Department of Economics and Finance
Degree Program in Economics and Finance major in Finance

STATISTICALLY DETECTABLE RATIONAL BUBBLES

Course of Econometric Theory

Supervisor:
Prof. Paolo Santucci de Magistris

Co - Supervisor:
Prof. Federico Carlo Eugenio Carlini

Candidate:
Viola Simonetti
ID: 759121

Academic Year 2023/2024

A papà, a nonno

e a Claudio.

Contents

Introduction

The research questions that this thesis tries to answer focus on financial bubbles, do they really exist? Are we able to detect them statistically? And most importantly, are we able to predict the subsequent crash?

Detecting bubbles in the stock market is important for many financial decision makers, such as central banks, financial institutions and regulators. The detection of these bubbles is particularly important and it can be achieved by exploiting time series methods for real time monitoring of structural breaks. Moreover, understanding the timing of a bubble allows for the reconciliation of the bubble's formation and conclusion with other macroeconomic events. This retrospective analysis provides insights into the evolution of the explosive behaviour throughout the economy, helping policymakers to comprehend the broader context in which financial bubbles develop. A bubble regime, characterized by rapid and abnormal growth in asset prices, holds significant implications for financial stability. Financial bubbles have the potential to create instability in the market, and understanding their emergence is essential for finance regulators, in particular to mitigate the risks associated with the sharp price decline that usually follows the price run-up (Harvey, Leybourne, and Sollis, 2017). The impact of financial bubbles on a country's macroeconomic performance can have detrimental consequences. As demonstrated by the experiences during the *dot-com bubble*, indecisiveness in the face of a bubble's development can have profound effects on monetary policy. Former Federal Reserve Chairman Alan Greenspan highlighted the challenges faced by the Fed committee during the dot-com bubble (Greenspan, 2007), where uncertainty about whether to increase or decrease interest rates prevailed due to a lack of clarity on the market dynamics. In conclusion, the detection of bubbles in the stock market is needed in order to be able to maintain financial stability, to inform policy decisions, and to understand their broader impact on the economy. It allows for improved decision-making in real time, it mitigates the risks associated with explosive behavior in asset prices, and it contributes to a more resilient and stable financial system.

In this thesis I utilize an econometric test to detect the presence of rational bubbles, exploiting their time series characteristics. I then use a date stamping algorithm to have a more precise estimate of their initiation, peak and conclusion. *Rational bubbles* are a particular kind of financial bubbles studied by Shiller (1981), which arise in a rational world, without the need of negating the efficient markets hypothesis. Following the rational bubble model, asset prices are determined by two components: the fundamental component and the bubble component. When the bubble component is different than zero, the price series with respect to its fundamental (the price-dividend ratio), has an explosive behavior. The econometric test from Phillips, Shi, and Yu (2015) takes advantage of this time series behavior, it is constructed as a recursive right-tailed unit root test, when the null hypothesis is rejected, the presence of a bubble is implied. The date stamping algorithm from Harvey, Leybourne, and Sollis (2017) considers a regime-switching data generating process where the start, the peak and the end of each bubble are estimated as parameters of the model by least squares.

A first empirical application uses the monthly observations of the S&P500 Index, where I use both the Phillips, Shi, and Yu (2015) test and the Harvey, Leybourne, and Sollis (2017) date stamping algorithm. In this application I am interested in the detection and timing of rational bubbles, I use a market index because of its broad coverage of different assets. A second empirical application considers 49 industry portfolios as classified by Fama and French. In this other application I am more interested in comparing the ability of the econometric test in the detection of bubbles with respect to a naive technique. The econometric test is the one from Phillips, Shi, and Yu (2015) that I use in the first empirical application. The naive technique is the one from Greenwood, Shleifer, and You (2019), who consider a bubble a mere price increase followed by a price decrease. To make this comparison I employ different logistic regressions, with and without some control variables.

This thesis is organized as follows. In Chapter 1 there is a broad and general overview of the discussion in the literature of the possibility of existence of bubbles in the market.

Chapter 2 describes the theory of the rational bubble model, whereas Chapter 3 describes the theory of the econometric test and of the date stamping algorithm. Chapter 4 shows the empirical results. The last chapter concludes.

Chapter 1

Can Bubbles Arise in Efficient Markets?

1.1 Efficient Capital Markets

The efficient capital market hypothesis simply states that prices reflect all available information (Fama, 1976). More formally,

$$f_m(p_{1t}, ..., p_{nt}|\phi^m_{t-1}) = f(p_{1t}, ..., p_{nt}|\phi_{t-1}), \tag{1.1}$$

where ϕ_{t-1} is the set of available information at time $t-1$ and $f(p_{1t}...p_{nt}|\phi_{t-1})$ is the joint probability density function for security prices at time t implied by the information available at time $t-1$. The subscript m stands for the market, so that ϕ^m_{t-1} is the set of information that the market uses to determine security prices and $f_m(p_{1t}, ..., p_{nt}|\phi^m_{t-1})$ is the joint probability density function assessed by the market. Equation 1.1 states that the information available in $t-1$ that the market uses to determine security prices at t includes all available information and that the market uses this information correctly.

An important implication of this model is that expected returns are constant, implying that the information available at time $t-1$ cannot be used to predict variation in expected returns. In particular, the efficient market hypothesis states that in a regression equation of the return of a security in t on the variables contained in the set ϕ_{t-1}, all the coefficients and the intercept should be zero. The set of information can contain anything, including past returns, so that it is not possible to use past returns to correctly assess an expected return between $t-1$ and t that deviates from the constant one. Assuming that the regression of $R_{j,t}$ on $R_{j,t-\tau}$ is linear:

$$\mathbb{E}(R_{j,t}|R_{j,t-\tau}) = \delta_\tau + \gamma_\tau R_{j,t-\tau}.$$

Let the lag $\tau = 1$, γ_τ is the autocorrelation coefficient. Market efficiency implies that the coefficients in this regression are zero for all values of the lag τ, including the autocorrelation. Thereby, returns follow a white noise process and prices follow a random walk.

Fama (2014) affirms that, assuming efficient markets, the predictability in the variation of expected returns of stocks and bonds results from the variation in risk or risk aversion. In particular, Fama considers the existence of a bubble detectable if the price run up leg and the subsequent crash are predictable, something which is not possible for two prominent reasons: the decline in prices that follows a price run-up is unpredictable and large swings in stock prices are rather given by large swings in the underlying economic conditions. For these reasons Fama considers the term *bubble* inconsistent in a world where markets are efficient. He reasons that maybe recessions are caused by major stock market oscillations and then market upturns bring them to an end. Moreover, if a bubble is defined as an irrational price increase that implies a strong decline, he asks which leg of the bubble is considered to be irrational, the upturn or the downturn. In an interview with Richard Thaler[1], Fama affirms that he wants evidence that people can statistically identify bubbles by predicting the crash that follows. A goal that according to him has not yet been achieved.

Greenwood, Shleifer, and You (2019) challenge Fama's propositions achieving three important results: a price increase does not predict unusually low returns afterwards, however it heightens the probability of a crash; the attributes of the increase in prices are good predictors of an eventual crash and of future returns; these characteristics are able to help investors to earn superior returns by timing the bubble. Greenwood, Shleifer, and You (2019) achieve these results with an empirical application on the prices of 48 industry portfolios (considering

[1] available on Chicago Booth: `https://www.chicagobooth.edu/review/are-markets-efficient`

the Fama-French classification). The authors consider a bubble episode as such if the industry experienced a price increase of at least 100% over two years and than a 40% drawdown within a two years period. With a sample size that goes from 1926 to 2014 they find 21 such episodes. They test whether a price run-up can heighten the probability of a crash afterwards with a simple regression and they show their results with a kernel density plot. They obtain that the stronger the price run-up is, the higher the probability of a crash afterwards. They conclude that Fama is correct by saying that a mere price increase does not predict low returns, however sharp price increases are able to predict a higher probability of a crash afterwards.

Moreover, Fama makes his claims considering only *irrational* bubbles and he does not take into account those bubbles that can arise even in an efficient market, such as *rational* bubbles. In this matter he only criticizes Shiller, in the above mentioned interview, by saying that his rational bubble model is based on the proposition that there is no variation through time of expected returns, where in the real world there is variation, both in expected returns and in risk aversion.

1.2 Rational Behavior and Bubbles

Shiller (2003) does not criticize the efficient market hypothesis and he affirms that it should be taken as approximately true. An argument in favor of the efficient market theory is that it is difficult in the stock market to earn superior returns just by buying low and selling high. Shiller (2003) considers the presence of the so called *smart money* who are those investors that look for opportunities in the market driving prices to their true values. The presence of these investors is implicit in the efficient market hypothesis, however it does not imply that the market cannot go through periods of significant mispricing. When this happens smart money are not able to exploit the under or overvaluation opportunities rapidly, so that the mispricing does not end immediately, increasing the uncertainty on when it will effectively end.

Miller (1977) shows with a simple supply-demand model that in a scenario in which investors have non-homogeneous expectations, the price of a security can be higher than what the market considers as fair value. He shows that when the divergence of opinions about the pricing of a particular asset between investors increases, the demand curve for that asset tends to become steeper. Thus, the price of the security increases, reflecting now only the expectations of the most optimistic investors. Only in the limit case, where there is no disagreement, the market price is determined by the average evaluation of all the potential investors. He also shows that, in the presence of short selling constraints, the market can be strongly overpriced. Consider the presence of a particularly fanatic kind of investors, which is not ruled out by the efficient market hypothesis. Consider the case in which these fanatics begin to aggressively buy a security so that they become the only holders. The security becomes overpriced. Smart money would like to short sell this security, but if they cannot find any to borrow, they can only buy it. In this situation the market is overpriced, the smart money knows it, but they cannot do anything about it.

Shiller (2003) presents some examples of what he calls *obvious mispricing*. The first example is about the eToys, a company that used to sell toys via the Internet. eToys sales and profits were much lower of the retailer Toys ”R” Us but despite this, its stock value was much higher. Apparently, the public had an exaggerated view on its potential, indeed, eToys filed for bankruptcy in 2001. Even if the eToys price seemed absurd, smart money did not correct it. To do so they should have tried to sell the stocks short, however, not everyone is always willing to, because there is always the possibility that the stock value will increase more. Another interesting example is about the 3Com sale of Palm. In March 2000, 3Com announced that its subsidiary Palm would be sold to the general public, first a fraction of it and the rest later. Palm stock value increased so much that surpassed the market value of the owner, which would not be possible in an efficient market. However, smart money could not exploit this opportunity because the interest cost of borrowing Palm shares was too high. Shiller and Miller show that even in an efficient market where smart money are present, the behavior of investors can lead to substantial mispricing.

Shiller (1981) develops the so called *rational* bubble model (see Chapter 2), without contradicting the efficient market theory. He shows that swings in stock prices are too big relative to actual subsequent events and too big relative to dividends and earnings swings.

Prices are considered to be equal to the present value of rationally expected real dividends discounted by a constant real discount rate. If the transversality condition (determined in Equation 2.3) fails, so that the present value of payments occurring infinitely far in the future is different than zero, a bubble component arises and prices appear to be higher than what implied by fundamentals. Giglio, Maggiori, and Stroebel (2016) challenges Shiller's proposition and find no evidence for the transversality condition to fail. The authors propose a model free direct test, verifying whether payments at infinite maturity do have a present value of zero. They do so by exploiting a characteristic of the UK and Singapore housing markets where there is a distinction between leaseholds and freeholds: leaseholds are finite-maturity ownership contracts whereas freeholds are infinite-maturity ownership contracts. They estimate the price difference between freeholds and very long maturity leaseholds, that should represent the bubble component, and test whether it is indeed positive. They find no significant difference between freeholds and leaseholds prices, thus cocluding that there is no evidence for the existence of rational bubbles.

