Bootstrap Unit Root Testing for Explosive Behaviour using Covariates

January 18, 2019

Ioannis Korkos, Sam Astill, Neil Kellard & A.M Robert Taylor
Essex Business School
University of Essex

Abstract

This paper proposes a right-tail bootstrap implementation of the covariate Augmented Dickey-Fuller (CADF) unit root test of Hansen (1995), motivated by the work of Chang, Siddes and Song (2017) who apply a left-tail bootstrap covariate Augmented Dickey-Fuller (BCADF) test when testing the null of a unit root against the alternative of stationarity. We apply the right-tail bootstrap BCADF test in a recursive manner as in Phillips, Shi and Yu (2015a) [PSY] to maximise power against an end-of-sample explosive alternative. We focus on testing the null hypothesis of a unit root against the alternative of an end-of-sample explosive episode in an attempt to examine the impact of including relevant covariates on the size and power properties of the tests of PSY. We provide evidence that the inclusion of relevant covariates in the conventional Augmented Dickey Fuller test regression leads to improved size control, while offering significant power gains when an end-of-sample explosive episode is present. An empirical application of the proposed methodology to the S&P 500 is conducted, utilising the Moody’s Seasoned Aaa Corporate Bond Yield, the VIX Volatility Index and the CDS bid-ask spread as covariates.

Keywords: Rational bubbles, Explosive behaviour, Right-tailed unit root testing, Covariates

JEL Classification C12;C13;C15;C22;C32;C63;G01
1 Introduction

Asset price bubbles have recently attracted significant interest in the finance literature, mainly due to the collapse of past bubbles having a significant impact on the real economy. According to economic theory, an asset bubble can be defined as a prolonged period of substantial price deviations from a fundamental value (see inter alia Blanchard and Watson 1982, Campbell, Lo and MacKinlay 1997 and Homm and Breitung 2012). Price misalignments lead investors to pay a higher price (than justified by fundamentals/fair price) for an asset, expecting to sell the asset at an even greater price in the future and generate a profit. In a bubble regime, there is a high volume of trading, in contrast to normal market conditions (Ofek and Richardson 2003). Subsequently, positive feedback mechanisms result in further inflation of the equilibrium price.

The identification of asset price bubbles is clearly of great interest for both theorists and empirical researchers. Recently many econometric techniques have been developed for bubble detection in the context of time series analysis. Cointegration analysis, for instance, has been applied as one of the main testing approaches assessing price deviations from equilibrium (see inter alia Campbell and Shiller 1987, 1988a and 1988b). In a rational bubble regime, the equilibrium condition between the asset price and its market fundamental is violated. Therefore, non-stationary deviations from the general equilibrium in the long term provide evidence for the existence of a bubble. Non-stationary behaviour is also examined in the logarithmic transformation of the price dividend ratio through unit root testing; if the dividend yield is integrated of order one, then this could be considered as a strong evidence of a rational bubble (Campbell and Shiller 1987). The rational expectations theory perceives (rational) bubbles as anticipated phenomena where their expected value of next period depends on the compounded value of the bubble at time zero. In other words, the value of a bubble today equals the discounted value of future bubble episodes.

Historical episodes of price bubbles have been well documented in literature, see inter alia Galbraith (1997), Kindleberger and Aliber (2005), and Sornette (2003b). The earliest known bubble episode in the financial history is known as Tulip Mania and took place in the Netherlands during the 17th century. In the 18th century, the first significant market crash in the British stock market occurred, driven by what is known as the South Sea Bubble; a result of excessive
speculation by the South Sea Company that had monopolistic rights to shipping and trading activities with South America (Sornette 2003b). A similar bubble episode occurred in France over the exact same period. In this case, banks excessively issued bank notes which were not equivalent to their gold and silver reserves. In 1720, the market crashed and this event is known historically as the Mississippi Bubble (Kindleberger and Aliber 2005).

Moving into the twentieth century and during the Roaring 20s, the U.S. economy was thriving mainly as a result of new technological innovations and industrialisation. In 1929, the Federal Reserve of the U.S. attempted to calm the market through implementing tight monetary policy. Panic resulted in a massive liquidation of shares, margin investors went bankrupt and major banks were driven into default since they had invested depositors' money, leading to a recession that lasted almost four years, known as Great Depression. The market crash of October 1987, referred to as Black Monday, came after a period of euphoria in capital markets resulting from low interest rates, mergers and acquisitions, hostile takeovers and leverage buyouts. The Federal Reserve of the U.S. increased interest rates and made access to funding extremely unaffordable. Computer trading (sell orders after losses), derivative securities, liquidity problems, huge trade and budget deficits and overvaluation combined with austere monetary policy led two of the largest capitalisation indexes in the U.S. (S&P500 and Dow Jones) to a decline of more than 20% of their value (Sornette 2003b).

The term "dot-com" bubble or "tech" bubble is widely used to describe the last few years of the 1990s, a decade when the stock prices of internet firms escalated to extremely high levels. In a short period of time, hundreds of thousands of small-medium sized firms raised capital through IPOs despite cash flow issues, taking advantage of the enthusiasm of capital markets participants to fund internet firms. During the early 2000s, investors realised that the price of many internet stocks were well above their fundamental value, with the price of these stocks subsequently crashing, resulting in a mild recession for the U.S economy, despite the effort of the Federal Reserve of the U.S. to decrease interest rates (Ofek and Richardson 2003, and Kindleberger and Aliber 2005).

More recently, the global financial crisis of 2008-2009 has been triggered by the subprime mortgage market crash (Akerlof and Shiller 2009). Due to the deep integration of the capital and money markets nowadays, the exuberance transmitted from the financial markets (com-
Modities, exchange rates, stock exchanges etc.) to the real economy. At the end of 2006, the mortgage backed security market reached extremely high levels of volume and at the same time, the majority of debtholders were unable to payback their debt leading to delinquencies and foreclosures. Investors lost trust, liquidity sank and the financial system, especially investment and commercial banks, collapsed. The contagion propagated to commodities, real exchange, fixed income and oil markets as investors selectively transferred their assets to other investment. This global financial crisis drove the majority of the developed countries into recession (Phillips & Yu, 2011). Akerlof and Shiller (2009) attribute the recent global financial crash to the breakdown of the financial system and especially of structured financial products, the high leverage and capital loss of the financial institutions and the already-agreed credit lines between the banks and their clients.

With the need for early detection of asset price bubbles apparent, this article intends to investigate whether the power of right-tail Dickey-Fuller unit root test procedures can be enhanced by the inclusion of relevant covariates if choosing to test for a bubble in a univariate framework examining a variable in isolation can be rather costly in terms of power since ignoring any correlation with other time series could possibly weaken the explanatory power of standard unit root tests leading to significant power losses (Hansen, 1995). Our proposed Covariate Augmented Dickey Fuller unit root test is applied in a backward supremum sequence as suggested by PSY to detect explosive episodes that occur at the end of the sample. Applying sub-sample techniques can, however, lead to imprecise estimation of the nuisance parameter introducing additional variability and causing severe size distortions (Chang, Sickles and Song 2017). To deal with the nuisance dependency problem we apply a parametric bootstrap procedure, ensuring the asymptotic validity of the critical values drawn from the bootstrap distribution of the test statistics. We concentrate on the case where the explosive episode takes place at the end of the sample to date-stamp bubbles in real time. The simulations show that the proposed bootstrap tests offer impressive size and power performance in finite samples. In particular, the bootstrap tests appear to be less size distorted compared to the non-bootstrap conventional unit root tests and the inclusion of covariates in the standard Augmented Dickey Fuller regression model offers significant power gains.

The remainder of this paper is organised as follows. In Section 2 we outline the explosive
financial bubble model and outline our proposed tests. In Section 3 the finite and sample size and power properties of our proposed tests are examined using Monte Carlo simulations. Section 4 presents an empirical application of our proposed tests to the S&P500 price dividend series. Section 5 concludes.

In what follows $\overset{p}{\rightarrow}$ denotes convergence in probability, $\overset{d}{\rightarrow}$ denotes convergence in distribution and $\lfloor . \rfloor$ denotes the integer part of its argument. We also denote $|.|$ as the Euclidean norm.

For a vector $z = z_i$, $|z|^2 = \sum_i z_i^2$ and for a matrix $A = (a_{i,j})$ $|A|^2 = \sum_{i,j} a_{i,j}^2$.

