ON THE ASYMPTOTIC BEHAVIOR OF BUBBLE DATE ESTIMATORS

EIJI KUROZUMIa AND ANTON SKROBOTOVb,c

aDepartment of Economics, Hitotsubashi University, Tokyo, Japan
bRussian Presidential Academy of National Economy and Public Administration, Moscow, Russia
cSaint Petersburg University (Center for Econometrics and Business Analytics), Moscow, Russia

In this study, we extend the three-regime bubble model of Pang et al. (2021, Journal of Econometrics, 221(1):227–311) to allow the forth regime followed by the unit root process after recovery. We provide the asymptotic and finite sample justification of the consistency of the collapse date estimator in the two-regime AR(1) model. The consistency allows us to split the sample before and after the date of collapse and to consider the estimation of the date of exuberation and date of recovery separately. We have also found that the limiting behavior of the recovery date varies depending on the extent of explosiveness and recovering.

Received 11 May 2022; Revised 5 November 2022; Accepted 7 November 2022

Keywords: change points; explosive autoregression; mildly explosive; mildly integrated; rational bubble; right-tailed unit root testing; time-varying volatility.

JEL. C12; C22.

MOS subject classification: 62F10; 62F12.

1. INTRODUCTION

The estimation of the dates of regime changes has a long history in the econometric literature (see, e.g. Casini and Perron, 2019 for a recent review). In this study, we address the issue of estimating the break dates in explosive bubble model. Such a bubble model is often generated (see Phillips et al., 2011) as a unit root process followed by explosive process, which in turn followed by a unit root regime (with possible stationary recovering regime after the collapse of the bubble). For estimation of bubble dates, Phillips et al. (2015a, 2015b) (PSY hereinafter) proposed the recursive algorithm based on the right-tailed ADF test. Generally speaking, the origination of the bubble is taken at the date for which the test statistic begins to exceed the critical value, and the date of collapse is taken at the date for which the test statistic subsequently falls below the critical value (see Skrobotov, 2021 for review).

However, as demonstrated in Harvey et al. (2017), the bubble dates estimates are more accurate if they are obtained by minimizing the sum of the squared residuals over all possible dates. For the model with only two regimes, Chong (2001) investigated consistency and the limiting distribution of the break date estimator for the AR(1) model with a break in the coefficient. The break date estimator is based on minimizing the sum of the squared residuals (SSR) over all possible break dates. For the model with only two regimes, Chong (2001) considered three cases, both regimes are stationary, the first regime is a unit root regime, and the second regime is stationary and vice versa. Pang et al. (2018) extended the result of Chong (2001), allowing the AR(1) coefficient to be dependent on the sample size. They also considered the case with change in persistence from a unit root to moderately explosive and vice versa. Pang et al. (2021) (PDC hereinafter) considered the model with three regimes that reflect the explosive financial bubble model: the first regime is a unit root regime, followed by an explosive regime and then by a stationary collapsing regime. PDC investigated the sample splitting strategy:
The first break date is estimated for the two regime AR(1) model. Results indicate that the break date estimator is consistent for the date of collapse (i.e., the second break). Then, one could consider the sample before the date of collapse and investigate the asymptotic behavior of the date of origination of the bubble – the results are the same as that in Pang et al. (2018). From a different perspective, the estimation procedure used by PDC is the same as that in Chong (1995) and Bai (1997), in which multiple breaks are estimated one at a time in the regime-wise stationary model with level shifts.

In this study, we extend the results of PDC by allowing the forth unit root regime. The advantage of this extension is that we can investigate not only the emerging and collapsing dates of the bubble but also the recovering date to the normal market. As a result, we can totally investigate the abnormal market behavior. In our procedure, we first prove the consistency of the collapse date estimator by minimizing the sum of the squared residuals using the two-regime model, allowing nonstationary volatility. Second, as the estimated break date is consistent with the collapsing date, one could split the sample at the estimated break date and consider the estimator of the innovation during each bubble episode. Subsequently, one can use our approach in the second step.

The remainder of this paper is organized as follows. Section 2 formulates the model and assumptions. In Section 3, we define the main procedure and provide the limiting behavior of the break date estimators. The model with serially correlated shocks is considered in Section 4. The finite sample performance of the estimated break dates is demonstrated in Section 5, and the empirical example is given in Section 6. Section 7 concludes the paper.

All proofs are collected in the Supporting Information.

