Financial Bubble Detection : A Non-Linear Method with Application to SP 500

Michaelides, Panayotis G. and Tsionas, Efthymios and Konstantakis, Konstantinos

2016

FINANCIAL BUBBLE DETECTION: A NON-LINEAR METHOD WITH APPLICATION TO S&P 500

Panayotis G. Michaelides[1]
National Technical University of Athens

Efthymios G. Tsionas
Lancaster University

Konstantinos N. Konstantakis
National Technical University of Athens

Abstract

The modeling process of bubbles, using advanced mathematical and econometric techniques, is a young field of research. In this context, significant model misspecification could result from ignoring potential non-linearities. More precisely, the present paper attempts to detect and date non-linear bubble episodes. To do so, we use Neural Networks tocapture the neglected non-linearities. Also, we provide a recursive dating procedure for bubble episodes. When using data on stock price-dividend ratio S&P500 (1871.1-2014.6), employing Bayesian techniques, the proposed approach identifies more episodes than otherbubble tests in the literature, while the common episodes are, in general, found to have a longer duration, which is evidence of an early warning mechanism (EWM) thatcouldhave important policy implications.

Keywords: Bubbles, Non-linearities, Neural Networks, EWM, S&P500

JEL Classification: C5, G1

[1] Laboratory of Theoretical and Applied Economics; School of Applied Mathematical and Physical Sciences; National Technical University of Athens; Heroon Polytechneiou 9, 157.80; Zografou Campus; Greece; Tel.: +302107721624; Fax: +302107721618; E-mail: pmichael@central.ntua.gr (*corresponding author*).

FINANCIAL BUBBLE DETECTION: A NON-LINEAR METHOD WITH APPLICATION TO S&P 500

1. Introduction

In August 2015,the Chinese stock market lost over 30% of its stock value, experiencing one of the worst stock market crashes in recent financial history. Despite the efforts made by the Chinese Government and the Chinese Central Bank to prevent the crash by implementing a strict legislatory framework on short selling as well as by providing huge cash injections to brokers so as to stimulate stock demand, the Shanghai Stock Exchange experienced an unprecedented crash. As a result, on the 24th of August, the Shanghai Stock Exchange experienced an overall devaluation of approximately 8% in stock prices, the so-called "Black Monday" of the Chinese Stock Market (The New York Times, 25 August 2015).

Despite the fact that in the long history of financial bubbles the Chinese case isnot the first and certainly not the last one, only limited attention has been paid by the scientific community tocreatingarigorous androbust framework for the detection of bubble formationbased on a credible Early Warning Mechanism (EWM).In general, EWMs are essential components of time-varying macroprudential policies that can help reduce the high losses associated with both banking and country specific crises. In this context, the EWMs employed should not only have sound statistical forecasting power, but also need to satisfy several additional requirements.

Analytically, the importance of bubble dating lies on the appropriate timing, which is a crucial requirement for EWMs. In this context, macroprudential policies need time before they become effective (Basel Committee 2010) and, hence, signals should need to arrive at a relatively early stage in order to prevent policy measures from being costly (Caruana 2010). The stability of the signal is a second, largely overlooked, requirement. More precisely, policy makers tend to base decisions on trends rather than reacting to changes in signaling variables immediately (Bernanke 2004). Meanwhile, the gradual implementation of policy measures may also allow policy makers to affect market expectations more efficiently and deal with uncertainties in the transmission mechanism (CGFS 2012). Finally, a last requirement is that EWM signals should be easy to interpret, as any signals that do not "make sense" are likely to be ignored by policy makers (Önkal et al 2002; Lawrence et al 2006). In sum, well designed EWIs, in terms of timing and signal processing, can reduce uncertainty and allow for more decisive policy action.

Thus far, one of the main reasons behind the inability of most models to capture the formation of bubbles,at a relatively early stage, is the fact that bubble formation has inherent non-linear characteristics,which are difficult to capture using standard linear models. This, clearly, implies that any econometric test that aims at capturing the formation of bubbles, especiallyat an early stage, should be able to capture their non-linear character.

Additionally, another equally important challenge for the econometric detection of bubbles is their dating, in the sense that an econometric test should be able to accurately date the bubble periods detected in the sample.Of course, early detection and accurate dating of financial bubbles could have important policy implications, especially for central bankers and policy makers since it could assist in the implementationofrelevant policy actionsthat could potentially ease the consequences of bubbles. More specifically, the importance of early identification lies in the timing of specific countermeasures that could potentially prevent: a) the magnitude of a potential collapse through regulatory interventions in the financial markets; b) the potential downturn effects of bubble collapse in the economy through appropriate inflation targeting, and c) the devastating spillover effects in the global economy through interest rate and/or exchange rate setting.

Due to the fact that, according to the recent financial history of bubbles, more than one bubble could occur in the same sample period (Ferguson 2008), any econometric test for bubble detection should be structured upon flexible backward and/or forward recursive estimation techniques.However, relatively limited research has been done in the literature using recursive estimation techniques for dating multiple bubble episodes. See Phillips and Yu (2011), and Phillips et al. (2011a, 2011b, 2013, 2014, 2015a) and Phillips et al. (2015b) [hereafter PSY].

Meanwhile, nonlinear economic models have become quite popular lately, because economic data exhibit significant non-linearities. To this end, in this paper, we propose a rigorous and robust mathematical and econometric framework for the detection of bubbles, which is structured upon ArtificialNeuralNetworks (ANN),that are perfectly capable of capturing any neglected non-linearity. In fact, this is the first paper in the relevant literature, to the best of our knowledge, which employs ANNs, to capture neglected non-linearities in bubbles.

After all, according to PSY, the use of computationally efficient dating methods "over long historical periods presents a more serious econometric challenge due to the complexity of the *nonlinear* structure and break mechanisms that are inherent in multiple-bubble phenomena within the same sample period". Finally, our approach provides a recursive algorithm for the accurate detection of bubbles, which serves as an EWM that could be used in order to guide a policy decision in an uncertain environment, without the need of taking into consideration the policy maker's preferences (e.g. Pesaranand Skouras 2002; Granger and Machina 2006; Baxa et *al.* 2013).

In brief, the present paper contributes to the literature in the following ways: (a) It establishes a rigorous framework, based on ANNs, under which bubble detection could be achieved, while emphasizingonthe presence of non-linearities; (b) It provides a new algorithm for the accurate and early detection of bubble formation, as well as for the identification of potential explosive behaviors; (c) it illustrates the proposed testby early detecting and capturing accurately the bubble episodes that are present in the S&P 500 index for the time period 1871 (M1)-2014 (M6), and by identifying more episodes compared to a competitive methodology in the literature.

This paper is structured as follows:insection 2, a review of the literature takes place; section3 presents the theoretical model; section 4 sets out the proposed non-linear test; section 5presents the empiricalanalysis;finally, section 6concludes.

2. Related Literature

According to Kindleberger (1978) a bubble is defined as "an upward price movement over an extended range that then implodes". Brunnermeier (2009) argued that bubbles "are typically associated with dramatic asset price increases followed by a collapse", whereas Garber (2000) defined a bubble as the part of the price movement that cannot be explained by fundamentals. Also, Barlevy (2007) described a bubble as "a situation where an asset´s price exceeds the fundamental value of the asset". In brief, a bubble occurs when the market value is higher than the fundamental (Diba and Grossman 1988). Some researchers (e.g. Wu 1997) define bubbles as the difference between the fundamental value and the market price allowing, thus, for negative bubbles.

Reasons for the occurrence of bubbles include, among other things, greed (Kindleberger 1978), introduction of breakthrough technologies or financial innovations (e.g. Perez 2009);existence ofrational and irrational traders (Dufwenberg, Lindqvist and Moore 2005; Hong, Scheinkman andXiong2007); institutional restrictions on short selling (Haruvy and Noussair, 2006); herding(DeMarzo, Kaniel and Kremer 2008), speculating investors (Greenwood and Nagel 2005; Scheinkman and Xiong2002), and "bubble riding" (Abreu and Brunnermeier 2003, and Temin and Voth 2003).

Despite the fact that several approaches, even seminal ones (e.g. Fama, 1965), have denied the possibility of bubbles in financial markets, the phenomenon has made its appearance long ago (e.g. Dutch Tulipmania [1634-

1637], Mississippi Bubble [1719–1720]) and has often led to generalized and deep economic recessions. As a result, Fama's*Efficient Market Hypothesis* and other similar theories have not always found so much support. After all, probably the most prominent economist, who considered the existence of bubbles in financial markets, was John Maynard Keynes (1936).

