Oxford Bulletin of Economics and Statistics

ORIGINAL ARTICLE OPEN ACCESS

Unit Root Tests for Explosive Financial Bubbles in the Presence of Deterministic Level Shifts

David I. Harvey[1] | Stephen J. Leybourne[1] | Benjamin S. Tatlow[1] | Yang Zu[2]

[1]School of Economics, University of Nottingham, Nottingham, UK | [2]Department of Economics, University of Macau, Macau, China

Correspondence: David I. Harvey (dave.harvey@nottingham.ac.uk)

Received: 13 February 2023 | **Revised:** 4 October 2024 | **Accepted:** 28 January 2025

Funding: The authors received no specific funding for this work.

Keywords: explosive autoregression | level shifts | right-tailed unit root testing | sign-based tests

ABSTRACT

This article considers the issue of testing for an explosive bubble in financial time series in the presence of deterministic level shifts. We demonstrate that the sign-based variants of the Phillips-Shi-Yu test retain their asymptotic validity in the presence of level shifts under a weak restriction on the number of shifts that occur. This is in contrast to the original Phillips-Shi-Yu test which only remains valid under a joint restriction involving both the number and magnitudes of the level shifts. We find, through Monte Carlo simulation, that the original test can display substantial over-size in the presence of level shifts, without a corresponding increase in power, while the sign-based variants are largely unaffected in both regards. The sign-based tests therefore offer robust and powerful methods for detecting an explosive autoregressive regime in a financial time series that potentially contains level shifts. Empirical applications of the different tests are provided using intraday Bitcoin log price data and daily Nasdaq price data.

JEL Classification: C12, C22, C58

1 | Introduction

Empirical identification of explosive behaviour in financial asset price series is closely related to the study of rational bubbles, with a rational bubble deemed to have occurred if explosive characteristics are manifest in the time path of prices, but not for the dividends. As a consequence, methods for testing for explosive time series behaviour have been a focus of much recent research. Phillips [1] [PSY] model potential bubble behaviour using a time-varying autoregressive specification, which allows for an explosive autoregressive regime (or possibly multiple such regimes) in an otherwise unit root autoregression. They suggest testing for such a property using a double supremum of forward and backward recursive right-tailed Dickey-Fuller unit root statistics. The PSY testing approach has rapidly gained status as a standard tool for the detection of rational bubbles.

Motivated by the well-known stylised fact that time-varying, and typically nonstationary, unconditional volatility is seen to be present in the first differences of many financial series (see, e.g., Rapach et al. [2]), Harvey et al. [3] [HLZ] suggest a modification of the PSY testing approach that, unlike the standard PSY test, is size-robust in the presence of time-varying volatility. Instead of forming the PSY statistic from the double sequence of sub-sample Dickey-Fuller statistics applied to the level of an observed series, y_t say, HLZ form the PSY statistic applied to the series of cumulated signs of the first differences of the data, that is a cumulation of $sign(\Delta y_t) = \Delta y_t / |\Delta y_t|$. Since $sign(\Delta y_t)$ is exact invariant to the variance of Δy_t, the HLZ implementation of the PSY test is, by construction, size robust to time-varying volatility present in Δy_t. As this simple sign-based procedure relies on an assumption of a zero median in the underlying innovation process, HLZ also propose a generalised version which is robust to departures from this assumption, based on cumulation of recursively demeaned $sign(\Delta y_t)$.

In this article, we investigate another robustness property of the sign-based approaches to PSY testing for explosive autoregressive behaviour. This is its robustness to the occurrence of *level shifts* in the observed series y_t. In the context of asset price series with high frequency data, such discontinuities, usually referred to as *jumps*, feature widely; see, *inter alia*, Eraker [4], Lee and Mykland [5], Lee [6], Bajgrowicz et al. [7] and Laurent and Shi [8]. Here they can be attributable to, for example, news or announcement effects. For lower frequency data, level shifts in prices are more typically associated with over-arching effects that tend to impact markets as a whole, commonly referred to as *structural changes* in a macroeconomic context. These may arise from, for example, policy changes, macroeconomic regime changes, or international conflict.

Given the prevalence of level shifts in financial time series, it is important to consider their impact on Dickey-Fuller type unit root tests. It is well documented that failure to account for level shifts in standard left-tailed unit root tests (against an alternative of stationarity) can lead to size distortions. The direction and severity of such size distortions involves a complex interplay between the number, magnitude and locations of the level shifts. In efforts to avoid size distortions due to neglected level shifts, the standard approaches involve first identifying the number and locations of level shifts and subsequently using dummy variables to isolate their effects, following on from Perron [9] and Vogelsang and Perron [10] and the many extensions and refinements thereof. However, all such approaches suffer from the drawback that it is entirely possible to under-estimate the number of level shifts and (even if that is not a consideration) poorly estimate their locations. In practical terms, such procedures really only work satisfactorily for a small number of fairly easily identifiable large shifts that are not in close proximity. When there is a large number of shifts of modest magnitude these dummy variable methods are liable to perform badly in terms of location estimation and, consequently, fail to correct the problem of size distortions.

A sign-based approach is almost entirely unaffected by the presence of level shifts. A level shift occurring at time t (i.e., one period shift in Δy_t) can at most only have the effect of changing the value of $sign(\Delta y_t)$, relative to its shift-free value, from +1 to −1 or −1 to +1. As a result of this, the magnitude of the level shift is effectively rendered irrelevant. In terms of PSY tests, this translates into asymptotic size robustness of the sign-based PSY tests in the presence of level shifts, a property not shared by the standard test. Only a weak condition on the allowable number of shifts is required. In contrast, the original test only remains asymptotically valid if a restriction is made involving a trade-off between the number of shifts allowed and their corresponding magnitudes. The sign-based approaches also bypass any need to attempt the potentially unsatisfactory procedure of detecting level breaks via searching for them and testing their individual significance, before constructing the PSY statistic from the dummy variable regression-adjusted data.

In this article, we quantify the extent to which the PSY and sign-based PSY tests are affected by the presence of level shifts through an analysis of their limiting null distributions. While HLZ discuss very briefly the impact of a finite number of level shifts on these tests, we permit the number of shifts to be dependent on the sample size and provide a formal analytical treatment of the role of level shifts in the asymptotic behaviour of the tests. We subsequently evaluate the finite sample sizes and powers of the procedures using Monte Carlo simulations, with generating processes that vary the number and magnitudes of the shifts. For the PSY tests, typically, the shifts lead to substantially increased size but without a corresponding increase in power. Encouragingly, the sign-based tests are much less affected by the presence of level shifts; instead, reliable size control and good levels of power are observed, with the level shifts having relatively little effect on the rejection frequency. We also provide empirical illustrations using intraday Bitcoin log price data and daily Nasdaq price data.

The rest of the article is organised as follows. Section 2 introduces the unit root null model allowing level shifts, and section 3 outlines the PSY and sign-based PSY tests. In section 4, we present the asymptotic properties of the tests under the null in the presence of level shifts (proofs of the asymptotic results being provided in an Appendix). Section 5 compares the finite sample size of the tests and also evaluates their power characteristics under an alternative that contains an explosive autoregressive regime. Our empirical illustrations are given in section 6, and some conclusions are offered in section 7. In this article, we adopt the following notation: $\mathbb{I}(.)$ denotes the indicator function; $\lfloor \cdot \rfloor$ the integer part; $\Rightarrow$ weak convergence and $\xrightarrow{p}$ convergence in probability.

2 | The Unit Autoregressive Root Model With Level Shifts

To characterise the null hypothesis for the tests, we will consider an observed time series process $\{y_t\}$ generated according to the following DGP:

$$y_t = \mu_0 + \sum_{i=1}^{n_T} \mu_{i,T} \mathbb{I}(t \geq t_i) + u_t, \qquad t = 1, \ldots, T,$$
$$u_t = u_{t-1} + \varepsilon_t, \qquad t = 2, \ldots, T. \tag{1}$$

We assume that the initial condition u_1 is such that $u_1 = o_p(T^{1/2})$, and ε_t is a zero-mean innovation process whose precise assumptions are detailed below. Under this DGP, u_t is a unit autoregressive root process throughout the full sample period. This model thereby represents the null hypothesis for PSY-type tests for a bubble; these tests have as their alternative hypothesis a model containing a regime (or more than one regime) where u_t follows an explosive autoregressive process (see section 5.2 for an example).

The model for the observed process y_t in (1) admits a constant term μ_0 and $n_T \geq 0$ deterministic level shifts (or jumps) at time periods t_i, $i = 1, \ldots, n_T$, with respective magnitudes $\mu_{i,T} \neq 0$, $i = 1, \ldots, n_T$. We assume, without loss of generality, that the level shifts follow the natural ordering $1 \leq t_1 < t_2 < \cdots < t_{n_T} \leq T$, so the level of y_t shifts from μ_0 to $\mu_0 + \mu_{1,T}$ at time $t = t_1$, then to $\mu_0 + \mu_{1,T} + \mu_{2,T}$ at time $t = t_2$ and so on. Importantly, we do not assume knowledge of either the number of level shifts, n_T, or the locations of the shifts, t_i. Moreover, we entertain the possibility that the number of level shifts and their magnitudes are not fixed but can depend on T. In what follows, we use n_t to denote the total number of level shifts occurring up to time t; that is $n_t = \sum_{i=1}^{n_T} \mathbb{I}(t \geq t_i)$. Our focus in this article will be testing the unit root null hypothesis using right-tailed Dickey-Fuller-type tests, cf. PSY, in the context of the observed process y_t undergoing n_T level shifts.

For the innovation process ε_t, we make the following assumptions:

Assumption 1. $\varepsilon_t \sim IID(0, \sigma^2)$ with $E(|\varepsilon_t|^j) < \infty$ for some $j \geq 4$.

Assumption 2. The CDF of ε_t, denoted $F(.)$, satisfies $F(0) = 1/2$.

Remark 1. Assumption 1 will be used to establish the limit distribution of the partial sum process [PSP] of ε_t, which is needed for the limit distribution of the PSY statistic. Assumptions 1 and 2 are used for the corresponding PSP of $sign(\varepsilon_t)$, as required for the basic non-recursively demeaned sign-based PSY statistic. Assumption 2 implies that $E(sign(\varepsilon_t)) = 0$, which is necessary for the invariance principle of the PSP of the uncentred signs to hold.[1] Assumption 1 is also used (but without requiring Assumption 2) for the corresponding PSP of $sign(\varepsilon_t) - E[sign(\varepsilon_t)]$, which is required for the recursively demeaned sign-based PSY statistic. These assumptions follow directly from HLZ, apart from our additional simplifying assumption of constant volatility in the ε_t process which is made primarily for expository purposes and could be relaxed along the lines of HLZ.

Under Assumption 1, the following invariance principle holds for the PSP of ε_t:

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \varepsilon_t \Rightarrow \sigma W(r)$$

where $W(r)$ is a standard Brownian motion process. Under Assumptions 1 and 2, a corresponding invariance principle for the PSP of $sign(\varepsilon_t)$ is given by

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} sign(\varepsilon_t) \Rightarrow W_1^s(r)$$

where $W_1^s(r)$ is also a standard Brownian motion process (correlated with $W(r)$). Lastly, under Assumption 1 alone, that is, where Assumption 2 is not imposed,

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \left\{ sign(\varepsilon_t) - E[sign(\varepsilon_t)] \right\} \Rightarrow \sigma_s W_2^s(r)$$

where $W_2^s(r)$ is again a standard Brownian motion process (correlated with $W(r)$), and $\sigma_s^2 = Var[sign(\varepsilon_t)]$.

As regards the level shifts, we present the following pairs of assumptions, with Assumption 3 relevant for the PSY test and Assumption 4 appropriate for the sign-based tests:

Assumption 3. (i) $\sup_{r \in [0,1]} \left| \sum_{i=1}^{n_{\lfloor rT \rfloor}} \mu_{i,T} \right| = O(T^{\alpha_{n,\mu}})$ with $0 \leq \alpha_{n,\mu} \leq 1/2$. (ii) $\sum_{i=1}^{n_T} \mu_{i,T}^2 = o(T)$.

Assumption 4. $n_T = O(T^{\alpha_n})$ with $0 \leq \alpha_n \leq 1/2$.

