SVR's performance depends on all parameters being set optimally. Numerous approaches for this optimization have been presented in literature, which can be summarized in the following trends:

- Setting ε as a non-negative constant for convenience ($\varepsilon=0$ or equal to a very small value) (Trafalis and Ince (2000)).
- Using data-driven approaches and standard parameterization of the SVR solution (Cherkassky and Ma, 2004).
- Estimating the parameters with the cross-validation technique (Cao *et al.*(2003) and Duan *et al.* (2003)).
- Controlling ε with *v*-SVR (Scholkopf *et al.* (1999)).

4.2 Feature Selection and Genetic Algorithms (GAs)

Feature selection is an optimization problem that refers to the search over a space of possible feature subsets in order to find those that are optimal with respect to specific criteria. Such a problem requires a search strategy that picks the feature subsets and an evaluation method that tests their goodness of fit. Many searching strategies have been proposed in literature, but those who seem to attract more attention

are the randomized searches, where probabilistic steps are applied (Sun *et al.*, 2004). Genetic Algorithms (GAs) are commonly used in such cases (Siedlecki and Sklansky (1989))

GAs, formerly introduced by Holland (1975), are search algorithms inspired by the principle of natural selection. They are useful and efficient if the search space is big and complicated or there is not any available mathematical analysis of the problem. A population of candidate solutions, called chromosomes, is optimized via a number of evolutionary cycles and genetic operations, such as crossovers or mutations. Chromosomes consist of genes, which are the optimizing parameters. At each iteration (generation), a fitness function is used to evaluate each chromosome, measuring the quality of the corresponding solution, and the fittest chromosomes are selected to survive. This evolutionary process is continued until some termination criteria are met. In general, GAs can deal with large search spaces and do not get trapped in local optimal solutions like other search algorithms.

5. HYBRID GA-SVR MODEL

In this section we present our hybrid Genetic Algorithm – Support Vector Regression (GA-SVR) model for optimal SVR parameter and macroeconomic variable selection. The proposed model genetically searches over a feature space (the pool of macroeconomic predictors as in Table 1) and then provides a single optimized SVR forecast for each series under study. In order to achieve this we use a simple GA where each chromosome comprises *feature genes* that encode the best feature subsets and *parameter genes* that encode the best choice of parameters.

The lack of information on the noise of the training datasets makes the *a priori* ε-margin setting of ε-SVR a difficult task. In order to overcome this and decrease the computational demands of our methodology, we decide to implement the RBF v-SVR approach in our hybrid GA-SVR model (see Section 4.1). A RBF kernel is in general specified as:

$$K(x_i, x) = \exp(-\gamma \left\| x_i - x \right\|^2), \gamma > 0 \tag{10}$$

where γ represents the variance of the kernel function. Consequently, the parameters optimized by the GA are C, v and γ.

The GA of our hybrid methodology uses the *one-point crossover* and the *mutation operators*. The one-point crossover creates two offspring from every two parents. The parents and a crossover point c_x are selected at random. The two offsprings are made by both concatenating the genes that precede c_x in the first parent with those that follow (and include) c_x in the second parent. The probability for selecting an individual as a parent for the crossover operator is called *crossover probability* and in our application is set to 0.85. Having a high crossover probability enables our model to keep some population for the next generation, hoping to create better new chromosomes from good parts of the old chromosomes. The offspring produced by the crossover operator replaces their parents in the population. On the other hand, the mutation operator places random values in randomly selected genes with a certain probability named as *mutation probability*. This operator is very important for avoiding local optima and exploring a larger surface of the search space. This probability is set to 0.15 in order to prevent our algorithm from performing a random search.

For the selection step of the GA, *the roulette wheel selection process* was used (Holland (1995)). In roulette wheel selection chromosomes are selected according to their fitness. The better the chromosomes are, the more chances to be selected they have. In our approach, *elitism* is used to raise the evolutionary pressure in better solutions and to accelerate the evolution. In that way, we assure that the best solution is copied without changes to the new population, so the best solution found can survive at the end of every generation. Similarly to the NNs, the GA-SVR model requires training and test subsets to validate the goodness of fit of each chromosome. The population of chromosomes is initialized in the training sub-period. The optimal selection of chromosomes is achieved when their forecasts minimize the MSE in the test-sub period (the last four years of the in-sample). Then, the optimized parameters and selected predictors of the best solution are used to train the SVR and produce the final optimized forecast, which is evaluated over the out-of-sample period. In genetic algorithm modeling, though, fitness functions need to be increasing functions. Therefore, our algorithm is minimizing the MSE by maximizing the following function:

$$\text{Fitness} = 1/(1+\text{MSE}) \tag{11}$$

The size of the initial population is set to 400 chromosomes while the maximum number of generations is set to 5000. Our algorithm though terminates when the number of generations is 3000 on average. This number must be reached in combination with a termination method that stops the evolution, when the population is deemed as converged. The population is deemed as converged when the average fitness across the current population is less than 5% away from the best fitness of the current population. More specifically, when it is less than 5% the diversity of the population is very low and evolving it for more generations is unlikely to produce different and better individuals than the existing ones or the ones already examined by the algorithm in previous generations. The summary of our GA's characteristics is presented in the following table.

[Insert Table 2]

The flowchart of the GA-SVR methodology is depicted in detail in figure 4:

[Insert Figure 4]

6. EMPIRICAL RESULTS

The empirical results of the proposed methodology are presented in this section. Here the adaptive selection of the macroeconomic variables is described for each forecasting exercise. Then, the statistical evaluation of the optimized GA-SVR forecasts follows in regard to its benchmarks and a robustness test.

6.1. Selection of Predictors

The macroeconomic contribution of this paper is based on the fact that GA-SVR algorithm is able to genetically adapt in the most relevant predictors for the US inflation and unemployment. The selected variables for both forecasting exercises and all out-of-sample periods are presented Table 3. This selection corresponds to the chromosomes that provide the best forecasts of CPI and UNEMP.