2. MODEL AND ASSUMPTIONS

Let us consider the following bubble’s emerging and collapsing model for $t = 1, 2, \ldots, T$:

$$
y_t = \begin{cases} 
  c_0 T^{-\eta_0} + y_{t-1} + \epsilon_t & : 1 \leq t \leq k_c, \\
  \phi_a y_{t-1} + \epsilon_t & : k_c + 1 \leq t \leq k_e, \\
  \phi_b y_{t-1} + \epsilon_t & : k_e + 1 \leq t \leq k_i, \\
  c_1 T^{-\eta_1} + y_{t-1} + \epsilon_t & : k_i + 1 \leq t \leq T,
\end{cases}
$$

(1)

where $y_0 = \alpha_0(T^{1/2})$, $c_0 \geq 0$, $\eta_0 > 1/2$, $\phi_a := 1 + c_a/T^a$ with $c_a > 0$ and $0 < a < 1$, $\phi_b := 1 - c_b/T^b$ with $c_b > 0$ and $0 < b < 1$, $c_1 \geq 0$, and $\eta_1 > 1/2$. In this model, the process evolves up to time $k_c$ as a unit root process with a possibly positive drift shrinking to zero, and then $y_t$ behaves mildly explosive and next starts collapsing at $t = k_c + 1$. Thereafter, the adjustment (collapsing and recovering) period lasts up to $k_e$, and then returns to a unit root process. We denote the mildly explosive regime with $\phi_a := 1 + c_a/T^a$ and the mildly stationary regime with $\phi_b := 1 - c_b/T^b$ for ease of exposition. However, we can also consider these AR(1) parameters in more general forms, such as $\phi_a := 1 + c_a/h_T$ and $\phi_b := 1 - c_b/k_T$, where $h_T + T/h_T \to \infty$ and $k_T + T/k_T \to \infty$. The break

---

1 The asymptotics for the date of the beginning of the explosive regime were considered in PDC.
Assumption 3. The following stronger assumption: when \( \sigma \) and strictly positive function on \([0, 1]\) satisfying \( \omega < \sigma(\cdot) < \overline{\omega} < \infty \).

Assumption 2 is general enough to allow for nonstationary unconditional volatility.

For model (1), we make the following assumption.

Assumption 1. \( 0 < \tau_e < \tau_c < \tau_r < 1 \).

Assumption 2. \( e_i := \sigma \varepsilon_i \), where \( \{e_i\} \sim i.i.d. (0, 1) \) with \( E[e_i^4] < \infty \) and \( \sigma_i := \omega(t/T) \) where \( \omega(\cdot) \) is a nonstochastic and strictly positive function on \([0, 1]\) satisfying \( \omega < \sigma(\cdot) < \overline{\omega} < \infty \).

Assumption 1 implies that break dates are distinct and not quite close each other, which is typically assumed in the existing literature. Assumption 2 is general enough to allow for nonstationary unconditional volatility.

Following Cavaliere and Taylor (2007a, 2007b), the functional central limit theorem (FCLT) holds for the partial sum process of \( \{e_i\} \) under Assumption 2:

\[
\frac{1}{\sqrt{T}} \sum_{i=1}^{[rT]} e_i \Rightarrow \tilde{\omega} W(\kappa(t)) =: W^\kappa(t),
\]

for \( 0 \leq r \leq 1 \), where \( \Rightarrow \) signifies weak convergence of associated probability measures, \( W(\cdot) \) is a standard Brownian motion, \( \kappa(t) := \int_0^t \omega(s)ds / \int_0^t \omega^2(s)ds \) is called the variance profile, and \( \tilde{\omega}^2 := \int_0^1 \omega^2(s)ds \). Note that, when \( \sigma_r \) is constant, \( W^\kappa(t) \) reduces to a constant volatility times a standard Brownian motion.

Assumption 2 is used mainly for the estimation of the collapsing date \( k_c \) and the market recovering date \( k_r \) for \( a < b \). For the estimation of the emerging date of a bubble \( k_e \) and the last break point \( k_r \) for \( a > b \), we impose the following stronger assumption:

Assumption 3. \( \{e_i\} \sim i.i.d. (0, \sigma^2) \) with \( E[e_i^4] < \infty \).

Under Assumption 3, the FCLT in (2) becomes simpler and is given by

\[
\frac{1}{\sqrt{T}} \sum_{i=1}^{[rT]} e_i \Rightarrow \sigma W(t).
\]

This result is used for deriving the limiting distributions of the estimators of \( k_e \) and \( k_r \) for \( a > b \) in Theorems 2(iii) and 3(ii), respectively.

3. INDIVIDUAL ESTIMATION OF BREAK DATES

Following PDC, we estimate the model with a one-time break. Note that model (1) can be expressed as

\[
y_t = \begin{cases} 
\phi_0 y_{t-1} + u_t & \text{if } k_0 = 1 \\
\phi_0 y_{t-1} + u_t & \text{if } k_1 = 1 - c_a/T^{\phi} \\
\phi_0 y_{t-1} + u_t & \text{if } k_2 = 1 \\
\phi_1 y_{t-1} + u_t & \text{if } k_3 = 1 - c_b/T^{\phi} \\
\phi_1 y_{t-1} + u_t & \text{if } k_4 = 1 
\end{cases}
\]

where \( c_a/T^{\phi} + \varepsilon_t \), \( \varepsilon_t \), \( \varepsilon_t \), \( c_1/T^{\phi} + \varepsilon_t \).

