Using covariates to improve the efficacy of univariate bubble detection methods☆

Sam Astill [a], A.M. Robert Taylor [a,*], Neil Kellard [a], Ioannis Korkos [b]

[a] *Essex Business School, University of Essex, United Kingdom*
[b] *Business School, London South Bank University, United Kingdom*

A R T I C L E   I N F O

A B S T R A C T

We explore how information additional to a specific price series can be used to improve the power of popular univariate autoregressive-based methods for detecting and dating speculative price bubble episodes. Following Phillips et al. (2011, 2015) we base our approach on sequences of sub-sample regression-based augmented Dickey–Fuller [ADF] statistics. Our point of departure from these extant procedures is to allow for additional information in the testing and dating procedures. To do so we follow the approach of Hansen (1995) and augment the sub-sample ADF regressions with covariate regressors. The limiting null distributions of the resulting statistics depend on the long-run squared correlation between the covariates and the regression error. We show that this dependence can be accounted for by using a residual bootstrap re-sampling method. Simulation evidence shows that including relevant covariates can significantly improve the efficacy of both the resulting bubble detection tests and the associated date-stamping procedure, relative to using standard sub-sample ADF statistics. An empirical application of the proposed methodology to monthly S&P 500 data is considered, using a variety of candidate covariates. Using these covariates, the onset of the dotcom bubble and the bubble associated with Black Monday are both identified significantly earlier than when using standard methods.

1. Introduction

Large price swings away from fundamentals characterised by bubbles (see Homm and Breitung, 2012; Shiller, 2015; Phillips and Shi, 2018) represent a misallocation of resources during the upward bubble phase and a subsequent destruction of value during the crash phase. Frequently, crashes are followed by a recession as deleveraging combines with a lack of confidence in the financial system to reduce the supply of credit to the real economy, as occurred in the aftermath of the Global Financial Crisis [GFC] of 2007/08. Policymakers reacted by considering new rules for macroprudential regulation and intervention including countercyclical capital and margin requirements.

A key element of a macroprudential policy approach is the rapid and accurate detection of ongoing asset price bubbles and the ability to identify the location and duration of historical bubble episodes. Accordingly, a large literature has developed around

testing for the presence of explosive rational asset bubbles. Recently, Phillips et al. (2011, PWY) and Phillips et al. (2015, PSY) propose right-tailed tests for the presence of bubble episodes based on the maximum of sequences of recursive univariate augmented Dickey–Fuller [ADF] unit root statistics applied to overlapping sub-samples of asset price series. The former is designed to detect a single bubble episode whose origination date is at (or before) the start of the available sample data, while the latter is designed for detecting multiple bubble regimes. Associated methods for date-stamping bubble episodes are also proposed in PSY, and can also be used to provide an ongoing approach for monitoring for the emergence of a bubble in real-time.

Further contributions using sub-sample univariate testing methods have been developed in Harvey et al. (2016), Astill et al. (2017) and Phillips and Shi (2018) among others. However, it seems very likely that information *additional* to a specific price series could help improve on the power of existing univariate bubble detection procedures and the associated date-stamping methods. Indeed, in the context of standard full sample left-tailed unit root tests, Hansen (1995) shows that examining a variable in isolation can be potentially very costly in terms of power and accordingly suggests a covariate-based version of the standard ADF unit root test whereby the usual ADF regression is augmented by stationary covariates.

Building on the approach of Hansen (1995), our contribution is to extend the bubble detection and date-stamping procedures proposed in PWY and PSY by developing analogous procedures based on the corresponding sequences of covariate augmented ADF (CADF) statistics from overlapping sub-samples of the data. Our approach does this by incorporating such covariates into the sub-sample ADF regressions used in the PWY and PSY procedures in the same way that the approach in Hansen (1995) incorporates covariates into the standard full sample ADF regression. Of course, in an empirical setting, the choice of covariate is key (Elliott and Jansson, 2003) and given the need to detect bubbles is most obviously seen in financial contexts, this also dictates the type of covariate one might use.

In an asset pricing context, the extant literature suggests several potential covariates that could aid in identifying periods of explosive, nonstationary behaviour. To mention only a few, for equities, dividend discount type models (Diba and Grossman, 1998; PSY) link prices to the risk-free rate of interest, whilst the capital asset pricing model (Kim and Kim, 2016) can embed time-varying volatility. Pricing equations for commodity spot prices (Tsvetanov et al., 2016) indicate inventories (Kilian and Murphy, 2014) and speculation measures (Singleton, 2014), whilst for cryptocurrencies, investor sentiment (Chen et al., 2019) may play an important role. Finally, given bubble behaviour in real estate may precede equity (Caballero et al., 2008) and commodity market bubbles (Phillips and Yu, 2011), potential housing market covariates such as interest rates, disposable income and mortgage finance (White, 2015) may be particularly useful. Overall, in the financial sphere, the availability of many potential covariates holds out the possibility of enhanced bubble monitoring in a range of systemically important markets and consequently, better-quality macroprudential policy, if a suitable covariate-based test can be devised.

It is important to note, that as with the full sample CADF statistics of Hansen (1995), the limiting null distributions of our proposed supremum-based statistics depend on the long-run squared correlation between the covariates and the regression error. While this nuisance parameter can be consistently estimated under the unit root null hypothesis from the data, Chang et al. (2017, CSS) demonstrate that this estimate tends to be heavily biased in finite samples leading to potentially severe size distortions in the CADF test. We find that, perhaps not surprisingly, this problem is somewhat amplified when considering tests based on the suprema of sequences of such statistics applied across sub-samples of the data. To rectify this issue we implement our tests using a bootstrap re-sampling procedure originally outlined in CSS. Monte Carlo simulations show that our proposed bootstrap covariate tests offer impressive size and power performance in finite samples relative to a univariate approach and also improve upon the accuracy of the associated date-stamping methods.

An empirical application to the S&P 500 is conducted, where we utilise a variety of covariates including Moody's Seasoned Aaa Corporate Bond Yield, the return on long-term AAA- and BAA-rated Corporate Bonds, the 10 Year US Treasury Constant Maturity Rate, the AAA credit spread and VXO volatility index. Our proposed tests are applied to data spanning the period associated with the lead up to Black Monday in October 1987 and the dotcom bubble in the late 1990s. In the vast majority of cases, our proposed tests would have detected the emergence of these two bubble episodes significantly earlier than the standard PSY tests. Given the severity of these crashes, this extra time would have been critical in implementing policies to ameliorate the significant real economic damage caused by the subsequent loss of investor confidence and associated credit tightening in the aftermath of the implosion of these bubbles.

The remainder of this paper is organised as follows. In Section 2 we outline the explosive asset price model and the relevant assumptions. Section 3 provides an overview of the conventional covariate augmented unit root tests of Hansen (1995) and the univariate bubble detection tests and bubble date-stamping procedures of PWY and PSY based on recursive estimation methods. Our proposed covariate augmented bubble detection procedures are detailed and their large sample properties derived in Section 4. Here we also discuss bootstrap implementation of the bubble detection and date-stamping methods. The finite sample properties of our proposed tests and date-stamping methods are examined through Monte Carlo simulation in Section 5 and compared with the corresponding univariate methods of PSY. Section 6 presents an empirical application of our proposed methods to the S&P 500 price dividend series previously analysed in PSY. Section 7 concludes. A full description of the bootstrap algorithm used to perform our proposed tests, discussion of how to allow for various specifications for the deterministic elements of the series used in the tests, and proofs of the main theorems which appear in the paper are relegated to a mathematical appendix.

In what follows $\xrightarrow{w}$, $\xrightarrow{p}$, and $\xrightarrow{w}_{p}$ denote weak convergence, convergence in probability, and weak convergence in probability, respectively, in each case as the sample size, T, diverges to infinity, $\lfloor \cdot \rfloor$ denotes the integer part of its argument, and $\mathbb{I}(\cdot)$ denotes the indicator function which returns the value 1 when its argument is true and 0 otherwise.

2. The model and assumptions

Consider the time series $\{y_t\}$ satisfying the data generating process (DGP):

$$\Delta y_t = \delta_t y_{t-1} + u_t, \quad t = 1, \ldots, T \tag{1}$$

with $\Delta := (1 - L)$, L the lag operator such that $L^k y_t = y_{t-k}$, $k = 0, \pm1, \pm2\ldots$.[1] We allow the error term, u_t, in (1) to be both serially correlated and to be related to other stationary covariates. This is done by assuming that u_t in (1) is generated according to

$$\alpha(L)u_t = b(L)'w_t + \varepsilon_t, \tag{2}$$

where ε_t is a martingale difference sequence with constant unconditional variance (precise assumptions on ε_t can be found in Assumption A.1 in Appendix A.3), and w_t is an m-vector of (mean zero) stationary covariates, $\alpha(L)$ is a pth order stationary lag polynomial, defined as $\alpha(z) := 1 - \sum_{k=1}^{p} \alpha_k z^k$, and $b(z) := \sum_{k=-q_1}^{q_2} \beta_k z^k$ is a lag polynomial allowing for, but not requiring, both leads and lags of the covariate w_t to enter the DGP. In many applications it will be natural for the covariates specified in w_t to constitute the first differences of other relevant financial and/or macroeconomic time series, although we will not impose this in our formulation which will simply require the elements of w_t to be weakly dependent ($I(0)$).

Our interest in this paper centres on testing the constant unit root null hypothesis

$$H_0 : \delta_t = 0, \text{ for all } t = 1, \ldots, T \tag{3}$$

in (1) against the alternative hypothesis that the series y_t displays explosive behaviour for some non-vanishing subset of the sample observations considered; that is,

$$H_1 : \delta_{\lfloor Tr \rfloor} = 0, r \notin \mathcal{R}_0; \quad \delta_{\lfloor Tr \rfloor} > 0, r \in \mathcal{R}_0 \tag{4}$$

where $\mathcal{R}_0$ is a non-empty convex subset of the unit interval $[0, 1]$; for example, unions of non-overlapping sub-intervals of $[0, 1]$. Under the null hypothesis, H_0, y_t follows a unit root ($I(1)$) process (subject to the technical conditions on u_t outlined in Appendix A.3 holding) throughout the sample period, while the alternative hypothesis, H_1, specifies the presence of either a single or multiple distinct bubble regimes (each of whose length is a fraction of the sample size, T) within the sample period, with y_t following a unit root process outside of these bubble regimes. Where a bubble regime terminates before the end of the sample period, T, a mechanism for the collapse of the bubble needs to be specified. This will be discussed later in Section 3.2.

Define $v_t := b(L)'w_t + \varepsilon_t$ and consider the long-run covariance matrix of $\zeta_t := (v_t, \varepsilon_t)'$:

$$\Omega = \begin{bmatrix} \sigma_v^2 & \sigma_{v\varepsilon} \\ \sigma_{v\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix} := \sum_{k=-\infty}^{\infty} E(\zeta_t \zeta_{t-k}') \tag{5}$$

Provided Ω is positive definite, we can then define the long-run squared correlation coefficient between v_t and ε_t as

$$\varrho^2 := \frac{\sigma_{v\varepsilon}^2}{\sigma_v^2 \sigma_\varepsilon^2}. \tag{6}$$

The parameter ϱ^2 is crucial to understanding the power gains that can be achieved by extending the bubble testing procedure of PSY and PWY to incorporate the covariates, w_t. As discussed in Hansen (1995), pp. 1150–1151, ϱ^2 measures the relative contribution of w_t to v_t at the long-run frequency. At one extreme if $b(L) = \mathbf{0}$, such that the covariates contain no useful information, then $\varrho^2 = 1$, while if the covariates jointly explain almost all of the long-run movement in v_t then ϱ^2 will be close to zero.

For the purposes of this paper we follow CSS and assume that the covariates w_t are generated by the $AR(\ell)$ process,

$$\Psi(L)w_{t+q_1+1} = \eta_t, \tag{7}$$

where η_t is an m-dimensional martingale difference sequence with constant variance matrix (precise assumptions on η_t can again be found in Assumption A.1 in Appendix A.3) and $\Psi(z) := I_m - \sum_{k=1}^{\ell} \Psi_k z^k$. Like CSS, who develop bootstrap implementations of the full sample CADF test of Hansen (1995), a parametric DGP is assumed for the covariates in (7) for the purposes of developing bootstrap implementations of the covariate augmented bubble detection tests proposed in this paper. Asymptotic versions of these tests (i.e. tests based on critical values from the limiting null distributions of the statistics) could equally well be based on the corresponding (non-parametric) strong mixing-based assumption employed in Assumption 1 of Hansen (1995, p.1151).

Remark 2.1. The DGP specified above implies that both y_t and w_t have (unconditional) mean zero and we outline our methods under this assumption. This assumption can be relaxed to allow for the possibility of a constant plus linear trend in the mean of the observed price series and covariates by specifying the price series in the latter case as $y_t^\dagger = a_0 + a_1 t + y_t$ and $w_t^\dagger = b_0 + b_1 t + w_t$, with y_t and w_t generated as previously specified, where a_0 and a_1 are scalar constants and b_0 and b_1 are m-vectors of constants. The special case thereof where the mean of each of the two series is a constant obtains by setting $a_1 = 0$ and $b_1 = \mathbf{0}$. The tests which we describe in the remainder of the paper are outlined for the case where y_t and w_t both have zero mean. Details of how the methods we outline can be generalised to allow for a constant and/or linear trend in the mean of y_t and w_t is provided in Appendix A.2.

---

[1] The initial condition, y_0, does not affect the large sample results given for our proposed tests provided it is stochastically bounded and so we set $y_0 = 0$ in what follows for expositional brevity.

3. Existing test procedures

In this section we briefly review the covariate augmented unit root tests of Hansen (1995) and the univariate bubble detection procedures of PWY and PSY.

3.1. Hansen's covariate augmented ADF unit root tests

In the context of the model in (1) under the parameter constancy restriction that $\delta_t = \delta$ for all $t = 1, \dots, T$, Hansen (1995) considers testing the null hypothesis that y_t is a unit root process, $H_0 : \delta = 0$ against the alternative of stationarity, $H_1 : \delta < 0$.

Hansen (1995) shows that, under this parameter constancy restriction, (1) and (2) can be combined to give the following single equation

$$\Delta y_t = \delta y_{t-1} + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + \sum_{k=-q_1}^{q_2} \beta_k' w_{t-k} + \varepsilon_t. \tag{8}$$

the parameters of which can be consistently estimated by OLS under Assumptions A.1 and A.2. Eq. (8) is essentially a standard ADF regression for y_t augmented by the covariates in w_t, together with the leads and lags thereof (as appropriate). Accordingly, Hansen (1995) proposes a CADF test of $H_0 : \delta = 0$ against $H_1 : \delta < 0$ based on the conventional OLS regression t-statistic on y_{t-1} in (8), $CADF(p, q_1, q_2) := \hat{\delta}/se(\hat{\delta})$, where $\hat{\delta}$ is the OLS estimate of δ and $se(\hat{\delta})$ is its standard error, using the full available data sample, $t = 1, \dots, T$.

Theorem 3, of Hansen (1995, p.1154) provides representations for the limiting distributions of the $CADF(p, q_1, q_2)$ statistics. These depend on the long-run correlation parameter ϱ^2. Consequently, to form a test of H_0 against H_1 left-tailed critical values from these distributions are needed. Table 1 of Hansen (1995, p.1155) provides these for $\varrho^2 = \{1.0, 0.9, 0.8, \dots, 0.1\}$. However, the value of ϱ^2 is not known in practice and so must be estimated from the data. Hansen (1995), pp. 1155–1156) suggests the following non-parametric estimator of ϱ^2:

$$\hat{\varrho}^2 := \frac{\hat{\sigma}_{v\varepsilon}^2}{\hat{\sigma}_{\varepsilon}^2 \hat{\sigma}_{v}^2} \tag{9}$$

where

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{v\varepsilon} \\ \hat{\sigma}_{v\varepsilon} & \hat{\sigma}_{\varepsilon}^2 \end{bmatrix} := \sum_{k=-M}^{M} g(k/M) T^{-1} \sum_t \hat{\zeta}_{t-k} \hat{\zeta}_t' \tag{10}$$

in which $\hat{\zeta}_t := (\hat{v}_t, \hat{\varepsilon}_t)'$, where $\hat{\varepsilon}_t$ are the OLS residuals obtained from estimating (8) and $\hat{v}_t := \sum_{k=-q_1}^{q_2} \hat{\beta}_k' w_{t-k} + \hat{\varepsilon}_t$, $\{\hat{\beta}_k\}_{k=-q_1}^{q_2}$ the OLS slope estimates from (8). In the context of (10), suitable choices of the kernel function, $g(\cdot)$, and the bandwidth, M, are discussed in Hansen (1995, p.1156). The test can then be implemented using the estimated $\hat{\varrho}^2$ to select the most appropriate asymptotic critical value from Table 1 of Hansen (1995).

Theorem 3 of Hansen (1995) also shows that the asymptotic local power of the CADF test based on the $CADF(p, q_1, q_2)$ statistic has the property that it increases as ϱ^2 decreases, other things equal. Monte Carlo simulations reported in Hansen (1995) and in CSS demonstrate that the inclusion of relevant covariates can lead to substantial increases in finite sample power relative to the corresponding standard ADF unit root test derived from (8). These simulation results also show that, unlike the property of the asymptotic local power function of the test noted above, the finite sample power of the CADF test is not necessarily higher the smaller the value of ϱ^2.

