Application of Stationary Wavelet Support Vector Machines for the Prediction of Economic Recessions

E. Giovanis

Abstract— This paper examines the efficiency of various approaches on the classification and prediction of economic expansion and recession periods in United Kingdom. Four approaches are applied. The first is discrete choice models using Logit and Probit regressions, while the second approach is a Markov Switching Regime (MSR) Model with Time-Varying Transition Probabilities. The third approach refers on Support Vector Machines (SVM), while the fourth approach proposed in this study is a Stationary Wavelet SVM modelling. The findings show that SW-SVM and MSR present the best forecasting performance, in the out-of sample period. In addition, the forecasts for period 2012-2015 are provided using all approaches.

Keywords— Discrete choice models, Stationary Wavelets Transform, Economic crisis, Markov Switching Regime, Support Vector Machines

I. Introduction

THE sub-prime financial crisis began with the bursting of the housing bubble in the United States (U.S.) and triggered a global financial crisis in 2007. In August 2007 the problems hit the financial world and caused enormous liquidity pressures within the interbank market. Due to the widespread dispersion of credit risk and the complexity of financial instruments, the mortgage crisis had a large impact on financial markets. Banks became reluctant to lend to each other. Share prices and stock market indices have massively declined since July 2007. Several large banks, credit insurances and mortgage companies have reported significant losses and have lost much of their market value. Up to today seven small U.S. banks have failed and were taken over by U.S. regulators. Moreover, central banks were substantially challenged by the crisis and had to take essential actions. Currency and economics crises occur as an outcome of unsustainable government policies and also as a consequence of speculative attack that is motivated by either self-fulfilling expectations or through contagion from crises occurring someplace else.

One major challenge of macroeconomists and financial managers is the prediction of financial crisis and economic recessions and expansions periods, as well as, early warning systems for financial distress of firms. Based on the identified indicators, different empirical approaches were developed to rank the vulnerability of countries and to predict future crises. Early-warning models exploit systematic relationships apparent in historical data between variables associated with the build-up to crises and the actual incidence of crisis. Their aim is to forewarn the policy makers about the future financial crisis and help them to take pre-emptive measures.

The purpose and contribution of this paper is the introduction of Support Vector Machines and Stationary Wavelets for the prediction of crisis periods. The structure of the study has as follows: In the second section a brief literature review is provided. In section three the main methodology followed in this study is presented, while in section four the data used in this paper are provided. In section five the main results are reported. Finally, in section six the concluding remarks of this study are presented.

II. Literature Review

Many papers and studies have been written in an effort to provide reliable approaches of financial crisis periods predictions. One approach is the first generation models [1]-[2], which describe that financial crisis are results of speculative attacks because of rational arbitrage. Krugman [1] was the main representative of currency crises generation. To explain the crisis, he focused and stressed that weak macroeconomic fundamentals of countries due to expansionary monetary and fiscal policy of governments were the main causes of crises. On the other hand, other models have shown that a crisis may develop without a significant change in the economic fundamentals. Then the second generation models have been developed [3]-[4], which show mechanisms that even sustainable currency pegs may be attacked and broken.

There are models beyond the second generation as the Moral Hazard models [5] which try to explain cycle of investment boom and bust as a result of important and rapid withdrawal from financial assets because the asset prices are declined sharply and suddenly, rather than a simple currency crisis. Specifically these models indicate that a possible index

E. Giovanis is with the Department of Economics, Royal Holloway University of London (phone: +442035565469; e-mail: giovanis95@gmail.com, Eleftherios.Giovanis.2010@live.rhul.ac.uk).

of financial system's fragility can be used as a reliable pre-warning indicator.

Two related approaches have been estimated in the literature for the designing of an early or pre warning system for financial crisis, the Probit and Logit discrete choice models and the signal to noise ratio. In the first approach the construction of a crisis dummy variable is required. There are various examples of application in currency crisis of discrete choice modelling [6]-[10].

Frankel and Rose [6] studied currency crashes from 1971 through 1992 and by using panel data has evaluated regression model to forecast currency crises; moreover by comparing countries that had suffered a repetitive pattern of currency crises paper also examine likewise economic indicators that led to collapse and reaction of policy changes. The authors [6] concluded that currency crash is rooted in the eliminated resources of foreign direct investment may have a direct effect onto country's reserves, consequently increasing domestic credit rates and overvaluing the real exchange rate. However, nor current account deviations, nor government budget deficit did not participate in the predication of the currency crashes. Berg and Pattillo [11] argued that despite the limitations of the Probit model analysis of proposed forecasting only for the currency crises; the authors found that there is an evidence of the nonlinearity in the relationship between predictive variables and the probability of the crises that may assist in the prediction of the banking crises, which differed from signal approach proposed by Kaminsky et al. [12]. The advantage of Probit and Logit models is that allow for statistical testing, identifying the sign, the magnitude and the marginal distributions of the explanatory variables to the onset of the crisis. On the other hand, this approach confronts the problem of misspecification errors and serial correlation.

Another attempt to analyse indicators of annual panel data was made by [13]; thus, the multinomial model was used to overcome limitations of the Probit or regression analysis in signalling early warning approach. The authors studied the macroeconomic and financial variable in identification of the best performing variable to predict financial distress and concluded that proxies for the vulnerability of the banking sector in Asian crises provided the accurate information. However some cases of significant financial distress are often preceded by the rapid credit expansion and growth in consumption, which is associated with an increase in domestic real interest rate.

The second approach optimises the signal to noise ratio for the potential crisis indicators [12], [14]. This approach is used in order to identify variables as strong potential indicators of crisis. The "noise-to-ratio" approach has been applied to currency and banking crises to examine the impact of deteriorating fundamentals and identify variables as strong potential indicators of crisis. These studies searched for origins causing a convergence of both types of crises: banking and currency. More specifically, Kaminsky and Reinhart [14] examined monthly data for 23 crises, and they found that the economic downturn precedes the banking crisis, where specifically, the output tends to peak roughly eight months before a crisis take place. Furthermore, Kaminsky et al. [12] found that exports, the deviations of the real exchange rates from the trend, the output and equity prices are strong potential indicators providing a warning signal that a crisis will take place. Additionally, when an indicator exceeds a specific threshold, then a banking crisis might take place in the following 24 months. The advantage of this approach is that the variables can be directly ranked as crisis indicators. On the other hand the one drawback of this approach is that it does not allow for statistical testing and also ignores about the possible correlations among the variables.

A different approach applied in currency crises prediction is the Markov Switching Regime (MSR) Model. The Markov switching model of Hamilton [15]-[16] gave rise to the exploration of the dynamics of exchange rate movements and one area is the study of currency crises. Engel and Hakkio [17] modeled the European Monetary System (EMS) currencies using MSR with Time-Varying Transition Probabilities (TVTP) where they considered two regimes, the stable and the volatile. Volatile periods occur during realignments of currencies in the EMS band where the transition probability depends on the location of the currency within the EMS band. They have shown that the probability of realignment depends on what regime the period belongs to. Cerra and Saxena [18] employed a MSR with Time-Varying Transition Probabilities (TVTP) to show evidence of contagion in the Asian Financial Crisis where an Index of Speculative Pressure (ISP) of Thailand and Korea driving the TVTP improved the estimation of the conditional probability of a crisis in Indonesia. Similarly, Abiad [19] and Mariano et al. [20] used an MSR with TVTP to develop an Early Warning System using Markov switching of the change in the nominal exchange rate with three categories of early warning indicators. Abiad's [19] indicators involved macroeconomic, capital flow and financial fragility variables. MRS allows for sharp movements between the regimes, so it is able to describe sudden shifts and changes in behavior. Furthermore, the model avoids the misclassification errors and serial correlation as in Probit and Logit models. On the other hand Markov switching model is still an econometrical approach based on statistical methodology, which is vulnerable as all econometric models are in relation with artificial intelligence procedures. However, MSR avoids the problems of data mining and actually is used for data mining purposes.