For a given \( 1 \leq k \leq T - 1 \), we denote the sum of the squared residuals as

\[
SSR(k/T) := \sum_{i=1}^k \left( y_i - \hat{\phi}_a(k/T)y_{i-1} \right)^2 + \sum_{t=k+1}^T \left( y_t - \hat{\phi}_b(k/T)y_{t-1} \right)^2.
\]
where $\sum_{t=0}^{m}$ is abbreviated just as $\sum_{t}$ and

$$\hat{\phi}_a(k/T) = \frac{\sum_{t=1}^{k} y_{t-1} y_t}{\sum_{t=1}^{k} y_{t-1}^2}$$

and

$$\hat{\phi}_b(k/T) = \frac{\sum_{t=k+1}^{T} y_{t-1} y_t}{\sum_{t=k+1}^{T} y_{t-1}^2}.$$

Defining the break point estimator as $\hat{k} := \arg \min_k SSR(k/T)$, we have the following theorem.

**Theorem 1.** Suppose that Assumptions 1 and 2 hold. Then, we have $\lim_{T \to \infty} P(\hat{k} = k_c) = 1$.

Theorem 1 is the same as Theorem 3(a) in PDC and implies that, even if the last regime with a unit root process is additionally included in the sample period, we can consistently estimate the bubble collapsing date. This may not be a surprising result because the explosive behavior in the second regime is quite different from those in the other regimes. Therefore, even if we estimate only a one time break date, the estimator is consistent with the collapsing date of the bubble. We also note that we allow nonstationary volatility in $\{\epsilon_t\}$, general enough to apply our method to practical analyses. It might be possible to estimate the break dates by minimizing the weighted sum of the squared residuals using the volatility estimated by Harvey et al. (2022), but this is beyond the scope of this paper.

The consistency of the collapsing date is intuitively explained as follows. As given in the proof of Lemma 2, during the emergence of the bubble in the second regime, the process can be expressed as

$$y_k = \phi_{a}^{(k-k_c)} y_{k_c} + \sum_{j=k_c+1}^{k} \phi_{a}^{(k-j)} u_j \sim_a \phi_{a}^{(k-k_c)} y_{k_c}$$

for $k_c + 1 \leq k \leq k_c$, where $y_{k_c}/\sqrt{T}$ weakly converges to a (time-transformed) Brownian motion. Therefore, the process evolves monotonically at a geometric rate (noting that $\phi_{a} > 1$) and attains its peak at $k = k_c$. Conversely, the process collapses at a geometrically decaying rate (noting that $\phi_{b} < 1$) because it can be expressed as

$$y_k = \phi_{b}^{(k-k_c)} y_{k_c} + \sum_{j=k_c+1}^{k} \phi_{b}^{(k-j)} u_j,$$

for $k_c + 1 \leq k \leq k_c$, where the first term on the right hand side can be shown to dominate the second term, at least, when $k$ is relatively close to $k_c$. These geometric explosive and collapsing behaviors help us identify the collapsing date $k_c$ consistently.

After we obtained the consistent estimator $\hat{k}$ of $k_c$, we decompose the whole sample into two sub-samples for $t = 1, \ldots, \hat{k}$ and $\hat{k} + 1, \ldots, T$ and estimate $k_c$ and $k$ from the first and second sub-samples, respectively. Because $\hat{k}$ is a consistent estimator of $k_c$, we can treat $k_c$ as a known break point. Thus, we use $k_c$, instead of $\hat{k}$, in the following.

Estimation of $k_c$ is based on the minimization of the sum of the squared residuals using the first sub-sample, and the estimator is defined as

$$\hat{k}_c := \arg \min_{\sum_{k/T} \leq \tau_c} SSR_1(k/T)$$

where $0 < \tau < \tau_c < \tau_c < \tau_c$ and

$$SSR_1(k/T) := \sum_{t=1}^{k} \left( y_t - \hat{\phi}_c(k/T) y_{t-1} \right)^2 + \sum_{t=k+1}^{T} \left( y_t - \hat{\phi}_c(k/T) y_{t-1} \right)^2.$$
The corresponding break fraction estimator is defined by \( \hat{\tau}_e := \hat{k}_e / T \).

The consistency of the \( \hat{\tau}_e \) has been already established in Theorem 1.3 of Pang et al. (2018) and stated in the next theorem.

**Theorem 2.** [Pang et al. (2018)] Suppose that Assumptions 1 and 3 hold. Then,

(i) when \( a < 1/2 \), \( \lim_{T \to \infty} P(\hat{k}_e = k_e) = 1 \).
(ii) when \( a = 1/2 \), \( |\hat{k}_e - k_e| = O_p(1) \).
(iii) when \( a > 1/2 \),

\[
(1 - \phi_b)^2 T^2 (\hat{\tau}_e - \tau_e) \Rightarrow \arg \max_{v \in R} \left\{ \frac{W^*(v)}{W(\tau_e)} - \frac{|v|}{2} \right\},
\]

where \( W^*(v) \) is a two-sided Brownian motion on \( R \) defined to be \( W^*(v) = W_t(-v) \) for \( v \leq 0 \) and \( W^*(v) = W_t(v) \) for \( v > 0 \) with \( W_t(\cdot) \) and \( W_t(\cdot) \) being two independent Brownian motion on \( R^* \).