[Insert Table 3]

Concerning the inflation exercise, the results show that the algorithm retains maximum seventeen time series from the overall one hundred ten as inputs. During the 1997-2000 and 2001-2004 those series are autoregressive terms up to the order of five of eight variables, the HOUSE, INDP, M1, EMPL, PCE, PI,

WAGE and DJIA. For the 2005-2008 sub-period, the GA-SVR discards the HOUSE, EMPL, PI, DJIA variables and adds the TBILL. It seems that in this period, there is a structural break for inflation and the set of variables that have explanatory value has changed. In the last sub-period, the algorithm selects eight inputs (autoregressive terms of the INDP, M1, PCE and PI variables). We note that the second lag of M1 is always selected as input from our model. The different set of inputs in each sub-period reveals that inflation is difficult to predict and models with a constant or a limited set of independent variables will have no value in the long-run.

In the case of unemployment, the GA-SVR selects more macroeconomic variables and respective lags than in the inflation exercise. This might indicate that forecasting the US unemployment is a more complex and demanding task that requires a larger set of independent variables than the US inflation. From the set of potential inputs, we note that the second lag of WAGE in always selected and that the first four autoregressive lags of IND are a popular choice from our algorithm. The set of inputs changes for each sub-period. This indicates that structural breaks dominate unemployment forecasting as the set of explanatory variables is constantly changing.

INDP, M1 and PCE are the only common economic indicators in the four periods under study for both inflation and unemployment. In each out-of-sample, though our algorithm accepts different autoregressive lags of them as common inputs. For example, the first three autoregressive terms of INDP, the third of M1 and the fourth of PCE are common predictors of inflation and unemployment during 1997-2000. In the period of 2001-2004 the algorithm selects the third and fourth lag of INDP, the second lag of M1 and the fourth lag of PCE for forecasting both series. Similarly during 2005-2008, the first two autoregressive terms of INDP, the third of M1 and the first of PCE qualify as potential predictors for both CPI and UNEMP. Finally, in 2009-2012 the first lag of INP and the second of M1 and PCE are kept in the inputs' pool for each exercise. The exchange rates are found to be irrelevant for all the out-of-samples. The HOUSE variable is pooled during 1997- 2000 and 2001-2004, but discarded for the periods 2005-2008 and 2009-2012 (during and after the US housing bubble burst). It is interesting to note that autoregressive terms of our potential inputs with order of six or higher have no value for our model. More specifically for the

periods 2005-2008 and 2009-2012 the majority of the selected inputs are first and second autoregressive lags of the respective macroeconomic variables.

From a technical point of view, the selection process of GA-SVR does not suffer from over-fitting since in both exercises and all out-of-sample periods the parameter γ is relatively small (see Table 3 note). Small values of γ are in general welcome because they result in smoother marginal decisions. The restrictiveness of the SVR 'tube' though depends on all three parameters and therefore it is difficult to assess if our model is more adaptive in the CPI or UNEMP forecasting exercise. In general, our algorithm requires more time (iterations[3]) to converge in UNEMP optimal chromosomes than CPI ones.

6.2 Statistical Performance

As it is standard in the literature, in order to evaluate statistically our forecasts, the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Theil-U statistics are computed. The mathematical formulas of these statistics are presented in Appendix B. For all four statistical measures retained, the lower the output the better the forecasting accuracy of the model concerned. The in-sample statistical performances of our models for the CPI and UNEMP during all relevant periods are presented in Table 4 below.

[Insert Table 4]

From Table 4, we note that GA-SVR presents the best in-sample statistical performance for both series under study for all the statistical measures. The second best model is GP. It outperforms both NNs and the traditional strategies, but it is always inferior to the GA-SVR. Although the models perform differently during each period in both forecasting tasks, the ranking of our models remains the same throughout 1974-2008. The worst performances are observed in the 1982-2004 and 1986-2008. Table 5 summarizes the statistical performances of the models in the relevant out-of-sample periods for CPI and UNEMP.

[Insert Table 5]

[3] *Iterations = Population * Generations*

From the results of Table 5, it is obvious that GA-SVR retains its forecasting superiority for the statistical measures applied in all four out-of-sample sub-periods. The statistical ranking of our models remains consistent with our in-sample results. Once more, the GP outperforms the MLP and RNN, while traditional models like RW, ARMA and MACD present the worst forecasts in term of statistical accuracy. The worst statistical results are attained in the 2005-2008 and 2009-2012 sub-periods. It seems the US subprime crisis increases the difficulty in this forecasting exercise. Nonetheless, the performance of the GA-SVR seems robust in both periods of economic instability.

In order to further verify the statistical superiority of our best proposed architecture, we calculate the Modified Diebold-Mariano (MDM) statistic for testing the equality of forecast accuracy, as proposed by Harvey *et al.* (1997). The MDM statistic is an extension of the Diebold-Mariano (1995) statistic (DM) and its formula is the following:

$$MDM = T^{-1/2}\left[T + 1 - 2k + T^{-1}k\left(k-1\right)\right]^{1/2} DM \tag{12}$$

where T the number of the out-of-sample observations and k the number of the step-ahead forecasts.

The use of MDM is common practice in forecasting because it is found to be robust in assessing the significance of observed differences between the performances of two forecasts (see Barhoumi *et al.* (2012) and Hassani *et al.* (2012)). MDM also overcomes the problem of over-sized DMs in moderate samples. The statistic is measured in each out-of-sample period and the MSE and the MAE are used as loss functions. The MDM test follows the Student's *t*-distribution with *f-1* degrees of freedom, where *f* is the number of forecasts. Table 6 below presents the values of the statistics, comparing the GA-SVR with its benchmarks.

[Insert Table 6]

From the above table it is obvious that the MDM null hypothesis of forecast encompassing is rejected for all comparisons and for both loss functions at the 1% confidence interval. The statistical superiority of the

GA-SVR forecasts is also confirmed as the realizations of the MDM statistic are always negative[4]. GP is found to have the closest forecasts with the GA-SVR model and remains the second best model in statistical terms. From the MDM values it is safe to claim that there is no conclusive evidence of encompassing between the GA-SVR forecasts of inflation and unemployment and their benchmarks. Finally, the in-sample and out-of-sample results indicate that the models implementing genetic approaches, GP and GA-SVR, project in general more accurate forecasts in comparison with popular NN techniques (MLP, RNN) or tradition linear models (RW, ARMA, MACD).