2 Model and Proposed Tests

2.1 The Model

Consider a time series process $\{y_t\}$ that consists of a purely deterministic component and a stochastic component:

$$y_t = d_t + S_t, \quad t = 1, ..., T$$  \hspace{1cm} (2.1)

where $d_t$ is the deterministic component and can be either equal to 0 (neither constant, nor trend), $\mu$ (constant but not trend) or $\mu + \theta t$ (constant and trend). The initial condition $y_0$ is stochastically bounded and therefore does not affect our subsequent analysis. The stochastic component is given by the following equation:

$$\Delta S_t = \delta_t S_{t-1} + u_t$$  \hspace{1cm} (2.2)

where $\delta_t$ is a time varying autoregressive parameter. The innovation sequence, $\{u_t\}$ is given by

$$\alpha(L)u_t = b(L)'\Delta x_t + \varepsilon_t$$  \hspace{1cm} (2.3)

where $\Delta x_t$ is an $m$-vector of stationary covariates, $\alpha(L)$ is a lag operator polynomial of order $p$: $\alpha(z) = 1 - \sum_{k=1}^p \alpha_k z^k$ and $b(k) = \sum_{k=-r}^q \beta_k z^k$ is a polynomial allowing for, but not requiring, both leads and lags of $\Delta x_t$ to enter the data generating process.

Under the null hypothesis of no explosivity, $H_0 : \delta_t = 0$ for all $t = 1, ..., T$ and, therefore,
\{y_t\} follows a unit root process for the entire sample. The alternative hypothesis is given by

\[ H_1 : \delta_t = 0 \text{ for } t = 2, \ldots, [r'T] \text{ and } \delta_t > 0 \text{ for } t = [r'T] + 1, \ldots, T. \]

Under the alternative of an end-of-sample explosive episode \{y_t\} follows a unit root process for the first \([r'T]\) observations and is then subject to explosive behaviour over the remaining observations.

Combining (2.2) and (2.3) we could then estimate the following regression

\[
\Delta y_t = d_t^* + \delta_t y_{t-1} + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + \sum_{k=-r}^{q} \beta_k' \Delta x_{t-k} + e_t =: CADF(p, r, q) \tag{2.4}
\]

where

\[
d_t^* = \begin{cases} 
0 & \text{if } d_t = 0 \\
-\delta \mu & \text{if } d_t = \mu \\
a(1)\theta - \delta \mu - \delta \theta & \text{if } d_t = \mu + \theta t
\end{cases}
\tag{2.5}
\]

Following Chang, Sickles and Song (2017) we assume that the stationary covariates \(\Delta x_t\) are generated by an AR(1) process given by

\[
\Psi(L) \Delta x_{t+r+1} = \eta_t \tag{2.6}
\]

where \(\Psi(z) = I_m - \sum_{k=1}^{l} \Psi_k \bar{z}^k\).

We also make the following assumptions on the innovation sequence \(\xi_t = (\epsilon_t, \eta_t')\) that defines the correlation between the stationary covariates \(\Delta x_t\) and the series of interest \(y_t\).

**Assumption 2.1.** (a) Let \(\{\xi_t\}\) be a martingale difference sequence such that \(E(\xi_t \xi_t') = \Omega\) and \((1/T) \sum_{t=1}^{T} \xi_t \xi_t' \overset{P}{\rightarrow} \Omega\) with \(\Omega > 0\) and \(E|\xi_t|^\gamma < K\) for some \(\gamma \geq 4\), where \(K\) is some constant depending only upon \(\gamma\).

(b) \(\alpha(z), \text{det}(\Phi(z)) \neq 0\) for all \(|z| \leq 1\)

**Remark 2.1.** As noted by Chang, Sickles and Song (2017) Assumption 2.1(a) allows for conditional heteroskedasticity, including GARCH behaviour, in all equations in the system including the covariates. By definition \((\epsilon_t)\) is uncorrelated with \((\eta_{t+k})\) for \(k \geq 1\). This condition implies that \((\epsilon_t)\) is uncorrelated with the lagged differences of the dependent variable \((\Delta y_{t-1}, \ldots, \Delta y_{t-p})\)
and the leads and lags of the covariates \((\Delta x_{t+r}, \ldots, \Delta x_{t-q})\) which is essential for (2.4) to be estimable by ordinary least squares.

Let \(z_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-p}, \Delta x_{t+r}, \ldots, \Delta x_{t-q})'\)

**Assumption 2.2.** \(\sigma^2_u > 0\) and \(E(z_t z_t') > 0\)

**Remark 2.2.** Assumption 2.2 ensures that the series of interest, \(\{y_t\}\), follows a unit root process under the null hypothesis of \(\alpha = 0\). \(E(z_t z_t') > 0\) ensures that the stationary regressors in (2.4) are asymptotically linearly independent, which is required along with Assumption 2.1(a) to ensure consistence of the least squares estimate of \(\alpha\).

### 3 Existing Tests

PSY(2015a) propose a univariate approach to testing for end-of-sample bubbles utilising the standard (non-covariate augmented) ADF regression given by

\[
\Delta y_t = \mu + \alpha y_{t-1} + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + e_t
\]  

where \(\mu\) is the intercept and \(p\) is the number of lags of the dependent variable \(\Delta y_t\).

An Augmented Dickey Fuller unit root test applied to the full sample size can be denoted as \(ADF_{10}(p)\). Full sample tests for explosive behaviour can be shown, however, to have very poor power to detect short lived explosive episode in a series that otherwise exhibit explosive behaviour. As such, PSY (2015a) consider test statistics that are functions of a sequence of ADF statistics applied to subsample of the data. Specifically, if we denote an ADF test procedure performed on the subsample \(t = r_0, \ldots, T\) as \(ADF(p)_{r_0}^{r}\), then PSY(2015a) propose the following test statistic to test for an end-of-sample explosive episode.

\[
SADF := \sup_{r \in [r_0, 1]} \{ADF(p)_{r_0}^{r}\}
\]  

where \(ADF(p)_{r_0}^{r}\) denotes the standard Augmented Dickey-Fuller \((ADF(p))\) test calculated using the observations 1, ..., \(rT\). This recursive regression technique constitutes a powerful tool of detecting periodically collapsing explosive behaviour and can also be utilised for confidence interval
construction and date-stamping of the origination and termination dates of the bubble episode. One drawback of this forward recursive unit root test is that it is not consistent on detecting multiple episodes of explosiveness. In addition, the SADF test is by construction less powerful on detecting explosive episodes at the end of the series since each sub-sample contains less and less explosive observations.

For that reason, PSY(2015a) propose the following test statistic to test for an end-of-sample explosive episode. If we denote an ADF test procedure performed on the subsample $t = r_1T, ..., r_2T$ as $ADF(p)_{r_1}^{r_2}$, then

$$BSADF := \sup_{r_1 \in [0,1-r_0]} \{ADF(p)_{r_1}^{1}\},$$

(3.3)

where $ADF(p)_{r_1}^{1}$ denotes the standard Augmented Dickey-Fuller test calculated using the observations $r_1T, ..., T$. The BSADF test, proposed by Phillips, Shi and Yu (2015a) is, thus, the supremum of right-tailed ADF statistics computed over all possible start dates (subject to a minimum sample size $r_0T$). The BSADF test is designed to detect end-of-sample explosive episodes and is used by PSY to date stamp past explosive episodes. It is particularly powerful when the bubble episode occurs at the end of the series since the BSADF test will be constructed in such a way that each subsample will contain a number of explosive observations. Critical values for the BSADF test can be found in PSY (2015a).

As an extension of the BSADF test, PSY(2015a) suggest the GSADF test that is a generalised version of the BSADF test computed over all possible start and end dates (subject to a minimum sample size). The GSADF test of PSY takes the form

$$GSADF := \sup_{r_1 \in [0,R_1]} \sup_{r_2 \in [r_0,1]} \{ADF(p)_{r_1}^{r_2}\}$$

(3.4)

where $ADF(p)_{r_1}^{r_2}$ denotes the standard Augmented Dickey-Fuller test calculated using the observations $r_1T, ..., r_2T$. The GSADF test is a double-recursive unit root test, designed to test for the presence of one or more explosive episodes in a financial time series; these are allowed to occur anywhere in the sample. As such, a rejection when using the GSADF test can only infer that a sample contains at least one explosive episode, but not where the explosive episode occurs.
PSY (2015a) find evidence of explosive behaviour in the S&P500 price-dividend ratio when applying the GSADF test. They then calculate the BSADF test for samples ending in each time period in order to date-stamp these events.