1.3 Literature Review on Rational Bubbles

Theoretical studies on rational bubbles can be found in Shiller (1981), Blanchard and Watson (1982), Diba and Grossman (1988), Froot and Obstfeld (1991), and Evans (1991). Where all these studies agree on the theoretical definition of a bubble, they differ on the statstical tests used to find evidence of them.

One of the first tests employed is a volatility test, used by Shiller (1981) and Blanchard and Watson (1982). This test arises from the observation that prices seem to be too volatile with respect to fundamentals. The authors estimate upper bounds for the unconditional and conditional (on the set of past information) variances of the price process. Since these are likely to be surpassed in the presence of bubbles, they see whether they are violated in the data. Both papers find evidence of episodes in which the volatility of prices surpasses the theoretical bound. West (1987) performs a different kind of test, he compares different sets of estimates of the parameters needed to calculate the present value of a stock's dividend series. Under the null hypothesis of absence of bubbles, the estimates should be equal. The author employs this test on US data and he finds enough statistical evidence to reject the null hypothesis, implying the presence of bubbles.

Froot and Obstfeld (1991) describe a particular model of rational bubbles, the intrinsic rational bubble model, where the bubble component is only driven by exogeneous fundamentals that determines asset prices, such as dividends. An important characteristic of this model is that the bubble remains constant over time for a given level of fundamentals, so that highly persistent fundamentals lead to highly persistent over or undervaluations. Moreover, this kind of bubble may cause asset prices to overreact to news about fundamentals.

Diba and Grossman (1988) employ stationary and cointegration tests to find evidence against the existence of rational bubbles. They reason that if the first differences process of dividends is stationary in the mean and rational bubbles do not exist, the first differences process of prices is also stationary. Moreover, if the first differences process of dividends is stationary in the mean and rational bubbles do not exist, stock prices and dividends are cointegrated of order (1, 1), which in practical terms means that prices and dividends move together[2]. With these tests they do not find evidence of bubbles. However, Evans (1991), describing a model of periodically collapsing bubbles, shows how in this scenario simple Dickey Fuller and Bhargava tests are not able to detect the presence of bubbles.

Others used Dickey Fuller tests to detect the presence of rational bubbles in the market, such as Hall, Psaradakis, and Sola (1999), Phillips, Wu, and Yu (2011), and Phillips, Shi, and Yu (2015). Hall, Psaradakis, and Sola (1999) use a generalization of the Augmented Dickey Fuller test using the class of dynamic Marlov-switching models, allowing regression parameters to switch values between different regimes. Phillips, Wu, and Yu (2011) and Phillips, Shi, and Yu (2015) use instead a recursive Augmented Dickey Fuller test to the price and dividend series (or to the price-dividend ratio series), firstly using an expanding-window procedure and then a rolling-expanding window procedure. All these papers find evidence for the presence of bubbles in the stock market. Harvey, Leybourne, Sollis, and Taylor (2016), Harvey, Leybourne, and Sollis (2017), and Harvey, Leybourne, and Whitehouse (2020) use the recursive unit root test to detect the presence of a bubble, but then they improve the

[2] In more statistical terms, this means that there exists at least one linear combination of prices and dividends that is stationary.

dating using a date-stamping algorithm based on the sum of sqaured residuals.

Other tests used are fractionally integration tests, used by Cuñado, Gil-Alana, and Grazia (2005), Koustas and Serletis (2005), and Sibbersten and Kruse (2009). These tests study the order of integration of the price and dividend series and the difference between the two variables. They differ from the previous tests because they allow for an order of integration d, with d not necessarily constrained to be 0 or 1.

Another branch of the literature focused on change point detection analysis for online detection (Chu, Stinchcombe, and White, 1996; Homm and Breitung, 2012; Górecki, Horváth, and Kokoszka, 2018; Horváth, Liu, et al., 2020; Astill et al., 2021; Horváth, Li, and Liu, 2022). These types of structural break tests fall under the scope of sequential monitoring. They are based on the estimation of a boundary function and some kind of detector (the most used one is a CUSUM based detector), as soon as the detector surpasses the boundary a structural break is detected. If a structural break is detected, then the null hypothesis of no structural break (no bubble) is rejected. This is considered an online monitoring procedure because the test is employed every time a new observation is available.

1.4 Bubbles and Behavioral Economics

In the literature there can be found different theories on how financial bubbles can develop in the market and how their existence can be detected. Besides the rational bubble model, there are those models coming from behavioral economics, in which bubbles arise from the irrationality of the agents in the market. In particular, there are the positive-feedback model and the disagreement model.

A description of the positive-feedback model can be found in Long et al. (1990). In this model there are two types of investors, rational speculators, who behave like the smart money, and positive feedback investors, who buy securities when prices are high and sell securities when prices decline. The authors show how in this scenario, when rational speculators are present, asset prices tend to go even higher than if they were not present, exacerbating the bubble. This kind of model differs from the rational bubble model in that expected return on stocks turn negative as the peak of the bubble.

Disagreement models can be found in Barberis et al. (2018), Sheinkman and Xiong (2003), and Abreu and Brunnermeier (2003). In this model two types of investors are present: rational investors and boundedly rational investors. In this scenario a bubble arises from the dispersion of opinions between the investors and the lack of synchronization between them. In particular, a source of disagreement can be a cognitive bias such as overreaction or overconfidence[3]. Disagreement models are based on the study of the behavior of agents in the market, and on how the presence of biases leads to the creation of bubbles even in a world where rational investors exist. They differ from the rational bubble model because they link the creation of a bubble not only with an increase in asset prices, but also with an increase in volatility and trade volume.

Another noteworthy model is the one from Tirole (1985). The author describes an overlapping-generations model, with an infinite number of finite lived agents. He shows how a bubble can only arise in a dynamically inefficient scenario, where agents have over-accumulated private capital, so that the interest rate becomes lower than the growth rate of the economy.

[3] For a more comprehensive study on how cognitive biases can affect the market see Bondt and Thaler (1985) and Thaler (2016).

Chapter 2

The Asset Pricing Model

2.1 Rational Bubble model

Considering the theoretical asset pricing model by Shiller (1981), the current stock price is related to the present value of next period's stock price and dividends:

$$P_t = \gamma \mathbb{E}_t[P_{t+1} + D_{t+1}], \tag{2.1}$$

where $\mathbb{E}_t[\cdot]$ represents the expected value conditional on the information available at time t. $0 < \gamma = 1/(1 + R) < 1$ is the real discount factor, assumed to be constant, and R is the constant real interest rate. Shiller (1981) proves how R in this model corresponds to the one-period holding return:

$$1 + R_t = \frac{P_{t+1} + D_{t+1}}{P_t}. \tag{2.2}$$

Define the transversality condition as:

$$\lim_{j \to \infty} \gamma^j \mathbb{E}_t[P_{t+j}] = 0. \tag{2.3}$$

This condition entails that the present value of payments occurring infinitely far in the future is equal to zero. If this condition holds, the forward looking looking solution to Equation 2.1 is:

$$P_t^f = \sum_{j=1}^{\infty} \gamma^j \mathbb{E}_t[D_{t+j}],$$

which is the standard no arbitrage condition, called market fundamental (Diba and Grossman, 1988) or unique equilibrium price (Froot and Obstfeld, 1991).

If the transversality condition does not hold, the general solution to Equation 2.1 is:

$$P_t = P_t^f + B_t. \tag{2.4}$$

B_t represents the rational bubble component and it is the solution of the homogeneous expectational difference equation:

$$\mathbb{E}_t[B_{t+1}] = \gamma^{-1} B_t. \tag{2.5}$$

Equation 2.5 represents the submartingale property (since $\gamma < 1$, it follows that $\mathbb{E}_t(B_{t+1}) > B_t$) and its solutions satisfy the stochastic difference equation:

$$B_{t+1} = \gamma^{-1} B_t + u_{t+1},$$

where u_{t+1} is a random variable such that $\mathbb{E}_t[u_{t+j}] = 0$ for every j (its expected future value is always zero). The variable u_{t+1} is an innovation that comprises the information available at time $t + 1$. This information can be intrinsically irrelevant and unrelated to the market fundamentals or it can be related to relevant variables through parameters that are not present in P_t^f (Diba and Grossman, 1988).

Consider D_t to follow an ARIMA process, so that its first difference follows a stationary ARMA process. If there are no bubbles and the transversality condition holds, it can be shown that also the first difference of P_t follows a stationary ARMA process (Evans, 1991). In this case, P_t and D_t are cointegrated of order $(1, 1)$, meaning that there exists a linear combination of P_t and D_t that is stationary. In particular, it can be shown that $P_t - R^{-1} D_t$

is stationary[1]. If instead, ΔD_t follows a stationary ARMA process but bubbles are present, then:

$$\lim_{j\to\infty} \mathbb{E}_t[P^f_{t+j}] = \lambda j + c_t,$$

for some constant c_t, where $\lambda = \mathbb{E}(\Delta P^f_t)$. Consider now Equation 2.5, rearrange and use the law of iterated expectation to obtain:

$$\mathbb{E}_t[B_{t+j}] = \gamma^{-j} B_t.$$

So, the conditional expectation of the market fundamental component grows linearly in the forecast horizon j. On the other hand, the conditional expectation of the bubble component contains the root $\gamma^{-1} > 1$. Thus, the conditional expectation of ΔP_{t+j} is explosive if a bubble is present. More simply put, if the bubble is present and it grows at an explosive rate, the observed price will behave like an explosive process even if the relevant fundamentals are stationary I(0) processes or, at most, integrated I(1) processes. Because of this, if explosive behaviour in the price-dividend ratio is found, the existence of a bubble can be inferred. (Phillips, Shi, and Yu, 2015).

2.2 Periodically Collapsing Bubbles

Bubbles can take the form of different stochastic processes, including those that consider different stages, of expansion and collapse. Even if they are not going to be considered in the rest of this thesis I will describe two exemplary models from Blanchard and Watson (1982) and Evans (1991). The model from Blanchard and Watson (1982) can be formalized as:

$$B_{t+1} = \begin{cases} (\pi\gamma)^{-1} B_t + \varepsilon_{t+1} & \text{with probability } \pi \\ \varepsilon_{t+1} & \text{with probability } 1-\pi \end{cases},$$

[1]for a proof see Campbell and Shiller (1987).

where ε_{t+1} is the error term with $\mathbb{E}_t(\varepsilon_{t+1}) = 0$. In each period, the probability of a bubble arising is π and the probability of a crash is $1 - \pi$. During the expansion phase the bubble grows at a rate of $(\pi\gamma)^{-1} > 1$ and its average duration is $(1 - \pi)^{-1}$. The conditional expectation of B_{t+1} can be obtained as:

$$\mathbb{E}_t(B_{t+1}) = \mathbb{E}_t \left\{ \pi[(\pi\gamma)^{-1} B_t + (1 - \pi)\varepsilon_{t+1}] \right\} = \gamma^{-1} B_t,$$

so it satisfies the submartingale property 2.5.

