Local Explosion Modelling by Noncausal Process

Gouriéroux, Christian and Zakoian, Jean-Michel

CREST and University of Toronto, CREST and Université de Lille

5 May 2016

Local Explosion Modelling by Noncausal Process

Christian Gouriéroux

CREST and University of Toronto

Jean-Michel Zakoïan

CREST and Université de Lille

May 5, 2016

Abstract. The noncausal autoregressive process with heavy-tailed errors possesses a nonlinear causal dynamics, which allows for local explosion or asymmetric cycles often observed in economic and financial time series. It provides a new model for multiple local explosions in a strictly stationary framework. The causal predictive distribution displays surprising features, such as the existence of higher moments than for the marginal distribution, or the presence of a unit root in the Cauchy case. Aggregating such models can yield complex dynamics with local and global explosion as well as variation in the rate of explosion. The asymptotic behavior of a vector of sample autocorrelations is studied in a semi-parametric noncausal AR(1) framework with Pareto-like tails, and diagnostic tests are proposed. Empirical results based on the Nasdaq composite price index are provided.

Keywords: Causal innovation, Explosive bubble, Heavy-tailed errors, Noncausal process, Stable process.

Address for correspondence: Jean-Michel Zakoïan, CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff cedex. Email: zakoian@ensae.fr

1. Introduction

A number of economic and financial time series possess nonlinear dynamic features such as asymmetric cycles [Ramsey and Rothman (1996)] or speculative bubbles.[1] It has been

[1]Two formal definitions of bubbles exist in the economic literature. In Rational Expectation (RE) models for valuing firms, the value can be written as the sum of the "fundamental value", defined as the discounted sum of future dividends, plus an additional term which is called economic bubble by some authors [e.g. Flood and Garber (1980)]. On the other hand, equilibrium RE models

noted that the mixed causal and noncausal linear autoregressive (AR) processes often provide a better fit to such time series than the standard causal autoregressive moving-average (ARMA) processes [see e.g. Lanne et al. (2012), Lanne, Saikkonen (2011), Davis, Song (2012), Chen et al. (2012)]. The traditional Box-Jenkins methodology based on Gaussian linear processes were found insufficient to capture such features. Indeed, Gaussian AR processes are the only processes with both causal and noncausal strong linear AR representations[2] [see Cheng (1992), Breidt, Davis (1992)]. In contrast, a non-Gaussian linear noncausal process admits a nonlinear dynamics in direct time, which may produce local explosion whenever the errors distribution has fat tails.

The aim of this paper is to analyze the dynamic properties of heavy-tailed noncausal linear AR(1) processes that do not admit a causal linear AR representation, and to understand their potential usefulness in applications. In particular, it provides a new model for multiple local explosions in a strictly stationary framework. The transition distribution in direct time displays surprising features, such as the existence of higher moments than for the marginal distribution, or the presence of a unit root in the Cauchy case. We will also show how such processes can be combined to disentangle local and global explosive patterns and for modeling recursive explosions with different rates.

The paper is organized as follows. Section 2 reviews the main properties of strong linear processes in the presence of heavy-tailed errors. In particular, we explain how local explosion can arise when the linear process has a noncausal component and heavy-tailed errors. Section 3 considers noncausal AR(1) processes with stable errors. We characterize their stationary distribution and the existence of conditional moments. In the cases of the noncausal AR(1) models with Cauchy and Lévy errors, we derive the closed form formula of the conditional density in direct time. These results are used to obtain semi-strong causal representations of the process. Aggregation of noncausal AR(1) processes is studied in Section 4. Such aggregated processes are used to model local explosions with different rates of explosion. We explain how to identify the different components and to disentangle local

admit a multiplicity of solutions, some of them featuring local explosions followed by a crash. Such phenomena are called explosive or speculative bubbles by other authors [see e.g. Diba and Grossman (1988), Evans (1991)]. In this article we only consider the second concept, interpreting bubbles in a statistical sense, as local explosive behaviours.

[2]Any purely nondeterministic, second-order stationary process admits both a backward and a forward looking weak moving average representation. However, in these representations, the errors are only a sequence of centered, uncorrelated variables with a constant variance (that is, a weak white noise). They are not independent in general.

explosion components and global trend. In Section 5 we derive the asymptotic properties of the sample autocorrelations for the noncausal Cauchy AR process. Next we explain why the standard unit-root tests based on the detection of global stochastic trends can be misleading when looking for local explosions. Finally, we study diagnostic tools. Monte-Carlo experiments are presented in Section 6. An application on real data is proposed in Section 7. Section 8 concludes. Proofs of the propositions and complementary results are gathered in an appendix.

2. Strong linear processes

We consider a strong linear process (Y_t), that is a two-sided moving average process:

$$Y_t = \sum_{h=-\infty}^{\infty} a_h \varepsilon_{t-h}, \tag{2.1}$$

where (ε_t) is a sequence of independent and identically distributed (i.i.d.) real random variables, (a_h) is a sequence of real coefficients, satisfying for some $s \in (0,1)$,

$$E|\varepsilon_t|^s < \infty \quad \text{and} \quad \sum_{h=-\infty}^{\infty} |a_h|^s < \infty. \tag{2.2}$$

It follows from Proposition 13.3.1 in Brockwell and Davis (1991) that the process (Y_t) in (2.1) is well defined. When $a_h = 0$ for all $h < 0$, the process (Y_t) is called *purely causal* (with respect to (ε_t)); when $a_h = 0$ for all $h > 0$, (Y_t) is called *purely noncausal*. The uniqueness of the MA representation in (2.1) with heavy-tailed errors was recently studied by Gouriéroux and Zakoian (2015).

The trajectory of a strong linear process can be considered as a stochastic combination of baseline deterministic functions.

i) Let us consider a strong purely causal (or backward looking) process. This process can be written as: $Y_t = \sum_{\tau=-\infty}^{+\infty} \varepsilon_\tau \mathbb{1}_{\tau \le t} a_{t-\tau}$. Thus, the path of process (Y_t) is a combination of baseline paths $Z_\tau(t) = \mathbb{1}_{\tau \le t} a_{t-\tau}$, with stochastic i.i.d. coefficients ε_τ. Figure 1, left panel, provides an illustration for a causal AR(1) process, $Y_t = \rho Y_{t-1} + \varepsilon_t$ with $|\rho| < 1$ (thus $a_h = \rho^h$ for $h \ge 0$). The baseline path shows an upward jump followed by an exponential decrease.

ii) If now (Y_t) is a strong purely noncausal (or forward looking) process, we have:

$$Y_t = \sum_{\tau=-\infty}^{+\infty} \varepsilon_\tau \mathbb{1}_{\tau \ge t} a_{\tau-t}. \tag{2.3}$$

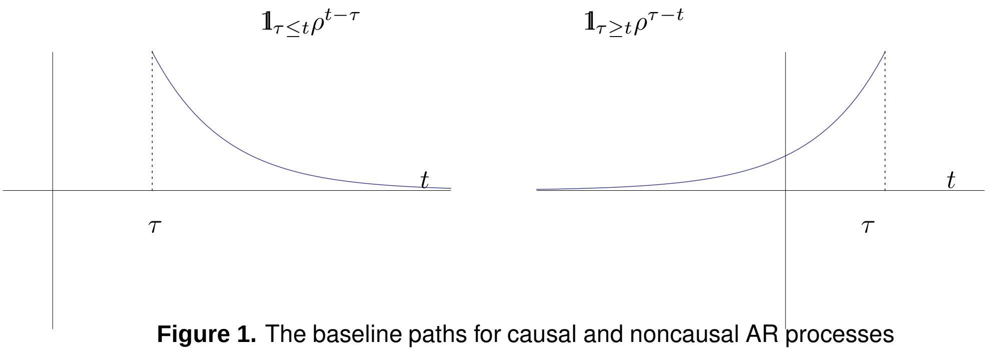

Figure 1. The baseline paths for causal and noncausal AR processes

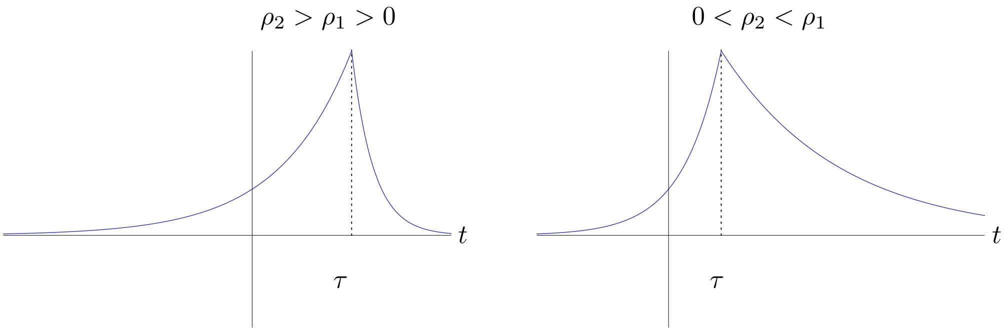

Figure 2. The baseline paths $1_{\tau\le t}\rho_1^{t-\tau} + 1_{\tau>t}\rho_2^{\tau-t}$ for mixed AR(2) processes

The path of process (Y_t) is now a combination of the baseline paths $Z_\tau(t) = \mathbb{1}_{\tau\ge t}a_{\tau-t}$, with stochastic i.i.d. coefficients ε_τ. Figure 1, right panel, illustrates this baseline path for a noncausal AR(1) process, $Y_t = \rho Y_{t+1} + \varepsilon_t$ with $|\rho| < 1$ (thus $a_{-h} = \rho^h$ for $h \ge 0$). The baseline path features an explosive growth followed by a vertical downturn at $t = \tau$.

iii) Let us finally consider a mixed (causal and noncausal) process. The path of (Y_t) is a combination of the baseline paths $Z_\tau(t) = \mathbb{1}_{\tau<t}a_{t-\tau} + \mathbb{1}_{\tau\ge t}a_{\tau-t}$, with stochastic i.i.d. coefficients ε_τ. For instance, if the model is the mixed AR(2): $(1-\rho_1 B)(1-\rho_2 F)Y_t = \varepsilon_t, \quad |\rho_1| < 1, |\rho_2| < 1$, where B and F are the backward and forward operators, respectively, we get $Z_\tau(t) = (1-\rho_1\rho_2)^{-1}\left(\mathbb{1}_{\tau<t}\rho_1^{t-\tau} + 1 + \mathbb{1}_{\tau>t}\rho_2^{\tau-t}\right)$. The baseline path features an explosive growth followed by an exponential decrease (see Figure 2).

The noncausal MA(∞) representation (2.3) helps understanding the formation of bubbles in the dynamics of noncausal processes. First note that the presence of potentially fat tailed error distributions is likely to produce extreme values of any sign over a finite time period. Now suppose that a very large, say positive, value ε_τ occurs at time τ. According to (2.3) if, for simplicity, the sequence (a_h) is strictly decreasing, for $t \leq \tau$ the weight of that extreme value increases as t approaches τ. This explains the growth phase of the bubble. At $t = \tau + 1$, the extreme value cancels out from the sum and the bubble bursts.

There are two types of asymmetries in the shape of a bubble. Longitudinal asymmetries arise in calendar time when the growth and downturn periods have different lengths, as illustrated in Figure 2. Transversal asymmetries arise when the curvature (resp. the magnitude) at a peak and at a trough are different due to the coefficients of the MA(∞) representation (resp. to the asymmetric tails of the error distribution).

3. The noncausal stable linear AR(1) process

In this section, we consider noncausal AR(1) processes with stable distributed errors. Let $X \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$ denote a variable following an α-stable distribution, where $\alpha \in (0, 2]$ is the index of stability, $\beta \in [-1, 1]$ is an asymmetry parameter, $\sigma \in (0, \infty)$ is a scale parameter, and $\mu \in \mathbb{R}$ is a location parameter.

In general, the probability density function (pdf) of a stable distribution is not known explicitly, but its characteristic function $\psi(s) = E(e^{isX})$ has the closed form:

$$\log \psi(s) = -\sigma^\alpha |s|^\alpha \left\{ 1 - i\beta \, (\text{sign } s) \tan\left(\frac{\pi\alpha}{2}\right) \right\} + i\mu s,$$

if $\alpha \neq 1$, and

$$\log \psi(s) = -\sigma |s| \left\{ 1 + i\beta \, (\text{sign } s) \, \frac{2}{\pi} \log |s| \right\} + i\mu s,$$

if $\alpha = 1$, where sign(x) denotes the sign of a real number x. The stable distribution with $\beta = 0$ and exponent $\alpha = 2$ (resp. $\alpha = 1$) is the Gaussian distribution $\text{N}(\mu, \sigma^2)$ (resp. the Cauchy distribution $\mathcal{C}(\mu, \sigma)$ whose pdf is $\frac{\sigma}{\pi\{(x-\mu)^2+\sigma^2\}}$) . The coefficient α determines the tails of the distribution of $X \sim S(\alpha, \beta, \sigma, \mu)$ in the sense that, when $\alpha < 2$,

$$E|X|^p < \infty \quad \text{if and only if} \quad p < \alpha. \tag{3.1}$$

See Samorodnitsky and Taqqu (1994) for further details on stable variables.

