Econometrics Journal (2019), volume **00**, pp. 1–35.
Article No. ectj??????

Probabilistic Forecasting of Bubbles and Flash Crashes

ANURAG BANERJEE[†], GUILLAUME CHEVILLON[‡] AND MARIE KRATZ[⋆]

[†] *Durham Business School*
E-mail: a.n.banerjee@durham.ac.uk
[‡] *ESSEC Business School (Corresponding author)*
E-mail: chevillon@essec.edu
[⋆] *ESSEC Business School, CREAR*
E-mail: kratz@essec.edu
Received: 24 July 2019

Summary 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 centred 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 time series and obtain related predictive probabilities. We study the asymptotic properties of the NERC and provide a procedure for inference on the parameters. In empirical illustrations, we estimate predictive probabilities of bubbles or flash crashes in financial asset prices.

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

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 industrialised 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. The literature does not appear to have converged to a commonly agreed definition of bubbles (e.g., Hamilton, 1986, and Granger and White, 2011) but a stylised feature is that bubbles in a time series y_t are characterised by sustained positive growth, and flash crashes through sharp decline. 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 characterised by mean reversion, stationarity or unit-root behaviour. 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.

The simplest model for bubble dynamics may be the AR(1) with an autoregressive coefficient, say ρ, which has to be above unity to generate an exponential growth. The coefficient ρ must also be close to unity to be empirically relevant. Hence, bubble dynamics can for instance be modelled by stochastic processes such as as in Examples (*iv*) and (*v*) in Table 1, which records a few common examples. Yet, the explosive autoregressive root must be temporary and the coefficient ρ must shift to a value at, or below, unity during normal times as in Table 1, Example (*vi*), which was proposed by Blanchard and Watson (1982). We will show that Examples (*iv*)-(*vi*) do generate bubbles according to the definition we will propose later. Yet these models generate explosiveness with probability one over any horizon. Hence they cannot distinguish between permanent and temporary bubbles as opposed to the Near Explosive Random Coefficient model (NERC) briefly described as Example (*vii*) in Table 1: this is the model we propose in this paper and that constitutes its focus.

To allow for temporary bubbles, some authors have assumed the presence of deterministic breaks in ρ (e.g., Phillips, Wu and Yu, 2011, Phillips, Shi and Yu, 2015, Harvey, Leybourne and Solis, 2013, and the example in equation (5) of Christensen, Oomen and Renò, 2018), as exemplified in Figure 2, Panel (c), but such breaks are by construction unpredictable. It is therefore necessary for forecasting purposes to assume that ρ shifts randomly: we denote the resulting process by ρ_t.[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 accommodate the existence of bubbles can generate both upward and downward trajectories. Hence, when the process y_t under consideration is negative, negative bubbles become what is known as a "flash crash" by considering negative changes $y_{t+h} - y_t < 0$, while ensuring some form of explosiveness ($y_{t+h}/y_t > 1$). This is not the same as the implosion of bubbles as studied by Phillips and Shi (2018). 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).

Most of the literature on bubble detection is concerned with detecting bubbles by rejecting the null hypothesis that there is no bubble. Hence the models generally proposed do not present "bubbly" properties but are rejected when bubbles form (such as, e.g., Phillips, Wu and Yu, 2011). Here, by contrast, we focus on models that can accommodate the presence of bubbles, yet contrary to approaches such as Engsted and Nielsen (2012),

[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.

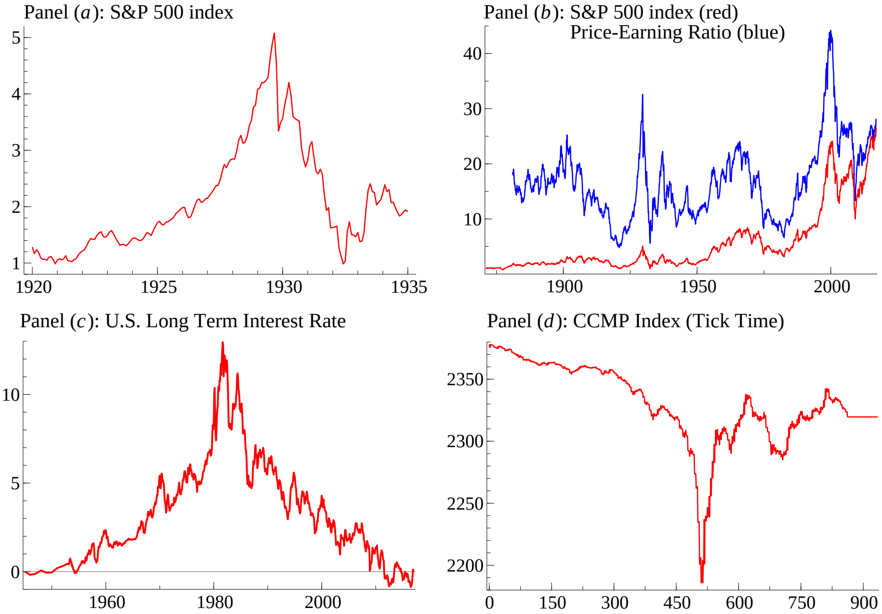

Figure 1. The figure presents examples of data series. Panels (a)-(c) produce monthly data from Robert Shiller's website: the S&P 500 index on (a) and (b) , the cyclically adjusted Price/Earning Ratio and the long term interest rate on Panel (c). Panel (d) is the Nasdaq index expressed in Tick time after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

Table 1. Examples of stochastic processes y_t with corresponding probabilities of bubbles or crashes. $\pi_{\tau,\chi}(\gamma)$ is the predictive probability given in in Definition 3.1, where τ is a parameter for the inception of the bubble, χ governs its duration and γ its magnitude.

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

Note: We assume that x_t is weakly stationary, η_t and u_t are independent *iid* processes. Details for (i)-(vi) are given in Appendix B. Example (vii) is the focus of this paper and is delineated in Section 2. The NERC model, (vii), with $\alpha = 1$ is nested within Example (ii).

our interest is to model temporary bubbles. We refrain from delineating our proposal for a bubble at this stage to avoid any unnecessary technicality but relegate it to Section 3 where, essentially, bubbles correspond to periods of explosive patterns.

We propose in this paper a simple time-series model that accommodates both 'quiet' and bubble periods, and where the bubble arises and disappears 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 that 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 autoregressive setting, $\rho_t \sim iid$ generates rich dynamics since the h-period growth 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. Whereas the movements in ρ_{t+j}, $j > 0$, are unpredictable, the process of interest in the forecasting exercise, $\prod_{i=1}^{h} \rho_{t+i}$, presents a high degree of persistence and can be subjected to a forecasting exercise, see Figure 2(*d*) for a simulated example where $\bar{\rho}_{t,h} = \sqrt[h]{\prod_{i=1}^{h} \rho_{t+i}}$.

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 its expectation, $\mathsf{E}[\rho_t]$, and variance, $\mathsf{V}[\rho_t]$, 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 modeller 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 suggest an empirical inference procedure and then provide, in Section 5, an empirical illustration to two of the datasets plotted in Figure 1. Section 6 concludes. Appendices provide (A) an additional discussion of the related literature, (B) examples of processes that

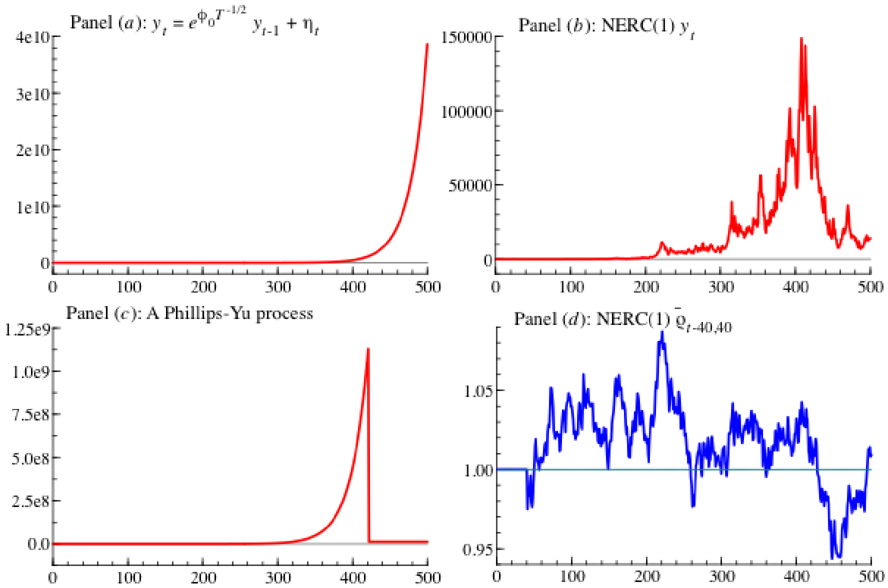

Figure 2. The figure presents simulated paths of near explosive AR(1) and NERC(1) processes driven by the same $\eta_t \overset{i.i.d}{\sim} N(0,1)$ with $T = 500$ and $\alpha = 1/2$. Panel (a): a near explosive AR(1) $y_t = e^{\phi_0/\sqrt{T}} y_{t-1} + \eta_t$ with $\phi_0 = .625$ so $\rho_{0,T} = 1.028$. Panel (b): a NERC(1) with innovations $u_t \overset{i.i.d}{\sim} N(0,1)$ and $(\phi, \lambda) = (.5, .5)$ so $E(\rho_t) = 1.028$. 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).

may or not exhibit bubbles, (C) further results on the inference procedure, and (D) the proofs of the propositions. A Supplementary Appendix containing further simulations, empirical applications and technical results is available online.[2] Throughout the paper, $\lfloor \cdot \rfloor$ denotes the integer part; $1_{\{A\}}$ the indicator function that takes value 1 if $\{A\}$ is true and zero otherwise; $\mathbb{R}_{+,*}$ the set of strictly positive real scalars; $\overset{p}{\to}$ and $\overset{L}{\to}$ convergences in probability and law; $\Rightarrow$ weak convergence of the associated probability measure under the Skorokhod topology.

2. A NEAR-EXPLOSIVE RANDOM COEFFICIENT AUTOREGRESSIVE PROCESS

This section presents the model we propose and studies its asymptotic properties. Because the literature on random coefficient autoregressive models (RCAR) is rich, the relevant references are presented in Appendix A to facilitate the exposition of our results below.

[2]http://guillaume-chevillon.faculty.essec.edu/research/working-papers/Supplement-BCK2019.pdf

2.1. The model

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

$$y_t = \rho_t y_{t-1} + \eta_t, \text{ with } \rho_t = \exp\left\{\frac{\phi}{T^\alpha} + \frac{\lambda}{T^{\alpha/2}} u_t\right\}, \tag{2.1}$$

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 (2.1) 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 $\lambda = 0$, the $\mathsf{V}[\rho_t] = 0$, 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 fact 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 (2.1).

Assumption A: The process admits an origin $y_0 = o_p(1)$ that is independent of (u_t, η_t) for $t > 0$.

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

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

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 A.1 in Appendix A).[3]

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 $\bar{\rho}_{t,h} = \exp\left\{\frac{\phi}{T^\alpha} + \frac{\lambda}{T^{\alpha/2}}\bar{u}_{t,h}\right\}$ such that

$$y_{t+h} = \bar{\rho}_{t,h}^h y_t + \bar{\eta}_{t,h}. \tag{2.2}$$

In the expression above, the multistep errors, $\bar{\eta}_{t,h}$, are defined implicitly and satisfy $\mathsf{E}_t[\bar{\eta}_{t,h}] = 0$. The process y_t may exhibit an explosive pattern between t and $t + h$ if

[3] 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.