The last approaches concern neural networks applications, where the most research papers present results of supporting them. Cheng et al. [21] propose a fuzzy regression for a pre-warning financial distress model with triangular membership function and they found that fuzzy regression outperforms Logit model's forecasting performance. Nachev and Stoyanov [22] propose that the prediction of corporate bankruptcy can be viewed as a pattern recognition problem and they estimate an ARTMAP neural networks model for financial diagnosis. Other researches compare the neural networks with traditional statistical approaches and their results show that neural networks outperform significant the other statistical methods [23]-[27]. Chen and Du [28] compare artificial neural networks (ANN) and data mining techniques developing a financial distress model and they found that ANN presents better prediction accuracies than data mining techniques, suggesting that artificial intelligence can be a more suitably methodology than traditional statistical approaches. It should be noticed that the proposed approaches are not compared to other techniques mentioned above, like neural networks, fuzzy logic, neuro-fuzzy systems and self-organising maps (SOMs) among others, as other studies have already used the above approaches [22]-[23],[25],[29]-[31] just to mention a few of them.

Regarding, Support Vector Machines the literature review provides evidence and findings only for the prediction of financial distress and bankruptcy of firms, while the classification and prediction of countries' financial and economic crisis has been neglected. Lee and To [32], examined the prediction of financial distress periods using data collected from a securities firm's database in Taiwan. They found that SVM outperforms the Multilayer Neural Networks with back propagation algorithm. Chen and Chen [33] proposed a particle swarm optimisation with support vector machines (PSO-SVM) using a dataset collected from Taiwanese listed companies composed of 68 companies during period 1996-2005. They found that PSO-SVM outperforms the simple SVM approach. Song et al. [34] propose a Genetic Algorithm (GA) based approach and statistical filter approaches are applied to identify the best features for the support vector machine (SVM) using a sample consisting of 194 non-ST (Specially treated) cases and 56 ST cases from high-tech manufacturing industries (information, mechanical, biological and pharmaceutical industry sectors) in China. The study's objective is to reduce (select) features for SVM-based classification models without degrading prediction accuracy, and thus to improve computational efficiency and identify critical ratios. Song et al. [34] found that GA-SVM and Pure SVM outperform the Logit approach in the out-of-sample or testing period. Ahn et al. [35] examined SVM performance using data for the economic financial condition indicator (Econ-FCI) in Korea for the period 1997-2002. The SVM is compared to multinomial logistic regression (MLR), decision tree (DT), case-based reasoning (CBR) and Neural Networks (NN) with three-layer connected back-propagation algorithm. Ahn et al. [35] found that SVM, DT and CBR are very efficient approaches in classifying financial crisis periods. On the other hand NN suffers from too much noise, which is typical when NN is trained on relatively small amount of, while MLR generates too many unnecessarily warning signals.

Martin and Milas [36] examined the causes of the financial crisis of 2007 using UK data and they conclude that the gross domestic product growth, loose monetary and the deterioration in current account balances were causes of the financial crisis. Bordo and Landon –Lane [37] examined banking, currency and debt crises and combinations of them across a large number of countries for the period from 1880 to 2010. The authors have considered several factors including globalisation, the gold standard and whether the USA was in crisis.

III. METHODOLOGY

A. Logit and Probit Models

In this section a brief description of the binary Logit model is provided. The logistic distribution is defined as [38]:

$$\Pr ob(Y=1\mid x)=\frac{e^{x'\beta}}{1+e^{x'\beta}}=\Phi(x'\beta) \quad (1)$$

The marginal partial effects of explanatory variables are given by:

$$\frac{\partial \mathrm{E}[y\mid x)}{\partial x}=\Phi(x'\beta)[1-\Phi(x'\beta)]\beta \quad (2)$$

Next the unknown probabilities are modelled as a linear function of variables x_i.

$$y_i=\ln\left(\frac{p_i}{1-p_i}\right)=\beta_0+\beta_1 x_{1,i}+\beta_2 x_{2,i}+\dots\dots+\beta_k x_{k,i} \quad (3)$$

The Logit model can be written in a general form regression as:

$$y=a+\sum_{i=1}^{n}\beta_i x_i+\varepsilon \quad (4)$$

Next the Probit model is presented [38], which is defined as:

$$\Phi^{-1}(p_i)=Z=a+\sum_{i=1}^{n}\beta_i x_i+\varepsilon \quad (5)$$

, where $\Phi^{-1}(pi)$ is the inverse cumulative distribution function (CD F) of the standard normal, a, β_i and x_i are defined as in (4).

One of the advantages and features of discrete choice models, like Logit and Probit, is the nonlinearity of the effect of the right-hand side variables on the left-hand side variable. This is an attractive property of the model because the literature on currency crises has demonstrated that such effects are at work [7]-[8]. Furthermore, there is complete freedom to choose which independent variables enter into the regression with the selection being based on various theories in the financial and economic crisis literature. Additionally, discrete choice models consider all variables together and look only at the marginal contributions of each indicator; the discrete choice models disregard variables that do not contribute information that is not already captured in the other variables. Lastly, these models allow for statistical testing, coefficient magnitude examination and interpretation.

On the other hand there are drawbacks about this approach. Kaminsky et al. [12] note that unlike the "noise-to-ratio" approach, one is unable to rank indicators on the basis of forecasting accuracy – they either enter the regression significantly, or do not. Moreover, measures of statistical significance cannot distinguish whether an indicator is good at correctly calling crises, or merely sends few false signals. The prediction or the classification percentage is done based on the estimated coefficients from the in-sample period using as the cutoff point the value of 0.5. For the forecasting and the classification performance of the binary Logit and Porbit models is:

If $y^* > 0.5$, then the economy is on economic recession period

If $y^* \leq 0.5$, then the economy is on non-economic recession period.

, where variable y^* denotes the predicted values.

The main problem with this measure is that the choice of the cutoff point is arbitrary. Traditionally, a cut-off point used has been 0.5. However, there is no reason why this cutoff is the appropriate one. Cramer [39] suggests that a more appropriate cutoff point is the sample frequency, which is defined as the sum of the percentage of correct prediction over the sum of the sample. There is no basis for this procedure, finding the appropriate cut-off point. More precisely, it is insoluble, and it is of no practical interest. Therefore, an open suggested research question is to examine and calculate the appropriate threshold for discrete choice models. However, this is out of the current study's scope.