3.1. The process

Let us consider the forward looking AR process:

$$Y_t = \rho Y_{t+1} + \varepsilon_t, \qquad |\rho| < 1, \qquad \varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, 0), \tag{3.2}$$

with i.i.d. backward "innovations" ε_t.[3] In view of (3.1) and condition (2.2), the strictly stationary solution of equation (3.2) is given by:

$$Y_t = \sum_{h=0}^{\infty} \rho^h \varepsilon_{t+h}. \tag{3.3}$$

The stationary distribution of (Y_t) is provided in the next proposition.

PROPOSITION 3.1. *The noncausal stable linear AR(1) process (3.2) has a stable stationary distribution given by:*

$$\begin{aligned}
Y_t &\sim \mathcal{S}\left(\alpha, \beta, \frac{\sigma}{(1-|\rho|^\alpha)^{1/\alpha}}, 0\right), && \text{if} \quad \alpha \neq 1, \rho \geq 0,\\
Y_t &\sim \mathcal{S}\left(\alpha, \beta\frac{1-|\rho|^\alpha}{1+|\rho|^\alpha}, \frac{\sigma}{(1-|\rho|^\alpha)^{1/\alpha}}, 0\right), && \text{if} \quad \alpha \neq 1, \rho \leq 0,\\
Y_t &\sim \mathcal{S}\left(1, \beta\frac{1-|\rho|}{1-\rho}, \frac{\sigma}{1-|\rho|}, -\beta\sigma\frac{2}{\pi}\frac{\rho\log|\rho|}{(1-\rho)^2}\right), && \text{if} \quad \alpha = 1.
\end{aligned}$$

In particular, the stationary distribution of the noncausal Cauchy linear AR(1) process ($\alpha = 1, \beta = 0$) is the Cauchy distribution $\mathcal{C}\left(0, \frac{\sigma}{1-|\rho|}\right)$. When $\rho \geq 0$, the asymmetry parameter of Y_t is that of the innovation ϵ_t; when $\rho < 0$, the sign of the asymmetry is unchanged, but the asymmetry coefficient is smaller. Finally, when $\alpha = 1$ and $\beta \neq 0$, a location parameter appears in the distribution of Y_t.

Now, let us consider the process in direct time. While for $\alpha < 2$ the backward stable transition pdf $f(Y_t|Y_{t+1})$ features fat tails, so that the p-th conditional moments do not exist for $p \geq \alpha$, the next proposition shows that the forward transition pdf at any horizon admits Pareto tails with tail parameter equal to $2\alpha + 1$, whenever $\rho \neq 0$. In the causal AR(1) framework, similar results were obtained by Cambanis and Fakhre-Zakeri (Theorem 3, 1994) for the one-step predictor ($h = 1$) with time reversed.

PROPOSITION 3.2. *The noncausal stable linear AR(1) process (3.2) is an homogeneous Markov process. Let $\alpha < 2$. For any $h \geq 0$, if $|\beta| \neq 1$ and $\rho \neq 0$, or if $|\beta| = 1$ and $\rho^{h+1} < 0$, we have*

$$E(|Y_{t+h}|^p \mid Y_{t-1}) < \infty, \quad a.s., \qquad \text{if and only if} \quad -1 < p < 2\alpha + 1. \tag{3.4}$$

[3]Strictly speaking, innovations are not defined when $\alpha \leq 1$ because $E(Y_t \mid Y_{t+1})$ does not exist but, when $\beta = 0$, replacing expectation by median we have: $\varepsilon_t = Y_t - \text{med}(Y_t \mid Y_{t+1})$.

If $|\beta| = 1$ *and* $\rho^{h+1} > 0$*, then* $E(|Y_{t+h}|^p \mid Y_{t-1}) < \infty, \quad$ *a.s. for all* $p > -1$*.*

In particular, the forward conditional expectation $E(Y_{t+h} \mid Y_{t-1})$ always exists, whereas the unconditional and backward conditional expectations exist only when $\alpha > 1$. The next result gives a closed-form expression of the conditional expectation for symmetric stable distributions.

PROPOSITION 3.3. *For* $\rho \neq 0$ *and* $\beta = 0$*, we have:*

$$E(Y_{t+h} \mid Y_{t-1}) \quad = \quad \text{sign}(\rho)|\rho|^{(h+1)(\alpha-1)} Y_{t-1}, \quad \forall h \geq 0.$$

In particular, for Cauchy processes ($\alpha = 1$*), when* $\rho \neq 0$ *we have:*

$$E(Y_{t+h} \mid Y_{t-1}) \quad = \quad \text{sign}(\rho) Y_{t-1}, \quad \forall h \geq 0.$$

This result is unexpected. Indeed, in the L^2 framework, if (X_t) is a stationary strong noncausal AR(1): $X_t = \rho X_{t+1} + W_t, \quad |\rho| < 1, \quad W_t \overset{iid}{\sim} (0, \sigma^2)$, then (X_t) can be expressed as a *weak causal* AR(1) with the same AR coefficient: $X_t = \rho X_{t-1} + W_t^*$, where (W_t^*) is a weak white noise with variance σ^2 (see e.g. Brockwell and Davis (1991), Proposition 4.4.2). It follows that if $E(X_t \mid X_{t-1})$ is linear in X_{t-1}, we must have $E(X_t \mid X_{t-1}) = \rho X_{t-1}$. In contrast, Proposition 3.3 shows that the first-order prediction of Y_t is $E(Y_t \mid Y_{t-1}) = |\rho|^{\alpha-1} Y_{t-1} \neq \rho Y_{t-1}$. The Cauchy process, for $\alpha = 1$, is even more intriguing as it has a unit root when $\rho > 0$, although the process is stationary.

REMARK 3.1 (**Stable AR(1) process with drift).** The introduction of a location parameter in the stable distribution of the innovations (ϵ_t^*) is tantamount to adding an intercept to Model (3.2). The noncausal stable linear AR process with drift μ writes

$$Z_t = \mu + \rho^* Z_{t+1} + \varepsilon_t^*, \qquad |\rho^*| < 1. \tag{3.5}$$

Studying the solutions to this model is straightforward, noting that there is a one-to-one relation between the solutions (Y_t) of Model (3.2) and (Z_t) of model (3.5) via the relation $Y_t = Z_t - \frac{\mu}{1-\rho^*}$. It thus follows from Proposition 3.5 that, for $\rho^* \neq 0$, we have, for $h \geq 0$:

$$E(Z_{t+h} \mid Z_{t-1}) = |\rho^*|^{(h+1)(\alpha-1)} Z_{t-1} + \frac{\mu}{1-\rho^*} \left(1 - |\rho^*|^{(h+1)(\alpha-1)}\right).$$

Interestingly, the adjunction of a constant does not remove the martingale property in the Cauchy case ($\alpha = 1$ and $\rho^* > 0$).

In the next section, we focus on the model with Cauchy and Lévy errors.

3.2. Causal predictive distributions in the Cauchy and Lévy cases

The predictive distributions of the process (Y_t) (and the causal first and second-order moments when they exist) generally do not admit closed form expressions. Two exceptions are the noncausal AR(1) processes with Cauchy and Lévy backward innovations, obtained for $(\alpha, \beta) = (1, 0)$ and $(\alpha, \beta) = (1/2, 1)$, respectively.

3.2.1. *Noncausal Cauchy linear AR(1) process*

The following result gives the Markov transition of the process (Y_t) in direct time.

PROPOSITION 3.4. *The causal transition pdf of the noncausal Cauchy linear AR(1) process at horizon h is given by:*

$$f_h(Y_t|Y_{t-h}) = \frac{1}{\pi\sigma_h} \frac{1}{1 + (Y_{t-h} - \rho^h Y_t)^2/\sigma_h^2} \frac{\sigma^2 + (1-|\rho|)^2 Y_{t-h}^2}{\sigma^2 + (1-|\rho|)^2 Y_t^2}, \qquad \textit{where } \sigma_h = \sigma \frac{1-|\rho|^h}{1-|\rho|}.$$

Therefore, the Pareto tail index of the predictive density is equal to 4, at any prediction horizon. This is only in the limiting case $h = \infty$ that we observe a discontinuity in the value of the tail index, that is, for the stationary distribution. Straightforward calculation shows that the conditional density $f_h(\cdot \mid Y_{t-h})$ is unimodal, for any value of Y_{t-h}.

Proposition 3.4 allows to obtain the causal conditional cdf of the noncausal Cauchy process (see Appendix B). The process also admits a strong causal nonlinear autoregressive representation which is derived in Appendix C.

The first and second-order causal conditional moments of the process exist by Proposition 3.4. The conditional mean was derived in Proposition 3.3. We now give a closed-form expression of the conditional second-order moment for the noncausal Cauchy process.

PROPOSITION 3.5. *For $\rho \neq 0$, when $\alpha = 1$ and $\beta = 0$ in Model (3.2), we have:*

$$E(Y_t \mid Y_{t-1}) = \text{sign}(\rho) Y_{t-1}, \qquad E(Y_t^2 \mid Y_{t-1}) = \frac{1}{|\rho|} Y_{t-1}^2 + \frac{\sigma^2}{|\rho|(1-|\rho|)}.$$

Despite the nonlinear causal autoregression, both processes (Y_t) and (Y_t^2) admit a semi-strong linear causal representation, that is, with linear causal innovations $Y_t - E(Y_t \mid Y_{t-1})$ and $Y_t^2 - E(Y_t^2 \mid Y_{t-1})$, respectively, that are martingale difference sequences, but not i.i.d.

In fact, (Y_t) has a semi-strong AR(1) representation of the form:

$$Y_t = \text{sign}(\rho) Y_{t-1} + \sigma_t \eta_t, \qquad \sigma_t^2 = \left(\frac{1}{|\rho|} - 1\right) Y_{t-1}^2 + \frac{\sigma^2}{|\rho|(1-|\rho|)}, \tag{3.6}$$

where $E(\eta_t \mid Y_{t-1}) = 0, \quad E(\eta_t^2 \mid Y_{t-1}) = 1$. When $\rho > 0$, the process (Y_t) is a conditionally heteroskedastic non integrable martingale sequence. Interestingly, the errors display

GARCH-type effects, which increase when ρ approaches 0, but vanish when $|\rho|$ increases to 1. In view of (3.6), (Y_t) might be called a double semi-strong AR(1) process (see Ling and Li (2008)). However, the innovations η_t are not independent. Indeed, by Proposition 3.4 for $h = 1$, the conditional density of η_t depends on the past: with Y_t given by (3.6),

$$f_\eta(\eta_t|Y_{t-1}) = \frac{1}{\sqrt{|\rho|(1-|\rho|)}\pi} \frac{\sigma\{\sigma^2 + (1-|\rho|)^2 Y_{t-1}^2\}^{3/2}}{\{\sigma^2 + (Y_{t-1} - \rho Y_t)^2\}\{\sigma^2 + (1-|\rho|)^2 Y_t^2\}}.$$

The recursive equations (3.6) associated with the first and second-order conditional causal moments seem explosive: the AR coefficient on the square Y_t^2 is $1/|\rho| > 1$ and the coefficient of Y_{t-1} in the expression of $E(Y_t \mid Y_{t-1})$ leads to a unit root phenomenon, if $\rho > 0$, and to a periodicity with period 2, if $\rho < 0$. This is not surprising since the unconditional first and second-order moments do not exist. However, the infinite moments of the unconditional distribution do not have the same implications on the process in reverse time. Indeed, in reverse time the process does not explose, since $|\rho| < 1$, and infinite moments are just a consequence of the fat tail of the backward innovations. Similar formulas can be derived at any prediction horizon $h > 0$ with ρ replaced by ρ^h and σ by σ_h, respectively.