Remark 2. Assumptions 3 and 4 are made for the PSY and sign-based PSY statistics, respectively, to ensure that the level shifts can feature (but not dominate) the relevant limit null distributions. We note that Assumption 3 for the PSY test involves requirements on both the number of level shifts n_T and the shift magnitudes $\mu_{i,T}$, whereas Assumption 4 for the sign-based variants is weaker in the sense that only the number of shifts is restricted, not the shift magnitudes. By way of simple examples, suppose the shift magnitudes are common ($\mu_{i,T} = \mu_T$) and the number of shifts is given by $n_T = O(T^{1/4})$, then Assumption 3(i) additionally imposes that the magnitudes μ_T are at most $O(T^{1/4})$; alternatively, if $n_T = O(1)$, Assumption 3(ii) restricts the shift magnitude to be at most $o(T^{1/2})$. No such constraints on the shift magnitudes are involved under Assumption 4.

3 | The PSY and Sign-Based PSY Tests

In this section, we briefly outline the test of PSY and the corresponding sign-based variants of HLZ. The PSY statistic is given by

$$PSY = \sup_{\lambda_1 \in [0, 1-\pi]} \sup_{\lambda_2 \in [\lambda_1 + \pi, 1]} DF(\lambda_1, \lambda_2) \tag{2}$$

where $DF(\lambda_1, \lambda_2)$ denotes the standard Dickey-Fuller statistic, that is, the t-ratio for $\hat{\phi}(\lambda_1, \lambda_2)$ in the fitted ordinary least squares (OLS) regression

$$\Delta y_t = \hat{\alpha}(\lambda_1, \lambda_2) + \hat{\phi}(\lambda_1, \lambda_2) y_{t-1} + \hat{\varepsilon}_t \tag{3}$$

calculated over the sub-sample period $t = \lfloor \lambda_1 T \rfloor, \ldots, \lfloor \lambda_2 T \rfloor$. That is

$$DF(\lambda_1, \lambda_2) = \frac{\hat{\phi}(\lambda_1, \lambda_2)}{\sqrt{\hat{\sigma}^2(\lambda_1, \lambda_2) \sum_{t=\lfloor \lambda_1 T \rfloor}^{\lfloor \lambda_2 T \rfloor} \left(y_{t-1} - \overline{y} \right)^2}}$$

where $\overline{y} = (\lfloor \lambda_2 T \rfloor - \lfloor \lambda_1 T \rfloor + 1)^{-1} \sum_{t=\lfloor \lambda_1 T \rfloor}^{\lfloor \lambda_2 T \rfloor} y_{t-1}$ and $\hat{\sigma}^2(\lambda_1, \lambda_2) = (\lfloor \lambda_2 T \rfloor - \lfloor \lambda_1 T \rfloor - 1)^{-1} \sum_{t=\lfloor \lambda_1 T \rfloor}^{\lfloor \lambda_2 T \rfloor} \hat{\varepsilon}_t^2$. The PSY statistic is therefore the supremum of a double sequence of statistics with minimum sample length $\lfloor \pi T \rfloor$, with right-tailed testing of PSY performed to distinguish between the null and an explosive autoregressive alternative.

The first sign-based analogue of (2) proposed by HLZ is based on the cumulative sum of signs, $C_t = \sum_{i=2}^{t} sign(\Delta y_i), t = 2, \ldots, T$, and is given by

$$sPSY = \sup_{\lambda_1 \in [0, 1-\pi]} \sup_{\lambda_2 \in [\lambda_1 + \pi, 1]} sDF(\lambda_1, \lambda_2)$$

where $sDF(\lambda_1, \lambda_2)$ denotes the t-ratio for $\hat{\rho}(\lambda_1, \lambda_2)$ in the fitted (without intercept) OLS regression

$$\Delta C_t = \hat{\rho}(\lambda_1, \lambda_2) C_{t-1} + e_t \tag{4}$$

calculated over the period $t = \lfloor \lambda_1 T \rfloor, \ldots, \lfloor \lambda_2 T \rfloor$, that is,

$$sDF(\lambda_1, \lambda_2) = \frac{\hat{\rho}(\lambda_1, \lambda_2)}{\sqrt{\hat{s}^2(\lambda_1, \lambda_2)\sum_{t=\lfloor \lambda_1 T \rfloor}^{\lfloor \lambda_2 T \rfloor} C_{t-1}^2}}$$

where $\hat{s}^2(\lambda_1, \lambda_2) = (\lfloor \lambda_2 T \rfloor - \lfloor \lambda_1 T \rfloor)^{-1} \sum_{t=\lfloor \lambda_1 T \rfloor}^{\lfloor \lambda_2 T \rfloor} e_t^2$.

The second sign-based analogue of (2) proposed by HLZ, denoted $\overline{s}PSY$, takes the same form as $sPSY$ above, but with C_t redefined as the recursively demeaned variant $C_t = \sum_{i=2}^{t} \left\{ sign(\Delta y_i) - (i-1)^{-1} \sum_{j=2}^{i} sign(\Delta y_j) \right\}, \quad t = 2, \ldots, T$. As with PSY, right-tailed testing is used for $sPSY$ and $\overline{s}PSY$.

4 | Large Sample Behaviour of PSY, $sPSY$ and $\overline{s}PSY$

We now consider the large sample behaviour of PSY, $sPSY$ and $\overline{s}PSY$ under the null model presented in section 2, where n_T level shifts are present in the series y_t. Our results are given in the next three theorems, noting that we rely on different assumptions regarding the number and magnitude of level shifts for PSY compared to $sPSY$ and $\overline{s}PSY$.

Theorem 1. *For the null model (1), under Assumptions 1 and 3,*

$$PSY \Rightarrow \sup_{\lambda_1 \in [0, 1-\pi]} \sup_{\lambda_2 \in [\lambda_1+\pi, 1]} L(\lambda_1, \lambda_2)$$

where

$$L(\lambda_1, \lambda_2) = \frac{\overline{H}(\lambda_2)^2 - \overline{H}(\lambda_1)^2 - (\lambda_2 - \lambda_1)}{2\sqrt{\int_{\lambda_1}^{\lambda_2} \overline{H}(r)^2 dr}},$$

$$\overline{H}(r) = H(r) - (\lambda_2 - \lambda_1)^{-1} \int_{\lambda_1}^{\lambda_2} H(s) ds$$

and

$$H(r) = \begin{cases} W(r) & 0 \le \alpha_{n,\mu} < 1/2 \\ W(r) + \sigma^{-1} J(r) & \alpha_{n,\mu} = 1/2 \end{cases}$$

with

$$J(r) = \lim_{T \to \infty} (T^{-1/2} \sum_{i=1}^{n_{\lfloor rT \rfloor}} \mu_{i,T})$$

Remark 3. We observe that the limit null distribution of PSY is dependent on both the number of level shifts and their magnitudes, via $\alpha_{n,\mu}$. PSY will obtain its usual null limit distribution provided $0 \le \alpha_{n,\mu} < 1/2$, since under that condition, $H(r) = W(r)$. If, however, $\alpha_{n,\mu} = 1/2$, an additive term proportional to $J(r)$ appears in $H(r)$. In essence, $J(r)$ is a limit measure of the total amount of asymptotically non-negligible level shift that has occurred up to time $\lfloor rT \rfloor$. Given that $L(\lambda_1, \lambda_2)$ essentially depends on the squares of $H(r)$, we would expect the limit distribution of PSY to be right-shifted in the case $\alpha_{n,\mu} = 1/2$, regardless of the sign of $J(r)$, relative to the $0 \le \alpha_{n,\mu} < 1/2$ case where $J(r)$ is absent. We would anticipate, therefore, that PSY is likely to be over-sized in the $\alpha_{n,\mu} = 1/2$ case, with the over-size increasing in the magnitude of $J(r)$. Note that, when $\alpha_{n,\mu} = 1/2$, $J(r)$ can still be zero for some values of r (e.g., the sum of positive and negative level shifts up to time $\lfloor rT \rfloor$ could be zero), but this cannot be the case for all r. We do not consider the case $\alpha_{n,\mu} > 1/2$, since here the deterministic level shifts component of the DGP would entirely dominate the limit behaviour of PSY.

Theorem 2. *For the null model (1), under Assumptions 1, 2 and 4,*

$$sPSY \Rightarrow \sup_{\lambda_1 \in [0, 1-\pi]} \sup_{\lambda_2 \in [\lambda_1+\pi, 1]} L^s(\lambda_1, \lambda_2)$$

where

$$L^s(\lambda_1, \lambda_2) = \frac{H^s(\lambda_2)^2 - H^s(\lambda_1)^2 - (\lambda_2 - \lambda_1)}{2\sqrt{\int_{\lambda_1}^{\lambda_2} H^s(r)^2 dr}}$$

and

$$H^s(r) = \begin{cases} W_1^s(r) & 0 \le \alpha_n < 1/2 \\ W_1^s(r) + K^s(r) & \alpha_n = 1/2 \end{cases}$$

with

$$K^s(r) = \lim_{T \to \infty} (T^{-1/2} \sum_{i=1}^{n_{\lfloor rT \rfloor}} sign(\varepsilon_{t_i} + \mu_{i,T}) - T^{-1/2} \sum_{i=1}^{n_{\lfloor rT \rfloor}} sign(\varepsilon_{t_i}))$$

where $|K^s(r)| \le 2\kappa_r \le 2\kappa_1$ *with* $\kappa_r = \lim_{T \to \infty} T^{-1/2} n_{\lfloor rT \rfloor}$.

Theorem 3. *For the null model (1), under Assumptions 1 and 4,*

$$\overline{s}PSY \Rightarrow \sup_{\lambda_1 \in [0, 1-\pi]} \sup_{\lambda_2 \in [\lambda_1+\pi, 1]} \overline{L}^s(\lambda_1, \lambda_2)$$

where

$$\overline{L}^s(\lambda_1, \lambda_2) = \frac{\overline{G}^s(\lambda_2)^2 - \overline{G}^s(\lambda_1)^2 - (\lambda_2 - \lambda_1)}{2\sqrt{\int_{\lambda_1}^{\lambda_2} \overline{G}^s(r)^2 dr}}$$

with

$$\overline{G}^s(r) = G^s(r) - \int_0^r x^{-1} G^s(x) dx$$

and

$$G^s(r) = \begin{cases} W_2^s(r) & 0 \le \alpha_n < 1/2 \\ W_2^s(r) + \sigma_s^{-1} K^s(r) & \alpha_n = 1/2 \end{cases}$$

with $\sigma_s^2 = Var[sign(\varepsilon_t)]$ *and* $K^s(r)$ *as defined in Theorem 2.*

Remark 4. In contrast to the limit null distribution of PSY, it is seen that the limit null distributions of $sPSY$ and $\overline{s}PSY$ are dependent on the number of level shifts, via α_n, but not on their orders of magnitude. When $0 \le \alpha_n < 1/2$, $sPSY$ and $\overline{s}PSY$ will obtain the same null limit distributions as in HLZ, since then $H^s(r) = W_1^s(r)$ and $G^s(r) = W_2^s(r)$. However, if $\alpha_n = 1/2$, the term $K^s(r)$ appears in $H^s(r)$ and $G^s(r)$, and therefore we would expect $sPSY$ and $\overline{s}PSY$ to be over-sized. Note that while the $\mu_{i,T}$ appear in the expression for $K^s(r)$, this term is bounded by $\pm 2\kappa_1$ for any orders of magnitude for the $\mu_{i,T}$, hence we would expect the degree of over-size to be largely unaffected by the magnitude of the level shifts. It can also be seen that when $\alpha_n = 1/2$, the limit behaviour of $\overline{s}PSY$ depends on $\sigma_s^2 = Var[sign(\varepsilon_t)]$ (via the dependence of $G^s(r)$ on this quantity), with $\sigma_s \neq 1$ in general when Assumption 2 is not imposed. Note that, when $\alpha_n = 1/2$, $K^s(r)$ can be zero for some values of r, but not for all r. Again, we do not consider the case $\alpha_n > 1/2$, since here the number of level shifts present in the DGP could dominate the behaviour of the statistic.

In summary, our theoretical results show that the sign-based PSY tests offer a natural advantage over the original PSY test when considering an environment with level shifts of unknown magnitude. More specifically, provided the number of level shifts is of a smaller order than $T^{1/2}$, then irrespective of the magnitude of the level shifts, the sign-based PSY tests will retain the same asymptotic null distributions as in the case of no level shifts, and therefore the tests can be applied using the standard critical values reported in HLZ. In contrast, asymptotic validity of the standard PSY test (developed in the case of no level shifts) only holds when the magnitude of the level shifts is restricted, with a trade-off between the number of level shifts and their permitted magnitudes. At a practical level, assuming the number of level shifts present is $o(T^{1/2})$, these findings suggest that the sign-based PSY tests will be size robust, whereas the PSY test would be expected to display size distortions, particularly when large magnitude level shifts are present.