On the other hand, for the estimation of \( k_e \), we minimize the sum of the squared residuals using the second subsample, and the estimator is defined as

\[
\hat{k}_e := \arg \min_{\sum_{t \leq T} \tau_e < \tau_e < T < \tau_e < 1} \sum_{T} \left( y_t - \hat{\phi}_e(k/T) y_{t-1} \right)^2
\]

where \( \tau_e < \tau_e < \tau_e < \tau_e < 1 \) and

\[
\sum_{T} \sum_{k+1}^T \left( y_t - \hat{\phi}_e(k/T) y_{t-1} \right)^2
\]

with \( \hat{\phi}_e(k/T) = \frac{\sum_{k+1}^T y_{t-1} y_t}{\sum_{k+1}^T y_{t-1}^2} \) and \( \hat{\phi}_e(k/T) = \frac{\sum_{k+1}^T y_{t-1} y_t}{\sum_{k+1}^T y_{t-1}^2} \)

The corresponding break fraction estimator is defined by \( \hat{\tau}_e := \hat{k}_e / T \).

To investigate the asymptotic property of \( \hat{k}_e \), we must distinguish the two cases: \( a < b \) and \( a > b \). In the case of \( a < b \), the explosive behavior is faster than the collapsing (recovering) speed, and we continue to make Assumption 2. On the other hand, to derive the limiting distribution, we need to impose stronger restrictions on the shocks given by Assumption 3 in the case of \( a > b \), in which the process recovers relatively faster than the evolution of the bubble.

The following theorem reveals that the convergence order of \( \hat{\tau}_e \) depends on whether \( a < b \) or \( a > b \).

**Theorem 3.** Suppose that Assumptions 1 and 3 hold. (i) When \( a < b \),

\[
\lim_{T \to \infty} P(\hat{k}_e = k_e) = 1.
\]

(ii) When \( a > b \),

\[
\left( 1 - \phi_b \right) T (\hat{\tau}_e - \tau_e) \Rightarrow \arg \max_{v \in R} \left\{ C^*_v (\hat{\tau}_e) - \frac{|v|}{2} \right\}
\]
where \( C_{c,v}(v) := 2 \int_{0}^{[v]} B_{v}(s)dB_{v}(s) - c_{v} \int_{0}^{[v]} \left( B_{v}^{2}(s) - 1/(2c_{v}) \right) ds \) for \( v < 0 \) and \( C_{c,v}(v) := -\left\{ B_{v}(v) + \int_{0}^{[v]} B_{v}(s)dB_{v}(s)/B_{v}(0) + c_{v} \int_{0}^{[v]} (B_{v}(s)/(2B_{v}(0)) + 1) B_{v}(s)ds \right\} / (c_{v}B_{v}(0)) \) for \( v \geq 0 \) with \( B_{v}(s) := \int_{s}^{\infty} \exp (-c_{v}(t-s)) dB_{1}(t) \) for \( s \geq 0 \) and \( B_{1}(\cdot) \) and \( B_{2}(\cdot) \) being two independent standard Brownian motions on \( \mathbb{R}^{*} \).

**Remark 1.** Evidently, Theorems 2 and 3 hold under Assumptions 1 and 2, except for the derivation of the limiting distributions. That is, the same convergence orders of the break date estimators will be obtained with independent and heteroskedastic errors. We made Assumption 3 to derive the limiting distributions in Theorems 2(iii) and 3(ii), which are the same as those obtained in the existing literature.

From Theorem 3, we can observe that the break date from the collapsing (recovering) regime to the normal market can be consistently estimated if \( a < b; \) that is, the explosive speed of the process is faster than the collapsing speed. In this case, as given in (6), the process decays at geometric rate for \( k_{c} + 1 \leq k \leq k_{r} \) (because the first term of (6) dominates the second one) and reaches at

\[
y_{k} = \phi_{b}^{(k-k_{c})} y_{k_{c}} + \sum_{j=k_{c}+1}^{k} \phi_{b}^{(k-j)} u_{j}
\]

\[\sim_{a} \phi_{a}^{(k-k_{c})} \phi_{b}^{(k-k_{c})} y_{k_{c}}.\]

Note that \( \phi_{a}^{(k-k_{c})} \phi_{b}^{(k-k_{c})} \to \infty \) when the explosive speed is faster \( (a < b) \) because

\[
\log \left( \phi_{a}^{(k-k_{c})} \phi_{b}^{(k-k_{c})} \right) = (r_{c} - r_{e}) T \log \left( 1 + \frac{c_{a}}{T^{\alpha}} \right) + (r_{e} - r_{c}) T \log \left( 1 - \frac{c_{b}}{T^{\beta}} \right)
\]

\[= (r_{c} - r_{e}) c_{a} T^{1-\alpha} (1 + o(1)) - (r_{e} - r_{c}) c_{b} T^{1-\beta} (1 + o(1)) \]

\[\to \infty\]

because \( a < b \). That is, the explosive effect induced by \( \phi_{a} \) remains in the process during the whole collapsing (recovering) regime, after which the process evolves as

\[
y_{k} = y_{k_{c}} + c_{i} \frac{k - k_{c}}{T^{\alpha}} + \sum_{j=k_{c}+1}^{k} \epsilon_{j} \sim_{a} y_{k_{c}}
\]

because we can observe from (7) that \( T^{-a} \phi_{a}^{(k-k_{c})} \phi_{b}^{(k-k_{c})} \to \infty \) for any \( \alpha > 0 \). This monotonically geometrical decay during the collapsing regime followed by the asymptotically flat behavior makes \( \hat{k}_{c} \) consistent.

