Probabilistic Forecasting of Bubbles and Flash Crashes.*

Anurag Banerjee
Durham Business School

Guillaume Chevillon
ESSEC Business School †

Marie Kratz
ESSEC Business School, CREAR

April 19, 2018

Abstract

We propose a near explosive random coefficient autoregressive model (NERC) to obtain predictive probabilities of the apparition and devolution of bubbles. The distribution of the autoregressive coefficient of this model is allowed to be centered at an $O(T^{-\alpha})$ distance of unity, with $\alpha \in (0, 1)$. When the expectation of the autoregressive coefficient lies on the explosive side of unity, the NERC helps to model the temporary explosiveness of various asset prices. We study the asymptotic properties of the NERC and provide a procedure for inference on the parameters. In empirical illustrations, we obtain predictive probabilities of bubbles or flash crashes in financial asset prices.

Keywords: Random Coefficient Autoregressive Model, Local Asymptotics, Predictive Density, Bubbles, Flash Crashes.
JEL Codes: C22, C53, C58, G12.

*The authors would like to thank Jim Hamilton, Kajal Lahiri, Kevin Lansing, Sebastiano Manzan, Sophocles Mavroeidis, Bent Nielsen, Huaming Peng, Barbara Rossi, Frank Schorfheide, Shu-Ping Shi, Allan Timmermann, Mark Watson, Jonathan Wright, the participants in the SNDE Symposium, the NBER/NSF Time Series conference, the NBER Summer Institute, ESEM, EC^2 as well as seminar/workshop participants at SUNY at Albany, Baruch College, CUNY, CREST, Durham Business School, ESSEC, Oxford University, HEC Montreal, Paris-Ouest Nanterre, Hannover. Chevillon also acknowledges research support from the Labex MME-DII (ANR-11-LBX-0023-01) and CREST.

†Corresponding author. Email: chevillon@essec.edu; Address: ESSEC Business School, Avenue B. Hirsch, 95021 Cergy-Pontoise, FRANCE.

1 Introduction and motivations

A distinctive feature of the last two decades has been the prolonged build-ups and sharp collapses in asset markets in the industrialized and developing worlds. Such patterns are often labeled 'bubbles', and a rich literature has developed that provides mechanisms for their formations and empirical techniques for their detection. The purpose of this paper is to complement the literature by providing a model for making probabilistic forecasts about the emergence, evolution and collapse of bubbles, together with so-called flash crashes which take the form of downward bubbles.

To achieve such a purpose, we need a working definition of bubbles and flash crashes. The literature does not appear to have converged to a commonly agreed definition (e.g. Hamilton, 1986, and Granger and White, 2011) but a stylized feature is that bubbles in a time series y_t are characterized by sustained positive growth, flash crashes by sharp negative growth. As commonly understood, a bubble eventually bursts so the sustained growth is only temporary and the subsequent collapse brings y_t to 'normal times'. It is often understood that 'sustained growth' refers to processes that exhibit exponential growth rates, whereas 'normal times' are characterized by mean reversion or stationarity for some transforms of the data. The duration of the bubble or flash crash must also relate to the frequency of the data and the available sample: equity price bubbles measured at monthly or quarterly frequencies generally last at most a few years – see Figure 1, Panel (a) – although some lasting bubbles have been detected over longer spans – e.g. Figure 1, Panel (b). As seen in Panel (c), some series experience a substantive growth over a few periods, but the ensuing devolution may not necessarily take the form of a sharp drop. Finally, stock market flash crashes are observed at intraday frequencies: see Panel (d) for an example of a crash over a few minutes.

Our interest in this paper is to obtain predictive probabilities for bubble and flash crash episodes in a local-asymptotic framework. Bubbles are, here, defined via their time series properties, hence following the time series approach Granger and White (2011) rather than relying on a structural decomposition that studies deviations from fundamental dynamics. Local-asymptotics, in which parameters relate to the available sample size, say T, have successfully been used many times in the econometrics literature, and recently in the context of bubble detections, see Phillips, Wu and Yu (2011). Here, we allow not only parameters, but also the timing and duration of events

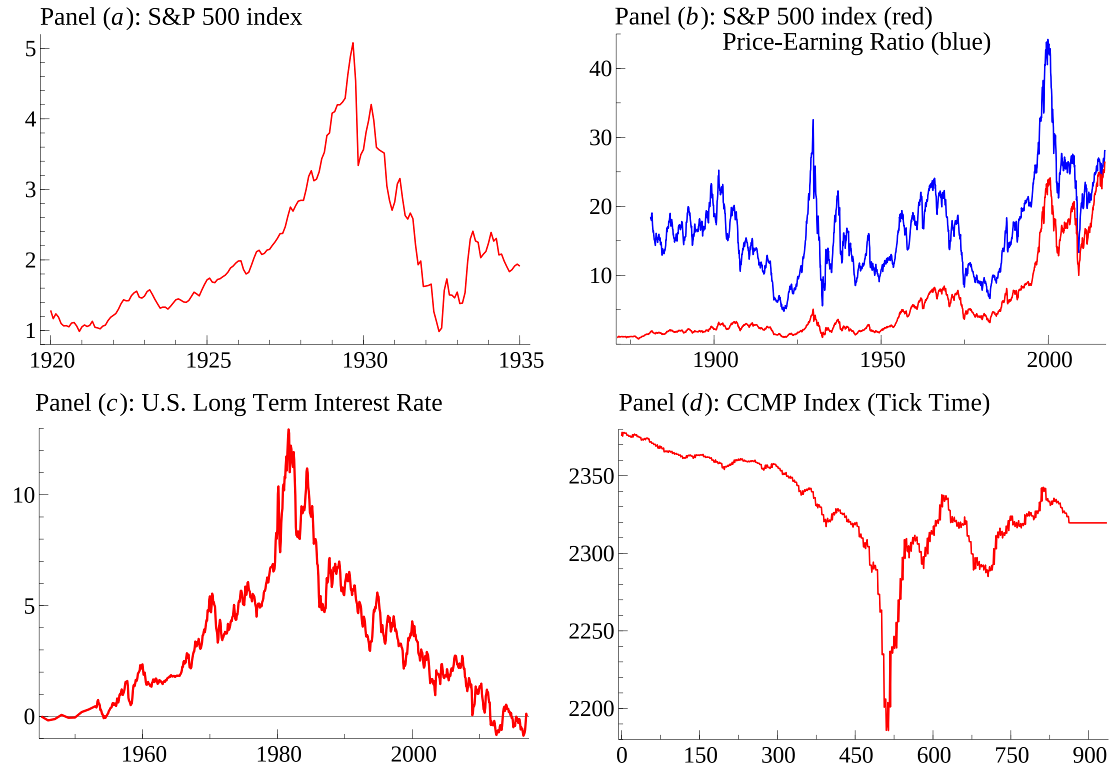

Figure 1: The figure presents examples of data series. Panels (a)-(c) produce monthly data from Robert Shiller's website: the S&P 500 index on (a) and (b) , the cyclically adjusted Price/Earning Ratio and the long term interest rate on Panel (c). Panel (d) is the Nasdaq index expressed in Tick time after 3:00PM EST on May 6, 2010. The data were downloaded from a Bloomberg Terminal as CCMP index (Bid).

to be expressed in relation to T: letting the latter tend to infinity allows to express magnitudes and derive distributions that prove useful in forecasting. This is a procedure which is common, for instance, when modeling and testing for structural breaks where the timing of the breaks is referred to in relation with the sample size (see, e.g., Andrews, 1993b, Perron, 1996, Bai and Perron, 1998, and Magnusson and Mavroeidis, 2014) and bears some resemblance with the literature on "in-fill" asymptotics (Jiang, Wang and Yu, 2017). As a consequence, we define probabilistic bubble episodes of duration h as exhibiting sustained growth, i.e. such that $g_{t,h} \equiv y_{t+h}/y_t$ is greater than unity. In a non-local asymptotic framework, this leads to defining the probability of a bubble episode at time t, of duration h and magnitude $\gamma > 1$ as the conditional probability $\Pr(g_{t,h} > \gamma | \mathcal{I}_t)$. When the data generating process or the timing and duration (possibly also the magnitude) are expressed as functions of the sample size in a local-asymptotic setting, the definition is better expressed in the limit as the sample size increases to infinity: the contrast between seemingly related finite sample dynamics become starker.

We start with defining the probability of occurrence of a bubble in a local-asymptotic framework: it allows to assess the relevance of models and compare their applicability in this context. It also justifies introducing a new modeling strategy, which is the main focus of the paper.

Definition 1 *Let the local-asymptotic process* $(y_{T,s})$ *be defined for* $0 \leq s \leq T$. *Consider two non-negative deterministic mappings for local time,* $\tau_{(T,t)}$, *and duration,* $\chi_{(T,h)}$ *for some parameters* $(t,h) \in \mathbb{R}_+^2$, *such that* $\tau_{(T,t)} \leq T$ *and, as* $T \to \infty$, $\left(\tau_{(T,t)}, \chi_{(T,h)}\right) \to (\infty, \infty)$ *with* $\chi_{(T,t)}/\tau_{(T,t)} \to 0$. *The probability of the process experiencing a bubble episode of magnitude* $\gamma > 1$ *at local time* $\tau_{(T,t)}$ *and of duration* $\chi_{(T,t)}$ *is defined as*

$$\pi_{t,h}(\gamma) = \lim_{T \to \infty} \Pr(g_{\tau_{(T,t)},\chi_{(T,h)}} > \gamma \Big| \mathcal{I}_{\tau_{(T,t)}}),$$

where $g_{s,u} \equiv y_{T,s+u}/y_{T,s}$ *is the growth between dates* s *and* $s+u$, *and* $\mathcal{I}_s$ *is the sigma-field generated by* $(y_{T,j})_{j \leq s}$.

The local time and duration mappings are specific to the models considered so, in empirical applications, the modeler who wishes to forecast the probability of bubbles will need to define these mappings appropriately. Typical mappings encountered in the econometric literature for local time and duration relate to the sample size, as parameterized by $(t,h) \in [0,1]^2$, i.e., $\tau_{(T,t)} = \lfloor tT^{\alpha} \rfloor$ and $\chi_{(T,h)} = \lfloor hT^{\beta} \rfloor$, for $(\alpha, \beta) \in [0,1]^2$ and where $\lfloor \cdot \rfloor$ denotes the integer part. In the literature on deterministic breaks detection, Bai and Perron (1998) for instance parameterize the break date as a fraction $\lfloor tT \rfloor$ of the sample size, i.e., with $\alpha = 1$. In the context of long run forecasting, predictive regressions and impulse responses, an extensive literature considers $\alpha = \beta = 1$ — see, inter alia, Richardson and Stock (1989), Stock (1996), Phillips (1998), Kemp (1999), Valkanov (2003), Gospodinov (2004) and Pesavento and Rossi (2006) — as opposed to the fixed horizon case $\beta = 0$. More recently, Mikusheva (2012) allows for intermediate impulse responses, such as when $\beta \in (0,1)$. In this paper we consider more general parameterizations, but Definition 1 assumes that the duration is such that, as $T \to \infty$, $\chi_{(T,t)}/\tau_{(T,t)} \to 0$, i.e., the duration of the episode is short to ensure that bubbles eventually end if a sample of sufficient length is observed.

Definition 1 enforces a dichotomy between processes that may generate bubbles and those that do not. To clarify this, we present a few examples in Appendix A.2 and summarize them in Table 1. In the first three examples of (*i*) deterministic polynomial growth, (*ii*) near stochastic trends or (*iii*) Froot-Obsfeld (1991) intrinsic bubbles, none of the processes can exhibit bubbles or flash crashes in the way we define them here, since they all yield $\pi_{t,h}(\gamma) = 0$ for $\gamma > 1$.

The simplest model for bubble dynamics may be the AR(1) with an autoregressive coefficient, say ρ, which has to be above unity – yet close to it to be empirically relevant – to generate an exponential growth, e.g. stochastic processes that can be modeled as in Examples (*iv*) and (*v*) in Table 1. Yet, the explosive autoregressive root must be temporary and the coefficient ρ must shift to a value below unity during normal times as in Example (*vi*) that was proposed by Blanchard and Watson (1982). Examples (*iv*)-(*vi*) do generate bubbles according to our definition. Yet these models generate bubbles with probability one since there exists $\gamma > 1$ such that $\pi_{t,h}(\gamma) = 1$. Hence these models cannot distinguish between permanent and temporary bubbles (as opposed to (*vii*) in Table 1, i.e. the model we propose in this paper). Some authors have assumed the presence of deterministic breaks in ρ (e.g. Phillips, Wu and Yu, 2011 or Phillips, Shi and Yu, 2015), as exemplified in Figure 2, Panel (*a*), but such breaks are by construction unpredictable. It is therefore necessary for forecasting purposes to assume that ρ shifts randomly, denote the resulting process ρ_t then.[1] Proposals that derive from Blanchard and Watson (1982) often require for tractability that regimes present fixed transition or unconditional probabilities making them periodic, in a sense. This is an assumption we want to avoid here as we do not want to specify ex ante the probability of observing bubbles. All the models above that accomodate the existence of bubbles can generate both upward and downward trajectories. Hence, when $y_t < 0$, our definition can be generalized to negative bubbles, also called "flash crashes" (not to be confused with the implosion of bubbles as studied by Phillips and Shi, 2017) by considering negative changes $y_{t+h} - y_t < 0$, while ensuring

[1] An additional benefit of specifying that the autoregressive coefficient is stochastic – as opposed to subject to deterministic breaks – is that we can draw inference on the whole sample and there is no need to resort to rolling or recursive windows to test the presence of a bubble and estimate its magnitude; the absence of deterministic breaks also avoids the usual trimming of observations at the beginning or end of the sample.

$g_{t,h} > 1$. In this sense, bubbles and flash crashes are symmetric in the sense that they pertain to the same dynamic models: they differ in that, during the episode, the change is positive or negative: a flash crash is a period of fast decay, which may differ from the dynamics that follow the burst of a bubble (we show in the empirical applications below that the restriction $y_t < 0$ needs not constrain the analysis, but requires a careful choice of a reference point).

We propose in this paper a simple time-series model that accommodates both 'quiet' and bubble periods, and where bubbles arise and disappear as a function of a latent process (which in standard pricing models may relate to the stochastic discount factor). A flexible way to achieve such a purpose is found in the literature on random coefficient autoregressive models (RCAR), see e.g. Nicholls and Quinn (1982) and Granger and Swanson (1997). The simplest version assumes that the variable of interest, y_t, follows an autoregressive process of order 1, $y_t = \rho_t y_{t-1} + \eta_t$, with random autoregressive coefficient ρ_t which is identically and independently distributed (iid). When ρ_t takes values on either side of unity with nonzero probabilities, (y_t) exhibits rich dynamics with 'mean reverting' and explosive periods, see Figure 2, Panel (b). The assumption that ρ_t is iid is made for simplicity and may appear unusual considering the literature on time-varying parameters which are often assumed to follow random walks. But in a dynamic setting, $\rho_t \sim iid$ generates rich dynamics since the h-period growth $g_{t,h} = y_{t+h}/y_t$ involves the product $\prod_{i=1}^{h} \rho_{t+i} = \exp \sum_{i=1}^{h} \log \rho_{t+i}$ when $\rho_{t+i} > 0$, i.e., the exponential of a partial sum which replicates the persistence that is often assumed in time-varying parameter models, see Figure 2(d) for a simulated example.

RCAR processes have been used to model bubbles but with limited success so far (see e.g. Hwang and Basawa, 2005 and Homm and Breitung, 2012). A reason is that to be empirically relevant, the distribution of ρ_t must not only remain very concentrated and close to unity but it should also depend on the size of the sample period, T. Indeed it is well known that explosive autoregressive coefficients, even with small departures from unity, may generate very large values of the process over short samples. Hence we borrow here from the literature on local-asymptotics (e.g. Aue, 2008): we model the distribution of ρ_t so the proximity of $\mathsf{E}[\rho_t]$ and $\mathsf{V}[\rho_t]$ to, respectively, unity and zero relates to the observable sample size and a concentration parameter α. Increasing α would generate higher concentration of the random coefficient ρ_t near unity. As mentioned above, the duration of bubble episodes must also relate to the sample size, hence the local asymptotic framework is well suited for this purpose. We design a parametric setting to ensure that the magnitude of $\mathsf{V}[\rho_t]$ remains large enough for the random nature of ρ_t to matter asymptotically (in contrast with Aue, 2008). A consequence is that we are able to make probabilistic forecasts about the evolution of bubbles that are direct functions of values taken by a latent Brownian motion. In empirical work, the modeler can test candidate variables for this latent process and thus possibly improve the probabilistic forecasts.

The content of the paper is as follows. In Section 2, we propose a model along the lines and with the purposes delineated above, then derive its asymptotic properties. In Section 3, we show how these properties can be used to provide probabilistic forecasts of bubbles or crashes conditional on the observed history. In Section 4, we delineate an empirical inference procedure and then provide an extensive empirical application to several datasets plotted in Figure 1. Section 5 concludes. The Appendix, Section A, provides (*i*) an additional discussion of the related literature, (*ii*) examples of processes that may or not exhibit bubbles, (*iii*) further results on the inference procedure; and Section B collects the proofs of the propositions. Supplementary Material containing further technical results is available online. Throughout the paper, $\lfloor\cdot\rfloor$ denotes the integer part; $1_{\{\cdot\}}$ the indicator function that takes value one if $\{\cdot\}$ is true and zero otherwise; and $\mathbb{R}_{+,*}$ the set of strictly positive real scalars.

Table 1: Examples of stochastic processes with corresponding probabilities of bubbles or crashes as given by $\pi_{t,h}(\gamma)$ in Definition 1. x_t is weakly stationary, η_t and u_t are independent *iid* processes. Details for (*i*)-(*vi*) are given in Appendix A.2. Example (*vii*) is the focus of this paper and is delineated in Section 2.