Remark 5. While our focus in this article is on test robustness to level shifts, HLZ show that the sign-based PSY tests embody inherent robustness to unconditional heteroskedasticity in the innovations ε_t, since $sign(\sigma_t \varepsilon_t) = sign(\varepsilon_t)$ for any positive σ_t. The same is not true of the PSY test, as HLZ demonstrate.

Remark 6. We have assumed thus far that ε_t is serially uncorrelated. If ε_t is a finite order stationary autoregressive process then the extension to the PSY procedure is to augment the fitted regression (3) with lagged values of Δy_t. For $sPSY$ and $\overline{s}PSY$, the method proposed in HLZ for dealing with serial correlation (see section 8.1 of HLZ) is not appropriate in the context of level shifts, since the correction involved uses augmented DF-type regressions which neglect the level shifts. Instead, the corresponding adjustment we propose is to augment the regression (4) with lagged values of ΔC_t. According to the discussion in Section 2.6.1 of Fan and Yao [11], ε_t is β-mixing with coefficients decaying to 0 exponentially fast and $\Delta C_t = sign(\Delta y_t)$ is a measurable transform of Δy_t that possesses the same mixing properties as ε_t and also trivially has β-mixing coefficients decaying at least as fast as that of Δy_t. Then by Theorem 2.1 of Paparoditis and Politis [12], the result of Theorem 2 still holds. On a cautionary note, when $\lfloor \lambda_1 T \rfloor, \ldots, \lfloor \lambda_2 T \rfloor$ is small, it is possible that the matrix $M = [\Delta C, C_{-1}, \Delta C_{-1}, \Delta C_{-2}, \ldots]$ can be less than (or close to less than) full rank; this can easily be checked by examining the eigenvalues of $M'M$. As a practical measure we suggest that in such cases calculation of $sDF(\lambda_1, \lambda_2)$ or $\overline{s}DF(\lambda_1, \lambda_2)$ is not attempted, effectively excluding it from consideration when calculating the supremum. This is only an issue we have encountered for small T and short intervals $\lambda_2 - \lambda_1$.

5 | Finite Sample Comparison of the Tests

In this section, we consider the finite sample sizes and powers of the PSY, $sPSY$ and $\overline{s}PSY$ tests in the presence of level shifts. Throughout this section, our simulations are based on 2000 Monte Carlo replications. We calculate the tests using $\pi = 0.1$, and conduct right-tail tests at the 0.05-level, using finite sample critical values generated under the null model (1) with $\varepsilon_t \sim IIDN(0,1)$ in the absence of any level shifts, for each value of T considered.

5.1 | Finite Sample Size

We first consider the size of the tests for a range of sample sizes, in relation to the asymptotic results of the previous section, using $T = \{100, 200, 400, 800, 1600\}$. Here, we initially set $\varepsilon_t \sim IIDN(0,1)$ (as in the computation of the critical values), and set $u_1 = \varepsilon_1$ and $\mu_0 = 0$ (all tests are invariant to μ_0). The level shifts in (1) are specified as follows. We introduce a total of $n_T = \lfloor kT^{\alpha_n} \rfloor$ level shifts, divided into $n_T^+ = pn_T$ (rounded to the nearest integer) positive shifts and $n_T^- = n_T - n_T^+$ negative shifts, with the settings $p = \{0.8, 0.6\}$.[2] The corresponding shift magnitudes are set to $\mu_{i,T} = \pm\mu_T$ with $\mu_T = \mu T^{\alpha_\mu}$, $\mu > 0$ (such that the n_T^+ positive level shifts have magnitude μ_T, and the n_T^- negative level shifts have magnitude $-\mu_T$). The locations, t_i, of the level shifts are generated as n_T independent drawings from a $\lfloor T \times U[0,1] \rfloor + 1$ distribution, excluding the possibility of repeated shift timings; these drawings are also independent across replications.

We consider the following three constellations of settings for the number and magnitude of the level shifts:

Case 1. $\{\alpha_n, \alpha_\mu\} = \{0.5, 0\}$, $k = \{1, 2\}$, $\mu = \{2.5, 5\}$.

Case 2. $\{\alpha_n, \alpha_\mu\} = \{0.25, 0.25\}$, $k = \{3, 6\}$, $\mu = \{1, 2\}$.

Case 3. $\{\alpha_n, \alpha_\mu\} = \{0, 0\}$, $k = \{5, 10\}$, $\mu = \{2.5, 5\}$.

The three cases are distinguished by the α_n and α_μ rates at which the number and magnitude of level shifts increases in T. Within each case, two settings for k and μ are chosen, to allow us to see the impact of smaller and larger numbers and magnitudes. Under Case 1, $\sup_{r\in[0,1]} \left|\sum_{i=1}^{n_{\lfloor rT \rfloor}} \mu_{i,T}\right| = O(T^{0.5})$, that is, $\alpha_{n,\mu} = 0.5$, hence we would expect the size of the PSY test to be affected by the level shifts, given the large sample results of Theorem 1. Since $\alpha_n = 0.5$ in this case, we would also expect the size of the $sPSY$ and $\overline{s}PSY$ tests to be affected, given the results of Theorems 2 and 3. Under Case 2, $\sup_{r\in[0,1]} \left|\sum_{i=1}^{n_{\lfloor rT \rfloor}} \mu_{i,T}\right| = O(T^{0.5})$, that is, $\alpha_{n,\mu} = 0.5$, and again the size of PSY would be expected to be impacted by the shifts. In contrast, in this case we would expect $sPSY$ and $\overline{s}PSY$ to be correctly sized in large samples, given that here $\alpha_n = 0.25$. Finally, under Case 3, $\sup_{r\in[0,1]} \left|\sum_{i=1}^{n_{\lfloor rT \rfloor}} \mu_{i,T}\right| = O(1)$, that is, $\alpha_{n,\mu} = 0$, and also $\alpha_n = 0$, hence we would expect all tests to be correctly sized asymptotically.

Tables 1–3 report the finite sample sizes of the tests corresponding to Cases 1-3 respectively. First consider Table 1, which provides results for Case 1. For both $p = 0.8$ and $p = 0.6$, we find that PSY is never correctly sized, and the degree of over-size is increasing in k and μ, the parameters controlling the number and magnitude of level shifts, respectively. This is consistent with the large sample theory of the previous section (see Remark 3), with the term $J(r)$ being larger in magnitude for larger k and/or μ, other things equal. These size distortions are clearly observed for all sample sizes considered, and are most apparent for $p = 0.8$,

TABLE 1 | Finite sample sizes of nominal 0.05-level tests: Case 1, $\alpha_n = 0.5$, $\alpha_\mu = 0$.

					$p = 0.8$					$p = 0.6$				
k	μ	T	n_T	μ_T	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$
1	2.5	100	10	2.5	8	2	0.073	0.064	0.063	6	4	0.063	0.049	0.058
		200	14	2.5	11	3	0.076	0.072	0.055	8	6	0.054	0.059	0.056
		400	20	2.5	16	4	0.097	0.070	0.048	12	8	0.059	0.052	0.046
		800	28	2.5	22	6	0.092	0.070	0.050	17	11	0.060	0.056	0.048
		1600	40	2.5	32	8	0.095	0.061	0.054	24	16	0.060	0.047	0.053
1	5	100	10	5	8	2	0.164	0.065	0.063	6	4	0.124	0.050	0.058
		200	14	5	11	3	0.166	0.073	0.056	8	6	0.112	0.059	0.058
		400	20	5	16	4	0.198	0.070	0.049	12	8	0.107	0.051	0.047
		800	28	5	22	6	0.182	0.070	0.051	17	11	0.096	0.054	0.049
		1600	40	5	32	8	0.196	0.063	0.055	24	16	0.083	0.047	0.054
2	2.5	100	20	2.5	16	4	0.123	0.126	0.070	12	8	0.071	0.046	0.053
		200	28	2.5	22	6	0.121	0.108	0.054	17	11	0.059	0.048	0.050
		400	40	2.5	32	8	0.166	0.119	0.048	24	16	0.070	0.052	0.048
		800	56	2.5	45	11	0.203	0.121	0.049	34	22	0.081	0.056	0.050
		1600	80	2.5	64	16	0.198	0.119	0.050	48	32	0.072	0.059	0.045
2	5	100	20	5	16	4	0.240	0.126	0.069	12	8	0.149	0.045	0.052
		200	28	5	22	6	0.268	0.111	0.054	17	11	0.131	0.048	0.051
		400	40	5	32	8	0.337	0.121	0.050	24	16	0.115	0.052	0.048
		800	56	5	45	11	0.414	0.126	0.048	34	22	0.122	0.057	0.050
		1600	80	5	64	16	0.425	0.119	0.048	48	32	0.105	0.059	0.045

TABLE 2 | Finite sample sizes of nominal 0.05-level tests: Case 2, $\alpha_n = 0.25$, $\alpha_\mu = 0.25$.

					$p = 0.8$					$p = 0.6$				
k	μ	T	n_T	μ_T	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$
3	1	100	9	3.162	7	2	0.089	0.059	0.046	5	4	0.079	0.048	0.046
		200	11	3.761	9	2	0.118	0.066	0.050	7	4	0.088	0.051	0.053
		400	13	4.472	10	3	0.109	0.061	0.053	8	5	0.090	0.056	0.051
		800	15	5.318	12	3	0.128	0.055	0.050	9	6	0.090	0.054	0.044
		1600	18	6.325	14	4	0.105	0.046	0.051	11	7	0.073	0.041	0.056
3	2	100	9	6.325	7	2	0.193	0.059	0.046	5	4	0.165	0.048	0.046
		200	11	7.521	9	2	0.259	0.066	0.050	7	4	0.201	0.051	0.053
		400	13	8.944	10	3	0.238	0.061	0.053	8	5	0.185	0.056	0.051
		800	15	10.637	12	3	0.266	0.055	0.050	9	6	0.165	0.054	0.044
		1600	18	12.649	14	4	0.227	0.046	0.051	11	7	0.141	0.041	0.056
6	1	100	18	3.162	14	4	0.127	0.099	0.071	11	7	0.085	0.052	0.062
		200	22	3.761	18	4	0.216	0.094	0.054	13	9	0.097	0.042	0.047
		400	26	4.472	21	5	0.219	0.083	0.043	16	10	0.098	0.055	0.044
		800	31	5.318	25	6	0.250	0.075	0.050	19	12	0.100	0.051	0.056
		1600	37	6.325	30	7	0.249	0.063	0.050	22	15	0.084	0.044	0.049
6	2	100	18	6.325	14	4	0.245	0.099	0.071	11	7	0.165	0.052	0.062
		200	22	7.521	18	4	0.394	0.094	0.054	13	9	0.198	0.042	0.047
		400	26	8.944	21	5	0.421	0.083	0.043	16	10	0.214	0.055	0.044
		800	31	10.637	25	6	0.441	0.075	0.050	19	12	0.195	0.051	0.056
		1600	37	12.649	30	7	0.433	0.063	0.050	22	15	0.154	0.044	0.049

TABLE 3 | Finite sample sizes of nominal 0.05-level tests: Case 3, $\alpha_n = 0$, $\alpha_\mu = 0$.