3.2.2. *Noncausal Lévy Autoregressive Process*

Let us now consider Model (3.2) under the assumption that $\varepsilon_t \sim \mathcal{S}(1/2, 1, 1, 0)$, that is when ε_t follows a Lévy (μ, c) distribution with parameters $\mu = 0, c = 1$. This means that ε_t only takes positive values, which is appropriate for modelling (positive) prices or nominal interest rates. More precisely, the density of ε_t is given by:

$$f_\varepsilon(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x^{3/2}} \exp\left(\frac{-1}{2x}\right) \mathbb{I}_{x>0}. \tag{3.7}$$

Let us assume $\rho > 0$ to ensure the positivity of Y_t. By Proposition 3.1, the stationary distribution of Y_t is also a Lévy distribution, with parameters $\mu = 0, c = 1/(1-\sqrt{\rho})^2$. Thus, the stationary density of the noncausal Lévy linear AR process is:

$$f(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{1-\sqrt{\rho}} \frac{1}{y^{3/2}} \exp\left(\frac{-1}{2y} \frac{1}{(1-\sqrt{\rho})^2}\right) \mathbb{I}_{y>0}. \tag{3.8}$$

PROPOSITION 3.6. *The causal transition pdf of the noncausal Lévy linear AR process, with $\rho > 0$, is given by:*

$$f(Y_t|Y_{t-1}) \quad = \quad \frac{1}{\sqrt{2\pi}} \left(\frac{Y_{t-1}}{Y_t(Y_{t-1} - \rho Y_t)}\right)^{3/2} \exp\left(\frac{-(Y_{t-1} - \sqrt{\rho} Y_t)}{2Y_{t-1}Y_t(1-\sqrt{\rho})^2(Y_{t-1} - \rho Y_t)}\right) \mathbb{I}_{0<\rho Y_t<Y_{t-1}}.$$

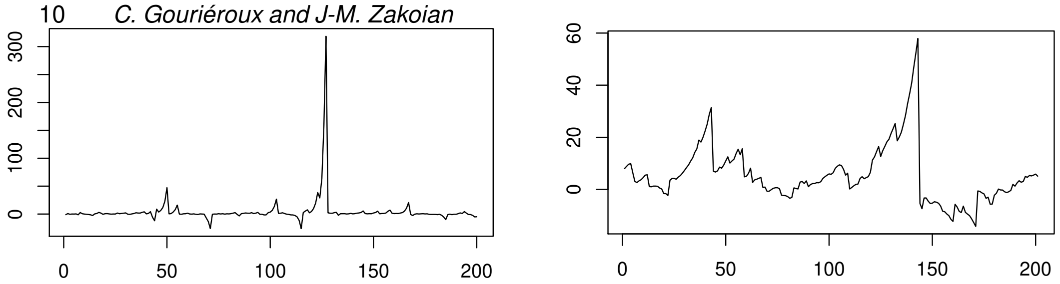

Figure 3. Simulations of Model (3.2) with $\alpha = 1, \beta = 0$, $\sigma = 0.5$ and $\rho = 0.5$ (left panel), $\rho = 0.9$ (right panel).

The support of the causal transition pdf is bounded from above. For this reason, the maximum rate of explosion of a bubble is $1/\rho$. It is also worth noting that the moments of Y_t conditional on Y_{t-1} exist at any order (which also follows from Proposition 3.2 in the case $\beta = 1$ and $\rho > 0$), whereas the unconditional expectation of Y_t does not exist.

By Proposition 3.3, the process has a semi-strong AR(1) causal representation: $Y_t = \frac{1}{\sqrt{\rho}} Y_{t-1} + v_t$, where $E(v_t \mid Y_{t-1}) = 0$. The bubble phenomenon can be more prominent than with Cauchy distributed errors, especially when ρ is close to zero.

3.3. Locally explosive phenomena

In Section 2 we have shown that the path of a noncausal linear AR process is a combination of right-censored increasing exponential curves (if $\rho > 0$) with i.i.d. stochastic coefficients. When the error distribution has fat tails, an extreme value from the right tail creates a jump of Y_t preceded by a local upward trend toward that extreme value. Conversely, an error from the left tail will create a symmetric pattern, able to represent a deflationary spiral. Depending on the purpose, the noncausal model can be adapted to only one type of locally explosive phenomenon, bubble or deflationary spiral, by choosing an error distribution on $\mathbb{R}^+$ with a fat right tail, or on $\mathbb{R}^-$ with a fat left tail.

We provide in Figure 3 the simulated paths of the noncausal Cauchy process based on the forward autoregression (3.2), where the ε_t are drawn independently in the Cauchy distribution (with $\sigma = 1$) and $\rho = 0.5$ (left panel) or $\rho = 0.9$ (right panel). For $\rho = 0.9$, we clearly observe the bubble phenomenon at regular intervals. Note that the bubble collapse can be sudden, i.e. a crash within single period, or it can take place gradually.

The heavy-tailed noncausal AR models can represent multiple local explosions and offer the possibility of predicting the times and sizes of the explosions. Bubbles are created when the noncausal innovation takes an extreme value, and the date at which that extreme value is reached is the date of bubble collapse. Therefore, it is possible to predict the date of a bubble collapse by considering the extreme behaviour of noncausal innovations ε_t.[4] At time t, we can compute the probability of a bubble collapse at $t+h$ as $P(Y_{t+h} - \rho Y_{t+h+1} > c \mid Y_t)$, for some extreme critical level c. By updating these probabilities and following their evolution, we obtain a tool for early warning of the bubble collapse. Thus, contrary to a belief common to economists (see e.g. Cooper (2008)), it seems possible to detect a bubble in its inflationary phase, and to predict the future bubbles (see Phillips et al. (2015) for a similar remark).

4. Aggregation of noncausal processes

4.1. Aggregation of a continuum of noncausal Cauchy AR(1)

Noncausal AR(1) processes with stable distributions can generate a series of local explosions, but only with the same (stochastic) rate of increase determined by the coefficient ρ. Aggregation of such basic processes allows us to get bubbles with different rates of increase. For simplicity, we focus on basic Cauchy AR(1) processes in this section.

Let us first consider an aggregate process from a finite set of noncausal Cauchy AR(1) processes. The process is defined by

$$\mathcal{Y}_t = c \sum_{j=1}^{J} \pi_j Y_{j,t}, \quad \text{with} \quad Y_{j,t} = \rho_j Y_{j,t+1} + \varepsilon_{j,t}, \qquad |\rho_j| < 1, \quad j = 1, \ldots, J, \tag{4.1}$$

where $(\varepsilon_{j,t})_{j=1,\ldots,J}$ is a family of i.i.d. standard Cauchy white noises, $c > 0$, $\pi_j \geq 0$ and $\sum_{j=1}^{J} \pi_j = 1$. By construction, this process can generate successive bubbles, with rates of increases $1/\rho_j$ (if $\rho_j > 0$) and occurrences governed by the weights π_j. However, the aggregation destroys the Markov property of the noncausal aggregate process in reverse time.

The aggregation can be extended to a continuum of values of parameter ρ, with a probability distribution π on $(-1, 1)$ (which can be either continuous or discrete). Let

$$\mathcal{Y}_t := \mathcal{Y}_t(c, \pi) = c \int_{(-1,1)} Y_{\rho,t} d\pi(\rho), \tag{4.2}$$

where $Y_{\rho,t} = \rho Y_{\rho,t+1} + \varepsilon_{\rho,t}, \quad |\rho| < 1, c > 0$, and $(\varepsilon_{\rho,t})$ are i.i.d. standard Cauchy white noises. Let us denote $E_\pi(f) = \int f(\rho) d\pi(\rho)$ for any function $f : (-1, 1) \mapsto \mathbb{R}$ such that the

[4]The predictive distributions at any horizon can be estimated and used to predict the dates and magnitudes of downturns at given horizons [see Gouriéroux and Jasiak (2015)].

integral exists. The following proposition gives a sufficient condition for the existence of process $\mathcal{Y}_t$.

Proposition 4.1. *Assume that, for some $s \in (0,1)$,*

$$\sum_{i=0}^{\infty} \left\{E_\pi(|\rho|^i|)\right\}^s < \infty. \tag{4.3}$$

Then, $\mathcal{Y}_t$ is well defined and we have:

$$\mathcal{Y}_t = \sum_{i\geq 0} \int \rho^i \varepsilon_{\rho,t+i} d\pi(\rho). \tag{4.4}$$

Moreover, the process $(\mathcal{Y}_t)_{t\geq 0}$ is strictly stationary and ergodic, and $\mathcal{Y}_t$ follows the Cauchy distribution with scale parameter $cE_\pi\left\{\frac{1}{1-|\rho|}\right\}$.

Example 4.1 (**Discrete measure).** Suppose that the support of probability measure π is $\{\rho_j\}_{j\in\mathbb{N}}$, where the ρ_j's belong to $(-1,1)$ and are ranked in increasing order. Let $\pi_j = \pi(\{\rho_j\})$ (with $\pi_j > 0$ and $\sum_{j\geq 0}\pi_j = 1$). Then, the condition in (4.3) becomes

$$\sum_{i=0}^{\infty}\left\{\sum_{j\geq 0}|\rho_j|^i\pi_j\right\}^s < \infty. \tag{4.5}$$

Using the elementary inequality $(x+y)^s \leq x^s + y^s$, for $x,y\geq 0$ and $s\in(0,1)$, we get a sufficient condition for strict stationarity: $\sum_{j\geq 0}\frac{\pi_j^s}{1-|\rho_j|^s} < \infty$, and a necessary condition for the existence of the sum in (4.5) is: $E_\pi\left(\frac{1}{1-|\rho|}\right) = \sum_{j\geq 0}\frac{\pi_j}{1-|\rho_j|} < \infty$. For instance if $\pi_j = \tilde{\pi}^j$ for $j>0$ and $|\rho_j| = 1-\tilde{\rho}^j$, for some $\tilde{\pi},\tilde{\rho}\in(0,1)$, a necessary condition for strict stationarity of $(\mathcal{Y}_t)$ is $\tilde{\pi} < \tilde{\rho}$.

The distribution of the strictly stationary aggregated process is characterized by its characteristic function $\Psi(u_0,\dots,u_k) = E[\exp\{i(\sum_{\ell=0}^k u_\ell\mathcal{Y}_{t+\ell})\}]$, for $(u_0,\dots,u_k)\in\mathbb{R}^{k+1}$.

Proposition 4.2. *Assume that (4.3) holds and, for any $t\in\mathbb{Z}$, assume that the process $\{\varepsilon_{\rho,t},(\rho,t)\in(-1,1)\times\mathbb{Z}^+\}$ is independent. We have, for $(u_0,\dots,u_k)\in\mathbb{R}^{k+1}$,*

$$\Psi(u_0,\dots,u_k) = \exp\left\{-cE_\pi\left(\sum_{h=0}^{k-1}\left|\sum_{\ell=0}^{h}\rho^{h-\ell}u_\ell\right| + \frac{\left|\sum_{\ell=0}^k\rho^{k-\ell}u_\ell\right|}{1-|\rho|}\right)\right\}.$$

Thus, the linear combination $\sum_{\ell=0}^k u_\ell\mathcal{Y}_{t+\ell}$ follows a Cauchy distribution, with scale parameter equal to the expectation in the brackets. The joint distribution of $(\mathcal{Y}_t,\mathcal{Y}_{t+k})$ is characterized below.

Corollary 4.1. *Under the assumptions of Proposition 4.2, we have*

$$E[\exp\{i(u_0\mathcal{Y}_t + u_k\mathcal{Y}_{t+k})\}] = \exp\left[-cE_\pi\left\{\frac{1}{1-|\rho|}\left(|u_0|(1-|\rho|^k) + |\rho^k u_0 + u_k|\right)\right\}\right].$$

4.2. Identification and estimation of the explosive patterns

Let us now show that the parameters characterizing the local explosive patterns are identifiable from the bivariate distribution of $(\mathcal{Y}_t, \mathcal{Y}_{t+1})$.

4.2.1. Identification

From Corollary 4.1, it follows that

$$\tilde{\Psi}(u) \quad := \quad \log E[\exp\{i(\mathcal{Y}_t + u\mathcal{Y}_{t+1})\}] = -cE_\pi \left(1 + \frac{|\rho + u|}{1 - |\rho|}\right). \tag{4.6}$$

The proposition below shows how to derive the local explosion parameters c and π from the joint characteristic function $\tilde{\Psi}$. Let $\tilde{\Psi}^{(2)}(u) = \frac{\partial^2 \tilde{\Psi}(u)}{\partial u^2}$.