(*i*) Trend	$y_t = \beta_t + x_t$, $\beta_t \sim c_\beta t^\nu$, $\nu > 0$,	$\pi_{t,h}(\gamma) = 0$
(*ii*) STUR	$y_t = \rho_t y_{t-1} + \eta_t$, $\rho_t - 1 = O_p(T^{-1})$,	$\pi_{t,h}(\gamma) = 0$
(*iii*) Froot-Obsfeld	$y_t = c_\eta(\sum_{j=1}^{t}\eta_j)^\lambda$, $\lambda > 0$,	$\pi_{t,h}(\gamma) = 0$
(*iv*) Explosive	$y_t = \rho_t y_{t-1} + x_t$, $\rho_s \geqslant 1+\epsilon$, $\epsilon > 0$, for $s \in [t, t+h]$,	$\pi_{t,h}(\gamma) = 1$
(*v*) Local Explosive	$y_t = \rho_t y_{t-1} + x_t$, $T^\alpha(\rho_s - 1) \to \phi > 0$, $s \geq t$, $\alpha \in (0,1)$,	$\pi_{t,h}\left(e^{\phi h}\right) = 1$
(*vi*) Blanchard-Watson	$y_t = \rho\pi_t y_{t-1} + \eta_t$, $\rho > 1$ and $\pi_t \overset{iid}{\sim} Bernoulli(\pi)$,	$\pi_{t,h}(\lim_{T\to\infty}(\rho\pi)^h) = 1$
(*vii*) NERC	$y_t = e^{\frac{\phi+\lambda T^{\alpha/2}u_t}{T^\alpha}} y_{t-1} + \eta_t$, $\alpha \in (0,1)$, $\phi + \lambda^2 \geq 0$,	$\pi_{t,h}(\gamma) \in [0,1]$

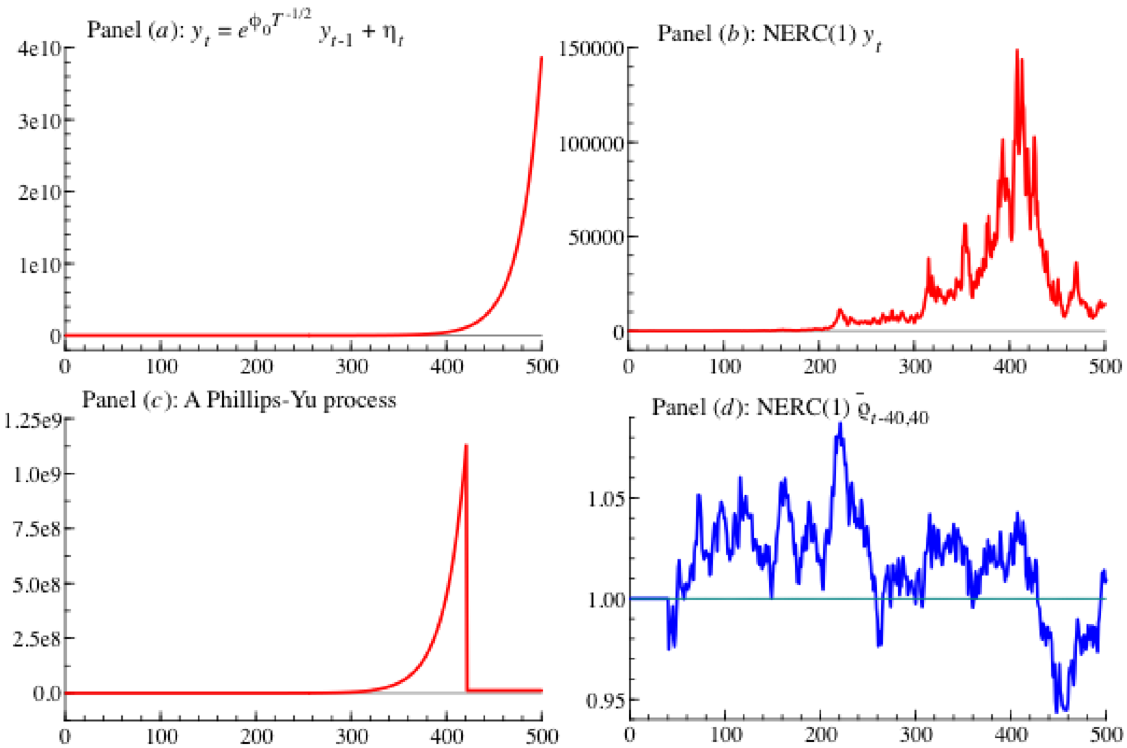

Figure 2: The figure presents simulated paths of near explosive AR(1) and NERC(1) processes driven by the same $\eta_t \overset{i.i.d}{\sim} N(0,1)$ with $T = 500$ and $\alpha = 1/2$. Panel (a): a near explosive AR(1) $y_t = e^{\phi_0/\sqrt{T}} y_{t-1} + \eta_t$ with $\phi_0 = .625$ so $\rho_{0,T} = 1.028$. Panel (b): a NERC(1) with $(\phi, \lambda) = (.5, .5)$ so $E(\rho_t) = 1.028$. The innovations $u_t \overset{i.i.d}{\sim} N(0,1)$. Panel (c): a bubble process with breaks as in PY, equation (15). The bubble starts at date $\tau_e = 320$ and ends at date $\tau_f = 420$ where the process reverts to the value observed at date τ_e. Prior to τ_e and after τ_f, the process follows a random walk driven by η_t. Panel (d): the stochastic slope parameter $\bar{\rho}_{t-40,40}$ of the NERC(1) simulated in Panel (b).

2 A near-explosive random coefficient autoregressive process

This section presents the model we propose and studies its asymptotic properties.

2.1 The model

We consider the data generating process (DGP) y_t as a Near Explosive Random Coefficient autoregressive process, NERC(1), defined, for $t = 1, \cdots, T + H$, by

$$\begin{aligned} y_t &= \rho_t y_{t-1} + \eta_t, \\ \rho_t &= \exp\left\{\frac{\phi}{T^\alpha} + \frac{\lambda}{T^{\alpha/2}} u_t\right\}, \end{aligned} \tag{1}$$

where $(\phi, \lambda, \alpha) \in \mathbb{R} \times \mathbb{R}_{+,*} \times (0, 1)$, u_t is *i.i.d* with zero mean and unit variance, T refers to the observable sample size and $H = O(T)$ is the maximum forecast horizon. Model (1) constitutes a double array where the distribution of y_t depends on the sample size T but for notational convenience, we omit the dependence on T unless explicitly required.

When $\lambda = 0$, $\mathsf{V}[\rho_t] = 0$ which corresponds to the model of Phillips and Magdalinos (2004, henceforth PM, and 2007), see also Phillips, Wu and Yu, 2011, and Phillips and Yu, 2011; respectively PWY and PY henceforth. Letting $\lambda \neq 0$ constitutes a nontrivial extension to PM. Also, under the standard RCAR model ($\alpha = 0$) consistency results *do not exist* under the assumption of explosive behavior, $\mathsf{E}(\rho_t^2) > 1$. Figure 2 illustrates the point that the NERC model can accommodate the inception and collapse of bubbles without resorting to deterministic breaks. When $\lambda > 0$, the cumulated stochastic innovations u_t induce $\bar{\rho}_{t,h} = \sqrt[h]{\prod_{i=1}^{h} \rho_{t+i}}$ to hover on either side of unity (Panel (d)), leading to the build-up or collapse of the NERC process (Panel (b)).

Throughout the paper, we make the following assumptions concerning Model (1).

Assumption A: The process admits an origin $y_0 = o_p(1)$.

Assumption B: The innovations u_t and η_t are mutually independent processes such that $u_t \sim i.i.d.(0, 1)$, $\eta_t \sim i.i.d.(0, \sigma_\eta^2)$, with

$$\begin{aligned} &\mathsf{E}[|\eta_t|^\nu] < \infty \quad \text{for} \quad \nu \geq \frac{2}{\alpha}; \\ &\mathsf{E}[e^{\lambda u_t}] < \infty \quad \text{for} \quad \lambda \geq 0 \quad \text{and} \quad \mathsf{E}[u_t^3] = 0. \end{aligned}$$

Assumption A is made for ease of exposition in the presence of explosive dynamics. It ensures that y_0 plays no role asymptotically, since $y_0 = o_p(y_1)$. It avoids imposing different assumptions on y_0 depending on model parameters and implies that y_t possesses finite second moments for finite T. Related assumptions are also found in Aue (2008) but with different consequences (see Remark 1 in the Appendix, Section A.1).[2]

In Assumption B, we specify u_t as *i.i.d.*, as opposed to the random walk often considered (e.g. by Stock and Watson, 1998). This is a simplification that ensures that the presence of explosiveness between y_t and y_{t+h} relates to the value of the moving average of the innovations $\bar{u}_{t,h} = \frac{1}{h}\sum_{i=1}^{h} u_{t+i}$ via the geometric average $\bar{\rho}_{t,h} = \sqrt[h]{\prod_{i=1}^{h} \rho_{t+i}} = \exp\left\{\frac{\phi}{T^{\alpha}} + \frac{\lambda}{T^{\alpha/2}}\bar{u}_{t,h}\right\}$ as in:

$$y_{t+h} = \bar{\rho}_{t,h}^{h} y_t + \bar{\eta}_{t,h}, \tag{2}$$

where the multistep errors $\bar{\eta}_{t,h}$ satisfy $\mathsf{E}_t[\bar{\eta}_{t,h}] = 0$. The process y_t may exhibit an explosive pattern between t and $t+h$ if $\phi + \lambda T^{\alpha/2}\bar{u}_{t,h} > 0$. This can happen even when the drift term ϕ is negative. The model is specified in (1) in terms of the proximity of $\log \rho_t$ to zero rather than of ρ_t to unity (as in some of the earlier literature) to restrict the support of ρ_t to $\mathbb{R}_+$ and clarify the mapping between the values of (ϕ, λ) and the corresponding properties of y_t.

Assumption B implies that all moments of u_t exist and that its distribution is symmetric. This implies that the parameters ϕ and λ^2 play similar roles in determining the magnitude of the expectation and variance of ρ_t:

$$\mathsf{E}[\rho_t] \equiv \rho = 1 + \frac{\phi + \frac{1}{2}\lambda^2}{T^{\alpha}} + O(T^{-2\alpha}) \tag{3}$$

$$\mathsf{V}[\rho_t] = \frac{\lambda^2}{T^{\alpha}} + O(T^{-2\alpha}) \tag{4}$$

Letting $\alpha \in (0,1)$ allows the process y_t to exhibit explosiveness in finite samples (as opposed to near unit root behavior when $\alpha = 1$; see PM). It allows also to derive an asymptotic distribution theory with consistent parameter estimators, where inference is feasible and which results in the normality of carefully chosen predictive densities, see below.

[2] In empirical work, we may assume that the process of interest is defined by $z_t = y_t + z_0$ with nonzero z_0 so z_t satisfies $z_t - z_0 = \rho_t(z_{t-1} - z_0) + \eta_t$, in which case y_t denotes the deviation of z_t from the origin of the sample under analysis.

Assumption B implies also that a strong approximation is possible, (see Csörgõ and Horváth, 1993, and PM) and there exist independent standard Brownian motions W, B such that, as $T \to \infty$,

$$\sup_{s \in [0, T^{1-\alpha}]} \left| T^{-\alpha/2} \sum_{t=1}^{\lfloor sT^\alpha \rfloor} u_t - W_s \right| = o_{a.s.}(1) \quad \text{and} \quad \sup_{r \in [0, T^{1-\alpha}]} \left| T^{-\alpha/2} \sigma_\eta^{-1} \sum_{t=1}^{\lfloor rT^\alpha \rfloor} \eta_t - B_r \right| = o_{a.s.}(1). \quad (5)$$

Throughout the paper, asymptotic behaviors depend on the sign of $\log \mathsf{E}[\rho_t^2] \sim 2T^{-\alpha}(\phi + \lambda^2)$ (see Granger and Swanson, 1997, for a discussion) so we define:

$$c = \phi + \lambda^2. \quad (6)$$

2.2 Asymptotic properties

The first step of our analysis is to provide the asymptotic distribution for the NERC model. We prove in the Appendix the following proposition.

Proposition 1 *Let the process y_t be defined for $t \geq 0$ by (1) under Assumptions A and B. Then, for $r \in [0, (T + H)\, T^{-\alpha}]$, as $T \to \infty$,*

$$T^{-\alpha/2} y_{\lfloor rT^\alpha \rfloor} \Rightarrow \sigma_\eta K_{\phi,\lambda}(r),$$

where $\Rightarrow$ denotes weak convergence of the associated probability measure and $K_{\phi,\lambda}$ is defined, for $(\phi, \lambda) \in \mathbb{R} \times \mathbb{R}_{+,*}$ and $r \in \mathbb{R}_+$, as

$$K_{\phi,\lambda}(r) = \int_0^r \exp\{(r - s)\phi + \lambda(W_r - W_s)\} dB_s, \quad (7)$$

with W and B independent standard Brownian motions defined as the limits in expression (5).

Proposition 1 shows that several cases arise depending on whether the distribution of $K_{\phi,\lambda}(r)$ remains bounded. Indeed, $\mathsf{V}[K_{\phi,\lambda}(r)] = \int_0^r e^{2cs} ds$. When $\lambda = 0$, $K_{\phi,\lambda}(r)$ reduces to the Ornstein-Uhlenbeck diffusion considered in PM. When $c \leq 0$, the magnitude of y_t, $t = \lfloor rT^\alpha \rfloor$ for $r \leq (T + H)\, T^{-\alpha}$, is similar to that which PM obtain, with $\mathsf{V}[y_t] = O(T^\alpha)$ if $c < 0$ and $\mathsf{V}[y_t] = O(T)$

if $c = 0$. When $c > 0$, the process with fixed origin exhibits explosiveness in its second moment as pointed out by Hwang and Basawa (2005): $\mathsf{V}[y_t] = O(T^\alpha e^{2cT^{1-\alpha}})$.

Although we do not consider it explicitly, the proposition above can be extended to cover $\alpha = 1$ as the following corollary shows.

Corollary 1 *Proposition 1 also holds when* $\alpha = 1$*:*
$T^{-1/2} y_{\lfloor rT \rfloor} \Rightarrow \sigma_\eta K_{\phi,\lambda}(r)$ *for* $r \in [0, 1 + H/T]$ *and as* $T \to \infty$*.*

Corollary 1 shows that explosive patterns may only arise in y_t if $\alpha < 1$: indeed when $\alpha = 1$ and for $r \in [0, 1]$, $y_{\lfloor rT \rfloor} = O_p\left(\sqrt{T}\right)$, i.e. the magnitude of a (stochastic) unit root process, see Granger and Swanson (1997). The local-asymptotic assumption $\alpha = 1$ is considered in Lieberman and Phillips (2017). In this paper, we restrict our attention to the case $\alpha \in (0, 1)$.

The following proposition gives the asymptotic behavior of the growth in y_t.

Proposition 2 *Let the process* y_t *be defined for* $t \geq 0$ *by (1) under Assumptions A and B. Then for* $(r, r+s) \in (0, (T+H)\, T^{-\alpha}]^2$ *with* $s > 0$*, as* $T \to \infty$*,*

$$g_{\lfloor T^\alpha r \rfloor, \lfloor T^\alpha s \rfloor} = \frac{y_{\lfloor T^\alpha r \rfloor + \lfloor T^\alpha s \rfloor}}{y_{\lfloor T^\alpha r \rfloor}} \Rightarrow \exp\{\phi s + \lambda(W_{r+s} - W_r)\} + \frac{\int_r^{r+s} e^{\phi(r+s-u)+\lambda(W_{r+s}-W_u)} dB_u}{\int_0^r e^{\phi(r-u)+\lambda(W_r - W_u)} dB_u}, \quad (8)$$

Conditional on $\mathcal{I}_{\lfloor T^\alpha r \rfloor}$*, the following limit holds*

$$T^{-\alpha/2} y_{\lfloor T^\alpha r \rfloor} (g_{\lfloor T^\alpha r \rfloor, \lfloor T^\alpha s \rfloor} - e^{\phi s + \lambda(W_{r+s} - W_r)}) \Big| \mathcal{I}_{\lfloor T^\alpha r \rfloor} \Rightarrow \int_r^{r+s} e^{\phi(r+s-u)+\lambda(W_{r+s}-W_u)} dB_u. \quad (9)$$

Proposition 2 provides the key results that we use in the next section when forecasting bubbles. Expression (8), which holds even for $\lambda = 0$ (as in PM) provides the asymptotic distribution of ratios of y_t separated by a time interval where both the dates and distances are expressed as a function of the sample size; expression (9) considers the same ratio whose distribution is conditional on information available at time $t = \lfloor T^\alpha r \rfloor$. In deriving the latter expression, we notice that numerator and denominator on the right hand side of expression (8) are independent and that $T^{-\alpha/2} y_{\lfloor rT^\alpha \rfloor}$ admits the limit given in Proposition 1. We show in the next section how the Proposition 2 can be used to make probabilistic forecasts.

3 Probabilistic Forecasts of Bubbles and Crashes

An attractive feature of the model we propose, is that it provides a distributional assumption on ρ_t contrary to models where ρ_t breaks deterministically. As a consequence, we can answer questions on the probability that a bubble forms, bursts, continues and so on. We show in this section, how the NERC model can be used to assess simply the probability that bubbles or crashes form over a given horizon. We only discuss bubble episodes here since the definition of flash crashes coincides with bubbles, except that they concern processes that take negative values.

3.1 Predictive probabilities

Definition 1 enforces a dichotomy between processes that generate bubbles and those that do not. As discussed in the introduction, we provide some examples (Appendix A.2) where the probability of bubble prediction $\pi_{t,h}(\gamma) = 0$ or 1 for all horizons $\chi_{(T,h)} = o(\tau_{(T,t)})$.

Building on the discussion and in line with Definition 1, we obtain, in the remainder of this paper, predictive probabilities of bubbles and crashes under the NERC model with $\tau_{(T,t)} = \lfloor tT^{\alpha} \rfloor$ and $\chi_{(T,h)} = \lfloor hT^{\alpha} \rfloor$ (for notational ease, we omit below the subscripts in τ and χ) Proposition 2 shows that the conditional distribution can be stated as $g_{t,h}|\mathcal{I}_{\tau} = \exp\left\{\phi\frac{\chi}{T^{\alpha}} + \lambda\left(W_{\frac{\tau+\chi}{T^{\alpha}}} - W_{\frac{\tau}{T^{\alpha}}}\right)\right\} + \frac{T^{\alpha/2}}{y_{\tau}}\int_{\frac{\tau}{T^{\alpha}}}^{\frac{t\tau\chi}{T^{\alpha}}} \exp\left\{\phi(\frac{\tau+\chi}{T^{\alpha}} - u) + \lambda\left(W_{\frac{\tau+\chi}{T^{\alpha}}} - W_u\right)\right\} dB_u + o_p(\frac{T^{\alpha/2}}{y_{\tau}}e^{c\frac{\chi}{T^{\alpha}}})$, which allows to provide the following Corollary to Proposition 2.