					$p = 0.8$					$p = 0.6$				
k	μ	T	n_T	μ_T	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$
5	2.5	100	5	2.5	4	1	0.060	0.055	0.056	3	2	0.062	0.048	0.059
		200	5	2.5	4	1	0.057	0.051	0.057	3	2	0.056	0.052	0.055
		400	5	2.5	4	1	0.054	0.053	0.051	3	2	0.050	0.051	0.049
		800	5	2.5	4	1	0.053	0.051	0.049	3	2	0.051	0.049	0.051
		1600	5	2.5	4	1	0.052	0.050	0.049	3	2	0.053	0.050	0.050
5	5	100	5	5	4	1	0.111	0.055	0.056	3	2	0.108	0.047	0.059
		200	5	5	4	1	0.095	0.051	0.056	3	2	0.089	0.052	0.054
		400	5	5	4	1	0.076	0.053	0.051	3	2	0.070	0.051	0.049
		800	5	5	4	1	0.068	0.051	0.049	3	2	0.065	0.049	0.051
		1600	5	5	4	1	0.059	0.050	0.049	3	2	0.056	0.049	0.050
10	2.5	100	10	2.5	8	2	0.073	0.064	0.063	6	4	0.063	0.049	0.058
		200	10	2.5	8	2	0.070	0.053	0.054	6	4	0.053	0.050	0.050
		400	10	2.5	8	2	0.061	0.059	0.049	6	4	0.055	0.054	0.051
		800	10	2.5	8	2	0.061	0.051	0.051	6	4	0.058	0.046	0.051
		1600	10	2.5	8	2	0.056	0.049	0.053	6	4	0.051	0.049	0.055
10	5	100	10	5	8	2	0.164	0.065	0.063	6	4	0.124	0.050	0.058
		200	10	5	8	2	0.131	0.054	0.054	6	4	0.101	0.050	0.050
		400	10	5	8	2	0.111	0.059	0.049	6	4	0.089	0.054	0.052
		800	10	5	8	2	0.085	0.051	0.051	6	4	0.068	0.047	0.051
		1600	10	5	8	2	0.070	0.049	0.053	6	4	0.057	0.049	0.055

where there are substantially more positive than negative level shifts. When $p = 0.6$, there is potential for a greater degree of cancellation of the level shifts within a given sub-sample window, resulting in less over-size than in the $p = 0.8$ case. As regards $sPSY$, there is evidence of some, more modest, over-size when $p = 0.8$, although there is very little in the way of size distortions when $p = 0.6$. As expected given our theoretical results, the size distortions associated with $sPSY$ are almost entirely unaffected by the magnitude of the level shifts, with very similar results obtained for the two different settings of μ considered. The $\bar{s}PSY$ test shows little evidence of size distortion for either $p = 0.8$ or $p = 0.6$, thereby displaying an inherently greater robustness to level shifts than $sPSY$. The fact that almost no size distortions are observed is somewhat surprising, given the results of Theorem 3, which show that the large sample distribution of $\bar{s}PSY$, like that for $sPSY$, does depend on the level shift-dependent term $K^s(r)$. Evidently, the recursive demeaning has the effect of reducing the impact of the level shifts on the size of the test. This feature is in part related to the uniformly distributed locations of the level shifts; in unreported simulations where non-uniform locations were generated, over-size in $\bar{s}PSY$ was observed. The overall picture from Table 1 is that, in this Case 1 setting where all tests are asymptotically affected by the presence of level shifts, PSY is by far the most size distorted procedure, with $sPSY$ and $\bar{s}PSY$ displaying a considerable greater degree of robustness.

Next consider Table 2. Here, the results are for Case 2, where our theoretical results suggest that PSY should be over-sized in large samples, whereas $sPSY$ and $\bar{s}PSY$ should be correctly sized. The finite sample size results in the table indeed show that PSY is over-sized, whereas both the sign-based variants control size extremely well. When $p = 0.8$, some modest over-sizing is seen for $sPSY$ and, to a lesser extent, $\bar{s}PSY$, for the smaller sample sizes and larger number of level shifts, but these distortions are diminishing as the sample size increases. In contrast, the over-size associated with PSY is evident for all sample sizes and level shift settings, and can be very substantial, particularly when $p = 0.8$. As was observed in Table 1, the over-size of PSY is increasing in both the number and magnitude of the level shifts. The general finding for Case 2 is again that $sPSY$ and $\bar{s}PSY$ offer considerably better size control than PSY.

In Table 3, the results are given for Case 3, where both the number and magnitude of the level shifts are not increasing in the sample size, and all three tests should have asymptotically correct size according to Theorems 1–3. Indeed we see this feature borne out in the results, with all tests close to nominal size for the largest sample size considered. In the smaller sample sizes, whereas $sPSY$ and $\bar{s}PSY$ retain excellent size control across the different level shift number and magnitude settings, some non-trivial over-size is observed for PSY for the larger magnitude level shifts. Once again, $sPSY$ and $\bar{s}PSY$ emerge as the tests with superior finite sample size properties relative to PSY.

Finally, given that $sPSY$ relies on Assumption 2, that is, that the distribution of the errors has zero median, it is of interest to evaluate the sizes of the procedures when this assumption is

TABLE 4 | Finite sample sizes of nominal 0.05-level tests: Case 2, $\alpha_n = 0.25$, $\alpha_\mu = 0.25$, $\chi^2(\nu)$ errors, $p = 0.8$.

							$\nu = 5$			$\nu = 10$		
k	μ	T	n_T	μ_T	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	PSY	$sPSY$	$\bar{s}PSY$
0	0	100	0	0	0	0	0.048	0.221	0.093	0.036	0.121	0.075
		200	0	0	0	0	0.041	0.361	0.062	0.040	0.199	0.065
		400	0	0	0	0	0.051	0.651	0.060	0.050	0.386	0.049
		800	0	0	0	0	0.056	0.941	0.050	0.049	0.655	0.048
		1600	0	0	0	0	0.051	1.000	0.055	0.045	0.944	0.050
6	2	100	18	6.325	14	4	0.248	0.048	0.056	0.238	0.040	0.060
		200	22	7.521	18	4	0.380	0.108	0.058	0.372	0.052	0.051
		400	26	8.944	21	5	0.423	0.370	0.055	0.420	0.158	0.047
		800	31	10.637	25	6	0.454	0.800	0.048	0.456	0.426	0.051
		1600	37	12.649	30	7	0.449	0.997	0.054	0.427	0.848	0.046

violated. To this end, we next consider finite sample size simulations where the critical values for all tests are obtained using $\varepsilon_t \sim IIDN(0,1)$, as above, but now where the simulated DGPs make use of an asymmetric error distribution; specifically we set $\varepsilon_t \sim IID\chi^2(\nu)$, with $\nu = \{5, 10\}$, standardised to have zero mean and unit variance. We consider Case 2 and in Table 4 report results both when no level shifts are present ($k = 0$) and a case where level shifts occur ($k = 6$ with $\mu = 2$). As might be expected, $sPSY$ lacks size control for these asymmetric errors, with over-size worsening in the sample size. This is essentially due to the fact that the test omits any form of mean correction to the sign process. In contrast, $\bar{s}PSY$ displays excellent size control, irrespective of whether level shifts are present or not, due to the recursive demeaning of the sign process in that procedure. The PSY test is correctly sized when no level shifts are present but, in line with the results for Table 2, is subject to substantial over-size when level shifts occur. It is clear then that if concern exists regarding asymmetry of the error distribution, $\bar{s}PSY$ represents the only reliable procedure in terms of size control in the possible presence of level shifts.

5.2 | Finite Sample Power

We next consider the finite sample powers of the tests, using the sample size $T = 200$. As regards the specification of the alternative hypothesis, we replace u_t in (1) with the following:

$$u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 2, \ldots, \lfloor \tau_1 T \rfloor, \\ (1+\delta)u_{t-1} + \varepsilon_t, & t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor, \\ u_{t-1} + \varepsilon_t, & t = \lfloor \tau_2 T \rfloor + 1, \ldots, T \end{cases} \qquad (5)$$

where $\delta \geq 0$ and $0 < \tau_1 < \tau_2 < 1$. When $\delta > 0$, the u_t process changes at time $\lfloor \tau_1 T \rfloor$ from unit root to explosive autoregressive dynamics (with explosive offset δ), providing a model of bubble behaviour. The explosive behaviour terminates at time $\lfloor \tau_2 T \rfloor$, at which point the process reverts to unit root behaviour. We set $\tau_1 = 0.3$, $\tau_2 = 0.7$ and $\delta \in \{0.00, 0.01, 0.02, 0.03, 0.04, 0.05\}$, with $\delta = 0$ representing the null model and the $\delta > 0$ values representing different alternatives (with increasing δ corresponding to increasing levels of explosivity). As before, we specify $u_1 = \varepsilon_1$ and $\mu_0 = 0$ (all tests are again invariant to μ_0). We set $\varepsilon_t \sim IIDN(0,1)$ and specify the level shifts in the same way as our size simulations, covering Cases 1-3.

Table 5 presents results for Case 1. In very general terms, we observe that the power levels associated with the different tests do not change much across the settings for the level shifts (k, μ and p). For PSY this is in contrast to the size results, where the number and magnitude of positive and negative level shifts has a substantial impact on the rejection frequency. Some exceptions to this general observation occur for PSY when $k = 2$ and $\mu = 5$, that is, where the level shifts are most pronounced; for example, when $p = 0.8$, power is higher for $\delta = 0.01$ than the corresponding power for other k and μ, driven by the large size found in this case. Comparing the powers of the different tests, we observe that the powers of PSY and $sPSY$ are largely similar, despite the fact that the sizes of the PSY test are generally well in excess of those for $sPSY$. The powers of $\bar{s}PSY$ are somewhat lower than the corresponding $sPSY$ powers for the smaller values of δ, which might be expected given that $\bar{s}PSY$ does not suffer from the modest over-size seen for $sPSY$. Exceptions to these comments again arise for the case of $k = 2$ and $\mu = 5$: when $p = 0.8$, PSY power exceeds that of $sPSY$ for small δ, due to its substantial over-size; when $p = 0.6$, the power of PSY falls below that of $sPSY$ for moderate δ.

The results for Case 2 are presented in Table 6. As was the case in Table 5, the powers for $sPSY$ and $\bar{s}PSY$ are pretty insensitive to the settings for the level shifts (k, μ and p). In contrast, the power of PSY for small and moderate δ varies more widely with the level shift settings. The changes in power across the different level shift settings correspond very closely to the pattern of size distortion for PSY across these settings. Generally, the only cases where PSY displays substantial power gains over $sPSY$ are cases where PSY suffers from substantial over-size and δ is relatively small. Moreover, $sPSY$ can have power levels that exceed those of PSY, while simultaneously possessing greater size control—see, for example, cases of moderate δ when $k = 6$ and $p = 0.6$. We observe that $\bar{s}PSY$ has power lower than $sPSY$ for smaller δ, with this modest power loss representing the cost

TABLE 5 | Finite sample powers of nominal 0.05-level tests: Case 1, $\alpha_n = 0.5$, $\alpha_\mu = 0$, $T = 200$, $\tau_1 = 0.3$, $\tau_2 = 0.7$.

					$p = 0.8$					$p = 0.6$				
k	μ	n_T	μ_T	δ	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$
1	2.5	14	2.5	0.00	11	3	0.076	0.072	0.055	8	6	0.054	0.059	0.056
				0.01	11	3	0.126	0.146	0.075	8	6	0.103	0.138	0.075
				0.02	11	3	0.393	0.420	0.241	8	6	0.389	0.423	0.239
				0.03	11	3	0.726	0.724	0.630	8	6	0.734	0.720	0.629
				0.04	11	3	0.888	0.881	0.842	8	6	0.888	0.883	0.847
				0.05	11	3	0.958	0.950	0.936	8	6	0.958	0.949	0.938
1	5	14	5	0.00	11	3	0.166	0.073	0.056	8	6	0.112	0.059	0.058
				0.01	11	3	0.193	0.147	0.074	8	6	0.139	0.139	0.074
				0.02	11	3	0.365	0.420	0.238	8	6	0.330	0.422	0.236
				0.03	11	3	0.684	0.722	0.620	8	6	0.671	0.717	0.622
				0.04	11	3	0.864	0.879	0.837	8	6	0.866	0.882	0.840
				0.05	11	3	0.950	0.949	0.933	8	6	0.951	0.947	0.936
2	2.5	28	2.5	0.00	22	6	0.121	0.108	0.054	17	11	0.059	0.048	0.050
				0.01	22	6	0.157	0.173	0.066	17	11	0.090	0.115	0.058
				0.02	22	6	0.371	0.410	0.221	17	11	0.344	0.390	0.219
				0.03	22	6	0.695	0.709	0.587	17	11	0.692	0.700	0.587
				0.04	22	6	0.875	0.866	0.825	17	11	0.874	0.867	0.821
				0.05	22	6	0.952	0.945	0.928	17	11	0.952	0.946	0.923
2	5	28	5	0.00	22	6	0.268	0.111	0.054	17	11	0.131	0.048	0.051
				0.01	22	6	0.286	0.175	0.066	17	11	0.147	0.114	0.059
				0.02	22	6	0.372	0.405	0.214	17	11	0.271	0.385	0.210
				0.03	22	6	0.626	0.699	0.563	17	11	0.590	0.697	0.567
				0.04	22	6	0.840	0.860	0.806	17	11	0.826	0.862	0.804
				0.05	22	6	0.938	0.936	0.917	17	11	0.932	0.940	0.911

of the recursive demeaning that achieves size robustness to asymmetric errors.

Finally, Table 7 reports the results for Case 3. Here, we find that PSY and $sPSY$ have generally similar levels of power to each other, with $\bar{s}PSY$ a little lower, whereas the powers of all tests vary little with the level shift settings. It is apparent that in this case of fixed number and magnitude of level shifts, there is little between the powers of PSY and $sPSY$, in line with there being little in the way of size distortion.

Overall we find the finite sample power of the sign-based tests to display a high degree of robustness to the presence and magnitude of level shifts, in contrast to the PSY test which can display changes in power when level shifts are present in the data. Together with the size results of the previous sub-section, the simulation results show that the $sPSY$ and $\bar{s}PSY$ tests possesses attractive properties relative to PSY when applied to series where level shifts can be prevalent.

