SPECIFICATION SENSITIVITY IN RIGHT-TAILED UNIT ROOT TESTING FOR EXPLOSIVE BEHAVIOR

By

Peter C.B. Phillips, Shu-Ping Shi, and Jun Yu

January 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1842

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behavior*

Peter C. B. Phillips
Yale University, University of Auckland,
University of Southampton & Singapore Management University

Shu-Ping Shi
The Australian National University

Jun Yu
Singapore Management University

November 23, 2011

Abstract

Right-tailed unit root tests have proved promising for detecting exuberance in economic and financial activities. Like left-tailed tests, the limit theory and test performance are sensitive to the null hypothesis and the model specification used in parameter estimation. This paper aims to provide some empirical guidelines for the practical implementation of right-tailed unit root tests, focussing on the sup ADF test of Phillips, Wu and Yu (2011), which implements a right-tailed ADF test repeatedly on a sequence of forward sample recursions. We analyze and compare the limit theory of the sup ADF test under different hypotheses and model specifications. The size and power properties of the test under various scenarios are examined in simulations and some recommendations for empirical practice are given. Empirical applications to the Nasdaq and to Australian and New Zealand housing data illustrate these specification issues and reveal their practical importance in testing.

Keywords: Unit root test; Mildly explosive process; Recursive regression; Size and power.
JEL classification: C15, C22

1 Introduction

In left-tailed unit root testing, results are often sensitive to model formulation. In effect, the maintained hypothesis or *technical lens* through which the properties of the data are explored can influence outcomes in a major way. Formulating a suitable maintained hypothesis is particularly

*We are grateful to Heather Anderson, Farshid Vahid, Hashem Pesaran, and Tom Smith for helpful discussions. Phillips acknowledges support from the NSF under Grant No. SES 09-56687. Shi acknowledges the Financial Integrity Research Network (FIRN) for funding support. Peter C.B. Phillips, Yale University, University of Auckland, University of Southampton & Singapore Management University; Email: peter.phillips@yale.edu. Shuping Shi, Research School of Economics, The Australian National University; Email: shuping.shi@anu.edu.au. Jun Yu, Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business, Singapore Management University, Singapore 178903; Email: yujun@smu.edu.sg.

difficult in the presence of nonstationarity because of the different roles that parameters can play under the null hypothesis of a unit root and the alternative of stationarity. Many of these issues of formulation have already been extensively studied in the literature on left-tailed unit root testing.

Suppose, for example, that the null hypothesis is that the data is difference stationary and the alternative is that the data is stationary. If we run the ADF regression

$$R_1 : y_t = \delta y_{t-1} + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right), \tag{1}$$

and test the null $\delta = 1$ against the alternative $\delta < 1$, the formulation (implicitly) assumes that the mean of y_t is zero under the alternative. Under this lens any evidence of a non-zero mean in the sample is likely to be interpreted as evidence in favor of the null and the test procedure tends to have poor power. A more suitable lens allows for a non zero mean in y_t under the alternative through the regression

$$R_2 : y_t = \alpha + \delta y_{t-1} + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right), \tag{2}$$

even though α may be zero under the null. Similarly, if the null is difference stationarity and the alternative trend stationarity, then the regression model (2) will be inappropriate because an empirical trend may be misinterpreted as evidence of a unit root, leading to the augmented formulation

$$R_3 : y_t = \alpha_0 + \alpha_1 t + \delta y_{t-1} + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right), \tag{3}$$

where we can test the null $\delta = 1$ against the alternative $\delta < 1$, even if $\alpha_1 = 0$ under the null. Use of the maintained hypothesis R_3 allows for both a unit root with drift ($\alpha_0 \neq 0$ and $\alpha_1 = 0$) under the null and trend stationarity ($\alpha_0 \neq 0$ and $\alpha_1 \neq 0$) under the alternative. Similar issues, of course, arise with more complex maintained hypotheses that allow for trend breaks and other deterministic components. The regression model of a left-tailed unit root test (against stationary or trend stationary alternatives) needs to nest the alternative hypothesis.[1]

Right-tailed unit root tests are also of empirical interest, particularly in detecting explosive or mildly explosive alternatives. For example, to find evidence of financial bubbles, Diba and Grossman (1988) applied right-tailed unit root tests to the fully sampled data. Phillips, Wu and Yu (2011, PWY hereafter) suggested sequential implementations of right-tailed unit root tests to recursive

[1]Similar arguments can be found in Dickey, Bell and Miller (1986) and Davidson and MacKinnon (2004).

subsamples; see also Phillips and Yu (2011). As in left-tailed unit root testing, the formulation of the null and alternative hypotheses and the regression model specification are important in right-tailed tests. Different suggestions appear in the literature and no empirical guidelines have yet been offered. For example, Diba and Grossman used the regression model (3) whereas PWY employed model (2). Further, Diba and Grossman did not allow for bubble crashes in the alternative whereas various collapse mechanisms were considered in Evans (1991) and Phillips and Yu (2009).

The present paper examines appropriate ways of formulating regressions for right-tailed unit root tests to assess empirical evidence for explosive behavior in the context of PWY test procedures. Other tests for explosive behavior are possible and many of these have been recently evaluated in extensive simulations by Homm and Breitung (2011). The simulations in that paper show that, while ex post analysis of the full sample data favors Chow type unit root tests for the detection of break points in the transition between unit root and explosive behavior, recursive tests such as those in PWY perform well in early detection of such transitions and are preferable in this anticipative role as a monitoring system. Homm and Breitung (2011) also confirm that the PWY tests are more robust in the detection of multiple bubble episodes than the other tests they considered. The primary intent of PWY was to develop recursive procedures that could assess whether Greenspan's remark on financial exuberance had empirical content at the time he made that statement in December 1995. It is in this context as an early warning device in market surveillance that the PWY tests were developed. The specification issues raised here apply equally well to other break tests for financial exuberance.

The rest of the paper is organized as follows. Section 2 discusses the appropriate choices for the null and alternative hypotheses and the regression model. Section 3 derives the limit distributions of the ADF statistic. Section 4 discusses several explosive models, all subject to crashes, for the alternative hypothesis. The sequential right-tailed ADF test, along with its finite sample and limit distributions, are explored in Section 5. Section 6 reports size and power properties for the sequential right-tailed ADF test. Using the proposed model formulations we apply the test to Nasdaq market data and to the Australian and New Zealand housing markets in Section 7. Section 8 concludes.

2 Formulating Hypotheses and Regression

The literature on right-tailed unit root testing has employed several different specifications for the null hypothesis. In PWY the null hypothesis is

$$H_{01} : y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \left(0, \sigma^2\right),$$

so that Δy_t has mean zero and y_t has no deterministic trend. On the other hand, Diba and Grossman (1988) used the null model

$$H_{03} : y_t = \widetilde{\alpha} + y_{t-1} + \varepsilon_t, \text{ with a constant intercept } \widetilde{\alpha},$$

so that y_t has deterministic trend behavior when $\widetilde{\alpha} \neq 0$ under this null.

A model that bridges these two null hypothesese involves a weak (local to zero) intercept with the form

$$H_{02} : y_t = \widetilde{\alpha} T^{-\eta} + y_{t-1} + \varepsilon_t \text{ with } \eta \geq 0. \tag{4}$$

Here y_t has an array formulation, the mean of Δy_t is $\widetilde{\alpha} T^{-\eta} = O(T^{-\eta})$, and y_t has a deterministic trend component of the form $\widetilde{\alpha} \frac{t}{T^\eta}$ whose magnitude depends on the sample size and the parameter η. The null model H_{02} becomes H_{01} when $\eta \to \infty$ and H_{03} when $\eta \to 0$.

Similarly, different alternative hypotheses have been used in the literature on the right-tailed unit root tests. The most obvious ones are the following explosive processes:

$$H_{A1} : y_t = \delta y_{t-1} + \varepsilon_t, \ \delta > 1, \tag{5}$$

$$H_{A2} : y_t = \widetilde{\alpha} + \delta y_{t-1} + \varepsilon_t, \ \delta > 1, \tag{6}$$

$$H_{A3} : y_t = \widetilde{\alpha} + \gamma t + \delta y_{t-1} + \varepsilon_t, \ \delta > 1. \tag{7}$$

These three models mirror alternatives considered in left-tailed unit root tests where $\delta < 1$. However, for left-tailed tests model (5) with $\delta < 1$ is rarely used because it restricts the mean of y_t to zero.

Explosive processes have a long history. In economics, Hicks (1950) suggested the possibility of explosive cyclical behavior contained by certain structural floors and ceilings with the cycles arising from multiplier-accelerator dynamics. In statistics, White (1958) and Anderson (1959) studied the asymptotic properties of the least squares (LS) estimator under (5). In recent work, Phillips and Magdalinos (2007) suggested a mildly explosive process of the type

$$H_{A4} : y_t = \delta_T y_{t-1} + \varepsilon_t \text{ with } \delta_T = 1 + cT^{-\theta}, \tag{8}$$

where $c > 0$, $\theta \in (0, 1)$, T is the sample size and δ_T is a moving parameter sequence. This model is called mildly explosive because the autoregressive coefficient δ_T is in an explosive region of unity (so that $\delta_T \to 1+$ as $T \to \infty$) that lies beyond the usual 'local to unity' interval where $\delta_T = 1 + \frac{c}{T}$ for which behavior of the process is similar to that of a unit root process. Under H_{A4}, the behavior of y_t resembles that of an explosive time series rather than that of a unit root process.

Model (6) is formulated with a non-zero intercept and produces a dominating deterministic component that has an empirically unrealistic explosive form (Phillips and Yu, 2009, PY hereafter). Similar characteristics apply a fortiori in the case of the inclusion of a deterministic trend term in model (7). These forms are unreasonable for most economic and financial time series and an empirically more realistic description of explosive behavior is given by models (5) and (8), which are both formulated without an intercept or a deterministic trend.

The empirical regression of the right-tailed unit root test given in Diba and Grossman (1988) is R_3. This regression has both a constant as well as a deterministic trend. Since the presence of either of these two terms is empirically unrealistic when $\delta > 1$, regression R_3 is not suitable for right-tailed unit root testing. On the other hand, due to the fact that neither a constant nor a deterministic trend is included in regression R_1, that model does not allow for deterministic-trend-like behavior when $\delta = 1$. Suppose we run R_1 to investigate evidence for mildly explosive behavior as in (8). Analogous to the effects of a left-tailed unit root test, in a regression of the form R_1 any evidence of non-zero mean in Δy_t may be misjudged as evidence in favor of the alternative - in this case, mildly explosive behavior. Thus, R_1 also seems inappropriate. By contrast, regression R_2 is empirically more realistic and PWY implemented a right-tailed unit root test using this regression formulation.

In view of the above discussion, we recommend that right-tailed unit root tests may be suitably formulated with a null hypothesis H_{02} and an empirical regression R_2. Since H_{02} depends on η, we discuss the asymptotic distribution of the test statistic and examine the size and the power properties of the the right-tailed unit root test for different settings of η in H_{02}. Simulation findings reported below provide further guidelines for the selection of the null and the regression model with associated test critical values.

3 Full-sample Right-Tailed Unit Root Tests

Right-tailed unit root tests, like their left-tailed counterparts, have asymptotic distributions that depend on the specification of the null hypothesis and regression model. As discussed above, one suitable regression model for right-tailed testing is R_2 and an empirically reasonable null is a unit root process with a drift of the form $\widetilde{\alpha}\frac{t}{T^{\eta}}$, arising from H_{02}. The right-tailed unit root test discussed in this section is the ADF test applied to the full sample. Other unit root tests can be studied in exactly the same manner. Note that the magnitude of the drift is inversely related to parameter η.