A serious practical drawback of the approach outlined in Hansen (1995), however, is that the estimate of ϱ^2 in (9) can be quite badly biased in finite samples; see the simulation results in Table 1 on page 145 of CSS. A consequence of this bias will be that the asymptotic critical value being used based on that estimate will likely be a poor approximation to the true critical value for the test and, as a result, the results in CSS demonstrate that the CADF test can be very badly sized in practice. Indeed, the results in CSS highlight that the finite sample bias in estimating ϱ^2 is worse the closer is ϱ^2 to zero, so that the CADF test becomes more unreliable in practice as the theoretical power gains it offers over the standard ADF test become larger. As a result CSS propose a bootstrap implementation of the CADF test which does not require an estimate of ϱ^2 and nor does it require tabulations of asymptotic critical values. We will subsequently utilise the re-sampling approach proposed in CSS in the context of the covariate augmented bubble tests proposed in Section 4 of this paper.

3.2. The PWY and PSY bubble detection procedures

Proceeding in the broad spirit of Diba and Grossman (1988) while taking account of Evans (1991) criticisms, PWY and PSY propose the use of univariate sub-sample estimation methods as a means to testing the unit root null hypothesis against the alternative hypothesis that explosive financial bubbles are present in the data; cf. (3) and (4).

To first understand the idea behind the PWY approach, consider the generic (non-covariate augmented) ADF regression,

$$\Delta y_t = \delta y_{t-1} + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k} + v_t. \tag{11}$$

For the remainder of this section we will assume, as in PWY and PSY, that $b(L) = \mathbf{0}$ in (2) such that $v_t = \varepsilon_t$ in (11).[2] Even under this restriction, the ADF regression in (11) is misspecified under the DGP in (1) and (2) because the parameter on y_{t-1} is taken to be fixed rather than time-varying; that is, the correct ADF re-parametrisation of (1) and (2) would have δ_t as the slope parameter on y_{t-1} in (11). The approach taken by PWY and PSY is based on the use of ADF statistics obtained from estimating (11) over sub-samples of the full available sample, $t = 1, \dots, T$, treating δ_t as if it were constant across that sub-sample.

To this end, define the sub-sample ADF statistic $ADF(p)_{r_1}^{r_2} := \tilde{\delta}(r_1, r_2)/se(\tilde{\delta}(r_1, r_2))$, where $\tilde{\delta}(r_1, r_2)$ is the OLS estimate of δ and $se(\tilde{\delta}(r_1, r_2))$ its standard error, from estimating (11) over the data sub-sample $t = \lfloor r_1 T \rfloor + 1, \dots, \lfloor r_2 T \rfloor$. Suppose it was suspected, for example, that a bubble episode occurred starting at $t = 1$ which lasted until $t = \lfloor r_b T \rfloor$. In this case one would clearly want to construct a right-tailed ADF test based on the single sub-sample ADF statistic $ADF(p)_0^{r_b}$ as this would be the largest possible sub-sample of purely explosive data available. In contrast, the full sample right-tailed ADF test based on $ADF(p)_0^1$ would contain a mix of explosive and non-explosive observations and so would be expected to be less powerful than the test based on $ADF(p)_0^{r_b}$, the more so the smaller the value of r_b. In practice, however, we would not know the location of the putative bubble episode and so PWY propose a right-tailed test based on the maximum of the sequence of ADF statistics obtained across the forward recursive sequence of sub-samples of the data, viz,

$$SADF := \sup_{r_2 \in [r_0, 1]} \{ADF(p)_0^{r_2}\}. \tag{12}$$

This test is therefore seen to be based on the supremum of the sequence of ADF statistics computed over all possible sub-samples of the data starting at $t = 1$, subject to a minimum sample size $\lfloor r_0 T \rfloor$. Notice that the full sample ADF statistic, $ADF(p)_0^1$, is the last element in this sequence. In the example above where an explosive bubble is present between $t = 1$ and $t = \lfloor r_b T \rfloor$, the supremum of this sequence would in theory be located at $t = \lfloor r_b T \rfloor$, and so the supremum test would effectively be based on the $ADF(p)_0^{r_b}$ statistic but without the practitioner needing to know r_b. Crucially, PWY show that this supremum test is much more powerful at detecting periodic episodes of explosive behaviour that occur towards the start of the available sample data than the corresponding right-tailed ADF test based on $ADF(p)_0^1$.

The $SADF$ statistic is constructed from ADF statistics calculated over sub-samples with the same start date, $t = 1$, and is therefore designed to detect explosive episodes that occur towards the start of the sample. If, however, interest centred on explosive episodes that occur towards the end of the sample, as would naturally be the case in implementing an on-going monitoring exercise for the emergence of an end of sample bubble, then a natural alternative would be to instead construct a test statistic from ADF statistics calculated on sub-samples of the data whose end date is the end of the sample. This approach is suggested by PSY who propose a right-tailed test based on the following test statistic,

$$BSADF := \sup_{r_1 \in [0, 1-r_0]} \{ADF(p)_{r_1}^1\}. \tag{13}$$

This test is therefore based on the supremum taken over the backward recursive sequence of ADF statistics computed on all possible sub-samples of the data ending at time $t = T$, subject to a minimum sample size $\lfloor r_0 T \rfloor$. The $BSADF$-based test is, therefore, designed to detect end-of-sample explosive episodes as it is computed using relatively more observations from the end of the sample than from the start.

PSY also propose a further (right-tailed) test based on the supremum of a sequence of ADF test statistics computed over all possible sub-samples of the data, again subject to a minimum sample size $\lfloor r_0 T \rfloor$. This test rejects for large positive values of

$$GSADF := \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \{ADF(p)_{r_1}^{r_2}\}. \tag{14}$$

The double recursive $GSADF$-based test is designed to test H_0 against the presence of one or more explosive episodes regardless of their location in the sample.

A rejection by any of these three tests can only signal that a sample contains at least one explosive episode, but not where these are located within the sample. Accordingly, PSY propose a wider approach for bubble detection and identification. Here one first applies the $GSADF$-based test to identify whether one or more bubble episodes are present in a series. If this test rejects H_0, then the next step in the PSY procedure is to date-stamp the bubble episode(s) in the data using sequential applications of the $BSADF$ statistic.

To illustrate PSY's date-stamping procedure, consider the case of a single explosive episode; the full procedure for date-stamping multiple possible bubbles is detailed in section 3 of PSY. Here one proceeds to estimate the origination date, t_e and the termination date, t_f, of the bubble. The origination date is estimated as the first date at which the $BSADF$-based test when applied sequentially to the data would signal a rejection of H_0. Specifically, consider the following statistic

$$BSADF_{r_2} := \sup_{r_1 \in [0, r_2 - r_0]} \{ADF(p)_{r_1}^{r_2}\}. \tag{15}$$

---

[2] Observe that (8) and (11) are equivalent representations and that the contribution from the covariates in the former, $\sum_{k=-q_i}^{q_2} \beta_k' w_{t-k}$, is subsumed into the error term, v_t, in the latter. An implication of this is that serial correlation in w_t will, in general, effect serial correlation in the error term in (11) when $b(L) \neq \mathbf{0}$. As such, for ADF-type tests obtained from (11) to have pivotal limiting null distributions the lag truncation order used in (11) would need to be set to some value, p_v say, where $p_v \geq p$, to ensure that the error in the ADF regression equation was approximately serially uncorrelated; if v_t contained a moving average component, then p_v would need to be such that $(1/p_v) + (p_v^2/T) \to 0$ as $T \to \infty$; cf. Chang and Park (2002).

The test statistic $BSADF_{r_2}$ is therefore seen to be simply the $BSADF$ statistic in (13) but applied only to the sample of data $t = 1, \ldots, \lfloor r_2 T \rfloor$ (treating this sub-sample as if it were the full available sample of data). Sequentially applying this statistic (i.e. for samples ending at $\lfloor r_2 T \rfloor + 1$, $\lfloor r_2 T \rfloor + 2$, etc. through finally to the full sample ending at T), the estimated origination date of the bubble is then estimated as $\hat{t}_e = \lfloor \hat{r}_e T \rfloor$, where

$$\hat{r}_e := \inf_{r_2 \in [r_0, 1]} \{ r_2 : BSADF_{r_2} > scv_{r_2}^{\theta_T} \}. \tag{16}$$

where $scv_{r_2}^{\theta_T}$ is the $100(1 - \theta_T)\%$ (asymptotic) critical value from the null distribution of the $BSADF_{r_2}$ test statistic. As PSY note, this approach could therefore be used as an on-going monitoring procedure to detect the onset of a bubble episode in real-time.

The estimated termination date of the bubble, $\hat{t}_f$, is analogously defined as the first date after $\hat{t}_e + L_T$ for which the sequentially applied $BSADF$ statistic falls below the critical value $scv_{r_2}^{\theta_T}$. PSY impose the condition that for a bubble to be identified its duration must exceed a minimum duration period given by $L_T := \gamma \log(T)$ where γ is a constant chosen based on the frequency of the data. In practical applications PSY suggest setting this minimum duration, L_T, equal to one calendar year. The estimated termination date of the bubble episode is therefore given by $\hat{t}_f = \lfloor \hat{r}_f T \rfloor$, where

$$\hat{r}_f := \inf_{r_2 \in [\hat{r}_e + \gamma \log(T)/T, 1]} \{ r_2 : BSADF_{r_2} < scv_{r_2}^{\theta_T} \}. \tag{17}$$

The significance level, θ_T, used in (16) and (17) can be chosen such that $\theta_T \to 0$ as $T \to \infty$. This ensures that $scv_{r_2}^{\theta_T}$ diverges to infinity with T, thereby eliminating Type I error as $T \to \infty$. In empirical applications PSY recommend fixing θ_T at some pre-determined significance level instead of using drifting significance levels, and it is this approach that we will adopt when performing our empirical application in Section 6.

PSY demonstrate that (16) and (17) provide consistent estimates of the bubble origination and termination dates, respectively, in the case of a DGP which contains a single bubble episode that is subject to *collapse* on termination; that is,

$$y_t = y_{t-1} \mathbb{I}(t < t_e) + \delta_T y_{t-1} \mathbb{I}(t_e \le t \le t_f) + \left(\sum_{k=1}^{t} v_k + y_{t_f}^* \right) \mathbb{I}(t > t_f) + v_t \mathbb{I}(t \le t_f)$$

where $\delta_T := 1 + cT^{-\alpha}$, $c > 0$ and $\alpha \in (0, 1)$, and where $y_{t_f}^* = y_{t_e} + y^*$ with $y^* = O_p(1)$.

PSY show that the consistency of the estimated origination and termination dates extends to the case where two or more collapsed bubble episodes occur. The consistency of PSY's date-stamping estimates can also be shown to hold for other possible transitional mechanisms between the explosive and non-explosive phases of the process; see, in particular, Harvey et al. (2017) and Phillips and Shi (2018).

4. Covariate augmented PWY and PSY procedures

A drawback of the PWY and PSY tests outlined in Section 3.2 is that, although they display increased power to detect the presence of bubble episodes in the data relative to a full sample ADF test, they are still based on the assumption of a univariate DGP for y_t. Given the complex relationships across multiple asset prices and the other inter-relationships that exist between price variables and other macro and financial variables, and the power gains demonstrated by Hansen (1995) from using covariate augmentation in the context of a full sample ADF regression, our contribution is to propose analogous covariate augmentation of the sub-sample ADF regressions used in constructing the PWY and PSY statistics in (12)–(14).

To that end, we consider applying the sub-sample regression based approach of PWY and PSY but in the context of the covariate augmented ADF regression in (8) rather than the standard ADF regression in (11) considered by PWY and PSY. Precisely, we propose right-tailed tests based on the following covariate augmented analogues of the PWY and PSY statistics from Section 3.2,

$$CSADF := \sup_{r_2 \in [r_0, 1]} \{ CADF(p, q_1, q_2)_0^{r_2} \}, \tag{18}$$

$$CBSADF := \sup_{r_1 \in [0, 1-r_0]} \{ CADF(p, q_1, q_2)_{r_1}^1 \}, \tag{19}$$

$$CGSADF := \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \{ CADF(p, q_1, q_2)_{r_1}^{r_2} \}. \tag{20}$$

where $CADF(p, q_1, q_2)_{r_1}^{r_2} := \hat{\delta}(r_1, r_2)/se(\hat{\delta}(r_1, r_2))$, with $\hat{\delta}(r_1, r_2)$ denoting the OLS estimate of δ and $se(\hat{\delta}(r_1, r_2))$ its standard error, from estimating (8) over the data sub-sample $t = \lfloor r_1 T \rfloor + 1, \ldots, \lfloor r_2 T \rfloor$. The proposed tests based on (18), (19) and (20) are, therefore, covariate augmented analogues of the right-tailed tests of PWY and PSY based on the statistics defined in (12), (13) and (14), respectively.

Remark 4.1. The date-stamping methods proposed in PSY discussed in Section 3.2 can similarly be adapted to be based on covariate augmented ADF regressions simply by replacing $BSADF_{r_2}$ in (16) and (17) by $CBSADF_{r_2}$, the $CBSADF$ statistic applied to data from $t = 1, \ldots, \lfloor r_2 T \rfloor$, and replacing $scv_{r_2}^{\theta_T}$ by the analogous (asymptotic) critical value, $cscv_{r_2}^{\theta_T}$ say, from the null distribution of the $CBSADF_{r_2}$ statistic. In the context of an historical date-stamping exercise, if leads of w_t are included in (8), so that $q_1 > 0$, data in advance of $t = \lfloor r_2 T \rfloor$ could be used in calculating the $CBSADF_{r_2}$ statistic, but this would of course not be possible if the procedure was being used in the context of a real-time monitoring exercise. ◇

Just as the limiting null distribution of Hansen's (1995) $CADF$ statistic differs from that of a standard ADF statistic, so will the limiting null distributions of the $CSADF$, $CBSADF$ and $CGSADF$ statistics differ from those of the standard PSY and PWY $SADF$, $BSADF$ and $GSADF$ statistics. To that end, in Theorem 1 we provide a representation for the limiting null distribution of the $CGSADF$ statistic, where following CSS we assume $\xi_t := (\varepsilon_t, \eta_t')'$ to be a $(m+1)$-dimensional martingale difference sequence with constant unconditional variance matrix (the precise set of assumptions on ξ_t are given in Assumptions A.1–A.2 in Appendix A.3). This limiting distribution is shown to depend on the long-run correlation parameter, ϱ^2. Corresponding representations for the limiting null distributions of the $CSADF$ and $CBSADF$ test statistics are discussed in subsequent remarks. All of these distributions depend on the long-run correlation parameter, ϱ^2.

Theorem 1. *Let $\{y_t, w_t\}$ be generated according to* (1) *and* (2) *and let Assumptions A.1 and A.2 hold. Then under the unit root null hypothesis, H_0 of* (3), *it holds that*

$$CGSADF \overset{w}{\to} \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ \frac{\int_{r_1}^{r_2} Q(s) dP(s)}{\left(\int_{r_1}^{r_2} Q(s)^2 ds \right)^{1/2}} \right\} =: CGSADF_\infty \tag{21}$$

where $Q(s) := b(1)' \Psi(1) B_\eta(s) + B_\varepsilon(s)$ and $P(s) := B_\varepsilon(s)/\sigma_\varepsilon$.