6 | Empirical Applications

The data series we consider in our applications are (i) the 5-minute Bitcoin log price and (ii) the daily Nasdaq 100 price. For each application, the PSY, $sPSY$ and $\bar{s}PSY$ tests are calculated with one lagged difference term incorporated in the augmented associated regressions. We compute p-values for each test using simulated null distributions of the tests, with one lagged difference term, for the specific sample size T, using a random walk simulation DGP with innovations $\varepsilon_t \sim IIDN(0,1)$ and no level shifts, cf. the null DGP in our size simulations. These simulations are based on 2000 Monte Carlo replications.

We assess level shift activity using the Lee and Mykland [5] jump detection procedure in two ways. For a given data series, we calculate the number of significant level shifts, which we denote as n, applying the standardised $|\mathcal{L}(i)|$ statistic (see Definition 1 and Lemma 1 of their article), detecting jumps when the standardised statistic exceeds the 0.05-level significance threshold. We also compute a measure of aggregate daily level shift activity as the absolute value of the sum of significant $\mathcal{L}(i)$ statistics, denoted $\left|\sum_n \mathcal{L}(i)\right|$. Specifics of the applications are given below.

TABLE 6 | Finite sample powers of nominal 0.05-level tests: Case 2, $\alpha_n = 0.25$, $\alpha_\mu = 0.25$, $T = 200$, $\tau_1 = 0.3$, $\tau_2 = 0.7$.

					$p = 0.8$					$p = 0.6$				
k	μ	n_T	μ_T	δ	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$
3	1	11	3.761	0.00	9	2	0.118	0.066	0.050	7	4	0.088	0.051	0.053
				0.01	9	2	0.158	0.149	0.071	7	4	0.125	0.135	0.069
				0.02	9	2	0.396	0.425	0.253	7	4	0.376	0.424	0.254
				0.03	9	2	0.729	0.723	0.635	7	4	0.731	0.723	0.634
				0.04	9	2	0.883	0.877	0.847	7	4	0.881	0.879	0.846
				0.05	9	2	0.952	0.951	0.939	7	4	0.955	0.955	0.940
3	2	11	7.521	0.00	9	2	0.259	0.066	0.050	7	4	0.201	0.051	0.053
				0.01	9	2	0.279	0.149	0.071	7	4	0.225	0.135	0.069
				0.02	9	2	0.408	0.425	0.253	7	4	0.358	0.424	0.254
				0.03	9	2	0.685	0.723	0.634	7	4	0.663	0.723	0.633
				0.04	9	2	0.863	0.876	0.846	7	4	0.861	0.879	0.846
				0.05	9	2	0.946	0.951	0.938	7	4	0.948	0.955	0.939
6	1	22	3.761	0.00	18	4	0.216	0.094	0.054	13	9	0.097	0.042	0.047
				0.01	18	4	0.241	0.168	0.069	13	9	0.119	0.117	0.060
				0.02	18	4	0.393	0.404	0.234	13	9	0.315	0.387	0.227
				0.03	18	4	0.696	0.706	0.589	13	9	0.669	0.707	0.587
				0.04	18	4	0.872	0.863	0.832	13	9	0.867	0.868	0.826
				0.05	18	4	0.952	0.942	0.929	13	9	0.951	0.948	0.930
6	2	22	7.521	0.00	18	4	0.394	0.094	0.054	13	9	0.198	0.042	0.047
				0.01	18	4	0.396	0.168	0.069	13	9	0.213	0.117	0.060
				0.02	18	4	0.449	0.404	0.234	13	9	0.289	0.387	0.227
				0.03	18	4	0.636	0.705	0.584	13	9	0.567	0.707	0.585
				0.04	18	4	0.841	0.863	0.824	13	9	0.810	0.867	0.825
				0.05	18	4	0.935	0.941	0.926	13	9	0.920	0.948	0.928

6.1 | Bitcoin Log Price Data

We use intraday Bitcoin log price data for the period September-November 2018. Bitcoin is a digital-only peer-to-peer asset designed to work as a medium of exchange. It uses cryptographic techniques to add verified transactions to a publicly distributed ledger known as the Blockchain, a task that is incentivised by a transaction fee along with newly created bitcoins. (A full breakdown of the technology behind the storage, creation, and transactions involving bitcoins can be found in Böhme et al. [13].) Since its initial introduction in Nakamoto [14], the cryptocurrency has gone on to have a reputation as a speculative asset among economists. We use Bitcoin price data obtained from Coinbase, one of the largest global online exchanges for cryptocurrencies, and the largest exchange that offers a direct pairing between Bitcoin and more standard currencies, rather than cryptocurrency alternatives. The data are obtained directly via Coinbase's Application Programming Interface, allowing direct access to historical price data, rather than a weighting of different exchanges. The close price is sampled at the 5-minute frequency for the Bitcoin-USD trading pair; this sampling frequency then consists of a total of 26,208 observations for the three month period.[3]

Rather than simply apply the tests to the full sample period, giving one single set of results, we make use of this long data set by splitting it into 91 sub-samples, each of which contains observations on one day (i.e., 91 non-overlapping sub-samples, each of $T = 288$ 5-minute observations). The daily series are plotted in Figure 1. On visual inspection, it appears that many of the series display evidence of level shifts, while there appears to be little visual evidence of explosive behaviour; indeed, the date range was chosen due to this being a period where bubble behaviour is largely considered to be absent. Within each sub-sample, we conduct the PSY, $sPSY$ and $\bar{s}PSY$ tests, and also measure the degree of level shift activity. Using the data in this way allows us to evaluate, across multiple samples, whether there is a pattern connecting the level shift activity with the outcomes of the different tests. Our supposition would be that, given the series do not generally appear to be explosive, rejections by any of the tests are likely to be spurious, and attributable to the effects of level shifts.

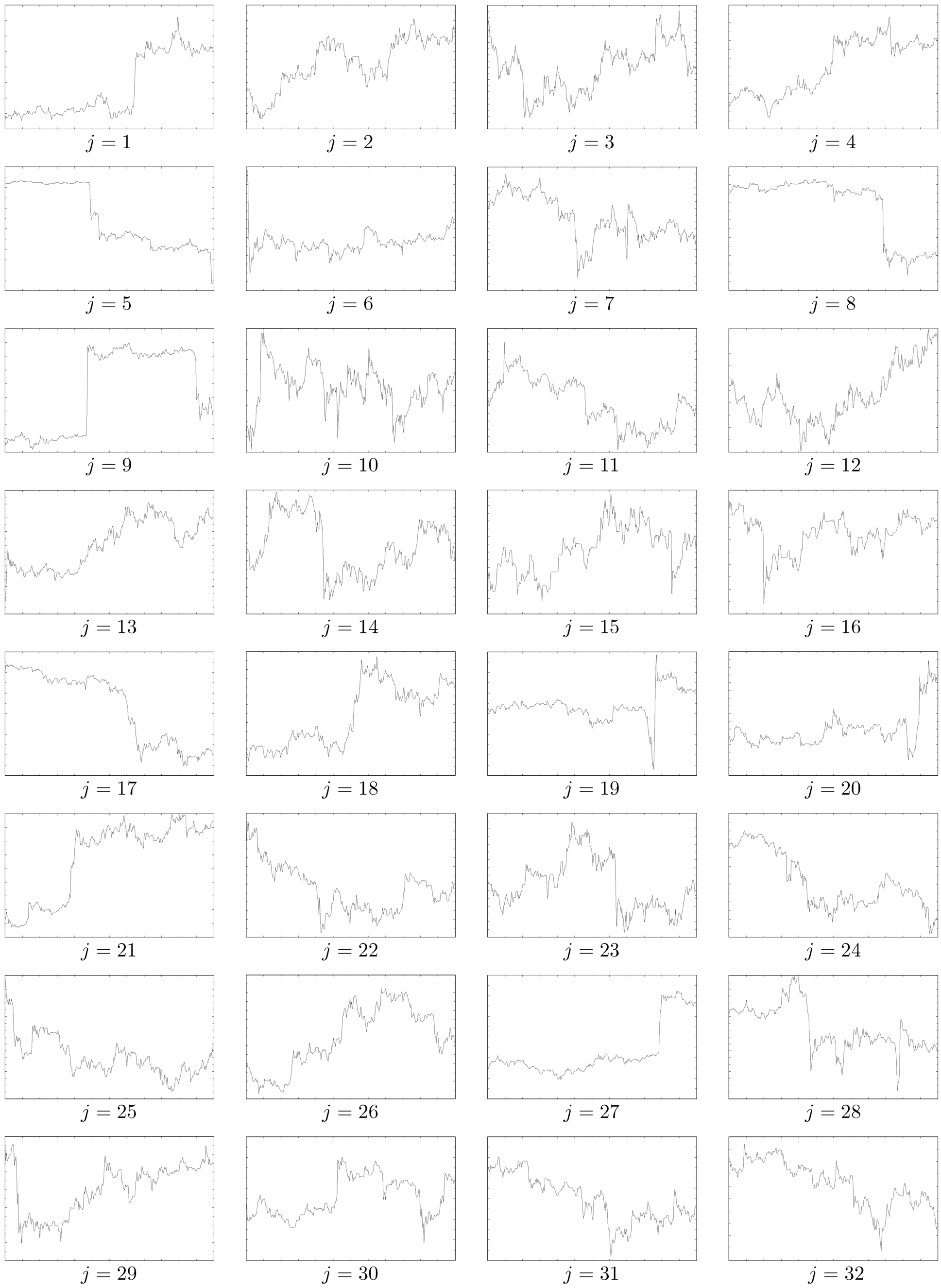
TABLE 7 | Finite sample powers of nominal 0.05-level tests: Case 3, $\alpha_n = 0$, $\alpha_\mu = 0$, $T = 200$, $\tau_1 = 0.3$, $\tau_2 = 0.7$.

					$p = 0.8$					$p = 0.6$				
k	μ	n_T	μ_T	δ	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$	n_T^+	n_T^-	PSY	$sPSY$	$\bar{s}PSY$
5	2.5	5	2.5	0.00	4	1	0.057	0.051	0.057	3	2	0.056	0.052	0.055
				0.01	4	1	0.110	0.149	0.074	3	2	0.103	0.147	0.073
				0.02	4	1	0.424	0.447	0.262	3	2	0.425	0.448	0.259
				0.03	4	1	0.759	0.738	0.642	3	2	0.761	0.738	0.637
				0.04	4	1	0.901	0.885	0.859	3	2	0.900	0.885	0.859
				0.05	4	1	0.963	0.958	0.942	3	2	0.962	0.957	0.942
5	5	5	5	0.00	4	1	0.095	0.051	0.056	3	2	0.089	0.052	0.054
				0.01	4	1	0.140	0.148	0.073	3	2	0.122	0.147	0.072
				0.02	4	1	0.398	0.447	0.260	3	2	0.394	0.448	0.258
				0.03	4	1	0.735	0.738	0.638	3	2	0.735	0.738	0.634
				0.04	4	1	0.890	0.884	0.857	3	2	0.889	0.884	0.858
				0.05	4	1	0.958	0.958	0.942	3	2	0.960	0.956	0.940
10	2.5	10	2.5	0.00	8	2	0.070	0.053	0.054	6	4	0.053	0.050	0.050
				0.01	8	2	0.113	0.143	0.070	6	4	0.102	0.141	0.070
				0.02	8	2	0.422	0.425	0.258	6	4	0.403	0.426	0.250
				0.03	8	2	0.742	0.728	0.636	6	4	0.744	0.729	0.639
				0.04	8	2	0.896	0.878	0.844	6	4	0.892	0.879	0.842
				0.05	8	2	0.962	0.955	0.942	6	4	0.958	0.955	0.942
10	5	10	5	0.00	8	2	0.131	0.054	0.054	6	4	0.101	0.050	0.050
				0.01	8	2	0.162	0.143	0.070	6	4	0.131	0.140	0.069
				0.02	8	2	0.371	0.424	0.256	6	4	0.352	0.425	0.247
				0.03	8	2	0.696	0.727	0.627	6	4	0.699	0.728	0.631
				0.04	8	2	0.883	0.878	0.842	6	4	0.875	0.878	0.841
				0.05	8	2	0.957	0.955	0.941	6	4	0.956	0.954	0.941

For the Dickey-Fuller tests, we set $\pi = 0.125$ such that the minimum sample length is 36 observations, corresponding to a three hour period. For the Lee and Mykland jump test, we use a window width of $K = \lfloor \pi T \rfloor + 2 = 38$ (which equates to 36 observations in the bipower variation formula). Table 8 presents the results for each daily sub-sample $j = 1, \ldots, 91$, with the results listed in descending order of the level shift magnitude measure $\left|\sum_n \mathcal{L}(i)\right|$. The p-values are highlighted using the colour coding of red, orange and yellow for rejections at the 0.01-, 0.05- and 0.10-levels, respectively, and green for non-rejections at the 0.10-level.