B. Markov Switching Autoregressive (MS-AR) Model

Markov-Switching Autoregressive (MS-AR) models have first been introduced by Hamilton [15] to analyse the rate of growth of USA GNP and then used in different fields to model time series subject to discrete regime shifts. The model of interest in this study is a Markov switching model with time varying transition probability (MS-TVTP model) with a mean-deviation, p^{th} order-autoregressive form:

$$\begin{aligned} & y_t - m(s_t) = \phi_1(y_{t-1} - m(s_{t-1})) + ... + \phi_p(y_{t-p} - m(s_{t-p})) \\ & + \varepsilon_t,\ \varepsilon_t \sim i.i.d.\ N(0,\sigma^2) \end{aligned} \tag{6}$$

, where $m(s_t)$ indicates that the parameter is dependent on the state t, $s_t=0,1$, which cannot be observed but it is assumed to follow a discrete Markov chain such that $\Pr(s_t|\ s_{t-1},\ s_{t-2},\ldots)=\Pr((s_t|\ s_{t-1})$. The value of s_t is determined by a latent variable s_t^* as the following rule:

$$\begin{aligned} & s_t = 1\ if\ \ s_t^* = z'_{t-1}\gamma^{(s_{t-1})} + u_t \geq 0, \\ & s_t = 0\ if\ \ s_t^* = z'_{t-1}\gamma^{(s_{t-1})} + u_t \leq 0 \end{aligned} \tag{7}$$

, where the value of s_{t-1} is given, z_{t-1} is a vector of exogenous variables available at $t-1$, and u_t is assumed to follow the standard normal distribution, i.e., $E(u_t) = 0$ and $Var(u_t) = 1$. Since u_t follows such a distribution, the transition probabilities from s_{t-1} to s_t for all the possible values of states are given by:

$$\begin{aligned} & s_t = 1\ and\ \ s_{t-1} = 1\ \ with\ \ \Phi(z'_{t-1}\gamma^{(1)}), \\ & s_t = 1\ and\ \ s_{t-1} = 0\ \ with\ \ \Phi(z'_{t-1}\gamma^{(0)}), \end{aligned} \tag{8}$$

, where $\Phi()$ is the cumulative distribution function for the standard normal distribution. Letting $m(s_t):=m_0 + m_1 s_t$ and $\varphi:=(\varphi_1.....\varphi_p)'$, the parameters to be estimated are:

$$\theta := \{m_0, m_1, \phi, \sigma^2, \gamma^{(0)}, \gamma^{(1)}, \{s_t\}_{t=p+1}^n, \{s_t^*\}_{t=p+1}^n \tag{9}$$

, where n denotes the number of observations.

C. Wavelets Transform

C.1. Discrete Wavelet Transform (DWT) and the Haar Wavelet

The Discrete Wavelet Transformation (DWT) is a linear signal processing technique that, when applied to a data vector X, transforms it to a numerically different vector X, of wavelet coefficients. The discrete wavelet transform uses the idea of dimensionality reduction, which is a multivariate statistical method that stores the compressed approximation of the data under the premise of little loss of information. A compressed approximation of the data can be retained by storing only a small fraction of the user-specified threshold wavelet coefficients and remaining data as zero. This technique also works well to remove noise and abnormal data without smoothing out the main features of data [40]. Using a wavelet expansion, any function in $L(R^2)$ can be expressed as a sum of the basis elements

$$f(t) = \sum_{k\in Z} c_k \phi(t-k) + \sum_{k\in Z}\sum_{j=0}^{\infty} d_{j,k} 2^{j/2}\psi(2^j t - k) \tag{10}$$

Parameter j determines the scale or the frequency range of each wavelet basis function ψ. Parameter k determines the time translation. The defining characteristic of a wavelet or multi-resolution system is that $\varphi(t)$ satisfies a scaling equation such as:

$$\phi(t) = \sum_{k\in Z} h[k]\sqrt{2}\phi(2t-k) \tag{11}$$

, for some sequence $h[k]$ that is usually finite. The wavelet function $\psi(t)$ is derived from $\varphi(t)$. Function ψ is called "mother wavelet", which can be often constructed from "father wavelet" φ. Each coefficient can be calculated as the inner product between $f(t)$ and the respective basis element. The L^2 inner products and alternative notation for the coefficients in (10) are:

$$c[k] = c_k = \int_{-\infty}^{\infty} f(t)\phi(t-k)dt,$$
$$d_j[k] = d_{j,k} = \int_{-\infty}^{\infty} f(t)2^{j/2}\psi(2^j t - k)dt \quad (12)$$

The *c[k]* are called approximation coefficients and the $d_j[k]$ are called detail coefficients. The Haar wavelet [41], used in this study, is the earliest and its basis elements are translated and scaled versions of the following functions:

$$\phi(x) = \begin{cases} 1 & x \in [0, 1) \\ 0 & otherwise \end{cases}$$
$$\psi(x) = \begin{cases} -1 & x \in [1/2, 1) \\ 1 & otherwise \end{cases} \quad (13)$$

Generally a Haar wavelet transformation is the simplest type of wavelet [40]. In discrete form, Haar wavelet transformation is related to a mathematical operation called the Haar Transformation. The Haar Transformation serves as a prototype for all other wavelet transforms. It decomposes an array into two halves of the original length of the array. One half is a running average, and the other half is a running difference. Then Haar Transformation performs an average and difference on a pair of values. The following steps show generally the process of Haar wavelet computation. A discrete signal can be expressed in the form:

$$f = (f_1, f_2 ....., f_N) \quad (14)$$

, where N is a positive even integer which we shall refer to as the length of **f**. The values of **f** are the N real numbers $f_1, f_2, . . . , f_N$. These values are typically measured values of an analog signal g, measured at the time values $t = t_1, t_2, . . . , t_N$. That is, the values of **f** are:

$$f_1 = g(t_1), f_2 = g(t_2), ..., f_N = g(t_N) \quad (15)$$

For simplicity, in this case the assumption that the increment of time that separates each pair of successive time values is always the same is taken into consideration, which is the case examined in this study. As it was mentioned above, the Haar Transformation decomposes a discrete signal into two sub-signals of half its length. One sub-signal is a running average or *trend;* the other sub-signal is a running difference or *fluctuation.* Therefore, if the process starts with the *trend*, then for the first trend $a^1=(a_1,a_2,\ldots a_N)$ the signal f is computed by taking the running average in the following way. Its first value a_1 is computed by taking the average of the first pair of values of f and then multiplying by the square root of 2. Thus it is $d_1=[(f_1+f_2)]/$*(square root of 2).* A similar process is followed for the next values. A formula for a^1 is:

$$a_k = \frac{f_{2k-1} + f_{2k}}{\sqrt{2}} \ for\ k = 1,2 ....., N/2. \quad (16)$$

The other sub-signal is called the *first fluctuation*. The first fluctuation of the signal f, which is denoted by $d_1 = (d_1, d_2, . . . , d_{N/2})$, is computed by taking a running difference in the following way. Its first value, d_1, is found by taking half the difference of the first pair of values of f and multiplying by the square root of 2. Thus it is $d_1=[(f_1-f_2)]/$*(square root of 2).* Similarly for the next values; therefore continuing in this way all of the values of d^1 are producing according to the following formula:

$$d_k = \frac{f_{2k-1} - f_{2k}}{\sqrt{2}},\ for\ k = 1,2 ....., N/2. \quad (17)$$

The Haar Transformation is performed in several stages, or levels. The first level is the mapping $\mathbf{H_1}$ defined by:

$$f \xrightarrow{H_1} (a^1 \mid d^1) \quad (18)$$

, from a discrete signal **f** to its first trend a^1 and first fluctuation d^1. The mapping $\mathbf{H_1}$ in (18) has an inverse. Its inverse maps the transform signal $(a^1 \mid d^1)$ back to the signal **f**, via the following formula:

$$f = (\frac{a_1 + d_1}{\sqrt{2}}, \frac{a_1 - d_1}{\sqrt{2}}, ..., \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}, \frac{a_{N/2} + d_{N/2}}{\sqrt{2}}) \quad (19)$$

An important property of the Haar Transformation is that it *conserves the energies of signals.* By the *energy* of a signal **f** we mean the sum of the squares of its values. That is, the energy ε_f of a signal **f** is defined by:

$$\varepsilon_f = f_1^{\ 2} + f_2^{\ 2} + ... + f_N^{\ 2} \quad (20)$$

Once a 1-level Haar Transformation has been performed, then it is easy to repeat the process and perform multiple level Haar transforms. After performing a 1-level Haar transformation of a signal **f** we obtain a first trend a^1 and a first fluctuation d^1. The second level of a Haar Transformation is then performed by computing a second trend a^2 and a second fluctuation d^2 for the first trend a^1 only.