3.1 Proposed Tests

A potential drawback of the BSADF test is that a univariate process is assumed. Given the complex relationships across multiple asset prices a multivariate approach could, potentially, be of much greater use. In the context of unit root testing Hansen (1995) first proposed adopting a multivariate approach when testing the null of a unit root against the alternative of stationarity, and found doing so could lead to substantial power gains relative to a univariate approach.

If a practitioner was interested in testing the null hypothesis of a unit root in \( \{ y_t \} \) against the alternative of explosivity for the entire sample (i.e. \( \delta_t = \delta > 0 \) for all \( t \)) then one could simply estimate regression (2.4) over the entire sample and perform a right tailed \( t \)-test of the null of \( \delta = 0 \) using critical values from Hansen (1995). Given the power improvements offered by the BSADF test relative to the full sample ADF test of PSY(2015a), however, we consider instead utilising the following test statistic:

\[
CBSADF := \sup_{r \in [0, 1 - r_0]} \{ CADF(p, r, q) \},
\]

that is the supremum of covariate augmented Dickey-Fuller test statistics computed over all possible start dates (subject to a minimum sample size \( r_0 T \)). Such an approach will be able to make use of the interdependencies between the series \( \{ y_t \} \) and any relevant covariates.

4 Limit Theory

We now derive the limiting null distribution of the CBSADF test statistic. The proof of the Theorem that follows requires the results of the following Lemma.

Lemma 4.1. Defining
\[ \Omega = \begin{pmatrix} \sigma^2 & \sigma_{\varepsilon u} \\ \sigma_{u\varepsilon} & \sigma^2_u \end{pmatrix} \quad (4.1) \]

and

\[ \varrho^2 = \frac{\sigma_{u\varepsilon}}{\sigma^2_u \sigma^2_{\varepsilon}} \quad (4.2) \]

Let data be generated according to (2.1) - (2.3) and additionally let Assumptions 1-2 hold. Then under the null hypothesis of no explosivity we have

\[ T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t \Rightarrow \sigma_u W_1(r) \quad (4.3) \]

\[ T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \Rightarrow \sigma_\varepsilon \left[ \rho W_1(r) + (1 - \varrho^2)^{1/2} W_2(r) \right] \]

\[ \equiv \sigma_\varepsilon W^*(r) \quad (4.4) \]

where \( W_1(r) \) and \( W_2(r) \) are independent standard Weiner processes.

**Remark 4.3.** The proof of Lemma 4.1 follows directly from Hansen (1995).

We are now in a position to outline the limiting null distribution of the CBSADF test statistic.

**Theorem 4.1.** Let data be generated according to (2.1) - (2.3) and additionally let Assumptions 1-2 hold. Then under the null hypothesis of no explosivity we have

\[ \text{CBSADF} \xrightarrow{d} \sup_{r \in [0,1-r_0]} \left\{ \frac{r_w \int_r^1 W_1 dW^* - \left[ \left( \int_r^1 W_1(s) ds \right) \left( W^*(1) - W^*(r) \right) \right]}{r_w^{1/2} \left( r_w \int_r^1 W_1(s)^2 ds - \left[ \int_r^1 W_1 ds \right]^2 \right)^{1/2}} \right\} \quad (4.6) \]

The proof of this result can be found in the Appendix.
The asymptotic critical values of the CBSADF test is a function of $\varrho^2$. As in Hansen (1995) the appropriate critical values are selected using a consistent nonparametric estimator of $\varrho^2$ given by

$$\hat{\varrho}^2 = \frac{\hat{\sigma}_{uc}^2}{\hat{\sigma}_u^2}$$

(5.1)

where

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_u^2 & \hat{\sigma}_{uc} \\ \hat{\sigma}_{uc} & \hat{\sigma}_e^2 \end{bmatrix} = \sum_{k=-M}^{M} w(k/M) T^{-1} \sum_t \hat{\gamma}_{t-k} \hat{\gamma}'_{t}$$

(5.2)

and $\hat{\gamma}_{t} = (\hat{u}_t, \hat{\epsilon}_t)'$ are least squares estimators of $\gamma_{t} = (u_t, \epsilon_t)'$, using the recommendation for $M$ and $w$ given by Hansen (1995). The function $w$ can be any kernel function that produces positive semidefinite covariance matrices (e.g. Parzen or Bartlett kernel) and $M$ is a bandwidth estimator.

### 6 Bootstrap Unit Root Tests with Covariates

In section 4 we presented the asymptotic distribution of the CBSADF test under the null hypothesis of a unit root. This asymptotic distribution of the test depends on the nuisance parameter $\varrho$ which is the long run correlation coefficient between the equation error and the covariate defined in equation (5.1). As mentioned earlier, $\varrho^2$ measures the relative contribution of the covariate $\Delta x_t$ to the error term $u_t$ and can take values between 0 (when the covariates fully explain the variability of the error term) and 1 (when the covariates have no explanatory power towards the error term). We expect that the lower the long run correlation between $\epsilon_t$ and $u_t$, the greater the power gains from the inclusion of the covariate into the standard ADF regression.

In order to estimate $\varrho^2$, we initially perform the proposed CBSADF test as defined in subsection 3.1 and then estimate the long run covariance matrix across all the recursive subsample sequences. Precisely, following Chang, Sickles and Song (2017) we used an optimisation process that minimises the $\varrho^2$ for the optimal combination of $p$ lags of the dependent variable $\Delta y_t$ and $q_1$ lags and $q_2$ leads of the covariate $\Delta x_t$. The lags of the dependent variable $\Delta y_t$, $p$
are chosen by using the BIC whereas the $q_1$ lags and $q_2$ leads of the covariate $\Delta x_t$ are chosen in such a way that minimise $\varrho^2$.

The methodology described above works quite well in full sample testing processes but leads to poor size control in a recursive sub-sample framework, mainly due to imprecise estimation of the nuisance parameter $\varrho^2$.

As we consider the bootstrap for the covariates ADF unit root tests, we will show that a bootstrap implementation of the CBSADF test can better control size than a test based on asymptotic critical values and offer significant power gains compared to the tests of PSY. In what follows we use the notation $^*$ to denote the bootstrap samples and $P^*$ and $E^*$ to denote the probability and expectation conditional to the original sample.

We need to generate the autoregressive time series of interest $y_t$ and a stationary covariate $\Delta x_t$ in order to construct the bootstrap covariate Augmented Dickey-Fuller test (hereafter BCADF). To do so, we formulate the fitted residuals assuming that under the null hypothesis the unit root restriction holds. Therefore, we let $u_t = \Delta y_t$ and by estimating the following regression using OLS

$$u_t = \sum_{k=1}^{p} \tilde{a}_k u_{t-k} + \sum_{k=-r}^{q} \tilde{\beta}_k^r \Delta x_{t-k} + \tilde{\varepsilon}_t$$

we get the fitted residual $\tilde{\varepsilon}_t$. Following Chang, Sickles and Song (2017) we base our bootstrap sampling on equation (3.1) imposing the unit root restriction $\alpha = 0$. To continue, we write the stationary covariate $\Delta x_t$ as an autoregression of order $q_2, q_1$ and we utilise the Yule-Walker method to estimate the fitted residual $\tilde{\eta}_t$:

$$\Delta x_{t+r+1} = \tilde{\Phi}_{1,n} \Delta x_{t+r} + ... + \tilde{\Phi}_{l,n} \Delta x_{t+r-l+1} + \tilde{\eta}_t$$

Then, we define $\tilde{\xi}_t = (\tilde{\varepsilon}_t, \tilde{\eta}_t)$ as a fitted residual matrix where $\tilde{\varepsilon}_t$ and $\tilde{\eta}_t$ the fitted residuals obtained from equations (6.1) and (6.3). Next, we generate bootstrap samples $(\tilde{\xi}_t^*)$ by resampling $\tilde{\xi}_t$ to obtain independent and identically distributed (iid) samples of $\tilde{\xi}_t$ from the empirical distribution of

$$\left(\tilde{\xi}_t - \frac{1}{n} \sum_{t=1}^{n} \tilde{\xi}_t\right)^n_{t-1}.$$  

\footnote{Brockwell & Davis (1991) suggest to utilise the Yule-Walker method for estimating ARMA($p,q$) models to obtain good estimates in small samples and/or when $q = 0$ to ensure stationarity.}
The bootstrapped residual matrix $\xi_t^*$ satisfies the following properties: $E^* \xi_t^* = 0$ and $E^* \xi_t^* \xi_t^{*\prime} = \hat{\Sigma}$ where $\hat{\Sigma} = (1/n) \sum_{t=1}^{n} \xi_t^{*\prime} \xi_t^*$. Now, we can recursively construct the bootstrap samples of the stationary covariate ($\Delta x_t^*$) by using the bootstrapped residuals $\eta_t^*$. Equation (6.3) can now be written as

$$\Delta x_{t+r+1}^* = \Phi_{1,n} \Delta x_{t+r}^* + \ldots + \Phi_{l,n} \Delta x_{t+r+l-1} + \hat{\eta}_{t}^*$$

(6.3)

with appropriately chosen $l$-initial values of $\Delta x_t^*$. We make the assumption that the initial value of the stationary covariate is stochastically bounded and we set is equal to zero in accordance to Chang, Sickles and Song (2017).