PROPOSITION 4.3. *Assume that the measure π admits a density, $d\pi(\rho) = \pi(\rho)d\rho$. Then, under (4.3), the parameters c, π are identifiable from the joint distribution of $(\mathcal{Y}_t, \mathcal{Y}_{t+1})$. We have, for $\rho \in (-1, 1)$,*

$$\pi(\rho) = \frac{(1 - |\rho|)\tilde{\Psi}^{(2)}(-\rho)}{\int_{-1}^{1}(1 - |u|)\tilde{\Psi}^{(2)}(-u)du}, \qquad c = -\frac{1}{2}\int_{-1}^{1}(1 - |u|)\tilde{\Psi}^{(2)}(-u)du.$$

4.2.2. Estimation of the local explosion structure

The relations in Proposition 4.3 can be used to estimate the density π and the constant c from observations $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$. First note that, because the distribution of $(\mathcal{Y}_t, \mathcal{Y}_{t+1})$ is symmetric, we have $\tilde{\Psi}(u) = \log E[\cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})]$. It follows that

$$\tilde{\Psi}^{(2)}(u) = -\frac{E[\mathcal{Y}_{t+1}^2 \cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})]}{E[\cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})]} - \left(\frac{E[\mathcal{Y}_{t+1} \sin(\mathcal{Y}_t + u\mathcal{Y}_{t+1})]}{E[\cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})]}\right)^2.$$

Note that: (i) the expectations appearing in the latter formula exist, by arguments given in the proof of Proposition 3.3, and using the fact that the distribution of $\mathcal{Y}_t$ is a Cauchy; (ii) $E[\cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})] = e^{\tilde{\Psi}(u)} > 0$. A consistent estimator of $\tilde{\Psi}^{(2)}(u)$ is thus, by applying the ergodic theorem to the process $(\mathcal{Y}_t)$ (see Proposition 4.1),

$$\hat{\tilde{\Psi}}^{(2)}(u) = -\frac{\sum_{t=1}^{n} \mathcal{Y}_{t+1}^2 \cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})}{\sum_{t=1}^{n} \cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})} - \left(\frac{\sum_{t=1}^{n} \mathcal{Y}_{t+1} \sin(\mathcal{Y}_t + u\mathcal{Y}_{t+1})}{\sum_{t=1}^{n} \cos(\mathcal{Y}_t + u\mathcal{Y}_{t+1})}\right)^2,$$

with by convention $\mathcal{Y}_{n+1} = 0$.

The parameters c, π can next be estimated, for $\rho \in (-1, 1)$, as follows:

$$\hat{\pi}(\rho) = \frac{(1 - |\rho|)\hat{\tilde{\Psi}}^{(2)}(-\rho)}{\int_{-1}^{1}(1 - |u|)\hat{\tilde{\Psi}}^{(2)}(-u)du}, \qquad \hat{c} = -\frac{1}{2}\int_{-1}^{1}(1 - |u|)\hat{\tilde{\Psi}}^{(2)}(-u)du.$$

The above estimators are derived under the assumption of a continuous distribution π. These results are modified when π is discrete. Let us, for example, consider the case where the support of π is a countable set, as in Example 4.1. By (4.6) we have

$$\tilde{\Psi}(u) = -c\left(1+\sum_{j\in\mathbb{N}}\pi_j\frac{|\rho_j+u|}{1-|\rho_j|}\right).$$

The first and second right-derivatives of $\tilde{\Psi}$ are

$$\frac{\partial\tilde{\Psi}(u)}{\partial u^+} = -c\sum_{j\in\mathbb{N}}\frac{\pi_j}{1-|\rho_j|}\{\mathbb{I}_{u\geq-\rho_j}-\mathbb{I}_{u<-\rho_j}\}, \quad \frac{\partial^2\tilde{\Psi}(u)}{\partial(u^+)^2}=0.$$

The second-order right derivative is no longer informative about parameters c and π. However, these parameters can be identified by considering the location and size of the jumps in the first-order right derivative.

4.2.3. Specification tests

Instead of considering the joint distribution of $(\mathcal{Y}_t, \mathcal{Y}_{t+1})$, let us now consider the joint distribution of $(\mathcal{Y}_t, \mathcal{Y}_{t+k})$. By Corollary 4.1 we have,

$$\tilde{\Psi}_k(u) := \log E[\exp\{i(\mathcal{Y}_t+u\mathcal{Y}_{t+k})\}] = -cE_\pi\left\{\frac{1-|\rho|^k+|\rho^k+u|}{1-|\rho|}\right\}.$$

Let us also assume in this section that the support of the measure π is $[0,1)$. We thus have $\tilde{\Psi}_k(u) = -cE_\pi\left\{\frac{1}{1-\rho}(1-\rho^k+|\rho^k+u|)\right\}$. We obtain alternative estimators of the bubble parameters, for $\rho\in(0,1)$,

$$\hat{\pi}_k(\rho) = \frac{(1-\rho)\rho^{k-1}\hat{\tilde{\Psi}}_k^{(2)}(-\rho^k)}{\int_{-1}^1(1-u)u^{k-1}\hat{\tilde{\Psi}}_k^{(2)}(-u^k)du}, \quad \hat{c}_k = -\frac{1}{2}\int_{-1}^1(1-u)u^{k-1}\hat{\tilde{\Psi}}_k^{(2)}(-u^k)du,$$

where $\hat{\tilde{\Psi}}_k^{(2)}$ is the sample counterpart of the second derivative of $\tilde{\Psi}_k$. We get a sequence of estimators of π indexed by k, which can be used to construct specification tests, since all the estimators converge to the same function if the underlying process is an aggregated noncausal Cauchy AR(1) process.

4.3. Locally explosive model with a Gaussian AR(1) component

The identification and estimation methods can be extended for applications to models that include a Gaussian AR(1) component. Let us now assume that the observations are generated by the following model

$$Z_t = Y_t + \mathcal{Y}_t, \tag{4.7}$$

where $(\mathcal{Y}_t)$ is the aggregated Cauchy AR(1) process defined in (4.2), and (Y_t) is the Gaussian AR(1) process

$$Y_t \quad = \quad rY_{t+1} + \sigma\eta_t, \quad (\eta_t) \overset{iid}{\sim} \mathcal{N}(0,1), \quad r \neq 1,$$

where (η_t) is i.i.d. and is independent from the noises $(\varepsilon_{\rho,t})$. Note that it is equivalent to write the Gaussian process in either the causal or noncausal forms which are distributionally equivalent. The characteristic function of (Z_t, Z_{t+1}) is, for $u_0, \lambda \in \mathbb{R}$,

$$\begin{aligned}\Psi(u_0, \lambda u_0) \quad &:= \quad E[\exp\{iu_0(Z_t + \lambda Z_{t+1})\}] \\ &= \quad \exp\left\{-\frac{u_0^2\sigma^2(1+2\lambda r+\lambda^2)}{2(1-r^2)} - c|u_0|E_\pi\left(1+\frac{|\rho+\lambda|}{1-|\rho|}\right)\right\}. \qquad (4.8)\end{aligned}$$

The argument u_0, that is the difference of tail magnitudes of the Gaussian and Cauchy distributions, can be used to disentangle both components. We have

$$\lim_{u_0\to 0} \frac{\log \Psi(u_0, \lambda u_0)}{|u_0|} \quad := \quad -cE_\pi\left(1+\frac{|\rho+\lambda|}{1-|\rho|}\right), \qquad (4.9)$$

whereas

$$\lim_{u_0\to \infty} \frac{\log \Psi(u_0, \lambda u_0)}{u_0^2} \quad := \quad -\frac{\sigma^2(1+2\lambda r+\lambda^2)}{2(1-r^2)}. \qquad (4.10)$$

We deduce the following proposition.

PROPOSITION 4.4. *Under the assumptions of Proposition 4.3, all parameters r, σ, c, π are identifiable in the aggregated noncausal Cauchy AR(1) with additional Gaussian AR(1) process defined in (4.7).*

This deconvolution result above is important from the modeling perspective. In practice, it is difficult to disentangle the global and local explosive patterns. Global explosive patterns are generally modeled by means of close to unit root (Gaussian) model, that is an AR model with coefficient r close to 1. In contrast, bubbles consist in local explosive patterns. Proposition 4.4 reveals the possibility of identifying the global and local explosive components.

Finally, the explicit form of the characteristic function (CF) in (4.8) suggests an estimation procedure based on the comparison of the empirical CF (ECF) and the theoretical CF of (Z_t, Z_{t+1}). Suppose for simplicity that the probability distribution π is a point mass at some AR coefficient ρ_0, that is $\mathcal{Y}_t = cY_{\rho_0,t}$. Given observations $(Z_1, \ldots, Z_n)$, the parameter $\boldsymbol{\theta} = (\rho, r, c, \sigma)'$ can be estimated by minimizing a distance between the CF and the ECF defined by

$$\widetilde{\Psi}_n(u, \lambda u) = \frac{1}{n}\sum_{t=1}^{n} \cos\{u(Z_t + \lambda Z_{t+1})\}. \qquad (4.11)$$

The CF estimator is defined by

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} \int_{\mathbb{R}^2} |\widetilde{\Psi}_n(u,\lambda u) - \Psi(u,\lambda u)|^2 dW(u,\lambda), \tag{4.12}$$

where Θ denotes the parameter space and $W(\cdot)$ is a nonnegative weighting function. There is an abundant literature on ECF-based estimation methods which goes back to Paulson et al. (1975) and Heathcote (1977) in the case of i.i.d. data. For dependent data, see for instance Yu (2004), Taufer and Leonenko (2009) and the references therein. We leave the theoretical properties of the estimator in (4.12) for further research.

5. Estimation and diagnostic checking in the noncausal heavy-tailed AR(1) model

We first derive the asymptotic properties of the sample autocorrelation function (ACF) for the noncausal AR(1) model under weak semi-parametric assumptions on the tail behaviour of the errors. Next, we discuss the unit root (UR) hypotheses and tests.

5.1. Estimating the AR coefficient

In this section, we consider estimating the AR coefficient in a non-causal AR(1) process with heavy-tailed errors whose distribution is not fully specified. More specifically, we assume

$$Y_t = \rho Y_{t+1} + \varepsilon_t, \qquad |\rho| < 1, \qquad (\varepsilon_t) \text{ i.i.d.}, \tag{5.1}$$

where, for simplicity, the distribution of ε_t is symmetric and the distribution of $|\varepsilon_t|$ is *regularly varying* with index $-\alpha$, that is,

$$P(|\varepsilon_t| > x) = x^{-\alpha} L(x), \tag{5.2}$$

with $\alpha \in (0,2)$ and $L(x)$ a slowly varying function at infinity. An estimator of the AR coefficient ρ in Model (5.1)-(5.2) is the first sample autocorrelation:

$$\hat{\rho}_n = \sum_{t=2}^{n} Y_t Y_{t-1} / \sum_{t=1}^{n} Y_t^2, \tag{5.3}$$

when the observations are $Y_1, \ldots, Y_n$. More generally, we define, for $\ell \geq 0$,

$$\hat{\rho}_n(\ell) = \sum_{t=\ell+1}^{n} Y_t Y_{t-\ell} / \sum_{t=1}^{n} Y_t^2, \tag{5.4}$$

with $\hat{\rho}_n(1) = \hat{\rho}_n$. Note that the theoretical autocorrelations of the process (Y_t) do not exist. However, we have the following result.

PROPOSITION 5.1 (DAVIS AND RESNICK, 1985). *Let (Y_t) be the strictly stationary solution of model (5.1)-(5.2). Then, the estimators defined in (5.4) are consistent, that is, for $\ell \geq 1$, $\hat{\rho}_n(\ell) \to \rho^\ell$, in probability.*

REMARK 5.1. Despite the fact that, for stable errors, $E(Y_t \mid Y_{t-1}) \neq \rho Y_{t-1}$ by Proposition 3.3, the sample AR coefficient defined in (5.3) converges to ρ. Thus the asymptotic analysis of the empirical ACF reveals serial dependence in reverse time, not in direct time.

The asymptotic distribution of a vector of sample autocorrelations of the noncausal AR process, in terms of stable variables, is the following. Let, for $M \geq 1$, $\hat{\boldsymbol{\rho}}_n = (\hat{\rho}_n(1), \ldots, \hat{\rho}_n(M))'$, $\boldsymbol{\rho} = (\rho, \ldots, \rho^M)'$ and let

$$a_n = \inf\{x : P(|\varepsilon_1| > x) \leq n^{-1}\}, \quad \tilde{a}_n = \inf\{x : P(|\varepsilon_1 \varepsilon_2| > x) \leq n^{-1}\}.$$

PROPOSITION 5.2 (DAVIS AND RESNICK, 1986). *Let (Y_t) be the strictly stationary solution of Model (5.1)-(5.2) with $E|\varepsilon_t|^\alpha = \infty$. Then*

$$\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) \xrightarrow{d} \mathbf{Z} := (Z_1, \ldots, Z_M), \tag{5.5}$$

where, $Z_\ell = \sum_{j=1}^{\infty}\{\rho^{j+\ell} - \rho^{|j-\ell|}\} S_j/S_0$, for $\ell = 1, \ldots, M$, and $S_0, S_1, S_2, \ldots$ are independent stable random variables; S_0 is positive with index $\alpha/2$ and S_j, for $j \geq 1$, has index α.

When $\alpha = 1$ the convergence in (5.5) holds with $a_n^2/\tilde{a}_n = n/\log n$. If the law of $|\varepsilon_t|$ is asymptotically equivalent to a Pareto, (5.5) holds with $a_n^2/\tilde{a}_n = (n/\log n)^\alpha$.