Corollary 2 *Under the assumptions of Proposition 2, and for* $(t, t+h) \in [0, (T+H)\,T^{-\alpha}]^2$ *with* $h > 0$, *predictive probabilities defined in Definition 1 satisfy, as* $T \to \infty$,

$$\pi_{t,h}(\gamma) \sim \mathsf{P}\left(e^{\phi h + \lambda(W_{t+h} - W_t)} + \frac{T^{\alpha/2}}{y_{\lfloor tT^{\alpha}\rfloor}}\int_t^{t+h} e^{\phi(t+h-u)+\lambda(W_{t+h}-W_u)} dB_u > \gamma \middle| \mathcal{I}_{\lfloor tT^{\alpha}\rfloor}\right). \tag{10}$$

Values of predictive probabilities $\pi_{t,h}(\gamma)$ can be obtain by simulation of the right-hand side of expression (10). Cases of interest comprise, in particular, $\pi_{t,h}(y_{\tau}/y_{\tau-\chi})$, which is the probability that the growth observed over the latest χ period carries over to the next χ periods. Also of interest is, e.g., $\pi_{t,h}(r_{\tau,\chi})$ when $y_{\tau+\chi}/y_{\tau}$ denotes χ-period log returns of assets and $r_{\tau,\chi}$ is the risk free rate of return and is part of the information set at time τ.

3.2 Simulation

To study the patterns generated by predictive probabilities, we simulate using 10^5 Monte Carlo replications the values of $\pi_{t,h}(\gamma)$ defined in (10) for $\gamma = 1.05$ and $\alpha = .8$. We consider $(\phi, \lambda) \in \{-1, 0, 1\} \times \{0, 5, 1, 2, 3\}$, and $h \in \{1, 6, 24\}$. In Figure 3, we report simulated $\pi_{t,h}(\gamma)$ for a large sample size $T = 1000$ where $900 \leq t \leq 1000$. Figure 4 records similar simulations for a smaller sample size $T = 200$ and we let $100 \leq t \leq 200$.

When $\phi = -1$ and for the range of (λ, h) considered, with $\gamma > 1$, both figures report probabilities that hover below .50 and that decrease with h since the latter impacts positively the variance of conditional forecasts. Probabilities reported in Figure 3 hardly vary with t, which shows that they are obtained from the ergodic distributions of the weakly stationary process. This is not the case when γ is time varying (e.g. a risk free rate) or for the smaller sample size reported in Figure 4 where predictive probabilities tend to decrease as t increases.

For positive values of ϕ, the probabilities process reported in Figure 3 tend to be more dispersed as of function λ and get closer to 0.5 as λ and h increases. They are also more stable as a function of t for larger values of λ. In the relatively large sample considered in the figure, the predictive probabilities are always below .5 except when $\phi = 1$ and $h = 24$. By contrast, in the smaller sample size recorded in Figure 4, $\pi_{t,h}(\gamma) > .5$ for smaller values of h when $\phi > 0$. This reflects the larger $\mathsf{E}\left(\rho_t^2\right)$ induced by the smaller T. Notice also that for so small a sample, the probabilities exhibit trending patterns as function of t.

4 Empirics

This section is concerned with putting the analytical results derived above into empirical use. We first provide an implementation scheme and then an empirical illustration to various datasets.

4.1 Implementation Scheme

To provide predictive probabilities $\pi_{t,h}(\gamma)$ as in Definition 1 and expression (10), which involves the parameters of the model, (ϕ, λ, α), we need to perform inference over the latter. The im-

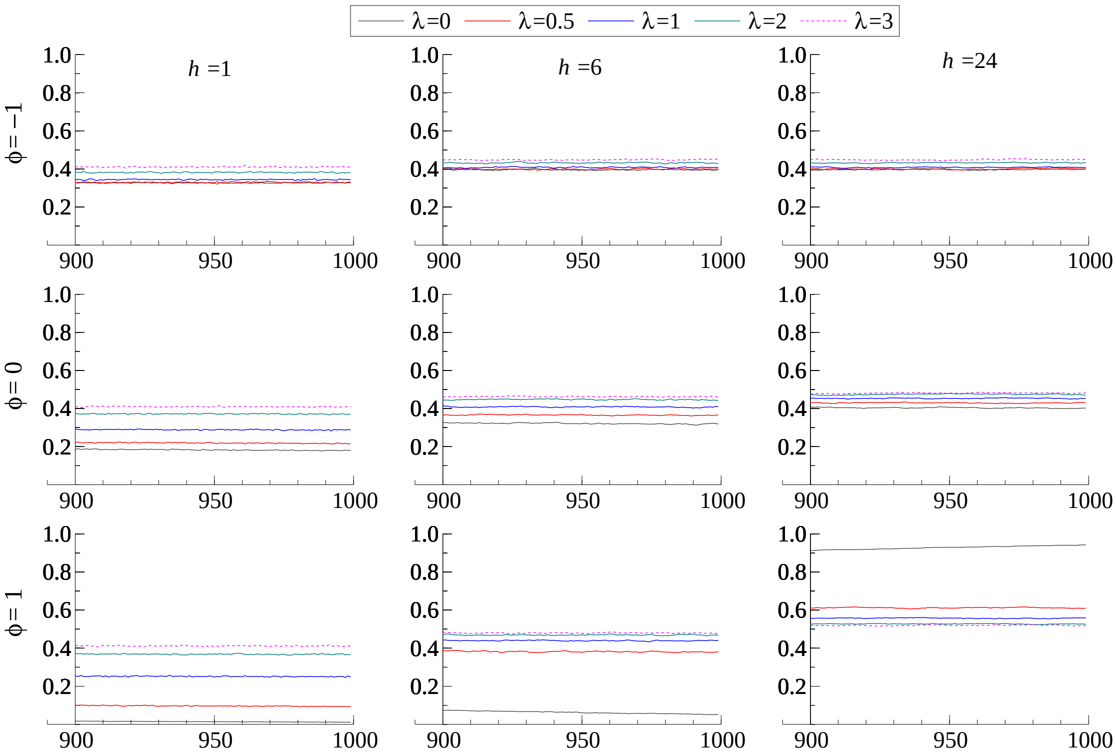

Figure 3: The figure presents records simulated probabilities $\pi_{t,h}(\gamma)$ for $\gamma = 1.05$ and $900 \leq t \leq 1000$ and $\alpha = .8$. Each row records a different value of $\phi \in \{-1, 0, 1\}$, and each column a value of the horizon $h \in \{1, 6, 24\}$; the graphs record different values of $\lambda \in \{0, .5, 1, 2, 3\}$. The number of Monte Carlo replications is 5×10^4.

plementation scheme we suggest therefore consists in obtaining confidence sets for the value of the parameters (ϕ, λ). They are then used for deriving the predictive probabilities of bubbles and crashes. Simulation evidence shows that fixing $\alpha \in (0, 1)$ bears little impact in finite samples on inference regarding the dynamic properties of y_t.[3]

As is common in the context of local asymptotics, consistent estimators of the localizing parameters (ϕ, λ) may be infeasible when $c \geq 0$.[4] Hence, we resort to the technique that is now standard under local asymptotics and consists in inverting a test statistic. There exists a significant literature where such an approach is used for inference in the near-unit root framework (originating in Stock,

[3]Indeed, α constitutes essentially a scaling parameter for the moments of ρ_t in (3)-(4).

[4]Hence, we do not consider the nonlinear Kalman or particle filters. We do not consider either quasi-maximum likelihood estimators (QMLE) as these are known to present consistency issues in the ERCA (i.e. when $\alpha = 0$, see Berkes et al., 2009) or to depend on nuisance parameters (σ_η^2, see Aue and Horváth, 2011 and Lieberman and Phillips, 2017) for which no estimator has been proved consistent in the explosive case.

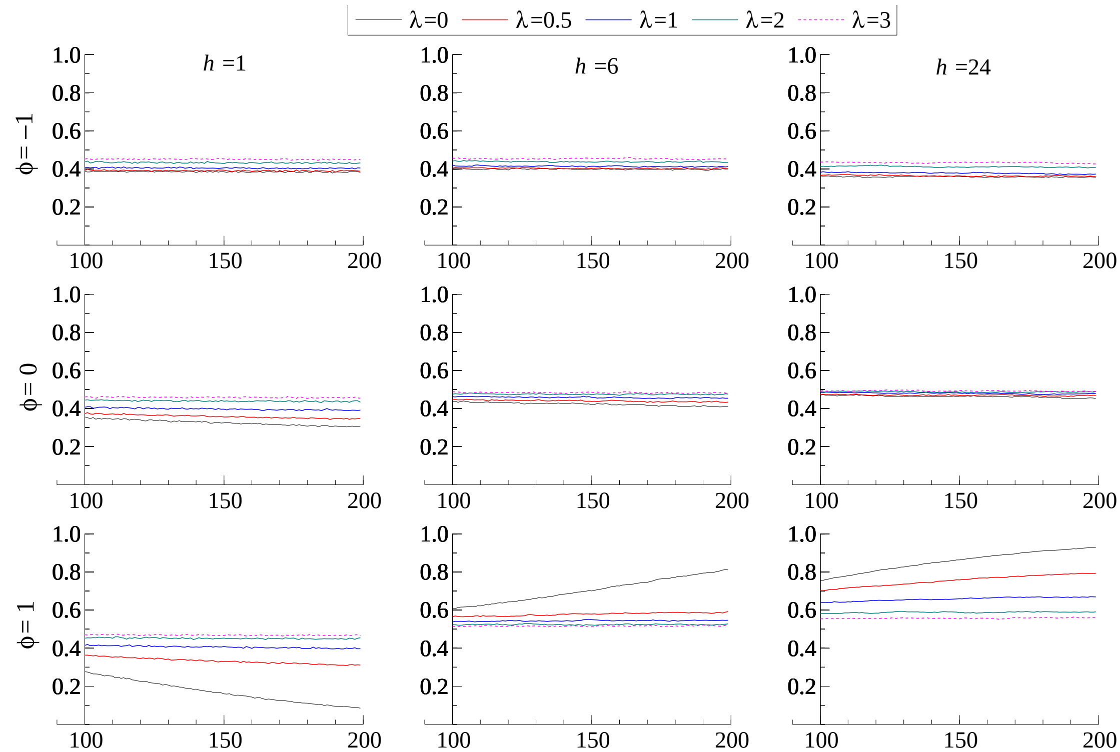

Figure 4: The figure presents records simulated probabilities $\pi_{t,h}(\gamma)$ for $\gamma = 1.05$, $100 \leq t \leq 200$ and $\alpha = .8$. Each row records a different value of $\phi \in \{-1, 0, 1\}$, and each column a value of the horizon $h \in \{1, 6, 24\}$; the graphs record different values of $\lambda \in \{0, .5, 1, 2, 3\}$. The number of Monte Carlo replications is 5×10^4.

1991 and Andrews, 1993a).[5] Inference is performed by constructing asymptotic confidence sets using grid search over the set of parameters that are not rejected under the null. The technique relies on introducing a scalar function $\delta_{\theta,T}$ of $y_1, ..., y_T$ (a test statistic) that satisfies $\delta_{\theta,T} \Rightarrow \mathcal{D}_\theta$, where $\theta = (\phi, \lambda)' \in \Theta$ constitutes the parameter of interest and $\mathcal{D}_\theta$ denotes a distribution that depends on θ. Under the null $\mathsf{H}_0 : \theta = \theta^0$, Stock (1991) constructs asymptotic $100 \times (1 - \omega)$ % confidence sets as $\Theta_\omega \subset \Theta$ consisting of the values θ^0 that are not rejected at the $100 \times \omega\%$ significance level by $\mathcal{D}_{\theta^0}$. The confidence sets are constructed in practice by grid search over the set Θ, computing for each $\theta \in \Theta$ the statistic $\delta_{\theta,T}$ and evaluating it against $\mathcal{D}_\theta$ (which is here obtained by simulation).[6] The

[5]This technique is also common in the context of weak instruments where there exists no fully robust estimation method, but robust tests can be constructed (see Dufour, 1997, and Staiger and Stock, 1997). For papers that discuss the mechanics of the inversion of robust tests to form confidence sets, see Andrews and Stock (2005) and references therein.

[6]In this setting, the least rejected parameter θ^* may constitute a biased estimator of θ but median-unbiased estimation is feasible under the weak convergence assumption, provided that the quantile function is monotonic

grid-bootstrap of Hansen (1999) (see also Mikusheva, 2007) improves on the inference by replacing the asymptotic distribution $\mathcal{D}_\theta$ by a bootstrap analog under the null.

Here we conduct inference using a unique moment condition under the null $\mathsf{H}_0 : (\phi, \lambda) = (\phi_0, \lambda_0)$, such that $\mathsf{E}_{\mathsf{H}_0}[\rho_T] = \rho^0$. The test we choose for simplicity follows from regressing y_t on y_{t-1} and we set the statistic $\delta_{\theta,T}$ to be the OLS estimator $\widehat{\rho} - \rho^0$ scaled by the asymptotic rate given the following proposition (proof in the Appendix).

Proposition 3 *Let the process y_t be defined for $t \geq 0$ by (1) under Assumptions A and B, with $\lambda \neq 0$. Letting $c = \phi + \lambda^2$, the OLS estimator $\widehat{\rho}$ in the regression of y_t on y_{t-1} satisfies, as $T \to \infty$:*
(i) if $c < 0$, $\delta_{\theta,T} = T^{\frac{1+\alpha}{2}} (\widehat{\rho} - \rho) \xrightarrow{L} \mathsf{N}\left(0, -2\phi + \lambda^2\right)$,
(ii) if $c \geq 0$, $\delta_{\theta,T} = T^{\alpha} (\widehat{\rho} - \rho) \Rightarrow \lambda\sqrt{\phi + 3\lambda^2}\, \dfrac{V_\theta}{Z_\theta}$,
where V_θ and Z_θ are uncorrelated variables defined as the limits of the standardized random variables $V_T = \int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)} dW_r$ and $Z_T = \int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)} dr$.

The statistic $\delta_{\theta,T} = \left[T^{\frac{1+\alpha}{2}} 1_{\{c<0\}} + T^{\alpha} 1_{\{c\geq 0\}}\right] (\widehat{\rho} - \rho)$ corresponds to using the moment condition $\mathsf{Cov}\left(y_t - \rho^0 y_{t-1}, y_{t-1}\right) = 0$ under H_0, which is defined for all parameter values owing to Assumption B. The limiting distribution $\mathcal{D}_\theta$ is given by Proposition 3 and can be simulated under H_0. The variance σ_η^2 constitutes a scaling parameter that does not affect the asymptotic distribution of $\widehat{\rho} - \rho^0$, so it is irrelevant in our inference technique (as opposed to other existing methods, see Berkes et al., 2009). Also, α is not identified using the method: it constitutes only a scaling parameter since it does not enter the asymptotic distributions in Proposition 3. Following Phillips (2014) we recognize that as $|\phi| \to \infty$ or $\lambda \to \infty$, the asymptotic distribution of the estimator becomes diffuse, so the confidence sets may become empty when the true data generating process does not present local parameters. We derive the asymptotic power of the test in the Appendix, Section A.3.1, and show that although we obtain valid asymptotic confidence sets under the null, the asymptotic power may be low and the proposed confidence sets may be too wide. We assess their coverage probabilities by simulation in Table 2. The table reports the simulated coverage for a nominal size of .8 (using the asymptotic distribution of the test statistic) when the test is computed

(Stock, 1991, Andrews, 1993a), see also Dufour, Khalaf and Maral (2006) for a discussion of its properties in relation to Hodges and Lehmann (1963) estimators. When $\delta_{\theta,T}$ is a Generalized Method of Moments (GMM) statistic, θ^* can be seen as the continuously-updated estimator (see Stock, Wright and Yogo, 2002) and it inherits its properties.

Table 2: The table reports the Monte Carlo coverage probabilities of confidence interval constructed using the test statistic over a sample of size $T = 1000$ for a nominal probability of 0.80 using the asymptotic distribution. Parameters are set to $\alpha = 0.9$; the rows report the value of ϕ and the columns report the values of λ. The number of Monte Carlo replications is 10^4 and asymptotic distributions are computed using samples of 5×10^4 observations.

	Coverage Probability for a nominal .80				
λ=	0	0.25	0.5	1	2
ϕ = -1	0.846	0.846	0.844	0.770	0.764
-0.5	0.837	0.847	0.851	0.774	0.760
0	0.900	0.761	0.757	0.763	0.772
0.5	0.955	0.758	0.740	0.762	0.783
1	1.000	0.722	0.741	0.765	0.778

over a sample of size $T = 1000$. Table 2 shows that the coverage of confidence sets is close to the nominal – although the test is slightly liberal for larger values of λ. When $\lambda = 0$, confidence sets are as conservative as reported in Phillips (2014).

Once an asymptotic $100 \times (1 - \omega)\,\%$ confidence set Θ_ω is obtained for (ϕ, λ), we can compute the set $\Pi^{\omega}_{t,h}(\gamma)$ of predictive probabilities $\pi_{t,h}(\gamma)$ obtained for each element of Θ_ω. The probabilities of interest are

$$\left(\widehat{\pi}^{\min}_{t,h}(\gamma), \widehat{\pi}^{\max}_{t,h}(\gamma)\right) = \left(\inf_{\pi_{t,h}(\gamma)\in\Pi^{\omega}_{t,h}(\gamma)} \pi_{t,h}(\gamma), \sup_{\pi_{t,h}(\gamma)\in\Pi^{\omega}_{t,h}(\gamma)} \pi_{t,h}(\gamma)\right),$$

as well as $\widehat{\pi}^{med}_{t,h}(\gamma)$ defined as the median of $\Pi^{\omega}_{t,h}(\gamma)$. Probabilities $\left(\widehat{\pi}^{\min}_{t,h}(\gamma), \widehat{\pi}^{\max}_{t,h}(\gamma)\right)$ correspond here to the standard procedure that consists in not reporting predictive intervals that are evaluated at the point estimates of the parameters of the DGP, but instead in reporting those that take full account of parameter estimation uncertainty.