Remark 4.2. The limiting distribution in (21) depends on both the minimum window size, r_0, and the long-run correlation coefficient ϱ^2. The latter dependence can be seen more clearly by writing the limiting distribution in (21) in an alternative form. To illustrate, the limiting null distribution of the full sample $CADF$ statistic of Hansen (1995) can be obtained from Eq. (A.17) in Appendix A.3 and is given by $CADF \overset{w}{\to} \frac{\int_0^1 Q(s) dP(s)}{\left(\int_0^1 Q(s)^2 ds \right)^{1/2}}$. An equivalent representation for this distribution is given in equation (15) of Hansen (1995, p.1154) which can be obtained from the form given here on noting that $b(1)' \Psi(1) B_\eta(s)$ is a scalar Brownian motion and using standard rotation arguments; see for example McCabe and Tremayne (1993, p.173). Applying the same arguments to the $CGSADF_\infty$ limit we can equivalently re-write the limiting null distribution of the $CGSADF$ statistic as

$$CGSADF \overset{w}{\to} \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ \varrho \frac{\int_{r_1}^{r_2} W_1(s) dW_1(s)}{\left(\int_{r_1}^{r_2} W_1^2(s) ds \right)^{1/2}} + (1 - \varrho^2)^{1/2} \frac{\int_{r_1}^{r_2} W_1(s) dW_2(s)}{\left(\int_{r_1}^{r_2} W_1^2(s) ds \right)^{1/2}} \right\} \tag{22}$$

where $W_1(\cdot)$ and $W_2(\cdot)$ are independent standard Brownian motions. ◇

Remark 4.3. The limiting null distributions of the $CSADF$ and $CBSADF$ statistics follow as immediate corollaries of Theorem 1 fixing $r_1 = 0$ and $r_2 = 1$, respectively, in (21), or equivalently in (22). In what follows we denote these distributions as $CSADF_\infty$ and $CBSADF_\infty$ respectively. These depend on both r_0 and ϱ^2. ◇

Remark 4.4. The limiting null distribution of the (covariate unaugmented) $GSADF$ statistic of PSY (where, in view of footnote , it is assumed that an autoregressive lag length of p_v is used in the context of the ADF regression, (11)) obtains as a special case of the limiting distribution in (22) setting $\varrho = 1$. Limiting null distributions for the $CSADF$ and $CBSADF$ statistics obtain under this restriction fixing $r_1 = 0$ and $r_2 = 1$, respectively in (22). ◇

Because of the dependence of the limiting null distributions of the $CGSADF$, $CSADF$ and $CBSADF$ statistics on the long-run correlation parameter, ϱ^2, and on the minimum window size, r_0, asymptotic tests based on these statistics will require tabulations of asymptotic critical values across a range of values of both ϱ^2 and r_0. While r_0 is a tuning parameter chosen by the practitioner, the true value of ϱ^2 is, however, unknown and so would need to be estimated. This can be done using the approach outlined in Section 3.1. However, as we will subsequently show in our Monte Carlo simulation results in Section 5, the resulting tests have very poor finite sample size properties, echoing the findings of CSS for the full sample asymptotic test based on the $CADF$ statistic. Consequently, we do not recommend the use of such asymptotic tests and so we also develop bootstrap-based implementations of these tests which do not require the practitioner to estimate the value of ϱ^2. A detailed description of how the covariate-augmented versions of the bubble detection tests of PWY and PSY developed in Section 4 can be implemented using bootstrap methods is outlined in Algorithm A.1 in Appendix A.1 where we denote the bootstrap analogues of the $CSADF$, $CBSADF$ and $CGSADF$ statistics as $CSADF^*$, $CBSADF^*$ and $CGSADF^*$, respectively.

Remark 4.5. With only minor modifications Algorithm A.1 can be used to implement bootstrap analogues of the (non-covariate augmented) original tests of PWY and PSY in (12)–(14). Specifically, this can be done by: omitting the covariate regressors, $\{w_{t-k}\}_{k=-q_1}^{q_2}$, from Eq. (A.1) in Step 1 and omitting their bootstrap analogues, $\{w_{t-k}^*\}_{k=-q_1}^{q_2}$, from Eq. (A.4) in Step 5; ignoring Steps 2 and 4; only re-sampling from $\tilde{\varepsilon}_t$ in Step 3; and then calculating bootstrap analogues of $SADF$, $BSADF$ and $GSADF$ statistics, denoted as $SADF^*$, $BSADF^*$ and $GSADF^*$, respectively, in Step 8. In the light of footnote , the lag length would need to changed from p to p_v in Steps 1, 6 and 8 and in the construction of the original $SADF$, $BSADF$ and $GSADF$ statistics. The asymptotic validity of the bootstrap PWY and PSY tests obtained in this way can be established in the same way as is done for the corresponding covariate augmented bootstrap tests below. ◇

In Theorem 2, we now detail the large sample behaviour of the bootstrap $CSADF^*$, $CBSADF^*$ and $CGSADF^*$ statistics from Algorithm A.1.

Theorem 2. *Let $\{y_t, w_t\}$ be generated according to* (1) *and* (2) *and let the conditions of Theorem* 1 *hold. It then holds that* $CGSADF^* \xrightarrow{w}_p CGSADF_\infty$, $CSADF^* \xrightarrow{w}_p CSADF_\infty$ *and* $CBSADF^* \xrightarrow{w}_p CBSADF_\infty$, *where* '$\xrightarrow{w}_p$' *denotes weak convergence in probability, as* $T \to \infty$.

A comparison of the results for the bootstrap $CSADF^*$, $CBSADF^*$ and $CGSADF^*$ statistics in Theorem 2 with those given for the $CSADF$, $CBSADF$ and $CGSADF$ statistics in Theorem 1 demonstrates the usefulness of the bootstrap; as the number of observations increases, the bootstrapped statistics have the same first-order limiting null distributions as the corresponding original test statistic. From this result it follows, using the same arguments as in the proof of Theorem 5 in Hansen (2000), that for each of the tests the bootstrap p-values are (asymptotically) uniformly distributed under the unit root null hypothesis, H_0 of (3), leading to tests with (asymptotically) correct size, thereby establishing the asymptotic validity of the bootstrap tests. It can therefore be seen that this validity result is achieved without the practitioner needing to know the true value of the long-run correlation parameter, ϱ^2, or having to estimate it. Moreover, the bootstrap automatically accounts for the dependence of the limiting null distributions of the $CSADF$, $CBSADF$ and $CGSADF$ statistics on the user-chosen minimum window size, r_0.

Remark 4.6. In practice the cdfs $G^*_{i,T}(\cdot)$, $i = 1, 2, 3$, required in Step 9 of Algorithm A.1 will be unknown but can be approximated in the usual way through numerical simulation. To illustrate for the case of the $CGSADF^*$ statistic, this is achieved by generating B bootstrap (conditionally) independent statistics, say $CGSADF^*_b$, $b = 1, \ldots, B$, each computed as in Algorithm A.1 above. The simulated bootstrap p-value for the test is then computed as $\tilde{p}^*_{3,T} = B^{-1} \sum_{b=1}^{B} \mathbb{I}(CGSADF^*_b > CGSADF)$, where $\mathbb{I}(\cdot)$ denotes the indicator function taking the value 1 when its argument is true and 0 otherwise, and is such that $\tilde{p}^*_{3,T} \xrightarrow{a.s.} p^*_{3,T}$ as $B \to \infty$, where $\xrightarrow{a.s.}$ denotes almost sure convergence. An approximate standard error for $\tilde{p}^*_{3,T}$ is given by $(\tilde{p}^*_{3,T}(1 - \tilde{p}^*_{3,T})/B)^{1/2}$; see Hansen (1996, p.419). For a discussion on the choice of B see, *inter alia*, Davidson and MacKinnon (2000). Simulated bootstrap critical values can be obtained for the tests. Again illustrating for the case of a test based on the $CGSADF$ statistic, a λ level empirical bootstrap critical value, $cv_{\lambda,B}$ say, can be calculated as the upper tail λ percentile from the order statistic formed from the B bootstrap statistics, $CGSADF^*_b$, $b = 1, \ldots, B$. The resulting bootstrap test, which rejects H_0 of (3) if $CGSADF > cv_{\lambda,B}$, will have asymptotic size that for sufficiently large B will be as close as desired to the given nominal level, λ. ⋄

Remark 4.7. Algorithm A.1 can be used to obtain bootstrap implementations of the covariate augmented versions of the date-stamping procedure of PSY outlined in Remark 4.1. Here, one replaces the asymptotic critical value $cscv^{\theta_T}_{r_2}$ by a corresponding bootstrap critical value obtained using Algorithm A.1 calculated in exactly the same manner as outlined in Remark 4.6. In the practical implementation of date-stamping to the real S&P 500 price dividend ratio PSY replace the asymptotic critical value $scv^{\theta_T}_{r_2}$ required in (16) and (17) with a finite sample critical value obtained by Monte Carlo simulation methods under the assumption that the error term u_t in (1) forms an independent sequence of standard normal random variables; i.e., $u_t \sim NIID(0, 1)$. Given that, like most price data, the S&P data appears to be both heavy-tailed (non-Gaussian) and to contain significant patterns of conditional heteroskedasticity, these simulated critical values are likely to deliver a poor approximation to the true finite sample critical values for the tests used in the sequential procedure, and so the resulting date-stamping estimates could well be biased as a result. Within the context of the standard PSY date-stamping method, therefore, we could replace the simulated critical values approach used by PSY by using bootstrap critical values obtained as above but using the simplified form of Algorithm A.1 detailed in Remark 4.5. As the bootstrap critical values obtained in this way are valid under both non-normality and conditional heteroskedasticity, the resulting date-stamping methods are likely to be more reliable in practice for use with financial price data. ⋄

In what follows, as shorthand notation, we will denote by $CGSADF^B$ the bootstrap test procedure outlined in Algorithm A.1, whereby the original statistic $CGSADF$ is compared to its empirical bootstrap critical value, $cv_{\lambda,B}$. Analogously, the corresponding bootstrap tests based on the $CSADF$ and $CBSADF$ statistics will be denoted as $CSADF^B$ and $CBSADF^B$, respectively.

5. Finite sample simulations

In this section we compare the finite sample size and power properties of the backward recursive $BSADF$-based test of PSY with the covariate augmented version of this test, based on the $CBASDF$ statistic, developed in this paper. We additionally compare the relative efficacy of PSY's date-stamping approach, as outlined in Section 3.2, with the corresponding covariate augmented approach outlined in Remark 4.1. We focus on backward recursive test procedures as these are designed to detect the emergence of an end-of-sample explosive episode which is arguably the application of these methods that will be of greatest interest to practitioners. These are also the tests which we will use in the empirical application we consider in Section 6. We additionally investigated the relative finite sample size and power properties of the covariate and non-covariate augmented versions of the $SADF$-based and $GSADF$-based tests but found these to be qualitatively similar to the results reported here for the $BSADF$-based tests.

5.1. Monte Carlo simulation design

We use Monte Carlo simulation methods to generate data according to the DGP:

$$\Delta y_t = \delta_t y_{t-1} + u_t, \quad t = -99, \ldots, 0, 1, \ldots, T \tag{23}$$

where the time-varying offset parameter is given by $\delta_t = c_t/T$. We discard the first 100 observations to eliminate any start up effects so that our data samples run from $t = 1, \ldots, T$.[3] In Section 5.2 we will report results relating to empirical size which is where the constant unit root null hypothesis, H_0 of (3), holds; that is, $c_t = 0$, $t = -99, \ldots, T$. In Section 5.3 we will investigate the power of the tests to identify locally explosive episodes (i.e. sub-samples where $c_t > 0$), and in Section 5.4 investigate how well covariate-based implementations of the date-stamping method of PSY compare with the basic PSY date-stamping method, both in terms of their frequencies of detecting the start of such an episode and how accurately they date this start point.

Adopting the same simulation design considered in CSS, the errors u_t are generated according to the following DGP:

$$u_t = \alpha_1 u_{t-1} + v_t \tag{24}$$

$$v_t = \beta w_t + \varepsilon_t, \tag{25}$$

$$w_{t+1} = \lambda w_t + \eta_t, \tag{26}$$

and where we generate $\xi_t := (\varepsilon_t, \eta_t)'$ as a sequence of bivariate $NIID(0, \Sigma)$ random variables with $\Sigma := \begin{bmatrix} 1 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & 1 \end{bmatrix}$

The finite sample size and power performance of all of the tests considered depend on the (long-run) squared correlation between the error term v_t and the stationary covariate w_t, and therefore on the parameters λ and β. We follow CSS and use values of $\lambda, \beta \in \{-0.8, -0.5, 0.5, 0.8\}$ and set $\alpha_1 = 0.2$ and $\sigma_{\varepsilon\eta} = 0.4$. We additionally report results for the case where $\beta = 0$, $\lambda = 0.8$ such that $\varrho^2 = 1$. This last case allows us to examine the impact on our procedures of including an irrelevant covariate.[4]

The bootstrap $CBSADF^B$ tests are implemented according to Algorithm A.1 with the omission of Step 6 as discussed in Remark A.2, as we found this step was only necessary when implementing the $BSADF^B$ test. In practice the number of lags of Δy_t, p, together with the number of leads and lags of the covariate w_t, q_1 and q_2 respectively, needed to estimate the $CADF(p, q_1, q_2)_{r_1}^{r_2}$ sub-sample statistics and to implement Step 1 of Algorithm A.1 are unknown. These are jointly determined from the full sample estimation of (8) using the conventional Bayes Information Criterion (BIC) setting a maximum lag length of 4 for p and a maximum lag length of 2 for q_1 and q_2. For the non-covariate augmented bootstrap test based on the $BSADF$ statistic proposed in Remark 4.5, denoted $BSADF^B$ with an obvious notation, the number of lags of Δy_t used in the $ADF(p_v)_{r_1}^{r_2}$ statistics is analogously determined by estimating (11) over the full available sample of data again using the BIC with a maximum lag length of 4.[5] The autoregressive lag order, ℓ, for w_t in Step 2 was chosen using the BIC with a maximum lag length of 4. For the non-bootstrap tests reported, asymptotic critical values were obtained by direct simulation of the limiting null distribution of the test statistic; we will denote the asymptotic tests based on the $BSADF$ and $CBSADF$ statistics as $BSADF^\infty$ and $CBSADF^\infty$, respectively. In the case of the $CBSADF^\infty$ test, these were based on the estimate of ϱ^2 defined in (9) implemented using a Parzen kernel and Andrews' (1991) automatic bandwidth estimator. Following PSY, the minimum window fraction used for all of the recursive test procedures is set at $r_0 = 0.01 + \frac{1.8}{\sqrt{T}}$.

The size and power simulations in Sections 5.2 and 5.3, respectively, are computed using 5000 Monte Carlo replications, while those relating to sequential monitoring in Section 5.4 are considerably more time-consuming and so are based on 2000 Monte Carlo replications. All bootstrap tests are performed with $B = 399$ bootstrap replications. All computations are performed on Gauss 17 using the RNDN random number generator. All tests are performed at a nominal 5% significance level. All of the reported results pertain to methods which allow y_t and w_t to have non-zero means (cf. Remark 2.1) by including a constant in the sub-sample ADF regressions used to compute the $ADF(p)_{r_1}^{r_2}$ statistics, and in the sub-sample covariate augmented ADF regression used to compute the $CADF(p, q_1, q_2,)_{r_1}^{r_2}$ statistic, and using demeaned values of y_t and w_t in Steps 1 and 2 of Algorithm A.1, as outlined in Appendix A.2.

5.2. Empirical size

We first examine the size properties of our proposed tests. Data are generated according to (23)–(26) with $c_t = 0$ for all $t = -99, \ldots, T$. Table 1 reports the size of both the asymptotic and bootstrap tests based on the $BSADF$ and $CBSADF$ statistics for sample sizes of $T = 250$ (Panel A) and $T = 400$ (Panel B). For reference we also report the value of the (long-run) squared correlation parameter, ϱ^2, implied by the parameters β and λ in each scenario.

The standard PSY test based on asymptotic critical values, $BSADF^\infty$, is in general seen to be under-sized, even for $T = 400$, with the degree of under-size clearly depending on the parameter λ characterising the degree of persistence of the covariate. While the degree of under-sizing is relatively modest for $\lambda = 0.8$, for $\lambda = 0.5$ the $BSADF^\infty$ test is significantly under-sized. As the power simulations in Section 5.3 will show, this under-sizing phenomenon leads to the $BSADF^\infty$ test having very poor power properties in particular for $\lambda = 0.5$. In contrast, our suggested bootstrap implementation of PSY's test, $BSADF^B$, displays empirical size much closer to the nominal 5% level throughout the nine cases and for both sample sizes considered, albeit it is slightly over-sized in most of the cases reported.

---

[3] The initial conditions on y_t and w_t are set as $w_{-100} = 0$ and $y_{-100} = 100$. The latter is done so that where bubble episodes are generated in our simulated data they will typically be upwardly explosive and therefore representative of the empirical series to which the tests will be applied in practice.

[4] We also considered the case where $\beta = 0$, $\lambda = 0.5$, but the results were near identical to the case where $\beta = 0$, $\lambda = 0.8$, which is unsurprising given that $\varrho^2 = 1$ in both cases. We therefore only report results for $\beta = 0$, $\lambda = 0.8$.

[5] In calculating bootstrap critical values for the $BSADF^B$ test, the value of the lag truncation parameter p_v in Step 1 of the modified algorithm outlined in Remark 4.5 is chosen according to the Schwert (1989) criterion $p_v = \lfloor 4(T/100)^{1/4} \rfloor$. For this DGP, v_t in (11) follows an $ARMA(1, 1)$ process and we found this to deliver considerably better size control than selecting the lag length by BIC (which resulted in very liberal bootstrap tests) but without sacrificing power; cf. Palm et al. (2008).

Table 1
Empirical size of nominal 0.05-level asymptotic and bootstrap $BSADF$ and $CBSADF$ tests.