$\phi + \lambda T^{\alpha/2}\bar{u}_{t,h} > 0$. This can happen even when the drift term ϕ is negative. The model is specified in (2.1) in terms of the proximity of $\log \rho_t$ to 0 rather than of ρ_t to 1 (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 (yet not necessarily unimodal). This implies that the parameters ϕ and λ^2 both matter in determining the magnitude of the expectation of ρ_t:

$$\mathsf{E}[\rho_t] = 1 + \frac{\phi + \frac{1}{2}\lambda^2}{T^\alpha} + O(T^{-2\alpha}), \tag{2.3}$$

$$\mathsf{V}[\rho_t] = \frac{\lambda^2}{T^\alpha} + O(T^{-2\alpha}). \tag{2.4}$$

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 and Lieberman and Phillips, 2017, and the discussion that follows Corollary 2.1 below). 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).

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

$$\sup_{s \in [0, T^{1-\alpha}]} \left| T^{-\alpha/2} \sum_{t=1}^{\lfloor T^\alpha s \rfloor} u_t - W_s \right| = o_{a.s.}(1),$$

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

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 use the notation:

$$c = \phi + \lambda^2 \quad \text{and} \quad \rho = \mathsf{E}[\rho_t], \tag{2.5}$$

where we recall that ρ implicitly depends on T.

2.2. Asymptotic properties

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

PROPOSITION 2.1. *Let the process y_t be defined for $t \geq 0$ by (2.1) under Assumptions A and B. Then, for $r \in [0, (T+H)\,T^{-\alpha}]$, as $T \to \infty$,*

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

where $K_{\phi,\lambda}$ is defined, for $(\phi, \lambda) \in \mathbb{R} \times \mathbb{R}_{+,*}$ and $r \in \mathbb{R}_+$, as

$$K_{\phi,\lambda}(r) = \int_0^r e^{(r-s)\phi + \lambda(W_r - W_s)} dB_s, \tag{2.6}$$

with W and B independent standard Brownian motions defined in Section 2.1.

Proposition 2.1 shows that several cases arise, depending on whether the distribution of $K_{\phi,\lambda}(r)$ remains bounded, i.e. on the value of c which governs the variance $\mathsf{V}[K_{\phi,\lambda}(r)] = \int_0^r e^{2cs}ds$. When $c < 0$, the magnitude of y_t is similar to that which PM obtain, i.e. $\mathsf{V}[y_t] = O(T^\alpha)$. When $c = 0$, $\mathsf{V}[y_t] = O(T)$ as in the near unit root when the coefficient is nonstochastic ($\lambda = 0$ so $K_{\phi,0}(r)$ reduces to the Ornstein-Uhlenbeck diffusion considered in PM). Finally, when $c > 0$, the process with fixed origin exhibits explosiveness in its second moment, as pointed out by Hwang and Basawa (2005), and $\mathsf{V}[y_t] = O(T^\alpha e^{2cT^{1-\alpha}})$.

Proposition 2.1 can be extended to cover $\alpha = 1$ as the following corollary shows.

Corollary 2.1. *Proposition 2.1 also holds when $\alpha = 1$: as $T \to \infty$, $T^{-1/2}y_{\lfloor rT \rfloor} \Rightarrow \sigma_\eta K_{\phi,\lambda}(r)$ for $r \in [0, 1 + H/T]$.*

Corollary 2.1 shows that explosive patterns may only arise in y_t if $\alpha < 1$. When $\alpha = 1$ the random coefficient stays within an $O_p\left(T^{-1}\right)$ neighbourhood of unity and, for $r \in [0,1]$, the corollary shows that $y_t = O_p\left(\sqrt{T}\right)$, for $t \le T + H$ with $H = O\left(T\right)$, i.e. the process retains the magnitude of a (stochastic) unit root process without explosive second moments, irrespective of the value of c, see also Granger and Swanson (1997) and Lieberman and Phillips (2017). We show in Corollary 3.1 below that the probability of bubble episodes is zero when $\alpha = 1$. By contrast, if $\alpha < 1$, second moments are explosive when $c > 0$, since $\mathsf{V}[y_T] = O(T^\alpha e^{2cT^{1-\alpha}})$. Hence, we focus our attention, in this paper, to the case $\alpha < 1$ to ensure the possibility of bubbles.

In order to study the presence of bubbles, we define the growth in process y_t between dates $t > 0$ and $t + h$ as

$$g_{t,h} = \frac{y_{t+h}}{y_t}. \tag{2.7}$$

The following proposition gives its asymptotic behaviour.

Proposition 2.2. *Let the process y_t be defined for $t \ge 0$ by (2.1) under Assumptions A and B with $\alpha \in (0,1]$. Then for $(r, r+s) \in (0, (T+H)\,T^{-\alpha}]^2$ with $s > 0$, the growth defined in expression (2.7) satisfies, as $T \to \infty$,*

$$g_{\lfloor T^\alpha r \rfloor, \lfloor T^\alpha s \rfloor} \Rightarrow \exp\{\phi s + \lambda(W_{r+s} - W_r)\} + \frac{\int_r^{r+s} e^{\phi(r+s-u)+\lambda(W_{r+s}-W_u)}dB_u}{\int_0^r e^{\phi(r-u)+\lambda(W_r-W_u)}dB_u}. \tag{2.8}$$

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

$$T^{-\alpha/2}y_{\lfloor T^\alpha r \rfloor}\left(g_{\lfloor T^\alpha r \rfloor, \lfloor T^\alpha s \rfloor} - e^{\phi s + \lambda(W_{r+s}-W_r)}\right)\Big| \mathcal{I}_{\lfloor T^\alpha r \rfloor} \Rightarrow \int_r^{r+s} e^{\phi(r+s-u)+\lambda(W_{r+s}-W_u)}dB_u. \tag{2.9}$$

Proposition 2.2 provides the key results used for forecasting bubbles in the next section. Expression (2.8), 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 (2.9) 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 (2.8) are independent and that $T^{-\alpha/2}y_{\lfloor T^\alpha r \rfloor}$ admits the limit given in Proposition 2.1. We show in the next section how the Proposition 2.2 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. In this section, we first provide a definition of bubbles in a local-asymptotic context and then show 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. Bubbles in a local-asymptotic context

We follow the time series approach of Granger and White (2011) and let bubbles be defined via their time series properties, rather than relying on a structural decomposition that studies deviations from fundamental dynamics. We consider a definition in a local-asymptotic context where not only model parameters relate to the available sample size T, but also the timing and duration of events. Letting $T \to \infty$ then allows to express magnitudes and derive distributions that prove useful in forecasting.[4] Probabilistic bubble episodes of duration h are defined as exhibiting sustained growth, i.e., such that, at time t, $g_{t,h}$ defined in (2.7) is greater than unity. In a non-local asymptotic framework, this leads to defining the probability of a bubble episode of duration h between times t and $t+h$ 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. We retain in the definition the notation $y_{T,t}$ for the process whose distribution at time t may depend on the sample size T.

Definition 3.1. *Let the local-asymptotic process $y_{T,t}$ be defined for $0 \leq t \leq T+H$, with $H = O(T)$. Consider two non-negative deterministic mappings for local time, $t_{(T,r)}$, and duration, $h_{(T,s)}$ for some parameters $(r,s) \in \mathbb{R}_+^2$, such that $t_{(T,r)} \leq T$ and, as $T \to \infty$, $\left(t_{(T,r)}, h_{(T,s)}\right) \to (\infty, \infty)$ with $h_{(T,s)} = o(T)$.*
The probability of the process experiencing a bubble episode of magnitude $\gamma > 1$ of duration $h_{(T,s)} > 0$ between local times $t_{(T,r)}$ and $t_{(T,r)} + h_{(T,s)}$ is defined as

$$\pi_{r,s}(\gamma) = \lim_{T \to \infty} \Pr(g_{t_{(T,r)},h_{(T,s)}} > \gamma \big| \mathcal{I}_{t_{(T,r)}}),$$

where $g_{t,h}$, is defined in expression (2.7) and $\mathcal{I}_t$ is the sigma-field generated by $(y_{T,j})_{j \leq t}$.

The local time and duration mappings are specific to the models considered so, in empirical applications, the modeller who wishes to forecast the probability of bubbles will

[4]This is a procedure which is common, for instance, when modelling 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).

need to define these mappings appropriately. Typical mappings encountered in the econometric literature for local time and duration relate to the sample size, as parameterised by $(r,s) \in [0,1]^2$, i.e., $t_{(T,r)} = \lfloor T^a r \rfloor$ and $h_{(T,s)} = \lfloor T^b s \rfloor$, for $(a,b) \in [0,1]^2$. In the literature on deterministic breaks detection, Bai and Perron (1998) for instance parameterize the break date as a fraction $\lfloor Tr \rfloor$ of the sample size, i.e., with $a = 1$. In the near or moderate deviations to unit roots, PM let $a \in (0,1)$ when considering the process of interest. In the context of long run forecasting, predictive regressions and impulse responses, an extensive literature considers $a = b = 1$ — 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 $b = 0$. Mikusheva (2012) allows for intermediate impulse responses, such as when $b \in (0,1)$. Definition 3.1 assumes that the duration is such that, as $T \to \infty$, $h_{(T,s)}/T \to 0$, i.e., the duration of the episode is short to ensure that bubbles eventually end if a sample of sufficient length is observed. This can also be strengthened to the situation where $h_{(T,s)}/t_{(T,r)} \to 0$ so the bubble is short with respect to the sample observed until its inception and termination. These two cases correspond, when $t_{(T,r)} = \lfloor T^a r \rfloor$ and $h_{(T,s)} = \lfloor T^b s \rfloor$ to assuming, respectively, $b < 1$ and $b < a$.

3.2. Predictive probabilities in the NERC

Definition 3.1 enforces a dichotomy between processes that may generate bubbles and those that do not, as discussed in the introduction. To clarify this, we present in Appendix B a few examples where the probability of bubble prediction $\pi_{r,s}(\gamma) = 0$ or 1 for all short horizons $h_{(T,s)} = o(t_{(T,r)})$, as summarised 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 $\pi_{r,s}(\gamma) = 0$ for $\gamma > 1$.

Building on the discussion and in line with Definition 3.1, we obtain, in the remainder of this paper, predictive probabilities of bubbles and crashes under the NERC. Throughout, we consider the mappings $t_{(T,r)} = \lfloor T^\alpha r \rfloor$ and $h_{(T,s)} = \lfloor T^\alpha s \rfloor$ so, for notational ease, we omit the subscripts in t and h. We are now ready for the following Corollary to Proposition 2.2.

Corollary 3.1. *Under the assumptions of Proposition 2.2, the predictive probabilities defined in Definition 3.1 with* $(t,h) = (\lfloor T^\alpha r \rfloor, \lfloor T^\alpha s \rfloor)$ *for* $(r, r+s) \in (0, (T+H)T^{-\alpha}]^2$, *satisfy, as* $T \to \infty$,

$$\text{if } \alpha \in (0,1): \pi_{r,s}(\gamma) = \lim_{T\to\infty} \Pr\left(e^{\phi s + \lambda(W_{r+s} - W_r)} + \frac{T^{\alpha/2}}{y_{\lfloor T^\alpha r \rfloor}} \int_r^{r+s} e^{\phi(r+s-u)+\lambda(W_{r+s}-W_u)} dB_u > \gamma \middle| \mathcal{I}_{\lfloor T^\alpha r \rfloor}\right) \tag{3.10a}$$

$$\text{if } \alpha = 1: \pi_{r,s}(\gamma) = 0. \tag{3.10b}$$

Proof. The case $\alpha \in (0,1)$ is immediate. When $\alpha = 1$, Definition 3.1 implies that, as $T \to \infty$, $h_{(T,s)} = o(T)$ so $s \to 0$ and $g_{t_{(T,r)}, h_{(T,s)}} \to 1 < \gamma$.