Following the related literature on financial bubble detection,Shiller (1981) and Lerroy and Porter (1981) were probably the first to develop variance bound tests for equity prices. Despite the fact that Shiller's (1981) variance bound test was not initially developed for bubble detection, the works of Blanchard and Watson (1982) and Tirole (1985) suggested that violation of variance bounds could be attributed to the presence of bubbles. Nevertheless, the variance bound tests were heavily criticized by a number of authors likeFlavin (1983), Mash and Merton (1983), Mankiw et al. (1985), Kleidon (1986) and Flood et al. (1994), due to the fact that the variance bound tests could fail not only if bubbles exist but also if any of the assumptions of the present value model is violated.

In a different approach, West (1987) developed a two-step test for the identification of bubbles in equity prices based on Euler's equation of no arbitrage process and the autoregressive process of dividends that governs the market fundamental stock price. Despite the fact that West's (1987) test was more attractive than the variance bound test as it explicitly incorporated the null hypothesis of no bubbles, once again Dezbakhsh and Demirguc-Kunt (1990), as well as Flood et al. (1994), criticized the econometric procedure of the test because it exhibited significant size distortions in small samples.

Another popular approach for bubble detection was the one proposed by Diba and Grossman (1987, 1988a, 1988b),who tried to exploit the theoretical properties of bubbles. Their test allowed for unobserved fundamentals in the market fundamental price and a bubble would exist if the dividends and stock prices did not have the same order of integration.

However, Evans (1991) criticized the test of Diba and Grossman (1988b) byarguingthat it was unable to capture a periodically collapsing bubble.

Following Evans (1991), a vast literature emerged concerning the detection of bubbles, like Hall and Sola (1993), van Norden (1996), van Norden and Vigfusson (1998), Driffil and Sola (1998), and Hall et al. (1999) who incorporated regime switching models for bubble detection. In the meantime,in a seemingly unrealtedapproach, Wu (1997) used Kalman filtering in an attempt to test for bubbles, while Wu and Xiao (2002) tried to establish a test for bubbles based on the residuals of the cointegrating equation between dividends and stock prices.

This signified the formation of the latest strand in the literature of bubble detection where researchers based the existence and detection of bubbles on the unit root behavior of key fundamental financial variables. In a prominent paper, Phillips and Yu (2011) introduced a recursive regression methodology in order to analyze the bubble characteristics of various financial time series during the subprime crisis. Phillips et al. (2011a) extended the work of Phillips and Yu (2011) by introducing a relevant econometric framework where more than one bubbles could exist in the same sample. Phillips et al. (2011b) provided the identification conditions regarding the explosive behavior of bubbles based on the unit root behavior of relevant financial time series.

Breitung and Holmes (2012) investigated the power properties of rational bubbles considering a large variety of testing alternatives, while Breitung and Kruse (2013) showed that structural break Chow-type tests have considerable power for the detection of bubbles. Again, Phillips et al. (2013) illustrated their proposed bubble specification and dating algorithm using data from S&P500 series, while Phillips et al. (2014) provided the asymptotic properties of the related bubble dating and identification conditions.Finally, in two seminalworks,Phillips et al. (2015a) and PSYprovided probably the only framework, thus far, in the existing literature, under which an EWM is established for the detection of multiple bubble episodes.

3. The Theoretical Model

From a technical point of view, probably the most important feature of bubbles is that they are characterized by explosive growth patterns, despite the fact that speculative movements are often assumed to follow a random walk process (e.g. Blanchard and Watson 1982, Campbell, Lo, and MacKinlay 1997). And it is exactly this, the most common way to identify a bubble, by applying tests for a structural change from a random walk regime to an explosive one. Such tests have been developed by Phillips, et al. (2011a), Phillips and Yu (2011), Homm and Breitung (2012), Phillips et al. (2014), and PSY.

3.1 Time Series Model

From a technical perspective, the identification of bubbles involves the use of key financial time series variables, such as dividends, stock prices, equity prices etc.

For any financial time series variable, $x_{t_j}, j \in J$, we will make a number of fairly standard assumptions:

Assumption 1: The time series $x_t$is assumed to conform to the standard additive component model, i.e. every financial time series variable $x_{t_i}, i \in I$, follows the process:

$x_{t_i} = s_{t_i} + g_{t_i} + c_{t_i} + \varepsilon_{t_i}, \ i \in I$ **(1)**

where: s_{t_i} is the seasonal component, g_{t_i} is the trend component, c_{t_i} is the cyclical component and $\varepsilon_{t_i} \sim N(0, \sigma^2)$ is the error term.

For the sake of simplicity, and without loss of generality, we also make the following assumption:

Assumption 2: The trend and constant term of the series $x_{t_i}, i \in I$, are both assumed to be equal to 0.

In case, (deterministic) terms are to be considered, the standard procedure is to apply demeaning and detrending procedures before computing the relevant test statistics.

Now, we present the general formulation of the unit-root test upon which the econometric testing of bubbleswill be based.

Assumption 3: The unit root detection is described by the following model: $\Delta x_{t_i} = \rho x_{t_i-1} \cdot G(x_{t_i-1}; \gamma) + \varepsilon_{t_i}, t_i = 1, \ldots, T, i \in I$ **(2)**

where$\varepsilon_{t_i} \sim NID\ (0, \sigma^2)$ and G is a sufficientlysmooth function.

With reference to the aforementioned general specification, without deterministic components, the most popular unit root test in the literature, i.e. the traditional Dickey Fuller (D.F.) test, is based on the t-statistic of $\rho$from the model:

$\Delta x_{t_i} = \rho x_{t_{i-1}} + \varepsilon_i, \ i \in I$ **(3)**

The null hypothesis, H_0, of a unit root is parameterized by $\rho = 0$.

The vast majority of empirical tests in the literatureare based on alternative forms of the D.F. test above (Equation 3). However, some other unit root testing attempts are also present in the literature, where researchers have attempted to capture bubbles based onsomenon-linearunit root specification.More precisely, Kapetanios et *al.* (2003) or KSS extended thestandardapproach on unit root testing through the introduction of a so-called exponentialsmooth transition autoregressive (ESTAR) model and decided to consider the following ESTAR process, emphasizing the expected low power of the linear augmented D.F. test, when applied to such a series:

$\Delta x_{t_i} = \gamma x_{t_i-1}\{1 - \exp(-\theta {x^2}_{t_i-1})\} + \varepsilon_{t_i}), \, i \in I$ **(4)**

The analysis of KSS focuses on θ, with $H0: \theta = 0$ and $H1: \theta > 0$. As γ is unidentified under $H0, \theta = 0$ cannot be tested. Hence, they based their work onLuukkonenet*al.* (1988) and employed a first-order Taylor series approximation to the ESTAR model under the null $H0: \theta = 0$. The relevant equation is:

$\Delta x_{t_i} = \rho {x^3}_{t_i-1} + \varepsilon_{t_i}, \, i \in I$ **(5)**

where the nonlinear test relies on the *t*-statistic of ρ from the O.L.S. regression on the previous equation.

However, it should be noted that the aforementioned models (i.e. linear, or ESTAR, etc)are not grounded on some formal mathematical or statistical criterion, but rather on the modeling choices of each individual researcher. Therefore, both attempts that are equivalent to the assumption that either $G(x_{t_i-1}; \gamma) \equiv 1$ or $G(x_{t_i-1}; \gamma) \equiv {x^3}_{t_i-1}, i \in I$, which are implied by the linear and ESTAR models, respectively, need to be reconsidered.

For instance, changing the degree of the implied polynomial assumed in the aforementioned ESTAR process would lead to another exponential power of the relevant test.Hence, misspecification issues arise from ignoring potential nonlinear terms. As a result, it would seem absolutely imperative to test for the presence of nonlinear terms.

In this work, in order to overcome these serious drawbacks which result from the arbitrarily assumptions about the processes to be followed, instead of fitting the *G*function with a pre-specified equation, we will use an Artificial Neural Network (ANN) to let the dataset itself serve as evidence to support the model's approximation of the underlying specification.

3.2ANNs Formulation

As we have seen, the main idea is to express the arbitrary specification $\Delta x_{t_i} = \rho x_{t_i-1} \cdot G(x_{t_i} - 1;\ \gamma), i \in I$not as a pre-specified form based on *a priori* assumptions, but rather let the dataset itself determine the specification of the underlying process. In other words, instead of fitting Δx_{t_i} with a pre-specified functional form, ANNs let the dataset itself serve as evidence to support the model's approximation of thespecification.In what follows, we proceed by providing a formal definition of ANNs.

