BSIC - Special Report

Markets Team November 2016

When the Bull Bursts the Bubble, the Bear Arrives

Forecasting Bear Markets

The main objective of our research was to find a model that could forecast or, at least, acknowledge the presence of a bear market. In order to estimate such a model, the first necessary step is to date past bear and bull markets. This can be done either through a parametric approach (Markov Switching Models) or through a non-parametric one. We chose to focus on the latter.

The non-parametric approach largely revolves around the algorithm developed by Bry and Boschan (1971). It was originally developed for and applied to the detection of business cycles, in particular for quantitatively replicating the contractions and expansions determined by the National Bureau of Economic Research (NBER). This computer program recognizes the patterns in the time series, detaches these patterns according to a sequence of rules, and locates the turning points (peaks and troughs) in the series. Following the business cycle literature, we assume that the duration of a complete cycle from the trough to the next trough (or alternatively peak to peak) must be at least 15 months. In addition, the time spent in a bear market (time from the peak to the next trough) or bull market (trough to peak) must be at least six months. Once identified the turning points we can build a binary time series where the value one signifies a bear market state.

The model we are going to estimate is called Probit Model and is a particular type of regression where the dependent variable can only take binary values. Given Y_t, X_t as the binary dependent variable and the regressors' matrix respectively, we assume the model takes the form $P_{t-1}(Y_t = 1|X_t) = \Phi(X'_{t-1}\beta)$ where P denotes probability, and Φ is the Cumulative Distribution Function (CDF) of the standard normal distribution. The parameters β are typically estimated by maximum likelihood. Therefore, after having estimated the parameters, we can make forecasts on the state variable Y for the next periods.

After a review of the current literature we decided to select as regressors the most significant ones: the 1-period lagged value of Y, the previous period log-return of the stock market and a macroeconomic indicator, which is the Term-Spread (i.e. the difference between the 10yr treasury and the 3months T-bill).

We analyzed 50 years of monthly S&P500 returns. As a first step, we divided the dataset in two subsets and then we estimated the model on the first half of the series and dedicated the second half of the series to out-of-sample forecasts. Forecasts are constructed using an expansive window of observations where the data from the start of the dataset through to the present forecast time are used in estimation to obtain a new forecast. This procedure is repeated until the end of the sample.

In the chart (*fig.1*) you can see, in the blue line, the forecasted probability of having a bear market the next month plotted against the S&P500 index.