Proposition 3.1 *If $\eta > 0.5$, the asymptotic distribution of the ADF statistic is*

$$ADF \xrightarrow{L} \frac{\frac{1}{2}\left[W^2(1) - 1\right] - W(1)\int_0^1 W(s)\,ds}{\left\{\int_0^1 W^2(s)\,ds - \left[\int_0^1 W(s)\,ds\right]^2\right\}^{1/2}} := F_1(W), \tag{9}$$

where W is a standard Wiener process and $\xrightarrow{L}$ denotes the convergence in distribution; If $\eta < 0.5$, then the asymptotic ADF distribution is

$$ADF \xrightarrow{L} \left[\int_0^1 s dW(s) - \int_0^1 W(s)\,ds\right]\left(\int_0^1 s^2 ds\right)^{-1/2} := F_2(W). \tag{10}$$

The proof of this proposition is nested in that of Proposition 5.1 (Appendix A).

Remark 3.1 *The asymptotic ADF distribution when $\eta > 0.5$ is identical to that of the PWY formulation despite the inclusion of an intercept in the null model. The reason the intercept does not affect the limit distribution is that the implied drift in the process has smaller order than the stochastic trend.*

Remark 3.2 *Suppose the null hypothesis is specified as H_{03}. The asymptotic ADF distribution in this case*[2] *is identical to that of the case when $\eta < 0.5$, (10). Here the implied drift has higher order of magnitude and behaves like a linear deterministic trend.*

Remark 3.3 *The asymptotic ADF distribution when $\eta = 0.5$ is*

$$ADF \xrightarrow{L} (D_\sigma - A_\sigma C_\sigma)\left(B_\sigma - A_\sigma^2\right)^{-1/2}, \tag{11}$$

[2]The asymptotic ADF distribution under this case is well documented in the unit root literature; see Phillips (1987) and Phillips and Perron (1988)

with $A_\sigma = \frac{1}{2} + \sigma \int_0^1 W(s)\,ds, B_\sigma = \frac{1}{3} + \sigma^2 \int_0^1 W(s)^2\,ds + 2\sigma \int_0^1 W(s)\,sds, C_\sigma = W(1)$ *and* $D_\sigma = \left[W(1) - \int_0^1 W(s)\,ds\right] + \frac{1}{2}\sigma \left[W(1)^2 - 1\right]$. *Importantly, the limit theory depends on the nuisance parameter* σ *and hence it is not invariant unless we include a trend in the regression or adjust for the trend in some other way (as, for example, in Schmidt and Phillips, 1992, and Phillips and Lee, 1996).*[3]

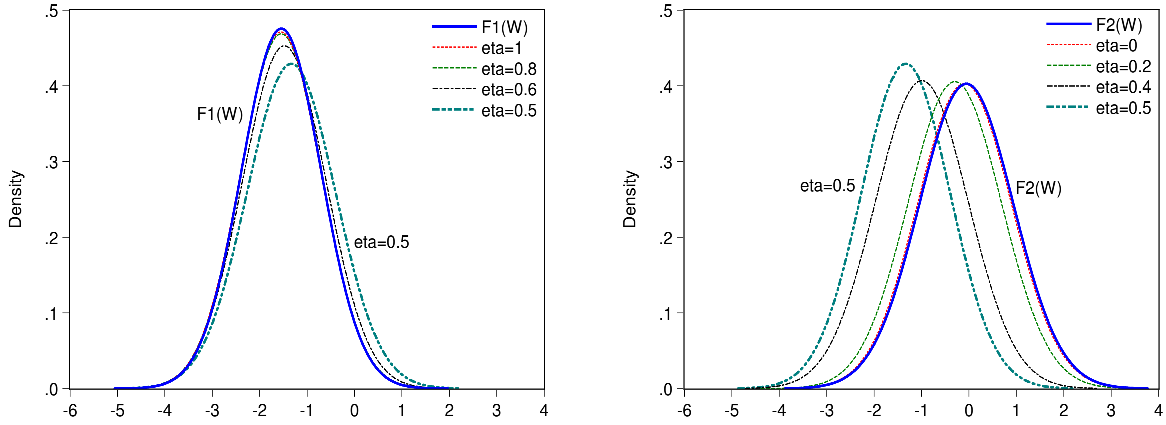
We now examine the finite sample distributions of the ADF statistic obtained by Monte Carlo simulations. The limit distributions are obtained by numerical simulation using Wiener process approximations based on partial sums of a standard normal with $5,000$ steps. In both cases 2000 replications are used.

Figure 1 displays the finite sample distributions of the ADF statistic when the sample size $T = 400$ and $\eta = \{1, 0.8, ..., 0.2, 0\}$. The dotted lines in the figure are the finite sample distributions and the bold solid lines are the limit distributions. When $\eta > 0.5$ the finite sample distribution moves towards the asymptotic distribution $F_1(W)$ as η increases. Nevertheless, the discrepancies among the finite sample distributions with $\eta = \{1, 0.8, 0.6\}$ are negligible. Second, when $\eta < 0.5$ the discrepancies among the finite sample distributions with $\eta = \{0.4, 0.2, 0\}$ are marked. Nonetheless, there is apparent convergence toward the limit distribution $F_2(W)$ as η decreases. Third, the finite sample distribution of the ADF statistic with $\eta = 0.5$ is significantly different from the corresponding distributions when $\eta \neq 0.5$.

4 Specifications for Explosive Behavior

Two specifications for the alternative hypothesis, both formulated without an intercept or a deterministic trend, are given by model (5) and model (8) in Section 2. Neither model has structural breaks. But as argued in Evans (1991, page 924) "bubbles do not appear to be empirically plausible unless there is a significant chance that they will collapse after reaching high levels." This argument is consistent with other models of explosive processes such as the business cycle model of Hicks (1950), where each cycle has an explosive expansion phase and a subsequent downswing due to disinvestment proceeding at the rate of deterioration of capital. Thus, more complete specification of the alternative hypothesis requires the inclusion of a downswing or bubble collapse process. This section considers such extensions within the context of some simple time series models.

[3] The asymptotic ADF distributions do not depend on $\tilde{\alpha}$. In what follows, we set $\tilde{\alpha}$ to unity.



(a) $\eta \geq 0.5$

(b) $\eta \leq 0.5$

Figure 1: The finite sample distribution of the ADF statistic when $T = 400$ and $\eta = \{1, 0.8, 0.6, 0.5, 0.4, 0.2, 0\}$.

4.1 A periodically collapsing explosive process

The DGP proposed by Evans (1991) consists of a market fundamental component P_t^f, which follows a random walk process

$$P_t^f = \tilde{u} + P_{t-1}^f + \sigma_f \varepsilon_t, \varepsilon_t \overset{iid}{\sim} N(0,1) \tag{12}$$

and a periodically collapsing explosive bubble component such that

$$B_{t+1} = \rho^{-1} B_t \varepsilon_{B,t+1}, \qquad \text{if} \;\; B_t < b \tag{13}$$

$$B_{t+1} = \left[\zeta + (\pi\rho)^{-1} \theta_{t+1} (B_t - \rho\zeta)\right] \varepsilon_{B,t+1}, \quad \text{if} \;\; B_t \geq b \;, \tag{14}$$

where $\rho^{-1} > 1$ and $\varepsilon_{B,t} = \exp\left(y_t - \tau^2/2\right)$ with $y_t \overset{iid}{\sim} N\left(0, \tau^2\right)$. θ_t follows a Bernoulli process which takes the value 1 with probability π and 0 with probability $1-\pi$. ζ is the remaining size after the bubble collapse. The bubble component has the property that $\mathbb{E}_t(B_{t+1}) = \rho^{-1} B_t$. By construction, the bubbles collapse completely in a single period when triggered by the Bernoulli process realization.

The market fundamental equation, (12), is equivalent to the combination of a random walk dividend process and the Lucas asset pricing equation

$$D_t = \mu + D_{t-1} + \varepsilon_{Dt}, \;\; \varepsilon_{Dt} \overset{iid}{\sim} N\left(0, \sigma_D^2\right) \tag{15}$$

$$P_t^f = \frac{\mu\rho}{(1-\rho)^2} + \frac{\rho}{1-\rho} D_t, \tag{16}$$

where μ is the drift of the dividend process and σ_D^2 the variance of the dividend innovations. The drift ($\tilde{u}$) of the market fundamental process is $\mu\rho\left(1-\rho\right)^{-1}$ and the standard deviation is $\sigma_f = \sigma_D\rho\left(1-\rho\right)^{-1}$. In Evans (1991), the parameter values for μ and σ_D^2 were matched to the sample mean and sample variance of the first differences of real S&P500 dividends from 1871 to 1980. The value for the discount factor ρ is equivalent to a 5% annual interest rate. So the parameter settings in Evans (1991) correspond to a yearly frequency. In accordance with our empirical application, we consider a set of the parameters calibrated to monthly data. Correspondingly, the parameters μ and σ_D^2 are set to be the sample mean and the sample variance of the monthly first differences of real Nasdaq dividends as described in the application section (normalized to unity at the beginning of the sample period). These are $\mu = 0.0020$ and $\sigma_D^2 = 0.0034$, respectively. The discount factor equals 0.985. We can then calculate the values of $\tilde{u}$, σ_f, P_0^f based on those of μ, σ_D^2, D_0.

The settings of the parameters in the bubble components (13) - (14) are the same as those in Evans (1991). The asset price P_t is equal to the sum of the market fundamental component and the bubble component, namely $P_t = P_t^f + \kappa B_t$, where κ controls the relative magnitudes of these two components. The parameter settings are given in Table 1 for *yearly* and *monthly* data.

Table 1: Parameter settings

	$\tilde{u}$	σ_f	P_0^f	ρ	b	B_0	π	ζ	τ	κ
Yearly	0.740	7.869	41.195	0.952	1	0.50	0.85	0.50	0.05	20
Monthly	0.131	3.829	94.122	0.985	1	0.50	0.85	0.50	0.05	150

Figure 2a shows a typical realization of this DGP with yearly parameter settings (sample size $T = 100$) and Figure 2b gives a corresponding realization for monthly data ($T = 200$).

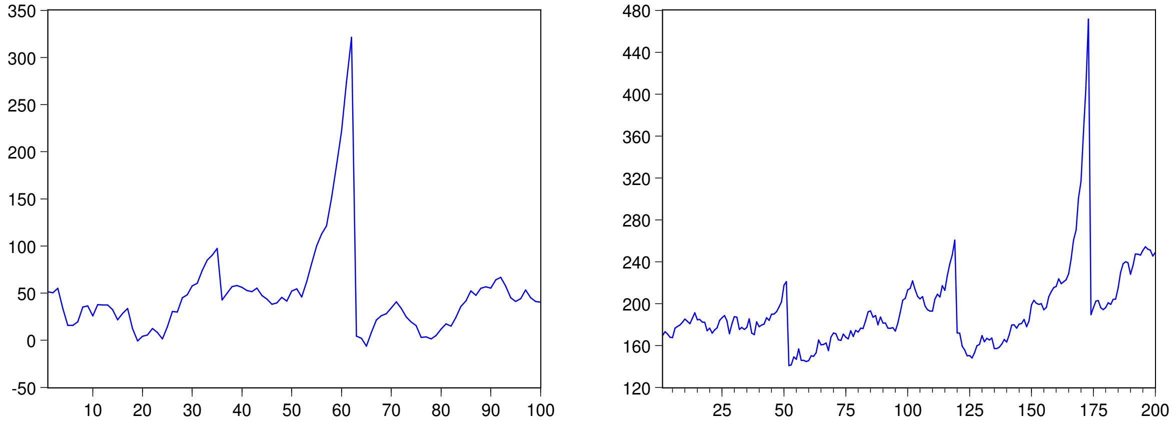
4.2 A locally explosive process

Locally explosive behavior can be expressed in terms of an AR process with time-varying coefficients of the form

$$y_t = u_t + \rho_t y_{t-1} + \sigma_t \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N\left(0,1\right), \tag{17}$$

where u_t is the intercept, ρ_t is the autoregressive coefficient and σ_t is the disturbance standard deviation.