Definition1: ANNs are collections of functions that relate an output variable Y to certain input variables $\boldsymbol{X}' = [X_1, \ldots, X_M]$. The input variables are combined linearly to form N intermediate variables $Z_1, \ldots, Z_N$: $\boldsymbol{Z}_N = X'\beta_n (k = 1, \ldots, N)$, where $\beta_n \in \mathbb{R}^N$are parameter vectors. The intermediate variables are combined non-linearly to produce Y:

$$Y = \sum_{n=1}^{N} a_n \varphi(Z_n) \quad \textbf{(6)}$$

where: φ is an activation function, the α_n 's are parameters and N is the number of intermediate nodes (Kuan and White 1994).

We make use of a single layer ANN to avoid computational and energetic requirements (see Sanger 1989).Hence, it is worth mentioning that the mechanism behind ANNs is that they combine simple units with intermediate nodes, so they can approximate any smooth nonlinearity (Chan and Genovese 2001). In fact, ANNs provide very good approximations to a large class of arbitrary functions while keeping the number of parameters to a minimum(Hornik et *al.* 1989, 1990). Also, they can approximate their derivatives, a fact which justifies their success (Hornik et *al.* 1990, Brasili and Siltzia 2003).

To sum up, ANNs are data-driven and self-adaptive, nonlinear methods that do not require specific assumptions about the underlying specification (Zhang and Berardi 2001). In addition, theyare universal approximators of functions.In this paper, we useaANN formulation in order tocaptureand model nonlinearities in bubbles.

3.3 Mathematical Properties

As we have seen in the previous section, the main idea for capturing a financial bubble episode is to thoroughly investigate the respective unit root behavior of the financial time series variable. To this end, using the general specification of unit root detection, i.e.$\Delta x_{t_i} = \rho x_{t_i - 1} \cdot G\left(x_{t_i - 1}; \gamma\right), j \in J$we will formally approximate the function G, using anANN. To do so, we will make use of the formal definitions of open set, open covering, compact set, dense set and closure (e.g. Rudin, 1976)that will help us formally stateour main Theorems,below. In what follows, we will make use of Hornik's (1991)Theorem (see Theorem 1, Appendix), whichstates the conditions under which anANN specification can approximate any given function.

In simple words, according to Hornik's (1991) Theorem,ANN's that are based on non-constant, continuous and bounded activation functions are capable of approximating any smooth function as long as the domain of the function is compact.Thus, we begin by formally definingthe set of times series

(Definition 2, Appendix), which constitutes the domain of the function, and then we prove that this set could be considered as being compact (see Theorem 2).

<u>Theorem 2</u>: If $x_{t_i}, i \in I,$ is an arbitrary time series, such that $x_{t_i} \in \mathbb{R}^{\mathrm{N}}$ $\forall i \in I$and $\forall t \in T$ andthe set of time series $\bigcup_{i \in I} x_{t_i} \subset \mathbb{R}^{\mathrm{N}}$, is closed and bounded, then$\bigcup_{i \in I} x_{t_i}$ is a compact subset of $\mathbb{R}^{\mathrm{N}}$.

Proof: See Appendix.

Please note that the implicit assumptions made for the time series set is that it is closed and bounded. The financial time series set could be considered as being closed since it could contain all its boundary points. Additionally, we consider the financial time series set to be bounded since all financial time series couldhave a finite time dimension.

Next, in order to be able to apply Hornik's(1991) Theorem, we also need to formally prove that the proposed specification, for the unknown function G of the general unit root specification, possesses all the mathematical properties that Theorem 1 explicitly states. Below, Theorem 3 formally presents the proposed functional specification and proves the relevant properties.

<u>Theorem 3</u>: If$x_{t_i}, i \in I$ is an arbitrary time seriesand the set of time series $\bigcup_{i \in I} x_{t_i} \subset \mathbb{R}^{\mathrm{N}}$is a compact subset of $\mathbb{R}^{\mathrm{N}}$, whereas $\varphi: \mathbb{R}^{\mathrm{N}} \to \mathbb{R}$ is a non-constant, bounded and continuous function, then any function $k: \mathbb{R}^{\mathrm{N}} \to \mathbb{R}$ of the form $k(x_{t_i-1}) \equiv \rho x_{t_i-1} \cdot F(x_{t_i-1}), \rho \in \mathbb{R}, t \in T$, where: $F\left(x_{t_i-1}\right) \equiv \sum_{n=1}^{N} a_n \varphi(\beta_n \cdot x_{t_i-1})$, with$a_n, \beta_n \in \mathbb{R}$ $\forall n \in \mathbb{N}$, and $a_n \neq 0$, for some $n \in \mathbb{N}$, is also continuous, bounded and non-constant.

Proof: See Appendix.

Having formally shown that the proposed specification is fully compatible with Hornik's(1991) Theorem, below we state our main result (Theorem 4), which states that the specification can formally approximate arbitrarily well the generalnon-linear specification.

Theorem 4

If the set $\bigcup_{i\in I} x_{t_i} \subset \mathbb{R}$, $t \in T$ is a compact subset of $\mathbb{R}$, then the family of functions $\mathcal{F} = \{k(x_{t_i-1}) \in \mathrm{C}\left(\bigcup_{j\in J} G_j\right): k(x_{t_i-1}) \equiv \rho x_{t_i-1} \cdot F(x_{t_i-1})$, $F\left(x_{t_i-1}\right) \equiv \sum_{n=1}^{N} a_n \varphi(\beta_n \cdot x_{t_i-1})$, with a_n , $\beta_n \in \mathbb{R} \forall n \in \mathbb{N}$, $\rho \in \mathbb{R}\}$ isdense in the set of functions $\mathcal{H} = \bigcup_{j\in J} G_j$

Proof: See Appendix.

In simple words, Theorem 4 implies that the proposed specification $k(x_{t_i-1}) \equiv \rho x_{t_i-1} \cdot F(x_{t_i-1})$, $F\left(x_{t_i-1}\right) \equiv \sum_{n=1}^{N} a_n \varphi(\beta_n \cdot x_{t_i-1})$, with a_n , $\beta_n \in \mathbb{R} \forall n \in \mathbb{N}$, $\rho \in \mathbb{R}$isa global approximator to any arbitrary specification $\rho x_{t-1} \cdot G(x_{t-1}; \gamma)$ and, hence, the proposed specification could approximate arbitrarily well the general non-linear unit root specification.

4. The Test

As PSY have emphatically pointed out,the econometric identification of multiple bubbles over time is difficult mainly because of the complex *non-*

linear structure involved in the multiple breaks that produce the bubble phenomena. This is the reason why a general *nonlinear* ANN approximation is used in this work as the main mechanism in the proposed econometric test.

4.1 *Formulation*

We have formally, shown thatthe proposed specification $k(x_{t_i-1}) \equiv \rho x_{t_i-1} \cdot F(x_{t_i-1})$, $F(x_{t_i-1}) \equiv \sum_{n=1}^{N} a_n \varphi(\beta_n \cdot x_{t_i-1})$, with a_n , $\beta_n \in \mathbb{R} \forall n \in \mathbb{N}, \rho \in \mathbb{R}$ is a global approximation to any arbitrarynon-linearunit root specification, i.e.$\rho x_{t-1} \cdot G(x_{t-1};\, \gamma)$. Therefore, $\forall i \in I$, the general unit root test of the form $\Delta x_{t_i} = \rho x_{t_i-1} \cdot G(x_{t_i-1};\, \gamma) + \varepsilon_t$ could be approximated arbitrarily well by the test $\Delta x_{t_i} = k(x_{t_i-1}) + \varepsilon_t$, where $\varepsilon_t$satisfies the usual assumptions.[2] In detail, exploiting the proposed NN specification, the relevant testing equation becomes:

Δx_{t_i}= $\sum_{n=1}^{N} \rho a_n x_{t_i-1} \cdot \varphi(x_{t_i-1};\, \beta_n)$, $\forall i \in I$ **(7)**

Now, without loss of generality,wecansafely make an additional simplifying assumption about the behavior of the employed time series.

Assumption 4: x_{t_i} represents time series of the form:$x_{t_i} = \ln\left(\frac{P_{t_i}}{P_{t_{i-1}}}\right)$.

For instance, x_{t_i}would naturally represent the logarithmic return of asset prices between two time periods in time *t* and *t-1,* e.g. daily. As a result, the quantity $x_{t_i} = \ln\left(\frac{P_{t_i}}{P_{t_{i-1}}}\right)$ hovers around zero, or $x_{t_i} \in B(0, \varepsilon)$.

[2] It should be noted that lag augmentation, in case of serial dependence, does not affect either the test or its mathematical derivation. On the contrary, lags of the dependent variable may indeed be included to eliminate serial correlation.