To continue, we obtain ($\Delta x_{t+r}^*, \ldots, \Delta x_{t-q}^*$) from the sequence $\Delta x_t^*$, we can generate the bootstrap samples of the innovation term, $v_t^*$

$$v_t^* = \sum_{k=-r}^{q} \hat{\beta}_k \Delta x_{t-k}^* + \varepsilon_{t}^*$$

(6.4)

using the OLS estimates $\hat{\beta}_k$, $-r \leq k \leq q$ from the fitted regression (6.1). Then, we recursively generate the bootstrap samples of the error term $u_t^*$ from $v_t^*$

$$u_t^* = \hat{\alpha}_1 u_{t-1}^* + \ldots + \hat{\alpha}_p u_{t-p}^* + v_t^*$$

(6.5)

with appropriately chosen $p$-initial values of $u_t^*$, and where $\hat{\alpha}_k$, $1 \leq k \leq p$ are the estimates for $\alpha_k$’s from the fitted regression (6.1). Finally, under the unit root constraint (null hypothesis) we can generate $y_t^*$ from $u_t^*$

$$y_t^* = y_{t-1}^* + u_t^* = y_0^* + \sum_{k=1}^{t} u_k^*$$

(6.6)

we set the initial value $y_0^*$ is equal to 100. The choice of the initial value may affect the finite sample performance of the bootstrap test but as it is stochastically bounded, the effect becomes asymptotically negligible. The bootstrap covariate Augmented Dickey-Fuller unit root test (hereafter $CADF^*$) is being derived by equation (2.4) and can be written as
\[ \Delta y_t^* = \alpha y_{t-1}^* + \sum_{k=-r}^{p} \alpha_k \Delta y_t^* + \sum_{k=-r}^{q} \beta_k \Delta x_{t-k}^* + \varepsilon_t^* \quad (6.7) \]

Therefore, to overcome with the nuisance parameter dependency and the large size distortions of the recursive CADF tests presented above, we propose utilising the \( CADF^*(p, q_1, q_2) \) test in place of the standard \( CADF(p, q_1, q_2) \) test in order to ensure the asymptotic validity of the critical values from the bootstrap distribution of the test. Our proposed tests are now given by

\[
\begin{align*}
CSADF^* &:= \sup_{r \in [r_0, 1]} \{ CADF^*(p, q_1, q_2)_r \} \\
CBSADF^* &:= \sup_{r \in [0, 1-r_0]} \{ CADF^*(p, q_1, q_2)_{1-r} \} \\
CGSADF^* &:= \sup_{r_1, r_2 \in [r_0, 1]} \{ CADF^*(p, q_1, q_2)_{r_1} \}
\end{align*}
\quad (6.8, 6.9, 6.10)
\]

where \( CSADF^* \) stands for the Bootstrap Covariate Supremum Augmented Dickey-Fuller test, \( CBSADF^* \) stands for the Bootstrap Covariate Backward Supremum Augmented Dickey-Fuller test and \( CGSADF^* \) stands for the Bootstrap Generalised Covariate Supremum Augmented Dickey-Fuller test.

The benefit now is that the critical values of the recursive bootstrap CADF tests, do not depend on the estimated \( \hat{\beta}^2 \) but bootstrapped critical values are used instead. The bootstrap CADF test deals with the nuisance dependency problem and ensures the asymptotic validity of the critical values of the bootstrap distribution of the test. Again, for convenience we will focus on the performance of the \( CBSADF^* \) test in the case that the bubble episode occurs at the end of the sample.

7 Simulation Results

7.1 Data Generating Process

In order to examine the empirical size and power properties of the proposed covariate tests in finite samples, we simulate data according to the following DGP

\[ y_t = \phi y_{t-1} + u_t, \quad t = 1, \ldots, 200 \quad (7.1) \]
with

\[ u_t = \alpha_1 u_{t-1} + \nu_t \quad (7.2) \]

\[ \nu_t = \beta \Delta x_t + \epsilon_t, \quad (7.3) \]

\[ \Delta x_{t+1} = \lambda \Delta x_t + \eta_t, \quad (7.4) \]

where

\[ \xi_t = \begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{\epsilon \eta} \\ \sigma_{\epsilon \eta} & 1 \end{bmatrix} \right) \quad (7.5) \]

with \( \phi_t = 1 \) for \( t = 1, \ldots, 200 \) under the null hypothesis of no explosive behaviour and \( \phi_t = 1 \) for \( t = 1, \ldots, 180 \) then \( \phi > 1 \) for \( t = 181, \ldots, 200 \) under the alternative of an end-of-sample explosive episode. Under the alternative hypothesis, \( y_t \) follows a unit root process for the first 180 observations and is then subject to, potentially, explosive behaviour for the final 20 observations.

The CADF regression is given by

\[ \Delta y_t = \mu + \alpha y_{t-1} + \alpha_1 \Delta y_{t-1} + \beta \Delta x_t + \epsilon_t \quad (7.6) \]

where \( \mu \) is the intercept, \( \Delta y_{t-1} \) the lag of the dependent variable and \( \nu_t \) the covariate.

### 7.2 Finite Sample Properties

In this subsection we examine the finite sample performance of the bootstrap BSADF* and CBSADF* sub sample tests proposed in subsection 3.3. Equations (7.3) and (7.4) show that the degree of correlation between the error term \( \epsilon_t \) and the covariate \( \Delta x_t \) depends on two parameter values; the AR coefficient of the covariate \( \lambda \) and the coefficient of the covariate \( \beta \). Following Chang, Sickles and Song (2017), we let these two parameters vary from \(-0.8\) to \(0.8\). We set the coefficient of the lagged difference term \( \alpha_1 = 0.2 \) and the contemporaneous covariance \( \sigma_{\epsilon \eta} = 0.4 \) and we drop the first 100 observations to minimise start-up effects, as in Hansen (1995). The sample size is 100 and 400 observations and the minimum window size of the backward tests is set according to
where $T$ is the sample size and therefore $r_0 = 19$ and $r_0 = 40$ respectively. Finally, we use a fixed lag order of 1 for the dependent variable $\Delta y_t^2$ and we refer to the case where there is one covariate at time $t$, $(\Delta x_t)$ but neither additional lags nor leads of the covariate. We examine the size and power properties of the $BCADF$ and $BCADF^*$ tests under both the null hypothesis of a unit root across all observations of the sample and the alternative hypothesis of explosive behaviour in the final 20 observations. All simulations that follow were conducted in GAUSS 17 using 10,000 Monte Carlo replications and 399 bootstrap resamples replications. All tests are performed at a 5% level of significance.

7.2.1 Empirical Size

To assess the size properties of the tests, data where generated according to the DGP given by equation (7.1), which under the null hypothesis of a unit root becomes:

$$
y_t = y_{t-1} + u_t, \quad t = 1, \ldots, 200
$$

(7.8)

where $u_t \sim iid N(0, 1)$ given by equation (7.2). Table 1 reports the empirical size of the proposed tests presented in subsection 3.3. In Table 2 we increase the sample size to 500 observations and again we drop the first 100 observations to minimise start-up effects.\footnote{As discussed by PSY (2015a) who argue that sequential significance testing for optimal lag selection can lead to significant size distortions in both the PWY and PSY processes and they recommend utilising a fixed small lag order.
}

Both the non-bootstrap $BSADF$ and $CBSADF$ tests show severe size distortions across multiple scenarios, with the size of these tests exceeding the 5% level of significance in all of the scenarios considered. One potential reason for these size distortions, can be the imprecise estimation of the nuisance parameter (denoted above as $\hat{\varphi}^2$). The limit distribution of these tests depend on the nuisance parameter and thus imprecise estimation of $\hat{\varphi}^2$ leads to incorrect critical values being utilised resulting in inevitable size distortions.