REMARK 5.2. In particular, for the noncausal Cauchy process, the estimator of the AR coefficient defined in (5.3) satisfies, when $|\rho| < 1$, $\frac{n}{\log n}(\hat{\rho}_n - \rho) \xrightarrow{d} (1+\rho) S_1/S_0$. It can be noted that S_1 is standard Cauchy-distributed, while S_0 has a Lévy distribution concentrated on $(0, \infty)$, with density given in (3.7). Therefore,

$$\frac{n}{\log n}(\hat{\rho}_n - \rho) \xrightarrow{d} (1+\rho) Y X, \tag{5.6}$$

where X, Y are independent with $Y \sim \mathcal{C}(0,1)$ and $X \sim \chi^2(1)$. Contrary to the standard situation where the rate of convergence is $\sqrt{n}$ and the limiting distribution is Gaussian, the above asymptotic distribution does not admit a finite expectation and is reached at a faster rate.

REMARK 5.3. The knowledge of index α is not required for the computation of the ordinary least-squares estimator of ρ, but its asymptotic distribution depends on α. After estimating ρ in a first step, the tail index α can be estimated from a standard approach in the second step. For instance, the Hill estimator can be used (for its main properties under various assumptions see Embrechts et al. (1997), Theorem 6.4.6).

Remark 5.4. If the errors distribution is completely known, a far more efficient estimator is the Maximum-Likelihood estimator (MLE). For α-stable causal and noncausal AR processes, the asymptotic distribution of the MLE was established by Andrews et al. (2009) [see also Lanne and Saikkonen (2011)]. With Cauchy errors, for instance, the MLE converges faster to ρ than the first sample autocorrelation (n instead of $n/\log n$), but the limiting distribution has no simple closed form.

5.2. Unit root hypothesis and test in the Cauchy case

The noncausal AR Cauchy process with $\rho > 0$ has the particularity of producing explosive features while being a stationary martingale. The speed of explosion of a bubble is in average strictly larger than 1, which is the unit root, to compensate the collapse behavior. This can be illustrated by plotting Y_t against Y_{t-1} for simulated paths of the noncausal AR(1) process with Cauchy errors (see Figure 4). The theoretical result in Proposition 3.5 concerning the conditional mean is confirmed by the simulated data, the slopes being almost equal to 1 and -1, respectively[5]. However, we also observe plots almost on the line $Y_t = 0$. They correspond to the dates at which the large (multiple) bubbles collapse.

The difference between the rate of explosion of the bubble for a given trajectory and the average rate of explosion of the process, equal to 1 under the martingale hypothesis, explains why the standard UR tests often provide strange results. See e.g. Homm and Breitung (2012) for a survey of UR tests for detecting bubbles, and Evans (1991), Charemza and Deadman (1995) for pitfalls in testing for explosive bubbles.

Standard UR tests, based on a difference between $\hat{\rho}_n$ and 1, will fail in detecting locally explosive features in the noncausal AR Cauchy model, although the martingale property is satisfied when $\rho > 0$ by Proposition 3.5. This is not surprising as the asymptotic properties of standard UR tests are often derived under assumptions which are not satisfied by the noncausal Cauchy AR(1) model, such as the non stationarity under the null hypothesis.

In our framework, a test for bubbles would need to detect successive transitory explosions along the observed trajectory, and not an explosive behavior in average, or a breakpoint in

[5]In fact slightly larger than 1 and slightly smaller than -1, respectively, to compensate the observations at collapse times.

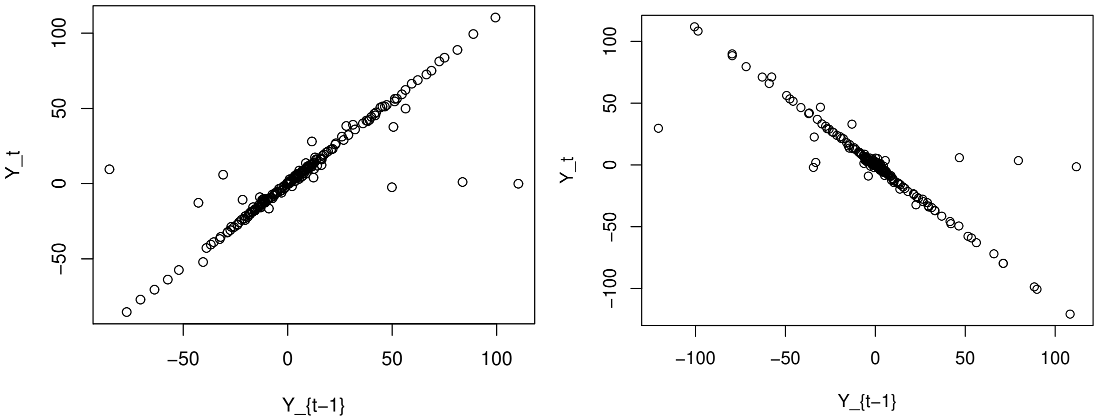

Figure 4. Scatterplot of Y_t against Y_{t-1} for the the Cauchy AR(1) simulation of Figure 3 with $\rho = 0.9$ (left panel) and $\rho = -0.9$ (right panel).

the explosive behavior in average, as in Busetti and Taylor (2004), Phillips et al. (2011) or Phillips et al. (2015) approaches.

5.3. Diagnostic checking

After estimating the coefficient ρ, it is important to verify the independence of the residuals in (5.1). In this respect, diagnostic checks can be based on the first-order residual autocorrelation. Let us define the backward residuals

$$\hat{\varepsilon}_t^* = Y_t - \hat{\rho}_n Y_{t+1}, \quad t = 1, \dots, n-1,$$

in the estimation of Model (5.1)-(5.2). Similarly, we define the forward residuals

$$\hat{\varepsilon}_t = Y_t - \hat{\rho}_n Y_{t-1}, \quad t = 2, \dots, n.$$

To check the white noise property, we consider the (backward and forward) residuals first-order autocorrelations defined by

$$R_n^* = \sum_{t=2}^{n-1} \hat{\varepsilon}_t^* \hat{\varepsilon}_{t-1}^* / \sum_{t=1}^{n-1} (\hat{\varepsilon}_t^*)^2, \quad R_n = \sum_{t=3}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} / \sum_{t=2}^{n} \hat{\varepsilon}_t^2.$$

5.3.1. *Asymptotic behaviour under correct specification*

The asymptotic distributions of statistics R_n and R_n^* under model (5.1)-(5.2), that is under correct specification, are as follows:

PROPOSITION 5.3. *Let (Y_t) be the strictly stationary solution of Model (5.1)-(5.2). Then*

$$\frac{a_n^2}{\tilde{a}_n} R_n^* \xrightarrow{d} \rho^2 S_1/S_0 - \{1-\rho^2\} \sum_{j=2}^{\infty} \rho^{j-1} S_j/S_0, \tag{5.7}$$

where the S_j are independent stable variables as described in Proposition 5.2. If $\alpha \geq 1$ and $|\varepsilon_t|$ is asymptotically equivalent to a Pareto, the statistic R_n has the same asymptotic distribution as R_n^*. In particular, for the noncausal Cauchy AR process we have:

$$\frac{n}{\log n} R_n^* \xrightarrow{d} R \quad and \quad \frac{n}{\log n} R_n \xrightarrow{d} R, \qquad as\ n \to \infty,$$

with $R = \rho(1+2\rho)YX$, where X, Y are independent with $Y \sim \mathcal{C}(0,1)$ and $X \sim \chi^2(1)$.

Similar results could be established for higher-order residual autocorrelations (see Lin and McLeod (2008) for portmanteau tests in the case of causal AR processes with stable errors).

Thus, at least when $\alpha \geq 1$ and the errors distribution is asymptotically equivalent to a Pareto, the empirical autocorrelations of the residuals $\hat{\varepsilon}_t^*$ and $\hat{\varepsilon}_t$ have the same asymptotic behaviour. This is a consequence of the weak causal linear representation:

$$Y_t = \rho Y_{t-1} + u_t, \tag{5.8}$$

where the u_t are "empirically uncorrelated" variables, in the sense that, for any $\ell > 0$,

$$\sum_{t=\ell+1}^{n} u_t u_{t-\ell} / \sum_{t=1}^{n} u_t^2 \to 0, \quad \text{in probability as } n \to \infty. \tag{5.9}$$

In the standard case where errors ε_t admit finite variance, (u_t) is a weak white noise and the same result is true. In our framework, the u_t's do not admit second-order moments. A surprising difference with the classical case is the coexistence of the "empirically" weak linear representation (5.8)-(5.9) and the semi-strong linear representation (3.6). Table 1 summarizes the different representations which can be defined for the noncausal Cauchy AR(1) process.

5.3.2. Asymptotic behaviour of statistics under a (near) random walk

Let us now discuss the behavior of the statistics R_n and R_n^* when the DGP is a strong AR(1) with root at or near unity. More precisely, we consider a time series that is generated by

$$Y_{n,t} = a_n Y_{n,t-1} + \xi_t, \quad t \geq 1, \quad a_n = \exp(c/n), \tag{5.10}$$

for some random initial value Y_0 whose distribution is independent of n and of (ξ_t), where c is a real constant. When $c = 0$, the model has a unit root; when $c < 0$, the model is stable

Name	Equation	Noise properties
Strong noncausal linear	$Y_t = \rho Y_{t+1} + \varepsilon_t$	$(\varepsilon_t/\sigma) \overset{i.i.d.}{\sim} \mathcal{C}(0,1)$
Semi-strong causal linear	$Y_t = \text{sign}(\rho) Y_{t-1} + \sigma(Y_{t-1})\eta_t$	$E(\eta_t \mid Y_{t-1}) = 0$ $E(\eta_t^2 \mid Y_{t-1}) = 1$
Strong causal nonlinear	$Y_t = G(Y_{t-1}, v_t)$	$(v_t) \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
Weak causal linear	$Y_t = \rho Y_{t-1} + u_t$	empirically uncorrelated: $\sum_{t=\ell+1}^{n} u_t u_{t-\ell} \to 0$

Table 1: Representations of the noncausal Cauchy AR(1) with $|\rho| < 1$

for finite n, and it has explosive features for $c > 0$. The errors may be dependent, but are assumed to satisfy the following conditions:

A0: $\xi = (\xi_t)$ is a strictly and second-order stationary process, $E(\xi_t) = 0$ and $E(\xi_t^2) > 0$.

A1: The strong mixing coefficients[6] of the process ξ are such that

$$\sum_{h=0}^{\infty} \{\alpha_\xi(h)\}^{\frac{\nu}{2+\nu}} < \infty \quad \text{for some } \nu > 0, \text{ with } E|\xi_t|^{2+\nu} < \infty.$$

Pham (1986) showed that **A0-A1** hold for a large class of processes, the strong mixing coefficients converging to zero exponentially fast. As far as GARCH are concerned, results on strict stationarity and strong mixing have been derived by numerous authors (see for instance Chapter 3 in Francq and Zakoïan (2010) and the references therein). Note that models (5.1)-(5.2) and (5.10) are non nested. Model (5.10), which was studied by Phillips (1987) under slightly more general conditions than **A0-A1**, might be erroneously estimated as a noncausal Cauchy AR in practice. The following result shows that the statistics R_n and R_n^* computed from the process $(Y_{n,t})$ have an asymptotic behavior which is very different from that obtained in Proposition 5.3.

Proposition 5.4. *For the (near) random walk (5.10) under* **A0-A1***, for any* $c \in \mathbb{R}$*:*

$$\frac{n}{\log n}|R_n^*| \longrightarrow \infty \quad \textit{and} \quad \frac{n}{\log n}|R_n| \longrightarrow \infty, \qquad \textit{in probability as } n \to \infty.$$

Therefore, tests based on R_n or R_n^*, to be defined in Section 7, can distinguish a unit root due to a stationary noncausal AR process from a unit root created by a (near) random walk model.

[6]defined by $\alpha_\xi(h) = \sup_{A \in \sigma(\xi_u, u \le t),\, B \in \sigma(\xi_u, u \ge t+h)} |P(A \cap B) - P(A)P(B)|$.

6. A Monte Carlo study

In this section, we study the behaviours of the estimator $\hat{\rho}_n$ and the statistics for diagnostic checking introduced in Section 5.3. We also compare $\hat{\rho}_n$ with the ML estimator of ρ.

6.1. Behaviour of $\hat{\rho}_n$ in finite samples

We simulated $N = 5,000$ paths of model (3.2) with $\alpha = 1$ and $\beta = 0$, for different values of ρ, ranging from 0.1 to 0.9, and different sample sizes (n=500, 2000, 5000). Table 2 shows characteristics of the empirical distribution of $\frac{n}{\log n}(\hat{\rho}_n - \rho)$ over the N simulated paths. Increasing the sample size does not entail much distorsion of the sample distributions, indicating that the normalization by $\frac{n}{\log n}$ is appropriate for finite sample sizes. The median is always very small, and the first and third quartiles are rather close in modulus, at least for ρ sufficiently far from 1, indicating that $\hat{\rho}_n$ is approximately symmetrically distributed around ρ. The probability of $\hat{\rho}_n$ exceeding 1 is extremely small, even for $\rho = 0.9$, showing that any standard unit-root test would reject the unit-root hypothesis. Comparison with the asymptotic distribution in (5.6), see the case $n = \infty$ in Table 2, leads to mixed conclusions. On the one hand, the center of the finite sample distributions (in the interquartile interval) appears to be well approximated by the asymptotic distribution, at least for n sufficiently large and ρ not too close to 1. On the other hand, for $\rho > 0$, the absolute values of the quantiles of the asymptotic distribution increase with ρ with a proportionality factor of $1 + \rho$. This pattern does not appear in finite sample size for large values of ρ, and this is particularly true for the extreme quantiles $q_{0.1}$ and $q_{0.9}$.