This is due to the fact that the quantity (before taking natural logarithms) : $\frac{P_{t_i}}{P_{t_{i-1}}} \in B(1, \varepsilon)$ hovers around unity, or $\frac{P_{t_i}}{P_{t_{i-1}}} \in B(1, \varepsilon)$, even for large daily fluctuations in pricesP_{t_i}. However, it should be noted that large daily fluctuations in pricesP_{t_i} are extremely improbable, even in developing markets.Additionally, we have to make an assumption about the activation function φ of the ANN.

Assumption 5: Without loss of generality,we may assume, that the activation function of the ANN has the following form:

$$\varphi(z_t) = e^{{z_t}^{\beta}} - 1 \quad \textbf{(8)}$$

It should be noted that$\varphi(z_t)$is continuous,non-constant and bounded when $z_{t_i} \in B(0, \varepsilon)$, and$\beta$ is a positive real number.

Of course, it should also be pointed out that other alternative activation functions could be used. See Bishop (1995). However, in general, the empirical results are robust, regardless of the activation function used (Haykin, 1999).

In this work, and given the complexity of the problem, the chosen function is able to transform the model to one which lends itself to empirical estimation, contrarily to other possible activation functions. In this sense, the argument by Kuan and White (1994) is in force: ‘‘given the popularity of linear models in econometrics, this form is particularly appealing, as it suggests that ANN models can be viewed as extensions of, rather as alternatives to, the familiar models”.

Now, based on equation (8), equation (7) takes the following form:

$$\Delta x_{t_i} = \sum_{i=1}^{N} \rho a_i x_{t_i-1} \cdot [e^{{x_{t_i-1}}^{\beta_i}} - 1] \quad \textbf{(9)}$$

In what follows, we will make use of Taylor's expansion Theorem, to get an equivalent but more convenient form, of the term:

$e^{x_{t_{i}-1}{}^{\beta_i}} - 1$**(10)**

Thus, by applying the aforementioned Theorem around $x_0 = 0$, we get that:

$e^{z_t{}^{\beta}} \approx 1+z_t{}^{\beta}$**(11)**

Hence, takinginto consideration equation (11), equation (9) becomes:

$\Delta x_{t_i}=\rho a_1 x_{t_i-1}\cdot\left[1+x_{t_i-1}{}^{\beta_1}-1\right]+\rho a_2 x_{t_i-1}\left[1+x_{t_i-1}{}^{\beta_2}-1\right]+\cdots+\rho a_k x_{t_i-1}\left[1+x_{t_i-1}{}^{\beta_k}-1\right]+\cdots+\rho a_N x_{t_i-1}\cdot\left[1+x_{t_i-1}{}^{\beta_N}-1\right]$

$\Delta x_{t_i} = \rho a_1 x_{t_i-1}{}^{\beta_1+1}+\rho a_2 x_{t_i-1}{}^{\beta_2+1}+\ldots+\rho a_N x_{t_i-1}{}^{\beta_N+1}$, $\forall i \in I$**(12)**

Now, without loss of generality,$\forall\, n \in \mathbb{N}$,let: $\rho a_n = \kappa_n$ and $\beta_n + 1=\delta_n$. Thus, we get:

$\Delta x_{t_i} = \kappa_1 \cdot x_{t_i-1}{}^{\delta_1} + \kappa_2 \cdot x_{t_i-1}{}^{\delta_2} + \ldots+\kappa_N \cdot x_{t_i-1}{}^{\delta_N}$,$\forall i \in I$**(13)**

With the inclusion of the error term, we have the following test:

Proposition 1: The null hypothesis, H_0, of a unit root is parameterized bya test of$\sum_{i=1}^{N}\kappa_i = 0,\ \delta_n \in |B(1,\varepsilon)|, \varepsilon > 0$, n = 1,2,…Nin:

$\Delta x_{t_i} = \kappa_1 \cdot x_{t_i-1}{}^{\delta_1} + \kappa_2 \cdot x_{t_i-1}{}^{\delta_2} + \ldots+\kappa_N \cdot x_{t_i-1}{}^{\delta_N} +\varepsilon_{t_i}$,$\forall i \in I$ **(14)**

Proof: See Appendix.

It is worth noting that equation (14)could beseen as a generalization of KSS.

Now, following PSY and the relevant strand in the literature, the previous model specification is complemented with transient dynamics, just as in standard ADF unit root testing. Hence,the proposed specification takes the form:

$\Delta x_{t_i} = \kappa_1 \cdot x_{t_i-1}{}^{\delta_1} + \kappa_2 \cdot x_{t_i-1}{}^{\delta_2} + \ldots+\kappa_N \cdot x_{t_i-1}{}^{\delta_N}+\sum_{p=1}^{P} b_p \Delta x_{t_i-p}$, $\forall i \in I$**(15)**

Of course, in order to allow application of the test with intercept, or intercept *and* trend terms included, these deterministic terms are removed via preliminary regression with the demeaned or detrended version of x_t.

4.2 *Existence of Bubbles*

In what follows,we propose a generalized *max* NNUnit Root (NNUR) test for the presence ofbubbles, as well as a recursive forward and backward technique,based on Bayesian Methods, to detect and time-stamp the bubble origination and termination dates, where flexible window widths are used in their implementation.

Instead of fixing the starting point of the recursion on the first observation, the proposed test extends the sample coverage by changing both the starting point and the ending point of the recursion over a feasible range of flexible windows and is, therefore, suited to analyzing long historical data (PSY).

Now, following the literature on the econometric detection of bubblesas set out earlier, we may make the following assumption:

Assumption 6: $\forall i \in I$ the error term, $\varepsilon_{t_i} \sim N(0, \sigma_{t_i}^2)$, where $\sigma_{t_i}^2$follows a GARCH process of the form:$\sigma_{t_i}^2 = g\left(\sigma_{t_i-1}^2, {\varepsilon_{t_i-1}}^2\right) = a_0 + a_1\sigma_{t_i-1}^2 + a_2{\varepsilon_{t_i-1}}^2$**(16)**

where: $a_0 > 0, a_1 > 0, a_2 > 0$.

In what follows, we performrepeated NNUR tests on sub-samples of the data on a recursive, backward and forward manner, changing the starting and ending points. We proceed by providing a simple algorithmfor the implementation of the test, regarding the detection of bubbles in a time frame. The following simple algorithm sets out the mechanism behind the proposed approach.

Step 1: Let$i \in I$, and x_{t_i} an arbitrary time series of length $T > 0$ and consider a sample of it, the so-called window W with length $0 < W < T$.

Step 2: Partitionthe sample W into all the possible sub-samples $r_{w_j} = [r_{1_j}, r_{2_j}] \subseteq W$ where r_{1_j} is the starting date of the j-th sub-sample and r_{2_j} the respective ending date. In this way, we obtain the set of all subsamples $r_w = \bigcup_{j \in J} r_{w\,j}$ in W.

Step 3: Compute the model's significance $Sig - NN_j$,corresponding to F-like tests, to obtain the set of Sig-s which refers to each window Was $Sig - NN^W = \bigcup_{j \in J \subseteq \mathbb{N}} Sig - NN_j$. Note that these models do not necessarily belong to a single sub-sample.

Step 4: For all the subsamples with the same starting point, choose the $Sig - NN_m, m \in M \subseteq J$ that are (equally or) more significant than their corresponding critical values$Sig - NN_m{}^*$,to obtain the set $Sig = \bigcup_{m \in M \subseteq J} Sig - NN_m$, which corresponds to the set of sub-samples $r_w{}^m = \bigcup_{m \in M \subseteq J} r_{w\,m}$. Note that this choice reduces the cost of keeping the non-significant values in the set.

Step 5: Compute the$max_{m \in M \subseteq J}\{Sig - NN_m\}$ on the set $\bigcup_{m \in M} Sig - NN_m$.

Step 6: **(a)** If there is only a single maximal point$max_{m \in M \subseteq J}\{Sig - NN_m\}$ for all the models with the same starting point, a unique bubble exists in the sub-samplem^*. **(b)**(i) If multiple maximal points exist in different neighborhoods of the same subsample, then multiple bubbles exist. (ii) If multiple maximal points exist in the same neighborhood of the same subsample, then one bubble exists: The one with the longer duration.

Step 7: Repeat steps (1)-(6) for allthe possible sets $Sig - NN_j$, $j \in J$.

Step 8: Repeat steps (1)-(7) for all the models with the same ending point.

Step 9: Repeat steps (1)-(8) above for all possible (rolling) windows W.

Note that the initial size of the window (w_0) is equal to the one suggested in PSY, namely: $w_0 = 0.01 + 1.8/\sqrt{T}$ .Of course, a parameter to account for data frequency could easily be includedin the model.The dating of bubbles is done trivially in the spirit of PSY.