On the other hand, when the collapsing speed is faster \( (a > b) \), the effect of the explosive component in \( y_{k} \) in the recovering regime diminishes to zero as \( k \) approaches \( k_{c} \) because \( \phi_{a}^{(k-k_{c})} \phi_{b}^{(k-k_{c})} \to 0 \) when \( a > b \), as can be observed in (7). That is, the initial value effect by \( y_{k_{c}} \) in the recovering regime gradually disappears and the process behaves as if it started from the small initial value. Therefore, we obtain the same result as Theorem 1 of Pang et al. (2018).

The asymptotically different convergence orders in Theorem 3 reflect the finite sample performance of \( \hat{k}_{c} \), as presented in Section 5.

**Remark 2.** It is possible to consider the case where \( a = b \). As is seen from the expansion given by (7), the case where \( (r_{c} - r_{e}) c_{a} > (r_{e} - r_{c}) c_{b} \) will have the same result as the case of \( a < b \) and the reverse relation corresponds to the case of \( a > b \). We can also consider the knife edge case in which \( (r_{c} - r_{e}) c_{a} = (r_{e} - r_{c}) c_{b} \). In this case, the asymptotic behavior of \( \phi_{a}^{(k-k_{c})} \phi_{b}^{(k-k_{c})} \) depends on the higher order expansion of the logarithms as well as the values of \( a = b \). Because this knife edge case will make our analysis complicated, we do not pursue this case and concentrate on the two cases where \( a < b \) or \( a > b \) to clearly state our main results.
4. SERIALLY CORRELATED CASE

In this section, we extend model (1) with \( \{ \varepsilon_t \} \) being serially correlated. We consider the case where \( \{ \varepsilon_t \} \) is a linear process given by

\[
\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{where} \quad \sum_{j=0}^{\infty} |\psi_j| < \infty \tag{8}
\]

For the innovations \( \{ \varepsilon_t \} \), we make the following assumption.

**Assumption 4.** \( \{ \varepsilon_t \} \sim \text{i.i.d.}(0, \sigma^2) \) with \( E|\varepsilon_t|^\gamma < \infty \) for some \( \gamma \) where \( \gamma > \max(4, 2/a, 2/b) \).

Linear process (8) can be expressed as, by the BN decomposition,

\[
\varepsilon_t = \psi \varepsilon_{t-1} - \bar{\varepsilon}_t \quad \text{where} \quad \bar{\varepsilon}_t := \sum_{\ell=0}^{\infty} \psi_{\ell-t} \varepsilon_{t-\ell} \quad \text{with} \quad \psi := \sum_{k=0}^{\infty} \psi_k \tag{9}
\]

and \( \psi = \sum_{j=0}^{\infty} |\psi_j| \). It is also well known that the summability condition in (8) implies \( \sum_{\ell=0}^{\infty} |\psi_{\ell-t}| < \infty \). In this case, we have the following corollary.

**Corollary 1.** Suppose that we estimate \( k_a, k_r, \) and \( c_b \) as in the previous section and that Assumptions 1 and 4 hold. Then, we obtain the same results as given in Theorems 1, 2, and 3(i), while Theorem 3(ii) holds with the limiting distribution replaced by

\[
\frac{1 - \phi_{b}}{c_{b}} T(\hat{\tau}_r - \tau_r) = T^{1-k}(\hat{\tau}_r - \tau_r) \Rightarrow \arg \max_{v \in \mathbb{R}} \left\{ C^*(v) - \frac{|v|}{2} \psi^* \right\},
\]

where \( \psi^* \) depends on whether \( \nu < 0 \) or \( \nu \geq 0 \) and is defined in Appendix C.

By Corollary 1, we can estimate the three break dates by using the misspecified AR(1) model with the same convergence orders of the estimators as before, although we may expect that the finite sample accuracy of the estimators would be improved by correctly specifying the correlation structure by considering the ADF type regression with an additional lag structure.

5. MONTE-CARLO SIMULATIONS

We examine the performance of the estimates of the bubble regimes dates in finite samples. Our purpose is to demonstrate how the finite sample properties of the estimated break dates reflect the asymptotic results obtained in the previous section.