C.2. Stationary Wavelet Transform (SWT)

The classical DWT is not shift invariant meaning that the DWT of a translated version of a signal is not the same as the same translation of the DWT of the original signal. In order to achieve shift invariance, the Stationary Wavelet Transform (SWT) with Haar wavelet is proposed in this study. SWT overcomes the absence of translation invariance of the DWT. Unlike the DWT which down-samples the approximation and detail coefficients at each decomposition level, in the case of SWT no down-sampling is performed. This means that the approximation and detail coefficients at each level have the same length as the original signal. This determines an increased number of coefficients at each scale and more accurate localization of signal features. In addition, SWT has the translation-invariance or shift-invariance property. Thus, SWT gives larger amount about the transformed signal compared to DWT. The shift-invariant property is especially important when statistical approaches are used for analyzing the wavelet coefficients, as well as , this property is important in feature extraction applications, de-noising and detection [42].

The basic idea of SWT is simple. More specifically, the appropriate high and low pass filters are applied to the data at each level to produce two sequences at the next level. There is not decimation and the two new sequences can have the same length as the original sequence. Instead, the filters at each level are modified, by padding them with zeroes [42]. Let $\Im$ be the operator that alternates a given sequence with zeroes so that for all integers j will be $(\Im_x)_{2j} = x_j$ and $(\Im_x)_{2j+1} = 0$

Then the filters $H^{[r]}$and $G^{[r]}$ are defined to have weights $\Im^r h$ and $\Im^r g$.

D. Support Vector Machines

The next step, after transforming the original input factors or variables using SWT with Haar transform, is the application of Support Vector Machines (SVM) as a tool for classifying. SVMs are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis and have been original developed by Vapnik and Lerner [43]. A binary class supervised classification problem is usually formulated in the following way: given n training samples $\{y_k, x_k\}_{k=1}^N$, where $x_k \in \Re^n$ is the k^{th} input pattern and $y_k \in \Re$ is k^{th} output pattern, the support vector method approach aims at constructing a classifier of the following form:

$$y(x) = sign\left[\sum_{k=1}^{N} a_k y_k \psi(x, x_k) + b\right] \tag{21}$$

, where a_k are positive real constants and b is a real constant. For $\psi(.,.)$ typically there are the following choices:

$$\psi(x, x_k) = x_k^T x, \quad \text{(linear SVM)} \tag{22.a}$$

$$\psi(x, x_k) = (x_k^T x + 1)^d, \ \text{(polynomial SVM of degree)} \tag{22.b}$$

$$\psi(x, x_k) = \exp\{-\| x - x_k \|_2^2 / \sigma^2\}, \ \text{(Radial Basis)} \tag{22.c}$$

$$\psi(x, x_k) = \tanh[kx_k^T x + \theta], \ \text{(two layer neural)} \tag{22.d}$$

The task of the discriminant function or a classifier is to learn the patterns in the training samples in such a way that at a later stage it can predict reliably a y_k for an unknown x_k. SVM is fundamentally developed for such binary classification case and is extendable for multi-class situation. Like other linear classifiers, it attempts to evaluate a linear decision boundary or a linear hyperplane between the 2-classes. The classifier is constructed as:

$$\begin{aligned} &w^T \phi(x_k) + b \geq 1, \ \ if \ \ y_k = +1, \\ &w^T \phi(x_k) + b \leq 1, \ \ if \ \ y_k = -1 \end{aligned} \tag{23}$$

, which is equivalent to:

$$y_k [w^T \phi(x_k) + b] \geq 1], \ l = 1,2,,,,,,N \tag{24}$$

, where φ(.) is a nonlinear function which maps the input space into a higher dimensional space. However, this function is not explicitly constructed. In order to have the possibility to violate (24), a case a separating hyperplane in this high dimensional space does not exist, variables ξ_k are introduced such that:

$$\begin{aligned} &y_k [w^T \phi(x_k) + b] \geq 1 - \xi_k, \ k = 1,2,,,,,,N, \\ &\xi_k \geq 0, \quad k = 1,2,.....N \end{aligned}$$

(25)

According to the structural risk minimisation principle, the risk bound is minimised by formulating the optimization problem:

$$\min_{w,\xi_k} \ell_l(w, \xi_k) = \frac{1}{2} w^T w + c \sum_{k=1}^{N} \xi_k \tag{26}$$

, subject to (25). Therefore the Lagrangian is:

$$\begin{aligned} &\ell_l(w, b, \xi_k; a_k, v_k) = \ell_l(w, \xi_k) - \\ &\sum_{k=1}^{N} a_k \{y_k [w^T \phi(x_k) + b] - 1 + \xi_k\} - \sum_{k=1}^{N} v_k \xi_k \end{aligned} \tag{27}$$

Then by introducing Lagrange multipliers such that $a_k \geq 0$, $v_k \geq 0$ *(k=1,2,…N).* The solution is given by the saddle point of the Lagrangian computing the following:

$$\max_{a_k, v_k} \min_{w,b,\xi_k} \ell_l(w, b, \xi_k; a_k, v_k) \tag{28}$$

, where the first partial derivatives with respect to w, b, ξ_k, lead to the following quadratic programming problem:

$$\begin{aligned} &\max_{a_k} Q_l(a_k; \phi(x_k)) = \\ &-\frac{1}{2} \sum_{k,l=1}^{N} y_k y_l \phi(x_k)^T \phi(x_l) a_k a_l + \sum_{k=1}^{N} a_k \end{aligned} \tag{29}$$

Such that:

$$\sum_{k=1}^{N} a_k y_k = 0,\ 0 \le a_k \le c,\ k = 1,....., N. \tag{30}$$

The function $\varphi(x_k)$ in (29) is related then to $\psi(x, x_k)$ by imposing by imposing:

$$\phi(x)^T \phi(x_k) = \psi(x, x_k) \tag{31}$$

The classifier (21) is designed by solving the following:

$$\max_{a_k} Q_l(a_k; \psi(x_k, x_l)) = -\frac{1}{2}\sum_{k,l=1}^{N} y_k y_l \psi(x_k, x_l) a_k a_l + \sum_{k=1}^{N} a_k \tag{32}$$

, subject to the constraints (29). Because the matrix associated with this quadratic programming problem is not indefinite, the solution to (32) will be global [44]. Two SVM methods are examined, the least squares (see for more details [45]) and the Sequential Minimal Optimization (see for more details 46). In addition, a polynomial kernel function Eq.(22.b) with order 3 and the Radial Basis Function (22.c) are used.