The model from Evans (1991) takes the form:

$$B_{t+1} = \begin{cases} \gamma^{-1} B_t z_{t+1} & \text{if } B_t \leq \alpha \\ [\delta + \pi^{-1}\gamma^{-1}\theta_{t+1}(B_t - \gamma\delta)] z_{t+1} & \text{if } B_t > \alpha \end{cases}. \tag{2.6}$$

δ and α are positive parameters such that $0 < \delta < \gamma^{-1}\alpha$. The error term z_{t+1} is multiplicative instead of additive and it is an exogeneous i.i.d. (independent and identically distributed) random variable satisfying $\mathbb{E}_t[z_{t+1}] = 1$. In particular z_{t+1} is lognormally distributed, scaled to have a unit mean, so that $z_{t+1} \sim \exp(\nu_{t+1} - \tau^2/2)$ with ν_{t+1} i.i.d. random variable distributed as a normal $\nu_{t+1} \sim \mathcal{N}(0, \tau^2)$. This ensures that the bubble process satisfies the submartingale property 2.5. θ_{t+1} is an exogeneous i.i.d. Bernoulli process that takes value equal to 1 with probability π and equal to 0 with probability $1 - \pi$. This process describes an always positive bubble, since $B_t > 0$ implies $B_s > 0$ for every $s > t$. As long as $B_t \leq \alpha$ the bubble grows at a mean rate of γ^{-1}, but when $B_t > \alpha$ it grows at an explosive rate of $(\gamma\pi)^{-1} > 1$ and it has a probability of $1 - \pi$ of collapsing. After the collapse, the bubble component falls to a mean value of δ and the process begins again.

2.3 Log-Linear Approximation

Consider $p_t = \log(P_t)$, $d_t = \log(D_t)$ and $r_t = \log(1 + R_t)$, the Equation 2.2 can be written as:

$$\log(P_{t+1} + D_{t+1}) - \log(P_t) = p_{t+1} - p_t + \log(1 + \exp(d_{t+1} - p_t) = r_{t+1}.$$

Applying a Taylor series expansion of $\log(1+\exp(d_{t+1}-p_{t+1}))$ at the sample mean of $p_t - d_t$, the return r_{t+1} can be expressed as:

$$r_{t+1} = \kappa + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t, \tag{2.7}$$

where,

$$\rho = \frac{1}{1+\exp(\overline{d-p})} < 1,$$
$$\kappa = -\rho\log\rho - (1-\rho)\log(1-\rho).$$

$\overline{d-p} = \frac{1}{n}\sum_{t=1}^{n}(d_t - p_t)$ is the average mean of the log dividend-price ratio based on a sample of size n.

By recursive substitution of 2.7 and taking conditional expectations, with respect to the set of available information at time t, the log price-dividend ratio can be written as the sum of a fundamental component f_t and a bubble component b_t:

$$p_t - d_t = f_t + b_t, \tag{2.8}$$

as in Equation 2.4 for price levels. The fundamental component can be written as:

$$f_t = \frac{\kappa}{1-\rho} + \sum_{k=0}^{\infty}\rho^k \mathbb{E}_t(\Delta d_{t+1+k} - r_{t+1+k}), \tag{2.9}$$

thus, it is determined in terms of a discounted present value, the second term on the right hand side, that involves the growth rate of dividends and returns. In the case of a constant discount factor Equation 2.9 can be simplified as:

$$f_t = \frac{\kappa + \log(\gamma)}{1-\rho} + \sum_{k=0}^{\infty}\rho^k \mathbb{E}_t(\Delta d_{t+1+k}),$$

where $\gamma = 1/(1+R)$ is the constant real discount factor. The bubble component of Equation

2.8 can be written as:

$$b_t = \lim_{i\to\infty} \rho^i \mathbb{E}_t(p_{t+i}),$$

thus, it is determined as the asymptotic discounted present expectation of future asset prices.

In the absence of bubbles $b_t = \lim_{i\to\infty} \rho^i \mathbb{E}_t(p_{t+i}) = 0$ and the transversality condition holds. Instead, in the presence of bubbles $b_t = \lim_{i\to\infty} \rho^i \mathbb{E}_t(p_{t+i}) \neq 0$ and the transversality condition does not hold. The bubble component then satisfies the submartingale property:

$$\begin{aligned}\mathbb{E}_t(b_{t+1}) &= \lim_{i\to\infty} \mathbb{E}_t(\rho^i p_{t+1+i}) \\ &= \frac{1}{\rho} \lim_{i\to\infty} \mathbb{E}_t(\rho^{i+1} p_{t+1+i}) \\ &= \frac{1}{\rho} b_t,\end{aligned}$$

so that, $\mathbb{E}_t(b_{t+1}) > b_t$, since $\frac{1}{\rho} > 1$. In other words, the bubble component follows an explosive process.

In the presence of bubbles, the log-linear approximation holds only under certain conditions. The log-linear approximation relies on the assumption that the sample mean of $p_t - d_t$ converges to the true population mean in the limit. However, when $b_t \neq 0$, the difference $p_t - d_t$ is explosive, leading to an explosive sample mean and:

$$\rho = \frac{1}{1 + \exp(\overline{d - p})} \to 1,$$

compromising the validity of the Equation 2.9 that determines the fundamental component f_t. Lee and Phillips (2016) show that the log-linear approximation and the present value identity 2.8 are still valid when the duration of the bubble is asymptotically negligible.

Chapter 3

Methodologies to Detect and Date Rational bubbles

3.1 A Recursive Right-Tailed Unit Root Test for Rational Bubbles Detection

Standard unit root tests are not able to capture explosive episodes in asset prices if these are periodically collapsing such as in Equation 2.6 (Evans, 1991). This happens because a periodically collapsing bubble process can behave like a unit root process or a stationary one when π is small. For this reason Phillips, Wu, and Yu (2011) and Phillips, Shi, and Yu (2015) decided to use an Augmented Dickey Fuller test recursively and they proved how this method is able to efficiently detect explosive episodes in prices (or price-dividend ratio) series. The test is employed as follows.

For each time-series y_t on which the test is going to be employed, the following autoregressive process is estimated by least squares:

$$y_t = \mu_t + \delta y_{t-1} + \sum_{k=1}^{K} \phi_k \Delta y_{t-k} + \varepsilon_{y,t}, \qquad \varepsilon_{y,t} \sim \mathcal{N}(0, \sigma_y^2),$$

where μ_t is the drift and K is the lag parameter (in the empirical application I omit the sum as in Phillips, Shi, and Yu (2015)). The unit root (martingale) null hypothesis is given by $H_0 : \delta = 1$ and the right-tailed alternative hypothesis is $H_1 : \delta > 1$. Phillips, Shi, and Yu (2015) consider the case in which there is a negligible drift:

$$y_t = dT^{-\eta} + \delta y_{t-1} + \sum_{k=1}^{K} \phi_k \Delta y_{t-k} + \varepsilon_{y,t}, \tag{3.1}$$

where d is a constant, T the sample size and η controls the magnitude of the intercept and the drift. In particular they consider the case in which $\eta > 1/2$, so that the order of magnitude of y_t is the same as that of a pure random walk.

The ADF test is performed repeatedly by expanding window. The first regression will consider a subset of data $\tau_0 = [Tl]$ for some fraction l of the total sample T. Phillips, Shi, and Yu (2015) suggest to use as initial window $l = 0.01 + 1.8/\sqrt{T}$. The subsequent regressions will consider $\tau = [Tn]$ for $n \in [l, 1]$. In doing so the result obtained is a series of ADF test statistics. The Supremum Augmented Dickey Fuller statistic (SADF hereafter) is defined as the supremum of this series:

$$SADF(l) = \sup_{n \in [l,1]} ADF_0^n. \tag{3.2}$$

The General Supremum Augmented Dickey Fuller test (GSADF hereafter) is performed in a similar way, but instead of just an expanding window approach, it considers a rolling-expanding window regression: set $m \in [0, n-l]$ and $n \in [l, 1]$, so that,

$$GSADF(l) = \sup_{\substack{n \in [l,1] \\ m \in [0, n-l]}} \{ADF_m^n\}, \tag{3.3}$$

with window width $[T(n-m)]$, the fraction l for the initial window is specified as above.

Phillips, Wu, and Yu (2011) proved that the limiting distribution of the statistic in Equation 3.2 is:

$$SADF(l) \to \sup_{n \in [l,1]} \frac{\int_0^n W^\mu(s)\, dW(s)}{\left(\int_0^n [W^\mu(s)]^2\right)^{1/2}},$$

where $W^{\mu}(n) = W(n) - \frac{1}{n}\int_0^1 W(s)\ ds$ is a demeaned Wiener process.

Phillips, Shi, and Yu (2015) proved that the limiting distribution of the statistic in Equation 3.3, with the assumption of negligible drift is,

$$GSADF(l) \to \sup_{\substack{n\in[l,1]\\ m\in[0,n-l]}} \left\{ \frac{\frac{1}{2}(n-m)[[W(n)]^2 - [W(m)]^2 - (n-m)] - \int_m^n W(s)\ ds[W(n) - W(m)]}{(n-m)^{1/2}\left\{(n-m)\int_m^n [W(s)]^2\ ds - [\int_m^n W(s)\ ds]^2\right\}^{1/2}} \right\},$$

where $W(\cdot)$ is a standard Wiener process.

Knowing the two limiting distributions, it is possible to compute asymptotic critical values to compare with the statistics found. If the statistics are above the critical values the martingale hypothesis is rejected, implying the presence of at least one bubble.

Phillips, Shi, and Yu (2015) also use a date stamping technique, that I am going to employ in the empirical application to have a preliminary estimate for the start and end date of the bubble(s). The strategy is a double recursive test procedure called Backward Supremum Augmented Dickey Fuller test (BSADF hereafter). This consists in a SADF test on a backward expanding window, where the endpoint of the sample is fixed at n and the start point varies from 0 to $n - l$. The BSADF statistic is defined as:

$$BSADF_n(l) = \sup_{m\in[0,n-l]} \{ADF_m^n\}.$$

The start date of the bubble is then estimated to be the first point observation for which the BSADF test statistic goes above the corresponding critical value and the end date is estimated to be the first point observation for which the BSADF test statistic goes below the corresponding critical value.

3.2 Date Stamping Algorithm

The date stamping algorithm from Harvey, Leybourne, and Sollis (2017) is based on a minimum sum of squared residuals estimator. As a data generating process they assume a DGP that imposes a unit root on the series y_t (prices or price-dividend ratio series) up until the time $[\tau_1 T]$, after which the series becomes explosive up until the time $[\tau_2 T]$ (this is the bubble regime), then the series becomes stationary up until $[\tau_3 T]$ (collapsing regime) and then it restores to a unit root process until the sample end, for T sample size and τ_j fraction of the sample, with $j = 1, 2, 3$.

$$y_t = \mu + u_t, \qquad u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 2, ..., [\tau_1 T] \\ (1+\delta_1)u_{t-1} + \varepsilon_t, & t = [\tau_1 T] + 1, ..., [\tau_2 T] \\ (1-\delta_2)u_{t-1} + \varepsilon_t, & t = [\tau_2 T] + 1, ..., [\tau_3 T] \\ u_{t-1} + \varepsilon_t, & t = [\tau_3 T] + 1, ..., T \end{cases}$$

Because of this specification there are four different DGPs possible:

1. unit root then bubble to sample end;
2. unit root, bubble, then unit root to sample end;
3. unit root, bubble, then collapse to sample end;
4. unit root, bubble, collapse, then bubble to sample end.