We find that PSY rejects (strongly) in just under half of the daily series, whereas $sPSY$ and $\bar{s}PSY$ very rarely reject at conventional significance levels. What is of particular interest is that most of the PSY rejections, especially the strongest rejections, are for series where the aggregate level shift magnitude measure $\left|\sum_n \mathcal{L}(i)\right|$ is greatest. This feature of PSY rejecting in series where there is a large degree of level shift activity, combined with non-rejections from $sPSY$ and $\bar{s}PSY$, suggest that no explosive behaviour is in fact present, and that the PSY rejections are spurious, driven by the effects of the level shifts. This accords precisely with what we would expect from our theoretical and simulation results, and visual inspection of the plots of the series. On the few occasions where $sPSY$ and $\bar{s}PSY$ reject the null, it is also interesting to observe that the number of detected level shifts n is typically large. Recall that while the $sPSY$ and $\bar{s}PSY$ tests do not require a condition on the magnitude of level shifts to be asymptotically correctly sized, the number of level shifts still requires a restriction, hence a large number of level shifts in a finite sample could result in these tests displaying spurious rejections of the null.

The overall picture from this first application is that use of a bubble detection procedure that is not robust to level shifts, such as PSY, runs the risk of spuriously detecting a bubble, when applied to a series without explosive characteristics, but subject to frequent and/or large magnitude level shifts. In contrast, use of the sign-based procedures $sPSY$ and $\bar{s}PSY$ offer a far greater degree of robustness to level shifts, and are considerably less likely to mistake level shift activity in a series for bubble behaviour.

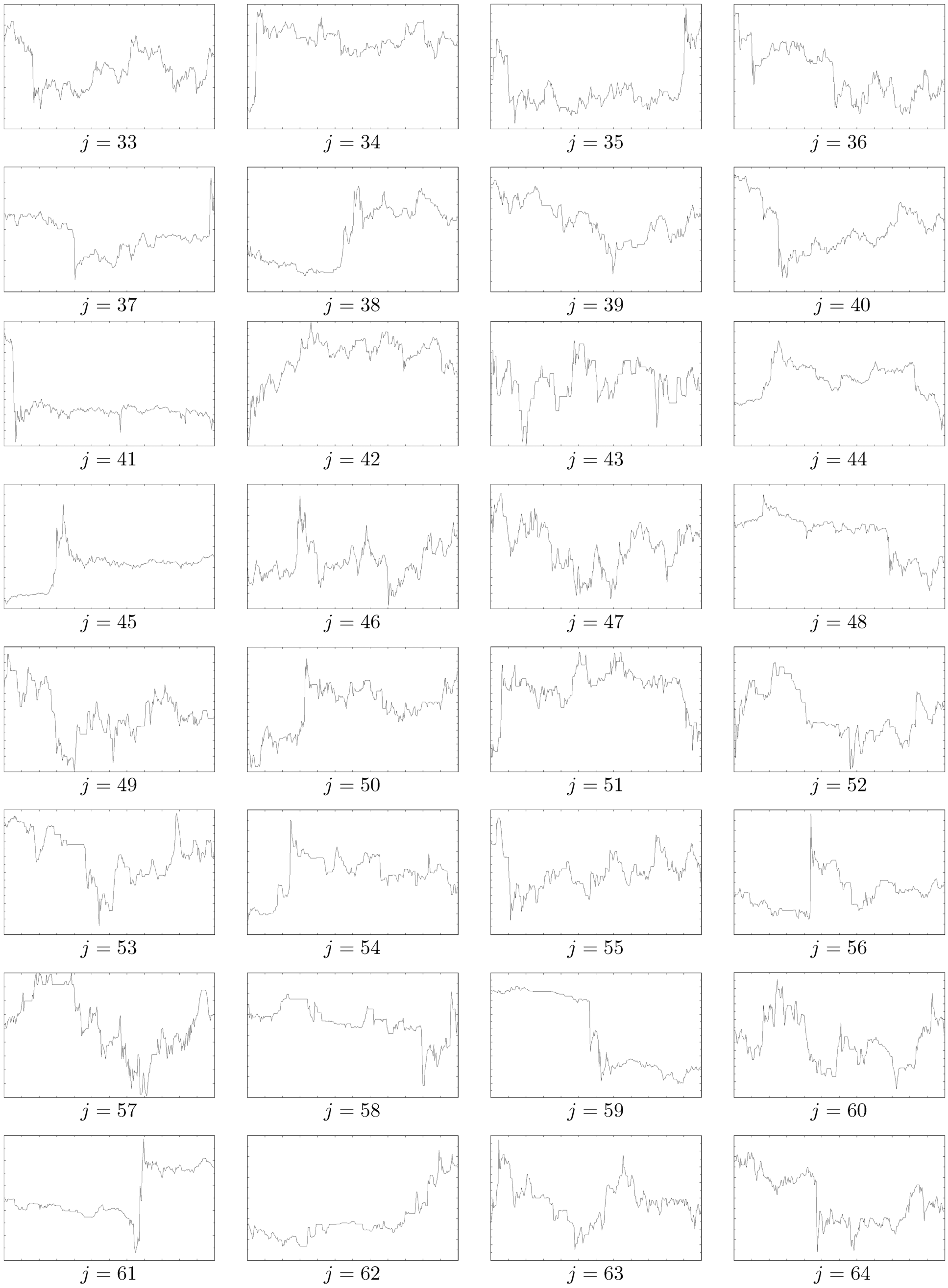


a. Intraday Bitcoin log price, day j within September-November 2018.

FIGURE 1 | Intraday Bitcoin log price, day j within September-November 2018.

6.2 | Nasdaq Price Data

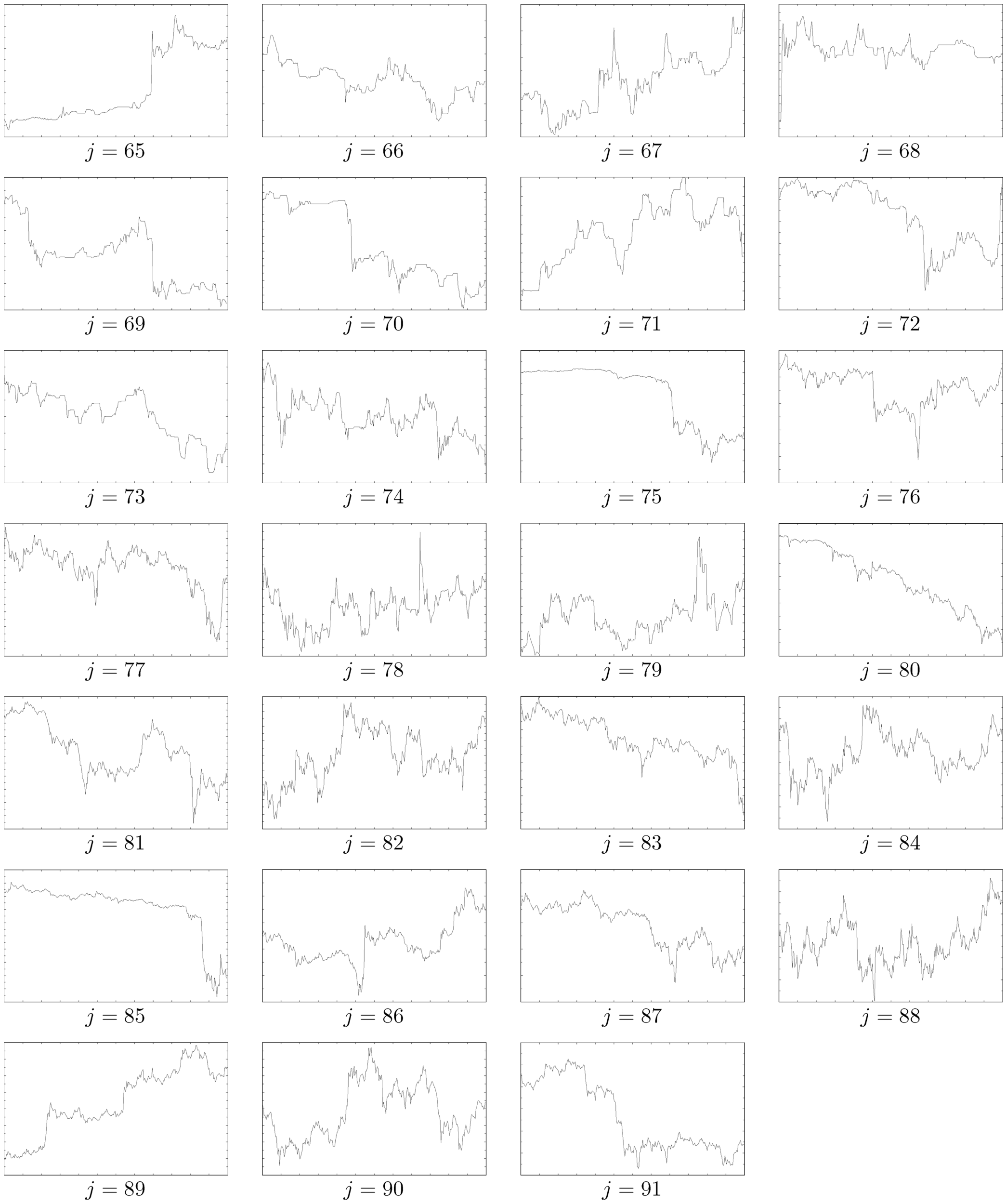
We next consider data for the Nasdaq 100 index daily closing price over the period 3 January 1995 to 31 December 2001, which yields $T = 1763$ observations. The Nasdaq 100 is an index of top 100 non-financial firms listed on the Nasdaq exchange, and is heavily concentrated on technology companies. We chose this period deliberately as it features the growth of the so-called "dot-com bubble" (together with its subsequent collapse). The data were obtained from the Investing.com website. Figure 2 plots this data series. Visually, and in sharp contrast to the Bitcoin series in Figure 1, the price series does indeed appear to contain a bubble regime ending in early March 2000, but, significantly, also appears largely free of any level shift activity. For conformity with the Bitcoin application above, we use $\pi = 0.125$, which gives a window width of $K = \lfloor \pi T \rfloor + 2 = 222$ in the Lee and Mykland test. The results are shown in Table 9. Here, extremely strong rejections of the unit root null are obtained from each of the PSY, $sPSY$ and $\overline{s}PSY$ tests, thereby confirming the presence of the "dotcom bubble". There is also only very modest evidence



b. Intraday Bitcoin log price, day j within September-November 2018.

FIGURE 1 | (Continued)

of level shift activity since the number of jumps and their aggregate measure are small—at least when measured relative to what we observed for the higher frequency Bitcoin log price series in Table 8 (which is what we might expect). We would expect conformity of bubble inference across tests in this case and have little concern here that the rejection by the PSY test is being spuriously induced by level shifts. Finally, we also examined the log of the Nasdaq price series. In terms of the level shift measures, here we find $n = 4$ and $\left|\sum_n \mathcal{L}(i)\right| = 1.943$, with even less jump activity apparent in the log data than its unlogged version. The p-value associated with PSY increases from 0.000 in the unlogged data to 0.222 for the log prices, overturning the rejection of the unit root null at conventional significance levels and thereby failing to find evidence of a bubble. We would not want to associate this sensitivity of inference with being a weakness of the PSY test, but just to note that transforming the data in this way does have consequences for inference. In contradistinction, the p-values associated with $sPSY$ and $\bar{s}PSY$ for the log data remain exactly as reported in Table 9 because the values of the $sPSY$ and $\bar{s}PSY$ statistics are exact invariant to strictly monotonic transformations of the data. We view this as another potentially appealing property of the sign-based tests.



c. Intraday Bitcoin log price, day j within September-November 2018.