Panel A: $T = 250$

β	λ	$BSADF^\infty$	$CBSADF^\infty$	$BSADF^B$	$CBSADF^B$	ϱ^2
0.8	0.8	0.038	0.129	0.057	0.062	0.335
0.5	0.8	0.044	0.103	0.056	0.067	0.432
−0.5	0.8	0.041	0.073	0.065	0.055	0.000
−0.8	0.8	0.047	0.091	0.055	0.053	0.026
0.8	0.5	0.011	0.061	0.055	0.052	0.556
0.5	0.5	0.011	0.054	0.057	0.053	0.700
−0.5	0.5	0.008	0.020	0.053	0.045	0.300
−0.8	0.5	0.010	0.040	0.053	0.045	0.057
0.0	0.8	0.009	0.025	0.062	0.047	1.000

Panel B: $T = 400$

β	λ	$BSADF^\infty$	$CBSADF^\infty$	$BSADF^B$	$CBSADF^B$	ϱ^2
0.8	0.8	0.033	0.105	0.055	0.050	0.335
0.5	0.8	0.038	0.081	0.056	0.052	0.432
−0.5	0.8	0.033	0.070	0.060	0.052	0.000
−0.8	0.8	0.038	0.089	0.057	0.057	0.026
0.8	0.5	0.009	0.050	0.049	0.043	0.556
0.5	0.5	0.011	0.043	0.051	0.043	0.700
−0.5	0.5	0.006	0.013	0.059	0.041	0.300
−0.8	0.5	0.009	0.034	0.055	0.045	0.057
0.0	0.8	0.006	0.016	0.055	0.040	1.000

Notes: Data simulated according to (23)-(26) where $\xi_t := (\varepsilon_t, \eta_t)'$ is a sequence of bivariate $NIID(0, \Sigma)$ random variables with $\Sigma := \begin{bmatrix} 1 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & 1 \end{bmatrix}$. Bootstrap critical values are obtained using 399 bootstrap replications.

Similarly to $BSADF^\infty$, the asymptotic critical value based $CBSADF^\infty$ test displays poor size control. For $\lambda = 0.8$, $CBSADF^\infty$ is heavily over-sized while for $\lambda = 0.5$ it is significantly under-sized for the negative values of β reported. This poor finite sample size control is largely driven by imprecise estimation of the nuisance parameter ϱ^2. In common with the findings of CSS for the full sample $CADF(p, q_1, q_2)$ statistic of Hansen (1995), we found that the asymptotic null distribution of the $CBSADF$ statistic is highly sensitive to the value of ϱ^2. As such, an imprecise estimate of ϱ^2 will likely lead to the use of a poor approximation to the true limiting critical value with inevitable size distortions. Because of this we will not consider the use of the $CBSADF^\infty$ test further in this paper. These size distortions are largely eliminated by using the bootstrap $CBSADF^B$ test which displays good size control throughout.

Overall, it can be seen that the bootstrap $BSADF^B$ and $CBSADF^B$ tests both control finite sample size well, each avoiding the often significant size distortions seen with the corresponding asymptotic tests, $BSADF^\infty$ and $CBSADF^\infty$, respectively.

5.3. Empirical power

We next turn to an examination of finite sample power. To that end, simulation data are generated according to Eqs. (23)–(26) under the alternative hypothesis of an end-of-sample explosive episode by setting $c_t = 0$ for $t = -99, \ldots, \lfloor 0.8T \rfloor$, and $c_t = c > 0$ for $t = \lfloor 0.8T \rfloor + 1, \ldots, T$. The series $\{y_t\}$ therefore follows a unit root process for $t = 1, \ldots, \lfloor 0.8T \rfloor$ and is then subject to locally explosive behaviour for the remainder of the sample. We compute finite sample power over relevant grids of 20 values of c for each of the 8 pairings of β and λ considered in the size simulations in the previous section.

Fig. 1 reports finite sample power curves for the bootstrap $CBSADF^B$ (solid black line) and $BSADF^B$ (solid red line) tests together with the asymptotic $BSADF^\infty$ (dashed red line) test, for each pairing of β and λ for $T = 250$. It is clearly seen from Fig. 1 that in the eight cases where $\beta \neq 0$, such that the covariate is relevant, the $CBSADF^B$ test displays superior power curves to both the $BSADF^B$ and $BSADF^\infty$ tests when $\beta \neq 0$ in all but the case of panel (g) where $\beta = -0.5$ and $\lambda = 0.5$. In this latter case the power curve of the $BSADF^B$ test lies very marginally above that of the $CBSADF^B$ test for values of c up to about $c = 1.3$ after which the power curves cross. The power gains from using the $CBSADF^B$ test are particularly strong for each of the cases graphed in panels (a)–(e) where $\lambda = 0.8$. As an example, for the case considered in panel (d) where $\beta = -0.8$ and $\lambda = 0.8$, we see that for $c = 2$ the $BSADF^B$ and $BSADF^\infty$ tests both have power of about 31% while the $CBSADF^B$ test has double the power at about 62%. The power curve for the case where $\beta = 0$, illustrated in panel (i), shows that the inclusion of an irrelevant covariate effects only a very small loss of power for the $CBSADF^B$ test vis-à-vis the $BSADF^B$ test. Fig. 1 also shows that the power of the asymptotic $BSADF^\infty$ test is very much lower than both the $CBSADF^B$ and $BSADF^B$ tests for the cases where $\lambda = 0.5$ (panels (e)–(g)); this effect is largely due to the considerable under-size of the $BSADF^\infty$ test in these cases noted in Section 5.2.

Fig. 2 reports equivalent finite sample power curves for each pairing of β and λ for a sample of size $T = 400$. The pattern of results for this sample size is qualitatively the same as those discussed in Fig. 1 for $T = 250$. In particular, across the eight cases where $\beta \neq 0$ the $CBSADF^B$ test again displays by far the best power performance overall in all but the case of panel (g) where $\beta = -0.5$ and $\lambda = 0.5$, although the power differential between $CBSADF^B$ and $BSADF^B$ in this scenario is really very small. Again including an irrelevant covariate, panel (i) where $\beta = 0$, only leads to a small loss in power compared to the $BSADF^B$ test.
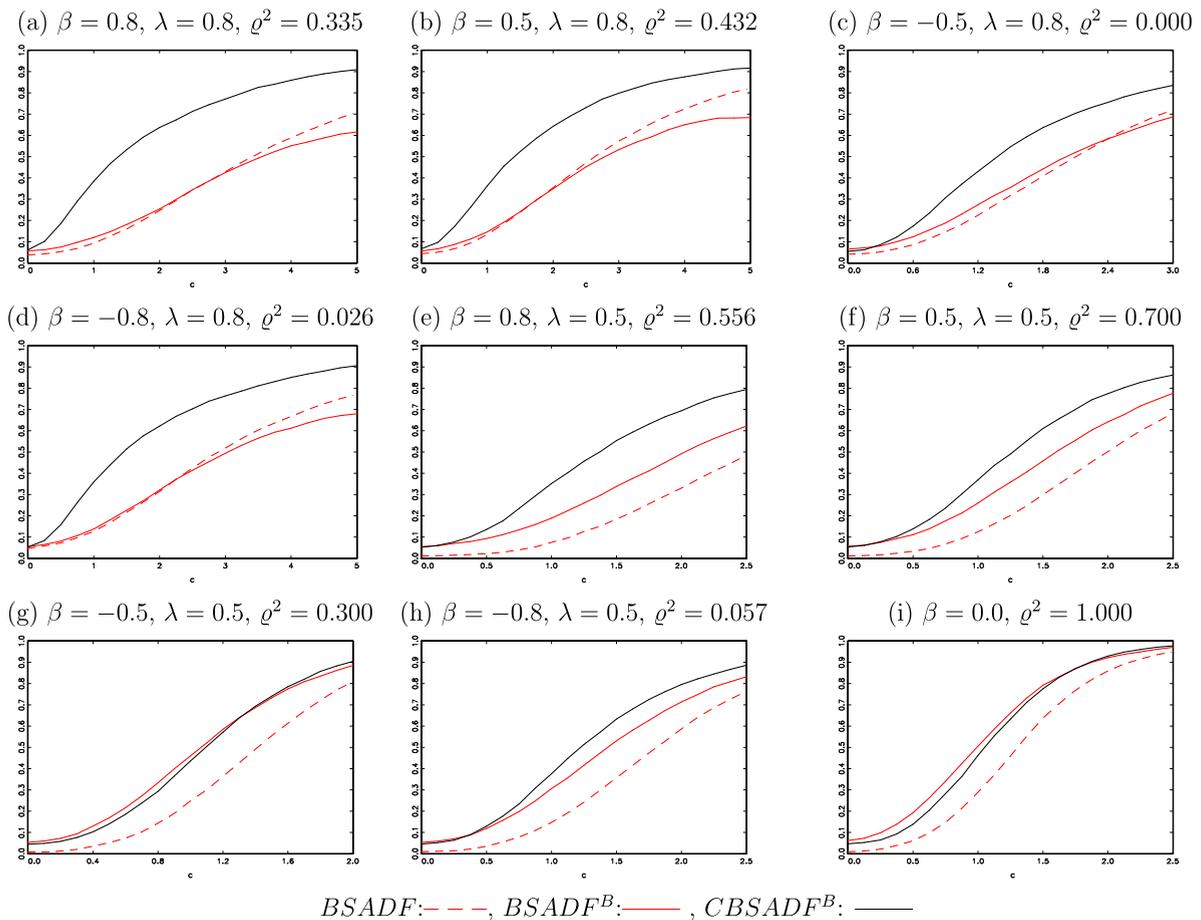
Fig. 1. Finite sample power of nominal 0.05-level tests.

Notes: Data simulated according to (23)–(26) where $\xi_t := (\varepsilon_t, \eta_t)'$ is a sequence of bivariate $NIID(0, \Sigma)$ random variables with $\Sigma := \begin{bmatrix} 1 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & 1 \end{bmatrix}$. Bootstrap critical values are obtained using 399 bootstrap replications. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As a final comment, it is worth noting that while the power gains offered by the $CBSADF^B$ test relative to $BSADF^B$ can be substantial, they are not as large as those reported in the simulation studies of CSS and Hansen (1995) in the context of a full sample test of the null hypothesis of a unit root against the alternative of stationarity. There are several reasons why this is the case. First, our DGP assumes a deviation from the null hypothesis for only a relatively small sub-sample of the data, whereas CSS and Hansen (1995) assume a deviation from the null hypothesis that holds across the full sample under the alternative. Second, our tests are based on the supremum of statistics from a sequence of sub-sample regressions rather than a single statistic based on a full sample regression. Moreover, due to the highly asymmetric property of the distribution of Dickey Fuller type statistics the power of left-tailed and right-tailed test procedures are not directly comparable.

5.4. Detection time

Finally, we turn to an examination of the relative efficacy of sequential procedures based on the $CBSADF^B$, $BSADF^B$ and $BSADF^\infty$ tests to date-stamp the onset of an explosive episode. As discussed in Section 3.2, one of the main advantages of the $BSADF$ and $CBSADF$ statistics is their ability to detect end-of-sample explosive episodes, therefore making them useful for detecting the inception date of an explosive bubble if applied in real-time in a sequential manner according to (16). To that end, we perform a Monte Carlo simulation study in which the $BSADF^\infty$, $BSADF^B$ and $CBSADF^B$ tests are applied sequentially, as if in a real-time monitoring exercise, to a sample of data containing an end-of-sample explosive episode in order to assess their relative efficacy in date-stamping the beginning of the episode.

Specifically, data were generated according to (23)–(26) with $T = 300$ setting $c_t = 0$ for $t = -99, \ldots, 250$, and $c_t = c > 0$ for $t = 251, \ldots, 300$, so that the data follows a unit root process from $t = 1, \ldots, 250$ before an explosive episode manifests in the data from $t = 251, \ldots, 300$. This was done for each of the 9 pairings of β and λ considered in Sections 5.2 and 5.3. So as to set a comparable
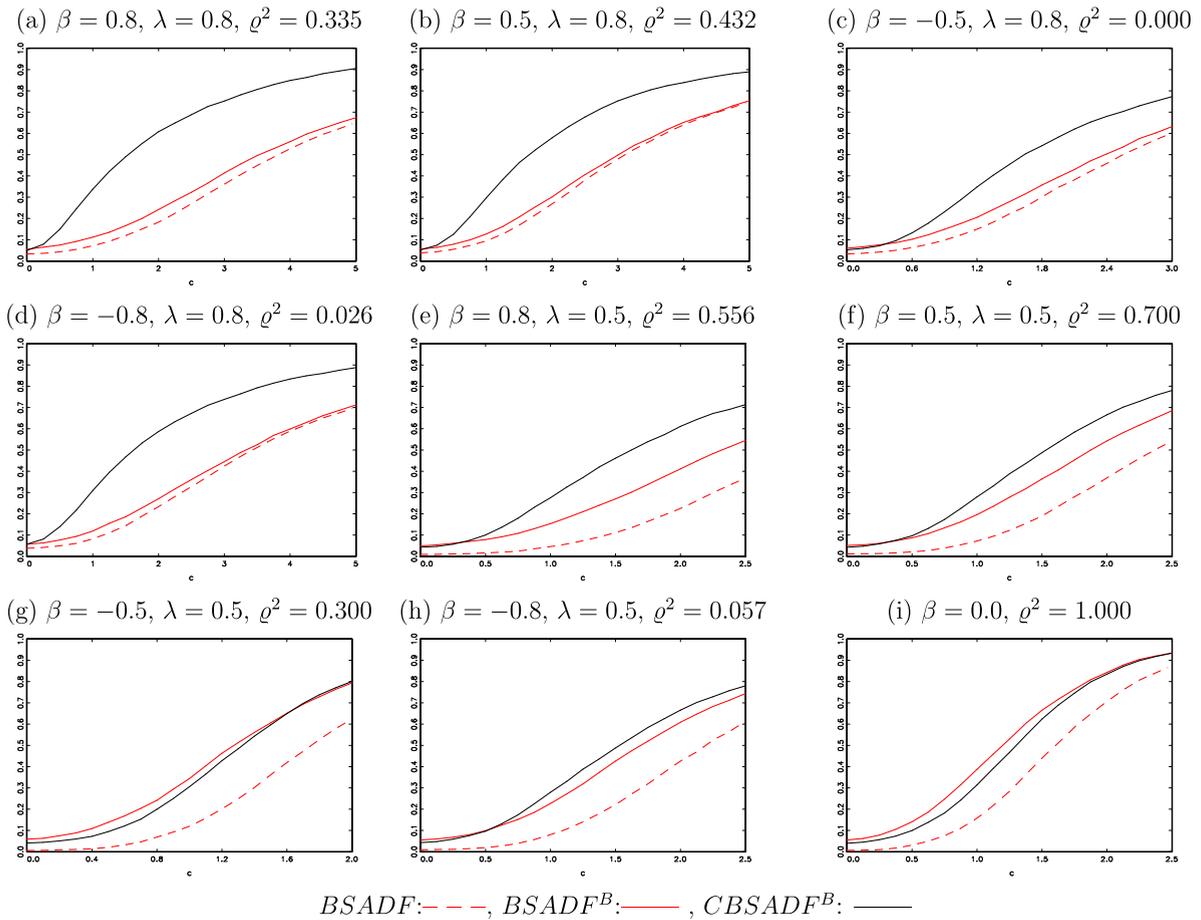
Fig. 2. Finite sample power of nominal 0.05-level tests with data simulated according to (23)–(26) with $T = 400$.

Notes: Data simulated according to (23)–(26) where $\xi_t := (\varepsilon_t, \eta_t)'$ is a sequence of bivariate $NIID(0, \Sigma)$ random variables with $\Sigma := \begin{bmatrix} 1 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & 1 \end{bmatrix}$. Bootstrap critical values are obtained using 399 bootstrap replications.

power benchmark across the various simulation DGPs considered, for a given (β, λ) pair, the value of c was chosen so that the value of the explosive offset $\delta_t = \delta > 0$ corresponded as closely as possible to the explosive offset which lead to the $CBSADF^B$ test having 80% power in the power curves in Fig. 1.[6] We begin by applying the tests to the sample indexed by $t = 1, \ldots, 251$, then on the sample from $t = 1, \ldots, 252$, and so on, until finally the tests are applied to the full sample of data. We record the first estimated rejection date, $\hat{t}_e$, implied by each test when run sequentially in this manner. For those cases where no test in this monitoring period rejects the null hypothesis we set $\hat{t}_e = 301$. For each (β, λ) pair and for each test, Table 2 reports the empirical rejection frequency [ERF] which is the frequency with which the null hypothesis is rejected by at least one test in the sequential procedure, together with the median (taken over the 2000 Monte Carlo replications) value of $\hat{t}_e$, denoted $\hat{t}_e^M$,[7] and (in parentheses) the associated median average deviation (MAD).

Unsurprisingly, the rankings of ERFs for the different sequential testing procedures is qualitatively similar to the ranking between these when applied as one-shot tests observed in Section 5.3. In particular, the procedure based on the $CBSADF^B$ test delivers at least one rejection within the monitoring period more frequently than do the other two procedures in all but the final three scenarios, the last of which is where an irrelevant covariate is included ($\beta = 0$). In these final three scenarios, corresponding to panels (g) - (i) of Figs. 1 and 2, the procedure based on the $BSADF^B$ test has a marginally higher ERF than the $CBSADF^B$ test. In all cases the procedures based on the $CBSADF^B$ and $BSADF^B$ tests both display higher ERFs than the procedure based on the $BSADF^\infty$ test.

What is of most relevance is a comparison of the median estimate of t_e across the three procedures. In all but two scenarios (one of which is where the covariate is irrelevant) the median detection date for the procedure based on the $CBSADF^B$ test is lower,

---

[6] We set this power benchmark reasonably high so that each of the sequential tests detects a bubble in the majority of the simulation replications. The relative performances of the three procedures for the values of c reported are representative of the results for other choices of c not reported here to save space.