On the other hand, the bootstrap $BSADF^*$ and $CBSADF^*$ tests display much better
size control. In particular, the $BSADF^*$ statistic has relatively controlled size across most of the scenarios considered, with some slight over rejection in three of the scenarios; $(\beta, \lambda) = (0.5, 0.8), (-0.5, 0.8), (-0.8, 0.8)$ and modest undersize in the remaining scenarios. In the case where we increase the sample size to 400 then in the first scenario $(\beta, \lambda) = (0.8, 0.8)$ size is equal to its nominal value. The $CBSADF^*$ test, on the other hand, does not over reject the null hypothesis and has controlled size across all scenarios considered, displaying a modest degree of undersize in each instance which fades away as we increase the sample size.

Overall, it can be seen that both the bootstrap $BSADF^*$ and $CBSADF^*$ tests control size to a much greater degree than their respective non-bootstrap ($BSADF$ and $CBADF$) tests at least in finite samples. Arguably, the inclusion of a covariate in our proposed tests does not significantly affect the size properties of the $BSADF$ statistic and the $CBSADF^*$ test substantially corrects the large size distortions of the $CBSADF$ test. In summary, we conclude that the bootstrap $BSADF^*$ and $CBSADF^*$ tests present sufficiently controlled finite sample size.

Table 1: Finite Sample Size, $T = 100$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\lambda$</th>
<th>$BSADF$</th>
<th>$CBSADF$</th>
<th>$BSADF^*$</th>
<th>$CBSADF^*$</th>
<th>$\hat{\rho}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.204</td>
<td>0.143</td>
<td>0.043</td>
<td>0.025</td>
<td>0.368</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>0.220</td>
<td>0.134</td>
<td>0.065</td>
<td>0.023</td>
<td>0.501</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.8</td>
<td>0.197</td>
<td>0.441</td>
<td>0.071</td>
<td>0.039</td>
<td>0.313</td>
</tr>
<tr>
<td>-0.8</td>
<td>0.8</td>
<td>0.215</td>
<td>0.517</td>
<td>0.066</td>
<td>0.039</td>
<td>0.166</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5</td>
<td>0.105</td>
<td>0.090</td>
<td>0.027</td>
<td>0.023</td>
<td>0.522</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.112</td>
<td>0.092</td>
<td>0.038</td>
<td>0.024</td>
<td>0.666</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.5</td>
<td>0.078</td>
<td>0.181</td>
<td>0.025</td>
<td>0.028</td>
<td>0.570</td>
</tr>
<tr>
<td>-0.8</td>
<td>0.5</td>
<td>0.091</td>
<td>0.275</td>
<td>0.034</td>
<td>0.034</td>
<td>0.331</td>
</tr>
</tbody>
</table>
7.2.2 Empirical Power

In order to examine the power performance of our proposed tests, data were generated under the alternative hypothesis of end-of-sample explosive episode according to the following DGP

\[
y_t = \begin{cases} 
y_{t-1} + u_t, & t = 1, \ldots, 180, \\
\phi y_{t-1} + u_t, & t = 181, \ldots, 200 \end{cases}
\]  

(7.9)

where \( \phi \geq 1 \) and \( u_t \sim i.i.d.N(0, 1) \). The series \( \{y_t\} \) follows a unit root process for the first 180 observations and is then subject to (potential) explosive behaviour over the remaining observations.

The finite sample power of the \( BSADF^* \) and \( CBSADF^* \) tests was computed for a grid of 50 values of \( \phi \) from \( \phi = 1.00 \) to \( \phi = 1.02 \) for each of the previous 8 scenarios considered, we do not report power results for the \( BSADF \) and \( CBSADF \) tests due to the severe size distortions these tests were found to exhibit under the null of no explosive episode. For each scenario considered, Figures 1 and 2 report finite sample power curves for the \( BSADF^* \) and \( CBSADF^* \) tests along with the corresponding values of \( \beta \) and \( \lambda \). It can be seen that the power of the two tests increases monotonically with \( \phi \).

In most of the scenarios, the \( CBSADF^* \) displays greater power overall than the \( BSADF^* \) test, although the power of the \( CBSADF^* \) test seems to be quite low for small values of \( \phi \) due to the undersizing exhibited by the \( CBSADF^* \) test in all scenarios. In general, however, the
power of the $CBSADF^*$ test exceeds that of the $BSADF^*$ test as $\phi$ increases.

When $(\beta, \lambda) = (0.8, 0.8), (0.8, 0.5), (0.5, 0.5)$ the $CBSADF^*$ test shows superior power compared to the $BSADF^*$ statistic, with the power differential between the two tests reaching almost 30% for some values of $\phi$. For $(\beta, \lambda) = (-0.5, 0.8)$ the $CBSADF^*$ test offers relatively low additional power compared to the $BSADF^*$ test, however we may argue that this could be reasonably attributed to the slightly higher finite sample empirical size of the latter test. In the case where $(\beta, \lambda) = (0.5, 0.8), (-0.8, 0.8)$ the power of the $CBSADF^*$ test is lower than that of the $BSADF^*$ test for small values of $\phi$ due to the oversize exhibited by the $BSADF^*$ test in this scenario, with the power of the $CBSADF^*$ test exceeding that of the $BSADF^*$ test for larger values of $\phi$. Finally, there is no significant contribution of the covariate when $(\beta, \lambda) = (-0.5, 0.5)$ since both tests appear to have similar power properties.

Overall, we argue that both the bootstrap $BSADF^*$ and $CBSADF^*$ tests show better size control than their respective non-bootstrap ($BSADF$ and $CBSADF$) tests across most of the parameter values with the best performance given by the $CBSADF^*$ test. Arguably, the inclusion of covariates in the $CBSADF^*$ test leads to greater power performance relative to the $BSADF^*$ test in finite samples, as well as offering slightly improved size control. We, therefore, recommend utilising the $CBSADF^*$ test in practice as it offers the best overall size control and power properties amongst the tests considered.

8 Empirical Application

Following PSY (2015a) empirical application, we consider the real S&P500 stock price index and the real S&P500 stock price index dividend$^4$ over the period January 1959 to June 2018 at a monthly frequency, constituting 714 observations. We utilise the same dataset with PSY (2015a) as it contains multiple historical bubble episodes and we estimate the present value of the real price-dividend ratio which is the real S&P500 stock price index over the real S&P500 stock price index dividend as outlined in PSY (2015a).

According to Shiller (2015) bonds are related to asset bubbles as when long-term interest rates decrease, bond prices increase creating enthusiasm in a similar way as in the stock market.

Data generated according to $y_t = y_{t-1} + u_t, \quad t = 1, ..., 180$ and $y_t = \phi y_{t-1} + u_t, \quad t = 181, ..., 200$ where $u_t \sim iid N(0,1)$. 

BSADF* test:  

CBSADF* test:  

Figure 1: Finite Sample Power
Figure 2: Finite Sample Power

\begin{align*}
\beta = 0.8 & \quad \lambda = 0.5 \\
\beta = 0.5 & \quad \lambda = 0.5 \\
\beta = -0.5 & \quad \lambda = 0.5 \\
\beta = -0.8 & \quad \lambda = 0.5
\end{align*}

BSADF* test: \quad CBSADF* test: ---

Data generated according to
\begin{align*}
y_t &= y_{t-1} + u_t, \quad t = 1, \ldots, 180 \\
y_t &= \phi_t y_{t-1} + u_t, \quad t = 181, \ldots, 200 \text{ where } u_t \sim \text{iid } N(0, 1).
\end{align*}
Vogel (2010) argues that during the end of double recessions of the 80s, the bond market had reached historical low levels as a consequence of the tight monetary policy and high interest rates FED implemented to deal with double-digit inflation. The "bond market conundrum", the inability of the monetary policy to affect long-term bond yields has been widely acknowledged in the literature (see *inter alia* Evano, Kaufman and Malliaris (2012), Bernanke (2005)).