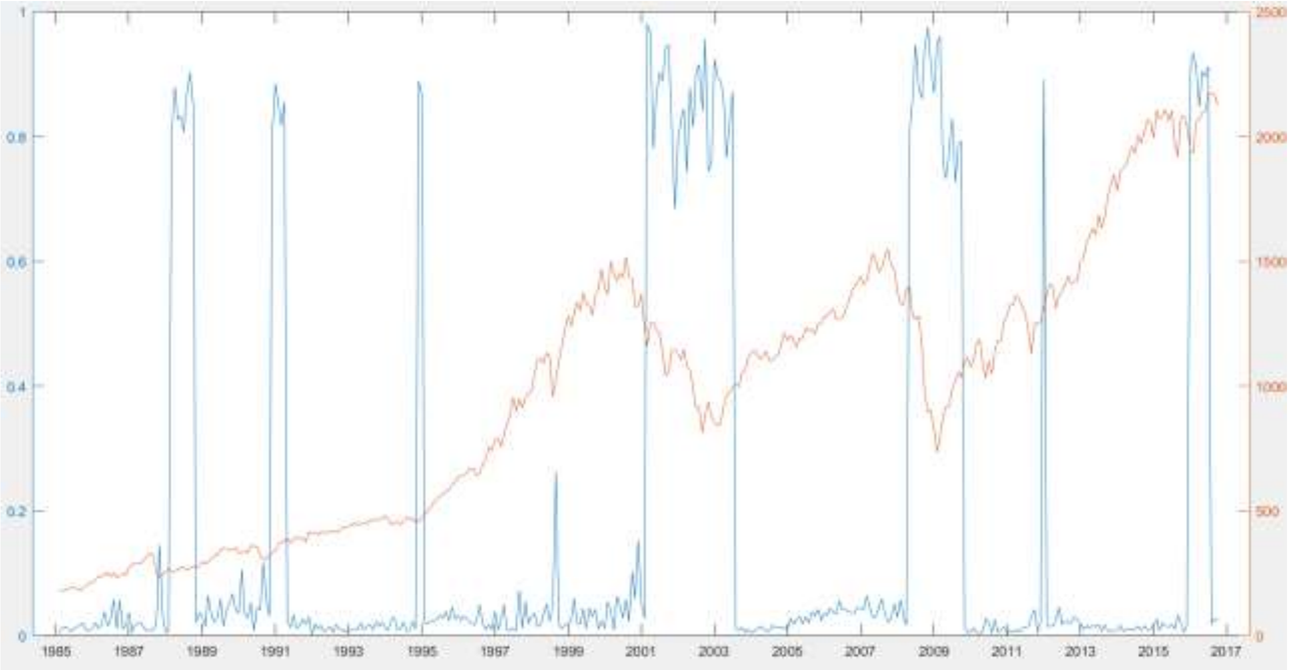

Fig.1 Probability forecast of a bear market (blue line) against S&P 500 (red line) (Source: yahoo.finance)

As you can see, besides some strange solitary extreme values, the forecast successfully predicted the 2 major bear markets of the last decades: the dot-com bubble burst and the 2008 financial crisis. Unfortunately, the model forecasts lag a few months behind the start of every bear market and this is due to the nature of the Bry Boschan algorithm, which requires data points before and after the peak (or troughs) in order to identify it and therefore it takes a few months to identify the most recent turning point. Nonetheless, it still turns out to be quite useful in predicting severe market crashes as stated before.

We developed a market timing strategy which is fully invested in the S&P500 each month whose next-month probability forecast of a bear market is below a certain threshold and fully invested in cash otherwise. In order to avoid unnecessary transactions due to single data point spikes in our probability forecasts we added another constraint to the strategy: it can unwind the stock position only if the probability of a bear market has been over the threshold for 3 consecutive months.

In the chart below (*fig.2*), you can see the growth of a 100$ capital invested in a buy and hold strategy (red line) vs our market timing strategy (blue line).