The predictive probabilities $\pi_{r,s}(\gamma)$ can be obtained by simulation of the right-hand side of expression (3.10a). With a slight abuse of notation we now refer to $\pi_{t,h}(\gamma)$

when the mappings (t,h) are implicitly defined. Cases of interest comprise, in particular, $\pi_{t,h}\left(y_t/y_{t-h}\right)$, which is the probability that the growth observed over the latest h period carries over to the next h periods. Also of interest is, e.g., $\pi_{t,h}\left(r_{t,h}\right)$ when y_{t+h}/y_t denotes h-period log returns of assets and $r_{t,h}$ is the risk free rate of return and is part of the information set at time t. We report a simulation exercise for predictive probabilities in the Supplementary Appendix.

4. INFERENCE SCHEME

This section is concerned with putting the analytical results derived above into empirical use. Throughout, we assume that u_t is not observed so we need to perform inference when ρ_t is latent.[5] To provide predictive probabilities $\pi_{t,h}(\gamma)$ as in Definition 3.1 and expression (3.10a), which involves the parameters of the model, (ϕ,λ,α), we need to perform inference over the latter. The implementation 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.

As is common in the context of local asymptotics, consistent estimators of the localizing parameters (ϕ,λ) may be infeasible when $c\geq 0$.[6] Hence, we resort to the technique, now standard under local asymptotics, which 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, 1991 and Andrews, 1993a).[7] 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).[8] The 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.

[5] We do not consider here filtering scheme that yield estimates of ρ_t since the literature has shown they do not perform well in the bubble context, see Wu (1995) and Gürkaynak (2008).

[6] 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*b*) for which no estimator has been proved consistent in the explosive case.

[7] 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.

[8] 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 (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.

Here we conduct inference using a unique moment condition under the null 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 Appendix D).

PROPOSITION 4.1. *Let the process y_t be defined for $t \geq 0$ by (2.1) under Assumptions A and B, with $\lambda \neq 0$ and $\alpha \in (4/5, 1)$. 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 $c < 0$, $T^{\frac{1+\alpha}{2}} (\widehat{\rho} - \rho) \xrightarrow{d} \mathsf{N}(0, -2\phi + \lambda^2)$,
(ii) if $c \geq 0$, $T^{\alpha} (\widehat{\rho} - \rho) \Rightarrow \left(\lambda \mathsf{D}^{(u,u)}_{2\phi,2\lambda} + \mathsf{D}^{(u,\eta)}_{\phi,\lambda} / \mathsf{D}^{(u,\eta)}_{-\phi,-\lambda}\right) / \mathsf{F}_{2\phi,2\lambda}$
where $\mathsf{D}^{(\cdot,\cdot)}_{\cdot,\cdot}$ and $\mathsf{F}_{\cdot,\cdot}$ denote nondegenerate random variables defined in Expressions (D.4) and (D.5) in the Appendix.

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 of $\delta_{\theta,T}$ is given by Proposition 4.1 and can be simulated under H_0. This limiting distribution does not depend on the parameter α which only governs the speed of convergence so α is not identified. The variance σ^2_η 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). Following Phillips (2014) we recognise 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 Appendix C, 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 where we set $\alpha = 0.9$ (the articles by Phillips and coauthors on testing for bubbles use a variety of values for $\alpha \in (.5, 1)$, the actual value does not seem to matter much in practice and α only seems to constitute a scaling factor). 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 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

Table 2. Monte Carlo estimates of coverage probabilities

	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

Note: Coverage probabilities of confidence intervals are 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.

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 $\alpha \geq 1/2$ (although we only prove the asymptotic validity of our estimation technique for $\alpha > 4/5$). The parameter α in practice 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 4.1 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 modeller to focus specifically to either type of episode (see the examples below).

5. EMPIRICAL ILLUSTRATION

We now turn to an empirical illustration of the proposed method for estimating predictive probabilities of bubbles and flash crashes.[9] We consider applications to prediction of growth in the Price Earning Ratio (PER) (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).[10] We provide additional applications to stock market returns and long term interest rates in the Supplementary Appendix. All series are presented in Figure 1. Datasets and codes for replication are available on request. 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

[9] As discussed in the introduction and shown in Table 1, other existing models of bubbles cannot be used to compute probabilities in the $(0, 1)$ range so we cannot compare our technique with alternatives.

[10] We use this dataset which is not recent but represent one of the historically well documented examples of a flash crash, see https://en.wikipedia.org/wiki/2010_Flash_Crash

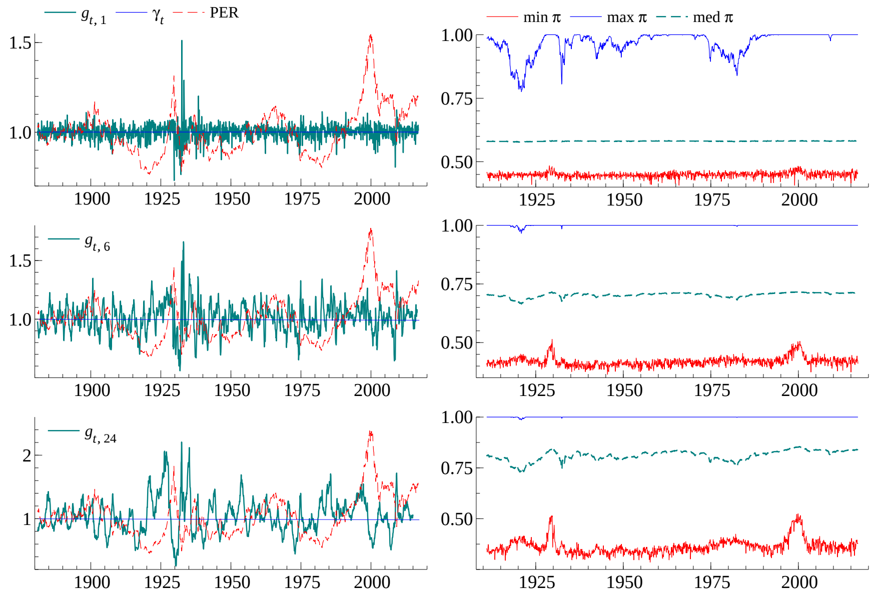

Figure 3. Predictive probabilities for y_t, the logarithm of the monthly Price Earning Ratio (PER). The left column reports the actual series as well as the growth $g_{t,h} = y_{t+h}/y_t$ for horizons $h = 1, 6$ and 24 in, respectively, the top, middle and bottom rows. The log PER data are scaled to match the mean and range of $g_{t,h}$. The column on the right reports predictive probabilities for $\gamma_t = 1$ where the confidence interval has nominal coverage 90%.

are computed recursively using only data observed at t and we report in each case the estimated $\left(\widehat{\pi}_{t,h}^{\min}(\gamma), \widehat{\pi}_{t,h}^{\max}(\gamma)\right)$ for $\omega = 0.10$, as well as our point estimate $\widehat{\pi}_{t,h}^{med}(\gamma)$. 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 3-4. In either 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. Figure 3 shows that we obtain, using the NERC, probabilities that are neither zero nor unity as in the examples of Table 1.

Figure 3 reports results for the logarithm of the average PER over a long period with $\gamma = 1$ (as a lower bound of the range of possible values for bubbles). 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. The following figure, Figure 4, 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.

Both applications above are performed with an expanding window, hence the relative stability in the predictive probability range. If the modeller is willing to use rolling win-

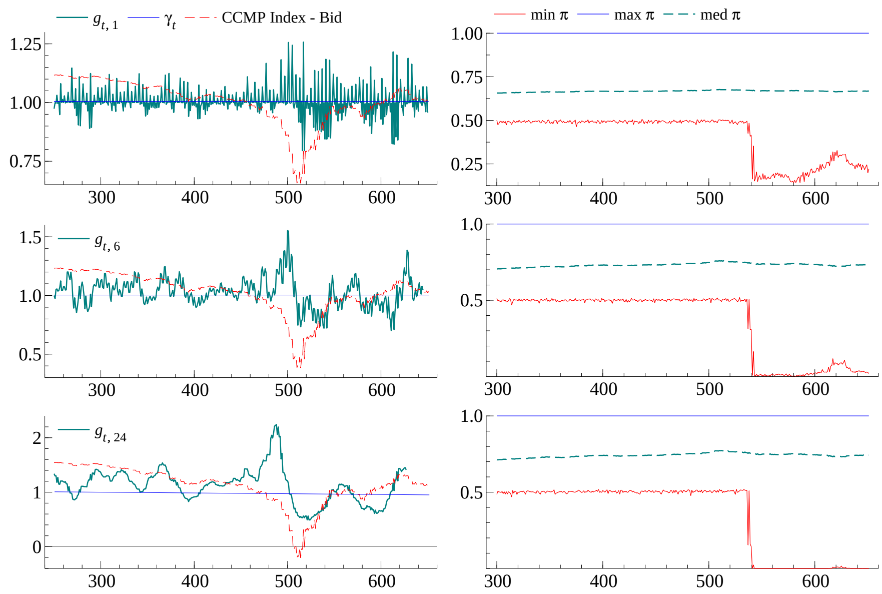

Figure 4. Predictive probabilities for the logarithm of the Nasdaq bid index (labeled CCMP on Bloomberg Terminals) in tick time from 3:00PM EST on May 6, 2010. The left column reports the actual series as well as the growth $g_{t,h}$. The log index data are scaled to match the mean and range of $g_{t,h}$. The column on the right reports predictive probabilities for $\gamma_t = 1$ where the confidence interval has nominal coverage 90%.

dows of observations or a time varying threshold γ, she will obtain more variation in the predictive range. The NERC could be extended to include additional autoregressive lags to generate more variability.

6. 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 flexibility 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 generalisation 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 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 use our results to test candidate variables for u_t, assuming it is observable, as in Lieberman and Phillips (2017*a*, 2018).

ACKNOWLEDGEMENTS

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, IESEG, Oxford University, HEC Montreal, Paris-Ouest Nanterre, Hannover. Chevillon and Kratz acknowledge research support from the Labex MME-DII (ANR-11-LBX-0023-01), Chevillon ackowledges research support from CREST.

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

Bertoin, J. and M. Yor (2002). On the entire moments of self-similar Markov processes and exponential functionals of Lévy processes. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, Volume 11, pp. 33–45.

Bertoin, J., M. Yor, et al. (2005). Exponential functionals of Lévy processes. *Probability Surveys 2*, 191–212.

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.

Christensen, K., R. Oomen, and R. Renò (2018). The drift burst hypothesis. Working paper, Available at SSRN 2842535.

Csörgő, 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.

Engsted, T. and B. Nielsen (2012). Testing for rational bubbles in a coexplosive vector autoregression. *The Econometrics Journal 15*(2), 226–254.