For each DGP an OLS regression is estimated:

$$\begin{aligned} DGP_1 &: \Delta y_t = \hat{\delta}_1 D_t(\tau_1, 1) y_{t-1} + \hat{\varepsilon}_{1,t}, \\ DGP_2 &: \Delta y_t = \hat{\delta}_1 D_t(\tau_1, \tau_2) y_{t-1} + \hat{\varepsilon}_{2,t}, \\ DGP_3 &: \Delta y_t = \hat{\delta}_1 D_t(\tau_1, \tau_2) y_{t-1} + \hat{\delta}_2 D_t(\tau_2, 1) y_{t-1} + \hat{\varepsilon}_{3,t}, \\ DGP_4 &: \Delta y_t = \hat{\delta}_1 D_t(\tau_1, \tau_2) y_{t-1} + \hat{\delta}_2 D_t(\tau_2, \tau_3) y_{t-1} + \hat{\varepsilon}_{4,t}, \end{aligned}$$

where D_t is a dummy variable defined as $D_t(\tau_j, \tau_{j+1}) = 1([\tau_j T] < t < [\tau_{j+1} T])$. For simplicity, the constant term μ is not considered.

The change point estimators are then computed as:

$$DGP_1 : \hat{\tau}_1 = \min_{\substack{0<\tau_1<1, \\ y_T > y_{[\tau_1 T]}}} SSR_1(\tau_1)$$

$$DGP_2 : (\hat{\tau}_1, \hat{\tau}_2) = \min_{\substack{0<\tau_1<\tau_2<1 \\ y_{[\tau_2 T]} > y_{[\tau_1 T]}}} SSR_2(\tau_1, \tau_2)$$

$$DGP_3 : (\hat{\tau}_1, \hat{\tau}_2) = \min_{\substack{0<\tau_1<\tau_2<1 \\ y_{[\tau_2 T]} > y_{[\tau_1 T]} \\ y_{[\tau_2 T]} > y_T}} SSR_3(\tau_1, \tau_2)$$

$$DGP_4 : (\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3) = \min_{\substack{0<\tau_1<\tau_2<\tau_3<1 \\ y_{[\tau_2 T]} > y_{[\tau_1 T]} \\ y_{[\tau_2 T]} > y_{[\tau_3 T]}}} SSR_4(\tau_1, \tau_2, \tau_3),$$

where $SSR_i(\cdot) = \sum_{t=2}^{T} \hat{\varepsilon}_{it}^2$ represents the sum of squared residuals. The constraints $y_{[\tau_2 T]} > y_{[\tau_1 T]}$ and $y_{[\tau_2 T]} > y_{[\tau_3 T]}$ are useful to ensure a positive bubble regime (between τ_1 and τ_2) and a downward stationary regime (between τ_2 and τ_3). Harvey, Leybourne, and Sollis (2017) proved that these estimators are consistent for the true change points. However, for DGP_4 this is true only if the condition $(1+\delta_1)^{\tau_2-\tau_1}(1-\delta_2)^{\tau_3-\tau_2} \geq 1$ holds. They also proved via simulations that this algorithm is more precise than the dating that results from the BSADF statistic.

Chapter 4

Empirical Application Results

4.1 S&P500 Application

As a first empirical application I use data on the S&P500 Index. I download the real price and the real dividend series (monthly observations) from the Robert Shiller website[1], from January 1926 to March 2024, totalling 1179 data points. I chose to use monthly observations because they are less noisy than daily prices, yet they still allow me to capture relatively short-term dynamics. Figure 4.1 shows the price of the S&P500 Index relative to its fundamental, i.e. the price-dividend ratio.

I firstly employ the SADF and GSADF tests to the price-dividend ratio. Table 4.1 presents the two test statistics. Together the finite sample critical values are reported, obtained from 1000 replications of 1179 observations, where I assumed that the model under the null hypothesis is the one described in Equation 3.1. The transient dynamic lag order is set to $K = 0$. The initial window size is set to $l = 0.01 + 1.8/\sqrt{1179} = 0.0624$. Both the SADF and GSADF statistics exceed their 99% critical value suggesting at least an explosive episode in the price-dividend series, implying the presence of at least one rational bubble in the sample. Moreover, these results are really similar from Phillips, Shi, and Yu (2015) for which the sample started in January 1871 and ended in December 2010.

[1]available at: `http://www.econ.yale.edu/~shiller/data.htm`

Figure 4.1: S&P500 Price-Dividend ratio, between January 1926 to March 2024.

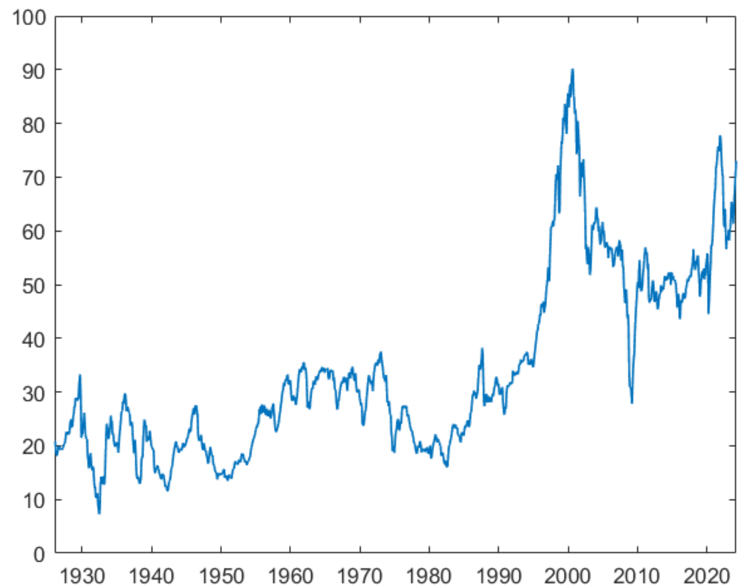


Table 4.1: SADF and GSADF test statistics and finite sample critical values. The finite sample critical values are obtained from Monte Carlo simulations, with 1000 replications (sample size $T = 1179$). The smallest window has 73 observations.

		Finite sample critical values		
	Test statistic	**90%**	**95%**	**99%**
SADF	3.1712	0.9774	1.2499	1.7492
GSADF	4.1603	2.1005	2.3156	2.9450

After having detected the presence of a bubble, I employ the BSADF test to the price-dividend ratio series for a preliminary estimate on the start and end dates of the bubble(s). Figure 4.2 shows the BSADF test statistics on the price-dividend ratio series against the corresponding 95% critical values. The estimate for the start date is the first point observation for which the statistic exceeds the critical value and the estimate for the end date is the first point observation for which the statistic goes below the critical value. I consider a bubble episode if the critical value is surpassed at least for three consecutive observations and I consider two different bubbles if there are at least three different observations below the critical value. From Figure 4.2 at least 7 episodes are observed, for which the preliminary dates are listed in Table 4.2.

Three of these episodes were also found in Phillips, Shi, and Yu (2015): the post-war boom

Figure 4.2: The BSADF test statistics against the corresponding 95% critical values.

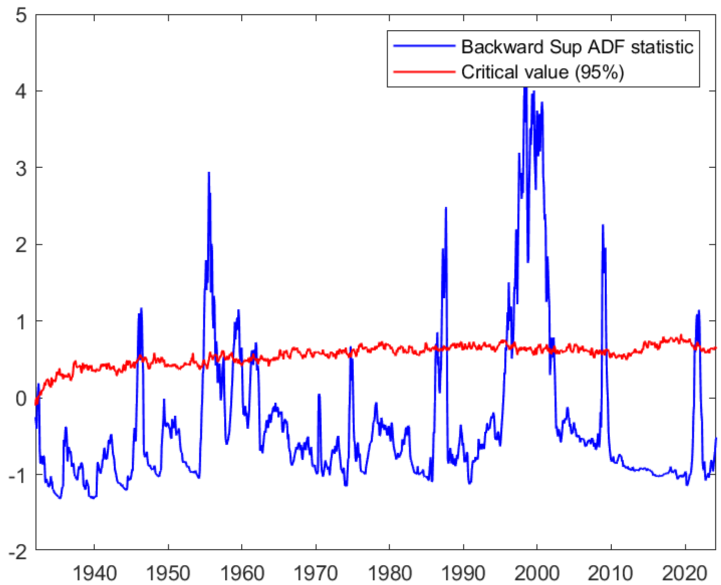


Table 4.2: Start date and end date for each episode found in the S&P500 with the BSADF estimator.

	Start date	End date
Episode A	Nov-45	Jul-46
Episode B	Nov-54	Aug-56
Episode C	Nov-58	Jan-60
Episode D	Apr-86	Sep-87
Episode E	Nov-95	Aug-01
Episode F	Oct-08	May-09
Episode G	Jun-21	Dec-21

(episode B), the Black Monday (episode D) and the dot-com bubble (episode E). Episode F, the subprime mortgage crisis, is not a bubble but a market downturn that is still detected by the test as explosive behaviour. Episode G is too recent to have been detected by the authors and it can be attributable to the Covid19-crisis. From this step onwards I only consider the bubble episodes that lasted at least 12 months.

Subsequently, I utilize the Harvey, Leybourne, and Sollis (2017) (HLS hereafter) minimum sum of squares estimator, to obtain a more precise estimate of the start and end date of the bubbles. Since this date stamping algorithm is constructed so that there is only one bubble, I need to divide the sample: the first subsample goes from January 1950 to May 1958, the

Table 4.3: Start date, end date and collapse end date estimated with the date stamping algorithm from Harvey, Leybourne, and Sollis (2017).

	Start date	End date	Collapse end date
Episode B	Sep-53	Mar-56	Nov-57
Episode D	Jun-85	Sep-87	Oct-88
Episode E	Jan-95	Aug-00	Sep-02

second subsample goes from January 1970 to December 1991, the third subsample goes from January 1992 to December 2005. Table 4.3 shows the results. For episode B both the start date and the end date are earlier than what the BSADF estimated. In 31 months the real price for the S&P500 Index increased by 104.84% and then declined during the collapse period, lasted 20 months, by 19.82%. For episode D the start date is earlier than what estimated by the BSADF but the end date is the same. In 28 months the price increased by 57.86% and then declined during the collapse period, lasted 13 months, by 16.72%. Also for episode E both the start date and the end dates are earlier than what the BSADF estimated. These tests and estimators consider the dot-com bubble to last 5 years and 8 months and in this period the real price for the S&P500 Index increased by 177.71%, from a level of 965.04 in January 1995 to a level of 2680 in August 2000. During the collapse period, lasted 2 years, the price dropped by 28%.

4.2 Fama-French 49 Industries

In this empirical application I will consider whether the econometric test used to detect the presence of bubbles has a predictive power for the crash that follows the abnormal price run-up. In particular, I compare the predictive power of the econometric test with the predictive power of the methodology used in Greenwood, Shleifer, and You (2019). To make an easier comparison with Greenwood, Shleifer, and You (2019) results, I consider 49 industries, identified as in the classification from Fama and French[2], without considering the *Others* sector so that the industries under study are 48. I download the data for each firm, for a total of 37747 firms, from the CRSP database, from January 1926 until December 2023,

[2]Details available at: `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_49_ind_port.html`.

both with daily observations and monthly observations. For each firm I consider, the SIC code (needed to separate the different sectors), the closing price, the dividend cash amount, the shares traded and the shares outstanding. For each firm I compute the daily and the monthly returns and the volatility of daily returns. I then aggregate the firms in the different sectors, using the mean for the prices and the dividends as aggregating measure, the returns are instead both value-weighted and equally-weighted. I then compute the variable I need for detecting bubbles with the BSADF statistic, that is the log price-dividend ratio, simply constructed as $\log(P_{i,t}) - \log(D_{i,t})$, for sector i and month t.