For expository reasons, we provide the following Data Generating Process (DGP),using standard notation. Consider a time series X_t, with length T>0. Let T be partitioned into $j \in J$sub-samples,$r_{w\,j}$. Let $r_{w\,j}{}^*$ be the only sub-samplewhere the bubble occurs. The DGP has the following representation:

$$X_t = X_{t-1}{}^{r_{wj}} 1\left\{r_{w\,j} \neq r_{w\,j}{}^*\right\} + \delta_T X_{t-1}{}^{r_{wj}{}^*} + 1 \sum_{k \neq r_{w\,j}{}^*} \varepsilon_k + \varepsilon^{r_{wj}{}^*}$$

In this scheme,in the pre-bubble period the series follows a pure random walk. The bubble expansion period is $r_{w\,j}{}^*$which involves a mildly explosive process with expansion rate δ_T. The process then collapses and continues its pure random walk behavior $\forall r_{w\,j}, j \in J$.

- Unit root behavior in $t_0$can be identified by:$\left.\frac{dX_t}{dX_{t-1}}\right|_{t=t_0} = 1$**(17)**
- An emerging bubble can be identified by: $\left.\frac{dX_t}{dX_{t-1}}\right|_{t=t_1} > 1$,$\left.\frac{dX_t}{dX_{t-1}}\right|_{t=t_2} \leq 1$**(18)**in the time period $[t_1, t_2]$
- A collapsing bubble can be identified by:$\left.\frac{dX_t}{dX_{t-1}}\right|_{t=\tau} < 1$**(19)**

  in the time period $[t_3, t_4]$.

5. Empirical Analysis and Discussion

Having analyzed the model and the proposed test, we continue by elaborating on the estimation techniqueand data used.

We use data on the stock price-dividend ratio S&P500 (1871.1-2014.6). The S&P 500, i.e. the Standard & Poor's 500, is a stock market index for the US and is based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. More specifically, the S&P 500 index components and their weightings are determined by S&P Dow Jones Indices. It is one of the most commonly followed equity indices, and many consider it as being one of the best representations of the US stock market, and a bellwether for the U.S. economy (Phillips et *al.* 2011a, 2011b).

The proposed approach uses a Bayesian approach because it has numerous advantages related to overcoming the over-fitting problem associated with the traditional approaches, but also due to its increased flexibility. Probably, the main advantage of our approach is the possibility of mixing different pieces of information (sample information, prior information, etc) in order to construct a model that accounts for the stochastic character of the variables.

Analytically, the main reason for using a Bayesian approach is that it facilitates representing and taking fuller account of the uncertainties related to model and parameter values. In contrast, most decision analyses based on maximum likelihood or least squares estimation involve fixing the values of parameters that may, in actuality, have an important bearing on the final outcome of the analysis and for which there is considerable uncertainty. Hence, one of the major benefits of the Bayesian approach is the ability to incorporate prior information, which, along with other numerical methods, makes computations tractable for virtually all parametric models.See,for instance, Carlin and Lewis (2000), Robert (2001) and Wasserman (2004).

We statistically assess, using Bayesian techniques, the following system of equations:

$$\left.\begin{array}{c} \Delta x_{t_i} = {\kappa_1}^{r_{w_j}} \cdot {x_{t_i-1}}^{{\delta_1}^{r_{w_j}}} + {\kappa_2}^{r_{w_j}} \cdot {x_{t_i-1}}^{{\delta_2}^{r_{w_j}}} + \cdots + {\kappa_N}^{r_{w_j}} \cdot {x_{t_i-1}}^{{\delta_N}^{r_{w_j}}} + \sum_{i=1}^{N} {b_i}^{r_{w_j}} \Delta x_{t_i-p} + {\varepsilon_t}^{r_{w_j}} \\ \sigma_{t_i}^2 = {a_0}^{r_{w_j}} + {a_1}^{r_{w_j}} \sigma_{t_i-1}^2 + {a_2}^{r_{w_j}} {\varepsilon_{t_i-1}}^2 \end{array}\right\} \quad \textbf{(19)}$$

The model needs an identification condition for κ_i's, since we are unable to identify them with any alternative procedure. In this context, we begin by imposingthe identification conditions$\kappa_1 < \kappa_2 < \kappa_3 < \ldots < \kappa_N$

We, then, approximate the marginal likelihood of the model using the Laplace approximation (DiCiccio et *al.* 1997).This procedure is fast and easy to apply, which is important in this context where repeated MCMC simulations have to be considered. It also has the advantage that it takes into consideration both the suitability of the model and the overfitting problem.The Laplace approximation to the log marginal likelihood of the model is:

$$L_K = -\frac{T+1}{2} log|\boldsymbol{A}| + \frac{d+L}{2} \log(2\pi) + \frac{1}{2} log\left|\widehat{\boldsymbol{\Delta_K}}\right| \textbf{(20)}$$

where: $\widehat{\boldsymbol{\Delta_K}}$ is an estimate of the covariance matrix of the ML estimator of $\boldsymbol{\Theta_K}$ (inverse Hessian of the log likelihood). This can be approximated by the covariance of the MCMC draws, after convergence and using thinning or an autocorrelation – consistent estimate.

Bayesian inference is performed through a Markov Chain Monte Carlo (MCMC) procedure (Tierney 1994)that resembles the Gibbs sampler using 1,500,000 iterations, the first 500,000 of which are discarded to mitigate start up effects. The long MCMC is needed to guarantee convergence starting from arbitrarily different initial conditions for the parameters. Convergence is assessed from ten different chains in terms of computed posterior probabilities for the different episodes as well as for the specific period during which the episodes occur.

Using the proposed specification for the detection of financial bubbles for each MCMC draw of parameters (Tierney 1994), we compute the derivatives of$k(x_{t_i-1}) \equiv g\left(x_{t_i-1}\right) \cdot F(x_{t_i-1})$that are used for the identification of unit root behavior and thus for the formation and collapse of bubbles.

The number of nodes is selected from all possible combinations using the marginal likelihood in (20), which can be computed relatively easily and efficiently. The model with the highest marginal likelihood is selected. In this context, by approximating the marginal likelihood of the model using the Laplace approximation followingDiCiccio et *al.* (1997), we finally select the number of nodes to be N=3.Next, we compute posterior probabilities that we have a bubble or collapse during certain periods.

It should be noted that the parameter estimates are updated from their previous values using sampling-importance resampling (Smith and Gelfand1992).The size of the resample in SIR was set to 10% of the original MCMC samples.Also, the length of the initial sub-sample r_{w_j}, i.e. $r_{w_0}$is 10, sufficiently small so as to ensure that no bubble will be missed and, meanwhile, that there are enough observations for estimation, in a Bayesian framework.

Of course,we need to ensure the robustness of our results, in the sense that they do not depend critically on the assumptions and calculation on which they were based. As a result, our analysis was applied to numerous logically and empirically plausible priors selected from relevant classes of priors (Berger 1985).In this context, in Table 1, wepresent the baseline priors of $\kappa's$, $\delta's$ and $a's$, as well as a set of alternative priors, which are centered at m and have standard deviations s.

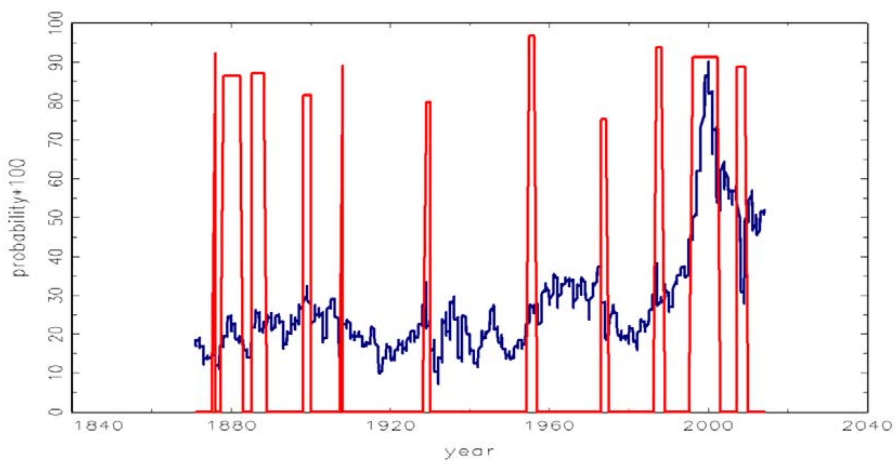
Table 1: Priors

Parameter	Baseline Priors	Alternative priors (m)	Alternative priors (s)
$\kappa_1, \kappa_2, \dots$	$N(0,10)$	$N(0,100)$	$\lvert N(0,100)\rvert$
$\delta_1, \delta_2, \dots$	$\lvert N(1,0.01)\rvert$	$\lvert N(1,0.1)\rvert$	$\lvert N(0,0.1)\rvert$
a_0, a_1, a_2	$\lvert N(0,10)\rvert$	$\lvert N(0,100)\rvert$	$\lvert N(0,100)\rvert$

We produced 10,000 computations under the specified alternative priors and the calculated results – which are available upon request by the authors – were not found to be sensitive to the alternative priors used. This clearly implies that we can safely proceed based on these findings. For a detailed discussion on the theoretical foundations of prior selection see, for instance, Kass and Wasserman (1996).