The CBOE volatility index measures market's expectation of 30-day volatility and it is constructed by using the implied volatilities on the S&P500 index options. In their seminal work, Fleming, Ostdiek and Whaley (1995) provide evidence in support of the argument that there is a tendency of the VIX to rise after large sell-offs and fall after large rallies. In 1998, during the dot.com bubble episode, VIX appeared having a quite wide range as 90% of the VIX levels were between 18.57% and 42.74% whereas after the collapse of the dot.com bubble the range according to Whaley (2000) narrowed to 11% (20% - 31%). The VIX index is also known as a *fear index* as it reflects investors expectations about future volatility and how much they are willing to pay in terms of implied volatility to hedge their stock portfolios as well (Whaley (2000)).

Credit default swaps (CDS) had been broadly traded in the capital markets of the United States (CDX) and Europe (iTraxx). It has been estimated that in 2007, the gross value of the CDS ranged between $45 and $62 trillion (Brunnermeier (2009). The securitised products frenzy and the low lending standards led to cheap credit and resulting in the housing and credit bubbles (Berman (2007)). Thus, the bid ask spread of the credit default swap provides valuable information on the origination of the subprime mortgage episode.

For all the above, we utilise as covariates the Moody's Seasoned Aaa Corporate Bond Yield\(^5\) that covers the full period between January 1959 and June 2018 (714 observations), the Chicago Board Options Exchange Volatility Index (CBOE VIX) from January 1990 to June 2018 (342 observations) and the bid-ask spread of the Credit Default Swap Investment Grade Index (CDX IG INDEX HIST BID-AS)\(^6\) over the period October 2005 to June 2018 (154 observations).

All covariates are sampled at a monthly frequency and due to different data availability of the Volatility Index and the CDS bid-ask spread, the present value of the real price-dividend ratio is re-estimated so that it is always equal to 100 at the beginning of each sample as in PSY (2015a).

---

5Retrieved from: the Federal Reserve Bank of St. Louis webpage https://fred.stlouisfed.org/series/AAA
6As obtained from the Bloomberg Database.
To continue, we apply the BSADF and CBSADF* tests as outlined in sections 3 and 6. In particular, we consider the price-dividend ratio as the dependent variable regressed on its lags and lags and leads of the covariates, namely the corporate bond yield, volatility index and CDS bid-ask spread, separately and in groups. All three covariates used in this paper have been tested for stationarity and, lagged differences have been used for the corporate bond yield to ensure stationarity.

Furthermore, we use the BIC to determine the lags of the differenced dependent variable. If a practitioner is interested in testing for bubbles in real time then one would only consider lags and not leads of the potential covariates. Since we are referring to past bubble episodes we choose to make use of leads as well, but our proposed tests can be equally useful on a real-time basis when using leads might not be feasible. For this reason, in the empirical application we focus on testing for explosive episodes both with and without including leads of the covariates. In order to decide on the number of lags and leads of the covariates we firstly set them to a maximum length of four and then we choose the combination of lags and leads that minimises $\hat{\varrho}^2$.

We run two empirical exercises. In the first one, we utilise all three covariates one-by-one to identify bubble episodes in historical prices. Particularly, we use the corporate bond yield for the period January 1959 to June 2018, the volatility index from January 1990 to June 2018 and the CDS bid-ask spread between October 2010 and June 2018. As such, the number of the lags of the price-dividend ratio is set to 1, the number of the lags of the corporate bond yield is equal to 4 and the number of leads is 4 as well. For the case that we do not include any leads of the corporate bond yield the number of lags that minimises $\hat{\varrho}^2$ is 4. In addition, the volatility index has 4 lags and 3 leads whereas when we set the number of leads to zero, the number of lags of the volatility index is 1. Finally, the number of the lags of the CDS bidask spread is 3 and the leads are equal to 4. When the number of leads of the CDS bid-ask spread is zero, the number of lags is still 3. Results are summarised in Panel A and B from Table 3.

In the second exercise, based on data availability we run the CBSADF* test by including both the corporate bond yield and the volatility index as covariates for the period January 1990 to June 2018 and all three covariates (corporate bond yield, volatility index and CDS bid-ask spread) for the period January 2005 to June 2018. In this case, the lags for the corporate bond yield and the volatility index are 4 each and the leads are 3, whereas in the absence of leads the
number of lags is reduced to 1 for both covariates. Lastly, when we consider all three covariates
the number of lags and leads is 1 and when the number of leads is zero, the lag of the three
covariates remains 1. Panel A and B from Table 4 report the results.

In addition, we compute right-tail finite sample critical values for both tests using 10,000
Monte Carlo and 9,999 bootstrap replications respectively. The minimum window size is deter-
mined as in equation (7.7) where $T$ is the sample size of the observations as outlined by PSY
(2015a). Both tests are performed at a 5% level of significance and a constant is included in the
models.

To investigate the accuracy of our proposed tests to detect bubble episodes in empirical data
series, we follow PSY (2015a) who perform a (pseudo) real-time bubble monitoring exercise on
the present value of the real S&P500 price-dividend ratio and apply a date-stamping strategy
to test for the presence of explosive behaviour. Particularly, we estimate both the $BSADF$ and
$CBSADF^*$ test statistics in a recursive framework and we identify the origination date of
the bubble episode as the first chronological observation whose test statistic is larger than the
simulated finite sample critical value therefore rejecting the null hypothesis of a unit root. On the
same spectrum, the termination date of the bubble episode is defined as the first chronological
observation whose test-statistic becomes less than the simulated finite sample critical value.

Tables 3 and 4 present three periods of bubble episodes (origination and termination dates)
as identified by the PSY (2015a) $BSADF$ test and our proposed $CBSADF^*$ test across three
covariates. We focus on the performance of the $BSADF$ and $CBSADF^*$ tests as detecting
explosive episodes that occur at the end of the sample in real-time, can be useful to regulators,
policy makers and central banks. For the $BSADF$ test we re-estimate the origination and
termination dates of the bubble episodes as we need to adjust the present value of the real
S&P 500 price-dividend ratio due to the different data availability of the covariates. Additionally,
Tables 3 and 4 report the estimated value of the long-run squared correlation coefficient $\hat{\varrho}^2$
that measures the degree of correlation between the stationary covariate $\Delta x_t$ and the error term $u_t$
of equation (2.3).

Although both the $BSADF$ and $CBSADF^*$ tests identify three periods of exuberance;
namely the Black Monday of October 1987, the dot-com bubble of 2000-2001 and the subprime
mortgage episode of 2007-2008 (at least when the data set is large enough e.g. when corporate
bond yield is used as a covariate) we present those explosive episodes that are detected earlier by the particular covariate compared to the standard PSY (2015a) \( BSADF \) test.

Table 3 reports the origination and termination dates of three bubble episodes as identified by the \( BSADF \) test and our proposed \( CBSADF^* \) test across different covariates. In particular, in Panel Ab of Table 3 we see that the \( BSADF \) test identifies May 1987 as the origination date of the Black Monday episode, five months prior to the crash, whereas when the corporate bond yield is used as a covariate, our proposed \( CBSADF^* \) test detects the bubble episode five months earlier compared to the \( BSADF \) test (December 1986) at least when leads of the covariate are not included. When leads of the corporate bond yield are included then the \( CBSADF^* \) test identifies the origination date of the explosive episode a month earlier than the \( BSADF \) test (see Panel B from Table 3).

Furthermore, as it can be seen in Panel B from Table 3, the \( CBSADF^* \) test identifies the origination date of the dot-com bubble episode on July 1995, a month earlier than the \( BSADF \) test by utilising the volatility index as a covariate and considering leads of the volatility index in the model. Finally, the \( BSADF \) test proposes that the subprime mortgage episode originated in August 2008, whereas the \( CBSADF^* \) identifies the origination date four months earlier, in April 2008 when we choose to include leads of the CDS bid-ask spread. However, when no leads of the CDS bid-ask spread are taken into consideration, the \( CBSADF^* \) test does not identify any explosiveness (Panel B from Table 3).