What I am interested in is the comparison of the ability of different methodologies not only to detect bubbles, but to detect those bubbles that have a crash afterwards. Is an econometric test really able to capture this better than a naive model? To answer this question I run a logistic regression and I consider four different methods to detect bubbles: the first two methodologies are based on Phillips, Shi, and Yu (2015) and the last two on Greenwood, Shleifer, and You (2019). In particular, I am interested in the probability of a crash in the following 2-years, given that in the month before, the industry was in an explosive period. That is, I am interested in $E[Y|X] = P(Y = 1|X)$, where $X = [\text{Bubble}_{i,t-1}; \text{Characteristics}_{i,t-1}]$ and Y represents the *Crash* dummy, constructed as:

$$\begin{cases} \text{Crash}_{i,t} = 1 & \text{if } R_{i,t\to t+24} \leq -40\%, \\ \text{Crash}_{i,t} = 0 & \text{Otherwise,} \end{cases} \tag{4.1}$$

where $R_{i,t\to t+24}$ are two-years raw returns for industry i. With a logistic regression I can estimate the probability of a crash given by the presence of a price peak as:

$$P(Y = 1|X) = \frac{\exp\left[\alpha + \beta_1 \cdot \text{Bubble}_{i,t-1} + \beta_2 \cdot \text{Characteristics}_{i,t-1}\right]}{1 + \exp\left[\alpha + \beta_1 \cdot \text{Bubble}_{i,t-1} + \beta_2 \cdot \text{Characteristics}_{i,t-1}\right]}. \tag{4.2}$$

β_1 is the coefficient that really matters, it represents the change in $\text{logit}[P(Y = 1|X)]$ when $\text{Bubble}_{i,t-1} = 1$ compared to the case in which $\text{Bubble}_{i,t-1} = 0$. More clearly, e^{β_1} represents how much higher (or lower) the probability of $Y = 1$ is when $\text{Bubble}_{i,t-1} = 1$ compared to the case in which $\text{Bubble}_{i,t-1} = 0$.

The *Bubble* dummy for the first two methodologies is constructed as:

$$\begin{cases} \text{Bubble}_{i,t} = 1 & \text{if } BSADF_{i,t}(l) > cv_t(\alpha), \\ \text{Bubble}_{i,t} = 0 & \text{Otherwise,} \end{cases} \tag{4.3}$$

where $BSADF_{i,t}(l)$ is the $BSADF$ statistic for month t and industry i. To avoid look-ahead bias, it is computed so that for month t I only consider observations up to month t. l is the fraction for starting window and it depends on the length of the sample as explained above (see Section 3.1). $cv_t(\alpha)$ is the critical value for month t, with α significance level. I consider $\alpha = 95\%$ and $\alpha = 99\%$.

The *Bubble* dummy for the third methodology is constructed as:

$$\begin{cases} \text{Bubble}_{i,t} = 1 & \text{if } R_{i,t-24\to t} \geq 100\%, \\ \text{Bubble}_{i,t} = 0 & \text{Otherwise,} \end{cases} \tag{4.4}$$

where $R_{i,t-24\to t}$ are two-years raw returns for month t and industry i.

The *Bubble* dummy for the fourth methodology is constructed as:

$$\begin{cases} \text{Bubble}_{i,t} = 1 & \text{if } R^{VW}_{i,t-24\to t} \geq 100\%, \\ \text{Bubble}_{i,t} = 0 & \text{Otherwise,} \end{cases} \tag{4.5}$$

where $R^{VW}_{i,t-24\to t}$ are two-years value-weighted raw returns for month t and industry i.

The *Characteristics* matrix contains some controlling variables: volatility, turnover, firm age and issuance. These characteristics are constructed following Greenwood, Shleifer, and You (2019) as:

- Volatility: I compute the percentile rank of volatility in the cross-section of firms each month for each sector, the industry volatility is then computed as the value-weighted mean of the rank for each industry, for example in January 1962, for the Bank sector the volatility is 0.42, meaning that 42% of firms had lower volatility than the average firm in that industry.

- Turnover: the turnover is defined as shares traded divided by shares outstanding, every month I compute the percentile rank of turnover in the whole cross-section of firms, the industry turnover is then computed as the value-weighted mean of the rank, for example for the Hardware sector in August 2000, the turnover was 0.69, meaning that the value-weighted turnover for that industry was higher than 69% of all listed stocks.

- Age: the age of a firm is measured as the number of years since the firm first appeared on CRSP, then to construct the industry age I compute the percentile rank of the age for every listed stock and I compute the value-weighted mean of this rank for each industry, for example, in March 1932 for the Aircraft sector, the age is 0.62 meaning that the value-weighted age for that industry was higher than 62% of all listed stocks.

- Issuance: considering the issuers those firms for which the shares increased by 5% in one year, this variable is constructed as the percentage of issuers in the industry in the past year.

I chose to use these characteristics as control variables in the logistic regression because Greenwood, Shleifer, and You (2019) found that they all have predictive power for subsequent returns. Moreover, these characteristics tend to increase abnormally during price run-up periods that are followed by crashes. Because of these reasons, it is possible that, after controlling for these variables, the coefficient on the *Bubble* dummy will not be significant anymore. This would imply that these variables have better predictive power for the subsequent crash than the presence of the bubble itself.

Tables 4.4 and 4.5 show some summary statistics for the control variables volatility, its 1-year change, turnover, its 1-year change, industry age and percentage of issuers. Table 4.4 considers the BSADF statistic as methodology to detect the presence of a bubble. All characteristics have a higher mean during a bubble episode, except for the percentage of issuers, however, between the bubble episodes that crash and those that do not there is not much difference. Table 4.5 considers two-years raw returns to detect price run-ups. Also in this case all characteristics have a higher mean during a price run up episode, however only volatility have a higher mean during those price run ups that have a subsequent crash than during those price run ups without a subsequent crash.

Table 4.4: **Summary statistics.** Summary statistics of different characteristics: volatility, the one-year change of volatility, turnover, the one-year change of turnover, industry age and percentage of issuers. The explosive episodes are detected with the BSADF statistic using 99% critical values. A crash is defined as a price drawdown of 40% in a two-years period. The columns of the second panel of this table show summary statistics for explosive episodes with a subsequent crash and for explosive episodes without a subsequent crash.

	All observations		Explosive Episodes	
	Mean	SD	Mean	SD
Volatility	0.320	0.131	0.464	0.197
1-year change in Volatility	-0.0004	0.120	0.046	0.245
Turnover	0.498	0.172	0.573	0.215
1-year change in Turnover	-0.002	0.099	0.072	0.154
Age	0.535	0.119	0.607	0.163
Issuance	0.165	0.871	0.030	0.149
	With Crash		**With no Crash**	
	Mean	SD	Mean	SD
Volatility	0.446	0.222	0.477	0.179
1-year change in Volatility	-0.047	0.214	0.114	0.247
Turnover	0.498	0.227	0.625	0.192
1-year change in Turnover	-0.016	0.111	0.138	0.149
Age	0.590	0.203	0.619	0.129
Issuance	0.000	0.000	0.051	0.192

Figure 4.3 shows a predicted probabilities plot of the probability of a crash conditional on two different variables: the blue line is estimated using the BSADF statistic as a predictor variable, whereas the red line is estimated using past value-weighted 2-years raw returns. Both predictor variables are standardized in order to make an easier comparison. It can be seen how the BSADF statistic corresponds to a higher probability of a crash in almost every point of the graph. The results from the logistic regressions are shown in the Tables 4.6, 4.7, 4.8 and 4.9 whereas Table 4.10 shows the percentage of all the explosive observations found that have a subsequent crash. First of all, I can observe in the logistic regression that controlling for the different variables does reduce the coefficients on all the Bubble variables but it does not make them insignificant. I can also observe that the econometric test is able to detect bubbles that have a subsequent crash better than the other models, but only when using a high critical value, the performance of the same model when using a lower critical value is worse than the naive model. The winner methodology captures much less explosive

Table 4.5: **Summary statistics.** Summary statistics of different characteristics: volatility, the one-year change of volatility, turnover, the one-year change of turnover, industry age and percentage of issuers. The price-run ups are detected when the two-year raw industry return is equal or higher to 100%. A crash is defined as a price drawdown of 40% in a two-years period. The columns of the second panel of this table show summary statistics for price run-up episodes with a subsequent crash and for price run-up episodes without a subsequent crash.

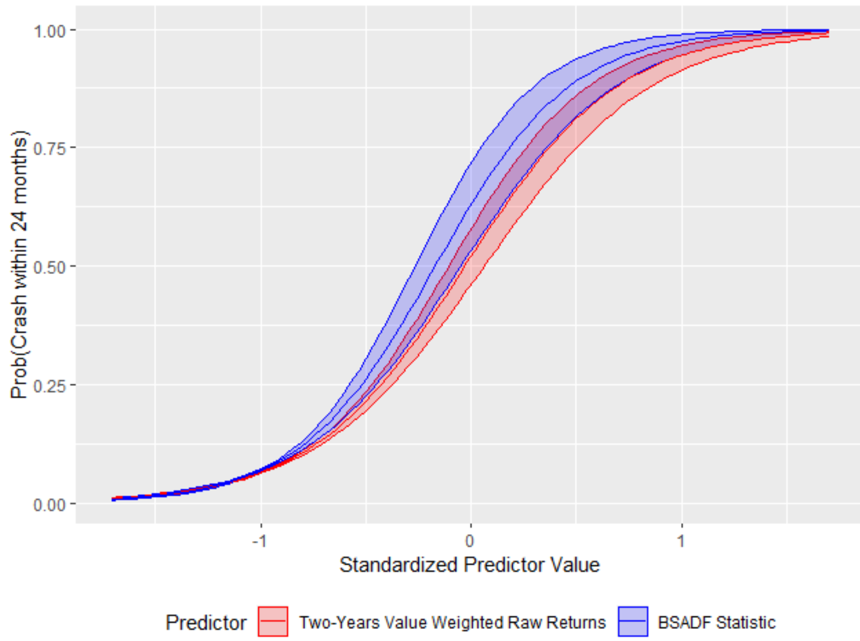
	All observations		Price Run-Up Episodes	
	Mean	SD	Mean	SD
Volatility	0.320	0.131	0.333	0.154
1-year change in Volatility	-0.0004	0.120	0.014	0.132
Turnover	0.498	0.172	0.551	0.173
1-year change in Turnover	-0.002	0.099	-0.019	0.143
Age	0.535	0.119	0.549	0.144
Issuance	0.165	0.871	0.361	2.52
	With Crash		**With no Crash**	
	Mean	SD	Mean	SD
Volatility	0.375	0.144	0.320	0.155
1-year change in Volatility	0.006	0.134	0.017	0.131
Turnover	0.542	0.194	0.554	0.165
1-year change in Turnover	-0.026	0.133	-0.017	0.146
Age	0.517	0.147	0.559	0.142
Issuance	0.092	0.403	0.446	2.88

price episodes, but more than 40% of them have a subsequent crash. On the other hand, the other methodologies capture more price run-up episodes, but only 20% of them have a subsequent crash. Thus, the econometric methodology is better able to capture bubble episodes.