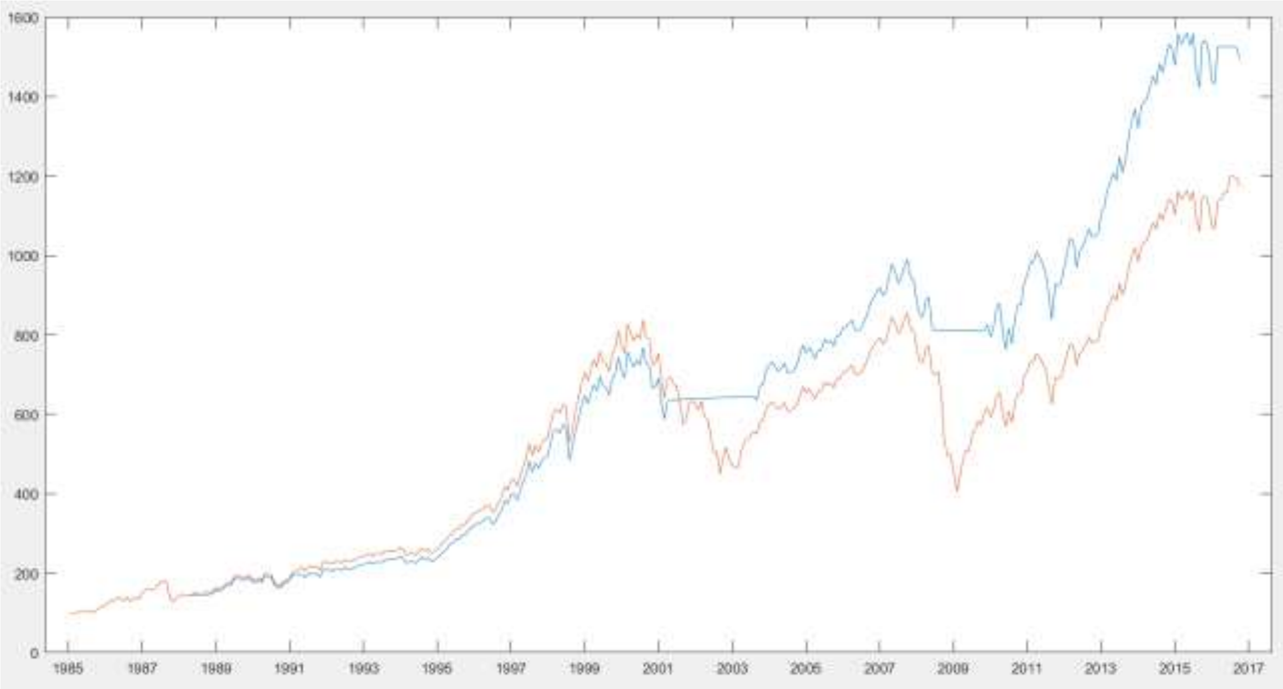

Fig.2 Growth of 100$ invested in buy and hold strategy (red line) vs market timing strategy (blue line) (Source: yahoo.finance; fred.stlouisfed.org)

As you can see, the main feature of our market timing strategy is the ability to avoid consistent losses during the most severe market crashes. And this is the advantage that on the long run makes it overperform greatly the buy and hold strategy. The Sharpe Ratio of this strategy is 0.646.

Although being a very basic strategy we believe it offers good results also considering that it has very low costs due to a very limited number of transactions (it is simply a buy and hold strategy that divests in anticipation of severe market crashes). Furthermore, there is ample room for improvement by adding some more complex feature to the trading strategy. For instance, a bear-market-probability dependent leverage.

Bubble Indicators

A speculative bubble is a situation in which the price of an asset increases above the intrinsic value of the asset. Such growth is triggered by expectation of price appreciation and by an imitating process between agents in the market. A bubble can end abruptly with a sharp drop in prices, so it is important to be able to detect them in order to profit from the ascending prices and to exit the market in time before the crash.

Here, we want to implement an algorithm which allows us to determine if a certain asset is experiencing a bubble.

First, we assume that during the growing of the bubble the asset price follows the Log-Periodic-Power-Law model, or LPPL. The logarithm $y = \ln(P_t)$ of price P_t at time t follows:

$$\ln(P_t) = y(t) = A + B(t_c - t)^z + C(t_c - t)^z \cos[\omega \ln(t_c - t) + \varphi] \qquad (1)$$

where t_c is the critical time of the bubble, the most probable time of the crash; A is the expected log-price at t_c; the exponent z defines the shape of the accelerating increase of y and lies between 0 and 1; ω is the angular frequency of the log periodic oscillations; φ is the phase of the oscillations; B and C measure respectively the incidence of the power law and log periodic components on the evolution of the price.

Fig.3 shows a simulation of the evolution of the price under this model.

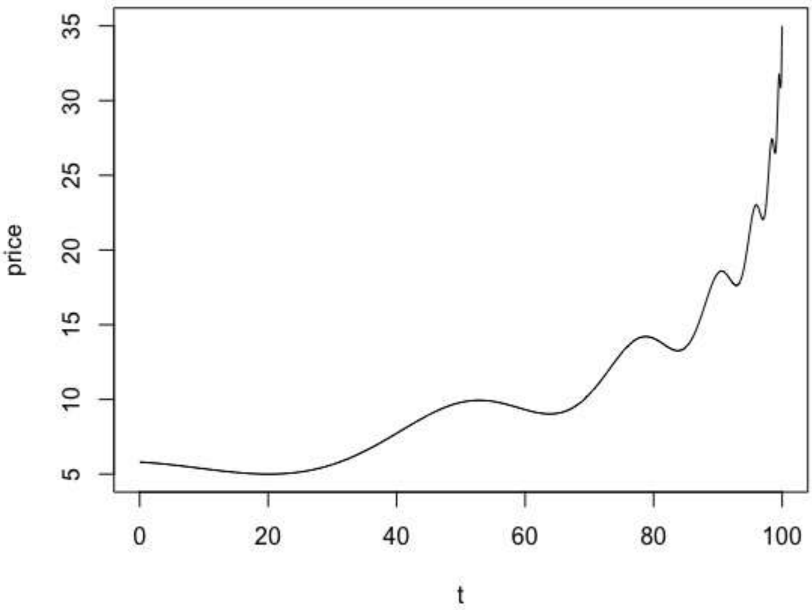

Fig.3 A simulation of the LPPL model; Source: BSIC

The model describes two behaviours:

$$A + B(t_c - t)^z \qquad (2)$$

a power-law growth in the logarithm of the price which ends at t_c, where the first derivative of y with respect to t (the expected return) tends to infinity. It is originated by the global herding and imitation of decision of the economic agents in the market during the bubble expansion.

$$C(t_c - t)^z \cos[\omega \ln(t_c - t) + \varphi] \qquad (3)$$

A periodic component with increasing frequency because it is expressed in terms of the logarithm of $t_c - t$, the time remaining before the burst of the bubble, rather than in terms of $t_c - t$ itself. It is generated by the hierarchical structure in the social network of investors (there is not the same degree of influence among investors) and by a nonlinear mean reversal behavior of fundamental investing styles.

Our analysis will then consist of two steps: first, we look for a more than linear increase in y, then we look for the log-periodic component.

Looking for Non-Linearity: Power Law

In the normal market conditions, the price is expected to grow exponentially, so the logarithm of the price can be modelled as

$$y = a + bt + \sigma\epsilon_t \qquad (4)$$

where b is the expected rate of return and ϵ_t is a stochastic noise component with mean=0 and variance=1, σ is the volatility of the price, a constant depending on y at time t=0.

As a bubble is a period of faster-than-exponential growth in the asset price, we can extend eq. (4) adding a quadratic term, so that

$$y = a + bx + cx^2 + \sigma\epsilon_t \qquad (5)$$

We want to measure at which extent we can assume that the price is increasing faster than exponentially or, equivalently, at which extent we can assume that $c \neq 0$.

After having fit both (4) and (5) using least-squares interpolation, we have to compare the two models. We cannot compare directly the goodness of fit of the two models because (4) has 3 degrees of freedom (a b and σ), while (5) has 4 degrees of freedom (a, b, σ and c). The higher number of degrees of freedom of (5) means that it would always fit more closely to the empirical data points than (4) would, but over-fit costs loss of generality. Eventually, we will use the Akaike Information Criterion.

The Akaike Information Criterion is a measure of relative quality of statistical models for a given set of data founded on information theory.

Practically speaking, a model with k parameters (k degrees of freedom) is fitted to a collection of data points so that the so-called likelihood function L is maximized. Given N data points (x_i, y_i) and a model $\hat{y}(x_i)$, the least-squares interpolation is equivalent to maximizing the function

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(y_i - \hat{y}(x_i))^2}{2\sigma^2}\right] \qquad (6)$$

Let L_{MAX} the maximum value of the likelihood function obtained when fitting the model, then the AIC value is defined as

$$AIC = 2k - 2\ln(L_{MAX}) \qquad (7)$$

For small data samples of numerosity N, the AIC must be substituted by

$$AICc = AIC + \frac{2K(K+1)}{N-K-1} \qquad (8)$$

The rule is to take the model with the lowest AICc. If we define SSR as the sum of squared residuals, for (4) we have

$$SSR_{linear} = \sum_{i=1}^{N} (y_i - a - bt_i)^2 \qquad (9)$$

and as well for (5)

$$SSR_{bubble} = \sum_{i=1}^{N} (y_i - a - bt_i - ct_i{}^2)^2 \qquad (10)$$

then, according to the AIC, we derive from (6) that the non-linear model (5) is to be preferred if

$$\mathrm{SSR_{bubble}} + 2\sigma^2 < \mathrm{SSR_{linear}} \qquad (11)$$

where the term $2\sigma^2$ accounts for the additional degree of freedom. Defining the improvement of model (5) relative to model (4) as

$$\mathrm{D} = \frac{\mathrm{SSR_{linear}} - \mathrm{SSR_{bubble}}}{\mathrm{SSR_{linear}}} \qquad (12)$$

(11) can be rewritten as

$$\mathrm{D} > \frac{2\sigma^2}{\mathrm{SSR_{linear}}} \qquad (13)$$

Because σ is the volatility of the stochastic component and is unknown, we proceed to calculate D and compare it with past values for the same asset or with the value derived from different assets of the same class to determine a threshold $\mathrm{D_{thr}}$, such that if $\mathrm{D} > \mathrm{D_{thr}}$, we can reject the null hypothesis that c=0. We will call this condition I.1, and is necessary but not sufficient for the price to be in a bubble regime.