In practice, our simulations and empirical evaluations show that although the choice of α matters for the computation of $\delta_{\theta,T}$, it does not seem to impact significantly the resulting predictive probabilities or the power of the testing procedure for values $\alpha \geq 1/2$: it acts as a scaling parameter for the range of parameter values searched over in the grid. A possible refinement would set α by a simulation aimed at maximizing the weighted average power of the test suggested in Proposition 3 over Θ. This would however induce a significant computational cost.

Whether the predictive probabilities are those of a bubble or a crash depends in practice on the sign of y_t. In the empirical application, we assume that y_t denotes the deviation of the observed data from its initial observation. Whether y_t is negative of positive will yield predictive probabilities of crashes or bubbles. A careful choice of the sample of interest may lead the modeler to target her interest specifically to either type of episode (see the examples below).

4.2 Application

We now turn to an empirical illustration of the proposed method for constructing ranges of predictive probabilities. We consider in turn applications to stock market returns, long term interest rates and PER (all downloaded at the monthly frequency from Robert Shiller's website) as well to the detection of flash crashes (NASDAQ Index downloaded from Bloomberg as CCMP index at the tick frequency starting at 3:00PM EST on May 6, 2010). All series are presented in Figure 1. Datasets and codes for replication are provided on the authors' websites. In these applications, we compute for each date t, the confidence sets for the probability that $g_{t,h}$ at horizon $h \in \{1, 6, 24\}$ is greater than some value γ (which is taken at specific values when these are natural in the context of each example). In all examples, probabilities are computed recursively using only data observed at t and we report in each case $\left(\widehat{\pi}_{t,h}^{\min}(\gamma), \widehat{\pi}_{t,h}^{\max}(\gamma)\right)$ for $\omega = 0.10$, as well as $\widehat{\pi}_{t,h}^{med}(\gamma)$. As in PWY, we proceed to a bias correction by simulating the finite sample counterpart of the distribution $\mathcal{D}_\theta$. The results are presented in Figures 5 to 8. In each figure, graphs on the left-hand side report the data and observed growth (adjusted scales) and graphs on the right-hand side report the statistics referring to growth probabilities. Figures 5 to 8 show that we obtain, using the NERC, probabilities that are neither zero nor unity as in the examples of Table 1.

Figure 5 considers the logarithm of the S&P 500 index since the early 1950s with $\gamma = 1.004$, which corresponds to an annual rate of growth of 5% – approximately the long run nominal growth of the economy. Throughout the sample, $\widehat{\pi}_{t,h}^{\min}(\gamma) < 0.5$ yet close to it (a random walk yields zero probability of a bubble according to our local-asymptotic definition but $\Pr(y_{t+h}/y_t > 1 | \mathcal{I}_t) = 0.5$ for t finite and $h \geq 1$), and $\widehat{\pi}_{t,h}^{\max}$ is strictly less than unity over the first two thirds of the sample, with $\widehat{\pi}_{t,h}^{med}(\gamma)$ close to 0.66. $\widehat{\pi}_{t,h}^{\max}(\gamma)$ increases over the sample to stabilize at 1, reflecting the growth of the data which does not preclude the possibility of explosive growth.

Comparable results are reported in Figure 6 for the logarithm of the long term interest rates since the early 1960s with $\gamma = 1$. We notice the stability of the estimated probabilities. The main difference with the previous figure is that since the interest rate reverts towards the end of the sample to the values observed at the beginning, minimum and maximum probabilities shift downwards, with $\widehat{\pi}_{t,h}^{\max}(\gamma) < 1$, towards the end of the sample, which widens the range of predictive probabilities.

The following figure, Figure 7, also presents a widening of the range of predictive probabilities. This figure reports the high frequency NASDAQ index in tick time with $\gamma = 1$. It is typical of a so-called "flash crash". In the first two thirds of the sample, the median predictive probability is drifting upward and the probability range is indicative of a negative bubble, with bounds that widen on the stationary side once the flash crash has been corrected. We also notice the range of predictive probabilities increases with the horizon h. Finally, Figure 8 reports results for the logarithm of the average PER over a long period with $\gamma = 1$. We notice several episodes of buoyancy where the predictive probability range narrows and shifts upwards. In those episodes, $\widehat{\pi}_{t,h}^{\min}(\gamma)$ moves to values above 0.5 as expected under (temporary) explosiveness.

All applications above are performed with an expanding window, hence the relative stability in the predictive probability range. If the modeler is willing to use rolling windows of observations or a time varying threshold γ, she will obtain more variation in the predictive range.

5 Conclusion

The paper proposes a local asymptotic model that builds on random coefficient autoregressive processes and shows how this NERC model can be applied to the modelling of asset prices.

We show that the process generated by a NERC converges towards the stochastic integral of a geometric Brownian motion, and derive the asymptotic distributions of OLS estimators of the first-order autocorrelation coefficient. We then provide a technique of inference on the parameters of the process based on inverting a test statistic.

As with some existing models for bubbles, the presence of a random coefficient introduces flex-

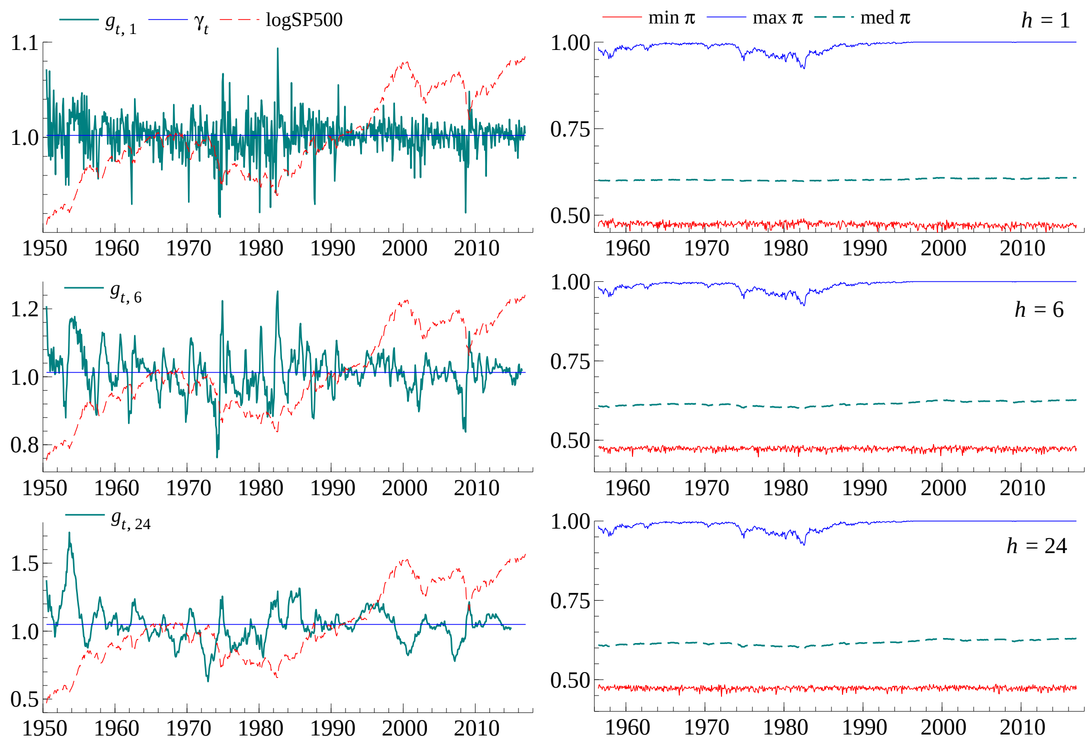

Figure 5: Predictive probabilities for the logarithm of the monthly S&P 500 stock index. The left column reports the actual series as well as the growth $g_{t,h} = y_{t+h}/y_t$ for horizons $h = 1, 6$ and 24 in, respectively, the top, middle and bottom rows. The log price data are scaled to match the mean and range of $g_{t,h}$. The benchmark γ_t for computing probabilities is set to 1.004 for all horizons. The column on the right reports probabilities $\widehat{\pi}^{\min}_{t,h}(\gamma_t)$, $\widehat{\pi}^{\max}_{t,h}(\gamma_t)$ and $\widehat{\pi}^{med}_{t,h}(\gamma_t)$. Minimum and maximum are computed over the set of parameters which are not rejected at a nominal size of 0.10.

ibility in the modelling of multiple bubbles. Here, bubbles may – or not – appear, and by avoiding regime switching, we do not imply that they regularly do. Instead, their existence depends on the values taken by a latent process that relates to the stochastic discount factor. The generalization we propose presents benefits that are similar to the univariate locally explosive AR(1) with breaks of Phillips, Wu and Yu (2011), while allowing for full-sample inference. Also our flexible model allows the so-called bubbles either to reflect nonstationary behavior or be caused by large deviations within a strictly stationary model.

Under the NERC, it is also possible to provide density forecasts and establish statements on the probability of bubbles. We apply our methodology to various U.S. financial datasets.

Possible extensions of the NERC(1) comprise multivariate models where a unique latent process

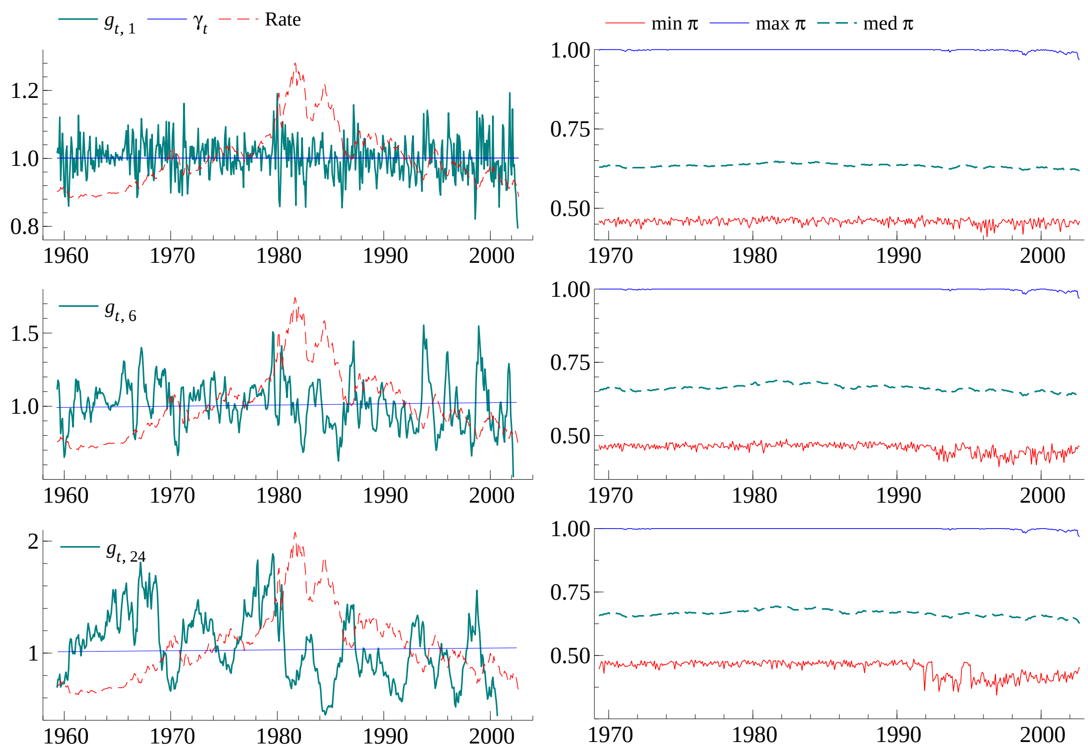

Figure 6: Predictive probabilities for the U.S. monthly long run interest rate. The left column reports the actual series as well as the growth $g_{t,h} = y_{t+h}/y_t$ for horizons $h = 1, 6$ and 24 in, respectively, the top, middle and bottom rows. The interest rate data are scaled to match the mean and range of $g_{t,h}$. The benchmark γ_t for computing probabilities is set to 1 for all horizons. The column on the right reports probabilities $\widehat{\pi}_{t,h}^{\min}(\gamma_t)$, $\widehat{\pi}_{t,h}^{\max}(\gamma_t)$ and $\widehat{\pi}_{t,h}^{med}(\gamma_t)$. Minimum and maximum are computed over the set of parameters which are not rejected at a nominal size of 0.10.

may be causing bubbles that spill over into different markets (as in PY). This might require relaxing the assumption that u_t is $i.i.d.$ In turn, it would then be possible to filter out an estimate of the latent process u_t or the stochastic discount factor. Alternatively, our results allow to postulate and test candidate variables for u_t.

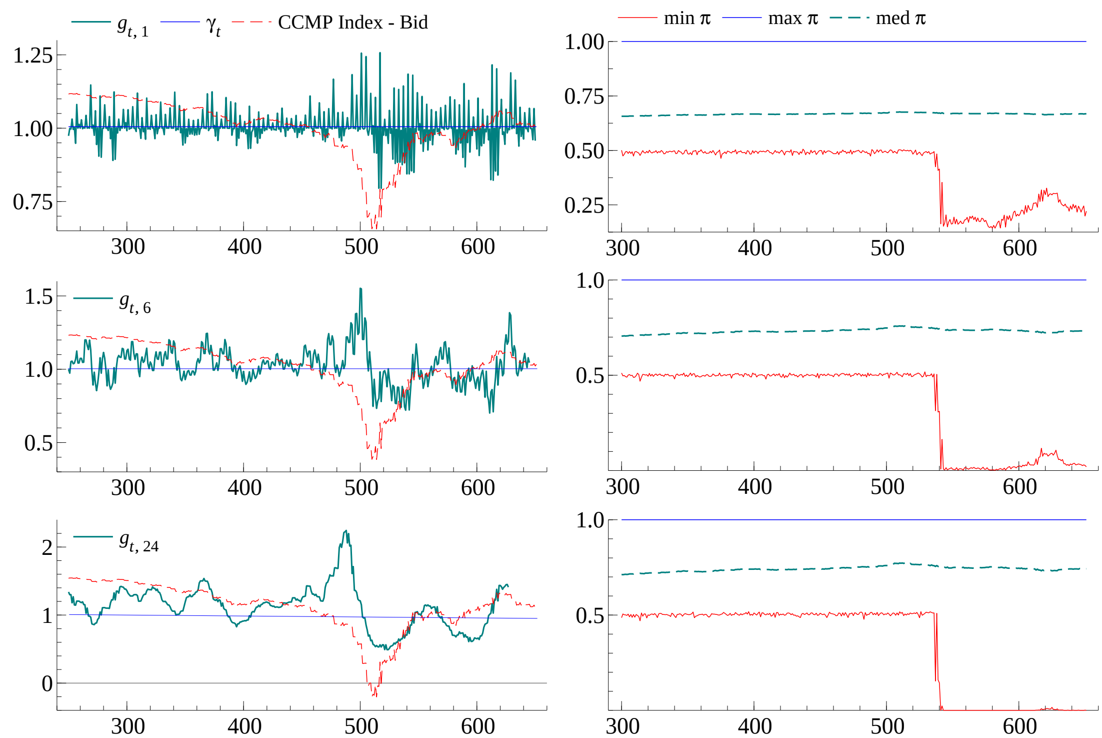

Figure 7: Predictive probabilities for the logarithm of the Nasdaq bid index (labeled CCMP on Bloomberg Terminals) in tick time from 3:00PM EST on May 6, 2010. The left column reports the actual series as well as the growth $g_{t,h} = y_{t+h}/y_t$ for horizons $h = 1, 6$ and 24 in, respectively, the top, middle and bottom rows. The log index data are scaled to match the mean and range of $g_{t,h}$. The benchmark γ_t for computing probabilities is set to 1 for all horizons. The column on the right reports probabilities $\widehat{\pi}_{t,h}^{\min}(\gamma_t)$, $\widehat{\pi}_{t,h}^{\max}(\gamma_t)$ and $\widehat{\pi}_{t,h}^{med}(\gamma_t)$. Minimum and maximum are computed over the set of parameters which are not rejected at a nominal size of 0.10.

A Appendix

A.1 Discussion of related literature

The model we study in this paper belongs to the class of RCAR models as proposed and studied by Andel (1976), Nicholls and Quinn (1982), Bougerol and Picard (1992), McCabe and Tremayne (1995) and Granger and Swanson (1997). The local asymptotic framework we use builds on Bobkoski (1983), Chan and Wei (1987), Phillips (1987) and the more recent work of Giraitis and Phillips (2006) and PM.

Several authors have studied nonstationary RCAR models under non local parameters (i.e.

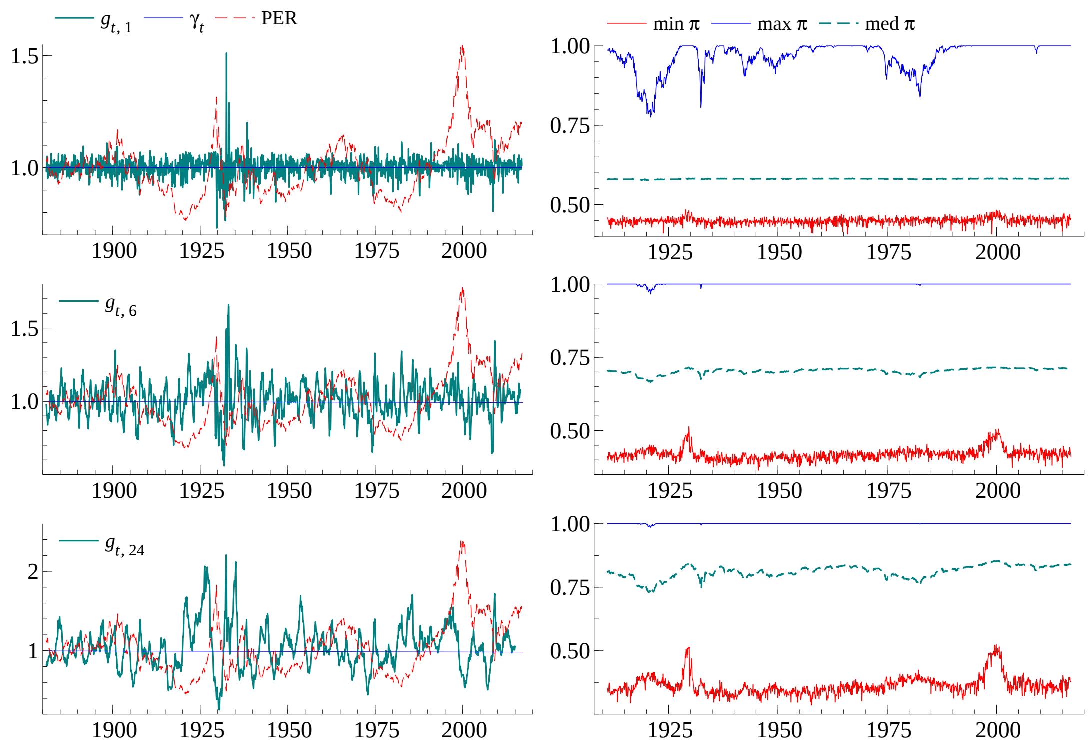

Figure 8: Predictive probabilities for the logarithm of the monthly Price Earning Ratio (PER). The left column reports the actual series as well as the growth $g_{t,h} = y_{t+h}/y_t$ for horizons $h = 1, 6$ and 24 in, respectively, the top, middle and bottom rows. The log PER data are scaled to match the mean and range of $g_{t,h}$. The benchmark γ_t for computing probabilities is set to 1 for all horizons. The column on the right reports probabilities $\widehat{\pi}^{\min}_{t,h}(\gamma_t)$, $\widehat{\pi}^{\max}_{t,h}(\gamma_t)$ and $\widehat{\pi}^{med}_{t,h}(\gamma_t)$. Minimum and maximum are computed over the set of parameters which are not rejected at a nominal size of 0.10.