The Monte-Carlo simulations reported in this section are based on the series generated by (1) with \( y_0 = 0 \) and \( \{ \varepsilon_t \} \sim \text{iIDN}(0, 1) \). Data were generated from this DGP for samples of \( T = (400, 800) \) with 50,000 replications.\(^2\)

We set the drift terms in the first and fourth regimes to \( c_a T^{-90} = 1/800 \) and \( c_b T^{-90} = 1/800 \), respectively, following PDC. In this experiment, we focus on two cases. In the first case, the explosive coefficient \( \phi_a \) changes with a fixed \( \phi_b \); \( \phi_b \) is set to 0.96 while \( \phi_a \) takes values among \( \{1.01, 1.05, 1.09\} \). In the second case, the collapsing coefficient \( \phi_b \) varies with a fixed \( \phi_a \); \( \phi_a \) is set to 1.05 while \( \phi_b \) varies among \( \{0.98, 0.96, 0.94\} \). We set the localizing parameters \( c_a = c_b = 1 \); then, the values of \( a \) and \( b \) are uniquely determined based on the definitions given by \( \phi_a = 1 + 1/T^a \) and \( \phi_b = 1 - 1/T^b \). For the dates of bubble regimes, we use \( (\tau_a, \tau_r, \tau_b) \) to be equal to \( \{0.4, 0.6, 0.7\} \).\(^3\)

---

\(^2\) All simulations were programmed in R with mnorm random number generator.

\(^3\) To compare the estimator of \( k_a \) with that of PDC, we also set \( (\tau_a, \tau_r, \tau_b) \) to be equal to \( \{0.4, 0.6, 1\} \), that is, without the last unit root regime (the normal market behavior after the collapse and recovering). The results are entirely similar and omitted for brevity.
Furthermore, in the minimization of $SSR(k/T)$, $SSR_1(k/T)$, and $SSR_2(k/T)$, we excluded the first and last 5% observations from the permissible break date $k$. For example, when estimating $k_c$ based on $SSR_2(k/T)$, the permissible break date $k$ ranges from $k_c + 0.05T + 1$ to $0.95T$. If the break date estimate $\hat{k}_c$ exceeds $0.95T$, then we cannot estimate $k_c$; we do not include such a case in any bins of the histogram and thus the sum of the heights of the bins does not necessarily equal one for $\hat{k}_c$ in some cases. Similarly, we cannot estimate $\hat{k}_c$ when $\hat{k}_c < 0.05T$.

In the following, we investigate only $\hat{k}_c$ and $\hat{k}_d$ because, once $\hat{k}_c$ is obtained, the estimate of $k_c$ in the four regime model is the same as that in the three regime model, which has been already investigated by PDC.

Figure 1 presents the histograms of $\hat{k}_c$ when $\phi_a = 0.96$ and $\phi_b = (1.01, 1.05, 1.09)$. As expected, when $\phi_a$ becomes larger and/or the sample size increases, the $\hat{k}_c$ becomes more accurate. When $a > b (|\phi_a - 1| < |\phi_b - 1|)$ in our setting) as in Figure 1(a) and (b), the frequency of selecting the true break date is not very high (it is around 30% for $T = 400$ and 65% for $T = 800$), but it increases to almost 100% in Figure 1(c)–(f).

In the second case (Figure 2) where $\phi_a = 1.05$ and $\phi_b$ varies among (0.98, 0.96, 0.94), the estimate $\hat{k}_c$ is quite accurate for all the three cases with the almost 100% frequency of selecting the true break date.

Figure 3 demonstrates the histograms of $\hat{k}_c$ for the fixed value of $\phi_a = 0.96$ and the different values of $\phi_b$. For the small value of $\phi_a = 1.01$, as shown in Figure 3(a) and (b), which corresponds to the case of $a > b$, we can observe that the accuracy of $\hat{k}_c$ deteriorates, while for $a < b (\phi_a = 1.05$ and $\phi_a = 1.09)$ the estimate becomes more accurate. In particular, for the large sample size of $T = 800$, the frequency of selecting the true break date is very close to 100%.

In the case where $\phi_a$ is fixed at 1.05 and $\phi_b$ varies, we can observe the good performance of the estimate $\hat{k}_c$, even for the case of $a > b (\phi_a = 1.05, \phi_b = 0.94)$ as shown in Figure 4(e) and (f). Reflecting the result in Theorem 3, the estimate $\hat{k}_c$ is most accurate when $\phi_a = 1.05$ and $\phi_b = 0.98$; the frequency of selecting the true break date is approximately 75% and 100% for $T = 400$ and 800, respectively.

We next consider the case where the explosive and collapsing regimes are shorter such as $(\sigma, \tau_c, \tau_r, \tau_e) = (0.5, 0.55, 0.6)$, which is relatively close to NASDAQ data in empirical application. Figures are collected in the Supporting Information and, as expected, the finite sample properties deteriorate in most case, although they improve as $T$ gets larger and/or the explosive speed gets relatively faster (Figures D.1–D.4).

We also experiment with shorter minimum separation periods; we exclude the first and last 1% samples from the permissible break date $k$. The results provided in the online appendix demonstrate that there is virtually no significant difference between two trimming parameters by comparing Figures 1–4 with Figures D.5–D.8.