In PY, it is assumed that $u_t = 0$ and $\sigma_t = \sigma$ for all $t = 1, \cdots, T$. For the bubble expansion period, the autoregressive coefficient ρ_t is exceeds unity and has the form $\rho_t = 1 + cT^{-\alpha}$ with $c > 0$ and $\alpha \in (0,1)$, but otherwise is equal to unity ($\rho_t = 1$). More specifically, PY used the following



(a) with yearly parameter setting

(b) with monthly parameter setting

Figure 2: Simulated time series based on Evans' DGP

model formulation with explicit break points at (T_e, T_f)

$$y_t = y_{t-1}\mathbf{1}(t < T_e) + \rho_T y_{t-1}\mathbf{1}(T_e \leq t \leq T_f) + \left(\sum_{k=T_f+1}^{t} \varepsilon_k + y_{T_f}^*\right)\mathbf{1}(t > T_f) + \varepsilon_t \mathbf{1}(t \leq T_f) \tag{18}$$

where $\rho_T = 1 + cT^{-\alpha}, y_{T_f}^* = y_{T_e} + y^*$ with $y^* = O_p(1)$, $\mathbf{1}(\cdot)$ is an indicator function, T_e is the origination date of the bubble and T_f is the termination date. In Model (18) y_t only crashes once at T_f. However, it is easy to generalize Model (18) to allow for more periods of explosive behavior and subsequent crashes, as discussed in Phillips, Shi and Yu (2011).

4.3 A modified locally explosive process

In Model (18) y_t is re-initialized to y_{T_e} (with an $O_p(1)$ perturbation y^*) upon the bubble collapse at T_f. This feature of a one-period crash is shared by the periodically collapsing model of Evans. Although bubbles frequently collapse rapidly, in most cases it is unrealistic to specify a complete collapse within a single period. For instance, according to PWY, the dot-com bubble began to collapse in March 2000 and the termination date was between September 2000 and March 2001. To accommodate a transitional rather than complete collapse, we may assume that y_t switches to a (mildly) stationary transition regime when the bubble starts to burst. The new DGP has the form

$$y_t = \begin{cases} u_1 + y_{t-1} + \sigma_1 \varepsilon_t, & t \in [1, T_e) \cup (T_c, T] \\ \phi_T y_{t-1} + \sigma_2 \varepsilon_t, & t \in [T_e, T_f] \\ \gamma_T y_{t-1} + \sigma_3 \varepsilon_t, & t \in (T_f, T_c] \end{cases}, \tag{19}$$

where T_c marks the conclusion of the bubble collapse period, $\phi_T = 1 + c_1 T^{-\alpha}$ and $\gamma_T = 1 - c_2 T^{-\beta}$ with $c_1, c_2 > 0$ and $\alpha, \beta \in [0, 1)$. The formulation of the AR coefficients ϕ_T and γ_T both involve mild deviations from unity in the sense of Phillips and Magdalinos (2007), one in the explosive direction for the bubble expansion, the other in the stationary direction for the bubble collapse. Equation (19) corresponds with (17) if we set

$$\begin{aligned} u_t &= s_{nt} u_1, \\ \rho_t &= s_{nt} + s_{bt} \phi_T + s_{ct} \gamma_T, \\ \sigma_t &= s_{nt} \sigma_1 + s_{bt} \sigma_2 + s_{ct} \sigma_3, \end{aligned}$$

where $s_{nt} = \mathbf{1}(t \in [0, T_e) \cup (T_c, T])$, $s_{bt} = \mathbf{1}(t \in [T_e, T_f])$, $s_{ct} = \mathbf{1}(t \in (T_f, T_c])$, which are the respective regime indicators for market fundamentals, bubble expansion, and bubble collapse episodes.

We illustrate the process (19) by setting the market fundamental regime as in Table 1 (monthly): $y_0 = 94.122$, $u_0 = 0.131$, $\sigma_1 = 3.829$. The other parameters relating to the bubble expansion and collapse regimes are set to be: $c_1 = c_2 = 1$, $\alpha = 0.6$, $\beta = 0.5$, $\sigma_2 = \sigma_1$, $\sigma_3 = 2\sigma_1, T_e = [0.6T]$, $T_f = [0.70T]$, $T_c = [0.75T]$. (Various additional settings for the parameters $\alpha, \beta, T_e, T_f, T_c$ are considered in size and power comparisons later in the paper.) The sample size $T = 200$. With these settings, the implied autoregressive coefficients are $\phi_T = 1.042$ and $\gamma_T = 0.929$. Figure 3 exhibits a realization of the DGP with these parameter settings. Compared with the PY and Evans DGPs, the bubble collapse period of this DGP is a gradual one and may be more realistic for empirical implementation.

5 The Sup ADF Test

Evans (1991) argued that right-tailed unit root tests, when applied to the full sample, have little power to detect periodically collapsing bubbles and demonstrated this effect in simulations. The low power of standard unit root tests is due to the fact that periodically collapsing bubble processes behave rather like an I(1) process or even a stationary linear autoregressive process when the probability of bubble collapse is non negligible. PY provided a theoretical underpinning of this

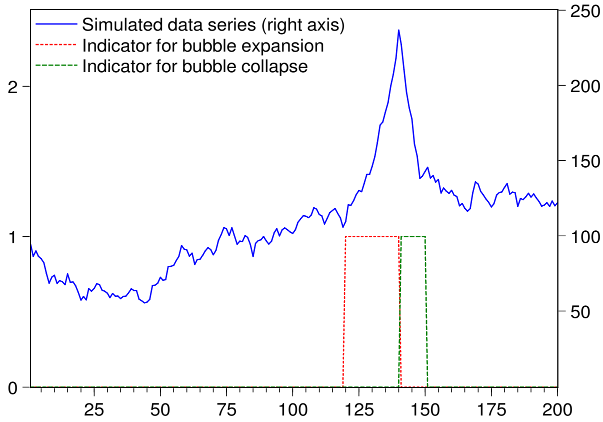


Figure 3: Simulated time series based on equation (19)

argument by deriving the order of the ADF t-statistic when the data are generated from the locally explosive process (18).

To overcome the problem identified in Evans, PWY proposed the sup ADF (SADF) statistic to test for the presence of explosive behavior in a full sample. In particular, the methods rely on forward recursive regressions coupled with sequential right-sided unit root tests. The sequential tests assess period by period evidence for unit root behavior against explosive alternatives. Suppose the right-tailed ADF test is employed in each period, the test statistic proposed by PWY is the sup value of the corresponding ADF sequence. In this setup, the alternative hypothesis of the test therefore includes both a periodically collapsing explosive behavior and a locally explosive behavior. The null hypotheses are exactly the same as that for the right-tailed unit root test in equation (4).

Suppose r is the window size of the regression (proportional to the full sample size) for the right-tailed unit root test. In the sup ADF test, the window size r expands from r_0 to 1 through recursive calculations. The smallest window size r_0 is selected to ensure that there are sufficient observations to initiate the recursion. The number of observations in the regression is $T_r = [Tr]$, where $[\cdot]$ signifies the integer part of its argument and T is the total number of observations.

The regression model for the sup ADF test is:

$$R_2 : y_t = \alpha + \delta y_{t-1} + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t, \tag{20}$$

where $t = 1, \cdots, T_r$ and k is the lag order. The corresponding ADF t-statistic is denoted by ADF_r. To test for the existence of bubbles, inferences are based on the sup ADF statistic $SADF(r_0) =$

$\sup_{r\in[r_0,1]} ADF_r$. This notation highlights the dependence of $SADF$ on the initialization parameter r_0.

5.1 The limiting distribution of sup ADF

Proposition 5.1 *If $\eta > 0.5$, the asymptotic distribution of the sup ADF statistic is*

$$SADF\left(r_0\right) \xrightarrow{L} \sup_{r\in[r_0,1]} \left\{ \frac{\frac{1}{2}r\left[W\left(r\right)^2 - r\right] - \int_0^r W\left(s\right)dsW\left(r\right)}{r^{1/2}\left\{r\int_0^r W\left(s\right)^2 ds - \left[\int_0^r W\left(s\right)ds\right]^2\right\}^{1/2}} \right\} := F_3\left(W, r_0\right); \quad (21)$$

If $\eta < 0.5$, the sup ADF statistic converges to

$$SADF\left(r_0\right) \xrightarrow{L} \sup_{r\in[r_0,1]} \left\{ \left[\int_0^r sdW\left(s\right) - \int_0^r W\left(s\right)ds\right]\left(\int_0^r s^2 ds\right)^{-1/2} \right\} := F_4\left(W, r_0\right). \quad (22)$$

Remark 5.1 *The proof of Proposition 5.1 is given in Appendix A. The asymptotic SADF distributions are obtained by standard methods using continuous maps. The result implies that the* lim sup *and the* sup lim *operations are equivalent, namely*

$$\lim_{T\to\infty} \sup_{r\in[r_0,1]} \left\{ADF_r\right\} = \sup_{r\in[r_0,1]} \left\{\lim_{T\to\infty} ADF_r\right\}, \quad (23)$$

for both cases.

Remark 5.2 *The asymptotic ADF_r distribution when $\eta < 0.5$ is*

$$ADF_r \xrightarrow{L} \left[\int_0^r sdW\left(s\right) - \int_0^r W\left(s\right)ds\right]\left(\int_0^r s^2 ds\right)^{-1/2}, \quad (24)$$

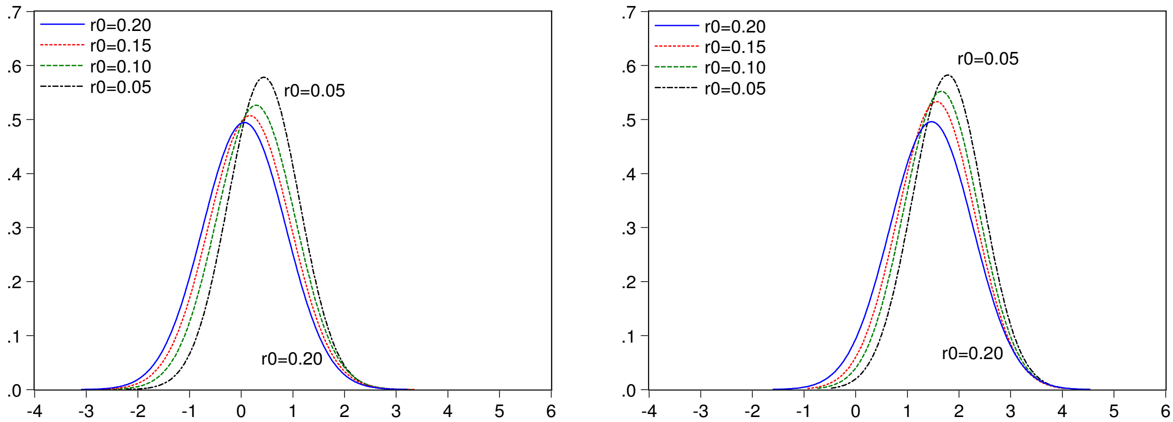
which is distributed as standard normal. Suppose $r_A, r_B \in [r_0, 1]$ and $r_A \neq r_B$, the asymptotic ADF_{r_A} distribution and the asymptotic ADF_{r_B} distribution are correlated due to the fact that both of them are functions of a standard Wiener process.