The results are illustrated in Figure 1, below.

Figure 1. Time series and posterior probabilities of episodes



As can be seen in Figure 1 and in Table 2, the proposed specification is able to identify eleven (11) bubble episodes or bubble formations in the S&P500 index in the sample period (1871.1-2014.6).

Table 2. Bubble periods and Posterior Probabilities

Bubble Period in years.months	**Probability (%)**	**Explanation**
1875.7 - 1876.10	92.32	"America's Almost Civil War", crisis
1877.8 - 1882.6	86.49	Banking panic (Post Long Depression Period)
1885.11- 1888.5	87.12	"Baltimore" Crisis
1898.12- 1900.11	81.55	Cuba War of independence, Crisis
1907.3-1908.1	89.13	Banking panic 1907

1928.8- 1930.10	79.67	Great crash
1954.6 -1956.12	96.81	Postwar boom
1973.1-1974.2	75.21	Oil shock
1986.7 - 1988.9	93.80	Black Monday
1995.6- 2002.6	91.32	dot-com boom
2007.1- 2009.6	88.77	Subprime crisis

In comparison to PSY, we are able to identify four (4) more bubble episodes in the S&P500 index and miss only one. See Table 3, below[3].

Table 3: Comparison for bubble detection

Bubble Period in years.months	**Bubble Explanation**	**Bubble detected in the present paper?**	**Bubble detected in PSY?**
1875.7 - 1876.10	"America's Almost Civil War", crisis	Yes	No
1885.11- 1888.5	"Baltimore" Crisis	Yes	No
1898.12- 1900.11	Cuba War of independence, Crisis	Yes	No
1973.1-1974.2	Oil shock	Yes	No
1917.08-1918.04	The 1917 Stock Market Crash	No	Yes

Another very interesting finding is that the bubbles do not have the same time duration, in comparison to PSY. See Table 4, below.

Table 4: Comparison between bubble durations

[3]We would like to thank an anonymous referee for suggesting the inclusion of Tables 3 and 4, below.

Bubble Period in years.months identified in the present paper	Bubble Period in years.months identified inPSY	Earlier Detection of Bubbles in the present paper compared to PSY?	How many months earlier was the bubble detected in the present paper compared to PSY?
1877.8 - 1882.6	1879.10-1880.4	Yes	14 months
1907.3-1908.1	1907.9-1908.2	Yes	6 months
1928.8- 1930.10	1928.11-1929.10	Yes	3 months
1954.6 -1956.12	1955.1-1956.4	Yes	7months
1986.7 - 1988.9	1986.6-1987.9	No	-1months
1995.6- 2002.6	1995.11-2001.8	Yes	5months
2007.1- 2009.6	2009.2-2009.4	Yes	25months

Hence, our bubble detection mechanism seems to be more sensitive to bubble formation.

As can be seen in Table 4,compared to PSY, the bubble episodes thatweidentify, in general,have longer duration. This means that the proposed specification is able to identify bubble episodes earlier, compared to PSY. Therefore, the proposed specification could be thought of as an EWM.

For instance, if we focus on the recent US subprime crisis, the proposed test indicates that the bubble started in January 2007 and ended in June 2009. According to official data (CIA World Factbook, 2011), the US subprime bubble started in December 2007, i.e. almost 10 months after ourproposed test suggests, i.e.$[2007.1 - 2009.6]$. However, the ending point of the identified bubble, and of the one provided by the official statistics, are exactly the same. This clearly implies that according to the proposed test, this 10-month period coincides with the build-up of the bubble.

Analytically, the proposed specification, based on the aforementioned dating algorithm, is capable of sufficiently answering the fundamental question of every EWM mechanism, which is the timing of detection, while taking into consideration the neglected non-linearities. The appropriate timing of an ideal EWM is crucial for policy makers as the EWMs need to signal the crisis early enough so that policy actions can be implemented in time to be effective. The time frame required to do so depends, *inter alia*, on the lead-lag relationship between changing a specific macroprudential tool and on the impact on the policy objective (CGFS 2012).

For instance, in contrast to monetary policy, where it takes at least a year for interest rates to impact on inflation, this relationship is less well understood for macroprudential instruments. Yet, it is likely to be at least as long. For instance, banks have one year to comply with increased capital requirements under the countercyclical framework of Basel III (Basel Committee, 2010). In addition, data are reported with lags and policy makers do not act immediately on developments but observe trends for some time before changing policies (Bernanke 2004). This urges EWMs to start issuing signals well before a crisis occurs as is the case with the suggested approach.

In fact, early bubble identification could substantially aid policy makers, worldwide. The validity of this argument lies of the fact that whilst tools and actual policies differ across countries and financial institution, the key objective of macroprudential policies, which is the reduction of systemic risk, remains the same(e.g. Borio 2009; Disyatat 2010). In this context, a crucial component of the macroprudential approach based on EWMs is to address the procyclicality of the financial system by, for example, stipulating the accumulation of buffers in "good times" so that these can be drawn down in "bad times". See, among others, White (2008). Tools, which are already used in this regard, include countercyclical capital buffers or dynamic provisioning. See Cukierman (2013). One key challenge for policy makers is the identification of the different states in real time, with particular emphasis on detecting unsustainable booms that may end up in a financial crisis.

6. Conclusion

Despite the fact that the history of financial bubbles is rather long, only limited attention has been paid by the scientific community to the creation of a rigorous econometric test for the early detection of bubble formation. Probably, one of the main reasons behind the inability of most models to efficiently capture the formation of bubbles, is the fact that bubble formation has inherent non-linear characteristic which are difficult to be captured using standard econometric models.

Additionally, another equally important challenge for the econometric detection of bubbles is the datingofbubbles' occurance, in the sense that an econometric test should be able to accurately date the bubble periods detected in the sample. Accurate dating of financial bubbles could have important policy implications, especially for central bankers and policy makers, since it could substantially aid the implementation of policy actions that could potentially ease the consequences of bubbles.

However, only few papers in the literature use recursive estimation techniques for dating multiple bubble episodes. More precisely, a recent strand in the literature, attempts to detect and date bubble episodes based on the unit root behavior of key financial variables. In this paper, we extendedthis strand of the literature by using ANNsin an attempt to approximate the basic unit root specification so as to account for neglected non-linearities.Moreover, we provided a recursive dating procedure for bubble episodes and we applied both our bubble detection test and its dating mechanism to the S&P500 index.

According to our findings, the proposed specification is fully capable of capturing the bubble episodes in the time sample examined. Additionally, the bubble periods identified are longerin comparison to PSY. More precisely, in all common bubble episodes our proposed specification identified the bubble, in the general case, earlier compared to PSY. In other words, our specification could be thought ofas an EWMfor bubble formation, which in turn could have important implications.

In brief, the early identification of bubblesis of outmost importance for policy makers and central bankers, as we have seen. The importance of early identification lies in the timing of implementation of specific countermeasures that could potential prevent: a) the magnitude of a potential collapse through regulatory interventions in the financial markets; b) the downturn effects of bubble collapse in the economy through appropriate inflation targeting; and c) the devastating spillover effects in the global economy through interest rate and/or exchange rate setting.

Of course, there are still numerous issues that could serve as examples for further investigation. For example, from a theoretical point of view, one could explore the limit theory characteristics of the proposed approach or, from an empirical point of view, onecould make anattempt to explore alternative NN architectures. Clearly, future research in capturing and modeling non-linearities in bubbles would be of great interest.

REFERENCES

Abreu D. andBrunnermeier K. M. (2003). ”Bubbles and crashes”, *Econometrica*, 71(1): 173–204.

Barlevy G. (2007). "Economic theory and asset bubbles," Economic Perspectives, Federal Reserve Bank of Chicago, issue Q III, pages 44-59.

Baxa J. and Horvath R. and Vasicek B., (2013), “Time varying monetary rules and financial stress: Does financial instability matter for monetary policy?”,*Journal of Financial Stability*, 9(1):117-138.