FIGURE 1 | (Continued)

TABLE 8 | Test p-values and level shift measures: daily sub-samples of 5-minute Bitcoin log price.

j	PSY	$sPSY$	$\bar{s}PSY$	n	$\lvert\sum_n \mathcal{L}(i)\rvert$	j	PSY	$sPSY$	$\bar{s}PSY$	n	$\lvert\sum_n \mathcal{L}(i)\rvert$
59	0.000	0.017	0.266	14	262.55	51	0.095	0.806	0.685	7	17.37
70	0.002	0.705	0.062	10	207.19	16	0.639	0.914	0.986	4	16.52
5	0.156	0.420	0.926	12	112.05	42	0.784	0.281	0.880	2	16.42
65	0.003	0.230	0.211	9	79.44	69	0.779	0.705	0.456	15	15.81
8	0.059	0.106	0.475	8	76.26	48	0.003	0.622	0.939	8	14.64
45	0.000	0.504	0.276	8	67.50	10	0.919	0.788	0.602	4	13.87
75	0.000	0.491	0.225	12	62.97	23	0.863	0.619	0.598	4	13.66
85	0.000	0.351	0.381	8	59.20	29	0.723	0.596	0.730	2	13.23
67	0.000	0.151	0.871	7	55.10	36	0.009	0.120	0.232	6	12.96
27	0.003	0.808	0.812	2	51.13	46	0.000	0.829	0.907	5	12.28
52	0.123	0.382	0.575	12	50.82	44	0.011	0.199	0.158	9	12.27
9	0.011	0.501	0.729	6	50.50	80	0.012	0.317	0.996	4	12.01
54	0.082	0.372	0.462	9	50.45	82	0.626	0.956	0.944	2	11.91
58	0.003	0.121	0.232	14	44.02	87	0.000	0.830	0.659	4	11.61
38	0.000	0.580	0.017	7	43.47	28	0.000	0.913	0.801	7	11.22
64	0.683	0.699	0.462	5	41.90	74	0.981	0.645	0.728	2	10.85
56	0.921	0.719	0.927	5	41.46	71	0.274	0.078	0.460	21	10.57
37	0.015	0.667	0.705	8	35.66	62	0.315	0.703	0.093	15	10.13
86	0.000	0.498	0.968	7	34.50	33	0.359	0.276	0.475	6	9.81
76	0.030	0.829	0.262	5	33.75	18	0.081	0.671	0.704	3	8.78
21	0.000	0.073	0.324	5	32.03	31	0.843	0.566	0.716	7	8.16
11	0.740	0.920	0.934	3	31.66	6	0.614	0.792	0.664	6	7.82
49	0.325	0.835	0.248	7	31.31	12	0.810	0.600	1.000	3	7.76
83	0.403	0.601	0.904	5	30.91	43	0.018	0.532	0.351	16	7.56
63	0.610	0.618	0.589	4	28.84	15	0.742	0.624	0.517	6	7.24
57	0.379	0.071	0.767	4	26.51	34	0.990	0.222	0.871	7	6.83
7	0.078	0.928	0.313	6	26.12	91	0.109	0.160	0.949	1	6.08
89	0.000	0.312	0.955	3	25.55	41	0.600	0.883	0.668	5	5.64
24	0.536	0.756	0.967	7	25.23	88	0.876	0.391	0.921	3	5.59
30	0.015	0.964	0.978	6	25.03	26	0.036	0.520	0.317	5	5.52
2	0.611	0.992	0.760	6	24.94	77	0.327	0.602	0.820	3	5.51
35	0.000	0.811	0.704	3	24.69	47	0.596	0.324	0.474	3	5.22
14	0.136	0.902	0.980	5	23.59	79	0.653	0.895	0.911	6	4.51
55	0.890	0.237	0.746	4	23.52	90	0.150	0.306	0.232	3	4.45
17	0.000	0.413	0.288	6	20.88	3	0.150	0.974	0.944	2	3.57
60	0.000	0.392	0.943	9	20.44	50	0.072	0.395	0.695	6	3.45
73	0.461	0.132	0.265	14	20.20	19	0.003	0.157	0.843	10	3.14
81	0.003	0.302	0.800	4	19.44	66	0.296	0.400	0.890	8	2.23
13	0.184	0.880	0.802	7	19.01	20	0.362	0.928	0.882	4	1.84
1	0.107	0.795	0.896	9	18.98	4	0.049	0.243	0.184	8	1.78
40	0.155	0.156	0.019	6	18.52	72	0.022	0.396	0.438	6	0.99
53	0.011	0.842	0.152	7	18.09	39	0.187	0.045	0.147	10	0.90
68	0.958	0.399	0.104	6	18.07	84	0.191	0.902	0.998	2	0.87
61	0.005	0.124	0.557	15	17.87	25	0.347	0.298	0.293	4	0.63
32	0.592	0.497	0.482	5	17.78	22	0.016	0.365	0.728	6	0.25
78	0.939	0.668	0.914	3	17.68						

Note: Cells highlighted in red, orange and yellow correspond to rejections at the 0.01-, 0.05- and 0.10-levels, respectively. Cells highlighted in green correspond to non-rejections at the 0.10-level.

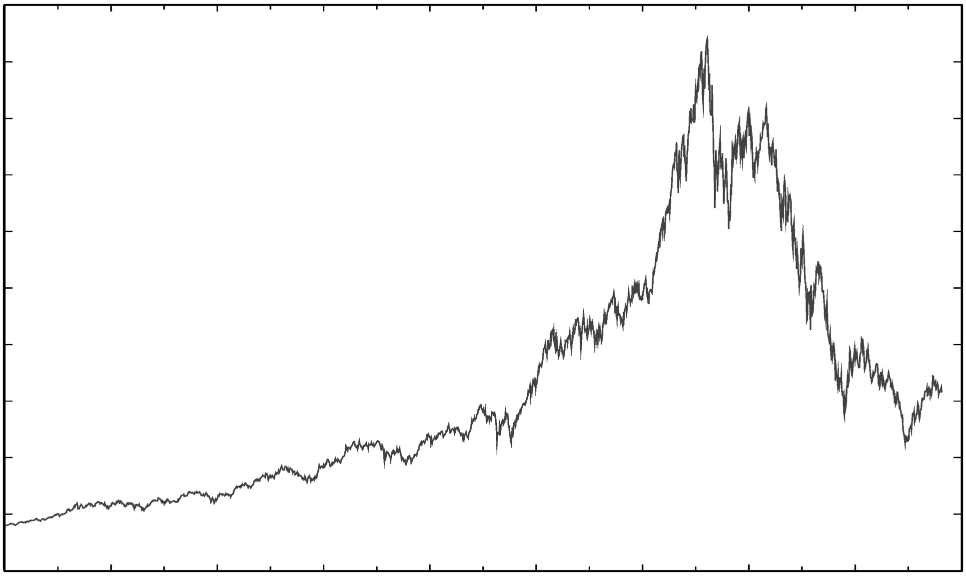


FIGURE 2 | Daily Nasdaq 100 price, 3 January 1995 to 31 December 2001.

TABLE 9 | Test p-values and level shift measures: daily Nasdaq 100 price.

PSY	$sPSY$	$\bar{s}PSY$	n	$\lvert \sum_n \mathcal{L}(i) \rvert$
0.000	0.000	0.004	4	14.702

Note: Cells highlighted in red correspond to rejections at the 0.01-level.

7 | Conclusion

In this article, we have analysed the impact of the number and magnitude of deterministic level shifts on the PSY test for explosive autoregressive behaviour and its sign-based variants. We find that the two sign-based tests offer a natural advantage over the original PSY test in the presence of level shifts of unknown magnitude. While the sign PSY tests retain their asymptotic validity under a restriction only on the *number* of level shifts, validity of the PSY test requires a joint restriction involving both the *number* and *magnitudes* of the level shifts. Our Monte Carlo simulations demonstrate that the PSY test can be badly over-sized in the presence of level shifts, while the sign-based variants offer far superior size control in such circumstances. Moreover, the sign-based tests remain competitive in terms of power with the PSY test, despite the latter's over-size when level shifts are present. The sign-based tests therefore offer a more robust and powerful method for detecting an explosive autoregressive regime in a financial time series that potentially contains level shifts. We applied the tests to daily sub-samples of recent intraday Bitcoin log price data, focusing on a period where explosive behaviour did not seem to be apparent. We found that level shift activity was prevalent in these samples, and in samples where this activity was particularly evident, the PSY test frequently appeared to (falsely) indicate the presence of explosive episodes, in contrast to the sign-based tests for which rejections were rarely found. Our second application was to Nasdaq data across the "dotcom bubble" period, a series for which explosive behaviour is apparent and level shift activity is found to be low. Here, all three tests provide similar inference, with strong rejections of the null confirming the presence of explosive behaviour.

Acknowledgements

We are very grateful to two anonymous referees for valuable comments on an earlier version of this article.

Endnotes

[1] Assumption 2 also implies the median of ε_t is zero, in addition to the zero mean assumption from Assumption 1; the imposed distributional assumption on ε_t is only slightly weaker than assuming the distribution of ε_t is symmetric about zero.

[2] Note that similar results are obtained on replacing p with $1 - p$, hence we only consider values of $p > 0.5$.

[3] Across this period, eleven observations were not recorded in the dataset; in these cases, we linearly interpolated neighbouring observations to proxy for these missing values.

References

1. P. C. Phillips, S. Shi, and J. Yu, "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500," *International Economic Review* 56 (2015): 1043–1078.

2. D. E. Rapach, J. K. Strauss, and M. E. Wohar, "Forecasting Stock Return Volatility in the Presence of Structural Breaks," in *Forecasting in the Presence of Structural Breaks and Model Uncertainty* (Emerald, 2008).

3. D. I. Harvey, S. J. Leybourne, and Y. Zu, "Sign-Based Unit Root Tests for Explosive Financial Bubbles in the Presence of Deterministically Time-Varying Volatility," *Econometric Theory* 36 (2020): 122–169.

4. B. Eraker, "Do Stock Prices and Volatility Jump? Reconciling Evidence From Spot and Option Prices," *Journal of Finance* 59 (2004): 1367–1403.

5. S. S. Lee and P. A. Mykland, "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," *Review of Financial Studies* 21 (2008): 2535–2563.

6. S. S. Lee, "Jumps and Information Flow in Financial Markets," *Review of Financial Studies* 25 (2012): 439–479.

7. P. Bajgrowicz, O. Scaillet, and A. Treccani, "Jumps in High-Frequency Data: Spurious Detections, Dynamics, and News," *Management Science* 62 (2016): 2198–2217.

8. S. Laurent and S. Shi, "Unit Root Test With High-Frequency Data," *Econometric Theory* 38 (2022): 113–171.

9. P. Perron, "Testing for a Unit Root in a Time Series With a Changing Mean," *Journal of Business and Economic Statistics* 8 (1990): 153–162.

10. T. J. Vogelsang and P. Perron, "Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time," *International Economic Review* 39 (1998): 1073–1100.

11. J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods* (Springer, 2005).

12. E. Paparoditis and D. N. Politis, "The Asymptotic Size and Power of the Augmented Dickey-Fuller Test for a Unit Root," *Econometric Reviews* 37 (2018): 955–973.

13. R. Böhme, N. Christin, B. Edelman, and T. Moore, "Bitcoin: Economics, Technology, and Governance," *Journal of Economic Perspectives* 29 (2015): 213–238.

14. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008, https://nakamotoinstitute.org/static/docs/bitcoin.pdf.

Appendix A

Proofs of Theorems

The PSY, $sPSY$ and $\bar{s}PSY$ statistics are invariant to μ_0 so in what follows we can set $\mu_0 = 0$ without loss of generality.