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 (2017). Bayesian analysis of bubbles in asset prices. *Econometrics 5*(4), 47–60.

Giraitis, L. and P. C. B. Phillips (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis 27*(1), 51–60.

Gospodinov, N. (2004). Asymptotic confidence intervals for impulse responses of near-integrated processes. *The Econometrics Journal 7*(2), 505–527.

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

Gürkaynak, R. S. (2008). Econometric tests of asset price bubbles: Taking stock. *Journal of Economic Surveys 22*(1), 166–186.

Hall, P. and C. C. Heyde (1980). Martingale Limit Theory and its Application. In Z. W. Birnbaum and E. Lukacs (Eds.), *Martingale Limit Theory and its Application.* Academic Press.

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.

Harvey, D. I., S. J. Leybourne, and R. Sollis (2013). Recursive right-tailed unit root tests for an explosive asset price bubble. *Journal of Financial Econometrics 13*(1), 166–187.

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.

Horváth, L. and L. Trapani (2019). Testing for randomness in a random coefficient autoregression model. *Journal of Econometrics 209*(2), 338–352.

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.

Jakubowski, A., J. Mémin, and G. Pagès (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbb{D}^1$ de Skorokhod. *Probability Theory and Related Fields 81*(1), 111–137.

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 (2014). Norming rates and limit theory for some time-varying coefficient autoregressions. *Journal of Time Series Analysis 35*(6), 592–623.

Lieberman, O. and P. C. B. Phillips (2017a). Hybrid stochastic local unit roots. CFDP 2113, Yale University.

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

Lieberman, O. and P. C. B. Phillips (2018). IV and GMM inference in endogenous stochastic unit root models. *Econometric Theory 34*(5), 1065–1100.

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. *Probability 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. 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*(1-2), 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. and S.-P. Shi (2018). Financial bubble implosion and reverse regression. *Econometric Theory 34*(4), 705–753.

Phillips, P. C. B., S.-P. 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.

Strassen, V. et al. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1.* The Regents of the University of California.

Tao, Y., P. C. B. Phillips, and J. Yu (2017). Random coefficient continuous systems: Testing for extreme sample path behaviour. CFDP 2114, Yale University.

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.

Wu, Y. (1995). Are there rational bubbles in foreign exchange markets? Evidence from an alternative test. *Journal of International Money and Finance 14*(1), 27–46.

APPENDIX A: 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), Granger and Swanson (1997) and Horváth and Trapani (2019). 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., $\alpha = 0$ 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.[11] 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).[12] 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 (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 (2017) where bubbles have constant expected duration. Yet, in our model the distribution of ρ_t is allowed to be bi- or multimodal 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 A.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 $\mathsf{E}[\rho_t] - 1 = O\left(T^{-\alpha}\right)$ with $\alpha \in (1/2, 1)$, and $\mathsf{V}[\rho_t] = o\left(T^{-1}\right)$ which implies that $\mathsf{V}[\rho_t]$ lies in a tighter neighbourhood of zero and so does not asymptotically impact[13] 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 A.2. *Proposals were developed in parallel to our work by Lieberman and Phillips (2014, 2017a, 2017b 2018).*
(i) In their 2014 article, they study a model where ρ_t (using the notation above) is time

[11]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).

[12]We assume η_t homoskedastic since expression (2.1) implies that y_t exhibits conditional heteroskedasticity: for $\rho_t \sim i.i.d.\left(\rho, \sigma_\rho^2\right)$ then

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

see inter alia Tsay (1987) and Hwang and Basawa (2005). These authors, as well as others, have also proposed functional forms that differ from (2.1) and that belong to the classes of double-autoregressive or bilinear processes.

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

varying but non-stochastic (such as would be in our setting, conditioning on the history $\{u_t\}_{1\leq t\leq T}$). 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[14]*, which makes extending their QMLE results to our setting difficult.*
(*ii*) *In Lieberman and Phillips (2017b), they build on the last part of the 2014 paper: 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, that $\phi = 0$ and $\alpha = 1$. Lieberman and Phillips (2018) extends the results to a model with non iid errors and they consider instrumental variable estimation. Lieberman and Phillips (2017a) allows for unrestricted values of ϕ and assumes u_t observed.*[15] *Here, by contrast, we assume $\alpha < 1$, so we consider cases where the variance of ρ_t is of higher magnitude, our focus being on dynamics that may possibly be characterised as near explosive. We show in this paper that the implications differ. Also, we do not assume that the modeler observes u_t or an instrument that correlates with it.*

APPENDIX B: 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 (r, s) and consider short lived bubbles where $h_{(T,s)}/t_{(T,r)} \to 0$, i.e., the bubble is short not only with respect to the sample size but also compared to the sample observed until its inception and termination.

Models that do not allow for bubble episodes In all examples below, for all $\gamma > 1$, $\lim_{T\to\infty} \Pr(g_{t_{(T,r)},h_{(T,s)}} > \gamma \big| \mathcal{I}_{t_{(T,r)}}) = 0$. 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, $t_{(T,r)}$ and $h_{(T,s)}$, are assumed to diverge to $+\infty$ as $T \to \infty$ with $h_{(T,s)}/t_{(T,r)} \to 0$. For notational simplicity we let t, h denote $t_{(T,r)}, h_{(T,s)}$ 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 $g_{t,h} \sim_p (1 + h/t)^\nu \overset{p}{\to} 1$.
2 *(Near-)Stochastic trend*: $y_t = \rho_t y_{t-1} + \eta_t$ with $\rho_t - 1 = O_p(T^{-1})$. If ρ_t is nonstochastic, y_t satisfies for $(t_{(T,r)}/T, \bar{t}_{(T,\bar{r})}/T) \to (r, \bar{r})$, $y_{\bar{t}}/y_t \Rightarrow 1 + (W_c(\bar{r}) - W_c(r))/W_c(r)$ where W_c denotes an Ornstein-Uhlenbeck process with parameter $c = \lim_{T\to\infty} T(\rho_T - 1)$. Letting $\bar{t} = t + h$, then $g_{t,h} = 1 + O_p(\sqrt{\frac{h}{t}})$ and $g_{t,h} \overset{p}{\to} 1$. If $\rho_t \sim iid$, the result is proven in Corollary 3.1 as the NERC with $\alpha = 1$.

[14]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 $T \to \infty$.
[15]This paper also relates to Tao, Phillips and Yu (2017) who consider estimating a continous time process using sequential in-fill & block asymptotics.

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 $(t_{(T,r)}/T, \bar{t}_{(T,\bar{r})}/T) \to (r, \bar{r})$, $y_{\bar{t}}/y_t \Rightarrow (1 + (W(\bar{r}) - W(r))/W(r))^\lambda$. Letting $\bar{t} = t + h$ then $g_{t,h}^{1/\lambda} = 1 + O_p(\sqrt{\frac{h}{t}})$ hence $g_{t,h} \overset{p}{\to} 1$ unless we assume that λ itself is a function of T and that $\lambda\sqrt{h/t} \not\to 0$, which requires $\lambda \to \infty$ as $T \to \infty$.

Models that allow for temporary bubble episodes

1 *(Temporary-) Explosive process*: $y_t = \rho_t y_{t-1} + x_t$ where, for $s \in [t, t+h], \rho_s \geqslant 1 + \epsilon$, $\epsilon > 0$, satisfies for all $\gamma > 1$, $\lim_{T\to\infty} \Pr(g_{t,h} > \gamma | \mathcal{I}_t) = 1$.
2 *Phillips-Wu-Yu* (2011) temporary near explosive bubble: $y_t = \rho_t y_{t-1} + x_t$ where, for $s \in [t, t+h], \rho_s = e^{\phi/T^\alpha}$, $\phi > 0$ and $\alpha \in (0,1)$, satisfies, for $s \in (0, (T+H)\,T^{-\alpha}]$ and $h = \lfloor T^\alpha s \rfloor$, $\lim_{T\to\infty} \Pr(g_{t,h} = e^{\phi s} | \mathcal{I}_r) = 1$
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 $\lim_{T\to\infty} \Pr(g_{t,h} = \phi | \mathcal{I}_r) = 1$, where $\phi = \lim_{T\to\infty} (\rho\pi)^h$. Bubbles arise if $\phi > 1$, i.e., if $h^{-1}(\rho\pi - 1)$ 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_{t+h}/y_t is degenerate under the assumption of a short-lived bubble episode $h/t \to 0$ as $T \to \infty$. This is not the case under the NERC model with $\alpha \in (0,1)$ and hence this allows to perform conditional probabilistic statements.

APPENDIX C: 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 (2.1) can be written as $y_t = \rho y_{t-1} + (\rho_t - \rho)\, y_{t-1} + \eta_t$. Hence, the OLS estimator satisfies

$$\widehat{\rho} - \rho = \frac{S_{yy\rho}}{S_{yy}} + \frac{S_{y\eta}}{S_{yy}}, \tag{C.1}$$

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, for generic parameters μ, ν and *iid* processes u_t, ε_t such that $T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} (u_t, \varepsilon_t)' \Rightarrow (W_r, \sigma_\varepsilon B_r)'$, where $(W_r, B_r)'$ denotes a bivariate standard motion with independent components, we let the Geometric Brownian Motion $G_{(\mu,\nu)}(r) = \exp(\mu r + \nu W_r)$. We prove in lemma D.2 the following result:

$$\mathsf{E}\left[\int_0^{T^{1-\alpha}} G_{(\mu,\nu)}^2(r)\, dr\right]^{-1} \int_0^{T^{1-\alpha}} G_{(\mu,\nu)}^2(r)\, dr \Rightarrow \mathsf{F}_{2\mu,2\nu}$$

$$\mathsf{E}\left[\int_0^{T^{1-\alpha}} G_{(\mu,\nu)}^2(r)\, dr\right]^{-1/2} \int_0^{T^{1-\alpha}} G_{(\mu,\nu)}(r)\, dB_r \Rightarrow \sigma_\varepsilon \mathsf{D}_{\mu,\nu}^{(u,\varepsilon)}$$

where $\sigma_\varepsilon \mathsf{D}_{\mu,\nu}^{(u,\varepsilon)}$ has characteristic function $\mathsf{E}\left[\exp\left(-\frac{1}{2}\mathsf{F}_{2\mu,2\nu}t^2\right)\right]$.[16] We can now provide our result on the weak convergence of the sample moments:

[16] Matsumoto and Yor (2005), Theorem 7.4, specify how the distribution of $\mathsf{F}_{2\phi,2\lambda}$ can be expressed (for some values of the parameters) in terms of transforms of Brownian motions involving a Gamma variable.

LEMMA C.1. *Let the process y_t be defined for $t \geq 0$ by (2.1) under Assumptions **A** and **B**, and where $\alpha \in (4/5, 1)$. Then as $T \to \infty$,*
(i) If $c < 0$,

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

(ii) If $c \geq 0$,

$$T^{-2\alpha}\left[\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-2} S_{yy} \Rightarrow \sigma_\eta^2 \mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2, \quad T^{-\alpha}\left[\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-1} S_{y\eta} \Rightarrow \sigma_\eta^2 \mathsf{D}_{\phi,\lambda}^{(u,\eta)} \mathsf{D}_{-\phi,-\lambda}^{(u,\eta)},$$

$$T^{-\alpha}\left[\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-1} S_{yy\rho} \Rightarrow \sigma_\eta^2 \lambda \mathsf{D}_{2\phi,2\lambda}^{(u,u)}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2,$$

with the scaling function

$$\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)} = \begin{cases} e^{\left(|\phi|+\lambda^2\right)\lfloor T^{1-\alpha}\rfloor}, & \text{if } \phi \neq 0, \\ e^{2\lambda^2\lfloor T^{1-\alpha}\rfloor}, & \text{if } \phi = 0. \end{cases} \tag{C.2}$$

Lemma C.1 implies that 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 (C.1). It also differs from the fixed-asymptotics framework of Hwang and Basawa (2005) where the ratio $S_{yy\rho}/S_{yy}$ diverges, making the OLS estimator inconsistent.