6.2. Statistics for diagnostic checking

Let us now study the statistics of Proposition 5.3. We only present, in Table 3, results for the sample size $n = 5000$. The behaviour of the statistics R_n and R_n^* is very similar, whatever the values of n and ρ. The previous comments concerning the convergence to the asymptotic distribution, that is the distribution of R, still apply. The two statistics R_n and R_n^* having the same asymptotic distribution, they cannot be used to distinguish between a causal and a noncausal AR process. For this purpose, it can be useful to consider the (backward and forward) serial correlation between a residual and a squared lagged

n	ρ	Mean	Std	$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$	$P[\hat{\rho}_n > 1]$
500	0.1	0.104	7.050	-1.547	-0.428	-0.000	0.446	1.563	0.0003
	0.3	0.101	5.324	-1.688	-0.482	0.000	0.488	1.659	0.0008
	0.5	0.058	4.479	-1.811	-0.515	0.001	0.517	1.709	0.0017
	0.7	-0.010	3.780	-1.791	-0.553	0.002	0.512	1.661	0.0033
	0.9	-0.060	2.866	-1.544	-0.512	-0.004	0.450	1.301	0.0123
2000	0.1	0.140	16.08	-1.690	-0.423	0.003	0.429	1.657	0.0000
	0.3	0.108	9.929	-1.838	-0.477	0.000	0.478	1.813	0.0002
	0.5	0.079	8.187	-1.936	-0.524	-0.000	0.513	1.907	0.0006
	0.7	0.021	6.650	-1.954	-0.549	-0.002	0.530	1.878	0.0010
	0.9	-0.066	4.626	-1.707	-0.519	-0.003	0.495	1.578	0.0029
5000	0.1	0.195	35.45	-1.756	-0.422	0.000	0.432	1.717	0.0000
	0.3	0.099	16.571	-1.928	-0.478	0.000	0.479	1.880	0.0001
	0.5	-0.138	11.518	-2.054	-0.525	0.000	0.522	1.983	0.0002
	0.7	0.021	9.255	-2.156	-0.551	0.000	0.544	2.032	0.0003
	0.9	-0.053	6.947	-1.906	-0.541	0.000	0.511	1.808	0.0011
∞	0.1	-	-	-2.634	-0.428	0.000	0.428	2.634	-
	0.3	-	-	-3.111	-0.506	0.000	0.506	3.111	-
	0.5	-	-	-3.592	-0.584	0.000	0.584	3.592	-
	0.7	-	-	-4.072	-0.662	0.000	0.662	4.072	-
	0.9	-	-	-4.549	-0.740	0.000	0.740	4.549	-

Table 2: Characteristics of the empirical distribution of $\frac{n}{\log n}(\hat{\rho}_n - \rho)$ over 50,000 simulated paths. The empirical α-quantile is denoted q_α. The last column gives the frequency of $\hat{\rho}_n$ exceeding 1. The results for $n = \infty$ are obtained by simulations of the asymptotic distribution in (5.6).

	ρ	$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$
R_n^*	0.1	-0.175	-0.045	0.000	0.047	0.195
	0.3	-0.664	-0.170	0.000	0.173	0.695
	0.5	-1.305	-0.337	0.000	0.345	1.339
	0.7	-2.058	-0.535	0.001	0.545	2.068
	0.9	-2.596	-0.724	0.001	0.749	2.731
R_n	0.1	-0.174	-0.045	0.000	0.047	0.194
	0.3	-0.663	-0.170	0.000	0.173	0.692
	0.5	-1.305	-0.337	0.000	0.341	1.327
	0.7	-2.045	-0.538	0.000	0.536	2.027
	0.9	-2.528	-0.719	0.000	0.719	2.599
R	0.1	-0.287	-0.047	0.000	0.047	0.287
	0.3	-1.148	-0.187	0.000	0.187	1.148
	0.5	-2.394	-0.390	0.000	0.390	2.394
	0.7	-4.024	-0.654	0.000	0.654	4.024
	0.9	-6.030	-0.981	0.000	0.981	6.030

Table 3: Characteristics of the empirical distributions of $\frac{n}{\log n} R_n^*$ and $\frac{n}{\log n} R_n$ over 50,000 simulated paths for $n = 5,000$. The asymptotic distribution in Proposition 5.3 (the law of R) is evaluated by simulation.

	ρ	$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$
T_n^*	0.1	-0.595	-0.127	0.000	0.130	0.610
	0.3	-0.606	-0.128	0.000	0.130	0.614
	0.5	-0.610	-0.130	0.000	0.129	0.616
	0.7	-0.635	-0.133	0.000	0.131	0.622
	0.9	-0.651	-0.135	0.000	0.134	0.662
T_n	0.1	-51.33	-44.54	-0.220	44.19	51.11
	0.3	-121.5	-105.3	0.139	104.8	121.0
	0.5	-144.5	-125.2	-0.162	124.5	144.0
	0.7	-82.62	-71.47	-0.272	71.07	82.16
	0.9	-8.570	-7.384	-0.054	7.348	8.518

Table 4: Characteristics of the empirical distributions of $\frac{n}{\log n}T_n^*$ and $\frac{n}{\log n}T_n$ over 5,000 simulated paths for $n = 5,000$.

residual. Let us consider the statistics:

$$T_n^* = \sum_{t=2}^{n-1} \hat{\varepsilon}_t^* (\hat{\varepsilon}_{t-1}^*)^2 / D_n^*, \quad T_n = \sum_{t=3}^{n} \hat{\varepsilon}_t (\hat{\varepsilon}_{t-1})^2 / D_n,$$

where $D_n^* = \sqrt{\left(\sum_{t=1}^{n-1} (\hat{\varepsilon}_t^*)^2\right)\left(\sum_{t=1}^{n-1} (\hat{\varepsilon}_t^*)^4\right)}$, and $D_n = \sqrt{\left(\sum_{t=1}^{n-1} (\hat{\varepsilon}_t)^2\right)\left(\sum_{t=1}^{n-1} (\hat{\varepsilon}_t)^4\right)}$. While (ε_t) is an i.i.d. sequence, the variables $u_t = Y_t - \rho Y_{t-1}$ are only "empirically uncorrelated" [see (5.9)] and this should reflect in the behaviour of T_n^* and T_n. The results presented in Table 4 confirm this intuition. The derivation of the asymptotic distributions of such statistics is left for further research.

6.3. ML estimation of the non causal AR(1) Cauchy model

To gauge the efficiency loss due to the LS estimation of the non causal AR(1) Cauchy model, we studied by Monte-Carlo experiments the properties of the ML estimator. Results are displayed in Table 5. The efficiency loss of the LSE appears clearly by comparing Tables 5 and 2, but the MLE requires knowledge of the errors distribution which is a strong assumption.

n	ρ	Mean	Std	$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$
500	0.1	-0.020	4.190	-3.411	-1.029	-0.001	0.996	3.414
	0.3	-0.049	3.633	-3.278	-1.010	-0.002	0.960	3.218
	0.5	-0.063	3.023	-2.968	-0.960	-0.003	0.889	2.872
	0.7	-0.073	2.334	-2.431	-0.824	-0.004	0.763	2.251
	0.9	-0.069	1.370	-1.472	-0.509	-0.005	0.454	1.296
2000	0.1	-0.013	4.075	-3.477	-1.040	-0.000	1.011	3.409
	0.3	-0.020	3.643	-3.280	-1.007	-0.003	0.984	3.221
	0.5	-0.020	3.042	-2.935	-0.946	-0.001	0.917	2.877
	0.7	-0.023	2.330	-2.378	-0.795	-0.000	0.775	2.319
	0.9	-0.024	1.348	-1.428	-0.495	-0.001	0.481	1.368
5000	0.1	-0.012	4.323	-3.485	-1.051	-0.001	1.007	3.490
	0.3	-0.018	3.721	-3.314	-1.019	-0.001	0.981	3.279
	0.5	-0.017	3.074	-2.953	-0.961	-0.000	0.927	2.899
	0.7	-0.012	2.358	-2.363	-0.807	-0.001	0.794	2.343
	0.9	-0.007	1.372	-1.422	-0.492	0.000	0.489	1.404

Table 5: Characteristics of the empirical distribution of $n(\hat{\rho}_{ML,n} - \rho)$ over 5,000 simulated paths of non causal AR(1) Cauchy models. The empirical α-quantile is denoted q_α.

7. An application

Many researchers found evidence of a speculative bubble in the series of the Nasdaq composite price index (see Homm and Breitung (2012), Phillips et al. (2011), and the references therein). Figure 5 plots the monthly time series of the Nasdaq real price[7] from February 1973 to December 2012. To gauge the adequacy of the noncausal Cauchy AR(1) model, we use the same data set as Phillips et al. (2011): the sample under study covers the period from February 1973 to June 2005 and comprises 389 observations. We introduce the statistics

$$Z_n^* = \frac{n}{\log n}\left|\frac{R_n^*}{\hat{\rho}_n(1+2\hat{\rho}_n)}\right|, \qquad Z_n = \frac{n}{\log n}\left|\frac{R_n}{\hat{\rho}_n(1+2\hat{\rho}_n)}\right|.$$

Proposition 5.3 shows that the adequacy of the model is rejected at level $\alpha \in (0,1)$ if $Z_n^* > \zeta_{1-\alpha}$, where $\zeta_{1-\alpha}$ is the $(1-\alpha)$ quantile of the variable $|XY|$, using for instance the statistics Z_n^*. The results displayed in Table 6 (left panel) show that the null hypothesis of a noncausal Cauchy AR(1) model cannot be rejected at any reasonable level for the tests based on Z_n or Z_n^*. The p-values are obtained from $1,000,000$ simulations of the variable $|Y|X$ of Proposition 5.3. In view of Proposition 5.4, if the DGP was a unit-root or near unit-root model of the form (5.10) under **A0-A1**, the statistics Z_n and Z_n^* would converge

[7] The Consumer Price Index (CPI), which can be obtained from the Federal Reserve Bank of St. Louis, was used to convert nominal prices into real prices.

to infinity in probability. Thus, the probability that the p-value be arbitrarily close to 1 should converge to 1. The results clearly do not support this hypothesis. In other words, the test procedures confirm what is seen on Figure 5. Indeed, the local explosive behaviour is more visible than the slight global stochastic trend.

Finally, we estimated the Cauchy AR(1) model with Gaussian component in (4.7). To compute the CF estimator, we used a discrete set up based on a grid $(u_1, \lambda_1), \ldots, (u_m, \lambda_m)$ for evaluating of the integral in (4.12). A preliminary estimator was obtained using a uniform weighting function $W(\cdot)$. The estimator in (4.12) is thus obtained as

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{m} |\widetilde{\Psi}_n(u_i, \lambda_i u_i) - \Psi(u_i, \lambda_i u_i)|^2.$$

We then considered the weight function

$$W_{\boldsymbol{\theta}}(u, \lambda) = [\mathrm{Var}(\cos\{u(Z_t + \lambda Z_{t+1})\})]^{-1} = \{0.5(1 + \Psi(2u, 2\lambda u)) - \{\Psi(u, \lambda u)\}^2\}^{-1}.$$

where the CF function Ψ is evaluated at the parameter $\boldsymbol{\theta}$. The second step estimator is thus obtained as

$$\widehat{\boldsymbol{\theta}}_n^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{m} |\widetilde{\Psi}_n(u_i, \lambda_i u_i) - \Psi(u_i, \lambda_i u_i)|^2 W_{\widehat{\boldsymbol{\theta}}_n}(u_i, \lambda_i).$$

The results of Table 6 (right panel) (obtained for the grid $u \in \{0.0005, 0.001, 0.01, 0.1, 0.5, 1, 2, 3, 5\}$ and $\lambda \in \{0.1, -0.2, 0.5, -1, -0.5, 1, 2, -2, 5\}$) confirm the existence of a non-causal Cauchy AR coefficient close to 1, while the Gaussian part presents a relatively small AR coefficient. The causal part is thus clearly stationary.