Finally, we investigate the finite sample behavior of the break date estimates when the volatility process $\sigma_t$ follows the single shift model, $\omega(s) = \sigma_0 + (\sigma_1 - \sigma_0)I(s > \tau_a)$, where $\tau_a = 0.5$ and $\sigma_1/\sigma_0 \in [1/3, 3]$. Note that $\sigma_1/\sigma_0 = 1$ corresponds to the case of constant unconditional volatility.

For $\sigma_1/\sigma_0 = 1/3$, with which the volatility in the first half of the sample is higher than in the later sample, $\hat{k}_c$ and $\hat{k}_c$ perform similarly to the case of the constant unconditional volatility; the frequency of correctly selecting the true collapsing date $\hat{k}_c$ is relatively high, whereas the performance of $\hat{k}_c$ is not satisfactory when $\phi_a = 1.01$ and $\phi_b = 0.96$ (Figures D.9–D.12).

In the case where the low volatility regime shifts to the high volatility one at the middle of the sample ($\sigma_1/\sigma_0 = 3$), it becomes more difficult to select the true collapsing and recovering dates than in the case of $\sigma_1/\sigma_0 = 1/3$, in particular, when $\phi_a = 1.01$ and $\phi_b = 0.96$. The possible reason is that both $\hat{k}_c$ and $\hat{k}_c$ are in the volatile regime when $\sigma_1/\sigma_0 = 3$ and thus it would be difficult to distinguish between the shifts in the parameters and the large shocks. However, the performance of the estimates improves for the larger values of $\phi_a$ and/or as $T$ gets larger (Figures D.13–D.16).

6. EMPIRICAL APPLICATION

In this section, we demonstrate the application of the bubble dates estimation method for the four regime model investigated in the previous sections to two time series data sets. It is worth noticing that even though we may detect emergence and collapse of a bubble by some tests, we are not sure whether the final observation available for estimation is included in the collapsing regime or the unit root regime. Therefore, before estimating the break
Figure 1. Histograms of $\hat{k}_c$ for $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$

(a) $T = 400, \phi_a = 1.01, \phi_b = 0.96$
(b) $T = 800, \phi_a = 1.01, \phi_b = 0.96$
(c) $T = 400, \phi_a = 1.05, \phi_b = 0.96$
(d) $T = 800, \phi_a = 1.05, \phi_b = 0.96$
(e) $T = 400, \phi_a = 1.09, \phi_b = 0.96$
(f) $T = 800, \phi_a = 1.09, \phi_b = 0.96$
Figure 2. Histograms of $\hat{\kappa}_c$ for $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$

(a) $T = 400, \phi_a = 1.05, \phi_b = 0.98$

(b) $T = 800, \phi_a = 1.05, \phi_b = 0.98$

(c) $T = 400, \phi_a = 1.05, \phi_b = 0.96$

(d) $T = 800, \phi_a = 1.05, \phi_b = 0.96$

(e) $T = 400, \phi_a = 1.05, \phi_b = 0.94$

(f) $T = 800, \phi_a = 1.05, \phi_b = 0.94$
Figure 3. Histograms of $\hat{k}_r$ for $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$

(a) $T = 400$, $\phi_a = 1.01$, $\phi_b = 0.96$

(b) $T = 800$, $\phi_a = 1.01$, $\phi_b = 0.96$

(c) $T = 400$, $\phi_a = 1.05$, $\phi_b = 0.96$

(d) $T = 800$, $\phi_a = 1.05$, $\phi_b = 0.96$

(e) $T = 400$, $\phi_a = 1.09$, $\phi_b = 0.96$

(f) $T = 800$, $\phi_a = 1.09$, $\phi_b = 0.96$
Figure 4. Histograms of $\hat{k}_r$ for $(\tau_e, \tau_c, \tau_r) = (0.4, 0.6, 0.7)$

(a) $T = 400$, $\phi_a = 1.05$, $\phi_b = 0.98$
(b) $T = 800$, $\phi_a = 1.05$, $\phi_b = 0.98$

(c) $T = 400$, $\phi_a = 1.05$, $\phi_b = 0.96$
(d) $T = 800$, $\phi_a = 1.05$, $\phi_b = 0.96$

(e) $T = 400$, $\phi_a = 1.05$, $\phi_b = 0.94$
(f) $T = 800$, $\phi_a = 1.05$, $\phi_b = 0.94$
Behavior of Bubble Date Estimators

Figure 5. The bubble origination, collapse and recovery dates of the NASDAQ composite index

Figure 6. The bubble origination, collapse and recovery dates of U.S. house price index

dates, we need to determine whether the three or four regime model is appropriate. Following Harvey et al. (2017), we implement the BIC with the penalty term given by the number of estimated parameters plus the number of breaks to determine the model. In the following, we define by $BIC_2$ the BIC obtained from the model without the collapsing regime; by $BIC_3$ the BIC obtained from the model without the fourth regime; by $BIC_4$ the BIC obtained from the full four-regime model.