The situation is different for the NERC model, as Proposition 4.1 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 $T^{\frac{1+\alpha}{2}}(\widehat{\rho} - \rho) \xrightarrow{d} \mathsf{N}(0, -2\phi)$. 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 4.1 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 $(2\phi)^{-1} T^\alpha e^{\phi T^{1-\alpha}}(\widehat{\rho} - \rho)$ tends to a standard Cauchy variable. Convergence can be arbitrarily slow here if α is close to zero: although we only prove the results for $\alpha \in (4/5, 1)$, notice 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 which are not Gaussian here. This implies that $\mathsf{F}_{2\phi,2\lambda}$ does not define a Cauchy variable contrary to the limiting distribution in PM.

Proposition 4.1 shows that $\widehat{\rho}$ allows to estimate $\phi + \lambda^2/2$ consistently when $c < 0$ since the higher order term in expression (2.3) 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 4.1 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).

Power

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

$$\delta_{\theta_0,T} \underset{\mathsf{H}_1}{=} \begin{cases} O_p\left(T^{\frac{1-\alpha}{2}}\right), & \text{if } c_0 < 0, \\ \phi_1 - \phi_0 + \left(\lambda_1^2 - \lambda_0^2\right)/2 + o_p(1), & \text{if } c_0 \geq 0 \text{ and } c_1 < 0, \\ \phi_1 - \phi_0 + \left(\lambda_1^2 - \lambda_0^2\right)/2 + \mathsf{D}_{\theta_1} + o_p(1), & \text{if } c_0 \geq 0 \text{ and } c_1 \geq 0, \end{cases} \tag{C.3}$$

where $c_1 = \phi_1 + \lambda_1^2$ and $\mathsf{D}_\theta = \left(\lambda \mathsf{D}^{(u,u)}_{2\phi,2\lambda} + \mathsf{D}^{(u,\eta)}_{\phi,\lambda} / \mathsf{D}^{(u,\eta)}_{-\phi,-\lambda} \right) / \mathsf{F}_{2\phi,2\lambda}$.

Expression (C.3) 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. (Expression (C.3)) 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).

□

APPENDIX D: PROOFS OF THE PROPOSITIONS

D.1. Proof of Proposition 2.1

Let y_t be defined for $t \geq 0$ by expression (2.1) under Assumptions **A** and **B**. Hence y_0 plays no role asymptotically, and we set it to zero in the following. The process y_t then satisfies

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

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, denoted PM in the proofs). Our proofs follow the lines of PM so we occasionally skip some of the details that can be found there, focusing instead on the specificities of our modelling strategy. Letting, for $t = 1$ to T, $t = \lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor$ for $j = 0, \cdots, \lfloor T^{1-\alpha} \rfloor - 1$, and for some $p \in [0, 1]$, we can write

$$T^{-\alpha/2} y_{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor}$$
$$= \sigma_\eta \sum_{i=1}^{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor} e^{\frac{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor - i}{T^\alpha} \phi + \lambda \frac{U_{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor} - U_i}{T^{\alpha/2}}} \frac{\eta_i}{\sqrt{\sigma_\eta^2 T^\alpha}}$$
$$= \sigma_\eta \int_0^{j + \lfloor T^\alpha p \rfloor / T^\alpha} e^{\frac{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor - \lfloor T^\alpha s \rfloor}{T^\alpha} \phi + \lambda \frac{U_{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor} - U_{\lfloor T^\alpha s \rfloor}}{T^{\alpha/2}}} dB_{T^\alpha}(s),$$

using Proposition A1 in PM – their Expression (46) – and where $B_{T^\alpha}(\cdot)$ denotes the partial sum on the Skorokhod space $D[0, \infty)$:

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

When applying the Functional Central Limit Theorem (FCLT) to the process W_{T^α} defined on $D[0, \infty)$ by $W_{T^\alpha}(s) \equiv T^{-\alpha/2} U_{\lfloor T^\alpha s \rfloor}$, 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 (D.1) 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. Convergence of $(W_{T^\alpha}, B_{T^\alpha})$ holds jointly so we can deduce, using Theorem 2.4 in Jakubowski *et al.* (1989), that, as $T \to \infty$,

$$\int_0^{j + \lfloor T^\alpha p \rfloor / T^\alpha} e^{\phi \frac{\lfloor T^\alpha j \rfloor + \lfloor T^\alpha p \rfloor - \lfloor T^\alpha s \rfloor}{T^\alpha} + \lambda (W_{T^\alpha}(j + \lfloor T^\alpha p \rfloor / T^\alpha) - W_{T^\alpha}(s))} dB_{T^\alpha}(s)$$

converges under Skorokhod's topology to

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

where W, B are the standard Brownian motions defined previously. Corollary 2.1 follows since the proof above also holds when $\alpha = 1$. □

D.2. Proof of Proposition 2.2

Consider the projection

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

Let $(r, r+s) \in (0, (T+H)\, T^{-\alpha}]^2$, with $s > 0$, then

$$\frac{y_{\lfloor T^{\alpha}(r+s)\rfloor}}{y_{\lfloor T^{\alpha} r\rfloor}} = e^{T^{-\alpha}\left((\lfloor T^{\alpha}(r+s)\rfloor - \lfloor T^{\alpha} r\rfloor)\phi+\lambda T^{\alpha/2}\sum_{j=T^{\alpha}r+1}^{\lfloor T^{\alpha}(r+s)\rfloor} u_j\right)}$$

$$+ \frac{1}{y_{\lfloor T^{\alpha} r\rfloor}} \sum_{i=\lfloor T^{\alpha} r\rfloor+1}^{\lfloor T^{\alpha}(r+s)\rfloor} e^{T^{-\alpha}\left((\lfloor T^{\alpha}(r+s)\rfloor - \lfloor T^{\alpha} r\rfloor - i)\phi+\lambda T^{\alpha/2}\sum\limits_{j=i+1}^{\lfloor T^{\alpha}(r+s)\rfloor} u_j\right)} \eta_i.$$

By definition of W, $T^{-\alpha/2}\sum_{j=\lfloor T^{\alpha} r\rfloor+1}^{\lfloor T^{\alpha}(r+s)\rfloor} u_j \Rightarrow W_{r+s} - W_r$,
so $e^{T^{-\alpha}\left((\lfloor T^{\alpha}(r+s)\rfloor - \lfloor T^{\alpha} r\rfloor)\phi+\lambda T^{\alpha/2}\sum_{j=\lfloor T^{\alpha}r\rfloor+1}^{\lfloor T^{\alpha}(r+s)\rfloor} u_j\right)} \Rightarrow e^{s\phi+\lambda(W_{r+s}-W_r)}$. Hence, using the same arguments as in the proof of Proposition 2.1, we obtain

$$T^{-\alpha/2}\left[y_{\lfloor T^{\alpha}(r+s)\rfloor} - e^{T^{-\alpha}\left((\lfloor T^{\alpha}(r+s)\rfloor - \lfloor T^{\alpha} r\rfloor)\phi+\lambda T^{\alpha/2}\sum_{j=\lfloor T^{\alpha}r\rfloor+1}^{\lfloor T^{\alpha}(r+s)\rfloor} u_j\right)} y_{\lfloor T^{\alpha} r\rfloor}\right]$$

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

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

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[17] a Cauchy variable C such that

$$\frac{y_{\lfloor T^{\alpha}(r+s)\rfloor}}{y\lfloor T^{\alpha} r\rfloor} \Rightarrow e^{\phi s+\lambda(W_{r+s}-W_r)} + D_{\theta}(r,s)\, C, \tag{D.2}$$

where $\theta = (\phi, \lambda)'$ and, using the notation of Appendix C,

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

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

$D_{\theta}(r,s)$ is independent of B, hence the second result in the proposition, conditionally on $\mathcal{I}_t$. The results above hold whether $\alpha \in (0,1)$ or $\alpha = 1$.

[17] Since C is Cauchy for all realisations of W.

D.3. Proof of Lemma C.1

Recalling that $c = \phi + \lambda^2$, we consider two cases depending on the sign of c, first $c < 0$ and then $c \geq 0$.

D.3.1. ***Case*** $\mathbf{c < 0}$. Recall that Proposition 2.1 gives, as $T \to \infty$, $T^{-\alpha/2} y_{\lfloor T^\alpha r \rfloor} \Rightarrow K_{\phi,\lambda}(r)$ with zero expectation and variance $\frac{e^{2cr}-1}{2c}\sigma_\eta^2$. Letting, for $c < 0$, $K^*_{\phi,\lambda}(r) = e^{cr} K^*_{\phi,\lambda}(0) + K_{\phi,\lambda}(r)$ such that $K^*_{\phi,\lambda}(r)$ is mixed Gaussian (conditional on W) and, unconditionally, with zero expectation and variance $-\frac{\sigma_\eta^2}{2c}$ and is stationary. Hence the proof in PM, Section 3, expression (13), applies here also. We can deduce from the definition of the Brownian motions B and W in Section 2, and via Chebyshev's inequality, that $T^{-(1+\alpha)} \sum_{t=1}^T y_t^2 \xrightarrow{p} \mathsf{E}[K^*_{\phi,\lambda}(r)^2] = \frac{-\sigma_\eta^2}{2c}$ 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 array $\xi_t \equiv T^{-\frac{1+\alpha}{2}} y_{t-1}^2 (\rho_t - \rho)$ which admits a conditional variance satisfying $\sum_{t=1}^T \mathsf{E}_{t-1}\left(\xi_t^2\right) = \frac{\lambda^2}{T^{1+\alpha}} \sum_{t=1}^T y_{t-1}^4 \xrightarrow{p} \frac{3\lambda^2\sigma_\eta^4}{4c^2}$, using the consistency of the empirical estimator of the kurtosis. Since $\sum_{t=1}^T y_{t-1}^4$ is of the same magnitude as $\sum_{t=1}^T y_{t-1}^2$, we can extend the results of PM, from $\lambda = 0$ to $\lambda \neq 0$. Indeed, the proof of Expression (5) in Phillips and Magdalinos (2007, p. 127) also applies here with our definition of ξ_t and appropriate implied modifications. They prove a Lindeberg condition that for all $\varepsilon > 0$, $\sum_{t=1}^T \mathsf{E}_{t-1}\left(\xi_t^2 1_{\{|\xi_t| > \varepsilon\}}\right) \xrightarrow{p} 0$ as $T \to \infty$. We adapt this result and prove it in the context of our framework in the Supplementary Appendix. It follows from Pollard (1984, Theorem VIII.1 p. 171) that, as $T \to \infty$,

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

D.3.2. ***Case*** $\mathbf{c \geq 0}$. The proof follows the main schemata given in PM, and retain their notation, namely

$$\kappa_T = T^\alpha \left\lfloor T^{1-\alpha} \right\rfloor \quad \text{and} \quad q_T = T^{1-\alpha} - \left\lfloor T^{1-\alpha} \right\rfloor. \tag{D.3}$$

We need to derive the asymptotic behaviors of the sample variance and covariances of y_t. For this, we provide three lemmas: the first lemma (Lemma D.2) provides an asymptotic convergence of a certain type of partial sum. Although this constitutes only an intermediate result, we provide this lemma here as it serves also to notation $\mathsf{D}^{(\cdot,\cdot)}_{\cdot,\cdot}$ for a class of distributions used extensively in the remainder. The subsequent two lemmas provide the asymptotic distributions of the sample variance of y_t, $\sum_{t=1}^T y_t^2$ in Lemma D.3, and of the sample covariances $\sum_{t=1}^T y_{t-1}\eta_t$ and $\sum_{t=1}^T y_{t-1}(\rho_t - \rho)$ in Lemma D.4. For readability, the proofs are relegated to Section D.5.