$\hat{\rho}_n$	Z_n^*	Z_n	pval(Z_n^*)	pval(Z_n)
0.998	0.980	1.030	0.341	0.333

$\hat{\rho}_n$	$\hat{r}_n$	$\hat{c}_n$	$\hat{\sigma}_n$
0.978	0.263	1.377	1.048

Table 6: NASDAQ index : Testing adequacy of the Cauchy AR(1) model (left panel); Cauchy AR(1) model with a Gaussian component (right panel).

8. Concluding remarks

By considering a noncausal AR(1) process, with α-stable noncausal innovations, we have derived special nonlinear features of its causal representation, such as locally explosive

[7] The rejection of nonstationarity was expected since the objective of the transformation of nominal prices to real prices is to eliminate global trend effects.

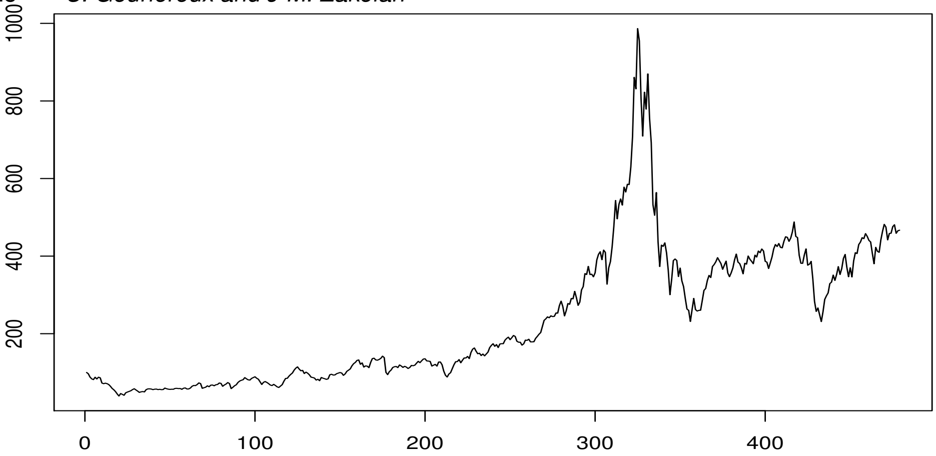

Figure 5. Real Nasdaq prices from February 1973 to December 2012.

features. The basic noncausal α-stable AR(1) process can be used as a cornerstone to create local explosions in dynamic models, with different magnitudes and rates of explosion by aggregation. We have seen that the noncausal Cauchy AR(1) process features a unit root. This questions the interpretation of the unit root hypothesis. Indeed, a unit root can represent a global explosive behaviour (stochastic trend) as well as a local explosion (bubble). We discussed the interpretation of the standard tests introduced in the literature to detect bubbles. These tests are often designed for detecting global explosions rather than local ones. We also highlighted the possibility to predict the times at which bubbles collapse. The analysis of our paper may help explain why mixed causal-noncausal linear AR models with heavy-tailed errors provide good fit on a large number of macroeconomic and financial time series. Indeed, these models are able to represent jumps, bubbles, and more generally asymmetric peaks with different speeds of increase and decrease. Such nonlinear features are often encountered in speculative markets, such as the market of physical commodities or the market of electronic currencies as the bitcoin (see e.g. Gouriéroux, Hencic (2015)).

Acknowledgements: We thank C. Francq and J.Jasiak for helpful comments. We gratefully acknowledge financial support of the Chaire ACPR "Regulation and Systemic Risk" and the Global Risk Institute. The second author acknowledges financial support of the French ANR via the Project PRAM ANR-10-ORAR-008-01.

Appendix: Proofs and complementary results

A. Proofs

A.1. Proof of Proposition 3.1

For $\alpha \neq 1$, the characteristic function of Y_t is

$$\begin{aligned} E[\exp(ivY_t)] &= \prod_{h=0}^{\infty} E[\exp\{iv\rho^h \varepsilon_{t+h}\}] \\ &= \exp \sum_{h=0}^{\infty} (-\sigma^\alpha |v|^\alpha |\rho|^{h\alpha} \left\{1 - i\beta\,(\text{sign } v)\left(\text{sign}(\rho)^h\right) \tan\left(\frac{\pi\alpha}{2}\right)\right\}) \\ &= \exp\left\{-\sigma^\alpha |v|^\alpha \left(\frac{1}{1-|\rho|^\alpha} - \frac{i\beta\,(\text{sign}(v))}{1-\text{sign}(\rho)|\rho|^\alpha} \tan\left(\frac{\pi\alpha}{2}\right)\right)\right\} \\ &= \exp\left\{-\left(\frac{\sigma}{(1-|\rho|^\alpha)^{1/\alpha}}\right)^\alpha |v|^\alpha \left(1 - \frac{i\beta\,(\text{sign}(v))\,(1-|\rho|^\alpha)}{1-\text{sign}(\rho)|\rho|^\alpha} \tan\left(\frac{\pi\alpha}{2}\right)\right)\right\}. \end{aligned}$$

This is the characteristic function of a stable distribution whose asymmetry parameter depends on the sign of ρ. For $\alpha = 1$, we have

$$\begin{aligned} E[\exp(ivY_t)] &= \prod_{h=0}^{\infty} E[\exp\{iv\rho^h \varepsilon_{t+h}\}] \\ &= \exp\left\{\sum_{h=0}^{\infty} -\sigma|v||\rho|^h - iv\beta\sigma\frac{2}{\pi}\sum_{h=0}^{\infty} \rho^h \log|v\rho^h|\right\} \\ &= \exp\left\{\frac{-\sigma|v|}{1-|\rho|} - iv\beta\sigma\frac{2}{\pi}\left(\frac{\log|v|}{1-\rho} + \frac{\rho\log|\rho|}{(1-\rho)^2}\right)\right\}. \end{aligned}$$

A.2. Proof of Proposition 3.2

i) Let us first show that the causal Markov property holds (see also Cambanis and Fakhre-Zakeri (1994), p. 217). Denote by f_* the transition pdf in direct time and by f^* the transition pdf in reverse time. For any lag p, we have:

$$\begin{aligned} f_*(Y_t|Y_{t-1},\ldots,Y_{t-p}) &= \frac{f(Y_t, Y_{t-1},\ldots,Y_{t-p})}{f(Y_{t-1},\ldots,Y_{t-p})} \\ &= \frac{f(Y_t)f^*(Y_{t-1}|Y_t)\ldots f^*(Y_{t-p}|Y_{t-p+1})}{f(Y_{t-1})f^*(Y_{t-2}|Y_{t-1})\ldots f^*(Y_{t-p}|Y_{t-p+1})} = \frac{f(Y_t)f^*(Y_{t-1}|Y_t)}{f(Y_{t-1})}. \end{aligned}$$

We deduce that the process (Y_t) is also causal Markov, with causal transition:

$$f_*(Y_t|Y_{t-1}) = \frac{f(Y_t)f^*(Y_{t-1}|Y_t)}{f(Y_{t-1})}. \tag{A.1}$$

ii) Now, the forward recursive equation at horizon $h+1$ is given by

$$\begin{aligned} Y_{t-1} &= \rho^{h+1} Y_{t+h} + \varepsilon_{t-1} + \rho \varepsilon_t + \ldots + \rho^h \varepsilon_{t+h-1} \\ &= \rho^{h+1} Y_{t+h} + \varepsilon_{t-1,h}, \text{ say.} \end{aligned}$$

The backward innovation $\varepsilon_{t-1,h}$ at lead $h+1$ follows a stable distribution with tail exponent α. Letting $f_{\varepsilon,h}$ denote the pdf of $\varepsilon_{t-1,h}$, the pdf of Y_{t-1} given Y_{t+h} is thus the function $y \mapsto f_{\varepsilon,h}\{y - \rho^{h+1} Y_{t+h}\}$. By the Bayes formula, the pdf of Y_{t+h} given $Y_{t-1} = y$ is thus the function

$$g : x \mapsto f_{\varepsilon,h}\{y - \rho^{h+1} x\} f_Y(x) / f_Y(y),$$

where f_Y denotes the marginal pdf of Y_t. If $|\beta| \neq 1$, the support of the stable pdf of $\varepsilon_{t-1,h}$ and Y_t is $\mathbb{R}$. It follows that when $x \to \pm\infty$,

$$g(x) \sim C(y)|x|^{-\alpha-1}|y - \rho^{h+1} x|^{-\alpha-1} \sim C^*(y)|x|^{-2(\alpha+1)},$$

where $C(y)$ and $C^*(y)$ are constants depending on y, which may change according to whether $x \to +\infty$ or $x \to -\infty$. Thus the integral of $|x|^p g(x)$ over any infinite interval excluding 0 (resp. over any finite interval including 0) is finite iff $p < 2\alpha + 1$ (resp. $p > -1$).

Now if $|\beta| = 1$, the support of the stable pdf of $\varepsilon_{t-1,h}$ and Y_t is either $\mathbb{R}^+$ or $\mathbb{R}^-$. It follows that when $\rho^{h+1} > 0$, the support of the density g is a compact; when $\rho^{h+1} < 0$ it is bounded below or above. Thus we have established the proposition.

A.3. Proof of Proposition 3.3

When $\beta = 0$, we have, by the arguments used to obtain Proposition 3.1,

$$Y_t \sim \mathcal{S}\left(\alpha, 0, \frac{\sigma}{(1-|\rho|^\alpha)^{1/\alpha}}, 0\right), \qquad \varepsilon_{t-1,h} \sim \mathcal{S}\left(\alpha, 0, \sigma\left(\frac{1-|\rho|^{(h+1)\alpha}}{1-|\rho|^\alpha}\right)^{1/\alpha}, 0\right).$$

It follows that, for any $u \in \mathbb{R}$,

$$\begin{aligned} E\left(e^{iuY_{t-1}} \mid Y_{t+h}\right) &= e^{iu\rho^{h+1} Y_{t+h}} E\left(e^{iu\varepsilon_{t-1,h}} \mid Y_{t+h}\right) \\ &= \exp\left\{iu\rho^{h+1} Y_{t+h} - |\sigma u|^\alpha \frac{1-|\rho|^{(h+1)\alpha}}{1-|\rho|^\alpha}\right\}, \end{aligned}$$

and thus for any $u, v \in \mathbb{R}$,

$$\begin{aligned} E\left(e^{iuY_{t-1} + ivY_{t+h}}\right) &= E\left\{E\left(e^{iuY_{t-1}} \mid Y_{t+h}\right) e^{ivY_{t+h}}\right\} \\ &= \exp\left\{-|\sigma u|^\alpha \frac{1-|\rho|^{(h+1)\alpha}}{1-|\rho|^\alpha}\right\} E\left\{e^{i\{v + u\rho^{h+1}\} Y_{t+h}}\right\} \\ &= \exp\left\{-\left(|u|^\alpha (1-|\rho|^{(h+1)\alpha}) + |v + u\rho^{h+1}|^\alpha\right) \frac{\sigma^\alpha}{1-|\rho|^\alpha}\right\}. \end{aligned}$$

Thus, for $u > 0$ and $\rho^{h+1} > 0$,

$$\begin{aligned}
& \left[\frac{\partial}{\partial u} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} \\
= & -E(e^{iuY_{t-1}}) \left(|u|^{\alpha-1}(1-|\rho|^{(h+1)\alpha}) + \rho^{h+1}|u\rho^{h+1}|^{\alpha-1}\right) \frac{\alpha\sigma^\alpha}{1-|\rho|^\alpha} \\
= & -E(e^{iuY_{t-1}})|u|^{\alpha-1}\frac{\alpha\sigma^\alpha}{1-|\rho|^\alpha}, \qquad (A.2)
\end{aligned}$$

and

$$\begin{aligned}
\left[\frac{\partial}{\partial v} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} &= -E(e^{iuY_{t-1}})|u\rho^{h+1}|^{\alpha-1}\frac{\alpha\sigma^\alpha}{1-|\rho|^\alpha} \qquad (A.3) \\
&= \rho^{(h+1)(\alpha-1)} \left[\frac{\partial}{\partial u} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0}.
\end{aligned}$$

On the other hand, for $u \neq 0$,

$$\begin{aligned}
\left[\frac{\partial}{\partial u} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} &= iE(Y_{t-1}e^{iuY_{t-1}}), \\
\left[\frac{\partial}{\partial v} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} &= iE(Y_{t+h}e^{iuY_{t-1}}).
\end{aligned}$$

Note that the latter expectations exist, by the Dirichlet's test[8] for improper integrals and using the equivalent of the density of an α-stable variable in the neighborhood of infinity, $f(x) \sim \frac{K}{|x|^{\alpha+1}}$. Therefore, for $u > 0$ and $\rho^{h+1} > 0$,

$$E\left\{\left(Y_{t+h} - \rho^{(h+1)(\alpha-1)}Y_{t-1}\right)e^{iuY_{t-1}}\right\} = 0. \qquad (A.4)$$

It can be checked that for $u < 0$ and $\rho^{h+1} > 0$ both derivatives in (A.2) and (A.3) have opposite signs, thus (A.4) continues to hold. If now $\rho^{h+1} < 0$, we obtain

$$\left[\frac{\partial}{\partial v} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} = -(-\rho)^{(h+1)(\alpha-1)} \left[\frac{\partial}{\partial u} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0}, \quad \text{if } \alpha \neq 1$$

and

$$\left[\frac{\partial}{\partial v} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} = -\left[\frac{\partial}{\partial u} E(e^{iuY_{t-1}+ivY_{t+h}})\right]_{v=0} \quad \text{if } \alpha = 1.$$

Finally, we have

$$E\left\{\left(Y_{t+h} - \text{sign}(\rho)|\rho|^{(h+1)(\alpha-1)}Y_{t-1}\right)e^{iuY_{t-1}}\right\} = 0, \quad \text{for any } u \in \mathbb{R}. \qquad (A.5)$$

The conclusion follows from Bierens (Theorem 1, 1982).