Bernanke, B S (2004): "Gradualism", Remarks at an economics luncheon cosponsored by the Federal Reserve Bank of San Francisco (Seattle Branch) and the University of Washington, Seattle, 20 May 2004.

Bishop, C.M. (1995), Neural Networks for Pattern Recognition, Clarendon Press, Oxford.

Borio, C and M Drehmann (2009): "Assessing the risk of banking crises - revisited", BIS Quarterly Review, March, 29-46.

Blanchard, O. and M. Watson.(1982). “Bubbles, Rational Expectations, and Financial Markets,” in Paul Wachter (ed.) Crises in the Economicand Financial Structure. Lexington, MA: Lexington Books: 295- 315.

Brunnermeier K.M. 2009.“Deciphering the Liquidity and Credit Crunch 2007–2008”, *Journal of Economic Perspectives*, 23(1):77-100.

Campbell Y.J.and Lo W. A. and MacKinlay A. C. 1997.”The Econometrics of Financial Markets”, Princeton University Press, Princeton, New Jersey.

Carlin, B. P. and Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, Second Edition, London: Chapman & Hall.

Caruana, J (2010): "The challenge of taking macroprudential decisions: Who will press which button(s)?". Speech at the 13th Annual International Banking Conference, Federal Reserve Bank of Chicago, in cooperation with the International Monetary Fund, Chicago, 24 September 2010.

Chan, N. H. and Genovese, C. R. 2001. “A Comparison of Linear and Nonlinear Statistical Techniques in Performance Attribution”, *IEEE Transaction on Neural Networks*, 12, 922–928.

Committee on the Global Financial System (CGFS) (2012): "Operationalising the selection and application of macroprudentialinstruments", CGFS Publications No. 48.

Cucierman A., (2013), “Monetary policy and Institutions before, during and after the global financial crisis”, *Journal of Financial Stability*, 9 (3): 373-384.

DeMarzo M. P. and Kaniel R. and Kremer I., 2008. "Relative Wealth Concerns and Financial Bubbles," *Review of Financial Studies*, 21(1):19-50.

Dezbakhsh, H. and A.Demirguc-Kunt.(1990). “On the Presence of Speculative Bubbles in Stock Prices,” *The Journal of Financial andQuantitative Analysis*,25:101-112.

Diba, B. and H. Grossman. (1988a). “The Theory of Rational Bubbles in Stock Prices,” *The Economic Journal*, 98(September): 746- 54.

Diba, B. and H.l Grossman.(1988b).”Explosive Rational Bubbles in Stock Prices?”*American Economic Review*,78(June): 520-30.

DiCiccio, T. J., Kass, R. E., Raftery, A. and Wasserman, L. (1997).“Computing Bayes factors by combining simulation and asymptotic approximation”, *Journal of American Statistical.Association*, 92, 903-915.

Disyatat P., (2010), “Inflation targeting, asset prices, and financial Imbalances: Contextualizing the Debate”, *Journal of Financial Stability,* 6(3):145-155.

Dufwenberg, Martin, Tobias Lindqvist, and Evan Moore. 2005. "Bubbles and Experience: An Experiment." *American Economic Review*, 95(5): 1731-1737.

Evans,G.(1991).“PitfallsinTestingforExplosiveBubblesinAssetPrices,”A*mericanEconomicReview*31(September): 922-30.

Fama, E. (1965).“The Behavior of Stock Market Prices", *Journal of Business*, 38: 34–105.

Flavin, M. (1983). “Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence,” *Journal of Political Economy*, 91(December): 929-956.

Flood P. R and Hodrick J. R. (1990). ”On testing for speculative bubbles”, *Journal of Economic Perspectives*, 4(2):85-101.

Flood, R. and R.Hodrick. (1986). “Asset Price Volatility, Bubbles and Process Switching,” *Journal of Finance*,41(September): 831-42.

Garber M.P. (2000). ”Famous First Bubbles; The Fundamentals of Early Manias”, The MIT Press Cambridge, Massachusetts London, England.

Granger, C and M Machina (2006): "Forecasting and Decision Theory," In Elliott, G, C Granger and A Timmermann (Eds.), Handbook of Economic Forecasting, Elsevier.

Greenwood, R. and Nagel, S. (2009). "Inexperienced investors and bubbles," *Journal of Financial Economics*, 93(2):239-258.

Hall, S. and M. Sola. (1993). "Testing for Collapsing Bubbles: An Endogenous Switching ADF Test," Discussion paper 15-93, London Business School.

Hall, S., Z.Psaradakis, and M. Sola. (1999). "Detecting Periodically Collapsing Bubbles: A Markov-Switching Unit Root Test," *Journal of Applied Econometrics,* 14: 143-154.

Haruvy E. and Noussair N. C. (2006), "The Effect of Short Selling on Bubbles and Crashes in Experimental Spot Asset Markets', Journal of Finance 61(3): 1119-1157.

Haykin, S. (1999), Neural Networks, Prentice-Hall, New Jersey.

Homm U. and Breitung J., (2012). "Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods," *Journal of Financial Econometrics*, 10(1): 198-231.

Hong H. and Scheinkman J. and Xiong W, (2006). "Asset Float and Speculative Bubbles," *Journal of Finance*, 61(3): 1073-1117.

Hornik, K. (1991). "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, 4(2):251-257.

Hornik, K., Stinchcombe, M., White, H. (1989)."Multilayer feedforward networks are universal approximators", *Neural Networks* 2, 359–366.

Hornik, K., Stinchcombe, M., White, H. (1990)."Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks", *Neural Networks* 3, 551–560.

Kapetanios G., Shin Y. and Snell A.(2003). "Testing for a unit root in the nonlinear STAR framework", *Journal of Econometrics*, 112 (2): 359-379.

Keynes, J. M, (1936). "The General Theory of Employment, Interest and Money", London: Macmillan.

Kindleberger, C. P., and Aliber, R. Z. (2005), "Manias, Panics and Crashes; A History of Financial Crises", Hoboken, New Jersey: John Wiley and Sons, Inc.

Kleidon, A. (1986). "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*,94(October): 953-1001.

Kuan, C.M., White, H. (1994). “Artificial neural networks: An econometric perspective”, *Econometric Reviews*, 13, 1–91.

Lawrence, M, P Goodwin, M O'Connor and D Onkal (2006): "Judgmental forecasting: A review of progress over the last 25years", *International Journal of Forecasting,* 22, 493-518.

LeRoy, S, and R.Porter(1981). “The Present-Value Relation: Tests based on Implied Variance Bounds,” *Econometrica,*49(May): 555- 574.

Luukkonen, R and Saikkonen, P. and Teriisvirta, T. 1988.“Testing linearity against smooth transition autoregressive models”, *Biometrika* 75, 491-99.

Mankiw, N. G, D.Romer, and M. Shapiro. (1985). “An Unbiased Reexamination of Stock Market Volatility,”*Journal of Finance*, 40(July): 677-687.

Önkal, D, M E Thomson and A A C Pollock (2002): "Judgmental forecasting", inClements, M P and Hendry, D F (eds), A companion to economic forecasting, Blackwell Publishers, Malden and Oxford.

Pesaran, M H and S Skouras (2002): "Decision-based methods for forecast evaluation". In: Clements, M P and D F Hendry (Eds.), A Companion to Economic Forecasting, Blackwell, Malden and Oxford.

Phillips P.C.B. and Shi S. and Yu J., (2014). "Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behaviour," *Oxford Bulletin of Economics and Statistics*, 76(3):15-333.

Phillips P.C.B. and Wu Y. and Yu J., (2011b). "Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?",*International Economic Review*, 52(1): 201-226.

Phillips P.C.B.and Yu J, (2011c). "Dating the timeline of financial bubbles during the subprime crisis," *Quantitative Economics*, 2(3):455-491.

Phillips, P. C. B., Shi, S. and Yu, J. (2011a).*Testing for Multiple Bubbles*, Singapore Management University, Working Papers No. 09-2011.

Phillips, P. C. B., Shi, S. and Yu, J. (2013).“Testing for Multiple Bubbles 1- Historical episodes of Exuberance and Collapse in the S&P 500m Singapore Management University, Working Papers No4-2013.

Phillips P. C. B., Shi, Shuping and Yu, Jun (2105a), TESTING FOR MULTIPLE BUBBLES: LIMIT THEORY OF REAL-TIME DETECTORS, International Economic Review, Vol. 56,No 4, 1079- 1134

Phillips P. C. B., Shi, Shuping and Yu, Jun (2105b), TESTING FOR MULTIPLE BUBBLES: HISTORICAL EPISODES OF EXUBERANCE AND COLLAPSE IN THE S&P 500, International Economic Review, Vol. 56, No 4, 1043-1078.