We now state our first intermediary result, where we must stress that we do not assume that u_t, ε_t are independent as we will use $\varepsilon_t = u_t$ in some applications of the lemma. We do not use the notation (ϕ, λ) in the lemma but refer to (μ, ν) as we apply it below to values $(\mu, \nu) \neq (\phi, \lambda)$.

Lemma D.2. *Let (u_t, ε_t) denote an iid sequence with zero expectation and variances $\sigma_u^2 = 1$, $\sigma_\varepsilon^2 > 0$ (the distribution of (u_t, ε_t) can be expressed as an array that depends on*

T) with $U_t = \sum_{j=1}^{t} u_j$.
Let, for $(\mu,\nu) \in \mathbb{R} \times \mathbb{R}_{+,*}$

$$\psi_T^{(\mu,\nu)} = \begin{cases} 1, & \text{if } \mu < 0, \\ e^{\left(\mu+\nu^2\right)T}, & \text{if } \mu \geq 0, \end{cases}$$

and

$$\zeta_{T,t}^{(\mu,\nu)} = T^{-\alpha/2}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-1}\exp\left(\frac{\mu}{T^{\alpha}}t + \frac{\nu}{T^{\alpha/2}}U_t\right)\varepsilon_t.$$

Then there exists a non degenerate random variable $\mathsf{D}_{\mu,\nu}^{(u,\varepsilon)}$ such that, as $T\to\infty$,

$$\sum_{t=1}^{T}\zeta_{T,t}^{(\mu,\nu)} \Rightarrow \sigma_\varepsilon \mathsf{D}_{\mu,\nu}^{(u,\varepsilon)}. \tag{D.4}$$

Proof. See Section D.5.

Lemma D.3. *Under the assumptions of Lemma C.1 and in the case $c \geq 0$, as $T\to\infty$,*

$$\sigma_\eta^{-2}T^{-2\alpha}\left[\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-2}\sum_{t=1}^{T}y_t^2 \Rightarrow \mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2,$$

where the scaling function $\chi_T^{(\phi,\lambda)}$ is given in Expression (C.2) and

$$T^{-\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-2}\sum_{t=1}^{\lfloor\kappa_T\rfloor}e^{2\phi T^{-\alpha}k+2\lambda T^{-\alpha/2}U_k} \Rightarrow \mathsf{F}_{2\phi,2\lambda} \tag{D.5}$$

which is nondegenerate. The scaling function $\psi_{\cdot}^{(\cdot,\cdot)}$ and the random variable $\mathsf{D}_{\cdot,\cdot}^{(\cdot,\cdot)}$ are defined in Lemma D.2.

Proof. See Section D.5.

Lemma D.4. *Under the assumptions of Lemma C.1 and in the case $c \geq 0$, as $T\to\infty$,*

$$\sigma_\eta^{-2}T^{-\alpha}\left[\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-1}\sum_{t=1}^{T}y_{t-1}\eta_t \Rightarrow \mathsf{D}_{\phi,\lambda}^{(u,\eta)}\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)} \tag{D.6}$$

and

$$\sigma_\eta^{-2}T^{-\alpha}\left[\chi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-2}\sum_{t=1}^{T}y_{t-1}^2\left(\rho_t-\rho\right) \Rightarrow \lambda\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2\mathsf{D}_{2\phi,2\lambda}^{(u,u)} \tag{D.7}$$

where the scaling function $\chi_{\cdot}^{(\cdot,\cdot)}$ is defined in Lemma D.3, and the random variable $\mathsf{D}_{\cdot,\cdot}^{(\cdot,\cdot)}$ in Lemma D.2.

Proof. See Section D.5.

We summarise in the tables below the results obtained above for the process y_t defined as in (2.1) and under Assumptions A and B. We consider the three cases $c < 0$, $c = 0$

and $c > 0$, and introduce the notation:

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

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}	$\mu^{y\eta}$	$\mu^{yy\rho}$
$c < 0$	$-2c$	$\sqrt{-2c}$	$-2c/\left(\sqrt{3}\lambda\right)$
$c \geq 0$	1	1	λ^{-1}

with

	ϕ_T^{yy}	$\phi_T^{y\eta}$	$\phi_T^{yy\rho}$
$c < 0$	$T^{-(1+\alpha)}$	$T^{-\frac{1+\alpha}{2}}$	$T^{-\frac{1+\alpha}{2}}$
$c \geq 0,\ \phi < 0$	$T^{-2\alpha}e^{-2\left(-\phi+\lambda^2\right)T^{1-\alpha}}$	$T^{-\alpha}e^{-2\left(-\phi+\lambda^2\right)T^{1-\alpha}}$	$T^{-\alpha}e^{-2\left(-\phi+\lambda^2\right)T^{1-\alpha}}$
$c \geq 0,\ \phi = 0$	$T^{-2\alpha}e^{-4\lambda^2 T^{1-\alpha}}$	$T^{-\alpha}e^{-4\lambda^2 T^{1-\alpha}}$	$T^{-\alpha}e^{-4\lambda^2 T^{1-\alpha}}$
$c \geq 0,\ \phi > 0$	$T^{-2\alpha}e^{-2\left(\phi+\lambda^2\right)T^{1-\alpha}}$	$T^{-\alpha}e^{-2\left(\phi+\lambda^2\right)T^{1-\alpha}}$	$T^{-\alpha}e^{-2\left(\phi+\lambda^2\right)T^{1-\alpha}}$

and

	D_{yy}	$D_{y\eta}$	$D_{yy\rho}$
$c < 0$	1	$\mathsf{N}(0,1)$	$\mathsf{N}(0,1)$
$c \geq 0$	$\mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2$	$\mathsf{D}_{\phi,\lambda}^{(u,\eta)}\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}$	$\mathsf{D}_{2\phi,2\lambda}^{(u,u)}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2$

The tables above directly provide Lemma C.1.

D.4. Proof of Proposition 4.1

The OLS estimator given by $\widehat{\rho} = \dfrac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2}$ satisfies

$$\widehat{\rho} - \rho = \frac{\sum_t y_{t-1}^2 (\rho_t - \rho)}{\sum_t y_{t-1}^2} + \frac{\sum_t y_{t-1}\eta_t}{\sum_t y_{t-1}^2} \tag{D.8}$$

where $\rho = \mathsf{E}(\rho_t)$. Hence, Proposition 4.1 follows directly from Lemma C.1. 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

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

which proves Proposition 4.1(i).

When $c \geq 0$, we obtain, since $\phi_T^{yy\rho} = \phi_T^{y\eta}$,

$$T^{\alpha}\left(\widehat{\rho}-\rho\right) \Rightarrow \frac{\lambda \mathsf{D}_{2\phi,2\lambda}^{(u,u)}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2}{\mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2} + \frac{\mathsf{D}_{\phi,\lambda}^{(u,\eta)}\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}}{\mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2} = \frac{\lambda \mathsf{D}_{2\phi,2\lambda}^{(u,u)} + \mathsf{D}_{\phi,\lambda}^{(u,\eta)}/\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}}{\mathsf{F}_{2\phi,2\lambda}},$$

where the various ratios are calculated from Lemma C.1. They provide the same results for all cases when $c \geq 0$. Proposition 4.1(ii) follows. □

D.5. Proof of Intermediary Lemmas

D.5.1. Proof of Lemma D.2 We use Theorem 3.5 of Hall and Heyde (1980). From the definition of $\zeta_{T,t}^{(\mu,\nu)}$ and $\psi_T^{(\mu,\nu)}$, we see that, denoting by $\mathcal{F}_{T,t}$ the filtration associated with $(u_s, \varepsilon_s)_{s\leq t}$,

$$\begin{aligned}
&\sum_{t=1}^{\lfloor\kappa_T\rfloor} \mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right] \\
&= \left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2} T^{-\alpha} \sum_{t=1}^{\kappa_T} e^{\frac{2\mu}{T^{\alpha}}t + \frac{2\nu}{T^{\alpha/2}}U_{t-1}} \mathsf{E}_{\mathcal{F}_{t-1}}\left[e^{\frac{2\nu}{T^{\alpha/2}}u_t}\varepsilon_t^2\right] \\
&= \left(\sigma_\varepsilon^2 + \gamma T^{-\alpha} + O\left(T^{-2\alpha}\right)\right)\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2} \int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s + 2\nu W_{T^{\alpha}}(s)} ds,
\end{aligned}$$

where γ is a constant that depends on the covariance between u_t and ε_t (it is equal to $2\sigma_\varepsilon^2\nu^2$ when u_t and ε_t are independent). Hence there exists $M > 0$ such that,

$$\begin{aligned}
&\left|\sum_{t=1}^{\lfloor\kappa_T\rfloor} \mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right] - \sigma_\varepsilon^2\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+2\nu W(s)}ds\right| \\
&\leq \frac{M}{T^{\alpha}}\left|\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+2\nu W(s)}ds\right| \\
&+ M\left|\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+\nu W_s}\left[e^{2\nu(W_{T^{\alpha}}(s)-W_s)}-1\right]ds\right| \\
&\leq \frac{M}{T^{\alpha}}\left|\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+2\nu W(s)}ds\right| \\
&+ M\left\lfloor T^{1-\alpha}\right\rfloor \sup_{0\leq s\leq\lfloor T^{1-\alpha}\rfloor}\left|e^{2\nu(W_{T^{\alpha}}(s)-W_s)}-1\right|\left|\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+\nu W_s}ds\right|
\end{aligned}$$

By assumption, u_t admits bounded fourth moments, so Strassen (1967), Theorem 1.5, applies, giving

$$\left|W_{T^{\alpha}}(s) - W_s\right| = O_{a.s.}\left(\left(T^{-\alpha}\log\log T\right)^{1/4}\left(\log T\right)^{1/2}\right).$$

Hence,

$$\left|\sum_{t=1}^{\lfloor\kappa_T\rfloor} \mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right] - \sigma_\eta^2\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+2\nu W(s)}ds\right|$$
$$\leq 2MT^{1-5\alpha/4}\left|\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+2\nu W(s)}ds\right|,$$

and the latter is $o_p(1)$ for $\alpha \in (4/5, 1)$.