To get a better understanding of which methodology is really able to capture the probability of the crash after the price run-up, I carry out a comparison between them. Thus, I run the logistic regression:

$$\begin{aligned}\text{logit}[P(Y=1|X)] &= \log\left[\frac{P(Y=1|X)}{1-P(Y=1|X)}\right] \\ &= \alpha + \beta_1 \cdot \text{Bubble}_{i,t-1}^{\text{BSADF99}} + \beta_2 \cdot \text{Bubble}_{i,t-1}^{\text{EW}} \\ &+ \beta_3 \cdot \text{Bubble}_{i,t-1}^{\text{VW}} + \beta_4 \cdot \text{Characteristics}_{i,t-1}.\end{aligned}$$

Figure 4.3: **Crash Predictability**. Probability of crash given by two different predictor variables: BSADF (blue) and value weighted raw returns (red). The line is the estimate and the colored area represents the 95% confidence interval. The predictor variables are standardized for easier comparison.



Y represents the *Crash* dummy constructed as in Equation 4.1, $\text{Bubble}_{i,t-1}^{\text{BSADF99}}$ represents the *Bubble* dummy constructed as in Equation 4.3 with $\alpha = 99\%$, $\text{Bubble}_{i,t-1}^{\text{EW}}$ as in Equation 4.4 and $\text{Bubble}_{i,t-1}^{\text{VW}}$ as in Equation 4.5. With regard to the dummy constructed considering the BSADF statistic, I chose to use only the one with $\alpha = 99\%$, because of its better performance with respect to the one with $\alpha = 95\%$. The *Characteristics* matrix contains the same control variables as in 4.2. The results from this regression are shown in Table 4.11. All coefficients on the various *Bubble* dummies are significant, but the highest coefficient is the one from the econometric statistic, confirming its better predictive power. Controlling for volatility, turnover, issuance and firm age does slightly decrease the coefficients but they all remain significant. To be sure about the difference between the coefficients I carry out two Wald tests, with $H_0 : \beta_1 = \beta_2$ and $H_1 : \beta_1 \neq \beta_2$ for the first one, and $H_0 : \beta_1 = \beta_3$ and $H_1 : \beta_1 \neq \beta_3$ for the second one. The first difference is significant at 5% and the second one at 1%. Including also the control variables, both differences are significant at 1%. The tests on the difference between the coefficients confirm the better performance of the econometric methodology in detecting bubbles.

Table 4.6: **Logistic Regression Coefficients.** Results from the logistic regression using the BSADF statistic to detect the presence of bubble episodes considering critical values at 99%. Maximum likelihood standard errors are in parenthesis. (***) significant at 99% (**) significant at 95% (*) significant at 90%.

	BSADF 99				
Intercept	-2.602***	-2.887***	-2.491***	-0.957***	-0.958***
	(0.018)	(0.048)	(0.065)	(0.109)	(0.109)
Bubble	2.245***	2.122***	2.179***	2.427***	2.417***
	(0.247)	(0.132)	(0.249)	(0.258)	(0.258)
Volatility	-	0.875***	0.992***	0.789***	0.786***
	-	(0.132)	(0.102)	(0.133)	(0.133)
Turnover	-	-	-0.891***	-0.924***	-0.903***
	-	-	(0.102)	(0.102)	(0.102)
Age	-	-	-	-2.799***	-2.793***
	-	-	-	(0.162)	(0.162)
Issuance	-	-	-	-	-0.080**
	-	-	-	-	(0.030)

From Table 4.12 to Table 4.15 all the bubble episodes are shown, in particular I consider only those episodes that have a subsequent crash and only for those sectors that have more than 10 firms by the time the episode started. There is alignment with what I find and Greenwood, Shleifer, and You (2019) results, but I find much more episodes. There is however great heterogeneity between the different methods, the only episode that is captured by every method is the one in the Personal Services industry in October 1968. This is due to the fact that the econometric test does not consider whether prices are high or low, but it is designed to detect rational bubbles, where the time series behavior of prices is more explosive than the time series behavior of the fundamentals (in this case, dividends). The two methods differ in the definition of the price run-up leg of the bubble, but I compare them using the same definition for the drawdown part of the bubble, and results suggest that the econometric test designed to detect rational bubbles has a better predictive power for the collapse following the explosive period.

Table 4.7: **Logistic Regression Coefficients.** Results from the logistic regression using the BSADF statistic to detect the presence of bubble episodes considering critical values at 95%. Maximum likelihood standard errors are in parenthesis. (***) significant at 99% (**) significant at 95% (*) significant at 90%.

	BSADF 95				
Intercept	-2.610***	-2.894***	-2.506***	-0.987***	-0.988***
	(0.018)	(0.048)	(0.065)	(0.109)	(0.109)
Bubble	1.355***	1.291***	1.288***	1.335***	1.340***
	(0.136)	(0.136)	(0.136)	(0.138)	(0.138)
Volatility	-	0.871***	0.988***	0.790***	0.7875***
	-	(0.132)	(0.131)	(0.133)	(0.134)
Turnover	-	-	-0.874***	-0.907***	-0.884***
	-	-	(0.102)	(0.102)	(0.102)
Age	-	-	-	-2.773***	-2.767***
	-	-	-	(0.162)	(0.162)
Issuance	-	-	-	-	-0.085**
	-	-	-	-	(0.031)

Table 4.8: **Logistic Regression Coefficients.** Results from the logistic regression, price run-up episodes are detected as an increase of 100% or more in 2-years equally-weighted raw returns. Maximum likelihood standard errors are in parenthesis. (***) significant at 99% (**) significant at 95% (*) significant at 90%.

	Equally-Weighted Returns				
Intercept	-2.666***	-2.952***	-2.506***	-0.967***	-0.970***
	(0.018)	(0.048)	(0.065)	(0.109)	(0.108)
Bubble	1.512***	1.501***	1.561***	1.609***	1.618***
	(0.069)	(0.069)	(0.069)	(0.070)	(0.071)
Volatility	-	0.879***	1.016***	0.821***	0.823***
	-	(0.132)	(0.130)	(0.133)	(0.133)
Turnover	-	-	-1.013***	-1.045***	-1.021***
	-	-	(0.103)	(0.103)	(0.103)
Age	-	-	-	-2.818***	-2.812***
	-	-	-	(0.162)	(0.162)
Issuance	-	-	-	-	-0.085**
	-	-	-	-	(0.028)

Table 4.9: **Logistic Regression Coefficients.** Results from the logistic regression, price run-up episodes are detected as an increase of 100% or more in 2-years value-weighted raw returns. Maximum likelihood standard errors are in parenthesis. (***) significant at 99% (**) significant at 95% (*) significant at 90%.

	Value-Weighted Returns				
Intercept	-2.677***	-2.939***	-2.454***	-0.946***	-0.948***
	(0.019)	(0.048)	(0.064)	(0.109)	(0.109)
Bubble	1.302***	1.279***	1.375***	1.381***	1.389***
	(0.061)	(0.062)	(0.062)	(0.063)	(0.0631)
Volatility	-	0.809***	0.955***	0.761***	0.761***
	-	(0.132)	(0.130)	(0.133)	(0.133)
Turnover	-	-	-1.106***	-1.142***	-1.119***
	-	-	(0.104)	(0.103)	(0.103)
Age	-	-	-	-2.751***	-2.745***
	-	-	-	(0.162)	(0.162)
Issuance	-	-	-	-	-0.089**
	-	-	-	-	(0.028)

Table 4.10: **Percentage of crash observations.** Percentage of the explosive observations found with each method that crashed in the following two-years. The methods are, respectively: BSADF statistic using 99% critical values, BSADF statistic using 95% critical values, two-years raw returns exceeding 100%, two-years value-weighted raw returns exceeding 100%.

BSADF 99	**BSADF 95**	**Raw Returns**	**Value-Weighted Raw Returns**
41.18%	22.19%	23.97%	20.19%

Table 4.11: **Comparison between methods: Logistic Regression Results.** This table summarises the results from the logistic regression including all methodologies used to detect the presence of potential bubbles and some control variables. Maximum likelihood standard errors are in parenthesis. (***) significant at 99% (**) significant at 95% (*) significant at 90%.

	Comparison between methods				
Intercept	-2.693***	-2.946***	-2.455***	-0.898***	-0.901***
	(0.019)	(0.048)	(0.064)	(0.109)	(109)
Bubble$^{\mathbf{BSADF99}}$	1.621***	1.519***	1.564***	1.817***	1.801***
	(0.263)	(0.265)	(0.264)	(0.266)	(0.266)
Bubble$^{\mathbf{EW}}$	0.926***	0.935***	0.941***	0.994***	1.000***
	(0.090)	(0.090)	(0.091)	(0.091)	(0.091)
Bubble$^{\mathbf{VW}}$	0.815***	0.791***	0.884***	0.871***	0.878***
	(0.080)	(0.081)	(0.081)	(0.082)	(0.082)
Volatility	-	0.781***	0.931***	0.722***	0.726***
	-	(0.133)	(0.131)	(0.133)	(0.133)
Turnover	-	-	-1.121***	-1.152***	-1.129***
	-	-	(0.103)	(0.103)	(0.104)
Age	-	-	-	-2.848***	-2.843***
	-	-	-	(0.163)	(0.162)
Issuance	-	-	-	-	-0.086**
	-	-	-	-	(0.027)

Table 4.12: **Bubble Episodes.** Explosive episodes that have at least 1 subsequent crash observation. The explosive episodes are found with the BSADF statistic using 99% critical values. For each panel the first column defines the sector and the second one when the price explosion is first observed. Panel A shows how many months the price explosion lasted, how many observations in that time span have a following crash and how many firms were in the sector by the time the bubble started. Panel B shows the performance of the sector after the bubble peak.

Panel A: Bubble Episodes

Sector	**Date of first observed explosive observation**	**Number of explosive observations**	**Number of subsequent crash observations**	**Number of firms**
banks	03/1961	1	1	18
banks	11/1961	1	1	19
fun	01/1961	1	1	12
banks	01/1962	3	3	20
rtail	01/1962	1	1	76

Panel B: Equally weighted and value weighted returns

Sector	**Date of first observed explosive observation**	**12-months raw returns**	**24-months raw returns**	**12-months raw value-weighted returns**	**24-months raw value-weighted returns**
banks	03/1961	-13.24%	-48.24%	-9.57%	-10.60%
banks	11/1961	-61.79%	-62.30%	-17.10%	-13.70%
fun	01/1961	-65.01%	-59.43%	-22.80%	-16.00%
banks	01/1962	-44.60%	-49.39%	-1.56%	-22.80%
rtail	01/1962	-49.54%	-49.69%	-18.50%	-12.90%

Table 4.13: **Bubble Episodes.** Explosive episodes that have at least 1 subsequent crash observation. The explosive episodes are found with the BSADF statistic using 95% critical values. For each panel the first column defines the sector and the second one when the price explosion is first observed. Panel A shows how many months the price explosion lasted, how many observations in that time span have a following crash and how many firms were in the sector by the time the bubble started. Panel B shows the performance of the sector after the bubble peak.