However, we can use the sample variance of the residuals as an estimator of $\sigma^2$and implement further the Akaike Information Criterion.

Since we assume that ϵ_t is a IID Gaussian noise and we use the least-squares interpolation, we can put (6) into (7) and obtain (remembering definitions (9) and (10))

$$\mathrm{AICc_{linear}} = 6 + \mathrm{N}\ln(2\pi) + 2\mathrm{N}\ln(\sigma) + \frac{\mathrm{SRR_{linear}}}{\sigma^2} + \frac{24}{\mathrm{N}-4} \qquad (14)$$

for model (4), and

$$\mathrm{AICc_{bubble}} = 8 + \mathrm{N}\ln(2\pi) + 2\mathrm{N}\ln(\sigma) + \frac{\mathrm{SRR_{bubble}}}{\sigma^2} + \frac{40}{\mathrm{N}-5} \qquad (15)$$

for model (5).

We then define $\Delta\mathrm{AICc_{linear}}$ and $\Delta\mathrm{AICc_{bubble}}$ as

$$\Delta\mathrm{AICc_{linear}} = \mathrm{AICc_{min}} - \mathrm{AICc_{linear}} \quad ; \quad \Delta\mathrm{AICc_{bubble}} = \mathrm{AICc_{min}} - \mathrm{AICc_{bubble}} \qquad (16)$$

where $\mathrm{AICc_{min}}$ is the minimum between $\mathrm{AICc_{linear}}$ and $\mathrm{AICc_{bubble}}$. These two values are used to compute the following two probabilities, $\mathrm{P_{linear}}$and $\mathrm{P_{bubble}}$:

$$\mathrm{P_{linear}} = \exp(\Delta\mathrm{AIC_{linear}}/2)/[\exp(\Delta\mathrm{AIC_{linear}}/2) + \exp(\Delta\mathrm{AIC_{bubble}}/2)] \qquad (17)$$

$$\mathrm{P_{bubble}} = \exp(\Delta\mathrm{AIC_{bubble}}/2)/[\exp(\Delta\mathrm{AIC_{linear}}/2) + \exp(\Delta\mathrm{AIC_{bubble}}/2)] \qquad (18)$$

The Akaike Information Criterion is derived from information theory because $\mathrm{P_{bubble}}$ is the probability that the non-linear model (5) minimizes the information we expect to lose if we use this model rather than model (4). On the other hand, $\mathrm{P_{linear}}$ express the same probability referred to model (4). So, the higher $\mathrm{P_{bubble}}$, the higher the probability that the logarithm of the price is experiencing non-linear growth typical of a bubble. We will define again a threshold $\mathrm{P_{thr}}$, so that if $\mathrm{P_{bubble}} > \mathrm{P_{thr}}$, we consider it to be a sign of a bubble. We will call this condition I.2, necessary but not sufficient for the price to be in a bubble regime.

Another sign of a bubble is the sign of c in model (5). According to (2), the second order derivative of y respect to t must be positive. So, we consider

$$c > 0 \quad (19)$$

as a sign of a bubble, as well. We will call this condition I.3

Log Periodicity

Once we have assured that the power law term in the LPPL model holds, we must look for the fluctuating behaviour. In order to do this, we need a time series of prices which has been de-trended from the power law.