When $\mu < 0$, Bertoin and Yor (2002, 2005) show the moments of $\frac{-2\mu}{\nu^2}\int_0^\infty \exp(\mu s + \nu W_s)\,ds$ are those of an Inverse-Gamma, $\mathsf{Inv}\text{-}\Gamma\left(\frac{-2\mu}{\nu^2}, \frac{\nu^2}{2}\right)$, distributed random variable.[18] When u_t and ε_t are independent, we can use Theorem 3.5 of Hall and Heyde (1980), conditionally on $\int_0^\infty e^{2(\mu s+\nu W_s)}ds$, which gives

$$\frac{\sqrt{-\mu}}{\sigma_\varepsilon \nu}T^{-\alpha/2}\sum_{t=1}^{\lfloor\kappa_T\rfloor} e^{\frac{\mu}{T^\alpha}t+\frac{\nu}{T^{\alpha/2}}U_t}\eta_t \xrightarrow{L} \mathsf{N}\left(0, \frac{-\mu}{\nu^2}\int_0^\infty e^{2(\mu s+\nu W_s)}ds\right). \tag{D.9}$$

The mixture of a normal distribution $\mathsf{N}(\pi, \sigma^2)$ with $\sigma^2 \sim \mathsf{Inv}\text{-}\Gamma(a, b)$ characterizes a non-standardised t-distribution $\pi + \omega\mathcal{T}$ where $\mathcal{T}$ follows a t_n distribution, $n = 2a$ and $\omega^2 = b/a$. Here, since $\frac{-\mu}{\nu^2}\int_0^\infty \exp(2(\mu s + \nu W_s))\,ds \sim \mathsf{Inv}\text{-}\Gamma\left(\frac{-\mu}{\nu^2}, 2\nu^2\right)$, the limiting distribution in (D.9) is non-standardised t distributed with parameters $(\pi, \omega^2, n) = \left(0, \frac{2\nu^4}{-\mu}, \frac{2\phi}{\nu^2}\right)$.

When $\mu \geq 0$, we use Theorem 3.5 of Hall and Heyde (1980). We therefore check below that their three conditions (3.33), (3.34) and (3.35) hold.

The scaling function $\psi_\cdot^{(\cdot,\cdot)}$ is defined such that $\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\mathsf{E}\left[\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2(\mu s+\nu W_s)}ds\right] \to \frac{1}{2(\mu+\nu^2)}$ so, applying Theorem 7.4 of Matsumoto and Yor (2005),[19] there exists $\mathsf{F}_{2\mu,2\nu}$ with explicit distribution such that, as $T \to \infty$,

$$\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2}\int_0^{\lfloor T^{1-\alpha}\rfloor} e^{2\mu s+2\nu W_s}ds \Rightarrow \mathsf{F}_{2\mu,2\nu}. \tag{D.10}$$

Hence, when $\nu \neq 0$ and for all μ, as $T \to \infty$,

$$\mathsf{E}\left|\sum_{t=1}^{\lfloor\kappa_T\rfloor}\mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right] - \sigma_\varepsilon^2\mathsf{F}_{2\mu,2\nu}\right| \to 0,$$

[18] Bertoin and Yor (2005) consider the case of $\lambda = 1$ in their example page 199 but it is easy to compute the general case from their Theorem 3: $\mathsf{E}\left[I_\infty^{-k}\right] = \frac{b}{\sigma^2/2}\left(\frac{\sigma^2}{2}\right)^k \Gamma\left(k + 2b/\sigma^2\right)/\Gamma\left(2b/\sigma^2\right)$ where $I_\infty = \int_0^\infty \exp(-[bs + \sigma W_s])\,ds$. The moments of $\frac{2b}{\sigma^2}I_\infty^{-1}$ match those of a $\Gamma\left(2b/\sigma^2, \sigma^2/2\right)$-distributed variable.

[19] Since we refer to this theorem several times in the proof, we delineate its result in the notation of this paper (we set $m = -1$ in their theorem). These authors define

$$A_t^{(\mu)} = \int_0^t e^{2(\mu s+W_s)}ds, \text{ and } \Delta_t^{(\mu,-1)} = E\left[A_t^{(\mu)}\right].$$

Their Theorem 7.4 proves the convergence and provides the limit (in the weak sense) of the probability measure associated with $\left[\Delta_t^{(\mu,-1)}\right]^{-1}A_t^{(\mu)}$ as $t \to \infty$ for values of μ. Extending their result to $\int_0^t e^{2(\mu s+\nu W_s)}ds$ for $\nu \neq 1$ is straightfoward and this results shows that the latter integral, scaled by its expectation, admits a limit.

which proves condition (3.33) p. 71 of Hall and Heyde (1980).

Now,

$$\mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right] = \left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2} T^{-\alpha} e^{\frac{2\mu}{T^\alpha}t+\frac{2\nu}{T^{\alpha/2}}U_{t-1}} \mathsf{E}_{\mathcal{F}_{t-1}}\left[e^{\frac{2\nu}{T^{\alpha/2}}u_t}\varepsilon_t^2\right]$$
$$= \left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-2} T^{-\alpha} e^{\frac{2\mu}{T^\alpha}t+\frac{2\nu}{T^{\alpha/2}}U_{t-1}} \left[\sigma_\varepsilon^2+\gamma\sigma_\varepsilon^2 T^{-\alpha}+O\left(T^{-2\alpha}\right)\right]$$

The Law of Iterated Logarithms implies that $\max_{1\le t\le T} U_{t-1} = O\left(\sqrt{2T\log\log T}\right)$ so when $\mu<0$, $\max_{1\le t\le T}\exp\left(\frac{2\mu}{T^\alpha}t+\frac{2\nu}{T^{\alpha/2}}U_{t-1}\right)$ is bounded and $\max_{1\le t\le T}\mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right]\to 0$. When $\mu\ge 0$, as $T\to\infty$,

$$\max_{1\le t\le T} e^{\frac{2\mu}{T^\alpha}t+\frac{2\nu}{T^{\alpha/2}}U_{t-1}} \sim e^{2\mu T^{1-\alpha}+2\nu\sqrt{2T^{1-\alpha}\log\log T}}$$

so

$$\max_{1\le t\le T}\mathsf{E}_{\mathcal{F}_{T,t-1}}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2\right] \sim \sigma_\varepsilon^2 T^{-\alpha} e^{2\nu\left(\sqrt{2T^{1-\alpha}\log\log T}-T^{1-\alpha}\right)} \to 0.$$

This proves condition (3.34) of Hall and Heyde (1980).

There remains only to prove condition (3.35) of Hall and Heyde (1980), namely that $\mathsf{E}\left|\sum_{t=1}^{\lfloor\kappa_T\rfloor}\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2-\sigma_\varepsilon^2\mathsf{F}_{2\mu,2\nu}\right|\to 0$. The proof follows the same lines as before, since U_i and ε_j are independent for all $i<j$. Theorem 3.5 of Hall and Heyde (1980) follows. It states in particular that the Lindeberg condition is satisfied, i.e., for $\xi>0$, $\sum_t\mathsf{E}\left[\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2 1_{|\zeta_t|>\xi}\right]\to 0$. Hence, as $T\to\infty$,

$$\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(\mu,\nu)}\right]^{-1}\sum_{t=1}^{\lfloor\kappa_T\rfloor} e^{\frac{\mu}{T^\alpha}t+\frac{\nu}{T^{\alpha/2}}U_t}\varepsilon_t \Rightarrow \sigma_\varepsilon\mathsf{D}_{\mu,\nu}^{(u,\varepsilon)}$$

where $\sigma_\varepsilon\mathsf{D}_{\mu,\nu}^{(u,\varepsilon)}$ has characteristic function $\mathsf{E}\left[\exp\left(-\frac{\sigma_\varepsilon^2}{2}\mathsf{F}_{2\mu,2\nu}t^2\right)\right]$. Since $\kappa_T/T\to 1$ as $T\to\infty$ and the sum $\sum_{t=1}^{\lfloor\kappa_T\rfloor} e^{\frac{\mu}{T^\alpha}t+\frac{\nu}{T^{\alpha/2}}U_t}\varepsilon_t$ diverges for all parameter combinations, it must hold that $\sum_{t=\lfloor\kappa_T\rfloor+1}^{T}\left(\zeta_{T,t}^{(\mu,\nu)}\right)^2=o_p(1)$. The lemma follows. □

D.5.2. Proof of Lemma D.3 We write

$$T^{-2\alpha}\sum_{t=1}^{T}y_t^2 = T^{-2\alpha}\underbrace{\sum_{j=0}^{\lfloor T^{1-\alpha}\rfloor-1}\sum_{k=1}^{\lfloor T^\alpha\rfloor} y_{\lfloor T^\alpha j\rfloor+k}^2}_{S_{1T}} + T^{-2\alpha}\sum_{t=\lfloor\kappa_T\rfloor}^{T}y_t^2+O_p\left(T^{-\alpha}\right),$$

where we recall that $\lfloor\kappa_T\rfloor = T^\alpha\left\lfloor T^{1-\alpha}\right\rfloor$. 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

$$\lfloor\kappa_T\rfloor-1\le\lfloor\kappa_T-T^\alpha\rfloor+\lfloor T^\alpha\rfloor\le\lfloor\kappa_T\rfloor.$$

In the following, we study the asymptotic behavior of S_{1T}, proving first that, as $T\to\infty$,

$$\sigma_\eta^{-2}T^{-2\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(-\phi,\lambda)}\right]^{-2}\left[\varphi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-1}S_{1T}\Rightarrow\mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2, \quad \text{(D.11)}$$

where

$$\varphi_T^{(\phi,\lambda)} = \begin{cases} 1, & \text{if } \phi < 0, \\ T, & \text{if } \phi \geq 0,\ c = 0, \\ e^{2(\phi+\lambda^2)T}, & \text{if } \phi \geq 0,\ c \neq 0, \end{cases}, \tag{D.12}$$

the scaling function $\psi_{\cdot}^{(\cdot,\cdot)}$ and the random variable $\mathsf{D}_{\cdot,\cdot}^{(\cdot,\cdot)}$ are defined in Lemma D.2 and $\mathsf{F}_{2\phi,2\lambda}$ is the nondegenerate limit defined in (D.10) such that it is also the limit $T^{-\alpha}\left[\varphi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-1}\sum_{k=1}^{\lfloor\kappa_T\rfloor}\exp\left(\frac{2\phi}{T^\alpha}k+\frac{2\lambda}{T^{\alpha/2}}U_k\right) \Rightarrow \mathsf{F}_{2\phi,2\lambda}$. Notice in practice that $\phi \geq 0$, $c = 0$ does not occur so $\varphi_T^{(\phi,\lambda)} = \psi_T^{(\phi,\lambda)}$. By definition of $\lfloor\kappa_T\rfloor$, it is obvious that $\sum_{t=\lfloor\kappa_T\rfloor}^T y_t^2$ is negligible compared to $\sum_{t=1}^T y_t^2$ when the latter is not $O_p(1)$. Hence, Expression (D.11) implies that S_{1T} is the dominant term in the asymptotic behaviour of $\sum_{t=1}^T y_t^2$ and Lemma D.3 follows.

We now turn to the study of S_{1T}, proving expression (D.11). We start by noticing that

$$y_t = \sum_{i=0}^{t-1} e^{\frac{\phi}{T^\alpha}i+\frac{\lambda}{T^{\alpha/2}}[U_t-U_{t-i}]}\eta_{t-i} = \sum_{i=1}^{t} e^{\frac{\phi}{T^\alpha}(t-i)+\frac{\lambda}{T^{\alpha/2}}[U_t-U_i]}\eta_i$$
$$= e^{\frac{\phi}{T^\alpha}t+\frac{\lambda}{T^{\alpha/2}}U_t}\sum_{i=1}^{t} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i$$

so we obtain

$$\sum_{t=1}^T y_t^2 = \underbrace{\left(\sum_{k=1}^{\lfloor\kappa_T\rfloor} e^{\frac{2\phi}{T^\alpha}k+\frac{2\lambda}{T^{\alpha/2}}U_k}\right)\left[\sum_{i=1}^{\lfloor\kappa_T\rfloor} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i\right]^2}_{S_{0T}} + R_{0T}, \tag{D.13}$$

where the contribution of R_{0T} is shown to be negligible in the Supplementary Appendix and S_{0T} can be expressed as a product whose elements we consider in turn.