Remark 5.3 *The asymptotic distribution of the SADF statistic when $\eta = 0.5$ is*

$$SADF\left(r_0\right) \xrightarrow{L} \sup_{r\in[r_0,1]} \left[r^{-1/2}\left(rD_{r,\sigma} - A_{r,\sigma}C_{r,\sigma}\right)\left(rB_{r,\sigma} - A_{r,\sigma}^2\right)^{-1/2}\right],$$

with $A_{r,\sigma} = \frac{1}{2}r + \sigma\int_0^r W\left(s\right)ds$, $B_{r,\sigma} = \frac{1}{3}r^3 + \sigma^2\int_0^r W\left(s\right)^2 ds + 2\sigma\int_0^r W\left(s\right)sds$, $C_{r,\sigma} = W\left(r\right)$ and $D_{r,\sigma} = \left[rW\left(r\right) - \int_0^r W\left(s\right)ds\right] + \frac{1}{2}\sigma\left[W\left(r\right)^2 - r\right]$. Similar to the ADF statistic, the limit theory depends on the nuisance parameters σ.

Figures 4 (a) and (b) examine the sensitivity of the asymptotic distributions of $SADF$ when $\eta > 0.5$ and $\eta < 0.5$ with respect to r_0. The distributions are obtained using $2,000$ replications, approximating the Wiener process by partial sums of standard normal variates with $5,000$ steps. The smallest window size r_0 is set to $\{0.2, 0.15, 0.10, 0.05\}$.



(a) $F_3(W, r_0)$ (b) $F_4(W, r_0)$

Figure 4: The asymptotic distributions of the SADF statistic with $r_0 = \{0.20, 0.15, 0.10, 0.05\}$.

Figure 4a displays the asymptotic distributions when $\eta > 0.5$ (i.e. $F_3(W, r_0)$) while Figure 4b is for the case $\eta < 0.5$ (i.e. $F_4(W, r_0)$). Under both cases, the asymptotic distributions of the SADF statistic move sequentially to the right as r_0 decreases.[4] In addition, the asymptotic distribution $F_3(W, r_0)$ has larger values for the 90%, 95% and 99% quantiles. For example, the 95% critical values of $F_3(W, r_0)$ with $r_0 = \{0.2, 0.15, 0.10, 0.05\}$ are respectively $1.39, 1.44, 1.54, 1.58$ while those of $F_4(W, r_0)$ are respectively $2.79, 2.86, 2.91, 2.96$. Obviously, the critical values are sensitive to r_0.

5.2 The finite sample distribution of sup ADF

The finite sample distribution of the SADF statistic depends on the sample size T, the value of the drift in the null hypothesis (depending on T and η) and the smallest window size r_0. Figure 5 describes the finite sample distributions of the SADF statistic when $T = 400$, $r_0 = 0.1$, $\widetilde{\alpha} = 1$, and $\eta = \{1, 0.8, 0.6, 0.5, 0.4, 0.2, 0\}$. The bold solid lines are the asymptotic distributions and the dotted lines are the finite sample distributions.

[4] Intuitively, when r_0 is smaller, the feasible range of r (i.e. $[r_0, 1]$) becomes wider and hence the parameter space of the distribution of $\lim_{T\to\infty} ADF_r$ expands. The asymptotic SADF distribution, which applies the sup function to the aforementioned distribution, then moves sequentially towards the right as r_0 decreases.

We observe a similar pattern as in Figure 1. For a given T and r_0, the finite sample distribution moves towards $F_3(W, 0.1)$ as η increases and shifts towards $F_4(W, 0.1)$ as η decreases. An obvious separation occurs when $\eta = 0.5$. The discrepancy among the finite sample distributions is negligible with $\eta = \{0.6, 0.8, 1\}$, but becomes considerably larger when $\eta = \{0.4, 0.2, 0\}$. Like the finite sample ADF distribution described in Figure 1, the finite sample SADF distribution is invariant to η when $\eta > 0.5$ while it varies significantly with η when η is less than 0.5.

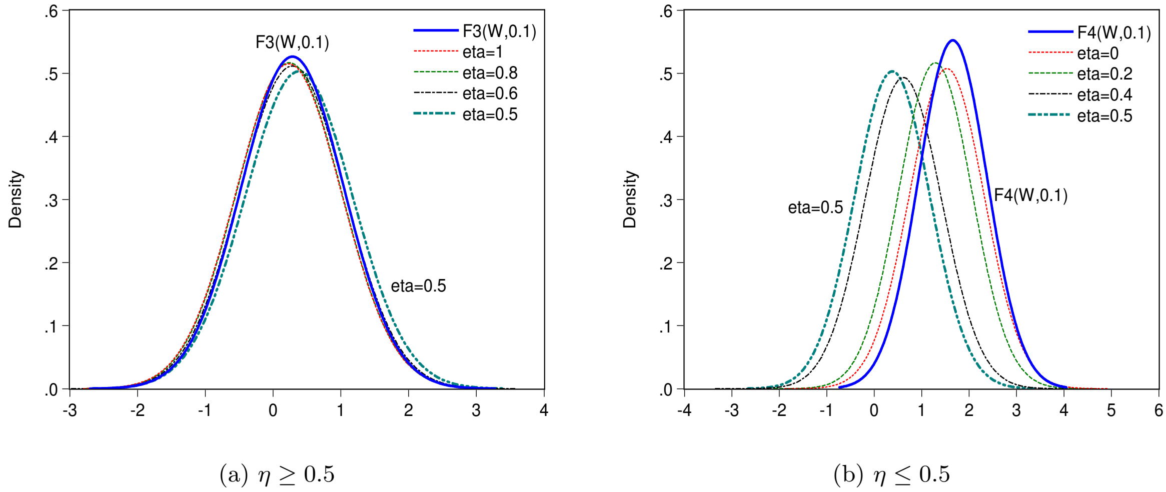


Figure 5: The finite sample distributions of the SADF statistic when $T = 400, r_0 = 0.1$ and $\eta = \{1, 0.8, 0.6, 0.5, 0.4, 0.2, 0\}$.

6 Size and Power Comparison

The 90%, 95% and 99% quantiles of the asymptotic distributions of the SADF statistic when $\eta > 0.5$ and $\eta < 0.5$ (i.e. $F_3(W, r_0)$ and $F_4(W, r_0)$) are presented in Table 2. As before, critical values are obtained by simulations with 2,000 replications of Wiener processes in terms of partial sums of standard normal variates with $5,000$ steps.

Table 3 gives sizes for the SADF test based on nominal asymptotic critical values with sample sizes $T = 100, 200$ and 400. The nominal size is 5%. The DGP is specified according to the respective null hypotheses with $\widetilde{\alpha} = 1$, $\eta = 1, 0.8, 0.6, 0.4, 0.2, 0$. The number of replications is 2,000. The lag order is determined by BIC with maximum lag length 12. The smallest window size has 40 observations. Table 3 shows that for all cases of $\eta > 0.5$ there are no obvious size distortion

Table 2: Asymptotic critical values of the SADF statistic (against explosive alternative)

	$F_3(W, r_0)$			$F_4(W, r_0)$		
	90%	95%	99%	90%	95%	99%
$r_0 = 0.4$	0.88	1.20	1.87	2.27	2.62	3.20
$r_0 = 0.2$	1.10	1.39	1.95	2.48	2.79	3.39
$r_0 = 0.1$	1.23	1.54	2.04	2.58	2.92	3.42

Note: asymptotic critical values are obtained using 2,000 replications and partial sums with 5,000 steps.

Table 3: Sizes of the SADF test (using asymptotic critical values). The data generating process is specified according to the respective null hypothesis. The nominal size is 5%.

	$\eta > 0.5$			$\eta < 0.5$		
	$\eta = 1$	$\eta = 0.8$	$\eta = 0.6$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
$T = 100$ and $r_0 = 0.4$	0.043	0.043	0.046	*0.003*	*0.018*	0.045
$T = 200$ and $r_0 = 0.2$	0.042	0.047	0.050	*0.004*	*0.022*	0.049
$T = 400$ and $r_0 = 0.1$	0.052	0.045	0.043	*0.003*	*0.020*	0.042

Note: size calculations are based on 2,000 replications.

when using the asymptotic critical values,[5] whereas there are significant size distortions for some cases with the value of η smaller than 0.5. On the latter point, it is noted that we did not observe obvious size distortion for the case of $\eta = 0$. However, the size distortion becomes progressively more severe when the value of η increases from 0 to 0.5. For example, the size of the SADF test is 0.042, 0.020 and 0.003 for $\eta = \{0, 0.2, 0.4\}$ respectively when the sample size $T = 400$.[6]

6.1 Periodically collapsing explosive behavior

To calculate power we specify several alternatives. First, we assume the DGP is Evans (1991) periodically collapsing explosive process, with both yearly and monthly parameters settings (see Table 1). The sample sizes considered for those two parameters settings are $T = \{100, 200\}$ and $T = \{100, 200, 400\}$ respectively. For each parameters and sample size setting, we calculate powers of the sup ADF test under four different specifications in the null hypothesis: $\eta > 0.5$,[7] $\eta = 0.4$, $\eta = 0.2$ and $\eta = 0$, all with $\widetilde{\alpha} = 1$. The powers under cases of $\eta > 0.5$ and $\eta = 0$ are calculated

[5] There are significant size distortions when using the significance test proposed by Campbell and Perron (1991) (with the maximum lag length 12) to determine the lag order. For example, the size of the SADF test when $\eta = 1$ is 0.115, 0.131 and 0.114 for $T = 100, 200, 400$ respectively.

[6] We observe similar patterns of size distortion when keeping the smallest fractional window size $r_0 = 0.4$ for all sample sizes. However, when T is large, there is some advantage to using a small value for r_0 so that the sup ADF test does not miss any opportunity to capture an explosive phase.

[7] This is due to the observation that as long as η is greater than 0.5, the discrepancy among the finite sample critical values of the SADF statistic is negligible.

from the 95% quantile of $F_3(W, r_0)$ and $F_4(W, r_0)$ respectively (Table 2). The power calculations for $\eta = 0.4$ and $\eta = 0.2$ are based on the 95% quantiles of the finite sample distributions (Table 4). The number of replications is 2,000.

Table 4: The finite sample critical values of the SADF statistic (against explosive alternative)

	$\eta = 0.4$			$\eta = 0.2$		
	90%	95%	99%	90%	95%	99%
$T = 100$ and $r_0 = 0.4$	1.26	1.57	2.32	1.84	2.22	3.03
$T = 200$ and $r_0 = 0.2$	1.44	1.72	2.35	2.11	2.42	3.04
$T = 400$ and $r_0 = 0.1$	1.57	1.88	2.56	2.26	2.62	3.31

Note: The finite sample critical values are obtained by simulation with $2,000$ replications.