In general, the success of GA-SVR against all the benchmarks and especially the computational intelligent ones is promising. The modelling of MLP, RNN and GP, though, is not so advanced compared to the proposed hybrid methodology. From that aspect, there is space for future work that would set the performance standards of the benchmarks even higher. For example, GA-SVR could be included in a forecasting competition with other state-of-the-art NNs, such as Error Correction Neural Networks (ECNNs) or Historical Consistent Neural Network (HCNNs). These architectures are superior to the traditional RNNs and have the ability to handle missing information and be more adaptive to shocks (Zimmermann *et al.*, 2012). Another extension of this work could be the statistical evaluation of the model on longer forecast horizons and in other countries than US. This could amplify the economic interpretability of the selected predictors and identify the policy implications of GA-SVR in a country per country case study.

7. CONCLUDING REMARKS

The motivation of this paper is to introduce a hybrid Genetic Algorithm – Support Vector Regression (GA-SVR) model in economic forecasting and macroeconomic variable selection. The proposed algorithm is applied to the task of forecasting the US inflation and unemployment. The GA-SVR genetically optimizes

[4] The MDM test is applied to couples of forecasts (GA-SVR vs. another forecasting model). A negative MDM value indicates that the first forecast (GA-SVR) is more accurate than the second forecast. The lower the negative value, the more accurate are the GA-SVR forecasts.

the SVR parameters and adapts to the optimal feature subset from a feature space of potential inputs. The feature space includes a wide pool of economic indicators that might affect the two series under study. The forecasting performance of the GA-SVR is benchmarked with a Random Walk model (RW), an Autoregressive Moving Average model (ARMA), a Moving Average Convergence/Divergence model (MACD), a Multi-Layer Perceptron (MLP), a Recurrent Neural Network (RNN) and a Genetic Programming (GP) algorithm. More specifically, the statistical performance of all models is investigated in four rolling samples during the period of 1974-2012.

In terms of our results, the GA-SVR outperforms all benchmark models for both forecasting exercises. Our model is able to genetically adapt to a small number of relevant variables and project superior forecasts at the same time. This performance is consistent also in periods of economic turmoil, which proves that the genetic SVR selection of the predictors is both computationally and statistically efficient. With this variable selection process, GA-SVR attempts to provide evidence on what inputs can be important predictors of US inflation and unemployment in the specific periods under study. The autoregressive lags of the past quarter are found to be of great importance, while information going back more than a semester seems irrelevant. The in-sample and out-of-sample results show that the models implementing genetic approaches, GP and GA-SVR, project the most accurate forecasts and outperform their benchmarks. This superiority is further validated by the MDM test. In general, the two forecasting exercises of this paper attempt to shed more light on the difficult quest of nonlinear mapping of macroeconomic variables over different sample periods.

REFERENCES

Ager P, Kappler M., Osterloh S. 2009. The accuracy and efficiency of the Consensus Forecasts: A further application and extension of the pooled approach. *International Journal of Forecasting* **25:** 167–181.

Ang A, Bekaert G, Wei M. 2007. Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better? *Journal of Monetary Economics* **54**: 1163–212.

Atkeson A, Ohanian, L. 2001. Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve*

Bank of Minneapolis Quarterly Review **25**: 2–11.

Barhoumi K, Darné O, Laurent Ferrara L. 2010. Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting, Special Issue: Advances in Business Cycle Analysis and Forecasting* **29**: 132–144.

Box GEP, Jenkins GM. 1976. *Time Series Analysis: Forecasting and Control*, Revised Edition, Oakland, CA: Holden-Day.

Cao LJ, Chua KS, Guan LK. 2003. C-ascending support vector machines for financial time series forecasting. In: *Computational Intelligence for Financial Engineering Proceedings* : 317-323.

Chen SH. 2002. *Genetic algorithms and genetic programming in computational finance*. Boston: Kluwer Academic Publishers.

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

Chua CL, Lim GC, Tsiaplias S. 2012. A latent variable approach to forecasting the unemployment rate. *Journal of Forecasting* **31:** 229–244.

Cogley T, Sargent TJ. 2005. Drifts and volatilities: monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics* **8:** 262–302.

Cogley T, Primiceri GE, Sargent TJ. 2010. Inflation-gap Persistence in the U.S. *American Economic Journal: Macroeconomics* **2:** 43-69.

D'Agostino A, Gambetti L, Giannone D. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* **28:** 82–101.

Diebold FX, Mariano RS. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* **13**: 253-263.

Duan K, Keerthi SS, Poo AN. 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing* **51**: 41-59.

Groen J, Paap R, Ravazzolo F. 2010. Real-time Inflation Forecasting in a Changing World, *Federal*

Reserve Bank of New York, Staff Report Number 388.

Harvey D, Leybourne S, Newbold P. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* **13:** 281–291.

Hassani H, Heravi S, Zhigljavsky A. 2012. Forecasting UK Industrial Production with Multivariate Singular Spectrum Analysis. *Journal of Forecasting, Published online in Wiley Online Library* **DOI**: 10.1002/for.2244

Holland J. 1995. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge, Mass: MIT Press.

Huang CF. 2012. A hybrid stock selection model using genetic algorithms and support vector regression.

Applied Soft Computing **12:** 807–818.

Inoue A, Kilian L. 2008. How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. *Journal of the American Statistical Association* **103:** 511–522.

Jacobson T, Karlsson S. 2004. Finding good predictors for inflation: a Bayesian model averaging approach. *Journal of Forecasting* **23**: 479–496.

Kim KJ. 2006. Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications* **30:** 519–526.

Kim HS, Sohn SY. 2010. Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research* **201:** 838–846.

Koop G, Korobilis D. 2012. Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review* **53**: 867–886.

Koza J, Adre D. 1996. Evolution of iteration in genetic programming. *In Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming. Cambridge:MIT Press.*

Li KW. 2012. A study on the volatility forecast of the US housing market in the 2008 crisis. *Applied Financial Economics* **22:** 1869-1880.

Lisboa PJG, Vellido A. 2000. Business Applications of Neural Networks. In *Business Applications of Neural Networks: The State-of-the-Art of Real-World Applications*, Lisboa PJG, Edisbury B, Vellido A (eds.) World Scientific: Singapore; vii-xxii.

Lu CJ, Lee TS, Chiu CC. 2009. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems* **47:** 115–125.McAdam P, McNelis P. 2005. Forecasting inflation with thick models and neural networks. *Economic Modelling* **22:** 848–867.

Milas C, Rothman P. 2008. Out-of-sample forecasting of unemployment rates with pooled STVECM forecasts. *International Journal of Forecasting* **24**: 101–121.