As (2) is an approximation of 1, we start by fitting

$$y(t) = A + B(t_c - t)^z \qquad (20)$$

to the set of prices collected from the market. The fitting is non-trivial because (20) is not linear and the Sum of the Squared Residuals function has many local minima, thus algorithms such as Simulated Annealing, Taboo Search or Genetic Algorithm should be preferred. We then define the de-trended logarithm of the price as

$$s_i = \frac{y_i - A}{(t_c - t)^z} \qquad (21)$$

and we normalize it

$$s'_i = \frac{s_i - \bar{s}}{\sigma_s} \quad \text{where} \quad \bar{s} = \frac{1}{N}\sum s_i \quad \sigma_s = \frac{1}{N-1}\sum (s_i - \bar{s})^2 \qquad (22)$$

We express s'_i in terms of τ_i, where

$$\tau_i = \ln(t_c - t) \qquad (23)$$

so, if LPPL holds, the pairs (s'_i, τ_i) form a sinusoidal pattern of constant angular frequency ω.

The easiest way to detect a periodic oscillation is by a periodogram. We can express a periodic function as a linear combination of sines and cosines of different frequencies.

$$f(t) = \sum_{k=1}^{+\infty}[a_k \cos(2\pi kt) + b_k \sin(2\pi kt)] \quad (24)$$

In practice, the periodogram $I(\omega)$ of a function shows the importance of frequency ω when the function is expressed as sines and cosines: the higher the value of $I(\omega)$, the greater is the incidence of $\sin(\omega x)$ and $\cos(\omega x)$.

We will use the Lomb-Scargle Periodogram, which is obtained using the least squares interpolation. For our set of N pairs (s'_i, τ_i), for a given ω the periodogram is calculated as

$$I(\omega) = \frac{\left[\sum_N s'_i \cos(\omega\tau_i + \tilde{\varphi})\right]^2}{\sum_N \cos^2(\omega\tau_i + \tilde{\varphi})} + \frac{\left[\sum_N s'_i \sin(\omega\tau_i + \tilde{\varphi})\right]^2}{\sum_N \sin^2(\omega\tau_i + \tilde{\varphi})} \qquad (25)$$

where $\tilde{\varphi}$ is defined as

$$\tan(2\tilde{\varphi}) = -\frac{\sum_N \sin(2\omega\tau_i)}{\sum_N \text{cov}(2\omega\tau_i)} \qquad (26)$$

Even if τ_i are not equally spaced, we would measure $I(\omega)$ at the natural frequencies anyways. We can use as natural frequencies ω_k

$$\omega_k = \frac{2\pi k}{T} \qquad k = 1, 2, \dots, \lfloor N/2 \rfloor \qquad (27)$$

where $\lfloor \cdot \rfloor$ is the floor function, N is the number of pairs (s'_i, τ_i) and T is the total time span of τ_i, $T = \max(\tau_i) - \min(\tau_i)$. $n = \lfloor N/2 \rfloor$ is the number of independent frequencies.

From the periodogram $I_k = I(\omega_k)$, we focus on the biggest one

$$I_{MAX} = \max\{I_k\} \;\; \text{for} \;\; k = 1, 2, \dots, n \qquad (28)$$

because we assume its angular frequency $\omega_{MAX} \Rightarrow I(\omega_{MAX}) = I_{MAX}$ to be the ω in (1). However, we are interested in determining if it is a consequence of random noise or it bears information. Without getting too technical, we want to perform a p-value test in order to reject the null hypothesis that the peak I_{MAX} is originated by noise alone. We use the Schwarzenberg-Czerny false alarm probability function defined as

$$\Pr[I_{MAX} > z] = 1 - \left[1 - \left(1 - \frac{z/\sigma^2}{n}\right)^n\right]^n = p(I_{MAX}) \qquad (29)$$

$p(I_{MAX})$ is our p-value. We define a threshold α: if $p(I_{MAX}) > \alpha$, then we can reject the null hypothesis, thus we can assume that the price is experiencing a Log-Periodic evolution. We call this condition II, and is a sign of a bubble.

Implementation

Here, we explain a possible implementation of this model to analyse a time series of daily prices.

Conditions I.1, I.2 and I.3 are based on two interpolations, so we have to choose a time interval of N subsequent days and interpolate model (4) and (5).