The first example is the close prices of monthly data from January 1985 to August 2013 of NASDAQ Composite Index. First, we select the model using the BICs; $BIC_2$, $BIC_3$, and $BIC_4$ equals 3409.296, 3404.268, and 3387.727, respectively. Thus, the four regime model has the minimum BIC and is selected. As explained in the previous section, we fit the AR(1) model with two regimes and the estimated break date corresponds to the collapsing date, which is February 2000, as depicted in Figure 5. Next, by splitting the whole sample at the estimated collapsing date, the date of origination of the bubble is detected at August 1998 from the first subsample, while the date of recovery is estimated at September 2001 from the second subsample. As is observed in Figure 5, our method can detect the explosive and collapsing behavior very well.

The second example is an application of our method to the logarithm of the U.S. house price index from January 1991 to December 2012, provided by the Federal Housing Finance Agency, adjusted by the consumer price index.

---

4 The data was downloaded from https://finance.yahoo.com

5 The house price index and the CPI are available from https://www.fhfa.gov/DataTools/Downloads/pages/house-price-index.aspx and https://fred.stlouisfed.org/
Again, because the BIC selects the four regime model \((BIC_2, BIC_3, \text{ and } BIC_4)\) equal \(-2809.591, -2895.905, \text{ and } -2918.223\), respectively, we first estimate the collapsing date, which is estimated at November 2006. As is seen in Figure 6, the explosive behavior becomes mild around early in 2006 but the collapsing date is estimated just before the series staring crashing. By splitting the whole sample at the estimated collapsing date, the emergence date of the explosive behavior is estimated at September 1997, while the recovering date is at May 2011. These estimated dates are consistent with the visual inspection and the proposed method works very well.

7. CONCLUDING REMARKS

In this paper, we considered the four regime bubble model and investigated the break date estimators using the sample splitting approach; the break dates are estimated one at a time. We showed that the break date estimated initially is consistent with the collapsing date of the bubble. We used the second subsample after the first estimated break date to estimate the break date returning to the normal market. The results revealed that this estimator is consistent with the recovering date if the explosive speed of the process during the bubble period is faster than the collapsing (recovering) speed, whereas in the case where the explosive speed is relatively slower, only the corresponding break fraction estimator is consistent. We need further investigation when the two speeds are the same.

Although we considered the case where the bubble occurred only once, we can extend our approach to the multiple bubble model in conjunction with the approach used by Harvey et al. (2020). In this case, we first identify the bubble regimes based on the PSY approach and then implement our method to each estimated regime.

ACKNOWLEDGEMENTS

We are grateful to two anonymous referees for helpful comments. All errors are our responsibility. We also thank the participants at the Center for Econometrics and Business Analytics (CEBA, St. Petersburg State University) seminar series and the transdisciplinary econometrics and data science seminar (TEDS) for helpful discussions and comments. Anton Skrobotov’ research for this paper was supported in part by a grant from the Russian Science Foundation (Project No. 20-78-10113). Eiji Kurozumi’s research was supported by JSPS KAKENHI Grant Number 19K01585 and 22K01422.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Bai J. 1997. Estimating multiple breaks one at a time. *Econometric Theory* **13**(3):315–352.
Casini A, Perron P. 2019. *Structural Breaks in Time Series*. Oxford Research Encyclopedia of Economics and Finance Oxford University Press, Oxford.
Cavaliere G, Taylor AMR. 2007a. Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics* **140**(2):919–947.
Cavaliere G, Taylor AMR. 2007b. Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis* **29**(2):300–330.
Chong TT-L. 1995. Partial parameter consistency in a misspecified structural change model. *Economics Letters* **49**(4):351–357.
Chong TT-L. 2001. Structural change in AR(1) models. *Econometric Theory* **17**(1):87–155.
Harvey D, Leybourne S, Sollis R. 2017. Improving the accuracy of asset price bubble start and end date estimators. *Journal of Empirical Finance* **40**:121–138.
BEHAVIOR OF BUBBLE DATE ESTIMATORS

Harvey D, Leybourne S, Whitehouse E. 2020. Date-stamping multiple bubble regimes. *Journal of Empirical Finance* **58**:226–246.

Harvey DI, Leybourne SI, Zu Y. 2022. Estimation of the variance function in structural break autoregressive models with non-stationary and explosive segments. *Journal of Time Series Analysis*. https://doi.org/10.1111/jtsa.12660.

Pang T, Chong TT-L, Zhang D, Liang Y. 2018. Structural change in nonstationary AR(1) models. *Econometric Theory* **34**(5):985–1017.

Pang T, Du L, Chong TT-L. 2021. Estimating multiple breaks in nonstationary autoregressive models. *Journal of Econometrics* **221**(1):277–311.

Phillips PCB, Shi S, Yu J. 2015a. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* **56**(4):1043–1078.

Phillips PCB, Shi S, Yu J. 2015b. Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review* **56**(4):1079–1134.

Phillips PCB, Wu Y, Yu J. 2011. Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review* **52**(1):201–226.

Skrobotov, A. (2021). Testing for explosive bubbles: a review. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3779111.