Montgomery AL, Zarnowitz V, Tsay RS, Tiao GC. 1998. Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association* **93**: 478-493.

Moshiri S, Brown L. 2004. Unemployment variation over the business cycles: a comparison of forecasting models. *Journal of Forecasting* **23**: 497–511.

Murphy JJ. 1999. *Technical Analysis of the Financial Markets: A Comprehensive Guide To Trading Methods And Applications*. New York Institute of Finance: Prentice Hall.

Olmedo E. 2013. Forecasting Spanish Unemployment Using Near Neighbor and Neural Net Techniques. *Computational Economics DOI 10.1007/s10614-013-9371-1*.

Rothman P. 1998. Forecasting Asymmetric Unemployment Rates. *The Review of Economics and Statistics* **80:** 164-168.

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

Scholkopf B, Smola A. 2002. *Learning with kernels*. Cambridge: MIT Press.

Sermpinis G, Laws J, Karathanasopoulos A, Dunis CL. 2012. Forecasting and trading the EUR/USD exchange rate with Gene Expression and Psi Sigma Neural Networks. *Expert Systems with Applications* **39:** 8865–8877.

Shapiro AF. 2000. A Hitchhiker's guide to the techniques of adaptive nonlinear models. *Insurance: Mathematics and Economics* **26**: 119-132.

Siedlecki W, Sklansky J. 1989. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters* **10**: 335-347.

Stock JH, Watson MW. 2003. Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature* **41**: 788–829.

Stock JH, Watson MW. 2004. Combination forecasts of output growth in a seven country dataset. *Journal of Forecasting* **23**: 204–430.

Stock JH, Watson MW. 2007. Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking* **39:** 3-33.

Stock JH, Watson MW. 2009. Phillips Curve Inflation Forecasts. *In Understanding Inflation and the Implications for Monetary Policy*. Fuhrer J, Kodrzycki Y, Little J, Olivei G (eds). Cambridge, MIT Press: 101-186.

Stock JH, Watson MW. 2012. Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business & Economic Statistics* **30:** 481-493.

Swanson NR, White H. 1997. A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics* **79**: 540-550.

Sun Z, Bebis G, Miller R. 2004. Object detection using feature subset selection. *Pattern Recognition* **37**: 2165–2176.

Trafalis TB, Ince H. 2000. Support vector machine for regression and applications to financial forecasting. In: *Neural Networks, IJCNN 2000* **1**: 348-353.

Vapnik VN. 1995. *The nature of statistical learning theory*. Springer.

Wang Z. 2010. Directed graphs, information structure and forecast combinations: an empirical examination of US unemployment rates. Journal of Forecasting **29:** 353–366.

Wright JH. 2009. Forecasting US Inflation by Bayesian Model Averaging. *Journal of Forecasting* **28**: 131–144.

Zhang M. 2009. *Artificial Higher Order Neural Networks for Economics and Business.* IGI Global: Hershey.

Zimmermann HG, Tietz C, Grothmann R. 2012. Forecasting with Recurrent Neural Networks: 12 Tricks. *Lecture Notes in Computer Science* **7700:** 687-707.

APPENDIX

A. Technical Characteristics

This appendix section includes the technical characteristics of the computational models used as benchmarks in this application. In the following table the design and training characteristics of the NNs are explained for each period.

[Insert Table A.1]

The selection of the NN's inputs is based on a sensitivity analysis on the in-sample period. We divided the in-sample period in to two sub-periods, the training and the test sub-periods. The test sub-period is consisted by the last four years of the in-sample. We experimented with the characteristics and inputs of our NNs in the training sub-period and we selected the architecture that provided the best statistical performance in the test sub-period. No part of the out-of-sample period was involved in the NN parameterization in any forecasting exercise. This approach is common in NN modelling and avoids problems such as the over-fitting and the data-snooping (Lisboa and Vellido (2000) and Zhang (2009)).

The parameter setting of the GP follows in table A.2.

[Insert Table A.2]

B. Statistical Performance Measures

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

[Insert Table B.1]

FIGURES

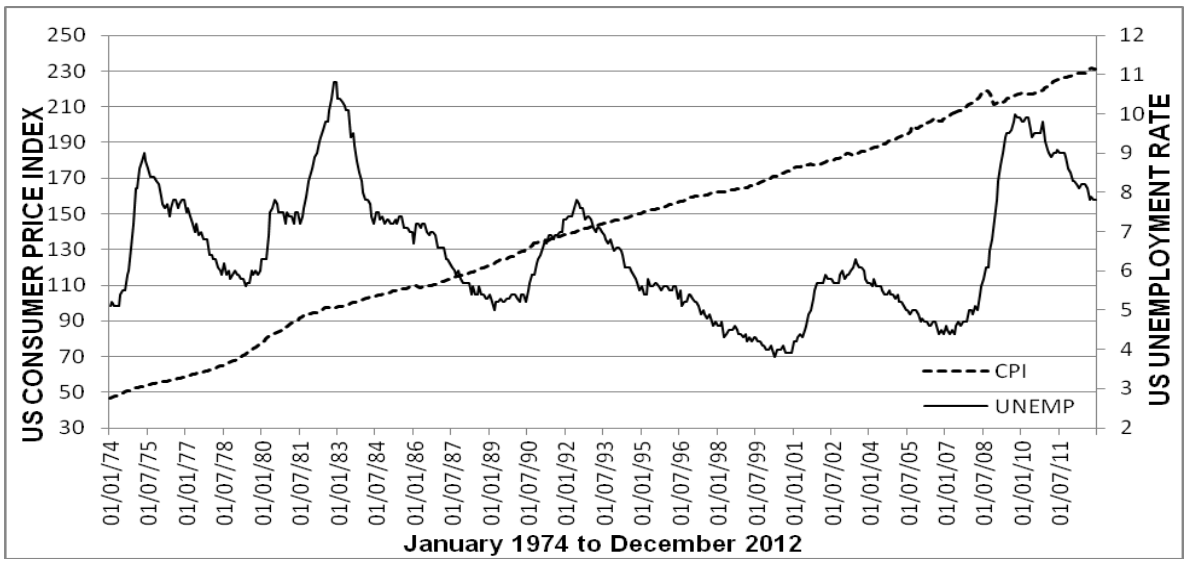
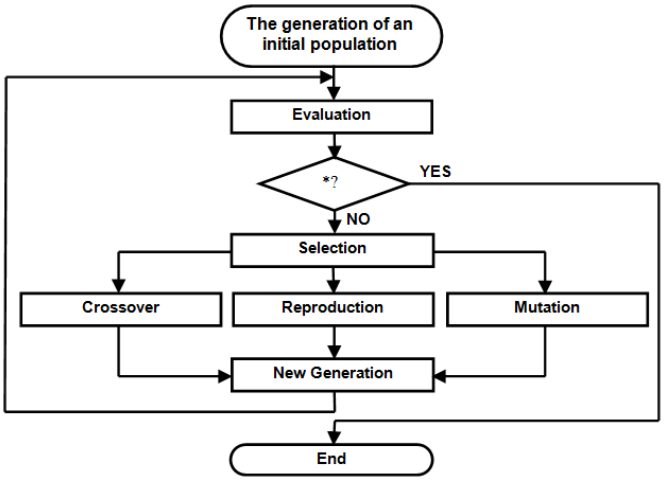