Table 5: Powers of the SADF test under Evans (1991) periodically collapsing explosive behavior

	Yearly parameter settings			
	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
$T = 100$ and $r_0 = 0.4$	0.44	0.37	0.28	0.24
$T = 200$ and $r_0 = 0.2$	0.62	0.58	0.49	0.45
	Monthly parameter settings			
$T = 100$ and $r_0 = 0.4$	0.59	0.51	0.34	0.26
$T = 200$ and $r_0 = 0.2$	0.75	0.69	0.55	0.48
$T = 400$ and $r_0 = 0.1$	0.86	0.81	0.71	0.68

Note: power calculations are based on 2,000 replications.

From Table 5 power of the test evidently increases with sample size. Under the yearly parameter setting and $T = 200$, power for $\eta > 0.5$, $\eta = 0.4$, $\eta = 0.2$ and $\eta = 0$ is 18%, 21%, 21% and 21% higher than when $T = 100$.

Furthermore, power for $\eta > 0.5$ is always higher than when $\eta < 0.5$. In addition, when $\eta < 0.5$, power decreases as $\eta \to 0$. From the lower panel of Table 5 (monthly parameters settings), when $T = 400$, for instance, the power of the test is 86% when $\eta > 0.5$ and then declines from 81% to 68% as η changes from 0.4 to zero.

6.2 Locally explosive behavior

The second alternative DGP is the locally explosive model (19). The parameter settings are the same as in Section 3.2. As mentioned, this DGP is more realistic than the PY and Evans models in the sense that explosive behavior does not collapse completely within one period. Instead, the collapse process is taken to be a (mildly) stationary process. The parameter β controls the contraction rate of the bubble, the duration of which is $T_c - T_f$. To explore the sensitivity of the SADF test to these two coefficients, we calculate powers of the test by setting β equal to $0.4, 0.5$ and 0.6 and $T_c - T_f$ equal to $[0.05T], [0.10T]$ and $[0.15T]$. In general, we find that the power of the SADF test is invariant to the contraction rate and the contraction duration of the bubble. For brevity, these results are not tabulated here.

Table 6: Powers of the SADF test for the locally explosive behavior (the rates of bubble expansion and contraction). Parameters are set as: $y_0 = 94.122, u_0 = 0.131, c_1 = c_2 = 1, \sigma_1 = \sigma_2 = 3.829, \sigma_3 = 2\sigma_1, \beta = 0.5, T_e = [0.6T], T_f = [0.7T], T_c = [0.75T], T = 200, r_0 = 0.2$.

	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
$\alpha = 0.60, \phi_T = 1.04$	0.57	0.48	0.29	0.21
$\alpha = 0.55, \phi_T = 1.05$	0.65	0.58	0.45	0.41
$\alpha = 0.50, \phi_T = 1.07$	0.77	0.75	0.70	0.69

Note: power calculations are based on 2,000 replications.

Table 7: Powers of the SADF test for the locally explosive behavior (the duration of bubble expansion and contraction). Parameters are set as: $y_0 = 94.122, u_0 = 0.131, c_1 = c_2 = 1, \sigma_1 = \sigma_2 = 3.829, \sigma_3 = 2\sigma_1, \alpha = 0.6, \beta = 0.5, T = 200, r_0 = 0.2, T_e = [0.6T], T_c - T_f = [0.05T]$.

	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
$T_f - T_e = [0.10T]$	0.57	0.48	0.29	0.21
$T_f - T_e = [0.15T]$	0.75	0.69	0.57	0.52
$T_f - T_e = [0.20T]$	0.89	0.86	0.79	0.78

Note: power calculations are based on 2,000 replications.

The explosive rate of the bubble is determined by the parameter α and the duration of the bubble expansion $T_f - T_e$. In simulations, we allow α to be $0.60, 0.55$ and 0.50 (Table 6) and $T_f - T_e$ to be $[0.10T], [0.15T]$ and $[0.20T]$ (see Table 7). From Table 6, we can see that, *ceteris paribus*, the power of the SADF test increases as α decreases. That is, the frequency of successfully detecting the existence of exuberant behavior is higher when the expansion rate is faster. For example, under the specification of $\eta > 0.5$, when $T = 200$ and α takes the values $0.6, 0.55$ and 0.5, the power is 57%, 65% and 77% respectively. Moreover, from Table 7 it is clear that, *ceteris paribus*, the power of the SADF test is higher when the duration of the bubble expansion is longer. For instance, when $T = 200$ the power for $\eta > 0.5$ with $T_f - T_e = [0.10T], [0.15T], [0.20T]$ is 57%, 75% and 89% respectively.

The location of the bubble episode is indicated by T_e. Table 8 illustrates the power of the SADF test with $T_e = [0.2T], [0.4T], [0.6T]$. We observe that given an identical expansion rate

Table 8: Powers of the SADF test for the locally explosive behavior (the location of the bubble episode). Parameters are set as: $y_0 = 41.195, u_0 = 0.740, c_1 = c_2 = 1, \sigma_1 = \sigma_2 = 7.869, \sigma_3 = 2\sigma_1, \alpha = 0.6, \beta = 0.5, T = 200, r_0 = 0.2, T_c - T_f = [0.05T]$.

			$T_e = [0.2T]$	
	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
$T_f - T_e = [0.10T]$	0.57	0.48	0.25	0.16
$T_f - T_e = [0.15T]$	0.72	0.63	0.40	0.30
$T_f - T_e = [0.20T]$	0.87	0.83	0.72	0.69
			$T_e = [0.4T]$	
$T_f - T_e = [0.10T]$	0.57	0.47	0.27	0.19
$T_f - T_e = [0.15T]$	0.74	0.65	0.48	0.42
$T_f - T_e = [0.20T]$	0.87	0.84	0.77	0.75
			$T_e = [0.6T]$	
$T_f - T_e = [0.10T]$	0.57	0.48	0.29	0.21
$T_f - T_e = [0.15T]$	0.75	0.68	0.57	0.52
$T_f - T_e = [0.20T]$	0.90	0.87	0.82	0.80

Note: power calculations are based on 2,000 replications.

and expansion duration of the bubble, if the bubble episode occurs at a later stage of the sample period, the frequency of successfully detecting a bubble episode is higher. For instance, when $T_f - T_e = [0.15T]$, the power of $\eta = 0.4$ is 63%, 65% and 68% for $T_e = [0.2T], [0.4T], [0.6T]$, respectively.

Table 9 illustrates the power of the SADF with different sample sizes. First, as expected, power rises with the sample size. Powers for $\eta > 0.5$ are 51%, 57% and 73% for $T = 100, 200, 400$. Second, the specification $\eta > 0.5$ always gives higher power than $\eta < 0.5$. The last observation applies to Table 6 - 9.

Table 9: Powers of the SADF test for the locally explosive behavior (the sample size). Parameters are set as: $y_0 = 41.195, u_0 = 0.740, c_1 = c_2 = 1, \sigma_1 = \sigma_2 = 7.869, \sigma_3 = 2\sigma_1, \alpha = 0.6, \beta = 0.5, T_e = [0.6T], T_f - T_e = [0.10T], T_c - T_f = [0.05T]$.

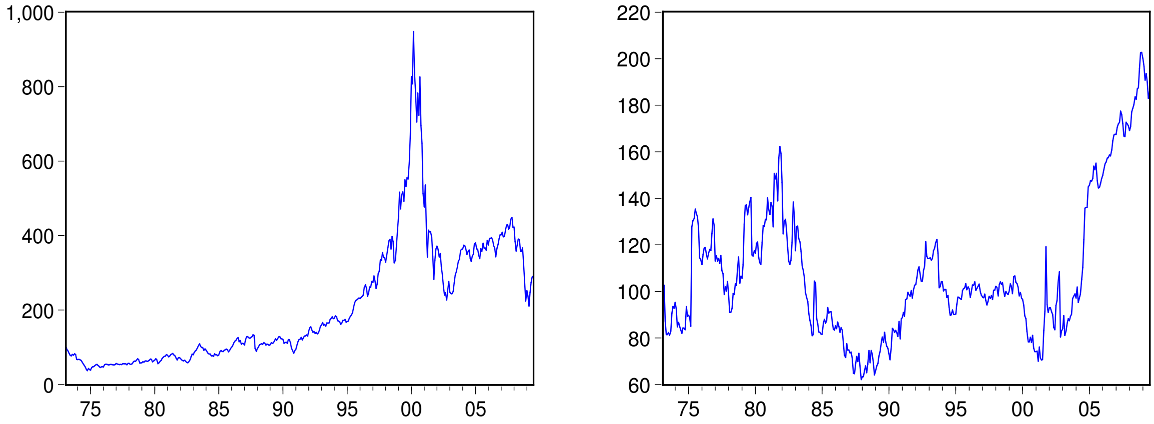
	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
$T = 100$ and $r_0 = 0.4$	0.51	0.40	0.22	0.13
$T = 200$ and $r_0 = 0.2$	0.57	0.48	0.29	0.21
$T = 400$ and $r_0 = 0.1$	0.73	0.66	0.49	0.44

Note: power calculations are based on 2,000 replications.

7 Empirical Applications

7.1 The Nasdaq

The first empirical application applies the sup ADF test to Nasdaq market data over the period from February 1973 to July 2009 (constituting 438 observations). The Nasdaq composite index and the Nasdaq dividend yield are obtained from DataStream International. The consumer price index, which is used to convert stock prices and dividends into real series, is downloaded from the Federal Reserve Bank of St. Louis.



(a) The real NASDAQ index

(b) The real NASDAQ index dividend

Figure 6: NASDAQ stock market sampled from February 1973 to September 2009 (normalized to 100 at the beginning of data series).

Figure 6 shows the time paths of the real Nasdaq index and the real Nasdaq dividend (normalized to 100 at the beginning of the data series) over the sample period. The real Nasdaq index grows steadily, manifesting an upward drift, until the early 90s. This is followed by a rapid increase to a peak that is 944.4 times larger than the starting point of the series. The Nasdaq index, then dropped quickly to a level of less than 248 times the starting point at April 2003. It recovers gradually until October 2008, however, followed by another sudden crash. Relative to the Nasdaq index, the dividend process changes are of a much smaller magnitude, although it is volatile throughout the sample, and shows some sustained growth since 2004.

Table 10 displays the SADF statistics for the logarithmic real Nasdaq index and the logarithmic real Nasdaq dividend, along with respective critical values under the specifications $\eta > 0.5$, $\eta = 0.4$,

$\eta = 0.2$ and $\eta = 0$. The lag order is determined by BIC with the maximum lag length 12. The smallest fractional window r_0 is set to be 0.1. Like the simulation experiments, we use asymptotic critical values for the specifications $\eta > 0.5$ and $\eta = 0$ and finite sample critical values for the specifications $\eta = 0.4$ and $\eta = 0.2$. The finite sample critical values are obtained from simulations with $2,000$ replications and sample size 438.

Table 10: The sup ADF test of the NASDAQ stock market

		SADF statistic		
Log Real NASDAQ Index		2.56		
Log Real NASDAQ Dividend		-1.07		
	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
90%	1.23	1.60	2.32	2.58
95%	1.54	1.90	2.59	2.92
99%	2.04	2.47	3.14	3.42

Note: Critical values of the sup ADF test under the specification of $\eta = 0.4$ and $\eta = 0.2$ are obtained by simulations with $2,000$ replications and sample size 438. The smallest fractional window r_0 is set to be 0.1.