IV. DATA

The period 1955 to the 2nd quarter of 2004 is set up as the estimation or the in-sample period. Then all the models are tested during the out-of-sample period, which starts from the third quarter of 2004 to 2011 in order to compare their predicting performance. The reason why the data samples are chosen in this way is because of the Stationary wavelet design. The data come directly from the Bank of England (www.bankofengland.co.uk) and the Office of National Statistics (www.ons.gov.uk). The data are based on quarterly frequency. It should be noticed that annual data are available since 1900 that could be used in order to include and capture the effects of the 1919-1921 crisis and the great depression of 1930s. However, quarterly changes, as for example the Gross Domestic Product growth, are more appropriate for forecasting purposes. Initially the variables reported in table 1 are examined. Other variables, like unemployment rate, capital formation and population could have been used, but there are only annual data for the period examined. Additionally, oil prices can be a driver which gives a warning signal of a recession. However, these are strongly correlated with inflation rate.

Generally, variables are excluded from the regressions as they are insignificant or are significantly strong correlated among them giving the wrong coefficient sign, which is one of the weaknesses of discrete choice models. On the other hand, the choice of the variables can be done also by applying factor analysis or principal components analysis. In the current study the choice is based on the best forecasting performance. Finally, obtaining forecasts (http://www.hm-treasury.gov.uk/data_forecasts_index.htm) an evaluation of the UK economy classification is applied for period 2012-2015. All the computations take place in MATLAB software.

TABLE 1
VARIABLES USED IN ESTIMATES

National Accounts		**Banking and Financial Variables**	
GDP Growth	Net Government Debt	M1 and M2 money supply	£/$ Exchange Rate
Public Sector Primary Surplus	Public Sector Net Lending	M3 and M4 money supply	FTSE 100 stock market index
Net Trade	Inflation rate	Yield curve	Bank rate

* Sources : Bank of England, Office of National Statistics

V. EMPIRICAL RESULTS

In table 2 the Logit and Probit estimates for the in-sample period are reported. The chosen variables are the Gross Domestic Product Growth and the Bank Rate. There is strong correlation among the variables leading to wrong estimated coefficient sign. For example the yield curve and the bank rate are strongly positively correlated. However, these variables have been chosen based on the optimum forecasting performance. The coefficients present the expected signs. More specifically, an increase in the Bank Rate leads to an increase of the probability of crisis occurrence, while the GDP Growth leads to decrease of the probability. Furthermore, the chosen variables are consistent with the study by Martin and Milas [36], who support that GDP growth is one of the main drivers for economic growth or recession.

TABLE 2
LOGIT AND PROBIT ESTIMATES

Panel A: Logit estimates			
Coefficients		Diagnostics	
Constant	-7.048 (1.388)***	Pseudo R square	0.4934
Gross domestic Product Growth	-1.358 (0.364)***	LR-chi square	63.33 [0.000]
Bank rate	0.504 (0.116)***	Log-Likelihood	-32.667
Panel B: Probit estimates			
Coefficients		Diagnostics	
Constant	-3.925 (0.743)***	Pseudo R square	0.4977
Gross domestic	-0.674	LR-chi square	64.18

Product Growth	(0.175)***		[0.000]
Bank rate	0.279 (0.063)***	Log-Likelihood	-32.390

Standard errors in brackets, p-vales between square brackets
*** denotes significance at 1% level

In addition the likelihood-ratio chi-square is displayed. More precisely, the likelihood ratio is defined as [47]:

$$LR_k(x^2) = 2\sum f \cdot \ln\left(\frac{f}{f_i}\right) \tag{33}$$

, where f and f_i indicate respectively the observed and the expected frequency of recession and non recession periods and k denotes the number of independent variables in the regression. Using the likelihood-ratio statistic of (33) the following hypotheses are tested: $H_0 : \beta_0 = \beta_1 = ... = \beta_k = 0$ indicating that the discrete choice models are statistically insignificant, against the alternative hypothesis $H_1 : \beta_0 \neq \beta_1 \neq ...... \neq \beta_k \neq 0$ indicating that the regressions are overall significant. In both Logit and Probit regressions the LR-chi square is higher than the respective values of critical tables, as well as, the p-value is lower than 0.05 and 0.01, for 95 per cent and 99 per cent of significance level, leading to the rejection of the null hypothesis indicating that the Logit and Probit estimates are overall statistically significant.

In table 3 the in-sample forecasts are reported. The forecasting performance between Logit and Probit models is very close. More specifically, the correct percentage classification for recession periods is 66.67 and 60.00 per cent respectively, while the correct percentage for the non-recession periods is 91.67 and 91.20. Logit model slightly outperforms the Probit model, with an overall correct classification of 90.90 to 90.40 per cent. However, based on Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) Probit's forecasting performance is slightly better to that of Logit. In fig. 1-2 the actual versus the predicted values of the Logit and Probit models respectively are presented. It should be noticed that even the non-significant variables are included into the discrete choice regressions, both in-sample and out-of-sample forecasts remain almost the same.

In panel C the classification of Markov Switching Regime model is reported. It becomes clear that the forecasting performance of MSR is superior to discrete choice models. More precisely, MSR predicts at a correct percentage the recession and non-recession periods at 94.73 and 98.89 per cent respectively, with an overall correct classification percentage equal to 98.48 per cent. In addition, MAE and RMSE are lower, to those of discrete choice models, with respective values equal to 0.0910 and 0.1625. Fig. 3 shows the actual against the forecasting values.

In panel D the forecasting performance of SVM approach is reported. It should be noticed that regarding SVM and SW-SVM approach the *GDP growth*, *Net Government Debt*, *Public Sector Primary Surplus*, *Public Sector Net Lending*, *Bank Rate* and the *£/$ Exchange Rate* have been considered as the best drivers of economic expansion and recession periods. SVM predicts correctly the recession periods at 80.00 per cent. On the other hand, regarding the non-recession periods SVM outperforms all the other approaches with a correct classification percentage equal to 100.00. The overall percentage is 97.47 per cent. Similarly, in panel E the forecasts of the proposed SW-SVM approach are reported. This approach predicts correctly the recession periods at 88.00 per cent, the correct percentage for the non-recession periods is 100.00, while its overall correct percentage is 98.48 per cent as in MSR. In fig. 4-5 the SVM and SW-SVM forecasts versus the actual values are respectively presented. It should be noted that the SMO algorithm and a Polynomial Kernel function of degree order equal to 3 have been applied in the case of the SVM and SW-SVM approaches.

TABLE 3
IN-SAMPLE FORECASTS Q1 1955- Q2 2004

Panel A: Logit				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	4	2	6	66.67%
No Recession	16	176	192	91.67%
Total	20	178		90.90%
MAE	0.1010	RMSE	0.1822	
Panel B: Probit				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	3	2	5	60.00%
No Recession	17	176	193	91.20%
Total	20	178		90.40%
MAE	0.0989	RMSE	0.1717	
Panel C: Markov Switching Regime				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	18	1	19	94.73%
No Recession	2	177	179	98.89%
Total	20	178		98.48%
MAE	0.0910	RMSE	0.1625	
Panel D: SVM				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	20	5	25	80.00%
No Recession	0	173	173	100.00%
Total	20	178		97.47%
MAE	0.0987	RMSE	0.1703	
Panel E: SW-SVM				
	True		Total	Correctly classified

Classified	Recession	No Recession		
Recession	22	3	25	88.00%
No Recession	0	173	173	100.00%
Total	20	178		98.48%
MAE	0.0963	RMSE	0.1653	