$\alpha = 0$ in our framework). The unit root hypothesis may then take several forms: $\mathsf{E}[\rho_t] = 1$, or $\mathsf{E}\left[\rho_t^2\right] = 1$; see Granger and Swanson (1997) for a discussion.[7] When $\log \mathsf{E}\left[\rho_t^2\right] > 0$, Hwang and Basawa (2005) name this model an Explosive Random Coefficient Autoregressive model (ERCA): they study processes such that $\log \mathsf{E}\left[\rho_t^2\right] \geq 0$ and $\mathsf{E}\left[\log \rho_t^2\right] < 0$ (which are strictly stationary but do not possess finite second moments).[8] An empirical analysis of the ERCA model with $\mathsf{E}[\rho_t] > 1$ was also made by Charemza and Deadman (1995) in the context of periodically collapsing bubbles

[7]Several Lagrange-Multiplier tests of the unit root hypothesis have been proposed in this framework, see Leybourne, McCabe and Tremayne (1996), Hwang and Basawa (2005), Distaso (2008) and Aue and Horváth (2011).

[8]We assume η_t homoskedastic since expression (1) implies that y_t exhibits conditional heteroskedasticity: for $\rho_t \sim i.i.d.\left(\rho, \sigma_\rho^2\right)$ then

$$\mathsf{E}\left[y_t | y_{t-1}\right] = \rho y_{t-1}, \quad \mathsf{Var}\left[y_t | y_{t-1}\right] = \sigma_\rho^2 y_{t-1}^2 + \sigma_\eta^2$$

see inter alia Tsay (1987) and Hwang and Basawa (2005). These authors, as well as others, have also proposed functional forms that differ from (1) and that belong to the classes of double-autoregressive or bilinear processes.

(see also, Aue and Horváth, 2011, and Wang and Ghosh, 2008).

The NERC also differs from Markov-Switching models such as considered by Hall, Psaradakis and Sola (1999) and Fulop and Yu (2014) where bubbles have constant expected duration. Yet, in our model the distribution of ρ_t is allowed to be bi- or multi-modal so that apparent "regimes" are not precluded.

Local-asymptotic RCARs have also been studied in the literature under differing parametric setting. The following two remarks review them.

Remark 1 *The model we propose deviates non-trivially from that of Aue (2008, Aue henceforth) in that we allow for a greater role played by the stochastic variation in ρ_t. In his setting $\mathsf{E}[\rho_t] - 1 = O(T^{-\alpha})$ with $\alpha \in (1/2, 1)$, and $\mathsf{V}[\rho_t] = o(T^{-1})$ which implies that $\mathsf{V}[\rho_t]$ lies in a tighter neighborhood of zero and so does not asymptotically impact*[9] *the tail distributions or explosiveness of y_t. In his framework, the asymptotic distribution of the least-squares estimator of the $AR(1)$ regression parameter coincides with PM. Our assumptions modify Aue to the situation where $\mathsf{V}[\rho_t]$ lies further away from zero and we show that this affects significantly the asymptotic distributions. For ease of exposition, we impose in turn the restriction, not present in Aue, that the first two moments of ρ_t shrink at the same rate.*

Remark 2 *Recent proposals were developed in parallel to our work by Lieberman and Phillips (2017). Their paper consists of two parts.*

(i) *First, they study a model where ρ_t (using the notation above) is time varying but non-stochastic. Extending the results of Lieberman (2012), under the assumption of positive lower bounds for the derivatives of ρ_t with respect to the parameters, Lieberman and Phillips (2017) develop Quasi-Maximum Likelihood Estimation (QMLE). It easy to see that our model does not satisfy such a requirement*[10]*, which makes the QMLE results inapplicable.*

(ii) *Second, they study a model where ρ_t is stochastic with shrinking variance. Their focus is the* near *stochastic unit root model and their parametric framework assumes, with our notation,*

[9] In Aue, conditions $\log \mathsf{E}[\rho_t^2] < 0$ and $\mathsf{E}[\log \rho_t^2] < 0$ are asymptotically equivalent.

[10] Their assumption A3 requires that $|\partial\rho_t/\partial\phi|$ and $|\partial\rho_t/\partial\lambda|$ admit bounded support with lower bound that is strictly positive. Here $\partial\rho_t/\partial\phi = T^{-\alpha}\rho_t$ and $\partial\rho_t/\partial\lambda = T^{-\alpha/2}u_t\rho_t$ so the assumption does not hold as $T \to \infty$.

that $(\phi, \lambda) = (0, 1)$ *and* $\alpha = 1$.[11] *Here, by contrast, we assume* $\alpha \in (0, 1)$, *so we consider cases where the variance of* ρ_t *is of higher magnitude, our focus being on dynamics that may possibly be characterized as* near explosive. *We show below that the implications differ.*

A.2 Examples

This definition enforces a dichotomy between processes that qualify as bubbles and those that do not. We provide some examples below as an illustration. In all examples, we work under the assumptions of Definition 1 for the orders of magnitude of (t, h).

Models that do not allow for bubble episodes In all examples below, for all $\gamma > 1$, $\lim_{T\to\infty} \mathsf{P}(g_{t_T,h_T} > \gamma | \mathcal{I}_t) = 0$. Throughout, x_t denotes a mean-zero covariance stationary process and η_t an $i.i.d.$ process with zero mean and constant variance such that $T^{-1/2} \sum_{j=1}^{\lfloor rT \rfloor} \eta_j \Rightarrow \sigma W(r)$, with $W(r)$ a Wiener process. All the local time and duration mappings, $\tau_{(T,t)}$ and $\chi_{(T,h)}$, are assumed to diverge to $+\infty$ as $T \to \infty$ with $\chi_{(T,h)}/\tau_{(T,t)} \to 0$. For notational simplicity we let τ, χ denote $\tau_{(T,t)}, \chi_{(T,h)}$ wherever possible and all limits are taken as $T \to \infty$.

1. *Deterministic trend*: $y_t = \beta_t + x_t$, where β_t is a deterministic process such that as $t \to \infty$, $\beta_t \sim c_\beta t^\nu$ with $c_\beta \neq 0$. Then for all local time and duration mappings $g_{\tau,\chi} \sim_p (1+\chi/\tau)^\nu \overset{P}{\to} 1$.

2. *(Near-)Stochastic trend*: $y_t = \rho_t y_{t-1} + \eta_t$ with $\rho_t - 1 = O_p(T^{-1})$ satisfies for $(\tau_{(T,t)}/T, \tau_{(T,\bar{t})}/T) \to (t, \bar{t})$, $y_{\bar{\tau}}/y_\tau \Rightarrow 1 + (W_c(\bar{t}) - W_c(t))/W_c(t)$ where $\bar{\tau} = \tau_{(T,\bar{t})}$ and W_c denotes a Ornstein-Uhlenbeck process with parameter $c = \lim_{T\to\infty} T(\rho_T - 1)$. Letting $\bar{\tau} = \tau + \chi$, then $g_{\tau,\chi} = 1 + O_p(\sqrt{\frac{\chi}{\tau}})$ and $g_{\tau,\chi} \overset{P}{\to} 1$.

3. *Froot-Obsfeld* (1991) intrinsic bubble: $y_t = c_\eta (\sum_{j=1}^t \eta_j)^\lambda$ with parameters $c_\eta, \lambda > 0$ satisfies, for $(\tau_{(T,t)}/T, \tau_{(T,\bar{t})}/T) \to (t, \bar{t})$, $y_{\bar{\tau}}/y_\tau = (1 + (W(\bar{t}) - W(t))/W(t))^\lambda$. Letting $\bar{\tau} = \tau + \chi$, then $g_{\tau,\chi}^{1/\lambda} = 1 + O_p(\sqrt{\frac{\chi}{\tau}})$ hence $g_{\tau,\chi} \overset{P}{\to} 1$ unless we assume that λ itself is a function of T and that $\lambda\sqrt{\chi/\tau} \not\to 0$, which requires $\lambda \to \infty$ as $T \to \infty$.

[11] In a subsequent paper, Lieberman and Phillips (2017), they extend their model to the multivariate case.

Models that allow for temporary bubble episodes

1. *(Temporary-) Explosive process*: $y_t = \rho_t y_{t-1} + x_t$ where, for $s \in [\tau, \tau+\chi]$, $\rho_s \geqslant 1+\epsilon$, $\epsilon > 0$, satisfies for all $\gamma > 1$, $\lim_{T\to\infty} \mathsf{P}(g_{\tau,\chi} > \gamma | \mathcal{I}_\tau) = 1$.

2. *Phillips-Wu-Yu* (2011) temporary near explosive bubble: $y_t = \rho_t y_{t-1} + x_t$ where, for $s \in [\tau, \tau+\chi]$, $\rho_s = e^{\phi/T^\alpha}$, $\phi > 0$ and $\alpha \in (0,1)$, satisfies, for $\chi \in (0, (T+H)\,T^{-\alpha}]$, $\lim_{T\to\infty} \mathsf{P}(g_{\tau,\chi} = e^{\phi\chi} \big| \mathcal{I}_\tau) = 1$

3. *Blanchard-Watson (1983) bubble*: $y_t = \rho_t y_{t-1} + \eta_t$ with $\rho_t = \rho\pi_t$, $\rho > 1$ and $\pi_t \overset{iid}{\sim} \mathsf{Bernoulli}(\pi)$ satisfies $\lim_{T\to\infty} \mathsf{P}(g_{\tau,\chi} = \phi | \mathcal{I}_\tau) = 1$, where $\phi = \lim_{T\to\infty} (\rho\pi)^\chi$. Bubbles arise if $\phi > 1$, i.e. if $\chi^{-1}(\rho\pi - 1)$ admits a strictly positive limit.

Although several models above allow for bubble episodes, an important remark is that, in all of the processes considered above, the limiting unconditional distributions of $y_{\tau+\chi}/y_\tau$ is degenerate under the assumption of a short-lived bubble episode $\chi/\tau \to 0$ as $T \to \infty$. This is not the case under the NERC model and hence this allows to perform conditional probabilistic statements.

A.3 Asymptotic distribution of the estimator

Let $\widehat{\rho}$ be the OLS estimator in the regression of y_t on y_{t-1} and let $\mathsf{E}(\rho_t) = \rho$. The model (1) can be written as $y_t = \rho y_{t-1} + (\rho_t - \rho)\, y_{t-1} + \eta_t$. Hence, the OLS estimator satisfies

$$\widehat{\rho} - \rho = \frac{S_{yy\rho}}{S_{yy}} + \frac{S_{y\eta}}{S_{yy}}, \tag{11}$$

where $S_{yy\rho} = \sum_{t=1}^T y_{t-1}^2 (\rho_t - \rho)$, $S_{y\eta} = \sum_{t=1}^T y_{t-1}\eta_t$ and $S_{yy} = \sum_{t=1}^T y_{t-1}^2$. Note that the asymptotic distribution of the OLS is driven by the sum with higher magnitude between $S_{yy\rho}$ and $S_{y\eta}$.

For this analysis, we introduce the following random variables:

$$V_T = \int_0^{T^{1-\alpha}} F_r^2 dW_r, \quad Z_T = \int_0^{T^{1-\alpha}} F_r^2 dr,$$
$$X_T = \int_0^{T^{1-\alpha}} F_r^{-1} dB_r, \quad Y_T = \int_0^{T^{1-\alpha}} F_r dB_r,$$

where $F_r = \exp(\phi r + \lambda W_r)$ is a Geometric Brownian Motion. We show in Appendix B.3 that

V_T, X_T, Y_T and Z_T, when centered and scaled by their standard deviations, converge weakly, as $T \to \infty$, to random variables V_θ, X_θ, Y_θ, and Z_θ, whose distribution beyond the first two moments depend on $\theta \equiv (\phi, \lambda)'$.[12] The variables X_θ and Y_θ are mixed Gaussian,[13] V_θ and Z_θ are uncorrelated random variables with zero expectation and unit variance.[14] We can now provide our result on the weak convergence of the sample moments:

Lemma 1 *Let the process y_t be defined for $t \geq 0$ by (1) under Assumptions **A** and **B**. Then as $T \to \infty$, we have:*

(i) If $c < 0$,

$$T^{-(1+\alpha)} S_{yy} \xrightarrow{p} \frac{\sigma_\eta^2}{-2c}, \quad T^{-\frac{1+\alpha}{2}} S_{y\eta} \xrightarrow{L} \mathsf{N}\left(0, \frac{\sigma_\eta^2}{-2c}\right), \quad T^{-\frac{1+\alpha}{2}} S_{yy\rho} \xrightarrow{L} \mathsf{N}\left(0, \frac{12\lambda^2\sigma_\eta^2}{c^2}\right).$$

(ii) If $c \geq 0$ then there exist $\left(\mu^{yx}, \phi_T^{yx}\right)$, functions of (ϕ, λ) and $x \in \{y\rho, \eta\}$, such that $\phi_T^{yx} \to 0$, $\phi_T^{yy\rho}/\phi_T^{y\eta} = o(1)$ and

$$T^{-\alpha}\phi_T^{yy\rho} S_{yy} \Rightarrow \frac{\sigma_\eta^2}{\lambda\mu^{yy\rho}\sqrt{c + 2\lambda^2}} X_\theta^2 Z_\theta, \quad \phi_T^{y\eta} S_{y\eta} \Rightarrow \frac{\sigma_\eta^2}{\mu^{y\eta}} X_\theta Y_\theta, \quad \phi_T^{yy\rho} S_{yy\rho} \Rightarrow \frac{\sigma_\eta^2}{\mu^{yy\rho}} X_\theta^2 V_\theta.$$

Lemma 1 implies the following: when $c < 0$ then both $S_{yy\rho}$ and $S_{y\eta}$ impact the asymptotic distributions; but when $c \geq 0$ and $\lambda \neq 0$, $S_{yy\rho}$ dominates. This setting differs markedly from that of Aue where the variance of ρ_t is of lower magnitude so $S_{y\eta}$ is the dominant term in the expansion (11). It also differs from the fixed-asymptotics framework of Hwang and Basawa (2005) where the ratio $S_{yy\rho}/S_{yy}$ diverges: making the the OLS estimator inconsistent.

[12] The random variables are defined as follows: X_θ in expression (18), Z_θ in (19), Y_θ in (29) and V_θ in (32).

[13] More specifically

$$X_\theta | W_{(\cdot)} \sim \mathsf{N}\left(0, \lim_{T\to\infty} \sqrt{\frac{2\left(\lambda^2 - \phi\right)}{e^{2\left(\lambda^2-\phi\right)T^{1-\alpha}} - 1} \int_0^{T^{1-\alpha}} F_r^{-2} dr}\right)$$
$$Y_\theta | W_{(\cdot)} \sim \mathsf{N}\left(0, \lim_{T\to\infty} \sqrt{\frac{2c}{e^{2cT^{1-\alpha}} - 1} \int_0^{T^{1-\alpha}} F_r^{2} dr}\right)$$

where both limiting conditional variances present unit unconditional expectations.

[14] Matsumoto and Yor (2005), Theorem 7.4, show how the distribution of Z_θ can be expressed (for some values of the parameters) in terms of transforms of Brownian motions involving a Gamma variable.

This is not the case here as Proposition 3 shows. When $c<0$, the asymptotic distribution of the OLS estimator $\widehat{\rho}-\rho$ is comparable to the results of PM and Aue, for whom $T^{\frac{1+\alpha}{2}}\left(\widehat{\rho}-\rho\right)\overset{L}{\to}\mathsf{N}\left(0,-2\phi\right)$: the presence of the stochastic root does not affect the asymptotic normality of $\widehat{\rho}$ nor the rate of convergence; the only difference is that the asymptotic variance is increased by λ^2 here.

Proposition 3 presents several key differences with the existing literature on near unit roots and random coefficients when $c\geq 0$ and the results are new. Here the OLS estimator converges more slowly than under the constant parameter AR(1) : it does not achieve the $O_p\left(T^{-1}\right)$ of unit root processes or the exponential rate of PM for whom $\left(2\phi\right)^{-1}T^{\alpha}e^{\phi T^{1-\alpha}}\left(\widehat{\rho}-\rho\right)$ tends to a standard Cauchy variable. Convergence can be arbitrarily slow here if α is close to zero: the limit $\alpha\to 0$ corresponds to the fixed-asymptotics of Hwang and Basawa (2005) where the estimator is shown to be inconsistent. Also, the limiting distribution is expressed, as in PM or Aue, as the ratio of two uncorrelated random variables. Yet, V_θ and Z_θ are not normal. This implies that V_θ/Z_θ does not define a Cauchy variable contrary to the limiting distribution in PM.

Proposition 3 shows that $\widehat{\rho}$ allows to estimate $\phi+\lambda^2/2$ consistently when $c<0$ since the higher order term in expression (3) is $o\left(T^{\frac{1+\alpha}{2}}\right)$, but this is not the case for $c\geq 0$ as the convergence of $\widehat{\rho}$ is then too slow.

Proposition 3 also shows that under the NERC model, the unit root problem does not exist when $c\geq 0$ since the asymptotic distribution does not exhibit the usual knife-edge issue as c tends to zero from above (see Berkes et al., 2009, for a discussion).