For the logarithmic real Nasdaq index, we reject the unit root null hypothesis against the explosive alternative at the 10% significance level under specifications with $\eta > 0.5$, $\eta = 0.4$ and $\eta = 0.2$, whereas we fail to reject the null hypothesis at the 10% significance level under the specification of $\eta = 0$ (although the difference between the test statistic and the critical value is very small). Furthermore, we cannot reject the null hypothesis of unit root at the 10% significance level for the logarithmic real Nasdaq dividend under all specifications considered.

In other words, with specifications of $\eta > 0.5$, $\eta = 0.4$ and $\eta = 0.2$, we find evidence of explosive behavior in Nasdaq using the sup ADF test. However, if the null hypothesis is specified as

$$H_{03} : y_t = 1 + y_{t-1} + \varepsilon_t, \varepsilon_t \overset{iid}{\sim} N\left(0, \sigma^2\right) \tag{25}$$

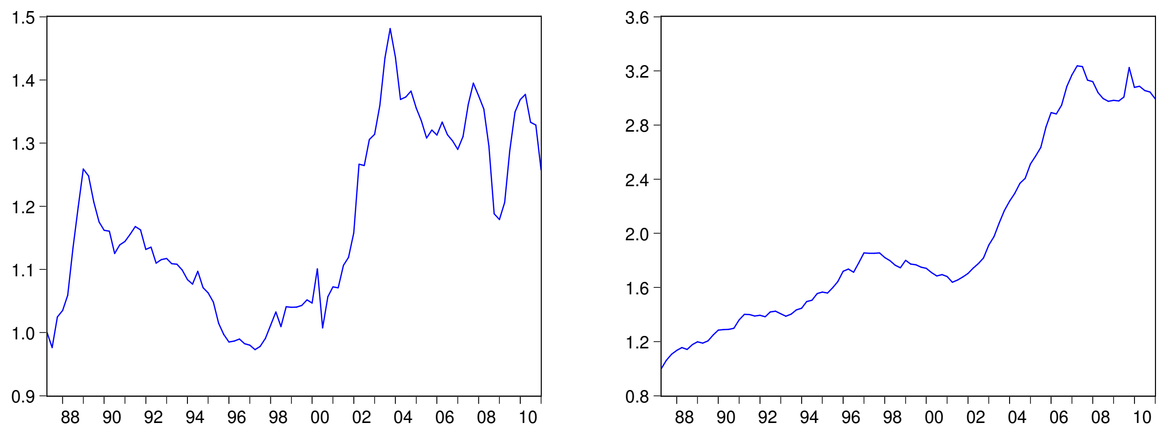
(i.e. the specification corresponding to $\eta = 0$), the sup ADF finds no evidence of bubble existence in the Nasdaq stock market during the sample period. This null hypothesis implies that the long-term average return of the Nasdaq stock index is 100%, which is obviously unrealistic and can be excluded on prior grounds.

Hence, the SADF test provides strong evidence for the presence of explosive behavior in the Nasdaq. The evidence is robust to specification of the null model with the exception of extreme models such as (25).

7.2 The Australia and New Zealand Housing Markets

The second appplication is to real estate markets where with data sampled over 1987Q2 to 2011Q1. The price index of established houses in Australia is taken at a quarterly frequency from the Australian Bureau of Statistics (ABS).[8] The New Zealand house price index (Quotable Value) is taken from the Reserve Bank of New Zealand (RBNZ). We use Australian household disposable income as a proxy for fundamentals in the Australia housing market and the series is obtained from the Reserve Bank of Australia (RBA). For the New Zealand housing market, we use national gross disposable income (per capita) taken from DataStream International.[9]

The Australian and New Zealand price-to-income ratios (normalized to unity at the beginning of the data series) are shown in Figure 7. From Figure 7a, it is clear that the Australian house price-to-income ratio fluctuates throughout the sample range. A considerable increase occurred over the period from 2001 to 2003. The peak of this increase was 1.48 times bigger than the starting point of the series. The house price-to-income ratio for New Zealand (Figure 7b) grows steadily until early 1997, followed by a mild downturn. The ratio then grows rapidly so that by the third quarter of 2007 it was 3.2 times bigger than the starting point. The magnitude of this expansion is much larger than that of the Australian housing market.



(a) Australia (b) New Zealand

Figure 7: The house price-to-income ratio sampled from 1987Q2 to 2011Q1 (normalized to unity at the beginning of data series).

[8]It is a weighted average of 8 capital cities (Sydney, Melbourne, Brisbane, Adelaide, Perth, Hobart, Darwin and Canberra).

[9]The quarterly household disposable income data series of New Zealand is discontinued in 2008Q4.

Table 11: The sup ADF test of the Australia and New Zealand housing markets

	SADF statistic			
Australia price-to-income ratio	1.23			
New Zealand price-to-income ratio	4.62			
	$\eta > 0.5$	$\eta = 0.4$	$\eta = 0.2$	$\eta = 0$
90%	0.88	1.26	1.84	2.27
95%	1.20	1.59	2.19	2.62
99%	1.87	2.33	3.04	3.20

Note: Critical values of the sup ADF test under the specification of $\eta = 0.4$ and $\eta = 0.2$ are obtained from simulations with $2,000$ replications and sample size 96. The smallest fractional window r_0 is set to be 0.4.

We apply the SADF test to the price-to-income ratios for these two markets. Table 11 presents the SADF statistics, along with respective critical values under different specifications of the null hypothesis. The smallest fractional window size r_0 equals 0.4. The finite sample critical values for the cases of $\eta = 0.4$ and $\eta = 0.2$ are obtained from simulations with $2,000$ replications and sample size 96.

From Table 11, the SADF statistic for the Australian price-to-income ratio is 1.23. We reject the null hypothesis when η is specified to be greater than 0.5 and fail to reject the null when it is smaller than 0.5. These results reveal that the empirical evidence of exuberance in the Australia housing market is sensitive to model specification.

For the New Zealand housing market, the SADF statistic is 4.62, which is greater than the 1% critical values under all four different null specifications. Hence, application of the SADF test confirms the existence of exuberance in the New Zealand housing market over the sample period and the confirmation is universal across specifications.

8 Conclusion

This paper has investigated various formulations of the null and alternative hypotheses in studying empirical evidence of exuberance in economic and financial time series. The formulations involve different specifications of the regression models used for the construction of empirical tests of exuberance, which are shown to impact both the finite sample and the asymptotic distributions of the tests.

Our findings suggest a model specification that should be reasonable in practical work. The empirical model does not include a linear deterministic trend in the regression but allows for some (small) deterministic drift in the process under the null hypothesis of a unit root. The test relies on the estimation (or recursive estimation) of the autoregressive coefficient in the model

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t.$$

The null hypothesis allows for an intercept α which is local to zero. In particular, α is a function of sample size with a power exponent parameter η, so that $\alpha = \widetilde{\alpha} T^{-\eta}$. The limiting distributions of the ADF statistic and the SADF statistic are derived for cases where $\eta > 0.5$, $\eta = 0.5$ and $\eta < 0.5$. The asymptotic critical values are obtained by simulation.

The size and power properties have been examined and compared. When asymptotic critical values are used, the SADF test does not show obvious size distortion under the cases where $\eta > 0.5$ and $\eta = 0$. There is significant size distortion when $0 < \eta < 0.5$ and the level of distortion increases with the value of η.

Power is assessed using the Evans (1991) periodically collapsing explosive process (with both yearly and monthly parameter settings) and the locally explosive process proposed in this paper (with monthly parameters setting). The conclusion drawn from these two DGPs is the same. The results indicate that the preferred procedure for practical implementation is to estimate the regression model of equation (2) and specify the null hypothesis to be an asymptotically negligible intercept (i.e. $\eta > 0.5$) in the right-tailed unit root test. The empirical application of these methods to Nasdaq market data and to the Australian and New Zealand housing markets demonstrates the importance of hypothesis and model specification in the right-tailed unit root test, revealing some sensitivity in the outcomes of the test to these modeling decisions.

REFERENCES

Anderson, T.W., 1959, On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics*, 30:676–687.

Davidson, R. and Mackinnon, J.G., 2004, *Econometric Theory and Methods*. Oxford University Press.

Diba, B.T., and Grossman, H.I., 1988, Explosive rational bubbles in stock prices? *The American Economic Review*, 78(3):520–530.

Dickey, D.A., William, R. B., and Miller, R.B., 1986, Unit roots in time series models: Tests and implications. *The American Statistician*, 40:12–26.

Evans, G.W., 1991, Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, 81(4):922–930.

Fuller, W.A., 1995, *Introduction to Statistical Time Series.* John Wiley & Sons, Inc..

Hicks, J., 1950, *A Contribution to the Theory of Trade Cycle.* Oxford University Press.

Homm, U. and J. Breitung, 2011, Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics,* forthcoming.

Ng, S. and Perron, P., 1995, Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90:268–281.

Phillips, P.C.B., 1987, Time series regression with a unit root. *Econometrica* 55, 277-301.

Phillips, P.C.B., and Lee, C.C., 1996, Efficiency gains from quasi-differencing under nonstationary. mimeographed, Yale University.

Phillips, P.C.B., and Magdalinos, T., 2007, Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136:115–130.

Phillips, P.C.B., and Perron, P., 1988, Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Phillips, P.C.B., Shi, S., and Yu, J., 2011, Testing for multiple bubbles. *Singapore Management University, Working Papers No. 09-2011.*

Phillips, P.C.B., and Solo, V., 1992, Asymptotics for linear processes. *The Annals of Statistics*, 20:971–1001.

Phillips, P.C.B., Wu, Y., and Yu, J., 2011, Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52:201–226.

Phillips, P.C.B., and Yu, J., 2009, Limit theory for dating the origination and collapse of mildly explosive periods in time series data. *Singapore Management University, Unpublished Manuscript.*

Phillips, P.C.B., and Yu, J., 2011, Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics*, 2:455–491.

Schmidt, P., and Phillips, P.C.B., 1992, LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, 54:257–287.

White, J.S., 1958, The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29:1188–1197.

A Appendix

Before proving Proposition 5.1, we list in the following Lemma some standard results that are useful in the proof.

Lemma A.1 *Let* $u_t = \psi(L)\varepsilon_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, *where* $\Sigma_{j=0}^{\infty} j|\psi_j| < \infty$ *and* $\{\varepsilon_t\}$ *is an i.i.d sequence with mean zero, variance* σ^2 *and finite fourth moment. Define* $M_T(r) = 1/T\sum_{s=1}^{[Tr]} u_s$ *with* $r \in [r_0, 1]$ *and* $\xi_t = \sum_{s=1}^{t} u_s$. *We have:*

(1) $\sum_{s=1}^{t} u_s = \psi(1)\sum_{s=1}^{t}\varepsilon_s + \eta_t - \eta_0$, *where* $\eta_t = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$, $\eta_0 = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{-j}$ *and* $\alpha_j = -\sum_{i=1}^{\infty}\psi_{j+i}$, *which is absolutely summable.*

(2) $\frac{1}{T}\sum_{t=1}^{[Tr]}\varepsilon_t^2 \xrightarrow{p} \sigma^2 r$.

(3) $T^{-1/2}\sum_{t=1}^{[Tr]}\varepsilon_t \xrightarrow{L} \sigma W(r)$.