Panel A: Bubble Episodes

Sector	**Date of first observed explosive observation**	**Number of explosive observations**	**Number of subsequent crash observations**	**Number of firms**
clths	11/1945	2	1	21
clths	01/1946	1	1	21
banks	02/1961	4	4	18
fun	03/1961	3	3	14
banks	08/1961	9	8	18
fun	01/1962	2	2	12
rtail	01/1962	3	3	76
persv	10/1968	3	3	14
softw	01/1983	6	6	50
banks	01/1986	3	3	332
food	05/1986	2	1	88

Panel B: Equally weighted and value weighted returns

Sector	**Date of first observed explosive observation**	**12-months raw returns**	**24-months raw returns**	**12-months raw value-weighted returns**	**24-months raw value-weighted returns**
clths	11/1945	-34.34%	-44.92%	-26.10%	-28.00%
clths	01/1946	-27.18%	-44.75%	-20.30%	-28.60%
banks	02/1961	-36.60%	-46.10%	-9.63%	-4.92%
fun	03/1961	-41.94%	-60.76%	-39.90%	-17.60%
banks	08/1961	-25.36%	-31.70%	12.10%	6.00%
fun	01/1962	-56.33%	-52.67%	-14.90%	-9.67%
rtail	01/1962	-49.18%	-49.18%	0.73%	4.32%
persv	10/1968	-44.62%	-50.19%	-29.30%	-19.80%
softw	01/1983	-60.62%	-54.41%	-31.90%	-22.20%
banks	01/1986	-22.31%	-40.45%	-0.09%	-22.40%
food	05/1986	-19.86%	-39.11%	-6.73%	-22.80%

Table 4.14: **Price Run-Up Episodes.** Explosive episodes that have at least 1 subsequent crash observation. The explosive episodes are defined as two-years raw returns that exceed a certain threshold X, with $X = 100\%$. The first column defines the sector, the second one when the price run-up is first observed, the third how many months the price run-up lasted, and the fourth how many observations in that time span have a following crash, the following two columns represents the performance of the sector after the bubble peak and the last column shows how many firms were in the sector by the time the bubble started.

Sector	**Date of first observed explosive observation**	**Number of explosive observations**	**Number of subsequent crash observations**	**12-months raw returns**	**24-months raw returns**	**Number of firms**
mines	05/1935	1	1	2.53%	-57.40%	20
mines	07/1935	1	1	4.89%	-56.05%	20
mines	11/1935	4	4	-22.18%	-80.03%	20
autos	03/1936	13	12	-7.69%	-66.67%	43
bldmt	03/1936	1	1	-1.34%	-53.29%	36
mach	05/1936	14	7	-23.16%	-33.22%	38
bldmt	07/1936	9	5	-49.42%	-45.39%	35
txtls	07/1936	11	9	-41.51%	-43.06%	18
hshld	08/1936	12	3	-18.75%	-2.13%	18
fin	09/1936	7	7	-60.89%	-65.83%	28
ships	09/1936	12	8	-35.78%	-0.71%	12
elceq	12/1936	7	3	-42.25%	-41.27%	20
steel	01/1937	5	5	-44.05%	-51.09%	62

Continued on next page

Table 4.14 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
trans	03/1937	2	2	-72.46%	-63.68%	78
aero	03/1940	1	1	-28.57%	-44.17%	13
autos	01/1946	1	1	-32.60%	-49.82%	54
txtls	01/1946	1	1	-32.38%	-51.58%	21
whlsl	01/1946	1	1	-20.42%	-49.54%	11
fun	04/1946	2	2	-64.30%	-73.09%	10
txtls	04/1946	1	1	-48.75%	-46.26%	22
whlsl	04/1946	3	3	-21.19%	-36.86%	10
chips	06/1967	1	1	-17.66%	-47.14%	89
meals	05/1968	9	9	-43.20%	-50.19%	22
cnstr	05/1968	8	8	-40.10%	-50.76%	21
persv	07/1968	5	5	-36.07%	-55.26%	12
rlest	08/1968	6	4	-26.05%	-21.43%	34
fun	09/1968	3	1	-35.20%	-44.63%	28
whlsl	09/1968	5	4	-18.96%	-37.55%	52
bussv	10/1968	1	1	-20.14%	-52.17%	57

Continued on next page

Table 4.14 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
rubbr	10/1968	2	2	-31.40%	-54.05%	29
txtls	11/1968	1	1	-47.82%	-47.39%	51
gold	01/1980	2	2	-20.38%	-40.27%	31
softw	06/1980	1	1	-25.98%	-36.14%	18
gold	06/1980	7	7	-68.65%	-61.95%	30
agric	02/1986	1	1	-56.54%	-73.51%	33
rlest	04/1986	24	18	-75.59%	-76.03%	116
rlest	10/1968	3	1	-6.24%	-74.81%	116
telcm	04/1999	1	1	-13.86%	-53.72%	279
chips	12/1999	5	5	-54.24%	-64.86%	323
softw	12/1999	3	3	-79.02%	-75.2%	698
hardw	12/1999	5	5	-58.03%	-61.67%	204
elceq	02/2000	2	2	-63.45%	-66.78%	249
chips	06/2000	4	4	-67.65%	-77.98%	346
drugs	08/2000	2	2	-34.31%	-55.11%	349
elceq	08/2000	2	2	-52.23%	-68.34%	129

Continued on next page

Table 4.14 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
labeq	08/2000	2	2	-52.23%	-68.34%	129
coal	10/2004	5	4	-1.07%	-28.89%	11
fabpr	08/2006	3	2	13.04%	-48.65%	14
coal	10/2010	6	3	-14.64%	-16.49%	18
gold	02/2011	1	1	-37.52%	-50.33%	72
gold	04/2011	1	1	-50.53%	-62.32%	76
coal	09/2017	12	9	1.29%	-24.65%	10
gold	07/2020	4	3	-26.77%	-41.34%	37
txtls	02/2021	2	2	-82.98%	-83.16%	11
beer	03/2021	3	3	-39.71%	-46.65%	14

Table 4.15: **Price Run-Up Episodes.** Explosive episodes that have at least 1 subsequent crash observation. The explosive episodes are defined as two-years value-weighted raw returns that exceed a certain threshold X, with $X = 100\%$. The first column defines the sector, the second one when the price run-up is first observed, the third how many months the price run-up lasted, and the fourth how many observations in that time span have a following crash, the following two columns represents the performance of the sector after the bubble peak and the last column shows how many firms were in the sector by the time the bubble started.

Sector	**Date of first observed explosive observation**	**Number of explosive observations**	**Number of subsequent crash observations**	**12-months raw returns**	**24-months raw returns**	**Number of firms**
elceq	02/1936	1	1	45.20%	-31.90%	17
mach	02/1936	3	2	34.70%	-42.20%	36
oil	02/1936	3	3	28.20%	-24.40%	45
autos	03/1936	1	1	-16.50%	-45.10%	43
autos	05/1936	11	10	-52.50%	-33.90%	44
fin	07/1936	2	2	-36.20%	-49%	28
txtls	08/1936	1	1	-21.90%	-33.20%	19
fin	10/1936	9	6	-32.80%	-46.90%	27
mines	10/1936	8	7	-28.80%	-31.60%	22
txtls	10/1936	6	6	-49.60%	-40.30%	18
bldmt	11/1936	5	5	-44.20%	-37.70%	39
elceq	12/1936	4	3	-41.20%	-30%	20
mach	12/1936	8	8	-32.90%	-43.90%	43

Continued on next page

Table 4.15 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
ships	12/1936	8	8	-36.70%	-41.70%	12
steel	01/1937	7	7	-23.80%	-28.30%	62
hshld	02/1937	3	1	-28.70%	-8.19%	18
trans	02/1937	4	4	-44.90%	-41.60%	77
aero	10/1939	7	1	-6.06%	-23.90%	12
beer	10/1945	5	5	-56.20%	-64.00%	11
trans	11/1945	1	1	-22.80%	-34.30%	76
txtls	01/1946	1	1	-22.00%	-32.80%	21
whlsl	01/1946	1	1	-13.20%	-25.80%	11
bussv	04/1946	4	3	-29.10%	-32.40%	10
txtls	04/1946	1	1	-40.90%	-20.50%	22
beer	04/1946	4	4	-34.20%	-36.20%	12
meals	05/1968	9	9	-21.20%	-23.70%	22
rlest	05/1968	7	7	-10.30%	-25.00%	33
rubbr	05/1968	8	8	16.50%	-27.50%	28
whlsl	05/1968	2	2	-13.40%	-29340%	51

Continued on next page

Table 4.15 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
persv	07/1968	6	6	-29.30%	-19.80%	12
elceq	08/1968	4	4	-17.80%	-10.70%	49
fun	08/1968	5	2	-4.95%	7.40%	27
whlsl	08/1968	6	5	-1.82%	-15.40%	51
bldmt	09/1968	1	1	-24.90%	-39.60%	128
bussv	09/1968	3	3	-16.70%	-25.60%	55
cnstr	10/1968	1	1	-34.00%	-50.70%	22
bldmt	11/1968	1	1	-24.00%	-17.80%	132
txtls	11/1968	1	1	-31.50%	-19.60%	51
toys	11/1968	1	1	-15.80%	-34.40%	44
meals	05/1972	1	1	-27.70%	-36.30%	41
toys	05/1972	3	3	-38.60%	-61.70%	49
smoke	06/1972	1	1	-7.65%	-44.20%	17
labeq	06/1972	1	1	0.61%	-15.50%	35
medeq	06/1972	1	1	-11.20%	-27.80%	25
gold	11/1979	19	18	-30.50%	25.90%	30

Continued on next page

Table 4.15 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
mines	01/1980	2	2	25.60%	-36.30%	47
agric	02/1986	3	1	-0.57%	-6.70%	33
soda	03/1986	1	1	-55.20%	-55.40%	12
gold	03/1987	7	6	-6.16%	18.40%	109
telcm	01/1999	2	2	27.80%	-30.30%	281
softw	01/1999	1	1	67.20%	-24.70%	579
toys	10/1999	6	6	-9.70%	2.69%	63
chips	11/1999	5	5	-56.30%	-57.60%	324
elceq	11/1999	4	4	-64.10%	-66.20%	251
labeq	12/1999	2	2	-23.10%	-39.90%	137
softw	12/1999	1	1	-26.60%	-38.30%	698
softw	02/2000	1	1	-47.20%	-43.80%	714
chips	06/2000	1	1	-49.90%	-63.50%	346
labeq	06/2000	1	1	-16.90%	-46.60%	132
chips	08/2000	2	2	-47.10%	-60.80%	384
labeq	08/2000	2	2	-26.60%	-45.80%	129

Continued on next page

Table 4.15 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
toys	08/2000	1	1	-23.40%	6.67%	61
hardw	08/2000	1	1	-41.10%	-46.30%	178
softw	08/2000	2	2	-27.10%	-34.20%	807
coal	10/2004	5	4	33.00%	-8.66%	11
cnstr	11/2006	2	2	-12.40%	-50.40%	58
agric	01/2007	10	4	-37.40%	-26.40%	17
cnstr	02/2007	4	4	-6.91%	-50.10%	58
steel	05/2007	2	2	8.80%	-39.60%	62
steel	08/2007	5	3	-58.20%	-34.70%	60
agric	12/2007	1	1	-40.10%	-24.20%	18
steel	05/2008	1	1	-50.00%	-44.10%	56
coal	06/2008	3	2	-16.70%	4.46%	16
coal	11/2010	6	3	-45.30%	-33.00%	18
gold	01/2011	4	4	-26.80%	-41.80%	71
steel	02/2018	8	4	-4.61%	-13.60%	40
gold	07/2020	6	3	-5.79%	-3.70%	37

Continued on next page

Table 4.15 continued from previous page

Sector	Date of first observed explosive observation	Number of explosive observations	Number of subsequent crash observations	12-months raw returns	24-months raw returns	Number of firms
softw	12/2020	1	1	20.50%	-36.10%	185
toys	01/2021	1	1	-0.39%	-20.40%	21
softw	06/2021	6	6	-52.90%	-32.20%	194

Conclusion

This thesis has focused in particular on two research questions: am I able to detect rational bubbles with statistical tests? Are econometric tests able to predict the crash that follows a bubble better than a simple naive model?