Figure 1: The historical monthly series of US CPI and Unemployment Rate in levels.



** The symbol '?' is the termination criterion which iterates or terminates the procedure of GP*

Figure 2: GP Architecture

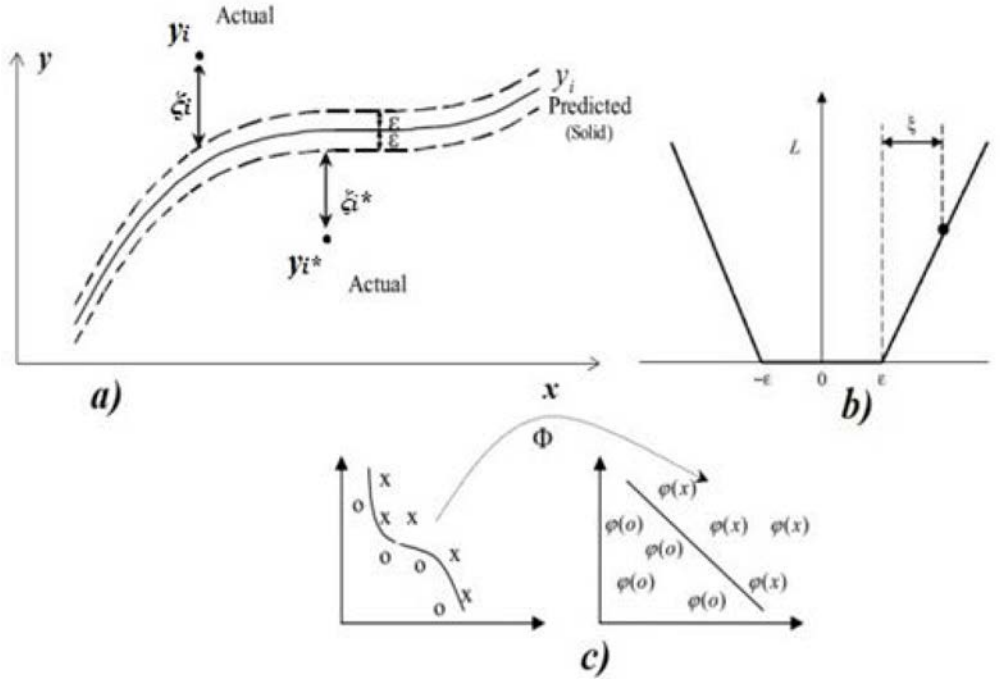


Figure 3: a) The f(x) curve of SVR and the ε-tube, b) plot of the ε-sensitive loss function and c) mapping procedure by $\varphi(x)$