We first consider the case $\phi > 0$. When $\lambda \neq 0$, Lemma D.2 proves that the second element in the definition of S_{0T} in expression (D.13) admits the following limit:

$$\frac{\sqrt{\phi}}{\sigma_\eta\lambda^2}T^{-\alpha/2}\sum_{i=1}^{\lfloor\kappa_T\rfloor} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i \Rightarrow \frac{\sqrt{\phi}}{\lambda^2}\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)} \tag{D.14}$$

where $\frac{\sqrt{\phi}}{\lambda^2}\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}$ is non-standardised t distributed with $\left(\pi,\omega^2,n\right) = \left(0, \frac{2\lambda^4}{\phi}, \frac{2\phi}{\lambda^2}\right)$.[20]

Now, regarding the first element on the right-hand side of expression (D.13), we saw previously that

$$T^{-\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(2\phi,2\lambda)}\right]^{-1}\sum_{k=1}^{\lfloor\kappa_T\rfloor} e^{\frac{2\phi}{T^\alpha}k+\frac{2\lambda}{T^{\alpha/2}}U_k} \Rightarrow \mathsf{F}_{2\phi,2\lambda},$$

[20] For the case $\lambda = 0$, MP obtained that $\frac{\sqrt{2\phi}}{\sigma_\eta}T^{-\alpha}\sum_{i=1}^{\lfloor\kappa_T\rfloor} e^{-\frac{\phi}{T^\alpha}i}\eta_i \xrightarrow{L} N(0,1)$.

from which we deduce

$$\left[\sigma_\eta^{-1}T^{-\alpha/2}\right]^2 T^{-\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(2\phi,2\lambda)}\right]^{-1}\sum_{t=1}^{\lfloor\kappa_T\rfloor} y_t^2 \sim \sigma_\eta^{-2}T^{-2\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(2\phi,2\lambda)}\right]^{-1} S_{0T}$$
$$\Rightarrow \mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2.$$

Now, if $\phi < 0$, Lemma D.2 implies that

$$T^{-\alpha/2}e^{-\left(-\phi+\lambda^2\right)T^{1-\alpha}}\sum_{i=1}^{\lfloor\kappa_T\rfloor} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i \Rightarrow \sigma_\eta \mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}$$

and, as in the proof of Lemma D.2, we use the result of Bertoin and Yor (2002, 2005) , namely that $T^{-\alpha}\sum_{k=1}^{\lfloor\kappa_T\rfloor} e^{2\phi T^{-\alpha}k+2\lambda T^{-\alpha/2}U_k} \Rightarrow \mathsf{F}_{2\phi,2\lambda}$ which follows an Inv-Γ distribution. Hence,

$$T^{-2\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(-\phi,\lambda)}\right]^{-2}\left[\varphi_{\lfloor T^{1-\alpha}\rfloor}^{(\phi,\lambda)}\right]^{-1} S_{0T} = T^{-2\alpha}e^{-2\left(-\phi+\lambda^2\right)T^{1-\alpha}} S_{0T} \Rightarrow \sigma_\eta^2 \mathsf{F}_{2\phi,2\lambda}\left[\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)}\right]^2.$$

Finally, if $\phi = 0$, $T^{-\alpha/2}e^{-\lambda^2 T^{1-\alpha}}\sum_{i=1}^{\lfloor\kappa_T\rfloor} e^{-\lambda T^{-\alpha/2}U_i}\eta_i \Rightarrow \sigma_\eta \mathsf{D}_{0,-\lambda}^{(u,\eta)}$ by Lemma D.2 and $\left[\varphi_{\lfloor T^{1-\alpha}\rfloor}^{(0,\lambda)}\right]^{-1} T^{-\alpha}\sum_{k=1}^{\lfloor\kappa_T\rfloor} e^{2\lambda T^{-\alpha/2}U_k} \Rightarrow \mathsf{F}_{0,2\lambda}$, we can conclude

$$T^{-2\alpha}\left[\psi_{\lfloor T^{1-\alpha}\rfloor}^{(0,\lambda)}\right]^{-2}\left[\varphi_{\lfloor T^{1-\alpha}\rfloor}^{(0,\lambda)}\right]^{-1} S_{0T} = T^{-2\alpha}e^{-4\lambda^2 T^{1-\alpha}} S_{0T} \Rightarrow \sigma_\eta^2 \mathsf{F}_{0,2\lambda}\left[\mathsf{D}_{0,-\lambda}^{(u,\eta)}\right]^2.$$

□

D.5.3. Proof of Lemma D.4 We start with the proof of Expression (D.6), noticing that, for a generic *i.i.d.* process ν_t with variance σ_ν^2, we can write

$$\sum_{t=1}^{\lfloor\kappa_T\rfloor} y_{t-1}\nu_t = \sum_{t=1}^{\lfloor\kappa_T\rfloor} e^{\frac{\phi}{T^\alpha}t+\frac{\lambda}{T^{\alpha/2}}U_t}\nu_t \sum_{i=1}^{t} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i$$

and

$$\sum_{t=1}^{T} y_{t-1}\nu_t = \left[\sum_{t=1}^{\lfloor\kappa_T\rfloor} e^{\frac{\phi}{T^\alpha}t+\frac{\lambda}{T^{\alpha/2}}U_t}\nu_t\right]\left[\sum_{i=1}^{\lfloor\kappa_T\rfloor} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i\right] + R_{0T}^*,$$

where R_{0T}^* can be shown to be negligible (see the Supplementary Appendix). Hence, Lemma D.2 implies that, as $T \to \infty$,

$$T^{-\alpha}\left[\chi_{T^{1-\alpha}}^{(\phi,\lambda)}\right]^{-1}\sum_{t=1}^{T} y_{t-1}\nu_t \Rightarrow \sigma_\nu\sigma_\eta \mathsf{D}_{\phi,\lambda}^{(u,\nu)}\mathsf{D}_{-\phi,-\lambda}^{(u,\eta)},$$

where $\chi_T^{(\phi,\lambda)} = \psi_T^{(\phi,\lambda)}\psi_T^{(-\phi,\lambda)}$ is defined by Expression (C.2). Expression (D.6) follows.

We now prove Expression (D.7). When $\lambda \neq 0$,[21] the summation $\sum_{t=1}^{\lfloor\kappa_T\rfloor} y_{t-1}^2\left(\rho_t-\rho\right)$ can be expressed as

$$\sum_{t=1}^{\lfloor\kappa_T\rfloor} y_{t-1}^2\left(\rho_t-\rho\right) = \sum_{t=1}^{\lfloor\kappa_T\rfloor} e^{\frac{2\phi}{T^\alpha}t+\frac{2\lambda}{T^{\alpha/2}}U_t}\left[\sum_{i=1}^{t} e^{-\frac{\phi}{T^\alpha}i-\frac{\lambda}{T^{\alpha/2}}U_i}\eta_i\right]^2 \Delta U_{t+1}^{\rho},$$

[21]When $\lambda = 0$, the expression $\sum_{t=1}^{\lfloor\kappa_T\rfloor} y_{t-1}^2\left(\rho_t-\rho\right) \equiv 0$ by definition of ρ.

and

$$\sum_{t=1}^{T} y_{t-1}^2 (\rho_t - \rho) = \left[\sum_{i=1}^{\lfloor \kappa_T \rfloor} e^{-\frac{\phi}{T^\alpha} i - \frac{\lambda}{T^{\alpha/2}} U_i} \eta_i \right]^2 \left(\sum_{k=1}^{\lfloor \kappa_T \rfloor} e^{\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k} \Delta U_{k+1}^{\rho} \right) + R_{0T}^{\times}, \quad (D.15)$$

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 $R_{0T}^{\times}$ is asymptotically negligible with respect to the other terms of (D.15), as we show in the Supplementary Appendix.

We introduce the *i.i.d.* process u_t^* with bounded moments and nonzero variance, defined as

$$u_t^* = T^\alpha \left(e^{\frac{\lambda u_t}{T^{\alpha/2}}} - \mathsf{E}\left[e^{\frac{\lambda u_t}{T^{\alpha/2}}} \right] - \frac{\lambda}{T^{\alpha/2}} u_t \right) = T^\alpha \sum_{k=2}^{\infty} \frac{\lambda^k}{k! T^{\alpha k/2}} \left(u_t^k - \mathsf{E}\left[u_t^k \right] \right).$$

Then $\mathsf{E} u_t = \mathsf{E} u_t^3 = 0$ implies that $\mathsf{Cov}(u_t, u_t^*) = O\left(\frac{\lambda^3}{3!} T^{-\alpha/2} \right)$. Hence, we can write,

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

Therefore,

$$\sum_{k=1}^{\lfloor \kappa_T \rfloor} e^{\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k} \Delta U_{k+1}^{\rho} = \lambda T^{-\alpha/2} \sum_{k=1}^{\lfloor \kappa_T \rfloor} e^{\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k} \Delta U_{k+1} + T^{-\alpha} \sum_{k=1}^{\lfloor \kappa_T \rfloor} e^{\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k} u_{k+1}^*.$$

Lemma D.2 shows that the two partial sums in the latter equation have equal magnitude so the first sum dominates since the second is premultiplied by a term of lower magnitude in T. We obtain

$$\left[\psi_{\lfloor T^{1-\alpha} \rfloor}^{(2\phi, 2\lambda)} \right]^{-1} \sum_{k=1}^{\lfloor \kappa_T \rfloor} e^{\frac{2\phi}{T^\alpha} k + \frac{2\lambda}{T^{\alpha/2}} U_k} \Delta U_{k+1}^{\rho} \Rightarrow \lambda \mathsf{D}_{2\phi, 2\lambda}^{(u,u)}.$$

where $\psi_{\lfloor T^{1-\alpha} \rfloor}^{(2\phi, 2\lambda)} = \left[\psi_{\lfloor T^{1-\alpha} \rfloor}^{(\phi, \lambda)} \right]^2$. Hence, Lemma D.2 gives, as $T \to \infty$,

$$\begin{aligned} &\left[T^{-\alpha/2} \psi_{\lfloor T^{1-\alpha} \rfloor}^{(-\phi, -\lambda)} \right]^{-2} \left[\psi_{\lfloor T^{1-\alpha} \rfloor}^{(\phi, \lambda)} \right]^{-2} \sum_{t=0}^{\lfloor \kappa_T \rfloor} y_t^2 (\rho_{t+1} - \rho) \\ &= T^{-\alpha} \left[\chi_{\lfloor T^{1-\alpha} \rfloor}^{(\phi, \lambda)} \right]^{-2} \sum_{t=0}^{\lfloor \kappa_T \rfloor} y_t^2 (\rho_{t+1} - \rho) \Rightarrow \left[\sigma_\eta \mathsf{D}_{-\phi, -\lambda}^{(u,\eta)} \right]^2 \left[\lambda \mathsf{D}_{2\phi, 2\lambda}^{(u,u)} \right]. \end{aligned}$$

As before, the asymptotic limit also holds when the summation is up to T instead of $\lfloor \kappa_T \rfloor$. □