From the interpolation of (5), we can directly get the sign of c, which is our condition I.3. Then, we use the residuals of the two interpolations to calculate SSR_{linear} and SSR_{bubble}, from which we get the value of D and the value of P_{bubble}. If the original data is a series longer than N, we can reiterate the calculation for the N past days of each day of the original data. We then get three series, for the sign of c, D and P_{bubble}, which we want to summarize in order to get a synthetic index, showing if there is a bubble or not. We define three binary variables, one for each condition, I.1, I.2 and I.3. I.3 will be 1 if c>0. With regards to, I.1 and I.2, we have to define some threshold levels. There are not predetermined levels, so we have to decide how to state them. For D, we can simply define a percentage k% and consider the (1-k%)-th percentile as D_{thr} (so our binary series will be 1 if $D > D_{thr}$ and 0 otherwise). Bearing in mind that P_{bubble} express a probability measure, we can soundly choose a P_{thr}: the binary series I.2 will be 1 if $P_{bubble} > P_{thr}$, 0 otherwise.

Then, we proceed to interpolate (20) to our series of N prices and we use the parameters obtained in this way to calculate s'_i. Than, we use (25) (26) and (27) to create the periodogram. As before, we have to determine a threshold α for the p-value.

Here are the results of the analysis of the S&P 500 index from 2013 up to now. We set k%=5%, P_{thr}=95% and α=5%. However, the binary series are heavily dependant on N, the time interval we consider for interpolation.

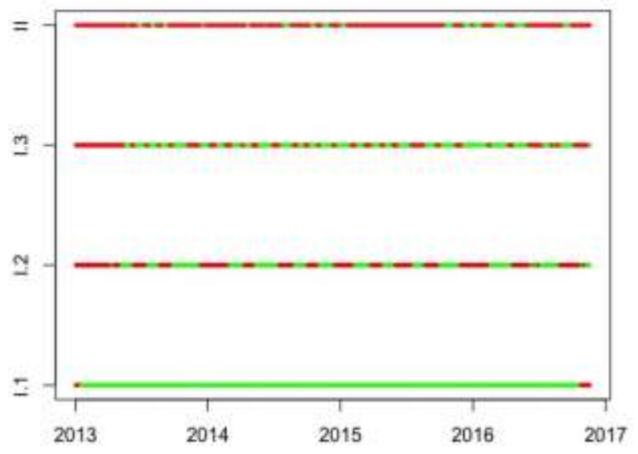

Fig.4 Green dots correspond to 1, while red dots correspond to 0. N=60

Source: BSIC

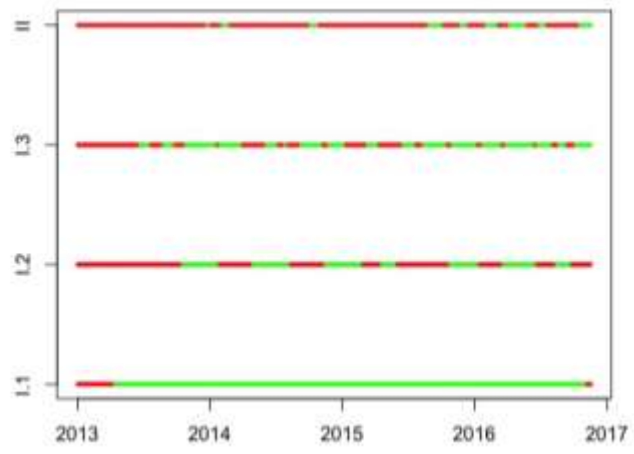

Fig.5 Green dots correspond to 1, while red dots correspond to 0. N=120

Source: BSIC

However, this algorithm is effective in analyzing different assets to check which of them is showing a sign of a bubble. We chose to focus on the first 250 components of the S&P 500 by market capitalization and analyze their behaviour for the last two years. Out of 250, we consider the 125 which showed an increase during the time considered. We can divide them:

- 38 have just one positive bubble indicator between I.1, I.2 and I.3;
- 68 have just two positive bubble indicator between I.1, I.2 and I.3;
- 19 have I.1, I.2, I.3 equal to 1.

We can consider the last group of stocks as being the most probable candidates of a bubble behavior. If we consider condition II (the one involving the log periodic behavior), all of them shows oscillation not originated by random noise with α=5%.