In Panel A from Table 4, we see that the \( BSADF \) test identifies the origination date of the Black Monday episode as in May 1987 whereas our proposed \( CBSADF^* \) test detects the bubble episode five months earlier relative to the \( BSADF \) test by utilising the corporate bond yield as a covariate when we do not include leads of the covariate. The result is the same as in Panel B from Table 3, since data for the corporate bond yield were only available for that time period. In Panel B from Table 4 we see that for the dot-com bubble, the \( CBSADF^* \) test identifies the origination date of the episode as in June 1995 when both the corporate bond yield and the volatility index are utilised as covariates and when we do not include any leads in the model, in contrast to the \( BSADF \) test that suggests August 1995 as the origination date of the bubble. Next, using all three covariates but no leads of them, namely the corporate bond yield, the volatility index and the CDS bid-ask spread, the \( CBSADF^* \) test results in identifying the
origination date of the subprime mortgage episode in December 2007, 8 months earlier than the
BSADF test. It is worth mentioning here that the CBSADF* test identifies the termination
date of the 2007-08 bubble episode on the date that the BSADF test identifies as the origination
date of the episode.

To conclude, our proposed CBSADF* test has identified the Black Monday of October 1987,
the dot-com bubble and the subprime mortgage episode earlier than the BSADF test, reflecting
the power of the CBSADF* test to quickly detect bubble episodes. The CBSADF* test seems
to be able to detect the bubble episodes earlier compared to the standard BSADF test for all
three covariates at least when we include leads. Finally, our empirical results are in accordance
to the theoretical evidence presented in section 7.2, where we argue that the CBSADF* test
shows better size control and power performance than the BSADF test and we provide empirical
evidence that the inclusion of covariates in the CBSADF* test leads to greater power properties
relative to the BSADF* test in finite samples and therefore, recommend utilising the CBSADF* test in practice.

9 Conclusion

The impact of asset bubbles on economic and financial activity has been widely acknowledged.
Bubble identification as a warning diagnostic by regulators is can be of great importance. In a
regulatory framework, the present value models persistently fail to explain bubble episodes and
dynamic models that have been used so far have been proved inadequate (see for example Shiller

The main assumption of this paper is that deviations of asset prices from their fundamentals
for a prolonged period of time lead to rational bubbles. The intuition behind this is that shifts
from a random walk to an explosive regime could be considered as a strong indication of an
explosive episode. However, as argued by Gürkaynak (2008) the hypothesis of no-bubble cannot
be easily rejected, since other determinants such as non-observable fundamentals and behavioural
aspects (e.g. market sentiment) should also be taken into consideration. Often, wrong statistical
inference may be derived by misspecified fundamentals and thus the researcher has to choose
Table 3: Bubble Date Stamping

**Panel A: leads of the covariates**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>BSADF</th>
<th>CBSADF*</th>
<th>$\hat{\varphi}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Bond Yield</td>
<td>1987M5 - 1987M9 (0.028)</td>
<td>1987M4 - 1988M1 (0.030)</td>
<td>0.94</td>
</tr>
<tr>
<td>Volatility Index (VIX)</td>
<td>1995M8 - 2000M9 (0.010)</td>
<td>1995M7 - 2000M4 (0.047)</td>
<td>0.53</td>
</tr>
<tr>
<td>CDS Spread (CDX IG)</td>
<td>2008M8 - 2009M3 (0.033)</td>
<td>2008M4-2008M7 (0.039)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

**Panel B: no leads of the covariates**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>BSADF</th>
<th>CBSADF*</th>
<th>$\hat{\varphi}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Bond Yield</td>
<td>1987M5 - 1987M9 (0.028)</td>
<td>1986M12 - 1987M9 (0.029)</td>
<td>0.95</td>
</tr>
<tr>
<td>Volatility Index (VIX)</td>
<td>1995M8 - 2000M9 (0.010)</td>
<td>1995M8 - 2000M8 (0.010)</td>
<td>0.51</td>
</tr>
<tr>
<td>CDS Spread (CDX IG)</td>
<td>2008M8 - 2009M3 (0.033)</td>
<td>-          (0.039)</td>
<td>0.18</td>
</tr>
</tbody>
</table>

*p-values on the origination date of the bubble episode in brackets.

between a bubble episode and any potential fundamental based explanation. Consequently, rejecting the alternative hypothesis of explosiveness or not depends on the initial assumptions (initial condition, volatility etc.) that have been made as well. Presuming a time-varying discount rate or structural breaks for instance, might loose the strictness of the primary conventions resulting in less evidence in support of the existence of a bubble (Gürkaynak 2008).

Unit root tests have gained a great deal of popularity in theoretical and empirical research the past three decades and it is a common fact that plenty economic and financial time series have unit root characteristics. However the standard univariate unit root tests suffer from low explanatory power (Hansen 1995). We consider a recursive regression methodology of right-tailed unit root test introduced by Phillips, Shi and Yu (2015a) by including covariates as firstly proposed by Hansen (1995) in order to improve the explanatory power of the standard ADF regression model.
### Table 4: Bubble Date Stamping

**Panel A: leads of the covariates**

<table>
<thead>
<tr>
<th>Covariates</th>
<th>BSADF</th>
<th>CBSADF*</th>
<th>$\hat{\varrho}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Bond Yield 1987M5 - 1987M9</td>
<td>1987M5 - 1987M9</td>
<td>1987M4 - 1988M1</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>Corporate Bond Yield &amp; Volatility</td>
<td>1995M8 - 2000M9</td>
<td>1995M7 - 1999M12</td>
<td>0.52</td>
</tr>
<tr>
<td>Index</td>
<td>(0.010)</td>
<td>(0.049)</td>
<td></td>
</tr>
<tr>
<td>Corporate Bond Yield, Volatility</td>
<td>2008M8 - 2009M3</td>
<td></td>
<td>0.025</td>
</tr>
<tr>
<td>Index &amp; CDS Spread</td>
<td>(0.033)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Panel B: no leads of the covariates**

<table>
<thead>
<tr>
<th>Covariates</th>
<th>BSADF</th>
<th>CBSADF*</th>
<th>$\hat{\varrho}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate Bond Yield 1987M5 - 1987M9</td>
<td>1987M5 - 1987M9</td>
<td>1986M12 - 1987M9</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.029)</td>
<td></td>
</tr>
<tr>
<td>Corporate Bond Yield &amp; Volatility</td>
<td>1995M8 - 2000M9</td>
<td>1995M6 - 1999M12</td>
<td>0.50</td>
</tr>
<tr>
<td>Index</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td></td>
</tr>
<tr>
<td>Corporate Bond Yield, Volatility</td>
<td>2008M8 - 2009M3</td>
<td></td>
<td>0.026</td>
</tr>
<tr>
<td>Index &amp; CDS Spread</td>
<td>(0.033)</td>
<td>(0.01)</td>
<td></td>
</tr>
</tbody>
</table>

*p-values on the origination date of the bubble episode in brackets.

The main hypothesis tested is the null of a unit root (random walk) against the alternative of explosiveness. The econometric techniques proposed in this article can offer valuable information on explosive behaviour identification not only in a multiple bubble framework but also on a real-time basis.

This article provides theoretical and empirical evidence that the inclusion of relevant covariates in the conventional Augmented Dickey Fuller test regression leads to improved size control, while offering significant power gains when an end-of-sample explosive episode is present. Specifically, we investigate whether the size and power properties of the tests of Phillips, Shi and Yu (2015a) can be enhanced by the inclusion of information in related time series that traditional univariate unit root tests tend to ignore. When working only in a univariate framework, the
choice to examine a variable in isolation can be rather costly in terms of power since ignoring any correlation with other relevant series leads to unnecessarily high standard errors when constructing the Dickey-Fuller t-statistic, leading to significant power losses. Our proposed test is applied using a recursive window sequence as suggested by Phillips, Shi and Yu (2015a) to detect explosive episodes that occur at the end of the sample.

To deal with potential bias we apply a parametric bootstrap version of the proposed test ensuring the asymptotic validity of the critical values drawn from the bootstrap distribution of the test. We concentrate on the case where the explosive episode takes place at the end of the sample as the detection of ongoing bubbles is of most importance to practitioners, with the tests also being useful in terms of date stamping past bubbles. Simulation results show that the proposed tests have improved size and power properties in finite samples relative to extant tests. In particular, the $CBSADF^*$ tests suffers less severe size distortions compared to conventional tests that do not utilise a bootstrap procedure or omit relevant covariates, whilst also displaying significantly better power properties.