[7] Setting $\hat{t}_e = 301$ for non-rejections and examining the median gives a small (unfair) advantage to the less powerful tests considered. However, if we were to use the mean, or to only report the median or mean taken across only those replications where a given procedure delivered a rejection in the monitoring period, this would give an even greater (unfair) advantage to the less powerful tests.

Table 2
Empirical rejection frequency, median first rejection date and median average deviation of monitoring procedures. Monitoring commences at $t = 251$.

β	λ	c	Empirical rejection frequency			$\hat{t}_e^M$ (MAD)		
			$BSADF^\infty$	$BSADF^B$	$CBSADF^B$	$BSADF^\infty$	$BSADF^B$	$CBSADF^B$
0.8	0.8	3.9	0.70	0.74	**0.87**	282 (19)	279 (19)	**265** **(10)**
0.5	0.8	3.6	0.78	0.80	**0.86**	276 (16)	275 (16)	**265** **(10)**
−0.5	0.8	3.0	0.78	0.83	**0.85**	276 (16)	272 (14)	**267** **(10)**
−0.8	0.8	4.2	0.78	0.80	**0.87**	276 (16)	274 (15)	**265** **(10)**
0.8	0.5	3.0	0.68	0.85	**0.87**	286 (15)	272 (13)	**267** **(10)**
0.5	0.5	2.4	0.71	0.86	**0.87**	284 (17)	271 (12)	**269** **(11)**
−0.5	0.5	1.9	0.78	**0.93**	0.92	279 (14)	**268** **(9)**	269 (9)
−0.8	0.5	2.4	0.75	**0.90**	0.89	280 (15)	269 (10)	**268** **(9)**
0.0	0.8	1.5	0.67	**0.89**	0.86	286 (15)	**271** **(12)**	273 (12)

Notes: Data simulated according to (23)–(26) with $T = 300$ setting $c_t = 0$ for $t = -99, \ldots, 250$, and $c_t = c > 0$ for $t = 251, \ldots, 300$.

and hence closer to the true bubble inception date, t_e, than that of the other two procedures. Comparing the procedure based on the $CBSADF^B$ test to the next most powerful procedure based on the $BSADF^B$ test we see that the median detection date for the former can be up to 14 periods sooner than the latter. In the one scenario where the procedure based on $CBSADF^B$ does not show a smaller median detection date than the procedure based on $BSADF^B$ when a relevant covariate is included the difference in the median rejection date for the tests is only a single period. When an irrelevant covariate is included ($\beta = 0$), the median rejection date for the $CBSADF^B$ test is only two periods behind that of $BSADF^B$.

Finally, the MAD of the three procedures is particularly striking. Not only does the procedure based on the $CBSADF^B$ test reject sooner on average than the other procedures, it also has much low variability in its estimated inception dates than do the other procedures in all but two scenarios, where in each of these two cases the MAD of the $CBSADF^B$- and $BSADF^B$-based procedures coincide.

Overall, therefore, where a relevant covariate is included the $CBSADF^B$-based procedure can detect an ongoing bubble both earlier, and with less variability, than both the $BSADF^\infty$-based and $BSADF^B$-based detection procedures tests.

6. Empirical application

We present results of an empirical application applying our covariate augmented bubble detection procedures to data on the real S&P 500 price dividend ratio obtained from Robert Shiller's website (http://www.econ.yale.edu/shiller/data.htm). The data are monthly (sampled, in common with the covariate data we will use, on the final day of each month) from January 1960–December 2010 ($T = 612$). A plot of the price dividend ratio is given in Fig. 3(a). This is a subset of the data utilised by PSY in their own empirical application chosen based on data availability of our chosen covariates.

The presence of two major historical bubbles in this dataset is widely accepted, namely the bubble in the lead-up to the Black Monday crash, and the dotcom bubble, with PSY identifying both of these episodes in their sequential date-stamping exercise. As noted by PSY, this can be viewed as a (pseudo) real-time monitoring exercise, with the estimated origination dates also indicating the date a test would have identified these episodes if applied in real-time. In particular, if the estimated origination date of a particular bubble identified by one procedure is N months earlier than that identified by another, this can be interpreted as the former procedure being able to detect the onset of an explosive episode N months earlier than the latter.

The first two potential covariates we consider are Moody's Seasoned Aaa Corporate Bond Yield (Aaa) and the 10 Year US Treasury Constant Maturity Rate (GS10). Monthly data on these variables from January 1960–December 2010 were obtained from the Federal Reverse Bank of St. Louis webpage (https://fred.stlouisfed.org). Many common models for asset price bubbles imply that the magnitude of a bubble episode will be a function of the risk-free rate of interest. As such, we might expect that these two variables that are closely related to the risk-free interest rate will be suitable covariates when testing for a bubble episode in the price dividend ratio.

A further potential covariate we consider is the VXO volatility index. Monthly data on this index from January 1986 to December 2010 were obtained from Yahoo Finance. The VXO index measures the market's expectation of 30-day volatility and it is constructed by using the implied volatilities on the S&P 100 index options. In their seminal work, Fleming et al. (1995) provide evidence in support of the argument that there is a tendency of implied volatility to rise after large sell-offs and fall after large rallies. As such, implied volatility could be a potentially useful covariate when testing for a bubble in the price dividend ratio.

In addition we also investigate using the predictor variables in the latest version of the Welch and Goyal (2008) data set[8] which are commonly employed in many empirical applications on return predictability. In particular, we consider Moody's Seasoned Baa Corporate Bond Yield (Baa), the Book-to-Market Ratio for the Dow Jones Industrial Average (BM), the return on long-term AAA- and BAA-rated Corporate Bonds (CORPR), the AAA Credit Spread (CREDSA),[9] the Long Term Rate of Returns on the Treasury Bond (LTR), the Long Term Yield on Long-Term U.S. Bonds (LTY), Stock Variance (SVAR), and the Treasury Bill rate (TBL). Like the AAa and GS10 covariates, we use data on the covariates from the Welch and Goyal (2008) dataset from January 1960 to December 2010.

---

[8] These data can be found at https://sites.google.com/view/agoyal145.

[9] Calculated as the difference between the AAA-rated Corporate Bond Yield (Aaa) and the Long Term Yield on Long-Term U.S. Bonds (LTY).
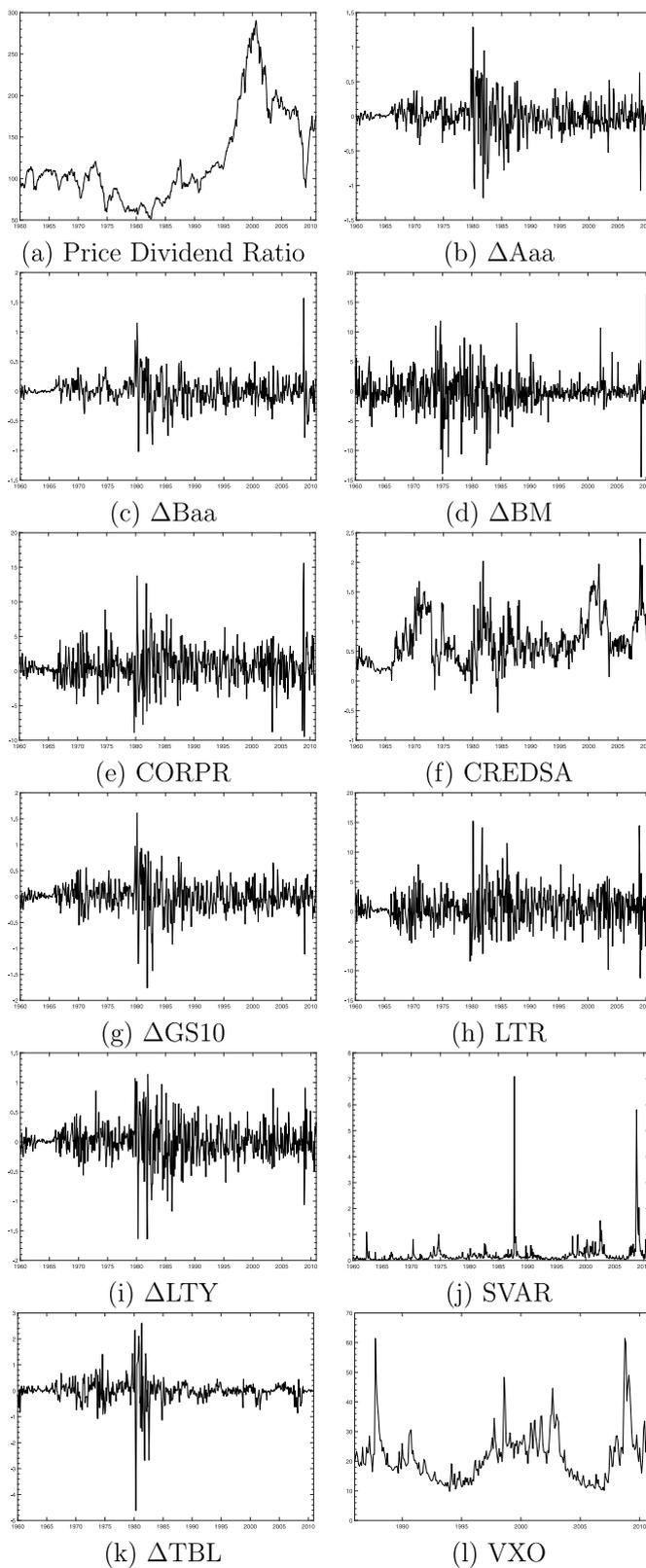
Fig. 3. Plots of variables used in empirical application.

Table 3
Standard unit root tests applied to covariates.

Variable	ADF Stat	p-value	Variable	ADF Stat	p-value
Aaa	−1.36	0.606	ΔGS10	−18.17	0.000
ΔAaa	−18.38	0.000	LTR	−24.02	0.000
Baa	−1.45	0.557	LTY	−1.49	0.538
ΔBaa	−15.71	0.000	ΔLTY	−23.13	0.000
BM	−1.29	0.637	SVAR	−11.22	0.000
ΔBM	−24.12	0.000	TBL	−2.29	0.177
CORPR	−18.49	0.000	ΔTBL	−6.09	0.000
CREDSA	−4.72	0.000	VXO	−5.16	0.000
GS10	−1.38	0.593			

Table 4
Full sample tests for bubbles in the price dividend ratio: January 1960–December 2010.

	Test Statistic	Bootstrap p-value
$GSADF^{\infty}$	3.171	0.022
$GSADF^{B}$	3.171	0.026
$CGSADF^{B}(\Delta$Aaa)	3.636	0.005
$CGSADF^{B}(\Delta$Baa)	3.827	0.003
$CGSADF^{B}(\Delta$BM)	4.885	0.000
$CGSADF^{B}($CORPR)	3.810	0.004
$CGSADF^{B}($CREDSA)	3.223	0.014
$CGSADF^{B}(\Delta$GS10)	3.614	0.007
$CGSADF^{B}($LTR)	3.591	0.006
$CGSADF^{B}(\Delta$LTY)	3.470	0.006
$CGSADF^{B}($SVAR)	4.187	0.001
$CGSADF^{B}(\Delta$TBL)	3.465	0.007
$CGSADF^{B}($VXO)	5.381	0.000

Notes: Asymptotic p-values based on 10,000 Monte Carlo replications and bootstrap p-values based on 1999 bootstrap replications.

The $CBSADF^{B}$ test requires the potential covariates to be weakly dependent. To check this condition we apply standard full sample left-tailed ADF tests to all candidate covariates. In all cases the ADF regressions include a constant and the lag order is determined using the BIC. The results of these tests are reported in Table 3, where the stated p-value is the default produced by EViews. These results indicate that Aaa, Baa, BM, GS10, LTY and TBL are all $I(1)$ in levels but $I(0)$ in first differences and so the first differences of these variables should be used. The CORPR, CREDSA, LTR, SVAR and VXO variables all appear to be $I(0)$ in levels and so do not require first differencing. Plots of the variables used as covariates can be found in Fig. 3(b)–(l).

We begin by testing for the presence of at least one bubble episode in the price dividend ratio using the $GSADF^{\infty}$, $GSADF^{B}$ and $CGSADF^{B}$ tests. Here and throughout the empirical application all test statistics are constructed using exactly the same settings as in the Monte Carlo simulations in Section 5, including allowing for non-zero means in y_t and w_t, with a minimum window width of $\lfloor r_0 T \rfloor = (0.01 + 1.8/\sqrt{T})T = 50$. The exception to this is when using VXO as a covariate where, due to this covariate only being available for the 300 observations beginning 01/86, the minimum window width is equal to $(0.01 + 1.8/\sqrt{300})300 = 34$. Table 4 reports the $GSADF$ and $CGSADF$ test statistics and the associated p-values.[10] All of the tests can be seen to reject the null hypothesis of no explosive behaviour in the price dividend ratio at the 5% level. In each case a considerably stronger rejection is seen for the covariate augmented $CGSADF^{B}$ tests than for either the $GSADF^{\infty}$ or $GSADF^{B}$ tests; indeed, excepting CREDSA, the covariate augmented tests are able to reject the null hypothesis at the 1% for all of the possible covariates considered.

Given the clear rejection of the null hypothesis of no bubbles by the full sample tests, we next apply, as if in a real-time monitoring procedure, the date-stamping procedures based on our new covariate augmented bootstrap $CBSADF^{B}$ tests, as outlined in Remark 4.7, to estimate the origination and termination dates of the bubbles present in the price dividend ratio and compare these with the estimates based on corresponding procedures using covariate unaugmented tests from PSY, the latter based on either bootstrap or asymptotic critical values, $BSADF^{B}$ and $BSADF^{\infty}$, respectively.[11] We also computed finite sample critical values assuming $u_t \sim NIID(0, 1)$ for 10,000 Monte Carlo replications in the same manner as PSY and denote a test in which the $BSADF$ test statistic is compared to these critical values as $BSADF^{\text{fscv}}$. Using the $CBSADF^{B}$ test to illustrate, we begin by applying the $CBSADF^{B}$ test to the data up to and including January 1980 which indicates whether the $CBSADF^{B}$ test would have rejected the null hypothesis of no bubble if performed at this particular date. We then move forward one month and test using data up to and including February 1980, and so on, until we eventually apply the $CBSADF^{B}$ test to data ending in December 2010. Like PSY we set the minimum bubble duration requirement to be a year and deem a bubble to have terminated only after observing a year of

---

[10] The p-values of the $CGSADF^{B}$ and $GSADF^{B}$ tests are calculated as described in Step 9 of Algorithm A.1 using $B = 1999$ bootstrap replications, and the p-value of the $GSADF^{\infty}$ test is calculated as $(1 - G_{\infty}(GSADF))$ where $G_{\infty}(\cdot)$ denotes the cumulative asymptotic null distribution of the $GSADF^{\infty}$ test simulated from 10,000 Monte Carlo replications.

[11] The bootstrap critical values for $CBSADF^{B}$ and $BSADF^{B}$ are again based on $B = 1999$ replications, while the asymptotic critical values for $BSADF^{\infty}$ are simulated using 10,000 Monte Carlo replications.

Table 5

Estimated origination and termination dates of bubbles identified in the price dividend ratio.

Test	Black Monday		Dotcom Bubble	
	$\hat{t}_e$	$\hat{t}_f$	$\hat{t}_e$	$\hat{t}_f$
$CBSADF^B(\Delta \text{Aaa})$	03/86	12/87	09/95	05/01
$CBSADF^B(\Delta \text{Baa})$	03/86	12/87	09/95	05/01
$CBSADF^B(\Delta \text{BM})$	01/87	12/87	06/95	05/01
$CBSADF^B(\text{CORPR})$	03/86	12/87	09/95	05/01
$CBSADF^B(\text{CREDSA})$	03/87	12/87	08/95	06/09
$CBSADF^B(\Delta \text{GS10})$	03/86	12/87	09/95	05/01
$CBSADF^B(\text{LTR})$	06/86	12/87	09/95	05/01
$CBSADF^B(\Delta \text{LTY})$	06/86	12/87	09/95	04/01
$CBSADF^B(\text{SVAR})$	02/86	01/88	01/96	10/08
$CBSADF^B(\Delta \text{TBL})$	03/87	12/87	07/95	05/01
$CBSADF^B(\text{VXO})$	–	–	07/95	04/02
$BSADF^B$	08/87	09/87	12/95	07/01
$BSADF^\infty$	08/87	09/87	02/96	10/00
$BSADF^{\text{fscv}}$	02/87	10/87	12/95	08/01
PSY	06/86	09/87	11/95	08/01

Notes: $\hat{t}_e$ is the first date a test rejects the null of no bubble when applied sequentially. $\hat{t}_f$ is the first date a test stops rejecting and continues to fail to reject for 12 consecutive months.

consecutive non-rejections with the first of those non-rejections being the estimated termination date.[12] We pay particular attention to the origination date which is informative about how soon each procedure would have been able to give an early warning signal of an ongoing bubble.