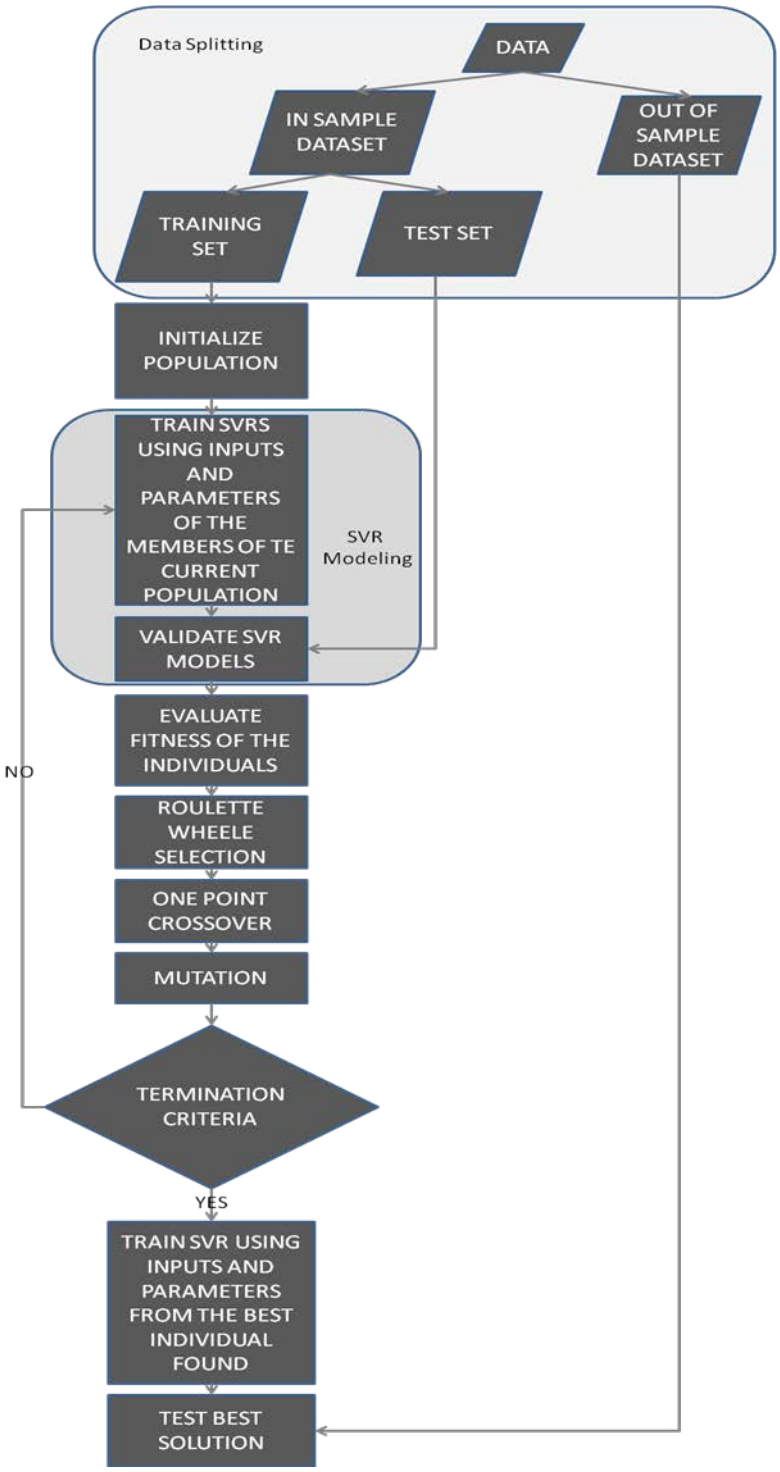


Figure 4: Hybrid GA-SVR flowchart

TABLES

No	MNEMONIC	DESCRIPTION	SOURCE
1	CPI	US Consumer Price Index for All Urban Consumers: All Items (SA)	FRED
2	UNEMP	US Civilian Unemployment Rate (SA)	FRED
3	JPY	JPY/USD Exchange Rate (NSA)	FRED
4	GBP	GBP/USD Exchange Rate (NSA)	BLOOM
5	HOUSE	US Housing Starts Total: New Privately Owned Housing Units Started (SA)	FRED
6	INDP	US Industrial Production Index (SA)	FRED
7	M1	US M1 Money Stock (SA)	FRED
8	EMPL	US All Employees: Total nonfarm (SA)	FRED
9	PCE	US Personal Consumption Expenditures (SA)	FRED
10	PI	US Personal Income (SA)	FRED
11	TBILL	US 3-Month Treasury Bill: Secondary Market Rate (NSA)	FRED
12	WAGE	US Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing (SA)	FRED
13	DJIA	Dow Jones Industrial Average (NSA)	BLOOM

Note: CPI and UNEMP2 are observed variables. The pool of predictors consists of the first ten autoregressive terms of variables 3-13 (110 series in total). FRED refers to the FRED database of the St. Louis Federal Reserve Bank, while BLOOM stands for Bloomberg. All series are in monthly percentage changes.SA and NSA means that the series is seasonally adjusted and not seasonally adjusted respectively.

Table 1: List of all the variables

Population Size	400
Maximum Generations	5000
Selection Type	Roulette Wheel Selection
Elitism	Best member of every population is maintained in the next generation.
Crossover Probability	0.85
Mutation Probability	0.15
Fitness Function	1/(1+MSE)

Table 2: GA Characteristics and Parameters

OUT-OF-SAMPLE PERIODS	ALL PREDICTORS	CPI PREDICTORS	SELECTED LAGS	UNEMP PREDICTORS	SELECTED LAGS
01/1997 – 12/2000*	JPY	-	-	-	-
	GBP	-	-	-	-
	HOUSE	HOUSE	1,2	HOUSE	2,3,4
	INDP	INDP	**1,2,3**	INDP	**1,2,3**,4
	M1	M1	2,**3**	M1	**3**,4
	EMPL	EMPL	1,4	EMPL	2,3
	PCE	PCE	**4**	PCE	**4**
	PI	PI	3,4	PI	1,3,5
	TBILL	-		TBILL	2,3
	WAGE	WAGE	1,2,3	WAGE	2,4,5
	DJIA	DJIA	5	DJIA	3
	TOTAL	8	16	9	21
01/2001 – 12/2004**	JPY	-	-	-	-
	GBP	-	-	-	-
	HOUSE	HOUSE	2,3,4	HOUSE	1,2
	INDP	INDP	**3,4**,5	INDP	1,2,**3,4**
	M1	M1	**2**	M1	**2**,3
	EMPL	EMPL	1,3,4	EMPL	3
	PCE	PCE	1,**4**	PCE	2,3,**4**
	PI	PI	4	PI	4,5
	TBILL	-	-	TBILL	1
	WAGE	WAGE	2,3	WAGE	1,2,4
	DJIA	DJIA	4,5	DJIA	2,5
	TOTAL	8	17	9	20
01/2005 – 12/2008***	JPY	-	-	-	-
	GBP	-	-	-	-
	HOUSE	-	-	-	-
	INDP	INDP	**1,2**	INDP	**1,2**,3,4
	M1	M1	2,**3**	M1	**3**
	EMPL	-	-	EMPL	2,3
	PCE	PCE	**1**,2	PCE	**1**,3
	PI	-	-	PI	1,2
	TBILL	TBILL	1,2	-	-
	WAGE	WAGE	1,2,3	WAGE	2,4
	DJIA	-	-	-	-
	TOTAL	5	11	6	13
01/2009 – 12/2012****	JPY	-	-	-	-
	GBP	-	-	-	-
	HOUSE	-	-	-	-
	INDP	INDP	**1**	INDP	**1**,2,3,4
	M1	M1	1,**2**	M1	**2**
	EMPL	-	-	EMPL	1
	PCE	PCE	1,**2**	PCE	**2**
	PI	PI	1,2,3	-	-
	TBILL	-	-	-	-
	WAGE	-	-	WAGE	2,3
	DJIA	-	-	-	-
	TOTAL	4	8	5	9

*Note: The bold predictors in the second column represent the commonly selected variables for both exercises regardless the out-of-sample period. The bold values in the fourth and sixth column are the common predictors for both forecasting exercises in the respective out-of-sample sub-periods. * CPI: Population=60, Generations=440, C=61.5, γ=0.015, v=0.47, UNEMP: Population=200, Generations=500, C=143.8, γ=0.91, v=0.54. **CPI: Population=110, Generations=280, C=54.3, γ=0.03, v=0.55, UNEMP: Population=300, Generations=400, C=121.4, γ=0.75, v=0.63. ***CPI: Population=80, Generations=220, C=37.8, γ=0.042, v=0.31, UNEMP: Population=140, Generations=250, C=94.6, γ=0.88, v=0.77. **** CPI: Population=75, Generations=550, C=51.5, γ=0.025, v=0.59, UNEMP: Population=130, Generations=430, C=135.3, γ=0.56, v=0.37.*