With the first empirical application I am more interested in answering the first question. To detect the rational bubbles I use the SADF (Supremum Augmented Dickey Fuller) test and the GSADF (Generalised Supremum Augmented Dickey Fuller) test from Phillips, Shi, and Yu (2015). These tests are recursive right-tailed unit root tests. The tests are employed on a measure of asset prices related to fundamentals, in this case the price-dividend ratio (or the log price-dividend ratio), because what is needed is the ability to capture the behavior of the time series of prices with respect to the behavior of the time series of dividends. The null hypothesis is a martingale hypothesis and the alternative hypothesis is right-tailed. More simply put, the test asks whether the price-dividend ratio series behave like a random walk or like an explosive process. If there is enough statistical evidence for the rejection of the null hypothesis, the presence of a bubble in the sample is implied. I then employ a date stamping algorithm from Harvey, Leybourne, and Sollis (2017) which estimates the start date, peak date and end date of the bubble with a minimum sum of squared residuals estimator. I employ these tests and algorithm on monthly observations of the real price-dividend ratio of the S&P500 Index from January 1926 until December 2023. I find six different explosive periods and a market downturn (the subprime mortgage crisis of 2008) that is still detected as explosive behavior. The longer episodes are the post-war boom (31 months to the peak then 20 months of collapse), the price run-up preceding the Black Monday of October 1987 (28 months to the peak then 13 months of collapse) and the dot-com bubble (68 months to the peak then 24 months of collapse). All these episodes experienced a price downturn during the collapse period, in the first episode the price declined by 19.82%, in the second episode it declined by 16.72% and in the third episode it declined by 28%. I also find a very recent bubble, from June 2021 to December 2021, soon after the Covid-19 crisis.

With the second empirical application I am more interested in answering the second question. To do so I firstly employ the BSADF (Backward Supremum Augmented Dickey Fuller) test from Phillips, Shi, and Yu (2015). Then, I run four different logistic regressions, where on the left hand side I have a *Crash* dummy, equal to 1 when prices decline by 40% in 2-years and equal to zero otherwise. On the right hand side I have a *Bubble* dummy, for the first two regressions I consider the BSADF statistic, for the last two regressions I consider two-years raw returns that increased more than 100% in two years, first equally-weighted returns and then value-weighted returns, as in Greenwood, Shleifer, and You (2019). I also consider some control variables, such as volatility, turnover, firm age and percentage of issuers, all variables that are proved to move abnormally during explosive episodes. From these regressions I find that the coefficients on the *Bubble* dummy are always significant. I run a second logistic regression as a comparison between the different methodologies. Also in this regression all coefficients are significant, none of the methodologies are able to make the others unimportant. However, the highest coefficient is the one from the dummy constructed with the BSADF statistic, and the difference between this coefficient and the others is significant. From these results I can conclude that the econometric test is really able to heighten the probability of a crash, thus, it is able to detect rational bubbles better than a naive technique.

Bibliography

Abreu, Dilip and Markus Brunnermeier (2003). "Bubbles And Crashes". In: *Econometrica* 71.1, pp. 173–204. URL: https://www.jstor.org/stable/3082044.

Astill, Sam et al. (2021). "CUSUM-Based Monitoring for Explosive Episodes in Financial Data in the Presence of Time-Varying Volatility". In: *Journal of Financial Econometrics* 21.1, pp. 187–221. URL: https://academic.oup.com/jfec/article/21/1/187/6268988.

Barberis, Nicholas et al. (2018). "Extrapolation And Bubbles". In: *Journal of Financial Economics* 129, pp. 203–227. URL: https://doi.org/10.1016/j.jfineco.2018.04.007.

Blanchard, Oliver and Mark Watson (1982). "Bubbles, Rational Expectations and Financial Markets". In: 0945. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=226909.

Bondt, Warner F. de and Richard Thaler (1985). "Does the Stock Market Overreact?" In: *The Journal of Finance* 40.3, pp. 793–805. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.1985.tb05004.x.

Campbell, John Y. and Robert J. Shiller (1987). "Cointegration and tests of present value models". In: *Journal of Political Economy* 95, pp. 1062–1088. URL: https://www.journals.uchicago.edu/doi/pdf/10.1086/261502.

Chu, Chia-Shang J., Maxwell Stinchcombe, and Halbert White (1996). "Monitoring Structural Changes". In: *Econometrica* 64.5, pp. 1045–1065. URL: https://www.jstor.org/stable/2171955.

Cuñado, Juncal, Luis Gil-Alana, and Fernando P. de Grazia (2005). "A Test for Rational Bubbles in the NASDAQ Stock Index: A Fractionally Integrated Approach". In: *Journal* *of Banking & Finance* 29, pp. 2633–2654. URL: https://www.sciencedirect.com/science/article/pii/S154461232100369X.

Diba, Behzad and Herschel I. Grossman (1988). "Explosive Rational Bubbles in Stock Prices?" In: *The American Economic Review* 78.3, pp. 520–530. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=309591.

Evans, George W. (1991). "Pitfalls in Testing for Explosive Bubbles in Asset Prices". In: *The American Economic Review* 81.4, pp. 922–930. URL: https://www.jstor.org/stable/2006651.

Fama, Eugene (1976). *Foundations of Finance*. Basic Books Inc.

— (2014). "Two Pillars of Asset Pricing". In: *The American Economic Review* 104.6, pp. 1467–1485. URL: https://www.jstor.org/stable/42920856.

Froot, Kenneth and Maurice Obstfeld (1991). "Intrinsic Bubbles: The Case of Stock Prices". In: *The American Economic Review* 81.5, pp. 1189–1214. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=304858.

Giglio, Stefano, Matteo Maggiori, and Johannes Stroebel (2016). "No Bubble Condition: Model-Free Tests in Housing Markets". In: *Econometrica* 84.3, pp. 1047–1091. URL: https://www.jstor.org/stable/43866458.

Górecki, Tomasz, Lajos Horváth, and Piotr Kokoszka (2018). "Change Point Detection in Heteroscedastic Time Series". In: *Econometrics and Statistics* 7, pp. 63–88. URL: https://www.sciencedirect.com/science/article/pii/S2452306217300667.

Greenspan, Alan (2007). *The Age of Turbulence: Adventures in a New World*. Penguin Press.

Greenwood, Robin, Andrei Shleifer, and Yang You (2019). "Bubbles for Fama". In: *Journal of Financial Economics* 131.1, pp. 20–43. URL: https://www.sciencedirect.com/science/article/pii/S0304405X1830254X?via%3Dihub.

Hall, Stephen G., Zacharias Psaradakis, and Martin Sola (1999). "Detecting Periodically Collapsing Bubbles: A Markov-Switching Unit Root Test". In: *Journal of Applied Econometrics* 14, pp. 143–154. URL: https://www.jstor.org/stable/223273.

Harvey, David, Stephen J. Leybourne, and Robert Sollis (2017). "Improving the Accuracy of Asset Price Bubble Start and End Date Estimators". In: *Journal of Empirical Finance* 40, pp. 121–138. URL: https://www.sciencedirect.com/science/article/pii/S0927539816301219.

Harvey, David, Stephen J. Leybourne, Robert Sollis, and Robert Taylor (2016). "Tests for Explosive Financial Bubbles in the Presence of Non-Stationary Volatility". In: *Journal of Empirical Finance* 38, pp. 548–574. URL: https://www.sciencedirect.com/science/article/pii/S0927539815000961.

Harvey, David, Stephen J. Leybourne, and Emily Whitehouse (2020). "Date-Stamping Multiple Bubble Regimes". In: *Journal of Empirical Finance* 58, pp. 226–246. URL: https://www.sciencedirect.com/science/article/pii/S0927539820300347.

Homm, Ulrich and Jorg Breitung (2012). "Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods". In: *Journal of Financial Econometrics* 10.1, pp. 198–231. URL: https://academic.oup.com/jfec/article/10/1/198/757787.

Horváth, Lajos, Hemei Li, and Zhenya Liu (2022). "How to identify the different phases of stock market bubbles statistically?" In: *Finance Research Letters* 46, p. 102366. URL: https://www.sciencedirect.com/science/article/pii/S154461232100369X.

Horváth, Lajos, Zhenya Liu, et al. (2020). "Sequential Monitoring for Changes from Stationarity to Mild Non-Stationarity". In: *Journal of Econometrics* 215, pp. 209–238. URL: https://www.sciencedirect.com/science/article/pii/S0304407619301964.

Koustas, Zisimos and Apostolos Serletis (2005). "Rational Bubbles or Persistent Deviations From Market Fundamentals?" In: *Journal of Banking & Finance* 29, pp. 2523–2539. URL: https://www.sciencedirect.com/science/article/pii/S0378426604002419.

Lee, Ji Hyung and Peter C. Phillips (2016). "Asset pricing with financial bubble risk". In: *Journal of Empirical Finance* 38, pp. 590–622. URL: https://www.sciencedirect.com/science/article/pii/S0927539815001206.

Long, Bradford de et al. (1990). "Positive Feedback Investment Strategies and Destabilizing Rational Speculation". In: *Journal of Finance* 45.2, pp. 379–395. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1990.tb03695.x.

Miller, Edward M. (1977). "Risk, Uncertainty and Divergence of Opinion". In: *The Journal of Finance* 32.4, pp. 1151–1168. URL: https://www.jstor.org/stable/2326520.

Phillips, Peter C., Shuping Shi, and Jun Yu (2015). "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P500". In: *International Economic Review* 56.4, pp. 1043–1078. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/iere.12132.

Phillips, Peter C., Yangru Wu, and Jun Yu (2011). "Explosive Behavior in the 1990s NASDAQ: When Did Exuberance Escalate Asset Values?" In: *International Economic Review* 52.1, pp. 201–226. URL: https://www.jstor.org/stable/23016628.

Sheinkman, Jose and Wei Xiong (2003). "Overconfidence and Speculative Bubbles". In: *Journal of Political Economy* 111, pp. 1183–1219. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=298865.

Shiller, Robert J. (1981). "Do Stock Price Move Too Much to Be Justified by Subsequent Changes in Dividends?" In: *The American Economic Review* 71.3, pp. 421–422. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=262076.

— (2003). *Irrational Exuberance*. Princeton University Press.

Sibbersten, Phillipp and Robinson Kruse (2009). "Testing for a Break in Persistence Under Long-Range Dependencies". In: *Journal of Time Series Analysis* 30.3, pp. 263–285. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.2009.00611.x.

Thaler, Richard (2016). "Behavioral Economics: Past, Present and Future". In: *American Economic Review* 106.7, pp. 1577–1600. URL: http://dx.doi.org/10.1257/aer.106.7.1577.

Tirole, Jean (1985). "Asset Bubbles and Overlapping Generations". In: *Econometrica* 53.6, pp. 1499–1528. URL: https://www.jstor.org/stable/1913232.

West, Kenneth D. (1987). "A Specification Test for Speculative Bubbles". In: *The Quarterly Journal of Economics* 102.3, pp. 553–580. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=307112.