A.3.1 Power

We derive in here the power of our proposed test statistic $\delta_{\theta_0,T}=\left[T^{\frac{1+\alpha}{2}}1_{\{c<0\}}+T^{\alpha}1_{\{c\geq 0\}}\right]\left(\widehat{\rho}-\rho\right)$ for the null $\mathsf{H}_0:\theta=\theta_0$ with $\theta=\left(\phi,\lambda\right)$. Then under $\mathsf{H}_1:\theta=\theta_1\neq\theta_0$ and, as $T\to\infty$,

$$\delta_{\theta_0,T}\underset{\mathsf{H}_1}{=}\begin{cases}O_p\left(T^{\frac{1-\alpha}{2}}\right), & \text{if } c_0<0;\\ \phi_1-\phi_0+\left(\lambda_1^2-\lambda_0^2\right)/2+o_p\left(1\right), & \text{if } c_0\geq 0 \text{ and } c_1<0;\\ \phi_1-\phi_0+\left(\lambda_1^2-\lambda_0^2\right)/2+\lambda_1\sqrt{c_1+2\lambda_1^2}\dfrac{V}{Z}+o_p\left(1\right), & \text{if } c_0\geq 0 \text{ and } c_1\geq 0.\end{cases}\tag{12}$$

where $c_1=\phi_1+\lambda_1^2$.

Expression (12) shows that the test based on the OLS estimator is asymptotically powerful under the null (ϕ_0, λ_0) such that $c_0 < 0$. Yet the test statistic does not diverge asymptotically (so the test has non-unit asymptotic power) when $c_0 \geq 0$. This holds irrespective of the alternative hypothesis within the class considered. It is also interesting to notice that the corollary sheds light on the reason why the simulations of Evans (1991) and Charemza and Deadman (1995) find that the Dickey-Fuller test has non-trivial yet low power in the presence of periodically collapsing bubbles.

Proof of Expression (12) We write

$$\widehat{\rho} - \mathsf{E}_{\mathsf{H}_0}[\rho_t] = \left(\widehat{\rho} - \mathsf{E}_{\mathsf{H}_1}[\rho_t]\right) + \left(\mathsf{E}_{\mathsf{H}_1}[\rho_t] - \mathsf{E}_{\mathsf{H}_0}[\rho_t]\right).$$

and consider the two elements of the sum in turn. The null and alternative hypotheses are local to each other:

$$\mathsf{E}_{\mathsf{H}_1}[\rho_t] - \mathsf{E}_{\mathsf{H}_0}[\rho_t] = \frac{\phi_1 - \phi_0 + \frac{1}{2}\left(\lambda_1^2 - \lambda_0^2\right)}{T^\alpha} + o\left(T^{-\alpha}\right),$$

hence $T^{\frac{1+\alpha}{2}}\left(\mathsf{E}_{\mathsf{H}_1}[\rho_t] - \mathsf{E}_{\mathsf{H}_0}[\rho_t]\right)$ diverges but $T^\alpha\left(\mathsf{E}_{\mathsf{H}_1}[\rho_t] - \mathsf{E}_{\mathsf{H}_0}[\rho_t]\right) = \phi_1 - \phi_0 + \left(\lambda_1^2 - \lambda_0^2\right)/2 + O\left(T^{-\alpha}\right)$ does not. Also, under the alternative, $T^{\frac{1+\alpha}{2}}\left(\widehat{\rho} - \mathsf{E}_{\mathsf{H}_1}[\rho_t]\right)$ diverges only if $\phi_1 + \lambda_1^2 \geq 0$ but $T^\alpha\left(\widehat{\rho} - \mathsf{E}_{\mathsf{H}_1}[\rho_t]\right)$ does not diverge. Finally, if both $T^{\frac{1+\alpha}{2}}\left(\widehat{\rho} - \mathsf{E}_{\mathsf{H}_1}[\rho_t]\right)$ and $T^{\frac{1+\alpha}{2}}\left(\mathsf{E}_{\mathsf{H}_1}[\rho_t] - \mathsf{E}_{\mathsf{H}_0}[\rho_t]\right)$ diverge, their sum is $O_p\left(T^{\frac{1-\alpha}{2}}\right)$ so they do not cancel each other. To conclude, $\tau_{0,T}$ diverges under H_1 only if $\phi_0 + \lambda_0^2 < 0$, irrespective of (ϕ_1, λ_1).

B Proofs

We collect here the proofs to the propositions.

B.1 Proof of Proposition 1

Let y_t be defined for $t \geq 0$ by expression (1) under Assumptions **A** and **B**. Then y_0 plays no role asymptotically, and set it to zero in the following:

$$
\begin{aligned}
y_t &= \sum_{i=0}^{t-1} \left(\prod_{j=0}^{i-1} \rho_{t-j} \right) \eta_{t-i} = \sum_{i=1}^{t} \left(\prod_{j=i+1}^{t} \rho_j \right) \eta_i \\
&= \sum_{i=1}^{t} \exp \left\{ \frac{(t-i)T^{-\alpha/2}\phi + \lambda(U_t - U_i)}{T^{\alpha/2}} \right\} \eta_i,
\end{aligned}
$$

where we set $\prod_{j=0}^{-1} \rho_j \equiv 1$ and U_t denotes the partial sum $U_t \equiv \sum_{k=1}^{t} u_k$. We evaluate the increment $y_t - y_0$ using the blocking method of Phillips and Magdalinos (2004). Letting, for $t = 1$ to T, $t = \lfloor jT^\alpha \rfloor + k$ for $j = 0, \cdots, \lfloor T^{1-\alpha} \rfloor - 1$, and $k = 1, \cdots, \lfloor T^\alpha \rfloor$, and letting $k = \lfloor pT^\alpha \rfloor$ for some $p \in [0, 1]$, we can write

$$
\begin{aligned}
& T^{-\alpha/2} y_{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor} \\
&= \sigma_\eta \sum_{i=1}^{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor} \exp \left\{ \frac{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor - i}{T^\alpha} \phi + \lambda \frac{U_{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor} - U_i}{T^{\alpha/2}} \right\} \frac{\eta_i}{\sqrt{\sigma_\eta^2 T^\alpha}} \\
&= \sigma_\eta \int_0^{j+p} \exp \left\{ \frac{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor - \lfloor sT^\alpha \rfloor}{T^\alpha} \phi + \lambda \frac{U_{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor} - U_{\lfloor sT^\alpha \rfloor}}{T^{\alpha/2}} \right\} dB_{T^\alpha}(s),
\end{aligned}
$$

using Proposition A1 in Phillips and Magdalinos (2004) in the last equality, where

$$
B_{T^\alpha}(s) \equiv \frac{1}{\sigma_\eta T^{\alpha/2}} \sum_{i=1}^{\lfloor sT^\alpha \rfloor} \eta_i. \tag{13}
$$

When applying the Functional Central Limit Theorem (FCLT) to the process W_{T^α} defined by $W_{T^\alpha}(s) \equiv T^{-\alpha/2} U_{\lfloor sT^\alpha \rfloor}$ $(0 \leq s \leq T^{1-\alpha})$, we obtain that W_{T^α} converges in distribution, as $T \to \infty$, to a standard Brownian motion (BM) on $\mathbb{R}_+$ that we denote by W.

The FCLT also implies that the process B_{T^α} defined in (13) converges in distribution, as $T \to \infty$, to a BM on $\mathbb{R}_+$, say B, which, by assumption on the sequences (u_i) and (η_j), is independent of W.

Then we can deduce, using e.g. Theorem 8.3.1 in Liptser and Shiryaev (1989), that

$$\int_0^{j+p} \exp\left\{\phi \frac{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor - \lfloor sT^\alpha \rfloor}{T^\alpha} + \lambda \frac{U_{\lfloor jT^\alpha \rfloor + \lfloor pT^\alpha \rfloor} - U_{\lfloor sT^\alpha \rfloor}}{T^{\alpha/2}}\right\} dB_{T^\alpha}(s)$$

converges, as $T \to \infty$, to

$$\int_0^r \exp\left\{\phi(r-s) + \lambda\left(W_r - W_s\right)\right\} dB_s, \quad \text{with} \quad r = j + p,$$

where W, B are the standard Brownian motions defined previously. Corollary 1 follows since the proof above also holds when $\alpha = 1$. □

B.2 Proof of Proposition 2

Consider the projection

$$y_{t+k} = \exp\left\{\frac{k\phi + \lambda T^{\alpha/2} \sum_{j=1}^{k} u_{t+j}}{T^\alpha}\right\} y_t + \sum_{i=1}^{k} \exp\left\{\frac{(k-i)\,\phi + \lambda T^{\alpha/2} \sum\limits_{j=i+1}^{k} u_{t+j}}{T^\alpha}\right\} \eta_{t+i}.$$

Let $(r, s) \in \left(0, T^{1-\alpha}\right)$, with $s > 0$, then

$$\begin{aligned} \frac{y_{[T^\alpha(r+s)]}}{y_{[T^\alpha r]}} &= \exp\left\{\frac{[T^\alpha s]\,\phi + \lambda T^{\alpha/2} \sum_{j=[T^\alpha r]+1}^{[T^\alpha(r+s)]} u_j}{T^\alpha}\right\} \\ &\quad + \frac{1}{y_{[T^\alpha r]}} \sum_{i=[T^\alpha r]+1}^{[T^\alpha(r+s)]} \exp\left\{\frac{\left([T^\alpha\,(r+s)] - [T^\alpha r] - i\right)\phi + \lambda T^{\alpha/2} \sum\limits_{j=i+1}^{[T^\alpha(r+s)]} u_j}{T^\alpha}\right\} \eta_i. \end{aligned}$$

Proposition 1 implies that, as $T \to \infty$, $\exp\left\{\frac{[T^\alpha s]\phi + \lambda T^{\alpha/2}\sum_{j=[T^\alpha r]+1}^{[T^\alpha(r+s)]} u_j}{T^\alpha}\right\} \Rightarrow \exp\left\{s\phi + \lambda\left(W_{r+s} - W_r\right)\right\}$, and

$$T^{-\alpha/2}\sum_{i=[T^\alpha r]+1}^{[T^\alpha(r+s)]}\exp\left\{\frac{\left([T^\alpha(r+s)] - [T^\alpha r] - i\right)\phi + \lambda T^{\alpha/2}\sum\limits_{j=i+1}^{[T^\alpha(r+s)]} u_j}{T^\alpha}\right\}\eta_i$$

$$\Rightarrow K_{\phi,\lambda}(r+s) - e^{\phi s + \lambda(W_{r+s} - W_r)}K_{\phi,\lambda}(r) = \int_r^{r+s} e^{\phi(r+s-u) + \lambda(W_{r+s} - W_u)}dB_u,$$

We see that $K_{\phi,\lambda}(r+s) - e^{\phi s + \lambda(W_{r+s} - W_r)}K_{\phi,\lambda}(r)$ can be written using integrals with respect to dW_u and dB_u for $u \geq r$, and hence is independent of $K_{\phi,\lambda}(r)$. Conditionally on W_u, $r \leq u \leq r+s$, it is normal with zero expectation. It follows that we can define[15] a Cauchy variable C such that

$$\frac{y_{[T^\alpha(r+s)]}}{y_{[T^\alpha r]}} \Rightarrow e^{\phi s + \lambda(W_{r+s} - W_r)} + D_\theta(r,s)\,C, \tag{14}$$

where $\theta = (\phi, \lambda)'$ and, using the notation of Section A.3,

$$D_\theta(r,s) = \left(\frac{\int_r^{r+s} e^{2\phi(r+s-u) + 2\lambda(W_{r+s} - W_u)}du}{\int_0^r e^{2\phi(r-u) + 2\lambda(W_r - W_u)}du}\right)^{1/2} = \frac{F_{r+s}}{F_r}\sqrt{\frac{\int_r^{r+s} F_u^{-2}du}{\int_0^r F_u^{-2}du}};$$

$$C = \frac{\left(\int_r^{r+s} e^{2\phi(r+s-u) + 2\lambda(W_{r+s} - W_u)}du\right)^{-1/2}\int_r^{r+s} e^{\phi(r+s-u) + \lambda(W_{r+s} - W_u)}dB_u}{\left(\int_0^r e^{2\phi(r-u) + 2\lambda(W_r - W_u)}du\right)^{-1/2}\int_0^r e^{\phi(r-u) + \lambda(W_r - W_u)}dB_u}.$$

$D_\theta(r,s)$ is independent of B, hence the second result in the proposition, conditionally on $\mathcal{I}_t$.

B.3 Proof of Proposition 3

We define $\rho = \mathsf{E}(\rho_t)$. The OLS estimator given by $\widehat{\rho} = \frac{\sum_{t=1}^T y_{t-1}y_t}{\sum_{t=1}^T y_{t-1}^2}$ satisfies

$$\widehat{\rho} - \rho = \frac{\sum_t y_{t-1}^2(\rho_t - \rho)}{\sum_t y_{t-1}^2} + \frac{\sum_t y_{t-1}\eta_t}{\sum_t y_{t-1}^2}. \tag{15}$$

[15]Since C is Cauchy for all realizations of W.

Hence the asymptotic distribution of the estimator is driven by the term with higher magnitude between $\sum_t y_{t-1}^2 (\rho_t - \rho)$ and $\sum_t y_{t-1}\eta_t$. So we need to study the three sums appearing in the expression of the OLS estimator. Recall that $c = \phi + \lambda^2$. We will consider two cases depending on the sign of c, first $c < 0$ and then $c \geq 0$.

B.3.1 Case $c < 0$

Recall that Proposition 1 gives $T^{-\alpha/2} y_{\lfloor rT^\alpha \rfloor} \Rightarrow K_{\phi,\lambda}(r) \sim \mathsf{N}\left(0, \frac{e^{2cr}-1}{2c}\sigma_\eta^2\right)$ as $T \to \infty$. Letting, for $c < 0$, $K^*_{\phi,\lambda}(r) = e^{cr} K^*_{\phi,\lambda}(0) + K_{\phi,\lambda}(r)$ such that $K^*_{\phi,\lambda}(r) \sim \mathsf{N}\left(0, -\frac{\sigma_\eta^2}{2c}\right)$ and is stationary. Hence the proof in PM, Section 3, applies here also. We can deduce from expression (5), via the Law of Large Numbers (LLN), that $T^{-(1+\alpha)} \sum_{t=1}^T y_t^2 \Rightarrow \mathsf{E}[K^*_{\phi,\lambda}(r)^2] = \frac{-\sigma_\eta^2}{2c}$ and that $T^{-\frac{1+\alpha}{2}} \sum_{t=1}^T y_{t-1}\eta_t \Rightarrow \mathsf{N}\left(0, -\frac{\sigma_\eta^4}{2c}\right)$.

The result concerning $\sum_{t=1}^T y_{t-1}^2 (\rho_t - \rho)$ follows similarly. Indeed, $\mathsf{V}\left[T^{\alpha/2}(\rho_t - \rho)\right] \to \lambda^2$ as $T \to \infty$, so we define the martingale difference sequence $\xi_t \equiv T^{-\frac{1+\alpha}{2}} y_{t-1}^2 (\rho_t - \rho)$ which admits a conditional variance satisfying $\sum_{t=1}^T \mathsf{E}_{t-1}\left(\xi_t^2\right) = \frac{\lambda^2}{T^{1+\alpha}} \sum_{t=1}^T y_{t-1}^4 \Rightarrow \frac{3\lambda^2\sigma_\eta^4}{4c^2}$, using the consistency of the empirical estimator of the kurtosis. Then a martingale analogue of the Lindberg condition (see e.g. Pollard, 1984) ensures that, as $T \to \infty$,

$$T^{-\frac{1+\alpha}{2}} \sum_{t=1}^T y_{t-1}^2 (\rho_t - \rho) \xrightarrow{L} \mathsf{N}\left(0, \frac{3\lambda^2\sigma_\eta^4}{4c^2}\right).$$

Combining these results as in (15) gives Theorem 3(i).

B.3.2 Case $c \geq 0$

The proof follows the main schemata given in PM, and retain their notation, namely

$$\kappa_T = T^\alpha \left\lfloor T^{1-\alpha} \right\rfloor \quad \text{and} \quad q_T = T^{1-\alpha} - \left\lfloor T^{1-\alpha} \right\rfloor. \tag{16}$$

We need to derive the asymptotic behaviors of the sample variance and covariances of y_t. Throughout, $\mathsf{D}(0,1)$ denotes a generic distribution with zero expectation and unit variance.

$\triangleright$ **Sample variance of y_t** We first consider the sample variance of y_t and prove the following lemma.