Table 3: The selected predictors for US inflation and unemployment (best CPI and UNEMP chromosome)

	IN-SAMPLE PERIODS							
CPI	**01/1974 – 12/1996**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0153	0.0151	0.0087	0.0058	0.0056	0.0051	**0.0047**
	MAPE	102.23%	101.86%	98.54%	62.67%	63.14%	59.79%	**53.44%**
	RMSE	0.0095	0.0091	0.0084	0.0069	0.0068	0.0061	**0.0055**
	Theil-U	0.8561	0.8456	0.6758	0.5881	0.5721	0.5377	**0.5112**
UNEMP	**01/1974 – 12/1996**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0171	0.0168	0.0093	0.0062	0.0059	0.0055	**0.0052**
	MAPE	101.21%	98.67%	87.94%	62.89%	61.34%	57.51%	**56.47%**
	RMSE	0.0175	0.0169	0.0144	0.0124	0.0099	0.0092	**0.0087**
	Theil-U	1.2595	1.2568	0.9884	0.8197	0.8155	0.7823	**0.7514**
CPI	**01/1978 – 12/2000**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0155	0.0154	0.0084	0.0055	0.0054	0.0052	**0.0046**
	MAPE	105.98%	102.15%	98.85%	63.03%	63.19%	58.99%	**53.17%**
	RMSE	0.0121	0.0102	0.0081	0.0068	0.0066	0.0063	**0.0056**
	Theil-U	0.8573	0.8511	0.6692	0.5775	0.5659	0.5412	**0.5067**
UNEMP	**01/1978 – 12/2000**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0188	0.0187	0.0089	0.0063	0.0058	0.0054	**0.0053**
	MAPE	98.15%	97.88%	86.53%	63.27%	62.14%	58.71%	**55.84%**
	RMSE	0.0166	0.0158	0.0139	0.0116	0.0098	0.0094	**0.0089**
	Theil-U	1.1884	1.1825	0.9992	0.8341	0.8216	0.8013	**0.7673**
CPI	**01/1982 – 12/2004**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0239	0.0233	0.0094	0.0074	0.0072	0.0069	**0.0064**
	MAPE	132.09%	131.84%	103.25%	69.88%	67.26%	64.21%	**61.42%**
	RMSE	0.0132	0.0129	0.0091	0.0076	0.0077	0.0072	**0.0069**
	Theil-U	0.9764	0.9715	0.8469	0.7198	0.7145	0.6755	**0.6211**
UNEMP	**01/1982 – 12/2004**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0202	0.0196	0.0098	0.0081	0.0079	0.0075	**0.0068**
	MAPE	124.11%	123.27%	90.37%	73.89%	71.26%	68.55%	**63.37%**
	RMSE	0.0215	0.0189	0.0157	0.0101	0.0099	0.0096	**0.0093**
	Theil-U	1.3965	1.3947	1.2558	0.9377	0.9358	0.9122	**0.8847**
CPI	**01/1986– 12/2008**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0191	0.0188	0.0092	0.0071	0.0069	0.0064	**0.0061**
	MAPE	118.64%	113.58%	99.56%	64.55%	63.17%	62.44%	**59.83%**
	RMSE	0.0129	0.0098	0.0085	0.0072	0.0072	0.0069	**0.0065**
	Theil-U	0.9322	0.9126	0.7941	0.6853	0.6751	0.6239	**0.5845**
UNEMP	**01/1986 – 12/2008**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0112	0.0106	0.0088	0.0075	0.0073	0.0068	**0.0064**
	MAPE	102.87%	100.68%	84.57%	68.12%	67.89%	64.58%	**60.29%**
	RMSE	0.0175	0.0168	0.0135	0.0097	0.0096	0.0093	**0.0091**
	Theil-U	1.2801	1.2745	0.9957	0.9254	0.9136	0.8947	**0.8667**

Table 4: Summary of In-Sample Statistical Performances

	OUT-OF-SAMPLE PERIODS							
CPI	**01/1997 – 12/2000**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0165	0.0162	0.0091	0.0059	0.0059	0.0052	**0.0049**
	MAPE	104.25%	103.58%	100.54%	66.15%	66.86%	62.47%	**57.34%**
	RMSE	0.0098	0.0094	0.0086	0.0071	0.007	0.0065	**0.0058**
	Theil-U	0.8755	0.8632	0.6955	0.6013	0.5971	0.5521	**0.5317**
UNEMP	**01/1997 – 12/2000**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0182	0.0178	0.0098	0.0065	0.0063	0.0059	**0.0055**
	MAPE	103.88%	100.12%	91.74%	65.38%	64.59%	61.13%	**59.11%**
	RMSE	0.0178	0.0174	0.015	0.0134	0.0107	0.0094	**0.009**
	Theil-U	1.2708	1.2647	1.0294	0.8465	0.8334	0.8037	**0.7867**
CPI	**01/2001 – 12/2004**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0161	0.0158	0.0086	0.0061	0.0057	0.0055	**0.0051**
	MAPE	105.87%	105.19%	99.65%	65.11%	64.83%	61.35%	**55.62%**
	RMSE	0.0128	0.0116	0.0084	0.007	0.0068	0.0065	**0.0059**
	Theil-U	0.8914	0.8845	0.7259	0.6018	0.5845	0.5633	**0.5297**
UNEMP	**01/2001 – 12/2004**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0264	0.0213	0.0092	0.0067	0.0061	0.0057	**0.0055**
	MAPE	102.67%	99.66%	89.45%	66.76%	64.88%	60.24%	**56.84%**
	RMSE	0.0167	0.0161	0.0142	0.0135	0.0102	0.0097	**0.0092**
	Theil-U	1.2298	1.2254	1.105	0.8656	0.8417	0.8229	**0.7837**
CPI	**01/2005 – 12/2008**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0325	0.0311	0.0112	0.0078	0.0075	0.0071	**0.0067**
	MAPE	146.15%	144.21%	105.83%	72.57%	70.64%	67.41%	**64.23%**
	RMSE	0.01447	0.0135	0.0096	0.0079	0.0078	0.0075	**0.0072**
	Theil-U	1.0156	1.0051	0.8657	0.7403	0.7367	0.7147	**0.6647**
UNEMP	**01/2005 – 12/2008**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0285	0.0215	0.0103	0.0084	0.0081	0.0078	**0.0073**
	MAPE	128.55%	125.64%	92.54%	75.21%	74.83%	71.75%	**67.28%**
	RMSE	0.0209	0.0197	0.0162	0.0129	0.0117	0.0099	**0.0096**
	Theil-U	1.4338	1.4269	1.2783	0.9531	0.9457	0.9314	**0.9158**
CPI	**01/2009– 12/2012**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0214	0.0208	0.0097	0.0074	0.0071	0.0068	**0.0064**
	MAPE	118.27%	116.17%	102.68%	69.82%	68.93%	66.71%	**62.67%**
	RMSE	0.0116	0.0108	0.0088	0.0075	0.0074	0.0070	**0.0068**
	Theil-U	0.9455	0.9384	0.8211	0.7139	0.6957	0.6483	**0.6144**
UNEMP	**01/2009– 12/2012**	**RW**	**ARMA**	**MACD**	**MLP**	**RNN**	**GP**	**GA-SVR**
	MAE	0.0119	0.0114	0.0092	0.0081	0.0078	0.0071	**0.0067**
	MAPE	108.84%	102.67%	88.98%	72.55%	71.39%	68.27%	**64.17%**
	RMSE	0.0194	0.0172	0.0145	0.0102	0.0099	0.0097	**0.0093**
	Theil-U	1.3274	1.3128	1.1296	0.9485	0.9318	0.9133	**0.8926**