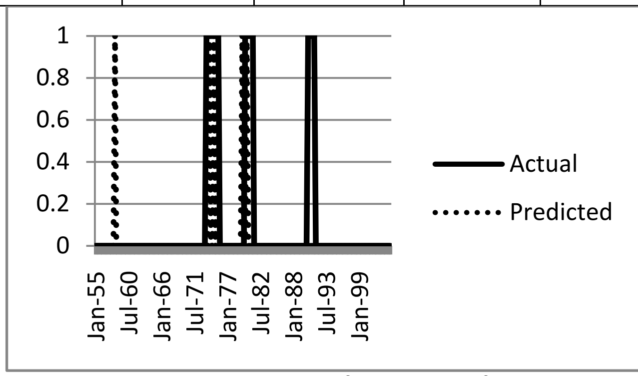

Fig. 1. Actual vs predicted values of Logit model for the in-sample period Q1 1955- Q2 2004

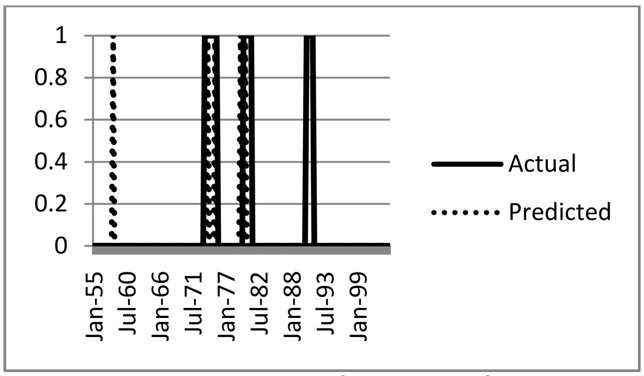

Fig. 2. Actual vs predicted values of Probit model for the in-sample period Q1 1955- Q2 2004

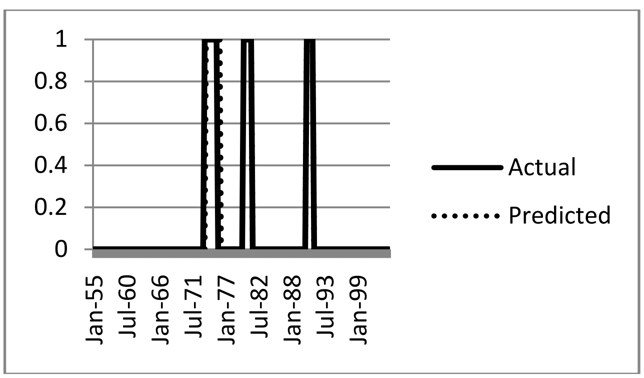

Fig. 3. Actual vs predicted values of MSR model for the in-sample period Q1 1955- Q2 2004

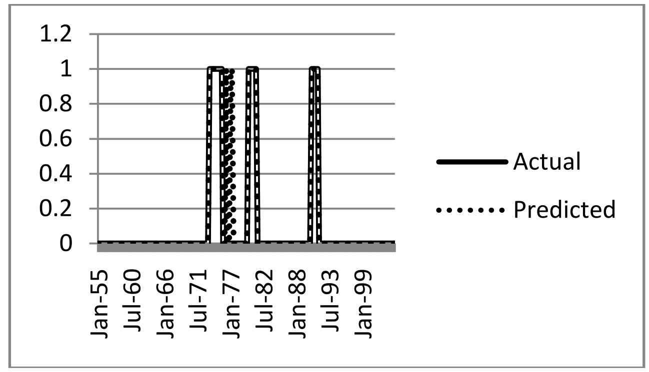

Fig. 4. Actual vs predicted values of SVM model for the in-sample period Q1 1955- Q2 2004

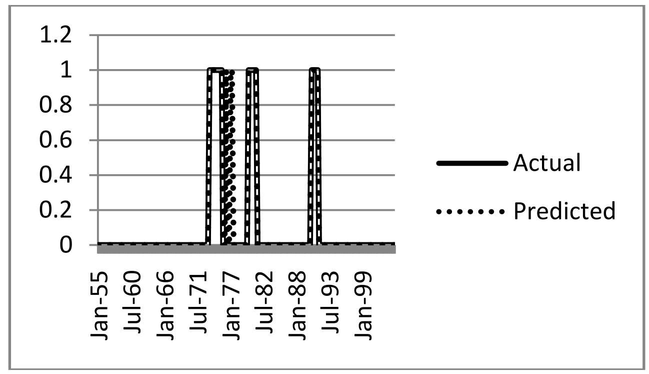

Fig. 5. Actual vs predicted values of SW-SVM model for the in-sample period Q1 1955- Q2 2004

In table 4 the forecasts for the out-of-sample period Q3 2004- Q4 2011 are reported. Both Logit and Probit present exactly the same classification percentage, as it can be seen in fig. 6. More specifically, discrete choice models predict correctly the recession and non-recession periods at 66.67 and 87.50 per cent respectively, with an overall correct classification percentage 83.33 per cent. Based on MAE and RMSE Probit model outperforms Logit.

Similarly in Panel C the out-of-sample forecasts using MSR are reported. MSR predicts correctly the recession and non-recession periods at 75.00 and 84.61 per cent respectively, while the overall classification percentage is 83.33 per cent. In addition Probit model outperforms Logit and MSR models based on MAE and RMSE.

In Panel D the forecasts of SVM approach are presented. Besides its poor performance in the in-sample period, SVM slightly outperforms the discrete choice models. This indicates that through the training process SVM is an efficient tool for the testing sample or the out-of-sample forecasts. More specifically, the correct classification percentage for recession and non-recession periods is 75.00 and 95.45 per cent respectively, and the overall classification percentage is 90.00 per cent. Furthermore, discrete choice models predict exactly the beginning of the crisis of 2007-2008, as fig. 8 shows, while MSR presents a delay of two quarters, based on fig. 7. In addition, SVM gives a warning signal one quarter before the double dip recession which took place in the last quarter of

2010. Finally, in panel E the forecasting classification using the proposed SW-SVM approach is reported. In that case this approach outperforms all models, with a correct classification percentage for recession and non-recession periods at 85.71 and 95.65 per cent respectively, and an overall correct percentage of 93.33. Additionally, based on fig. 8 SW-SVM approach gives a warning signal one quarter before the financial crisis of 2007-2008, as well, as one quarter before the double dip recession which took place on the last quarter of 2011, while it predicts perfectly the negative performance of the UK economy in the last quarter of 2010.

TABLE 4
OUT-OF-SAMPLE FORECASTS Q3 2004- Q4 2011

Panel A: Logit				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	4	2	6	66.67%
No Recession	3	21	24	87.50%
Total	7	23		83.33%
MAE	0.2014	RMSE	0.3070	
Panel B: Probit				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	4	2	6	66.67%
No Recession	3	21	24	87.50%
Total	7	23		83.33%
MAE	0.1837	RMSE	0.2736	
Panel C: Markov Switching Regime MSR				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	3	1	4	75.00%
No Recession	4	22	26	84.61%
Total	7	23		83.33%
MAE	0.1844	RMSE	0.2886	
Panel D: SVM				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	6	2	8	75.00%
No Recession	1	21	22	95.45%
Total	7	23		90.00%
MAE	0.1868	RMSE	0.2839	
Panel E: SW-SVM				
	True		Total	Correctly classified
Classified	Recession	No Recession		
Recession	6	1	7	85.71%
No Recession	1	22	23	95.65%
Total	7	23		93.33%
MAE	0.1733	RMSE	0.2678	

Table 5 shows the forecasts of the examined models for the period 2012-2015. Discrete choice and MSR models give a signal of non-recession. SVM gives a recession signal for the whole period 2012-2014, while SW-SVM gives an early warning signal for the first and second quarters of 2012 and the second and third quarters of 2013. However, the reliability of the prediction is mainly based on estimated forecasts for the independent variables or the inputs. The bank rate probably will remain constant at 0.5 during 2013, with a possible cut in 2014 at 0.4 and then an increase again at 0.5 in 2015.