Lemma 2 *Define*

$$\phi_{T^{1-\alpha}} = \begin{cases} \left(\exp\left[2cT^{1-\alpha}\right] - 1\right) / (2c), & \text{if } c \neq 0; \\ T^{1-\alpha}, & \text{if } c = 0; \end{cases}$$

$$\psi_{T^{1-\alpha}} = \begin{cases} \frac{1}{2}\exp\left[2(c+\lambda^2)T^{1-\alpha}\right] / \sqrt{(c+2\lambda^2)(c+\lambda^2)}, & \text{if } \ c \neq 0; \\ \frac{1}{2\sqrt{2}}\exp\left[2\lambda^2 T^{1-\alpha}\right] / \lambda^2, & \text{if } \ c = 0; \end{cases} \quad \textit{and}$$

$$\varphi_{T^{1-\alpha}} = \begin{cases} \left[2\left(\phi - \lambda^2\right)\right]^{-1}, & \text{if } \lambda^2 - \phi < 0; \\ \sqrt{T^{1-\alpha}}, & \text{if } \lambda^2 - \phi = 0; \\ \exp\left[(\lambda^2 - \phi)T^{1-\alpha}\right] / \sqrt{2\left(\lambda^2 - \phi\right)}, & \text{if } \lambda^2 - \phi > 0. \end{cases}$$

Then, as $T \to \infty$,

$$\sigma_\eta^{-2} T^{-2\alpha} \psi_{T^{1-\alpha}}^{-1} \varphi_{T^{1-\alpha}}^{-2} \sum_{t=1}^{T} y_t^2 \Rightarrow X_\theta^2 Z_\theta, \tag{17}$$

where the random variables X_θ and Z_θ are defined, respectively, by

$$\sigma_\eta^{-1} \frac{\varphi_{\lfloor T^{1-\alpha} \rfloor}^{-1}}{T^{\alpha/2}} \sum_{t=1}^{\lfloor \kappa_T \rfloor} \exp\left(-\frac{\phi}{T^\alpha} t - \frac{\lambda}{T^{\alpha/2}} U_t\right) \eta_t \Rightarrow X_\theta \sim \mathsf{D}\left(0, 1\right), \tag{18}$$

with $U_t = \sum_{k=1}^{t} u_k$, and

$$\psi_{\lfloor T^{1-\alpha} \rfloor}^{-1} \left(\int_0^{\lfloor T^{1-\alpha} \rfloor} e^{2(\phi s + \lambda W_s)} ds - \phi_{\lfloor T^{1-\alpha} \rfloor} \right) \Rightarrow Z_\theta \sim \mathsf{D}\left(0, 1\right). \tag{19}$$

Proof of Lemma 2. We write

$$T^{-2\alpha} \sum_{t=1}^{T} y_t^2 = T^{-2\alpha} S_{1T} + T^{-2\alpha} S_{2T} + O_p\left(T^{-\alpha}\right), \tag{20}$$

with $S_{1T} \equiv \sum_{j=0}^{\lfloor T^{1-\alpha} \rfloor - 1} \sum_{k=1}^{\lfloor T^\alpha \rfloor} y_{\lfloor T^\alpha j \rfloor + k}^2$, and $S_{2T} \equiv \sum_{t=\lfloor \kappa_T \rfloor}^{T} y_t^2$. Note that the index of the last summation term in the definition of S_{1T}, given by $\lfloor \kappa_T - T^\alpha \rfloor + \lfloor T^\alpha \rfloor$, is bounded by $\lfloor \kappa_T \rfloor - 1 \leq \lfloor \kappa_T - T^\alpha \rfloor + \lfloor T^\alpha \rfloor \leq \lfloor \kappa_T \rfloor$.

In the following, we study the asymptotic behavior of these two sums S_{1T} and S_{2T}, proving that, as $T \to \infty$,

$$\sigma_\eta^{-2} T^{-2\alpha} \psi_{\lfloor T^{1-\alpha} \rfloor}^{-1} \varphi_{\lfloor T^{1-\alpha} \rfloor}^{-2} S_{1T} \Rightarrow X_\theta^2 Z_\theta, \tag{21}$$

where the random variables X_θ and Z_θ are defined in Lemma 2, and

$$\begin{aligned} T^{-2\alpha} S_{2T} &= \left(\sigma_\eta \int_0^{T^{1-\alpha}} e^{-\phi s - \lambda W_s} dB_s \right)^2 \int_0^{T^{1-\alpha}} e^{2(\phi s + \lambda W_s)} ds \\ &\quad - \frac{1}{T^{2\alpha}} \left(\frac{\int_0^{T^{1-\alpha}} e^{-\phi s - \lambda W_s} dB_s}{\int_0^{\lfloor T^{1-\alpha} \rfloor} e^{-\phi s - \lambda W_s} dB_s} \right)^2 S_{1T} + o_p(1). \end{aligned} \tag{22}$$

Combining (20), (22) and the asymptotic equivalence $\psi_{T^{1-\alpha}}^{-1} \psi_{\lfloor T^{1-\alpha} \rfloor} = 1$ shows $T^{-2\alpha} \sum_{t=1}^T y_t^2 = T^{-2\alpha} S_{1T} + o_p(1)$. Together with (21), this allows to conclude to Lemma 2. ■

We now turn to the study of S_{1T} and S_{2T}, proving expressions (21) and (22) respectively.

Proof of Expression (21). Notice that

$$\begin{aligned} y_k &= \sum_{i=0}^{k-1} \exp\left(\frac{\phi}{T^\alpha} i + \frac{\lambda}{T^{\alpha/2}} \sum_{j=k-i+1}^{t} u_j \right) \eta_{k-i} \\ &= \exp\left(\frac{\phi}{T^\alpha} k + \frac{\lambda}{T^{\alpha/2}} U_k \right) \sum_{i=1}^{k} \exp\left(-\frac{\phi}{T^\alpha} i - \frac{\lambda}{T^{\alpha/2}} U_i \right) \eta_i, \end{aligned}$$

so

$$\begin{aligned} \sum_{k=1}^t y_k^2 &= \sum_{k=1}^t \exp\left(\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k \right) \left[\sum_{i=1}^k \exp\left(-\frac{\phi}{T^\alpha} i - \frac{\lambda}{T^{\alpha/2}} U_i \right) \eta_i \right]^2 \\ &= \sum_{k=1}^t \exp\left(\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k \right) \\ &\quad \times \left[\sum_{i=1}^t \exp\left(-\frac{\phi}{T^\alpha} i - \frac{\lambda}{T^{\alpha/2}} U_i \right) \eta_i - \sum_{i=k+1}^t \exp\left(-\frac{\phi}{T^\alpha} i - \frac{\lambda}{T^{\alpha/2}} U_i \right) \eta_i \right]^2 \\ &= \left(\sum_{k=1}^t \exp\left(\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k \right) \right) \left[\sum_{i=1}^t \exp\left(-\frac{\phi}{T^\alpha} i - \frac{\lambda}{T^{\alpha/2}} U_i \right) \eta_i \right]^2 + R_t \end{aligned}$$

where

$$R_t = \sum_{k=1}^{t} \exp\left(\frac{2\phi}{T^\alpha}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)\left[\sum_{i=1}^{k} \exp\left(-\frac{\phi}{T^\alpha}i - \frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i\right]^2 - \left(\sum_{k=1}^{t} \exp\left(\frac{2\phi}{T^\alpha}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)\right)\left[\sum_{i=1}^{t} \exp\left(-\frac{\phi}{T^\alpha}i - \frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i\right]^2.$$

Therefore we obtain

$$S_{1T} = \left(\sum_{k=1}^{\lfloor \kappa_T \rfloor} \exp\left(\frac{2\phi}{T^\alpha}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)\right)\left[\sum_{i=1}^{\lfloor \kappa_T \rfloor} \exp\left(-\frac{\phi}{T^\alpha}i - \frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i\right]^2 + R_{\lfloor \kappa_T \rfloor}. \tag{23}$$

where $R_{\lfloor \kappa_T \rfloor}$ can be shown to be negligible in this expression (see the Supplementary Appendix, where we provide the proof).

In expression (23) we see that S_{1T} can be expressed as a product. The second element on the right-hand side admits the limit:

$$\sigma_\eta^{-1}\varphi_{\lfloor T^{1-\alpha} \rfloor}^{-1} T^{-\alpha/2} \sum_{i=1}^{\lfloor \kappa_T \rfloor} \exp\left(-\frac{\phi}{T^\alpha}i - \frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i \Rightarrow X_\theta. \tag{24}$$

Indeed, we can write $\frac{1}{T^{\alpha/2}}\sum_{i=1}^{\lfloor \kappa_T \rfloor} \exp\left(-\frac{\phi}{T^\alpha}i - \frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i = \sigma_\eta \int_0^{\lfloor T^{1-\alpha} \rfloor} e^{-\phi s - \lambda W_{T^\alpha}(s)} dB_{T^\alpha}(s) + o_p(1)$, where B_{T^α} and W_{T^α} are defined in Section B.1. When $\lambda^2 < \phi$, it converges weakly to $\sigma_\eta \int_0^\infty e^{-\phi s - \lambda W_s} dB(s)$.

When $\lambda^2 \geq \phi$, the stochastic integral is not defined, but since $\int_0^{\lfloor T^{1-\alpha} \rfloor} e^{-(\phi s + \lambda W_s)} dB_s$ is mixed-normally distributed, it will be enough to scale it by its standard deviation, using

$$\mathsf{V}\left[\int_0^{\lfloor T^{1-\alpha} \rfloor} e^{-(\phi s + \lambda W_s)} dB_s\right] = \begin{cases} \frac{e^{2\left(\lambda^2-\phi\right)\lfloor T^{1-\alpha} \rfloor}-1}{2(\lambda^2-\phi)}, & \text{if } \lambda^2 > \phi; \\ \lfloor T^{1-\alpha} \rfloor, & \text{if } \lambda^2 = \phi; \\ \frac{1-e^{-2\left(\phi-\lambda^2\right)\lfloor T^{1-\alpha} \rfloor}}{2(\phi-\lambda^2)}, & \text{if } \lambda^2 < \phi. \end{cases} \tag{25}$$

Hence, whether $\lambda^2 < \phi$ or $\lambda^2 \geq \phi$, the limit X_θ has zero expectation and unit variance, which we write $X_\theta \sim \mathsf{D}(0, 1)$.

Now, regarding the first element on the right-hand side of expression (23), we have

$$\psi_{\lfloor T^{1-\alpha}\rfloor}^{-1} T^{-\alpha} \sum_{k=1}^{\lfloor \kappa_T \rfloor} \exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right) = \psi_{\lfloor T^{1-\alpha}\rfloor}^{-1}\left(\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\phi s + 2\lambda W_s} ds - \phi_{\lfloor T^{1-\alpha}\rfloor}\right) + o_p(1),$$

with $\psi_{\lfloor T^{1-\alpha}\rfloor}^{-1}\phi_{\lfloor T^{1-\alpha}\rfloor}$ tending to 0 as $T \to \infty$. Note that the expectation of $\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\phi s + \lambda W_s)} ds$ is given by

$$\phi_{\lfloor T^{1-\alpha}\rfloor} \equiv \mathsf{E}\left[\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\phi s + 2\lambda W_s} ds\right] = \int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2cs} ds = \begin{cases} \frac{e^{2c\lfloor T^{1-\alpha}\rfloor}-1}{2c}, & \text{if } c \neq 0; \\ \lfloor T^{1-\alpha}\rfloor, & \text{if } c = 0; \end{cases} \tag{26}$$

and that the rate $\psi_{\lfloor T^{1-\alpha}\rfloor}$ comes from the second moment of $\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\phi s + \lambda W_s)} ds$. Indeed, straightforward computations lead, for $c \geq 0$, to

$$\mathsf{E}\left[\left(\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\phi s + \lambda W_s)} ds\right)^2\right] = \mathsf{E}\left[\left(\int_0^{\lfloor T^{1-\alpha}\rfloor}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\phi(s+r) + 2\lambda(W_s + W_r)} ds dr\right)\right]$$
$$= \int_0^{\lfloor T^{1-\alpha}\rfloor}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\phi(s+r) + 2\lambda^2(s + 2\min(r,s) + r)} ds dr$$
$$= \int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2cr}\int_0^{r} e^{2(c+2\lambda^2)s} ds dr + \int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(c+2\lambda^2)r}\int_r^{\lfloor T^{1-\alpha}\rfloor} e^{2cs + 2(c+2\lambda^2)r} ds dr$$
$$= \begin{cases} \frac{e^{4\left(c+\lambda^2\right)\lfloor T^{1-\alpha}\rfloor}}{4(c+2\lambda^2)(c+\lambda^2)} - \frac{e^{2cT^{1-\alpha}}}{2c(c+2\lambda^2)} + \frac{1}{4c(c+\lambda^2)} = \frac{e^{4\left(c+\lambda^2\right)\lfloor T^{1-\alpha}\rfloor}}{4(c+2\lambda^2)(c+\lambda^2)} + O\left(e^{2cT^{1-\alpha}}\right), & \text{if } c \neq 0; \\ \frac{e^{4\lambda^2\lfloor T^{1-\alpha}\rfloor}}{8\lambda^4} + O\left(T^{1-\alpha}\right), & \text{if } c = 0. \end{cases}$$

Now, Theorem 7.4 of Matsumoto and Yor (2005) implies that $\left(\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\phi s + \lambda W_s)} ds\right)^2$ divided by its expectation converges weakly to a random variable as $T \to \infty$. The continuous mapping theorem implies that the square root thereof also admits a weak limit. Hence there exists Z_θ with unit variance and zero expectation such that $\psi_{\lfloor T^{1-\alpha}\rfloor}^{-1}\left(\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\phi s + \lambda W_s)} ds - \phi_{\lfloor T^{1-\alpha}\rfloor}\right) \Rightarrow Z_\theta \sim \mathsf{D}(0,1)$. From this last result, together with (24) and (23), we deduce (21). ■

Proof of Expression (22). We have $\frac{1}{T^{2\alpha}} S_{2T} = \frac{1}{T^{2\alpha}} \sum_{j=0}^{T-[\kappa_T]} y_{j+[\kappa_T]}^2 = \int_0^{q_T} \left(\frac{1}{T^{\alpha/2}} y_{[\kappa_T]+[T^{\alpha}p]}\right)^2 dp + O_p\left(T^{-2\alpha}\right)$, where κ_T and q_T are defined in (16). For all $j = 0, ..., \lfloor T^{1-\alpha}\rfloor - 1$, as $T \to \infty$,

$T^{-\alpha/2} y_{[T^\alpha j]+[T^\alpha p]} \Rightarrow \sigma_\eta e^{\phi(j+p)+\lambda W_{j+p}} \int_0^{j+p} e^{-\phi s-\lambda W_s} dB_s$. Then it comes

$$\begin{aligned}
\frac{1}{T^{2\alpha}} S_{2T} &= \int_0^{q_T} e^{2\phi\left(\lfloor T^{1-\alpha}\rfloor+p\right)+2\lambda W_{\lfloor T^{1-\alpha}\rfloor+p}} \left(\sigma_\eta \int_0^{\lfloor T^{1-\alpha}\rfloor+p} e^{-\phi s-\lambda W_s} dB_s\right)^2 dp + o_p(1) \\
&= \left(\sigma_\eta \int_0^{\lfloor T^{1-\alpha}\rfloor+q_T} e^{-\phi s-\lambda W_s} dB_s\right)^2 \int_0^{q_T} e^{2\left(\phi\left(\lfloor T^{1-\alpha}\rfloor+s\right)+\lambda W_{\lfloor T^{1-\alpha}\rfloor+s}\right)} ds + o_p(1) \\
&= \left(\sigma_\eta \int_0^{\lfloor T^{1-\alpha}\rfloor+q_T} e^{-\phi s-\lambda W_s} dB_s\right)^2 \\
&\times \left(\int_0^{\lfloor T^{1-\alpha}\rfloor+q_T} e^{2(\phi s+\lambda W_s)} ds - \int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\phi s+\lambda W_s)} ds\right) + o_p(1),
\end{aligned}$$

hence S_{2T} satisfies (22). ■

$\triangleright$ **Sample covariances of y_t** We now prove that the covariance terms satisfy the following:

Lemma 3 *We have, as $T \to \infty$,*

$$\sigma_\eta^{-2} T^{-\alpha} \varphi_{T^{1-\alpha}}^{-1} \phi_{T^{1-\alpha}}^{-1} \sum_{t=1}^{T} y_{t-1}\eta_t \Rightarrow X_\theta Y_\theta \tag{27}$$

and

$$\sigma_\eta^{-2} \lambda^{-1} \varphi_{T^{1-\alpha}}^{-2} \varkappa_{T^{1-\alpha}}^{-1} T^{-\alpha} \sum_{t=1}^{T} y_{t-1}^2 (\rho_t - \rho) \Rightarrow X_\theta^2 V_\theta, \tag{28}$$

where $X_\theta \sim \mathsf{D}(0,1)$, $Y_\theta \sim \mathsf{D}(0,1)$, $V_\theta \sim \mathsf{D}(0,1)$,
and with $\varkappa_{T^{1-\alpha}} = \frac{1}{2}\exp\left[2(c+\lambda^2)T^{1-\alpha}\right]/\sqrt{c+\lambda^2}$ and $(\varphi_{T^{1-\alpha}}, \phi_{T^{1-\alpha}})$ given in Lemma 2.

Proof of Lemma 3. We start with the proof of Expression (27).

Note that

$$\mathsf{V}\left(\int_0^{T^{1-\alpha}} e^{\phi s+\lambda W_s} dB_s\right) = \mathsf{E}\left[\left(\int_0^{T^{1-\alpha}} e^{\phi s+\lambda W_s} dB_s\right)^2\right] = \mathsf{E}\left[\int_0^{T^{1-\alpha}} e^{2\phi s+2\lambda W_s} ds\right] = \phi_{T^{1-\alpha}},$$

using in this last equality the computation of $\phi_{T^{1-\alpha}}$ made in expression (26). So we have

$$\phi_{T^{1-\alpha}}^{-1} \int_0^{T^{1-\alpha}} e^{\phi s+\lambda W_s} dB_s \Rightarrow Y_\theta \sim \mathsf{D}(0,1). \tag{29}$$

Hence we can write

$$\frac{\varphi_{T^{1-\alpha}}^{-1}\phi_{T^{1-\alpha}}^{-1}}{T^{\alpha}}\sum_{t=1}^{T}y_{t-1}\eta_t=\left(\sigma_\eta\varphi_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}}e^{-(\phi s+\lambda W_s)}dB_s\right)\left(\sigma_\eta\phi_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}}e^{\phi r+\lambda W_r}dB_r\right)+I_T,$$

where I_T can be shown to be negligible, referring to PM. Then (27) follows.

We now prove Expression (28). When $\lambda\neq 0$, the summation $\sum_{t=1}^{T}y_{t-1}^2(\rho_t-\rho)$ can be expressed as

$$\begin{aligned}\sum_{t=0}^{T-1}y_t^2(\rho_{t+1}-\rho)&=\sum_{t=1}^{T-1}\exp\left(\frac{2\phi}{T^{\alpha}}t+\frac{2\lambda}{T^{\alpha/2}}U_t\right)\left[\sum_{i=1}^{t}\exp\left(-\frac{\phi}{T^{\alpha}}i-\frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i\right]^2\Delta U_{t+1}^{\rho}\\&=\left[\sum_{i=1}^{T}\exp\left(-\frac{\phi}{T^{\alpha}}i-\frac{\lambda}{T^{\alpha/2}}U_i\right)\eta_i\right]^2\sum_{k=1}^{T-1}\exp\left(\frac{2\phi}{T^{\alpha}}k+\frac{2\lambda}{T^{\alpha/2}}U_k\right)\Delta U_{k+1}^{\rho}+T^{-\alpha/2}R_t^*,\end{aligned}\tag{30}$$

where $\Delta U_{k+1}^{\rho}\equiv U_{k+1}^{\rho}-U_k^{\rho}$ with $U_t^{\rho}=\sum_{j=1}^{t}(\rho_j-\rho)$, and where $T^{-\alpha/2}R_t^*$ is asymptotically negligible with respect to the other terms of (30) (see the technical results in the Supplementary Appendix).