Proof of Theorem 1. We first note that a little manipulation shows that $DF(\lambda_1, \lambda_2)$ can be written in the form

$$DF(\lambda_1,\lambda_2) = \frac{T^{-1}(y_{\lfloor\lambda_2 T\rfloor}-\bar{y})^2 - T^{-1}(y_{\lfloor\lambda_1 T\rfloor}-\bar{y})^2 - T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(\Delta y_t)^2}{2\sqrt{\hat{\sigma}^2(\lambda_1,\lambda_2)T^{-2}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(y_{t-1}-\bar{y})^2}}$$

We now examine the behaviour of $y_{\lfloor rT\rfloor}$:

$$T^{-1/2}y_{\lfloor rT\rfloor} = T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\varepsilon_t + T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}}\mu_{i,T} \tag{A1}$$

Considering the second term of (A1),

$$T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}}\mu_{i,T} = \begin{cases} o(1) & 0 \le \alpha_{n,\mu} < 1/2 \\ O(1) & \alpha_{n,\mu} = 1/2 \end{cases}$$

So,

$$\begin{aligned} T^{-1/2}y_{\lfloor rT\rfloor} &\Rightarrow \begin{cases} \sigma W(r) & 0 \le \alpha_{n,\mu} < 1/2 \\ \sigma W(r) + J(r) & \alpha_{n,\mu} = 1/2 \end{cases} \\ &= \sigma H(r) \end{aligned} \tag{A2}$$

where $J(r) = \lim_{T\to\infty}(T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}}\mu_{i,T})$. Applying (A2) we therefore find, via the CMT, that

$$T^{-1}(y_{\lfloor\lambda_2 T\rfloor}-\bar{y})^2 \Rightarrow \sigma^2\overline{H}(\lambda_2)^2 \tag{A3}$$

$$T^{-1}(y_{\lfloor\lambda_1 T\rfloor}-\bar{y})^2 \Rightarrow \sigma^2\overline{H}(\lambda_1)^2 \tag{A4}$$

$$T^{-2}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(y_{t-1}-\bar{y})^2 \Rightarrow \sigma^2\int_{\lambda_1}^{\lambda_2}\overline{H}(r)^2dr \tag{A5}$$

where

$$\overline{H}(r) = H(r) - (\lambda_2-\lambda_1)^{-1}\int_{\lambda_1}^{\lambda_2}H(s)ds$$

Next, using $\Delta y_t = \varepsilon_t + \mu_{i,T}\mathbb{I}(t_i = t)$, we have

$$\begin{aligned} T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(\Delta y_t)^2 &= T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\varepsilon_t^2 + T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\mu_{i,T}^2\mathbb{I}(t=t_i) \\ &\quad + 2T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\varepsilon_t\mu_{i,T}\mathbb{I}(t=t_i) \end{aligned} \tag{A6}$$

Regarding the second term of (A6),

$$\begin{aligned} T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\mu_{i,T}^2\mathbb{I}(t=t_i) &\le T^{-1}\sum_{i=1}^{n_T}\mu_{i,T}^2 \\ &= o(1) \end{aligned}$$

by virtue of Assumption 3(ii).

The third term of (A6) is $o_p(1)$ since it has zero mean, while

$$\begin{aligned} Var\left[T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\varepsilon_t\mu_{i,T}\mathbb{I}(t=t_i)\right] &= \sigma^2T^{-2}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\mu_{i,T}^2\mathbb{I}(t=t_i) \\ &= o(1) \end{aligned}$$

Hence,

$$T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(\Delta y_t)^2 = T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}\varepsilon_t^2 + o(1) + o_p(1) \xrightarrow{p} (\lambda_2-\lambda_1)\sigma^2 \tag{A7}$$

Lastly, it is easily shown that

$$\hat{\sigma}^2(\lambda_1,\lambda_2) = (\lfloor\lambda_2 T\rfloor - \lfloor\lambda_1 T\rfloor + 1)^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(\Delta y_t)^2 + o_p(1) \xrightarrow{p} \sigma^2 \tag{A8}$$

in view of (A7).

Combining (A3), (A4), (A5), (A7) and (A8), we obtain

$$\begin{aligned} DF(\lambda_1,\lambda_2) &\Rightarrow \frac{\overline{H}(\lambda_2)^2 - \overline{H}(\lambda_1)^2 - (\lambda_2-\lambda_1)}{2\sqrt{\int_{\lambda_1}^{\lambda_2}\overline{H}(r)^2dr}} \\ &= L(\lambda_1,\lambda_2) \end{aligned}$$

and the result of Theorem 1 then follows.

Proof of Theorem 2. We can write $sDF(\lambda_1, \lambda_2)$ in the form

$$sDF(\lambda_1,\lambda_2) = \frac{T^{-1}C_{\lfloor\lambda_2 T\rfloor}^2 - T^{-1}C_{\lfloor\lambda_1 T\rfloor}^2 - T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}(\Delta C_t)^2}{2\sqrt{\hat{s}^2(\lambda_1,\lambda_2)T^{-2}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor}C_{t-1}^2}}$$

Next,

$$\begin{aligned} T^{-1/2}C_{\lfloor rT\rfloor} &= T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}sign(\Delta y_t) \\ &= T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}sign(\varepsilon_t + \mu_{i,T}\mathbb{I}(t_i=t)) \end{aligned}$$

Now,

$$\begin{aligned}&T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor} sign(\varepsilon_t+\mu_{i,T}\mathbb{I}(t_i=t))\\&=T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} sign(\varepsilon_t)+T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i}+\mu_{i,T})\\&-T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i})\end{aligned}$$

Here,

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} sign(\varepsilon_t)\Rightarrow W_1^s(r)$$

while

$$\begin{aligned}&\left|T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i}+\mu_{i,T})-T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i})\right|\\&\leq 2T^{-1/2}n_{\lfloor rT\rfloor}\\&=\begin{cases}o(1) & 0\leq\alpha_n<1/2\\ O(1) & \alpha_n=1/2\end{cases}\end{aligned}\tag{A9}$$

Now we can write

$$\begin{aligned}&\lim_{T\to\infty}(T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i}+\mu_{i,T})-T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i}))\\&=\begin{cases}0 & 0\leq\alpha_n<1/2\\ K^s(r) & \alpha_n=1/2\end{cases}\end{aligned}\tag{A10}$$

with $|K^s(r)|$ bounded by $2\lim_{T\to\infty}T^{-1/2}n_{\lfloor rT\rfloor}=2\kappa_r$ from (A9). Hence,

$$\begin{aligned}T^{-1/2}C_{\lfloor rT\rfloor}&\Rightarrow\begin{cases}W_1^s(r) & 0\leq\alpha_n<1/2\\ W_1^s(r)+K^s(r) & \alpha_n=1/2\end{cases}\\&=H^s(r)\end{aligned}\tag{A11}$$

Using (A11), the CMT then shows that

$$T^{-1}C^2_{\lfloor\lambda_2T\rfloor}\Rightarrow H^s(\lambda_2)^2\tag{A12}$$

$$T^{-1}C^2_{\lfloor\lambda_1T\rfloor}\Rightarrow H^s(\lambda_1)^2\tag{A13}$$

$$T^{-2}\sum_{t=\lfloor\lambda_1T\rfloor}^{\lfloor\lambda_2T\rfloor}C^2_{t-1}\Rightarrow\int_{\lambda_1}^{\lambda_2}H^s(r)^2dr\tag{A14}$$

Next,

$$\begin{aligned}T^{-1}\sum_{t=\lfloor\lambda_1T\rfloor}^{\lfloor\lambda_2T\rfloor}(\Delta C_t)^2&=T^{-1}\sum_{t=\lfloor\lambda_1T\rfloor}^{\lfloor\lambda_2T\rfloor}\{sign(\Delta y_t)\}^2\\&=T^{-1}(\lfloor\lambda_2T\rfloor-\lfloor\lambda_1T\rfloor+1)\to\lambda_2-\lambda_1\end{aligned}\tag{A15}$$

It is also easily shown that

$$\hat{s}^2(\lambda_1,\lambda_2)=(\lfloor\lambda_2T\rfloor-\lfloor\lambda_1T\rfloor)^{-1}\sum_{t=\lfloor\lambda_1T\rfloor}^{\lfloor\lambda_2T\rfloor}\{sign(\Delta y_t)\}^2+o_p(1)\xrightarrow{p}1\tag{A16}$$

in view of (A15).

Taken together, (A12), (A13), (A14), (A15) and (A16), we find

$$DF(\lambda_1,\lambda_2)\Rightarrow\frac{H^s(\lambda_2)^2-H^s(\lambda_1)^2-(\lambda_2-\lambda_1)}{2\sqrt{\int_{\lambda_1}^{\lambda_2}H^s(r)^2dr}}$$

leading to the result of Theorem 2.

Proof of Theorem 3. The proof follows along the same lines as that of Theorem 2. Here,

$$\begin{aligned}T^{-1/2}C_{\lfloor rT\rfloor}&=T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\left\{sign(\Delta y_t)-(t-1)^{-1}\sum_{j=2}^{t}sign(\Delta y_j)\right\}\\&=T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\left\{sign(\varepsilon_t+\mu_{i,T}\mathbb{I}(t_i=t))\right.\\&\left.-(t-1)^{-1}\sum_{j=2}^{t}sign(\varepsilon_j+\mu_{i,T}\mathbb{I}(t_i=j))\right\}\\&=T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\left\{sign(\varepsilon_t+\mu_{i,T}\mathbb{I}(t_i=t))-E[sign(\varepsilon_t)]\right\}\\&-T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\left\{(t-1)^{-1}\sum_{j=2}^{t}\{sign(\varepsilon_j+\mu_{i,T}\mathbb{I}(t_i=j))\right.\\&\left.-E[sign(\varepsilon_j)]\}\right\}\end{aligned}$$

Now,

$$\begin{aligned}&T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\left\{sign(\varepsilon_t+\mu_{i,T}\mathbb{I}(t_i=t))-E[sign(\varepsilon_t)]\right\}\\&=T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\left\{sign(\varepsilon_t)-E[sign(\varepsilon_t)]\right\}\\&+T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i}+\mu_{i,T})-T^{-1/2}\sum_{i=1}^{n_{\lfloor rT\rfloor}} sign(\varepsilon_{t_i})\end{aligned}$$

Here,

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\left\{sign(\varepsilon_t)-E[sign(\varepsilon_t)]\right\}\Rightarrow\sigma_sW_2^s(r)$$

and, using (A10), we can then write

$$\begin{aligned}&T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\{sign(\varepsilon_t+\mu_{i,T}\mathbb{I}(t_i=t))-E[sign(\varepsilon_t)]\}\\&\Rightarrow\sigma_s\begin{cases}W_2^s(r) & 0\leq\alpha_n<1/2\\ W_2^s(r)+\sigma_s^{-1}K^s(r) & \alpha_n=1/2\end{cases}\\&=\sigma_sG^s(r)\end{aligned}$$

Similarly,

$$\begin{aligned}&T^{-1/2}\sum_{t=2}^{\lfloor rT\rfloor}\left\{(t-1)^{-1}\sum_{j=2}^{t}\{sign(\varepsilon_j+\mu_{i,T}\mathbb{I}(t_i=j))-E[sign(\varepsilon_j)]\}\right\}\\&\Rightarrow\sigma_s\int_0^r x^{-1}G^s(x)dx\end{aligned}$$

and so

$$\begin{aligned}T^{-1/2}C_{\lfloor rT\rfloor}&\Rightarrow\sigma_s\left\{G^s(r)-\int_0^r x^{-1}G^s(x)dx\right\}\\&=\sigma_s\overline{G}^s(r)\end{aligned}$$

The CMT then shows that

$$T^{-1}C^2_{\lfloor\lambda_2 T\rfloor} \Rightarrow \sigma_s^2 \overline{G}^s(\lambda_2)^2,$$
$$T^{-1}C^2_{\lfloor\lambda_1 T\rfloor} \Rightarrow \sigma_s^2 \overline{G}^s(\lambda_1)^2,$$
$$T^{-2}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor} C^2_{t-1} \Rightarrow \sigma_s^2 \int_{\lambda_1}^{\lambda_2} \overline{G}^s(r)^2 dr$$

Next,

$$\begin{aligned}
&T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor} (\Delta C_t)^2 = \\
&T^{-1}\sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor} \left\{ sign(\Delta y_t) - (t-1)^{-1}\sum_{j=2}^{t} sign(\Delta y_j) \right\}^2 \\
&= T^{-1}(\lfloor\lambda_2 T\rfloor - \lfloor\lambda_1 T\rfloor + 1) \\
&\times \Big[(\lfloor\lambda_2 T\rfloor - \lfloor\lambda_1 T\rfloor + 1)^{-1} \\
&\times \sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor} \left\{ sign(\Delta y_t) - (t-1)^{-1}\sum_{j=2}^{t} sign(\Delta y_j) \right\}^2 \Big] \\
&\xrightarrow{p} (\lambda_2 - \lambda_1)\sigma_s^2
\end{aligned}$$

It is also easily shown that

$$\begin{aligned}
&\hat{s}^2(\lambda_1, \lambda_2) = (\lfloor\lambda_2 T\rfloor - \lfloor\lambda_1 T\rfloor)^{-1} \\
&\quad\times \sum_{t=\lfloor\lambda_1 T\rfloor}^{\lfloor\lambda_2 T\rfloor} \left\{ sign(\Delta y_t) - (t-1)^{-1}\sum_{j=2}^{t} sign(\Delta y_j) \right\}^2 + o_p(1) \\
&\quad\xrightarrow{p} \sigma_s^2
\end{aligned}$$

Collecting results we find

$$\overline{s}DF(\lambda_1, \lambda_2) \Rightarrow \frac{\overline{G}^s(\lambda_2)^2 - \overline{G}^s(\lambda_1)^2 - (\lambda_2 - \lambda_1)}{2\sqrt{\int_{\lambda_1}^{\lambda_2} \overline{G}^s(r)^2 dr}}$$

leading to the result of Theorem 3.