(4) $T^{-1}\sum_{t=1}^{[Tr]}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t \xrightarrow{L} \frac{1}{2}\sigma^2\left[W(r)^2 - r\right]$.

(5) $T^{-3/2}\sum_{t=1}^{[Tr]}\varepsilon_t t \xrightarrow{L} \sigma\left[rW(r) - \int_0^r W(s)\,ds\right]$.

(6) $T^{-1}\sum_{t=1}^{[Tr]}(\eta_{t-1} - \eta_0)\varepsilon_t \xrightarrow{p} 0$.

(7) $T^{-1/2}\left(\eta_{[Tr]} - \eta_0\right) \xrightarrow{p} 0$.

(8) $\sqrt{T}M_T(r) \xrightarrow{L} \psi(1)\sigma W(r)$.

(9) $T^{-1}\sum_{t=1}^{[Tr]}\xi_{t-1}\varepsilon_t \xrightarrow{L} \frac{1}{2}\psi(1)\sigma^2\left[W(r)^2 - r\right]$.

(10) $T^{-3/2}\sum_{t=1}^{[Tr]}\xi_{t-1} \xrightarrow{L} \psi(1)\sigma\int_0^r W(s)\,ds$.

(11) $T^{-5/2}\sum_{t=1}^{[Tr]}\xi_{t-1}t \xrightarrow{p} \psi(1)\sigma\int_0^r W(s)\,sds$.

(12) $T^{-2}\sum_{t=1}^{[Tr]}\xi_{t-1}^2 \xrightarrow{L} \sigma^2\psi(1)^2\int_0^r W(s)^2\,ds$.

(13) $T^{-3/2}\sum_{t=1}^{[Tr]}\xi_{t-1}u_{t-j} \xrightarrow{p} 0, \forall j \geq 0$.

All of these results can be found or easily derived from Phillips (1987), Phillips and Perron (1988), and Phillips and Solo (1991).

Lemma A.2 *Define* $y_t = \alpha_T t + \sum_{s=1}^{t} u_s$, $\alpha_T = \psi(1) T^{-\eta}$ *and* $u_t = \psi(L)\varepsilon_t = \Sigma_{j=0}^{\infty}\psi_j \varepsilon_{t-j}$, *where* $\Sigma_{j=0}^{\infty} j |\psi_j| < \infty$ *and* $\{\varepsilon_t\}$ *is an i.i.d sequence with mean zero, variance* σ^2 *and finite fourth moment. Then, if* $\eta > 1/2$, *we have*

$$(a1)\ \ T^{-1}\sum_{t=1}^{[Tr]} y_{t-1}\varepsilon_t \xrightarrow{L} \frac{1}{2}\sigma^2\psi(1)\left[W(r)^2 - r\right],$$

$$(b1)\ \ T^{-3/2}\sum_{t=1}^{[Tr]} y_{t-1} \xrightarrow{L} \psi(1)\sigma\int_0^r W(s)\,ds,$$

$$(c1)\ \ T^{-2}\sum_{t=1}^{[Tr]} y_{t-1}^2 \xrightarrow{L} \sigma^2\psi(1)^2\int_0^r W(s)^2\,ds,$$

$$(d1)\ \ T^{-3/2}\sum_{t=1}^{[Tr]} y_{t-1}u_{t-j} \xrightarrow{p} 0,,j = 0,1,\cdots;$$

if $\eta = 1/2$, *we have*

$$(a2)\ \ T^{-1}\sum_{t=1}^{[Tr]} y_{t-1}\varepsilon_t \xrightarrow{L} \psi(1)\sigma\left\{\left[rW(r) - \int_0^r W(s)\,ds\right] + \frac{1}{2}\sigma\left[W(r)^2 - r\right]\right\},$$

$$(b2)\ \ T^{-3/2}\sum_{t=1}^{[Tr]} y_{t-1} \xrightarrow{L} \psi(1)\left[\frac{1}{2}r + \sigma\int_0^r W(s)\,ds\right],$$

$$(c2)\ \ T^{-2}\sum_{t=1}^{[Tr]} y_{t-1}^2 \xrightarrow{L} \psi(1)^2\left[\frac{1}{3}r^3 + \sigma^2\int_0^r W(s)^2\,ds + 2\sigma\int_0^r W(s)\,sds\right],$$

$$(d2)\ \ T^{-3/2}\sum_{t=1}^{[Tr]} y_{t-1}u_{t-j} \xrightarrow{p} 0,,j = 0,1,\cdots;$$

and if $\eta < 1/2$, *we have*

$$(a3)\ \ T^{-3/2}\sum_{t=1}^{[Tr]} y_{t-1}\varepsilon_t \xrightarrow{L} \alpha_T\sigma\left[rW(r) - \int_0^r W(s)\,ds\right].$$

$$(b3)\ \ T^{-2}\sum_{t=1}^{[Tr]} y_{t-1} \xrightarrow{p} \frac{\alpha_T}{2}r^2.$$

$$(c3)\ \ T^{-3}\sum_{t=1}^{[Tr]} y_{t-1}^2 \xrightarrow{p} \frac{\alpha_T^2}{3}r^3.$$

$$(d3)\ \ T^{-2}\sum_{t=1}^{[Tr]} y_{t-1}u_{t-j} \xrightarrow{p} 0, \forall j \geq 0.$$

The proof of Lemma A.2 is straightforward and is therefore omitted here for brevity. (It may be obtained from the authors upon request.)

We now derive the asymptotic distributions of SADF.

Proof of Proposition 5.1. The regression model is

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{k=1}^{p-1} \phi_k \Delta y_{t-k} + \varepsilon_t.$$

Under the null hypothesis that $\alpha = T^{-\eta}$ and $\beta = 0$, we have $y_t = \alpha_T t + \sum_{s=1}^{t} u_s$, where $\alpha_T = \psi(1)\alpha$ and $u_t = \psi(L)\varepsilon_t$ with $\psi(L) = \left(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_{p-1} L^{p-1}\right)^{-1}$.

The deviation of the OLS estimate $\hat{\theta}_r$ from the true value θ is given by

$$\hat{\theta}_r - \theta = \left[\sum_{t=1}^{[Tr]} X_t X_t'\right]^{-1} \left[\sum_{t=1}^{[Tr]} X_t \varepsilon_t\right], \tag{26}$$

where $X_t = [\alpha_T + u_{t-1}\ \ \alpha_T + u_{t-2}\ \ldots\ \alpha_T + u_{t-p+1}\ \ 1\ \ y_{t-1}]'$ and $\theta = [\phi_1\ \ \phi_2\ \ldots \phi_{p-1}\ \ \alpha\ \ \beta]'$. We know that the probability limit of $\sum_{t=1}^{[Tr]} X_t X_t'$ is block diagonal from $(d1)$, $(d2)$ and $(d3)$ of Lemma A.2. Therefore, we only need to obtain the last 2×2 components of $\sum_{t=1}^{[Tr]} X_t X_t'$ and the last 2×1 component of $\sum_{t=1}^{[Tr]} X_t \varepsilon_t$, which are

$$\begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \text{ and } \begin{bmatrix} \Sigma \varepsilon_t \\ \Sigma y_{t-1}\varepsilon_t \end{bmatrix},$$

respectively, where Σ denotes summation over $t = 1, 2, \cdots, [Tr]$. Pre-multiplying equation (26) by a scaling matrix Υ_T, results in

$$\Upsilon_T \begin{bmatrix} \widehat{\alpha}_r - \alpha \\ \widehat{\beta}_r - \beta \end{bmatrix} = \left\{ \Upsilon_T^{-1} \left[\sum_{t=1}^{[Tr]} X_t X_t'\right]_{(-2)\times(-2)} \Upsilon_T^{-1} \right\}^{-1} \left\{ \Upsilon_T^{-1} \left[\sum_{t=1}^{[Tr]} X_t \varepsilon_t\right]_{(-2)\times 1} \right\}.$$

If $\eta > 1/2$, based on (3) of Lemma A.1 and $(a1)$ of Lemma A.2, the scaling matrix should be $\Upsilon_T = diag\left(T^{1/2}, T\right)$. Consider the matrix $\Upsilon_T^{-1}\left[\sum_{t=1}^{[Tr]} X_t X_t'\right]_{(-2)\times(-2)} \Upsilon_T^{-1}$,

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix} \xrightarrow{L} \begin{bmatrix} r & \psi(1)\sigma \int_0^r W(s)\,ds \\ \psi(1)\sigma \int_0^r W(s)\,ds & \sigma^2 \psi(1)^2 \int_0^r W(s)^2\,ds \end{bmatrix},$$

and the matrix $\Upsilon_T^{-1}\left[\sum_{t=1}^{[Tr]} X_t \varepsilon_t\right]_{(-2)\times 1}$,

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \varepsilon_t \\ \Sigma y_{t-1}\varepsilon_t \end{bmatrix} = \begin{bmatrix} T^{-1/2}\Sigma \varepsilon_t \\ T^{-1}\Sigma y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma W(r) \\ \frac{1}{2}\sigma^2 \psi(1)\left[W(r)^2 - r\right] \end{bmatrix}.$$

Under the null hypothesis that $\alpha = T^{-\eta}$ with $\eta > 1/2$ and $\beta = 0$,

$$\left[\begin{array}{c} T^{1/2}\left(\hat{\alpha}_r-\alpha\right) \\ T\hat{\beta}_r \end{array}\right] \overset{L}{\to} \left[\begin{array}{cc} r & A_{0,r} \\ A_{0,r} & B_{0,r} \end{array}\right]^{-1} \left[\begin{array}{c} C_{0,r} \\ D_{0,r} \end{array}\right] = \frac{1}{-B_{0,r}r+A_{0,r}^2} \left[\begin{array}{c} -B_{0,r}C_{0,r}+A_{0,r}D_{0,r} \\ A_{0,r}C_{0,r}-rD_{0,r} \end{array}\right]$$

with

$$\begin{aligned} A_{0,r} &= \psi\left(1\right)\sigma\int_0^r W\left(s\right)ds, B_{0,r} = \sigma^2\psi\left(1\right)^2\int_0^r W\left(s\right)^2 ds, \\ C_{0,r} &= \sigma W\left(r\right), D_{0,r} = \frac{1}{2}\sigma^2\psi\left(1\right)\left[W\left(r\right)^2-r\right]. \end{aligned}$$

Therefore, we have

$$T\hat{\beta}_r \overset{L}{\to} \frac{A_{0,r}C_{0,r}-rD_{0,r}}{-B_{0,r}r+A_{0,r}^2}.$$

To calculate the t-statistic of $\hat{\beta}_r$, we need to find the standard error of $\hat{\beta}_r$. We know that

$$var\left(\left[\begin{array}{c} \hat{\alpha}_r \\ \hat{\beta}_r \end{array}\right]\right) = \sigma^2\left[\begin{array}{cc} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{array}\right]^{-1}$$

so the variances of $T\hat{\beta}_r$ can be calculated as follows:

$$\begin{aligned} var\left(\left[\begin{array}{c} T^{1/2}\left(\hat{\alpha}_r-\alpha\right) \\ T\hat{\beta}_r \end{array}\right]\right) &= \sigma^2\left\{\left[\begin{array}{cc} T^{1/2} & 0 \\ 0 & T \end{array}\right]^{-1}\left[\begin{array}{cc} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{array}\right]\left[\begin{array}{cc} T^{1/2} & 0 \\ 0 & T \end{array}\right]^{-1}\right\}^{-1} \\ &\overset{L}{\to} \frac{\sigma^2}{-B_{0,r}r+A_{0,r}^2}\left[\begin{array}{cc} -B_{0,r} & A_{0,r} \\ A_{0,r} & -r \end{array}\right]. \end{aligned}$$