To study the other terms of (30) we notice that $\mathsf{E}\left(\int_0^{T^{1-\alpha}}e^{4(\phi r+\lambda W_r)}dr\right)=\frac{e^{4\left(c+\lambda^2\right)T^{1-\alpha}}}{4(c+\lambda^2)}-\frac{1}{4(c+\lambda^2)}=\varkappa_{T^{1-\alpha}}^2+O(1)$. Again, we use a Lindberg Condition, this time regarding

$$\zeta_{k+1}\equiv\varkappa_{T^{1-\alpha}}^{-1}\exp\left(\frac{2\phi}{T^{\alpha}}k+\frac{2\lambda}{T^{\alpha/2}}U_k\right)\Delta U_{k+1}^{\rho}\tag{31}$$

which admits conditional variance such that

$$\begin{aligned}\sum_{k=1}^{T-1}\mathsf{E}_k\left[\zeta_{k+1}^2\right]&=\varkappa_{T^{1-\alpha}}^{-2}\sum_{k=1}^{T-1}\exp\left(\frac{4\phi}{T^{\alpha}}k+\frac{4\lambda}{T^{\alpha/2}}U_k\right)\lambda^2\\&=\varkappa_{T^{1-\alpha}}^{-2}\lambda^2\int_0^{T^{1-\alpha}}e^{4(\phi r+\lambda W_r)}dr+o_p(1)=O_p(1).\end{aligned}$$

We introduce the *i.i.d.* process u_t^* with bounded moments, and which does not correlate with u_t, such that

$$\Delta U_k^{\rho} = \rho_t - \rho = \exp\left\{\frac{\phi}{T^{\alpha}}\right\}\left(\exp\left\{\frac{\lambda u_t}{T^{\alpha/2}}\right\} - \mathsf{E}\left[\exp\left\{\frac{\lambda u_t}{T^{\alpha/2}}\right\}\right]\right) = e^{\frac{\phi}{T^{\alpha}}}\left(\frac{\lambda}{T^{\alpha/2}}u_t + \frac{u_t^*}{T^{\alpha}}\right).$$

Therefore we can write

$$\begin{aligned}\sum_{k=1}^{T-1}\frac{\exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)}{\varkappa_{T^{1-\alpha}}}\Delta U_{k+1}^{\rho} &= \lambda\sum_{k=1}^{T-1}\frac{\exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)}{\varkappa_{T^{1-\alpha}}} \times \frac{\Delta U_{k+1}}{T^{\alpha/2}} \\ &\quad + T^{-\alpha/2}\sum_{k=1}^{T-1}\frac{\exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)}{\varkappa_{T^{1-\alpha}}} \times \frac{\Delta U_{k+1}^*}{T^{\alpha/2}}.\end{aligned}$$

where $U_k^* \equiv \sum_{t=1}^{k} u_t^*$. The Lindberg condition (31) also holds when ΔU_{k+1}^{ρ} is replaced with ΔU_{k+1} or ΔU_{k+1}^*, hence

$$\sum_{k=1}^{T-1}\frac{\exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)}{\varkappa_{T^{1-\alpha}}}\frac{\Delta U_{k+1}}{T^{\alpha/2}} = \varkappa_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)}dW_r + o_p(1)$$

and, introducing W_r^*, $r \in [0,1]$, such that $T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor} u_t^* \Rightarrow W_r^*$ as $T \to \infty$,

$$\sum_{k=1}^{T-1}\frac{\exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)}{\varkappa_{T^{1-\alpha}}}\frac{\Delta U_{k+1}^*}{T^{\alpha/2}} = \varkappa_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)}dW_r^* + o_p(1).$$

It follows that $\sum_{k=1}^{T-1}\frac{\exp\left(\frac{2\phi}{T^{\alpha}}k + \frac{2\lambda}{T^{\alpha/2}}U_k\right)}{\varkappa_{T^{1-\alpha}}}\frac{\Delta U_{k+1}^{\rho}}{\lambda} = \varkappa_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)}dW_r + o_p(1)$ from which we deduce that

$$T^{-\alpha}\varphi_{T^{1-\alpha}}^{-2}\varkappa_{T^{1-\alpha}}^{-1}\sum_{t=1}^{T} y_{t-1}^2(\rho_t - \rho) = \left(\varphi_{T^{1-\alpha}}^{-1}\sigma_{\eta}\int_0^{T^{1-\alpha}} e^{-(\phi s + \lambda W_s)}dB_s\right)^2 \varkappa_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)}dW_r + o_p(1)$$

and, defining V_{θ} such that

$$\varkappa_{T^{1-\alpha}}^{-1}\int_0^{T^{1-\alpha}} e^{2(\phi r + \lambda W_r)}dW_r \Rightarrow V_{\theta} \sim \mathsf{D}(0,1), \qquad (32)$$

we obtain (28). ■

Proof of Lemma 1 and Conclusion. We summarize in the table below the results obtained above for the process y_t, $t \geq 0$, defined as in (1) and under Assumptions A and B. We consider the three cases $c < 0$, $c = 0$ and $c > 0$, and introduce the notation:

$$S_{yy} = \sum_{t=1}^{T} y_t^2, \quad S_{y\eta} = \sum_{t=1}^{T} y_{t-1}\eta_t, \quad \text{and} \quad S_{yy\rho} = \sum_{t=1}^{T} y_{t-1}^2 \left(\rho_t - \rho\right).$$

As $T \to \infty$ and for $x \in \{yy, y\eta, yy\rho\}$,

$$\sigma_\eta^{-2}\ \mu^x\ \phi_T^x\ S_x \Rightarrow D_x,$$

where (μ^x, ϕ_T^x, D_x) are defined as follows (assuming $(\phi, \lambda) \neq (0, 0)$).

	μ^{yy}	$\mu^{y\eta}$	$\mu^{yy\rho}$
$c < 0$	$-2c$	$\sqrt{-2c}$	$-2c/\left(\sqrt{3}\lambda\right)$
$c = 0$	$8\sqrt{2}\ \lambda^4$	2λ	$8\lambda^2$
$c > 0,\ \lambda^2 < \phi$	$8(c - 2\lambda^2)^2\ \sqrt{(c + 2\lambda^2)\,(c + \lambda^2)}$	$4c\left(c - 2\lambda^2\right)$	$8(c - 2\lambda^2)^2\ \sqrt{c + \lambda^2}/\lambda$
$c > 0,\ \lambda^2 = \phi$	$2\sqrt{(c + 2\lambda^2)\,(c + \lambda^2)}$	$2c$	$2\sqrt{c + \lambda^2}/\lambda$
$c > 0,\ \lambda^2 > \phi$	$4(2\lambda^2 - c)\sqrt{(c + 2\lambda^2)\,(c + \lambda^2)}$	$2c\sqrt{2\,(2\lambda^2 - c)}$	$4\left(2\lambda^2 - c\right)\sqrt{c + \lambda^2}/\lambda$

with

	ϕ_T^{yy}	$\phi_T^{y\eta}$	$\phi_T^{yy\rho}$
$c < 0$	$T^{-(1+\alpha)}$	$T^{-\frac{1+\alpha}{2}}$	$T^{-\frac{1+\alpha}{2}}$
$c = 0$	$T^{-2\alpha}e^{-6\lambda^2 T^{1-\alpha}}$	$T^{-1}e^{-2\lambda^2 T^{1-\alpha}}$	$T^{-\alpha}e^{-6\lambda^2 T^{1-\alpha}}$
$c > 0,\ \lambda^2 < \phi$	$T^{-2\alpha}e^{-2\left(c+\lambda^2\right)T^{1-\alpha}}$	$T^{-\alpha}e^{-2cT^{1-\alpha}}$	$T^{-\alpha}e^{-2\left(c+\lambda^2\right)T^{1-\alpha}}$
$c > 0,\ \lambda^2 = \phi$	$T^{-(1+\alpha)}e^{-2\left(c+\lambda^2\right)T^{1-\alpha}}$	$T^{-\frac{1+\alpha}{2}}e^{-2cT^{1-\alpha}}$	$T^{-1}e^{-2\left(c+\lambda^2\right)T^{1-\alpha}}$
$c > 0,\ \lambda^2 > \phi$	$T^{-2\alpha}e^{-6\lambda^2 T^{1-\alpha}}$	$T^{-\alpha}e^{-\left(c+2\lambda^2\right)T^{1-\alpha}}$	$T^{-\alpha}e^{-6\lambda^2 T^{1-\alpha}}$

and

	D_{yy}	$D_{y\eta}$	D_{yyu}
$c < 0$	1	$\mathsf{N}(0,1)$	$\mathsf{N}(0,1)$
$c \geq 0$	$X_\theta^2 Z_\theta$	$X_\theta Y_\theta$	$X_\theta^2 V_\theta$

where $X_\theta \sim \mathsf{D}(0,1)$, $Y_\theta \sim \mathsf{D}(0,1)$, $V_\theta \sim \mathsf{D}(0,1)$ and $Z_\theta \sim \mathsf{D}(0,1)$ such that $X_\theta \perp V_\theta$, and $Z_\theta \perp V_\theta$. The tables above directly provide Lemma 1.

Now, Proposition 3 can then be directly deduced from the results of this table. Indeed, in the case $c < 0$, we can write, after noticing that $T^{-\frac{1+\alpha}{2}} \sum_{t=1}^{T} y_{t-1}^2 (\rho_t - \rho)$ is asymptotically uncorrelated with $T^{-\frac{1+\alpha}{2}} \sum_{t=1}^{T} y_{t-1}\eta_t$, that

$$\begin{aligned} T^{\frac{1+\alpha}{2}} (\widehat{\rho} - \rho) &= \frac{T^{-\frac{1+\alpha}{2}} \sum_t y_{t-1}^2 (\rho_t - \rho)}{T^{-1-\alpha} \sum_t y_{t-1}^2} + \frac{T^{-\frac{1+\alpha}{2}} \sum_t y_{t-1}\eta_t}{T^{-1-\alpha} \sum_t y_{t-1}^2} \\ &= \frac{\mu^{yy}}{\mu^{yy\rho}} \frac{\sigma_\eta^{-2}}{\sigma_\eta^{-2}} \frac{\mu^{yy\rho} T^{-\frac{1+\alpha}{2}} \sum_t y_{t-1}^2 (\rho_t - \rho)}{\mu^{yy} T^{-1-\alpha} \sum_t y_{t-1}^2} + \frac{\mu^{yy}}{\mu^{y\eta}} \frac{\sigma_\eta^{-2}}{\sigma_\eta^{-2}} \frac{\mu^{y\eta} T^{-\frac{1+\alpha}{2}} \sum_t y_{t-1}\eta_t}{\mu^{yy} T^{-1-\alpha} \sum_t y_{t-1}^2} \\ &\xrightarrow{L} \mathsf{N}\left(0, -2\phi + \lambda^2\right). \end{aligned}$$

Assume now that $c \geq 0$. We can write

$$T^{\alpha/2} \frac{\phi_T^{yy\rho}}{\phi_T^{yy}} (\widehat{\rho} - \rho) \Rightarrow \frac{\mu^{yy}}{\mu^{yy\rho}} \frac{D_{yy\rho}}{D_{yy}},$$

where the various ratios are calculated using the previous table and provide the same results for all cases when $c \geq 0$, namely

$$\phi_T^{yy\rho} / \phi_T^{yy} = T^\alpha \qquad \text{and} \qquad \mu^{yy\rho} / \mu^{yy} = \frac{1}{\lambda^2 \sqrt{\phi + 3\lambda^2}},$$

hence the result. ■

References

Andel, J. (1976). Autoregressive series with random parameters. *Math. Operationsforsch. u. Statist. 7*, 735–41.

Andrews, D. W. (1993a). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica 61*(1), 139–65.

Andrews, D. W. (1993b). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. and J. H. Stock (2005). Inference with Weak Instruments. NBER Technical Working Papers 0313, National Bureau of Economic Research, Inc.

Aue, A. (2008). Near-integrated random coefficient autoregressive time series. *Econometric Theory 24*(05), 1343–1372.

Aue, A. and L. Horváth (2011). Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica 21*, 973–999.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Berkes, I., L. Horvth, and S. Ling (2009). Estimation in nonstationary random coefficient autoregressive models. *Journal of Time Series Analysis 30*(4), 395–416.

Blanchard, O. and M. Watson (1982). Bubbles, Rational Expectations and Financial Markets. In P. Wachtel (Ed.), *Crises in the Economic and Financial Structure*, pp. 295–316. Lexington Books.

Bobkoski, M. (1983). Hypothesis testing in nonstationary time series. Unpublished PhD thesis, Dept. of Statistics, University of Wisconsin, Madison.

Bougerol, P. and N. Picard (1992). Stationarity of garch processes and of some nonnegative time series. *Journal of econometrics 52*(1-2), 115–127.

Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics 15*(3), 1050–1063.

Charemza, W. W. and D. F. Deadman (1995). Speculative bubbles with stochastic explosive roots: The failure of unit root testing. *Journal of Empirical Finance 2*(2), 153–163.

Csörgo, M. and L. Horváth (1993). *Weighted Approximations in Probability and Statistics.* New York: Wiley.

Distaso, W. (2008). Testing for unit root processes in random coefficient autoregressive models. *Journal of Econometrics 142*(1), 581 – 609.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica 65*(6), 1365–1387.

Dufour, J.-M., L. Khalaf, and M. Kichian (2006). Inflation dynamics and the New Keynesian Phillips Curve: an identification robust econometric analysis. *Journal of Economic Dynamics and Control 30*(9), 1707–1727.

Evans, G. W. (1991). Pitfalls in testing for explosive bubbles in asset prices. *American Economic Review 81*(4), 922–30.

Froot, K. A. and M. Obstfeld (1991). Intrinsic bubbles: The case of stock prices. *American Economic Review 81*(5), 1189–214.

Fulop, A. and J. Yu (2014). Bayesian analysis of bubbles in asset prices. Mimeo, ESSEC Business School.

Giraitis, L. and P. C. B. Phillips (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis 27*(1), 5160.

Granger, C. W. and N. R. Swanson (1997). An introduction to stochastic unit-root processes. *Journal of Econometrics 80*(1), 35–62.

Hall, S. G., Z. Psaradakis, and M. Sola (1999). Detecting periodically collapsing bubbles: a Markov-switching unit root test. *Journal of Applied Econometrics 14*(2), 143–154.

Hamilton, J. D. (1986). On testing for self-fulfilling speculative price bubbles. *International Economic Review 27*(3), 545–52.

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. *The Review of Economics and Statistics 81*(4), 594–607.

Hodges, J. L. and E. L. Lehmann (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics 34*(2), 598–611.

Homm, U. and J. Breitung (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Hwang, S. Y. and I. V. Basawa (2005). Explosive random-coefficient AR(1) processes and related asymptotics for least-squares estimation. *Journal of Time Series Analysis 26*(6), 807–824.

Jiang, L., X. Wang, and J. Yu (2017). In-fill asymptotic theory for structural break point in autoregression: A unified theory. Mimeo, Singapore Management University.

Kemp, G. C. R. (1999). The behavior of forecast errors from a nearly integrated AR(1) model as both sample size and forecast horizon become large. *Econometric Theory 15*, 238–256.

Leybourne, S. J., B. P. M. McCabe, and A. R. Tremayne (1996). Can economic time series be differenced to stationarity? *Journal of Business & Economic Statistics 14*(4), pp. 435–446.

Lieberman, O. (2012). A similarity-based approach to time-varying coefficient non-stationary autoregression. *Journal of Time Series Analysis 33*(3), 484–502.

Lieberman, O. and P. C. B. Phillips (2017). A multivariate stochastic unit root model with an application to derivative pricing. *Journal of Econometrics 196*(1), 99–110.

Liptser, R. and A. Shiryayev (1989). *Theory of Martingales*. Mathematics and its Applications. Springer.

Magnusson, L. and S. Mavroeidis (2014). Identification using stability restrictions. *Econometrica 82*(5), 1799–1851.

Matsumoto, H. and M. Yor (2005). Exponential functionals of Brownian motion, I: Probability laws at fixed time. *Probab. Surveys 2*, 312–347.

McCabe, B. P. M. and A. R. Tremayne (1995). Testing a time series for difference stationarity. *The Annals of Statistics 23*(3), 1015–1028.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica 75*(5), 1411–1452.

Mikusheva, A. (2012). One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica 80*(1), 173–212.

Nicholls, D. F. and B. G. Quinn (1982). *Random Coefficient Autoregressive Models: An introduction*. New York: Springer-Verlag.

Perron, P. (1996). The adequacy of asymptotic approximations in the near-integrated autoregressive model with dependent errors. *Journal of Econometrics 70*(2), 317–350.

Pesavento, E. and B. Rossi (2006). Small-sample confidence intervals for multivariate impulse response functions at long horizons. *Journal of Applied Econometrics 21*(8), 1135–1155.

Phillips, P. C. and S.-P. Shi (2017). Financial bubble implosion and reverse regression. *Econometric Theory forthcoming*, 1–49.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. (1998). Impulse response and forecast error variance asymptotics in nonstationary VARs. *Journal of Econometrics 83*, 21–56.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica 82*(3), 1177–1195.

Phillips, P. C. B. and T. Magdalinos (2004). Limit theory for moderate deviations from a unit root. Working papers, Cowles Foundation, Yale University.

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

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

Phillips, P. C. B., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review 52*, 201–26.

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

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer.

Richardson, M. and J. H. Stock (1989). Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics 25*(2), 323–348.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics 28*(3), 435–459.

Stock, J. H. (1996). VAR, error correction, and pretest forecasts at long horizons. *Oxford Bulletin of Economics and Statistics 58*, 685–701.

Stock, J. H. and M. W. Watson (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association 93*(441), 349–358.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics 20*(4), 518–29.

Tsay, R. S. (1987). Conditional heteroscedastic time series models. *Journal of the American Statistical Association 82*(398), 590–604.

Valkanov, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics 68*(2), 201–232.

Wang, D. and S. K. Ghosh (2008). Bayesian estimation and unit root tests for random coefficient autoregressive models. *Model Assisted Statistics and Applications 3*(4), 281–295.

White, H. and C. W. Granger (2011). Consideration of trends in time series. *Journal of Time Series Econometrics 3*(1), 1941–28.