Table 5 reports the estimated origination and termination dates of the of Black Monday and dotcom bubbles from the procedures based on the $CBSADF^B$, $BSADF^B$, $BSADF^\infty$ and $BSADF^{\text{fscv}}$ tests applied sequentially in this manner, where the notation $CBSADF^B(x)$ is used to indicate that the covariate x has been used. No other bubbles were detected in addition to the Black Monday and dotcom bubbles. Due to data availability we were only able to use the VXO index when dating the dotcom bubble as the VXO index was only created in 1986.[13] Also reported in Table 5 are the estimated origination and termination dates of these bubbles identified by PSY in their empirical exercise. As discussed in Remark 4.7, the estimates reported in PSY, like $BSADF^{\text{fscv}}$, are based on finite sample critical values which assume that $u_t \sim NIID(0, 1)$. Moreover, the results in PSY are based on sub-sample test regressions all of which impose a lag order of $p_v = 0$, inconsistent with the BIC which we found to select $p_v = 1$ in a vast majority of sub-sample regressions. We report in Figs. 4 and 5 the sequence of test statistics and critical values calculated for each procedure used in the monitoring exercise in the neighbourhood of the Black Monday and dotcom bubble episodes, respectively. The estimated origination and termination dates for all of the procedures are also plotted on a graph of the price dividend ratio in Fig. 6.

Consider first the results in Table 5 pertaining to the Black Monday bubble. We observe that the $CBSADF^B(\text{SVAR})$ test first rejects the null hypothesis of no bubble 18 months earlier than the $BSADF^B$ or $BSADF^\infty$ tests. Furthermore, the $CBSADF^B(\Delta \text{Aaa})$, $CBSADF^B(\Delta \text{Baa})$, $CBSADF^B(\text{CORPR})$ and the $CBSADF^B(\Delta \text{GS10})$ tests reject the no-bubble null hypothesis 17 months earlier than either the $BSADF^B$ or $BSADF^\infty$ tests. When utilising the $CBSADF^B(\text{LTR})$ or $CBSADF^B(\Delta \text{LTY})$ tests a rejection is signalled 14 months earlier than the $BSADF^B$ or $BSADF^\infty$ tests. Finally, the $CBSADF^B(\Delta \text{BM})$ test rejects 7 months earlier than $BSADF^B$ and $BSADF^\infty$, and $CBSADF^B(\text{CREDSA})$ and $CBSADF^B(\Delta \text{TBL})$ also reject 5 months earlier than $BSADF^B$ and $BSADF^\infty$. We note that both the $BSADF^B$ and $BSADF^\infty$ tests only reject the null hypothesis in a single month, August 1987. Consequently neither the $BSADF^B$ or $BSADF^\infty$ tests actually identify the Black Monday bubble because the minimum required length of one year between $\hat{t}_e$ and $\hat{t}_f$ is not met. This is also true for the $CBSADF^B(\text{CREDSA})$ and $CBSADF^B(\Delta \text{TBL})$ tests, albeit these tests do reject for more than a single month.

While it may appear strange that the first rejection dates seen for the bootstrap $BSADF^B$-based procedure occur some time after the bubble origination dates identified by PSY, the finite sample critical values used by PSY are almost identical to those we calculated graphed on Fig. 4 and can be seen to considerably more liberal than the bootstrap critical values for $BSADF$ graphed on Fig. 4; as discussed in Remark 4.7, their reliability is highly questionable for the real S&P 500 price–dividend ratio data. This is also evidenced by the $BSADF^{\text{fscv}}$-based procedure which rejects ahead of both the $BSADF^B$ and $BSADF^\infty$ based procedures. The $BSADF^{\text{fscv}}$ procedure does still reject later than the date reported by PSY. This is likely due to PSY not including any lagged dependent variables in their sub-sample ADF regressions which tends to inflate the value of the $BSADF$ statistic relative to when lag augmentation is included. The termination date for the Black Monday bubble episode is identified as December 1987 by all $CBSADF^B$ procedures except where SVAR is utilised as a covariate where the estimated termination date is a month later in January

---

[12] The latter condition ensures that we do not spuriously indicate the end of a bubble before it has actually terminated. This could happen for at least two reasons: first, as each of the test statistics in the sequence is the outcome of a random variable, even within a genuine bubble episode the outcomes of some statistics in the sequence can lie below the critical value in finite samples by chance; second, bubble episodes often contain short periods of downward movement before the underlying upward movement continues (these might be very short-lived mini-collapses or temporary market corrections) and these can be sufficient to trigger non-rejections in the tests in the sequence.

[13] We begin monitoring using the $CBSADF^B(\text{VXO})$ test in April 1994 to ensure the test is initially performed on a reasonably large sample size of 100.
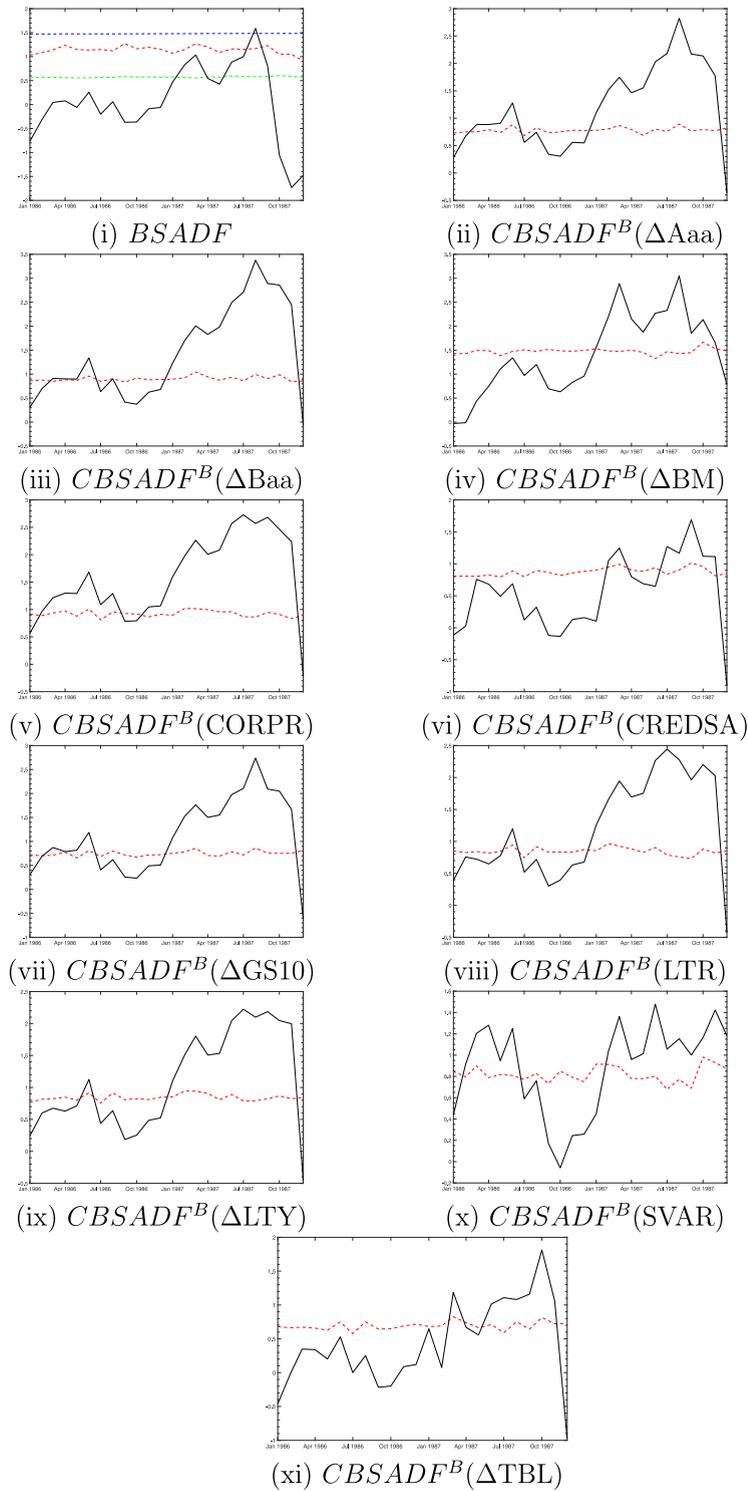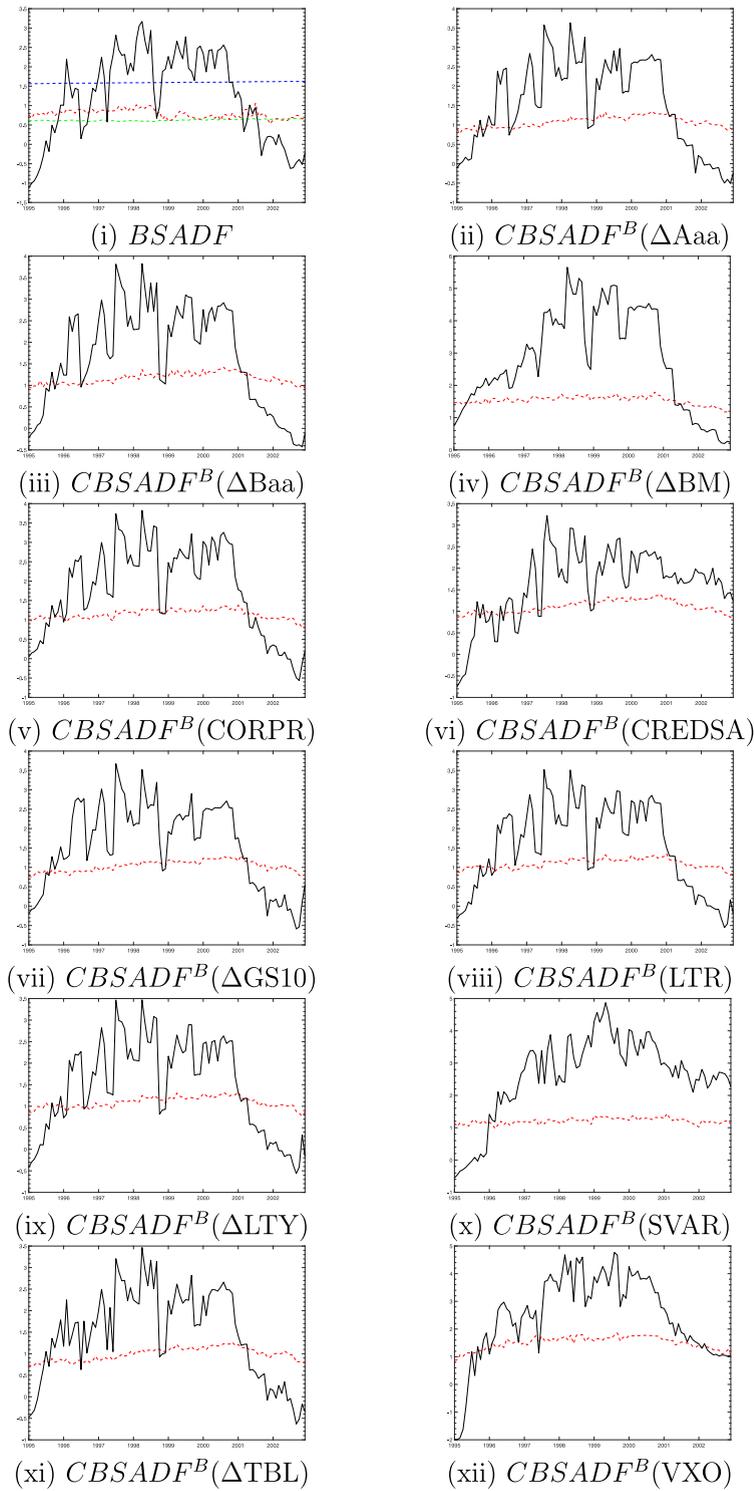
(i) $BSADF$

(ii) $CBSADF^B(\Delta\text{Aaa})$

(iii) $CBSADF^B(\Delta\text{Baa})$

(iv) $CBSADF^B(\Delta\text{BM})$

(v) $CBSADF^B(\text{CORPR})$

(vi) $CBSADF^B(\text{CREDSA})$

(vii) $CBSADF^B(\Delta\text{GS10})$

(viii) $CBSADF^B(\text{LTR})$

(ix) $CBSADF^B(\Delta\text{LTY})$

(x) $CBSADF^B(\text{SVAR})$

(xi) $CBSADF^B(\Delta\text{TBL})$

Test Statistic: ——
Bootstrap cv: − − −, Asymptotic cv: − − −, Finite Sample cv: − − −

Fig. 4. Sequential test statistics and critical values: January 1986–December 1987 (Black Monday bubble).

(i) $BSADF$

(ii) $CBSADF^B(\Delta\text{Aaa})$

(iii) $CBSADF^B(\Delta\text{Baa})$

(iv) $CBSADF^B(\Delta\text{BM})$

(v) $CBSADF^B(\text{CORPR})$

(vi) $CBSADF^B(\text{CREDSA})$

(vii) $CBSADF^B(\Delta\text{GS10})$

(viii) $CBSADF^B(\text{LTR})$

(ix) $CBSADF^B(\Delta\text{LTY})$

(x) $CBSADF^B(\text{SVAR})$

(xi) $CBSADF^B(\Delta\text{TBL})$

(xii) $CBSADF^B(\text{VXO})$

Test Statistic: ———

Bootstrap cv: $- - -$, Asymptotic cv: $- - -$, Finite Sample cv: $- - -$

Fig. 5. Sequential test statistics and critical values: January 1995–December 2002 (dotcom bubble).

(b) ΔAaa

(c) ΔBaa

(d) ΔBM

(e) CORPR

(f) CREDSA

(g) ΔGS10

(h) LTR

(i) ΔLTY

(j) SVAR

(k) ΔTBL

(l) VXO
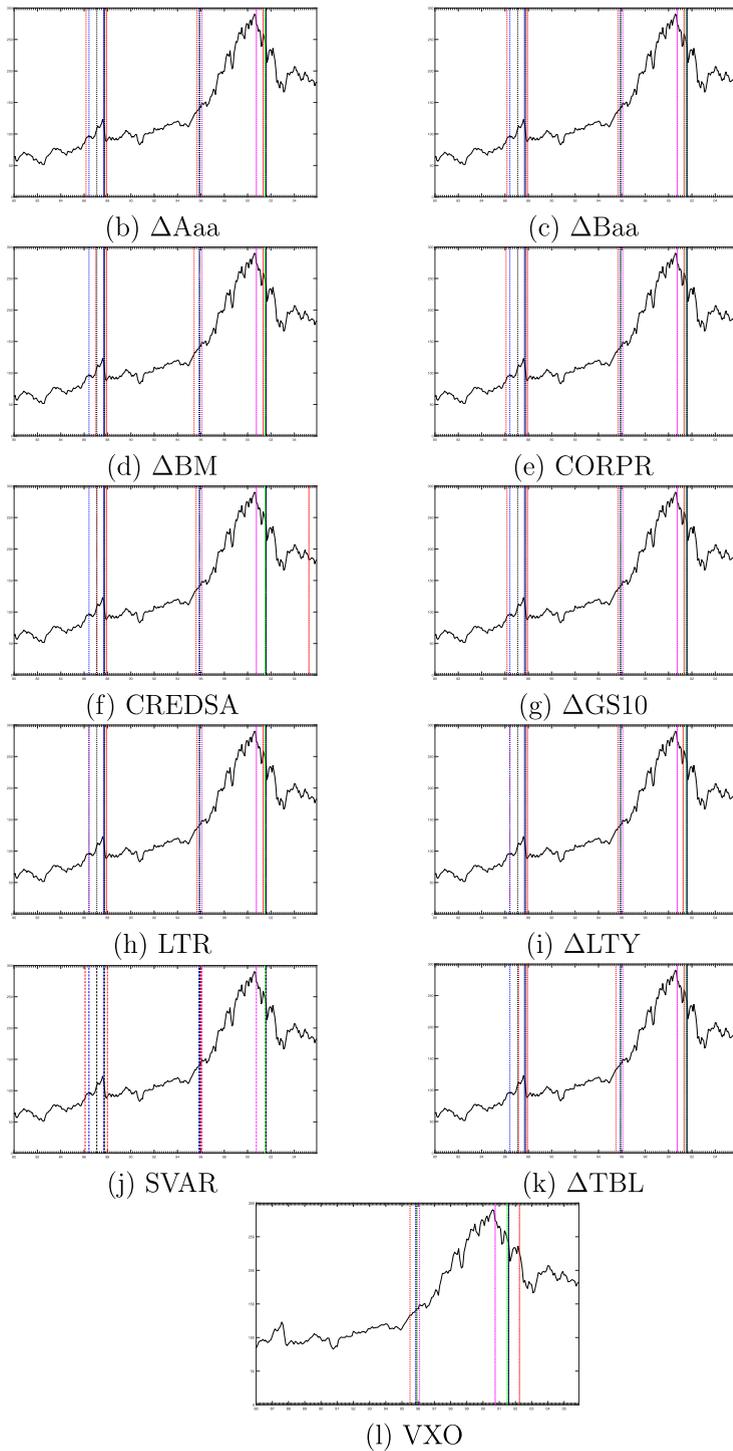
Price Dividend Ratio: ——, $\boldsymbol{CBSADF}$(x): $\hat{t}_e$ – – – $\hat{t}_f$ – · –,
$\boldsymbol{BSADF^B}$: $\hat{t}_e$ – – – $\hat{t}_f$ – · –, $\boldsymbol{BSADF^\infty}$: $\hat{t}_e$ – – – $\hat{t}_f$ – · –,
$\boldsymbol{BSADF}^{\text{fscv}}$: $\hat{t}_e$ – – – $\hat{t}_f$ – · –, PSY: $\hat{t}_e$ – – – $\hat{t}_f$ – · –

Fig. 6. Origination and termination dates of identified bubble episodes.