Hence, the t-statistic of $\hat{\beta}_r$ is

$$ADF_r = \frac{T\hat{\beta}_r}{se\left(T\hat{\beta}_r\right)} \overset{L}{\to} \frac{\frac{1}{2}r\left[W\left(r\right)^2-r\right]-\int_0^r W\left(s\right)dsW\left(r\right)}{r^{1/2}\left\{\int_0^r W\left(s\right)^2 dsr-\left[\int_0^r W\left(s\right)ds\right]^2\right\}^{1/2}}.$$

By the CMT, we have

$$SADF\left(r_0\right) \overset{L}{\to} \sup_{r\in[r_0,1]}\left\{\frac{\frac{1}{2}r\left[W\left(r\right)^2-r\right]-\int_0^r W\left(s\right)dsW\left(r\right)}{r^{1/2}\left\{\int_0^r W\left(s\right)^2 dsr-\left[\int_0^r W\left(s\right)ds\right]^2\right\}^{1/2}}\right\}.$$

If $\eta = 1/2$, based on (3) of Lemma A.1 and (a2) of Lemma A.2, the scaling matrix should be $\Upsilon_T = diag\left(\sqrt{T}, T\right)$. Consider the matrix $\Upsilon_T^{-1}\left[\sum_{t=1}^{[Tr]} X_t X_t'\right]_{(-2)\times(-2)}\Upsilon_T^{-1}$,

$$\begin{aligned} &\left[\begin{array}{cc} \sqrt{T} & 0 \\ 0 & T \end{array}\right]^{-1}\left[\begin{array}{cc} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{array}\right]\left[\begin{array}{cc} \sqrt{T} & 0 \\ 0 & T \end{array}\right]^{-1} \\ &\overset{L}{\to}\left[\begin{array}{cc} 1 & 0 \\ 0 & \psi\left(1\right) \end{array}\right]\left[\begin{array}{cc} r & \frac{1}{2}r+\sigma\int_0^r W\left(s\right)ds \\ \frac{1}{2}r+\sigma\int_0^r W\left(s\right)ds & \frac{1}{3}r^3+\sigma^2\int_0^r W\left(s\right)^2 ds+2\sigma\int_0^r W\left(s\right)sds \end{array}\right]\left[\begin{array}{cc} 1 & 0 \\ 0 & \psi\left(1\right) \end{array}\right], \end{aligned}$$

and the matrix $\Upsilon_T^{-1}\left[\sum_{t=1}^{[Tr]} X_t\varepsilon_t\right]_{(-2)\times 1}$,

$$\left[\begin{array}{cc}\sqrt{T} & 0\\ 0 & T\end{array}\right]^{-1}\left[\begin{array}{c}\Sigma\varepsilon_t\\ \Sigma y_{t-1}\varepsilon_t\end{array}\right]\xrightarrow{L}\left[\begin{array}{cc}\sigma & 0\\ 0 & \sigma\psi_r(1)\end{array}\right]\left[\begin{array}{c}W(r)\\ \left[rW(r)-\int_0^r W(s)\,ds\right]+\frac{1}{2}\sigma\left[W(r)^2-r\right]\end{array}\right].$$

Under the null hypothesis that $\alpha=T^{-\eta}$ and $\beta=0$,

$$\left[\begin{array}{c}\sqrt{T}(\hat{\alpha}_r-\alpha)\\ T\hat{\beta}_r\end{array}\right]\xrightarrow{L}\left[\begin{array}{cc}\sigma & 0\\ 0 & \sigma\psi(1)^{-1}\end{array}\right]\left[\begin{array}{cc}r & A_{r,\sigma}\\ A_{r,\sigma} & B_{r,\sigma}\end{array}\right]^{-1}\left[\begin{array}{c}C_{r,\sigma}\\ D_{r,\sigma}\end{array}\right]$$

where

$$\begin{aligned}A_{r,\sigma}&=\frac{1}{2}r+\sigma\int_0^r W(s)\,ds,\\ B_{r,\sigma}&=\frac{1}{3}r^3+\sigma^2\int_0^r W(s)^2\,ds+2\sigma\int_0^r W(s)\,sds\\ C_{r,\sigma}&=W(r),D_{r,\sigma}=\left[rW(r)-\int_0^r W(s)\,ds\right]+\frac{1}{2}\sigma\left[W(r)^2-r\right].\end{aligned}$$

We can see that $\hat{\beta}_r$ converges at rate T to the following distribution

$$T\hat{\beta}_r\xrightarrow{L}\left[\begin{array}{cc}0 & \sigma\psi(1)^{-1}\end{array}\right]\left[\begin{array}{cc}r & A_{r,\sigma}\\ A_{r,\sigma} & B_{r,\sigma}\end{array}\right]^{-1}\left[\begin{array}{c}C_{r,\sigma}\\ D_{r,\sigma}\end{array}\right]=\frac{\sigma}{\psi(1)}\frac{rD_{r,\sigma}-A_{r,\sigma}C_{r,\sigma}}{B_{r,\sigma}r-A_{r,\sigma}^2}.$$

We know that

$$\begin{aligned}var\left(\left[\begin{array}{c}\sqrt{T}(\hat{\alpha}_r-\alpha)\\ T\hat{\beta}_r\end{array}\right]\right)&=\sigma^2\left\{\left[\begin{array}{cc}\sqrt{T} & 0\\ 0 & T\end{array}\right]^{-1}\left[\begin{array}{cc}\Sigma 1 & \Sigma y_{t-1}\\ \Sigma y_{t-1} & \Sigma y_{t-1}^2\end{array}\right]\left[\begin{array}{cc}\sqrt{T} & 0\\ 0 & T\end{array}\right]^{-1}\right\}^{-1}\\ &\xrightarrow{L}\sigma^2\left[\begin{array}{cc}1 & 0\\ 0 & \psi(1)\end{array}\right]^{-1}\left[\begin{array}{cc}r & A_{r,\sigma}\\ A_{r,\sigma} & B_{r,\sigma}\end{array}\right]^{-1}\left[\begin{array}{cc}1 & 0\\ 0 & \psi(1)\end{array}\right]^{-1}.\end{aligned}$$

Hence, the t-statistic of $\hat{\beta}_r$ is:

$$ADF_r=\frac{T\hat{\beta}_r}{se\left(T\hat{\beta}_r\right)}\xrightarrow{L}\frac{rD_{r,\sigma}-A_{r,\sigma}C_{r,\sigma}}{r^{1/2}\left(B_{r,\sigma}r-A_{r,\sigma}^2\right)^{1/2}}.$$

By CMT, we have

$$SADF(r_0)\xrightarrow{L}\sup_{r\in[r_0,1]}\left\{\frac{rD_{r,\sigma}-A_{r,\sigma}C_{r,\sigma}}{r^{1/2}\left(B_{r,\sigma}r-A_{r,\sigma}^2\right)^{1/2}}\right\}.$$

If $\eta<1/2$, based on (3) of Lemma A.1 and (a) of Lemma A.2, the scaling matrix should be $\Upsilon_T=diag\left(T^{1/2},T^{\frac{3}{2}}\right)$. Consider the matrix $\Upsilon_T^{-1}\left[\sum_{t=1}^{[Tr]} X_tX_t'\right]_{(-2)\times(-2)}\Upsilon_T^{-1}$,

$$\left[\begin{array}{cc}T^{1/2} & 0\\ 0 & T^{3/2}\end{array}\right]^{-1}\left[\begin{array}{cc}\Sigma 1 & \Sigma y_{t-1}\\ \Sigma y_{t-1} & \Sigma y_{t-1}^2\end{array}\right]\left[\begin{array}{cc}T^{1/2} & 0\\ 0 & T^{3/2}\end{array}\right]\xrightarrow{L}\left[\begin{array}{cc}r & \frac{1}{2}\alpha_T r^2\\ \frac{1}{2}\alpha_T r^2 & \frac{1}{3}\alpha_T^2 r^3\end{array}\right], \qquad (27)$$

and the matrix $\Upsilon_T^{-1}\left[\sum_{t=1}^{[Tr]} X_t\varepsilon_t\right]_{(-2)\times 1}$,

$$\left[\begin{array}{cc} T^{1/2} & 0 \\ 0 & T^{3/2} \end{array}\right]^{-1}\left[\begin{array}{c} \Sigma\varepsilon_t \\ \Sigma y_{t-1}\varepsilon_t \end{array}\right] \xrightarrow{L} \left[\begin{array}{c} \sigma W(r) \\ \alpha_T\sigma\left[rW(r)-\int_0^r W(s)\,ds\right] \end{array}\right].$$

Under the null hypothesis that $\alpha = T^{-\eta}$ and $\beta = 0$,

$$\left[\begin{array}{c} T^{1/2}(\hat{\alpha}_r-\alpha) \\ T^{3/2}\hat{\beta}_r \end{array}\right] \xrightarrow{L} \left[\begin{array}{c} 2r^{-2}\sigma\left[-rW(r)+3\int_0^r W(s)\,ds\right] \\ 6r^{-3}\alpha_T\sigma\left[rW(r)-2\int_0^r W(s)\,ds\right] \end{array}\right].$$

We can see that

$$T\hat{\beta}_r \xrightarrow{L} 6r^{-3}\alpha_T\sigma\left[rW(r)-2\int_0^r W(s)\,ds\right].$$

The variances of $T\hat{\beta}_r$ can be calculated as follows:

$$var\left(\left[\begin{array}{c} T^{1/2}(\hat{\alpha}_r-\alpha) \\ T^{3/2}\hat{\beta}_r \end{array}\right]\right) = \sigma^2\left\{\left[\begin{array}{cc} T^{1/2} & 0 \\ 0 & T^{3/2} \end{array}\right]^{-1}\left[\begin{array}{cc} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{array}\right]\left[\begin{array}{cc} T^{1/2} & 0 \\ 0 & T^{3/2} \end{array}\right]^{-1}\right\}^{-1}$$

$$\xrightarrow{p} \sigma^2\left[\begin{array}{cc} r & \frac{1}{2}\alpha_T r^2 \\ \frac{1}{2}\alpha_T r^2 & \frac{1}{3}\alpha_T^2 r^3 \end{array}\right]^{-1} = \sigma^2\left[\begin{array}{cc} 4r^{-1} & -6r^{-2}\alpha_T^{-1} \\ -6r^{-2}\alpha_T^{-1} & 12r^{-3}\alpha_T^{-2} \end{array}\right]$$

Hence, the t-statistic of $\hat{\beta}_r$ is

$$ADF_r = \frac{T^{3/2}\hat{\beta}_r}{se\left(T^{3/2}\hat{\beta}_r\right)} \xrightarrow{L} \frac{\int_0^r s\,dW(s) - \int_0^r W(s)\,ds}{\left(\int_0^r s^2 ds\right)^{1/2}}.$$

By CMT, we have

$$SADF(r_0) \xrightarrow{L} \sup_{r\in[r_0,1]}\left\{\frac{\int_0^r s\,dW(s) - \int_0^r W(s)\,ds}{\left(\int_0^r s^2 ds\right)^{1/2}}\right\}.$$