Empirical work explores the effectiveness of the proposed tests as early warning mechanisms, informing policy makers of a bubble episode that might occur, which should provide evidence that the tests might be useful to structuring macroprudential policy. Specifically, we examine whether our proposed tests would have detected known past bubbles in the S&P500 price dividend series before the tests of Phillips, Shi and Yu (2015a) when used as an early warning mechanism, utilising the Moody’s Seasoned Aaa Corporate Bond Yield, the Volatility Index and the CDS bid-ask spread as covariates. The superiority of our proposed test is reflected on the earlier detection of three major explosive episodes: Black Monday of October 1987, the dot-com bubble and the subprime mortgage episode.

The proposed bootstrap unit root testing for explosive behaviour by using covariates may constitute a conservative and strict tool of macro-prudential policy and surveillance. Macroprudential regulation may focus on dealing with bubble episodes using tools that are specifically structured to do so (such as counter-cyclical capital requirements, credit constraints, credit-to-GDP ratio monitoring and margin requirements etc. (Borio 2003)) than monetary policy instruments that might fuel the bubble. Overall, there is a great challenge for both theorists and empirical researchers to understand the magnitude of asset price bubbles, investigate their origin
and causes, track their evolution and handle these phenomena.

There is a great need for deeper understanding by the central banks and policy makers of the mechanisms asset bubbles form, how they grow and collapse and how they contaminate other markets and the real economy as well. Fiscal regulators and institutional surveillance mechanisms require tools with low false detection rate to implement macro-prudential policy implementation to address bubble episodes in financial markets (PSY 2015a). The question whether these bubbles have rational or behavioural determinants is still left to be answered. There is great scope for further research.
Appendix

Proofs

Before proceeding to derive the Limit Theory of the CBSADF test we note that the following results follow trivially from results in Hansen (1995).

Where $W_1$ and $W_2$ are as defined in Lemma 4.1.

Proof of Theorem 4.1

In order to get a limit distribution for the GSADF and BSADF test statistics in this setup we are interested in the limit behaviour of the OLS estimate of $\beta$ from the following subsample regression

$$\Delta y_t = \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + \sum_{k=-r}^{q} \beta_k' \Delta x_{t-k} + \hat{\mu} + \hat{\delta} y_{t-1} + \hat{\epsilon}_t = X_t \theta + \hat{\epsilon}_t$$

estimated for $t = |T_r|, \ldots, |T|$. Henceforth, we will use $\sum$ to denote summation over $|T_r|, \ldots, |T|$. The estimation error in the least squares estimator $\hat{\theta}_{r,1}$ from its true value is given by

$$\hat{\theta}_{r,1} - \theta_{r,1} = \left[ \sum X_t X_t' \right] \left[ \sum X_t \epsilon_t \right]$$

The limit of the signal matrix $\sum X_t X_t'$ is block diagonal with the limit behaviour of $\hat{\delta}$ depending only on the lower $2 \times 2$ block of $\sum X_t X_t'$ and, therefore, the final $2 \times 1$ block of $\sum X_t \epsilon_t$. Defining the scaling matrix $\Upsilon_T = diag(\sqrt{T}, T)$ we can write

$$\Upsilon_T \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\delta} - \delta \end{pmatrix} = \left( \Upsilon_T^{-1} \left[ \sum X_t X_t' \right]_{(2) \times (2)} \Upsilon_T^{-1} \right) \left( \Upsilon_T^{-1} \left[ \sum X_t \epsilon_t \right]_{(2) \times 1} \right)$$

where the notation $(2) \times (2)$ and $(2) \times 1$ means the lower $2 \times 2$ and $2 \times 1$ block, respectively. The matrix $\Upsilon_T^{-1} \left[ \sum X_t X_t' \right]_{(2) \times (2)} \Upsilon_T^{-1}$ has partitioned form
\[
\begin{pmatrix}
\sqrt{T} & 0 \\
0 & T
\end{pmatrix}^{-1}
\begin{pmatrix}
\sum 1 & \sum y_{t-1} \\
\sum y_{t-1} & \sum y_{t-1}^2
\end{pmatrix}
\begin{pmatrix}
\sqrt{T} & 0 \\
0 & T
\end{pmatrix}^{-1}
= \begin{pmatrix}
T^{-1} \sum 1 & T^{-3/2} \sum y_{t-1} \\
T^{-3/2} \sum y_{t-1} & T^{-2} \sum y_{t-1}^2
\end{pmatrix}
\rightarrow
\begin{pmatrix}
T^{-1/2} \sum \varepsilon_t & \sigma \varepsilon W^*(1) - W^*(r) \\
T^{-1} \sum y_{t-1} \varepsilon_t & \sigma \varepsilon \int_r^1 W_1(s) ds
\end{pmatrix}
\]

from (??) and (??) and where \(r_w = 1 - r\).

the vector \(Y_T^{-1} [\sum X_t \varepsilon_t]_{2 \times 1}\) has components

\[
\begin{pmatrix}
T^{-1/2} \sum \varepsilon_t \\
T^{-1} \sum y_{t-1} \varepsilon_t
\end{pmatrix}
\rightarrow
\begin{pmatrix}
\sigma \varepsilon (W^*(1) - W^*(r)) \\
\sigma \varepsilon \int_r^1 W_1 dW^*
\end{pmatrix}
\]

from (??) and (??). So under the null we have

\[
\begin{pmatrix}
\sqrt{T} (\hat{\mu} - \mu) \\
T \hat{\beta}
\end{pmatrix}
\rightarrow
\begin{pmatrix}
A & B
\end{pmatrix}^{-1}
\begin{pmatrix}
C \\
D
\end{pmatrix}
\]

where

\[
A = \sigma \int_r^1 W_1(s) ds
\]
\[
B = \sigma^2 \int_r^1 W_1(s)^2 ds
\]
\[
C = \sigma \varepsilon (W^*(1) - W^*(r))
\]
\[
D = \sigma \varepsilon \int_r^1 W_1 dW^*
\]

Therefore

\[
T \hat{\beta} \rightarrow \frac{AC - r_w D}{A^2 - r_w B}
\]

The \(t\)-statistic for testing the null of \(\delta = 0\), \(t_{\delta - 0}\), uses the standard error \(s_\delta\) defined by

32
\[ s_δ^2 = \frac{\hat{\sigma}_\epsilon^2 \left( \sum_{1} y_{t-1} \sum_{y_{t-1}} \right)^{-1}}{\sum y_{t-1}^2 - (\sum y_{t-1})^2 / \sum 1} \]

where \( \hat{\sigma}_\epsilon^2 = (1/Tr_w) \sum \hat{\sigma}_t^2 \rightarrow \sigma^2_\epsilon \), so that

\[ T^2 s_δ^2 \rightarrow \frac{\sigma^2_\epsilon}{B - A^2/r_w} \]

So finally we obtain the following limit for the \( t \)-statistic \( t_\delta = 0 \)

\[ t_\delta = \frac{T \hat{\delta}}{(T^2 s_δ^2)^{1/2}} \rightarrow \left( \frac{AC - r_w D}{A^2 - r_w B} \right) \left( \frac{B - A^2/r_w}{\sigma^2_\epsilon} \right)^{1/2} \]

\[ = \frac{r_w D - AC}{(r_w B - A^2)^{1/2} \sigma^2_\epsilon r_w^{1/2}} \]

\[ = \frac{r_w \sigma_u \sigma_\epsilon}{\sigma_u \sigma_\epsilon} \left( r_w \int_r^1 W_1 dW^* - \sigma_u \sigma_\epsilon \left( \int_r^1 W_1(s) ds \right) (W^*(r) - W^*(1)) \right) \]

\[ = \frac{r_w \int_r^1 W_1 dW^* - \left( \int_r^1 W_1(s) ds \right) (W^*(1) - W^*(r))}{r_w^{1/2} \left( r_w \int_r^1 W_1(s)^2 ds - \left( \int_r^1 W_1(s) ds \right)^2 \right)^{1/2}} \]

This is the limit for one subsample only. As the CBSADF test is the supremum of all possible subsamples with minimum window width \( r_0 \) the limit distribution of the CBSADF, therefore, takes the form

\[ \sup_{r \in [0,1-r_0]} \left\{ \frac{r_w \int_r^1 W_1 dW^* - \left( \int_r^1 W_1(s) ds \right) (W^*(1) - W^*(r))}{r_w^{1/2} \left( r_w \int_r^1 W_1(s)^2 ds - \left( \int_r^1 W_1(s) ds \right)^2 \right)^{1/2}} \right\} \]
References