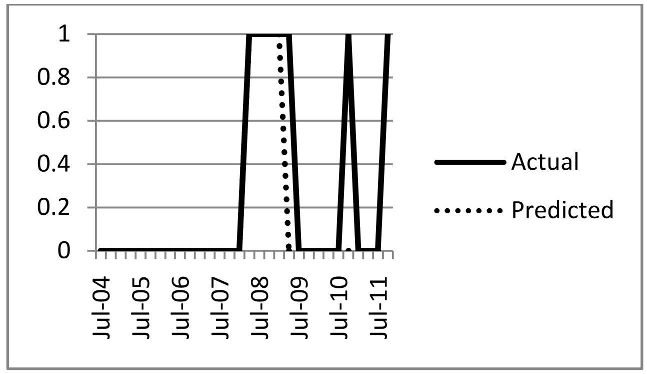

Fig. 6. Actual vs predicted values of discrete choice models for the out-of-sample period Q3 2004- Q4 2011

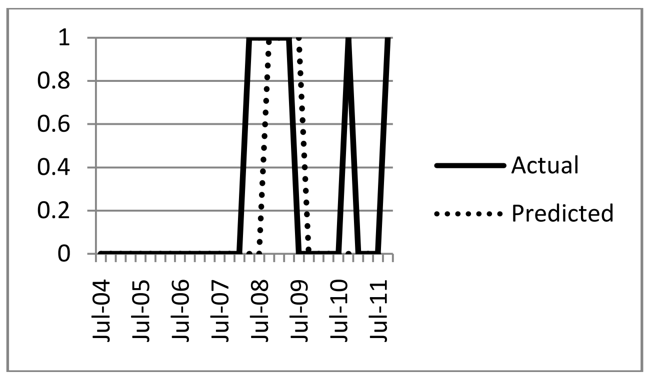

Fig. 7. Actual vs predicted values of MSR model for the out-of-sample period Q3 2004- Q4 2011

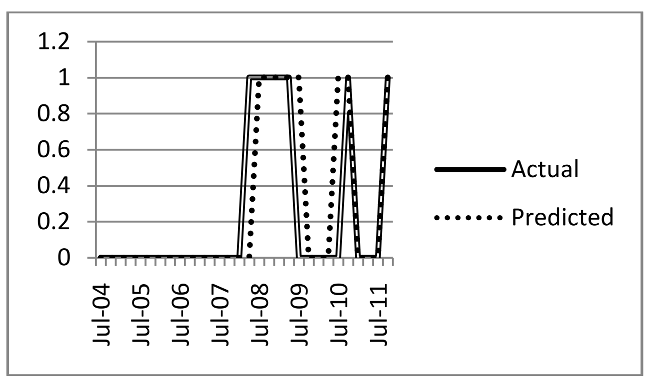

Fig. 8. Actual vs predicted values of SVM model for the out-of-sample period Q3 2004- Q4 2011

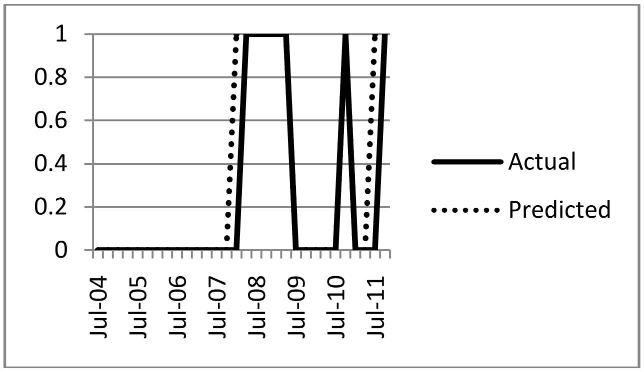

Fig. 9. Actual vs predicted values of SW-SVM model for the out-of-sample period Q3 2004- Q4 2011

Additionally, the GDP growth has been revised at around 1-1.5 per cent for the period 2013-2015 after the initial forecasts which were estimating a growth of more than 2.5 per cent. Therefore, if the actual GDP growth reaches 2.5 per cent or the Net Government debt will be increased at lower levels than the forecasting then there will be no recession for the UK economy. On the other hand, if the GDP growth will be lower than the expected levels, around 0.5-1.0 per cent, and if at the same time the Net Government debt reach the 90.00 per cent of GDP, which is expected by 2015, then the SW-SVM gives a warning signal of a recession period in 2014. Generally, SVM and SW-SVM approaches suggest that UK economy will come back to growth in 2015.

TABLE 5
FORECASTS FOR 2012-2015

Period	Logit-Probit-MSR	SVM	SW-SVM
2012	No recession	Recession	Recession at Quarters 1 and 2
2013	No recession	Recession	Recession at Quarters 2-3
2014	No recession	Recession	No recession
2015	No recession	No recession	No recession

Even if SW-SVM and other techniques are able to predict the crisis periods this is not enough, because the main responsibility of the governments, the financial managers, and economists is to guarantee that a crisis like that will not be repeated in the future. In addition, there are other factors which can be used, like unemployment rate and productivity. Therefore, even if the GDP will be actually increase at 2.5 per cent this is not necessary enough to create jobs and reduce the unemployment rate and if there would be any recovery this would be very flat.

VI. CONCLUSION

This study examined various models for the prediction of economic expansions and recession in the economy of United Kingdom. Four different approaches have been applied, the discrete choice models, Logit and Probit, a Markov Switching Autoregressive models, Support Vector Machines and Stationary Discrete Wavelets Transform with Support Vector Machine classifiers. The findings support that the MSR, SVM and SW-SVM approaches outperform the forecasting performance of the discrete choice econometric modelling. In addition, SVM and SW-SVM shows a pre-warning signal that a new recession in UK economy could take place in 2013-2014.

REFERENCES

[1] P. Krugman, "A Model of Balance of Payment crises," Journal of Money, Credit and Banking, vol. 11, pp. 311-325, 1979.

[2] R.P. Flood, P.M. Garber, "Collapsing Exchange Rates Regime: Some Linear Examples," *Journal of International Economics*, vol. 17, pp. 1-13, 1984.

[3] M. Obstfeld, "The logic of currency crises," *Cahiers Economiques et Monetaires*, vol. 43, pp. 189-213, 1994.

[4] P. Krugman, "Currency Crises" Mimeo. Massachusetts Institute of Technology, Cambridge. Massachusetts; available at web.mit.edu/krugman/www/crises.html, 1998.

[5] F. Allen, D. Gale, "Bubbles and Crisis," *Economic Journal*, vol. 110, pp. 236-255, 2000.

[6] J. Frankel, A.K. Rose, "Currency crashes in emerging Markets: An Empirical Treatment," International Finance discussion Papers 534. Board of Governors of the Federal Reserve System. Washington. D.C., 1996.

[7] A. Demirguc-Kunt, E. Detragiache, "The determinants of Banking crises in developing and developed countries," *IMF Staff Papers*, 45, pp. 81-109, 1998.