Table 5: Summary of Out-of-Sample Statistical Performances

PERIODS	VARIABLES	STATISTICS	RW	ARMA	MACD	MLP	RNN	GP
01/1997 –12/2000	**CPI**	MDM_1	-7.22	-7.19	-6.65	-4.54	-3.92	**-3.15**
		MDM_2	-9.81	-9.45	-8.77	-7.19	-6.98	**-5.19**
	UNEMP	MDM_1	-5.21	-5.09	-4.91	-4.13	-4.08	**-3.02**
		MDM_2	-7.71	-7.68	-7.43	-5.51	-4.98	**-4.51**
01/2001 – 12/2004	**CPI**	MDM_1	-7.07	-6.96	-6.53	-4.24	-3.97	**-3.22**
		MDM_2	-9.34	-8.97	-8.58	-7.81	-7.12	**-6.79**
	UNEMP	MDM_1	-5.23	-5.03	-4.84	-4.19	-4.10	**-3.26**
		MDM_2	-7.90	-7.82	-7.48	-5.40	-4.97	**-4.87**
01/2005 – 12/2008	**CPI**	MDM_1	-8.14	-8.05	-7.54	-7.19	-6.88	**-6.34**
		MDM_2	-9.84	-9.80	-9.30	-8.95	-8.80	**-8.62**
	UNEMP	MDM_1	-9.25	-9.16	-8.73	-8.15	-7.51	**-7.07**
		MDM_2	-10.37	-10.27	-9.81	-9.54	-9.17	**-8.75**
01/2009 – 12/2012	**CPI**	MDM_1	-6.46	-6.27	-5.92	-5.71	-5.48	**-4.77**
		MDM_2	-7.95	-7.88	-7.71	-7.44	-7.32	**-7.04**
	UNEMP	MDM_1	-7.82	-7.66	-7.08	-6.77	-6.33	**-6.09**
		MDM_2	-8.75	-8.54	-8.28	-7.83	-7.39	**-7.25**

Note: $MDM_1$and MDM_2 are the statistics computed for the MSE and MAE loss function respectively.

Table 6: Modified Diebold-Mariano statistics for MSE and MAE loss functions

	PARAMETERS	01/1974 – 12/2000		01/1978 – 12/2004		01/1982-12/2008		01/1986/12/2012	
		MLP	RNN	MLP	RNN	MLP	RNN	MLP	RNN
CPI	Learning algorithm	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent
	Learning rate	0.003	0.002	0.005	0.002	0.004	0.003	0.002	0.002
	Momentum	0.004	0.003	0.006	0.003	0.005	0.005	0.004	0.003
	Iteration steps	50000	45000	50000	40000	35000	25000	60000	40000
	Initialisation of weights	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)
	Input nodes	9	7	8	7	7	7	9	8
	Hidden nodes	6	5	6	6	4	3	5	6
	Output node	1	1	1	1	1	1	1	1
UNEMP	Learning algorithm	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent
	Learning rate	0.002	0.003	0.002	0.002	0.004	0.002	0.003	0.002
	Momentum	0.005	0.005	0.004	0.003	0.006	0.005	0.005	0.004
	Iteration steps	35000	30000	35000	30000	35000	30000	35000	30000
	Initialisation of weights	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)	N(0,1)
	Input nodes	7	6	8	6	9	6	8	7
	Hidden nodes	6	4	5	5	7	3	4	4
	Output node	1	1	1	1	1	1	1	1

Table A1: Neural Network Design and Training Characteristics for all periods under study

GENETIC PROGRAMMING PARAMETERS			
Population Size	200	**Fitness evaluation function**	MSE
Termination Criterion	75000	**Tournament Size**	20
Max. tree depth	12	**Crossover trials**	1
Function Set	+, -, *, /, ^, ^2, ^3, ^1/2, ^1/3, Exp, If, sin, cos, tan	**Mutation Probability**	0.8

Table A.2: GP parameters setting

STATISTICAL PERFOMANCE MEASURES	DESCRIPTION
Mean Absolute Error	$MAE = (\frac{1}{n}) \sum_{\tau=t+1}^{t+n} \left\| \hat{Y}_\tau - Y_\tau \right\|$ *with* Y_τ *being the actual value and* $\hat{Y}_\tau$ *the forecasted value*
Mean Absolute Percentage Error	$MAPE = \frac{1}{n} \sum_{\tau=t+1}^{t+n} \left\| \frac{Y_\tau - \hat{Y}_\tau}{Y_\tau} \right\|$
Root Mean Squared Error	$RMSE = \sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} (\hat{Y}_\tau - Y_\tau)^2}$
Theil-U	$Theil - U = \frac{\sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} (\hat{Y}_\tau - Y_\tau)^2}}{\sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} \hat{Y}_\tau^2} + \sqrt{\frac{1}{n} \sum_{\tau=t+1}^{t+n} Y_\tau^2}}$

Table B.1: Statistical Performance Measures and Calculation