1988. PSY, on the other hand, report an estimated termination date of September 1987. It is unsurprising that these estimated termination dates are quite close together given the sharp crash in world stock markets following Black Monday.

We next turn to the estimated origination dates for the dotcom bubble. The $CBSADF^B(\Delta BM)$ test first rejects the no bubble null hypothesis 6 months earlier than $BSADF^B$ (which in turn first rejects one month later than the estimated origination date reported in PSY) and 8 months ahead of $BSADF^\infty$. The $CBSADF^B(\Delta TBL)$ and $CBSADF^B(VXO)$ tests first reject the null hypothesis in July 1995, 5 months ahead of $BSADF^B$ and 7 months ahead of $BSADF^\infty$. In addition, $CBSADF^B(CREDSA)$ first rejects the null hypothesis 4 months ahead of $BSADF^B$ and 6 months earlier than $BSADF^\infty$. Finally, the $CBSADF^B(\Delta Aaa)$, $CBSADF^B(\Delta Baa)$, $CBSADF^B(CORPR)$, $CBSADF^B(\Delta GS10)$, $CBSADF^B(LTR)$ and $CBSADF^B(\Delta LTY)$ tests all reject 3 months earlier than $BSADF^B$ and 5 months ahead of $BSADF^\infty$. Fig. 5 shows that the $CBSADF^B$ tests also displays a more consistent pattern of rejections over time than both $BSADF^B$ and $BSADF^\infty$. The $BSADF^{fscv}$ procedure first rejects a month after the estimated origination date reported by PSY.

Somewhat more variability is observed in the estimated termination date of the dotcom bubble across the procedures than was seen for the Black Monday episode, with $CBSADF^B(\Delta LTY)$ estimating the earliest termination date and $CBSADF^B(SVAR)$ estimating the latest termination date. That there is more variability in the estimated termination date of the dotcom bubble relative to the Black Monday episode is unsurprising given the relative manner in which these two bubbles collapsed. The market collapse following Black Monday was short and sharp, leading the procedures to identify the termination of this bubble within a very small window of observations. The decline in stock prices following the dotcom bubble, whilst unprecedented, occurred over a period of more than one year, leading to somewhat more variability across the estimated termination dates in Table 5.

Our empirical results show that by employing empirically relevant covariates, date-stamping procedures based on the $CBSADF^B$ tests typically estimate the origination date of both the bubble preceding Black Monday and the dotcom bubble well in advance of the standard PSY procedure. An implication is that the covariate augmented methods proposed in this paper would have identified these bubbles several months earlier than the existing procedures if used as part of a real-time monitoring exercise.

7. Conclusions

Reliable bubble detection procedures provide crucial information for policy makers looking to implement macroprudential policy aimed at mitigating the damage to the economy that can be caused by bubble episodes in financial markets. Such methods need to have simultaneously low false detection rates and high true detection rates. We have contributed to this development by extending the most popular univariate bubble detection and associated date-stamping procedures in the literature developed in PWY and PSY to allow for the incorporation of additional information from stationary covariates. This leads to a more precise estimate of the leading autoregressive coefficient upon which bubble detection and date-stamping procedures are based. To account for the dependency of the null distribution of the covariate augmented statistics to a nuisance parameter (which measures the relevance of the covariates), asymptotically valid bootstrap implementations of the tests were employed. Monte Carlo simulation results have shown that the covariate-augmented bootstrap tests display well controlled false detection properties, and that the inclusion of relevant covariates can lead to considerable gains in the efficacy of both the bubble detection and date-stamping procedures, relative to the corresponding univariate procedures.

In an empirical application we have explored the use of our proposed covariate-based methods in settings where they have the potential to be used as early warning mechanisms, informing policy makers in real-time of the emergence of a bubble episode; information crucial to structuring macroprudential policy. Specifically, we have examined whether or not our proposed covariate-based methods would have detected the bubble preceding Black Monday and the dotcom bubble in the S&P 500 price dividend series previously identified by PSY at an earlier date than by the use of the corresponding univariate procedures used in PSY. In this exercise we investigated the use of multiple candidate covariates, with our results showing that using these covariates would indeed have allowed for significantly earlier detection of both of these bubble episodes.

CRediT authorship contribution statement

Sam Astill: Conceptualization, Supervision, Software, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **A.M. Robert Taylor:** Conceptualization, Supervision, Software, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Neil Kellard:** Conceptualization, Supervision, Methodology, Writing – original draft, Writing – review & editing. **Ioannis Korkos:** Methodology, Software, Investigation, Formal analysis, Writing – original draft, Writing – review & editing.

Appendix

This Appendix contains three sections. In Appendix A.1 we outline the residual bootstrap used to generated *p*-values for the covariate-augmented bubble detection tests, and provide some discussion relating to the bootstrap methods. In Appendix A.2 we outline how the methods outlined in the main text can be generalised to allow for either a constant or linear trend in the data. Finally, Appendix A.3 provides proofs of Theorems 1 and 2.

A.1. Bootstrap implementation

   Our proposed bootstrap implementation of the covariate-augmented versions of the bubble detection tests of PWY and PSY developed in Section 4 is first outlined in Algorithm A.1. This is based on the IID residual bootstrap re-sampling approach developed in CSS for use with the full sample CADF tests of Hansen (1995). Some discussion follows Algorithm A.1.

Algorithm A.1 (*Bootstrap CADF tests*).
Step 1: Estimate the following regression by OLS,

$$\Delta y_t = \sum_{k=1}^{p} \tilde{\alpha}_k \Delta y_{t-k} + \sum_{k=-q_1}^{q_2} \tilde{\beta}'_k w_{t-k} + \tilde{\varepsilon}_t. \tag{A.1}$$

Step 2: Estimate the following regression by Yule–Walker[14] (see, for example, Brockwell and Davis (1991), pp.239-241),

$$w_{t+q_1+1} = \tilde{\Psi}_1 w_{t+q_1} + \cdots + \tilde{\Psi}_\ell w_{t+q_1-\ell+1} + \tilde{\eta}_t. \tag{A.2}$$

Step 3: Define $\tilde{\xi}_t := (\tilde{\varepsilon}_t, \tilde{\eta}_t)'$, $t = t_1, ..., T-q_1$, where $t_1 := \max(\ell+1, 2+\max(p,q_2))$ and where $\tilde{\varepsilon}_t$ and $\tilde{\eta}_t$ are the fitted residuals obtained from equations (A.1) and (A.2), respectively, and generate the bootstrap sample $\xi_t^* := (\varepsilon_t^*, \eta_t^*)'$, $t = 1, ..., T$, from the re-centred residuals, $\tilde{\xi}_t^c := \tilde{\xi}_t - (T-q_1-t_1+1)^{-1} \sum_{s=t_1}^{T-q_1} \tilde{\xi}_s$, using a standard residual bootstrap re-sampling (with replacement) scheme; that is, $\xi_t^* := \tilde{\xi}_{\mathcal{U}_t}^c$, where $\mathcal{U}_t$, $t = 1, ..., T$, is an independent and identically distributed (IID) sequence of (discrete) uniformly distributed random variables on $\{t_1, ..., T-q_1\}$.
Step 4: Construct the bootstrap sample of the stationary covariate, $\{w_t^*\}$, $t = 1-q_2, ..., T+q_1$, according to the recursion

$$w_{t+q_1+1}^* = \tilde{\Psi}_1 w_{t+q_1}^* + \cdots + \tilde{\Psi}_\ell w_{t+q_1+1-\ell}^* + \eta_t^* \tag{A.3}$$

where $\tilde{\Psi}_k$, $1 \le k \le \ell$ are the Yule–Walker estimates obtained in Step 2, and where we follow CSS and initialise the bootstrap sample by setting $\eta_t^* = w_{t+q_2}$ for $t = 1-q_2, ..., q_1+1$ and $w_{1-q_2-\ell}^* = \cdots = w_{-q_2}^* = 0$.
Step 5: Construct the bootstrap sample $\{v_t^*\}$ according to

$$v_t^* = \sum_{k=-q_1}^{q_2} \tilde{\beta}'_k w_{t-k}^* + \varepsilon_t^* \tag{A.4}$$

where $\tilde{\beta}_k$, $-q_1 \le k \le q_2$ are the OLS estimates obtained in Step 1.
Step 6: Generate the bootstrap errors $\{u_t^*\}$, according to the recursion

$$u_t^* = \tilde{\alpha}_1 u_{t-1}^* + ... + \tilde{\alpha}_p u_{t-p}^* + v_t^* \tag{A.5}$$

where the p initial values of u_t^* are set to zero, and where $\tilde{\alpha}_k$, $1 \le k \le p$ are the OLS estimates obtained in Step 1.
Step 7: Generate the bootstrap sample $\{y_t^*\}$ by cumulating the bootstrap errors $\{u_t^*\}$;

$$y_t^* = y_{t-1}^* + u_t^* = y_0^* + \sum_{k=1}^{t} u_k^* \tag{A.6}$$

initialised at $y_0^* = 0$. Notice therefore that the unit root null hypothesis is imposed on the bootstrap sample data $\{y_t\}$.
Step 8: Using the bootstrap sample data $\{y_t^*, w_t^*\}$, construct the bootstrap analogues of the $CGSADF$, $CSADF$ and $CBSADF$ statistics as

$$CSADF^* := \sup_{r\in[r_0,1]} \{CADF^*(p,q_1,q_2)_0^r\} \tag{A.7}$$

$$CBSADF^* := \sup_{r\in[0,1-r_0]} \{CADF^*(p,q_1,q_2)_r^1\} \tag{A.8}$$

$$CGSADF^* := \sup_{\substack{r_2\in[r_0,1]\\r_1\in[0,r_2-r_0]}} \{CADF^*(p,q_1,q_2)_{r_1}^{r_2}\} \tag{A.9}$$

where $CADF^*(p,q_1,q_2)_{r_1}^{r_2}$ denotes the bootstrap analogue of the $CADF(p,q_1,q_2)_{r_1}^{r_2}$ statistic and is obtained from estimating the regression

$$\Delta y_t^* = \alpha y_{t-1}^* + \sum_{k=1}^{p} \alpha_k \Delta y_{t-k}^* + \sum_{k=-q_1}^{q_2} \beta'_k w_{t-k}^* + \varepsilon_t^* \tag{A.10}$$

over the bootstrap data sub-sample $t = \lfloor r_1 T \rfloor + 1, ..., \lfloor r_2 T \rfloor$.
Step 9: Bootstrap p-values for the three tests are then computed as: $p_{1,T}^* := 1 - G_{1,T}^*(CSADF)$, $p_{2,T}^* := 1 - G_{2,T}^*(CBSADF)$ and $p_{3,T}^* := 1 - G_{3,T}^*(CGSADF)$, where $G_{1,T}^*(\cdot)$, $G_{2,T}^*(\cdot)$ and $G_{3,T}^*(\cdot)$ denote the conditional (on the original sample data) cumulative density functions (cdf) of $CSADF^*$, $CBSADF^*$ and $CGSADF^*$, respectively. Notice, therefore, that the bootstrap tests, run at the λ significance level, based on $CSADF$, $CBSADF$ and $CGSADF$ are then defined such that they reject the constant unit root null hypothesis, H_0 of (3), if $p_{1,T}^* < \lambda$, $p_{2,T}^* < \lambda$ and $p_{3,T}^* < \lambda$, respectively.

---

[14] Following CSS, the Yule–Walker estimation method is used to guarantee that the estimated autoregression fitted to w_t has only stable roots.

Remark A.1. Step 1 of Algorithm A.1 estimates the CADF Eq. (8) under the constant unit root null hypothesis, H_0, to obtain the residuals $\tilde{\varepsilon}_t$. This will clearly lead to more efficient estimates of the parameters of (8) when H_0 is true. Alternatively, one could obtain $\tilde{\varepsilon}_t$ in Step 1 from the unrestricted estimation of (8). Doing so would not alter the large sample properties of the bootstrap statistics from Algorithm A.1 detailed in Appendix A.3. ◇

Remark A.2. A simplified version of Algorithm A.1 may be considered whereby the so-called re-colouring aspect of Step 6 can be omitted. In this case in Step 6 set $u_t^* = v_t^*$. The large sample properties of the bootstrap statistics from Algorithm A.1 detailed in Appendix A.3 are unaffected by this simplification. ◇

Remark A.3. Where the innovations ξ_t follow a specific GARCH model, CSS (p.142) investigate a possible extension of their basic IID bootstrap re-sampling algorithm which explicitly models the GARCH component of the data. Even where the GARCH model adopted for the innovations in the algorithm is correctly specified, the simulation results in CSS suggest that the finite sample performance of this extended bootstrap is no better than that of the basic IID bootstrap. In unreported simulations, we found the same to be true in the setting considered in this paper and so we will only report results in what follows for the standard IID residual bootstrap outlined in Algorithm A.1. As discussed in CSS (p.150), it is not entirely surprising that this extended bootstrap performs no better than the basic IID bootstrap because the limiting null distributions of the test statistics involved are not pivotal (recall that they depend on ρ^2) and, hence, the GARCH modification to the bootstrap algorithm will not provide an asymptotic refinement. ◇

A.2. Allowing for a non-zero mean and/or linear trend in y_t and w_t

Each of the $SADF$, $BSADF$ and $GSADF$ statistics in (12)–(14) can be generalised to allow for the presence of a constant or constant plus linear trend in y_t (cf. Remark 2.1). As with the full sample CADF tests of Hansen (1995), this can be done by augmenting the sub-sample ADF regression used in computing the $ADF(p)_{r_1}^{r_2}$ statistic with either a constant or constant plus linear trend, respectively.

Similarly, each of the $CSADF$, $CBSADF$ and $CGSADF$ statistics can be generalised to allow for the presence of a constant or constant plus linear trend in y_t and w_t by augmenting the sub-sample covariate augmented ADF regression used for computing the $CADF(p, q_1, q_2)_{r_1}^{r_2}$ statistic with either a constant or constant plus linear trend, respectively. The limiting distributions of these change from those given for the no deterministics case in Theorem 1; see Remark A.6 for further details.

For bootstrap implementation of the tests, Algorithm A.1 is presented for the case where no deterministics are allowed for in computing the $CGSADF$, $CSADF$ and $CBSADF$ statistics on the original data. In order to allow for a constant or constant plus linear trend in the data simply replace y_t and w_t in Steps 1 and 2 of Algorithm A.1 by their full sample demeaned or detrended counterparts, denoted $\hat{y}_t$ and $\hat{w}_t$ respectively, obtained as the OLS residuals from regressing y_t and the elements of w_t on either a constant or a constant and linear trend, over $t = 1, \ldots, T$, and commensurately in Step 8 of Algorithm A.1 include either a constant or constant plus linear trend in (A.10), the bootstrap analogue of the CADF regression. The bootstrap validity results outlined in Theorem 2 continue to hold when allowance is made in this way for either a constant or constant plus linear trend in the analysis.

A.3. Proof of Theorems 1 and 2

In what follows we will follow CSS and impose the following assumptions on the innovations $\xi_t := (\varepsilon_t, \eta_t')'$ and on the variables u_t and $z_t := \left(\Delta y_{t-1}, \ldots, \Delta y_{t-p}, w_{t+r}', \ldots, w_{t-q}' \right)'$.

Assumption A.1 (*a*). Let $\{\xi_t\}$ be a martingale difference sequence [MDS] such that $E(\xi_t \xi_t') = \Sigma := \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \Sigma_\eta \end{pmatrix}$ and $(1/T) \sum_{t=1}^T \xi_t \xi_t' \underline{P} \Sigma$ with $\Sigma > 0$ and $E|\xi_t|^\gamma < K$ for some $\gamma \geq 4$, where K is some constant depending only upon γ, and $|.|$ denotes the Euclidean norm; and (b) $\alpha(z) \neq 0$ and $\det(\Psi(z)) \neq 0$ in each case for all $|z| \leq 1$.