[8] B. Eichengreen, A.K. Rose, "Staying Afloat When the Wind Shifts: External Factors and Emerging-Market Banking crises," National Bureau of Economic Research Working Papers 6370, 1998.

[9] R. Glick, A.K. Rose, "Contagion and trade: Why currency crises are regional?," National Bureau of Economic Research. Working Papers 6806, 1998.

[10] R. Glick, R. Moreno, "Money and credit, competitiveness and currency crises in Asia and Latin America," Federal Reserve Bank of San Francisco papers BP99-01. Center for Pacific Basin Money and economic Studies, 1999.

[11] A. Berg, C. Pattillo, "Are currency crises predictable: A test," *International Monetary Fund Staff Paper*, vol. 46, no. 2, pp. 107-138, 1998.

[12] L.G. Kaminsky, S. Lizondo, C.M. Reinhart, "Leading Indicators of currency crises," *IMF Staff Papers*, vol. 45, pp. 1-48, 1998.

[13] D. Hardy, C. Pazabasioglu, "Determinants and leading indicators of banking crises: Further Evidence.," *International Monetary Fund paper*, vol. 46, no., pp. 247-258, 1999.

[14] L.G. Kaminsky, C.M. Reinhart, "The Twin crises: The causes of Banking and Balance of Payments Problems," Federal reserve Board discussion Papers 544. Board of Governors of the Federal Reserve System, Washington. D.C., 1996.

[15] J. Hamilton, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, vol. 57, no. 2, pp.357-384, 1989.

[16] J. Hamilton, “Analysis of Time Series Subject to Changes in Regime,” *Journal of Econometrics*, vol. 45, pp. 39-70, 1990.
[17] C. Engel, C.S. Hakkio, “The Distribution of Exchange Rates in the EMS,” *International Journal of Finance and Economics*, vol. 1, pp. 55-67, 1996.
[18] V. Cerra, S.C. Saxena, “Contagion, Monsoons and Domestic Turmoil in Indonesia’s Currency Crisis,” *Review of International Economics*, vol. 10, no. 1, pp. 36-44, 2002.
[19] A. Abiad, “Early-Warning System: A Survey and a Regime-Switching Approach,” IMF Working Paper WP/03/32, 2003.
[20] R.S. Mariano, A.G. Abiad, A.G. Gultekin, T. Shabbir, A.H.H. Tan, “Markov Chains in Predictive Models of Currency Crises – with Applications to Southeast Asia,” *Taiwan Economic Review*, vol. 31, no. 4, pp. 401-436, 2003.
[21] W.Y. Cheng, E. Su, S.J. Li, “A Financial distress pre-warning study by fuzzy regression model of TSE-listed companies,” *Asian Academy of Management Journal of Accounting and Finance,* vol. 2, no. 2, 75-93, 2006.
[22] A. Nachev, B. Stoyanov, “Default ARTMAP Neural Networks for Financial Diagnosis,” in *DMIN Proceedings* CSREA Press, pp. 3-9, 2007.
[23] P.K. Coats, F.L. Fant, “Recognizing financial distress patterns using a neural network tool,” Financial Management, vol. 22, pp.142-155, 1993.
[24] E. Fernandez, I. Olmeda, “Bankruptcy prediction with artificial neural networks,” *Lecture Notes of Computer Science*, pp. 1142-1146, 1995.
[25] G. Zhang, Y. HU, E.B. Patuwo, “Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis,” *European Journal of Operation Research*, vol. 116, pp. 16-32, 1999.
[26] P.L. Brockett, L.L. Golden, J. Jang, C. Yang, “A comparison of neural network, statistical methods and variable choice for life insurers’ financial distress prediction,” *Journal of Risk and Insurance*, vol. 73, pp. 397-419, 2006.
[27] H., Ni, H. Yin, “Exchange rate prediction using hybrid neural networks and trading indicators,” *Neurocomputing*, vol. 72, pp. 2815-2823, 2009.
[28] W.S. Chen, Y.K. Du, “Using neural networks and data mining techniques for the financial distress prediction model,” *Expert Systems with Applications: An International Journal*, vol. 26, no. 2, pp. 4075-4086, 2009.
[29] M. Odom, R. Sharda, “A neural networks model for bankruptcy prediction,” *Proceedings of the IEEE International Conference on Neural Network*, vol. 2, pp.163–168, 1990.
[30] E. Giovanis, “Application of Logit Model and Self-organizing Maps (SoM) for the prediction of Financial Crisis Periods in US Economy,” *Journal of Financial Economic Policy*, vol. 2, pp. 98-125, 2010.
[31] E. Giovanis, “Study of Discrete Choice Models and Adaptive Neuro-Fuzzy Inference System in the Prediction of Economic Crisis Periods in USA,” *Economic Analysis and Policy (EAP)*, vol. 42, no. 1, pp. 79-96, 2012
[32] M.C. Lee, C. To, “Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress,” *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 1, no. 3, pp. 31-43, 2010.
[33] B.T. Chen, M.Y. Chen, “Applying particles swarm optimization for support vector machines on predicting company financial crisis,” *International Conference on Business and Economics Research*. IACSIT Press. Kuala Lumpur, Malaysia, vol. 1, pp.301-305, 2010.
[34] X. Song, Y. Ding, J. Huang, Y. Ge, “Feature selection for support vector machine in financial crisis prediction: a case study in China,” *Expert Systems, The Journal of Knowledge Engineering*, vol. 27, pp.299-310, 2010.
[35] J.J. Ahn, K.J., Oh, T.Y. Kim, D.H. Kim, “Using Support Vector Machine to Development Early Warning System for Financial Crisis,” Expert Systems with Applications: An International Journal, vol. 38, no.4, 2966-2973, 2011.
[36] C. Martin, C. Milas, “Causes of the Financial Crisis: an Assessment Using UK Data,” Working Paper Series. 10_09, Rimini Centre for Economic Analysis. Rimini, Italy, 2009.
[37] M.D. Bordo, J.S. Landon Lane, “The Global Financial Crisis: Is It Unprecedented?,” NBER Working Paper No. 16589, 2010.
[38] W.H. Greene, *Econometric Analysis*. 6th ed.. NJ: Prentice Hall, 2008.
[39] J.S. Cramer, *Logit Models from Economics and Other Fields*. Cambridge: Cambridge University Press, 2003.
[40] J.S. Walker, *A Primer on Wavelets and Scientific Applications*, 2nd ed. Chapman and Hall/CRC Taylor and Francis Group. LLC, 2008.
[41] A. Haar, Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, vol. 69, no. 3, pp. 331-371, 1910.
[42] Y. Meyer, *Wavelets and Operators*. Cambridge University Press. Cambridge, 1992
[43] V. Vapnik, A. Lerner, Pattern Recognition Using Generalized Portrait Method, *Automation and Remote Control*, vol. 24, 774–780, pp. 1963.
[44] R. Fletcher, *Practical Methods of Optimization*. 2nd ed. John Wiley and Sons, 1987.
[45] J.A.K. Suykens, J. Vandewalle, *Least Squares Support Vector Machine Classifiers*, Kluwer Academic Publishers, 1998.
[46] J. Platt, “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines,” *Microsoft Research. Technical Report MSR-TR-98-14,* April 21, 1998.
[47] A. Agresti, *Categorical Data Analysis. Second Edition*. NJ: John Wiley & Sons, 2002.