Robert, C. P. (2001), *The Bayesian Choice*, Second Edition, New York: Springer-Verlag.

Rudin, W. (1976). "*Principles of Mathematical Analysis*", McGraw-Hill, New York, International Edition.

Sanger, T.D. (1989), Optimal Unsupervised learning in a Single Layer Linear Feedforward Neural Network, *Neural Networks*, 2: 495-479.

Scheinkman J. and Xiong W.(2002). "Overconfidence, Short-Sale Constraints and Bubbles," Princeton Economic Theory Working Papers 98734966f1c1a57373801367f, David K. Levine.

Shiller, R. (1981). "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*71(June): 421-436.

Smith, A. F. M. and Gelfand, A. E. (1992), Bayesian statistics without tears: A sampling–resampling perspective,*American Statistician* 46, 84–88

Temin P. and Voth H-J., 2004. "Riding the South Sea Bubble," *American Economic Review*, 94(5):1654-1668.

Tierney L. 1994. "Markov Chains for exploring posterior distributions", *Annals of Statistics*, 22(4):1701-1728.

Tirole, J. (1982). "On the Possibility of Speculation under Rational Expectations," *Econometrica*,50: 1163-1182.

Van Norden, S. and R.Vigfusson. (1998). "Avoiding the Pitfalls: Can Regime-Switching Tests Reliably Detect Bubbles?" *Studies in NonlinearDynamics and Econometrics*,3:1-22.

Van Norden, S. (1996). "Regime Switching as a Test for Exchange Rate Bubbles," *Journal of Applied Econometrics*,11(July): 219-51.

Wasserman, L. (2004), *All of Statistics: A Concise Course in Statistical Inference*, New York: Springer-Verlag.

West, K. (1987). "A Specification Test for Speculative Bubbles,",*The Quarterly Journal of Economics*102(August):553-580.

White W. R., (2008), “Past financial crises, the current financial turmoil and the need for a new macrofinancial stability theory”, *Journal of Financial Stability*, 4(4):307-312.

Wu, G and Z Xiao. (2002). “Are there Speculative Bubbles in Stock Markets? Evidence from an Alternative Approach,” working paper, Michigan Business School.

Wu, Y., (1997), "Rational Bubbles in the Stock Market: Accounting for the U.S. Stock-Price Volatility," *Economic Inquiry*, 35(2):309-319.

Zhang, G.-P., and Berardi, V. (2001).“Time series forecasting with neural network ensembles: an application for exchange rate prediction”, *Journal of the Operational Research*, 52:652–664.

Appendix

<u>Theorem 1</u>: Consider $\mathrm{X} \subseteq \mathbb{R}^{\mathrm{N}}$ a compact subset of $\mathbb{R}^{\mathrm{N}}$ and C(X) the space of all real valued functions defined on X. Let $\varphi: \mathrm{X} \rightarrow \mathbb{R}$ be a non-constant, bounded and continuous function. Then, the family:

$\mathcal{F} = \{\mathrm{F(x)} \equiv \sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{a_i} \varphi\left(\mathrm{w_i^T x} + \mathrm{b_i}\right), \mathrm{a_i}, \mathrm{b_i} \in \mathbb{R}, \mathrm{w_i} \in \mathbb{R}^{\mathrm{N}}\}$ is dense on $\mathrm{C(X)}$.

Proof: See Hornik (1991).

<u>Definition 2:</u>If $x_{t_i}, i \in I$ is an arbitrary time series such that $x_{t_i} \in \mathbb{R}^{\mathrm{N}} \forall i \in I$,and$\forall i \in I, \forall t \in T$,we define$\bigcup_{i \in I} x_{t_i} \subset \mathbb{R}^{\mathrm{N}}$ to be the time series set.

<u>Proof of Theorem 2</u>

The proof is trivial and is based on the fact that any closed and bounded subset of $\mathbb{R}^{\mathrm{N}}$ is compact (e.g. Rudin 1976).

<u>ProofofTheorem 3</u>

Without loss of generality, let $g: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ be a function of the form $g\left(x_{t_i-1}\right) = \rho x_{t_i-1}$. Then,the function $k: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is defined asthe product of functions $g: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$and $F\left(x_{t_i-1}\right): \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$, i.e. $k(x_{t_i-1}) \equiv g\left(x_{t_i-1}\right) \cdot F(x_{t_i-1})$.

(i) Let $i \in I$ and $t \in T$. $F\left(x_{t_i-1}\right): \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$is non-constant by definition when a_n≠0, for some $n \in \mathbb{N}$. In order to prove that $k: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is also non-constant, it suffices to prove that $g: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is non constant. But, by definition, $\rho \in \mathbb{R}$ and $x_{t_i-1} \neq 0$ for some $t \in T$,and, hence $g: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is non constant.

(ii) Let $i \in I$ and $t \in T$. Since $F\left(x_{t_i-1}\right): \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is bounded, in order to prove that $k: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is bounded, it suffices to prove that $g: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is bounded i.e. $\left|g\left(x_{t_i-1}\right)\right| < M$, $M \in \mathbb{R}$. By construction, $g: \mathbb{R}^{\mathrm{N}} \rightarrow \mathbb{R}$ is bounded since $\rho \in \mathbb{R} \forall t \in$

T. Hence, there exists $M \in \mathbb{R}$ such that $|g(x_{t_i-1})| < M, \forall i \in I$. Hence, $g: \mathbb{R}^N \to \mathbb{R}$ is bounded.

(iii) Let $i \in I$ and $t \in T$. The function $k: \mathbb{R}^N \to \mathbb{R}$ is continuous as the product of the continuous functions $F(x_{t_i-1}): \mathbb{R}^N \to \mathbb{R}$ and $g(x_{t_i-1}): \mathbb{R}^N \to \mathbb{R}$.

Proof of Theorem 4

From Theorem 2, the set of time series is compact. From Theorem 3, any function of the form $k(x_{t_i-1}) \equiv \rho x_{t_i-1} \cdot F(x_{t_i-1}), \rho \in \mathbb{R}$ is continuous, bounded and non-constant. Hence, from Theorem 1, the family: $\mathcal{F} = \{k(x_{t_i-1}) \in \mathrm{C}(\bigcup_{j\in J} G_j): k(x_{t_i-1}) \equiv \rho x_{t_i-1} \cdot F(x_{t_i-1}), F(x_{t_i-1}) \equiv \sum_{n=1}^{N} a_n \varphi(\beta_n \cdot x_{t_i-1})$, with a_n, $\beta_n \in \mathbb{R} \forall n \in \mathbb{N}, \rho \in \mathbb{R}\}$ is dense in $\mathrm{C}(\bigcup_{j\in J} G_j)$.

Proof of Proposition 1

Let x_{t_i}, $i \in I$ be an arbitrary time series of length T>0. Then the proposed specification implied by equation (12) for x_{t_i} is:

$$x_{t_i} = \rho a_1 x_{t_i-1}{}^{\beta_1+1} + \rho a_2 x_{t_i-1}{}^{\beta_2+1} + \cdots + x_{t_i-1} + \varepsilon_t$$

By application of the lag operator L, we get:

$$x_{t_i} = \rho a_1 L x_{t_i}{}^{\beta_1+1} + \rho a_2 L x_{t_i}{}^{\beta_2+1} + \cdots + L x_{t_i} + \varepsilon_{t_i}$$

Using the linearity of the lag operator, we get:

$$x_{t_i} = \rho a_1 L x_{t_i} x_{t_i}{}^{\beta_1} + \rho a_2 L x_{t_i} x_{t_i}{}^{\beta_2} + \cdots + L x_{t_i} + \varepsilon_{t_i}$$

$$x_{t_i}\left(1 - L\left(\rho \alpha_1 x_{t_i}{}^{\beta_1} + \rho \alpha_2 x_{t_i}{}^{\beta_2} + \cdots + 1\right)\right) = \varepsilon_{t_i}$$

Therefore, x_{t_i} is a stationary process of the form $x_{t_i} = \frac{\varepsilon_{t_i}}{1 - L(\rho\alpha_1 + \rho\alpha_2 + \cdots + 1)}$ when $1 - L(\rho\alpha_1 + \rho\alpha_2 + \cdots + 1) \neq 0$, $\beta_n \in |B(0, \varepsilon)|$, $\varepsilon > 0$. This, in turn, implies that: $\rho \sum_{n=1}^{N} \alpha_n \neq 0$. Thus: $\sum_{n=1}^{N} \kappa_n \neq 0$, since: $\rho a_n = \kappa_n \forall\, n \in \mathbb{N}$. This completes the proof.