Assumption A.2. The following additional conditions are assumed to hold in (1) and (2): (a) $\sigma_v^2 > 0$; and (b) $E(z_t z_t') > 0$

Remark A.4. A full discussion of Assumptions A.1 and A.2 can be found on page 138 of CSS and page 1151 of Hansen (1995). Importantly, Assumption A.1(a) allows for conditional heteroskedasticity, including GARCH behaviour, in all equations in the system including the covariates. The MDS assumption implies that ε_t is serially uncorrelated with η_{t+k} for all $k \geq 1$, and hence is orthogonal to the lagged differences of the dependent variable, $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$, and to the leads and lags of the covariates, $w_{t+q_1}, \ldots, w_{t-q_2}$. Taken together Assumptions A.1 and A.2 ensure that, under H_0, y_t is an $I(1)$ process (so that we may interpret our tests of H_0 as tests for a constant parameter unit root in y_t; see footnote 1 of Hansen (1995), for further discussion on this point) and that the parameters $\{\alpha_j\}_{j=1}^p$ and $\{\beta_k\}_{k=-q_1}^{q_2}$, can be consistently estimated by standard ordinary least squares (OLS) estimation. ◇

Remark A.5. Under Assumption A.1(a) the innovations, ξ_t, satisfy the multivariate invariance principle $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T \cdot \rfloor} \xi_t \xrightarrow{w} B(\cdot)$, where $B(r) = (B_\varepsilon(\cdot), B_\eta'(\cdot))'$ is an $(m+1)$-dimensional Brownian motion with covariance matrix Σ. ◇

A.3.1. Proof of Theorem 1

Consider to begin with the case where r_1 and r_2 are fixed. Here the data sub-sample is $t = \lfloor r_1 T \rfloor + 1, \ldots, \lfloor r_2 T \rfloor$. Defining $T_1 := \lfloor r_1 T \rfloor + 2 + \max(p, q_2)$ and $T_2 := \lfloor r_2 T \rfloor - q_1$, we can define the corresponding sub-sample regression quantities

$$A_{r_1,r_2} := \sum_{t=T_1}^{T_2} y_{t-1}\epsilon_t - \left(\sum_{t=T_1}^{T_2} y_{t-1}z_t'\right)\left(\sum_{t=T_1}^{T_2} z_t z_t'\right)^{-1}\left(\sum_{t=T_1}^{T_2} z_t \epsilon_t\right) \tag{A.11}$$

$$B_{r_1,r_2} := \sum_{t=T_1}^{T_2} y_{t-1}^2 - \left(\sum_{t=T_1}^{T_2} y_{t-1}z_t'\right)\left(\sum_{t=T_1}^{T_2} z_t z_t'\right)^{-1}\left(\sum_{t=T_1}^{T_2} z_t y_{t-1}\right) \tag{A.12}$$

$$C_{r_1,r_2} := \sum_{t=T_1}^{T_2} \epsilon_t^2 - \left(\sum_{t=T_1}^{T_2} \epsilon_t z_t'\right)\left(\sum_{t=T_1}^{T_2} z_t z_t'\right)^{-1}\left(\sum_{t=T_1}^{T_2} z_t \epsilon_t\right). \tag{A.13}$$

By trivially adapting results given in Lemma 2.1 of Park and Phillips (1989) to the sub-sample data considered here, we immediately have the results that $\sum_{t=T_1}^{T_2} z_t z_t' = O_p(T)$, $\sum_{t=T_1}^{T_2} z_t \epsilon_t = O_p(T^{1/2})$ and $\sum_{t=T_1}^{T_2} y_{t-1}z_t' = O_p(T)$. It therefore follows that the second terms appearing on the right hand sides of (A.11), (A.12) and (A.13), respectively, satisfy

$$\left(\sum_{t=T_1}^{T_2} y_{t-1}z_t'\right)\left(\sum_{t=T_1}^{T_2} z_t z_t'\right)^{-1}\left(\sum_{t=T_1}^{T_2} z_t \epsilon_t\right) = O_p(T^{1/2})$$

$$\left(\sum_{t=T_1}^{T_2} y_{t-1}z_t'\right)\left(\sum_{t=T_1}^{T_2} z_t z_t'\right)^{-1}\left(\sum_{t=T_1}^{T_2} z_t y_{t-1}\right) = O_p(T)$$

$$\left(\sum_{t=T_1}^{T_2} \epsilon_t z_t'\right)\left(\sum_{t=T_1}^{T_2} z_t z_t'\right)^{-1}\left(\sum_{t=T_1}^{T_2} z_t \epsilon_t\right) = o_p(T).$$

Using these results, we therefore have that

$$(T_2 - T_1 + 1)^{-1} A_{r_1,r_2} = (T_2 - T_1 + 1)^{-1}\sum_{t=T_1}^{T_2} y_{t-1}\epsilon_t + o_p(1) \tag{A.14}$$

$$(T_2 - T_1 + 1)^{-2} B_{r_1,r_2} = (T_2 - T_1 + 1)^{-2}\sum_{t=T_1}^{T_2} y_{t-1}^2 + o_p(1) \tag{A.15}$$

$$(T_2 - T_1 + 1)^{-1} C_{r_1,r_2} = (T_2 - T_1 + 1)^{-1}\sum_{t=T_1}^{T_2} \epsilon_t^2 + o_p(1). \tag{A.16}$$

Defining $\hat{\sigma}_{r_1,r_2}^2$ as the corresponding OLS regression standard error, the $CADF(p, q_1, q_2)_{r_1}^{r_2}$ statistic can be written, under the unit root null hypothesis, H_0 of (3), as

$$\begin{aligned}
CADF(p, q_1, q_2)_{r_1}^{r_2} &= \frac{1}{\hat{\sigma}_{r_1,r_2}}\frac{A_{r_1,r_2}}{B_{r_1,r_2}^{1/2}} \\
&= \frac{1}{\hat{\sigma}_{r_1,r_2}}\left(\frac{(T_2 - T_1 + 1)^{-1}\sum_{t=T_1}^{T_2} y_{t-1}\epsilon_t}{\left((T_2 - T_1 + 1)^{-2}\sum_{t=T_1}^{T_2} y_{t-1}^2\right)^{1/2}}\right) + o_p(1) \\
&\overset{w}{\to} \frac{\int_{r_1}^{r_2} Q(s)dP(s)}{\left(\int_{r_1}^{r_2} Q(s)^2 ds\right)^{1/2}}
\end{aligned} \tag{A.17}$$

where the second line follows using the results in (A.14) and (A.15) and the last line follows using a trivial adaptation of Lemma A.2 of Chang et al. (2013) to the data sub-sample $t = \lfloor r_1 T \rfloor + 1, \ldots, \lfloor r_2 T \rfloor$ used to construct A_{r_1,r_2} and B_{r_1,r_2} and the fact that $\hat{\sigma}_{r_1,r_2}^2 \overset{p}{\to} \sigma_\epsilon^2$, which follows using (A.16). The large sample result in (A.17) holds formally only for fixed r_1, r_2. However, following the same approach (which is based on the proof strategy adopted by Zivot and Andrews (1992), to prove their Theorem 1) as that used to establish Equation (A.6) on p.1072 in the proof of Theorem 1 in PSY (pp. 1072–1075), the stated result for the limiting null distribution of the $CGSADF$ statistic can be shown to follow by means of the Continuous Mapping Theorem [CMT] from the fixed r_1, r_2 representation in (A.17).

Remark A.6. Where a constant is included in the sub-sample CADF regressions the formula for the limiting distribution in (21) holds with $Q(s)$ replaced by $Q^\mu(s)$ where for a generic Brownian motion process, $W(s)$, we define the sub-sample demeaned process $W^\mu(s) := W(s) - \frac{1}{(r_2-r_1)}\int_{r_1}^{r_2} W(t)dt$. Where a constant and linear trend are included in the CADF regression, $Q(s)$ is replaced by $Q^\tau(s)$ where, generically, $W^\tau(s) := W^\mu(s) - 12(r_2 - r_1)^{-3}(s - \frac{(r_2-r_1)}{2})\int_{r_1}^{r_2}(t - \frac{(r_2-r_1)}{2})W^\mu(t)dt$. Analogously in (22), replace $W_1(s)$ by $W_1^\mu(s)$ or $W_1^\tau(s)$, as appropriate. These resulting limiting distributions depend on both r_0 and ϱ^2. ◇

A.3.2. Proof of Theorem 2

As with the proof of Theorem 1, consider to begin with the case where r_1 and r_2 are fixed. Correspondingly, define the bootstrap sample moment analogues of A_{r_1,r_2} and B_{r_1,r_2} defined in the proof of Theorem 1 as

$$A^*_{r_1,r_2} := \sum_{t=T_1}^{T_2} y^*_{t-1}\varepsilon^*_t - \left(\sum_{t=T_1}^{T_2} y^*_{t-1}z^{*\prime}_t\right)\left(\sum_{t=T_1}^{T_2} z^*_t z^{*\prime}_t\right)^{-1}\left(\sum_{t=T_1}^{T_2} z^*_t \varepsilon^*_t\right) \tag{A.18}$$

$$B^*_{r_1,r_2} := \sum_{t=T_1}^{T_2} y^{*2}_{t-1} - \left(\sum_{t=T_1}^{T_2} y^*_{t-1}z^{*\prime}_t\right)\left(\sum_{t=T_1}^{T_2} z^*_t z^{*\prime}_t\right)^{-1}\left(\sum_{t=T_1}^{T_2} z^*_t y^*_{t-1}\right). \tag{A.19}$$

By trivially adapting results given in Lemma 3 of Chang and Park (2003) to the data sub-sample $t = \lfloor r_1 T\rfloor + 1,\dots,\lfloor r_2 T\rfloor$, we immediately have that

$$\sum_{t=T_1}^{T_2} z^*_t z^{*\prime}_t = O^*_p(T), \quad \sum_{t=T_1}^{T_2} z^*_t \varepsilon^*_t = O^*_p(T^{1/2}) \quad \text{and} \quad \sum_{t=T_1}^{T_2} y^*_{t-1}z^{*\prime}_t = O^*_p(T) \tag{A.20}$$

where we use the notation O^*_p and o^*_p to denote the bootstrap analogues of the standard O_p and o_p stochastic order symbols for the original sample asymptotics (for a formal definition and further discussion of these, see Chang and Park (2003), pp. 386–387). Using the results in (A.20), it therefore follows immediately that the second terms appearing on the right hand sides of (A.18) and (A.19), respectively, satisfy

$$\left(\sum_{t=T_1}^{T_2} y^*_{t-1}z^{*\prime}_t\right)\left(\sum_{t=T_1}^{T_2} z^*_t z^{*\prime}_t\right)^{-1}\left(\sum_{t=T_1}^{T_2} z^*_t \varepsilon^*_t\right) = O^*_p(T^{1/2})$$

$$\left(\sum_{t=T_1}^{T_2} y^*_{t-1}z^{*\prime}_t\right)\left(\sum_{t=T_1}^{T_2} z^*_t z^{*\prime}_t\right)^{-1}\left(\sum_{t=T_1}^{T_1} z^*_t y^*_{t-1}\right) = O^*_p(T).$$

Using these results we therefore have that

$$(T_2 - T_1 + 1)^{-1} A^*_{r_1,r_2} = (T_2 - T_1 + 1)^{-1}\sum_{t=T_1}^{T_2} y^*_{t-1}\varepsilon^*_t + o^*_p(1) \tag{A.21}$$

$$(T_2 - T_1 + 1)^{-2} B^*_{r_1,r_2} = (T_2 - T_1 + 1)^{-2}\sum_{t=T_1}^{T_2} y^{*2}_{t-1} + o^*_p(1). \tag{A.22}$$

Defining $\hat{\sigma}^{*2}_{r_1,r_2}$ to be the bootstrap counterpart of $\hat{\sigma}^2_{r_1,r_2}$ and noting that Step 7 of Algorithm A.1 imposes H_0 of (3) on the bootstrap data $\{y^*_t\}$, the bootstrap $CADF^*(p, q_1, q_2)^{r_2}_{r_1}$ statistic can be written as

$$\begin{aligned} CADF^*(p, q_1, q_2)^{r_2}_{r_1} &= \frac{1}{\hat{\sigma}^*_{r_1,r_2}}\frac{A^*_{r_1,r_2}}{B^{*1/2}_{r_1,r_2}} \\ &= \frac{1}{\hat{\sigma}^*_{r_1,r_2}}\left(\frac{(T_2 - T_1 + 1)^{-1}\sum_{t=T_1}^{T_2} y^*_{t-1}\varepsilon^*_t}{\left((T_2 - T_1 + 1)^{-2}\sum_{t=T_1}^{T_2} y^{*2}_{t-1}\right)^{1/2}}\right) + o^*_p(1) \\ &\xrightarrow{w}_p \frac{\int_{r_1}^{r_2} Q(s)\,dP(s)}{\left(\int_{r_1}^{r_2} Q(s)^2\,ds\right)^{1/2}} \end{aligned}$$

where the second line follows using the results in (A.21) and (A.22), and the last line follows using a trivial adaptation of Lemma A.4 of Chang et al. (2013) to the bootstrap data sub-sample $t = \lfloor r_1 T\rfloor + 1,\dots,\lfloor r_2 T\rfloor$ used to construct $A^*_{r_1,r_2}$ and $B^*_{r_1,r_2}$, and the fact that $\hat{\sigma}^{*2}_{r_1,r_2} = \sigma^2_\varepsilon + o^*_p(1)$, the proof of which follows immediately from Equation (20) of CSS page 154. As in the proof of Theorem 1, the stated result for the bootstrap $CGSADF^*$ statistic can then be shown to follow using an application of the CMT by again using the same arguments as in the proof of Theorem 1 of PSY.

References

Astill, S., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2017. Tests for an end-of-sample bubble in financial time series. Econometric Rev. 36, 651–666.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second ed. Springer-Verlag, New York.

Caballero, R.J., Farhi, E., Gourinchas, P.-O., 2008. Financial crash, commodity prices and global imbalances. In: Brookings Papers on Economic Activity. pp. 1–55.

Chang, Y., Park, J.Y., 2002. On the asymptotics of ADF tests for unit roots. Econometric Rev. 21, 431–447.

Chang, Y., Park, J.Y., 2003. A sieve bootstrap for the test of a unit root. J. Time Series Anal. 24, 379–400.

Chang, Y., Sickles, R.C., Song, W., 2013. Bootstrapping Unit Root Tests with Covariates. Working Paper, Rice University.

Chang, Y., Sickles, R.C., Song, W., 2017. Bootstrapping unit root tests with covariates. Econometric Rev. 36, 136–155.

Chen, C.Y.-H., Despres, R., Guo, L., Renault, T., 2019. What makes cryptocurrencies special? Investor sentiment and return predictability during the bubble. Available at SSRN: https://ssrn.com/abstract=3398423.

Davidson, R., MacKinnon, J., 2000. Bootstrap tests: how many bootstraps? Econometric Rev. 19, 55–68.

Diba, B.T., Grossman, H.I., 1988. Explosive rational bubbles in stock prices? Am. Econ. Rev. 78, 520–530.

Elliott, G., Jansson, M., 2003. Testing for unit roots with stationary covariates. J. Econometrics 115, 75–89.

Evans, G.W., 1991. Pitfalls in testing for explosive bubbles in asset prices. Am. Econ. Rev. 81, 922–930.

Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: A new measure. J. Futures Mark. 15, 265–302.

Hansen, B.E., 1995. Rethinking the univariate approach to unit root testing: Using covariates to increase power. Econom. Theory 11, 1148–1171.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

Hansen, B.E., 2000. Sample splitting and threshold estimation. Econometrica 68, 575–603.

Harvey, D.I., Leybourne, S.J., Sollis, R., 2017. Improving the accuracy of asset price bubble start and end date estimators. J. Empir. Financ. 40, 121–138.

Harvey, D.I., Leybourne, S.J., Sollis, R., Taylor, A.M.R., 2016. Tests for explosive financial bubbles in the presence of non-stationary volatility. J. Empir. Finance 38, 548–574.

Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: a comparison of alternative methods. J. Financ. Econom. 10, 198–231.

Kilian, L., Murphy, D.P., 2014. The role of inventories and speculative trading in the global market for crude oil. J. Appl. Econometrics 29, 454–478.

Kim, K.H., Kim, T., 2016. Capital asset pricing model: A time-varying approach. J. Empir. Financ. 37, 268–281.

Palm, F.C., Smeekes, S., Urbain, J.-P., 2008. Bootstrap unit-root tests: comparison and extensions. J. Time Series Anal. 29, 371–401.

Park, J.Y., Phillips, P.C.B., 1989. Statistical inference in regressions with integrated processes: Part 2. Econom. Theory 5, 95–131.

Phillips, P.C.B., Shi, S.-P., 2018. Financial bubble implosion and reverse regression. Econom. Theory 34, 705–753.

Phillips, P.C.B., Shi, S.-P., Yu, J., 2015. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the SP500. Internat. Econom. Rev. 56, 1043–1078.

Phillips, P.C.B., Wu, Y., Yu, J., 2011. Explosive behavior in the 1990s Nasdaq: when did the exuberance escalate asset values? Int. Econ. Rev. 52, 201–226.

Phillips, P.C.B., Yu, J., 2011. Dating the timeline of financial bubbles during the subprime crisis. Quant. Econ. 2, 455–491.

Shiller, R.J., 2015. Irrational Exuberance, third ed. Princeton University Press, New Jersey.

Singleton, K.J., 2014. Investor flows and the 2008 boom/bust in oil prices. Manage. Sci. 60, 300–318.

Tsvetanov, D., Coakley, J., Kellard, N., 2016. Bubbling over! the behaviour of oil futures along the yield curve. J. Empir. Financ. 38, 516–533.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.

White, M., 2015. Cyclical and structural change in the UK housing market. J. Eur. Real Estate Res. 8, 85–103.

Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. J. Bus. Econ. Stat. 10, 251–270.