[8]Let f and g denote two real functions defined on $[a,\infty)$ and regulated on every interval $[a,b]$ with $b > a$ (that is, admitting left-hand and right-hand limits at all points $x > a$ and a right-hand limit at a). If f is decreasing and $\lim_{x\to\infty} f(x) = 0$, if $\sup_b |\int_a^b g(x)dx| < \infty$, then the integral $\int_a^\infty f(x)g(x)dx$ exists.

A.4. Proof of Proposition 3.4

From the Proof of Proposition 3.2, the pdf of Y_t given $Y_{t-h} = y$ is the function

$$g : x \mapsto f_{\varepsilon,h-1}\{y - \rho^h x\} f_Y(x)/f_Y(y),$$

where f_Y is the marginal pdf of Y_t and $f_{\varepsilon,h-1}$ is the pdf of $\sum_{i=0}^{h-1} \rho^i \varepsilon_{t-i}$. By Proposition 3.1, $Y_t \sim \mathcal{C}\left(0, \frac{\sigma}{1-|\rho|}\right)$, and $f_{\varepsilon,h-1}$ is the pdf of the $\mathcal{C}(0, \sigma_h)$. The conclusion follows.

A.5. Proof of Proposition 3.5

The proof of Proposition 3.5 uses the special form of the transition pdf and is given for $\sigma = 1$ and $\rho \neq 0$.

i) Let us first compute the conditional moment of $1 + (1 - |\rho|)^2 Y_t^2$. We get:

$$\begin{aligned} E_{t-1}[1 + (1 - |\rho|)^2 Y_t^2] &= \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{1}{1 + (Y_{t-1} - \rho Y_t)^2} [1 + (1 - |\rho|)^2 Y_{t-1}^2] dY_t \\ &= \frac{1}{\pi} [1 + (1 - |\rho|)^2 Y_{t-1}^2] \int_{-\infty}^{+\infty} \frac{1}{1 + (Y_{t-1} - \rho Y_t)^2} dY_t \\ &= \frac{1}{|\rho|} [1 + (1 - |\rho|)^2 Y_{t-1}^2]. \end{aligned}$$

We deduce the second equality in Proposition 3.5:

$$E(Y_t^2 \mid Y_{t-1}) = \frac{1}{|\rho|} Y_{t-1}^2 + \frac{1}{|\rho|(1 - |\rho|)}. \tag{A.6}$$

ii) By the same method, we can retrieve the conditional mean already obtained in Proposition 3.3 using characteristic functions. Let us symmetrically compute:

$$\begin{aligned} E_{t-1}[1 + (Y_{t-1} - \rho Y_t)^2] &= \frac{1}{\pi} [1 + (1 - |\rho|)^2 Y_{t-1}^2] \int_{-\infty}^{-\infty} \frac{1}{1 + (1 - |\rho|)^2 Y_t^2} dY_t \\ &= \frac{1}{1 - |\rho|} [1 + (1 - |\rho|)^2 Y_{t-1}^2] \\ &= \frac{1}{1 - |\rho|} + (1 - |\rho|) Y_{t-1}^2. \end{aligned}$$

We deduce that:

$$\begin{aligned} 2\rho Y_{t-1} E(Y_t \mid Y_{t-1}) &= 1 + Y_{t-1}^2 + \rho^2 E(Y_t^2 \mid Y_{t-1}) - \frac{1}{1 - |\rho|} - (1 - |\rho|) Y_{t-1}^2 \\ &= -\frac{|\rho|}{1 - |\rho|} + |\rho| Y_{t-1}^2 + \rho^2 E(Y_t^2 \mid Y_{t-1}) = 2|\rho| Y_{t-1}^2, \end{aligned}$$

by applying expression (A.6). Therefore, we have $E(Y_t \mid Y_{t-1}) = \text{sign}\rho Y_{t-1}$, which provides the first formula in Proposition 3.5.

A.6. Proof of Proposition 3.6

Since the transition pdf in reverse time being is given by:

$$f^*(Y_{t-1}|Y_t) = \frac{1}{\sqrt{2\pi}} \frac{1}{(Y_{t-1} - \rho Y_t)^{3/2}} \exp\left(\frac{-1}{2(Y_{t-1} - \rho Y_t)}\right) \mathbb{I}_{Y_{t-1} - \rho Y_t > 0},$$

we deduce, by (A.1) and (3.7), the transition pdf in direct time is

$$\begin{aligned} f_*(Y_t|Y_{t-1}) &= \frac{1}{\sqrt{2\pi}} \left(\frac{Y_{t-1}}{Y_t(Y_{t-1} - \rho Y_t)}\right)^{3/2} \exp\left(\frac{-1}{2(1-\sqrt{\rho})^2}\left(\frac{1}{Y_t} - \frac{1}{Y_{t-1}}\right)\right) \\ &\quad \times \exp\left(\frac{-1}{2(Y_{t-1} - \rho Y_t)}\right) \mathbb{I}_{0 < \rho Y_t < Y_{t-1}}. \end{aligned}$$

The conclusion follows.

A.7. Proof of Proposition 4.1

We have

$$|\mathcal{Y}_t| = \left| \int \sum_{i \geq 0} \rho^i \varepsilon_{\rho,t+i} d\pi(\rho) \right| = \left| \sum_{i \geq 0} X_{t,i} \right|, \qquad X_{t,i} = \int \rho^i \varepsilon_{\rho,t+i} d\pi(\rho).$$

To verify that $|\mathcal{Y}_t|$ is a.s. finite, it suffices to check that $E|\mathcal{Y}_t|^s < \infty$. Note that, for any integrable function H (with respect to π), the random variable $\int \varepsilon_{\rho,t} H(\rho) d\pi(\rho)$ follows a Cauchy distribution with scale parameter $E_\pi|H(\rho)|$, provided the latter expectation is finite. Therefore, the variable $X_{t,i}$ follows a Cauchy distribution with scale parameter $E_\pi(|\rho|^i)$. Thus $E|X_{t,i}|^s = \{E_\pi(|\rho|^i)\}^s m_s$ where, for $s \in (0,1)$, m_s is the s-th order moment of the $\mathcal{C}(0,1)$. Using the elementary inequality $(x+y)^s \leq x^s + y^s$ for $x, y \geq 0$ and $s \in (0,1)$, we then have

$$E|\mathcal{Y}_t|^s \leq \sum_{i \geq 0} E\left|X_{t,i}\right|^s = \sum_{i \geq 0} \{E_\pi(|\rho|^i)\}^s m_s < \infty.$$

Thus $\mathcal{Y}_t$ is well defined and (4.4) holds. The Cauchy distribution of $\mathcal{Y}_t$ follows, noting that $E_\pi \left\{\frac{1}{1-|\rho|}\right\}$ exists under (4.3).

The strict stationarity of $(\varepsilon_{\rho,t})_{t \in \mathbb{Z}}$, for any $\rho \in (-1,1)$, entails that the joint distribution of any n-uple $(\mathcal{Y}_{t_1+h}, \ldots, \mathcal{Y}_{t_n+h})$ does not depend of h. The strict stationarity of $(\mathcal{Y}_t)$ follows.

Let us now turn to ergodicity. Denote by $(\Omega, \mathcal{B}, P)$ the underlying probability space and let $\Omega^+ = \mathbb{R}^{\mathbb{Z}^+}$ denote the space of all functions $\mathbb{Z}^+ \mapsto \mathbb{R}$, endowed with the product

σ-algebra $\mathcal{A} = \mathcal{B}(\mathbb{R})^{\mathbb{Z}^+}$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. A set $B \in \mathcal{B}$ is called invariant for $(\mathcal{Y}_t)_{t\geq 0}$ if there exists $A \in \mathcal{A}$ such that, for all $t \in \mathbb{Z}$,

$$B = \{(\mathcal{Y}_t, \mathcal{Y}_{t+1}, \dots) \in A\}. \tag{A.7}$$

The process $(\mathcal{Y}_t)_{t\geq 0}$ is ergodic if any invariant set $B \in \mathcal{A}$ satisfies $P(B) = 0$ or $P(B) = 1$ (see for instance Krengel (1985), p. 26). We have, by (4.4),

$$\mathcal{Y}_t = \sum_{i\geq 0} X_{t+i}^{[i]}, \quad \text{where} \quad X_{t+i}^{[i]} = \int \rho^i \varepsilon_{\rho,t+i} d\pi(\rho)$$

are independent (though not identically distributed) random variables. Let $A = A_0 \times A_1 \times \dots \in \mathcal{A}$ such that (A.7) holds for all $t \in \mathbb{Z}$. It follows that

$$\begin{aligned} B &= \{(\mathcal{Y}_0, \mathcal{Y}_1, \dots) \in A\} \cap \{(\mathcal{Y}_1, \mathcal{Y}_2, \dots) \in A\} \\ &= \{\mathcal{Y}_0 \in A_0, \mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots\} \\ &= \{X_0^{[0]} + \sum_{i\geq 1} X_i^{[i]} \in A_0, \mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots\}. \end{aligned}$$

Denoting by g the density of $\sum_{i\geq 1} X_i^{[i]}$ with respect to the Lebesgue measure on $\mathbb{R}$, we thus have, using the independence between the $X_i^{[i]}$,

$$\begin{aligned} P(B) &= \int P\left[X_0^{[0]} + z \in A_0, \mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots \middle| \sum_{i\geq 1} X_i^{[i]} = z\right] g(z)dz \\ &= \int P\left[X_0^{[0]} + z \in A_0\right] P\left[\mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots \middle| \sum_{i\geq 1} X_i^{[i]} = z\right] g(z)dz \\ &:= \int P\left[X_0^{[0]} + z \in A_0\right] G(z)dz. \end{aligned} \tag{A.8}$$

It follows that

$$\begin{aligned} P(B) &\leq \int P\left[\mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots \middle| \sum_{i\geq 1} X_i^{[i]} = z\right] g(z)dz \\ &= P\left[\mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots\right]. \end{aligned}$$

But $B = \{\mathcal{Y}_1 \in A_0, \mathcal{Y}_2 \in A_1, \dots\}$ entails

$$P(B) \geq P\left[\mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots\right].$$

Hence,

$$P(B) = P\left[\mathcal{Y}_1 \in A_0 \cap A_1, \mathcal{Y}_2 \in A_1 \cap A_2, \dots\right] = \int G(z)dz. \tag{A.9}$$

Assume that $P(B) \neq 0$. Then, the set $E = \{z, G(z) > 0\}$ has positive Lebesgue measure and, by (A.8) and (A.9), we get $P\left[X_0^{[0]} + z \in A_0\right] = 1$ for almost all $z \in E$. But since $X_0^{[0]}$ has a continuous distribution, this entails $A_0 = \mathbb{R}$. By induction, we can prove that $A_k = \mathbb{R}$ for all $k \geq 0$, from which we deduce that $P(B) = 1$. To conclude, note that sets of the form $A_0 \times A_1 \times \ldots$ generate $\mathcal{A}$. The ergodicity of $(\mathcal{Y}_t)_{t\geq 0}$ is thus established.

A.8. Proof of Proposition 4.2

We have,

$$
\begin{aligned}
&\Psi(u_0, \ldots, u_k)\\
= \; & E\left[\exp\left\{i\left(\sum_{\ell=0}^{k} u_\ell \int Y_{\rho,t+\ell} d\pi(\rho)\right)\right\}\right]\\
= \; & E\left[\exp\left\{i\left(\sum_{\ell=0}^{k} u_\ell \int \sum_{j\geq 0} \rho^j \varepsilon_{\rho,t+\ell+j} d\pi(\rho)\right)\right\}\right]\\
= \; & E\left[\exp\left\{i\int\left[\sum_{h=0}^{k-1} \varepsilon_{\rho,t+h}\left(\sum_{\ell=0}^{h} \rho^{h-\ell} u_\ell\right) + \left(\sum_{\ell=0}^{k} \rho^{k-\ell} u_\ell\right) \sum_{h=k}^{\infty} \rho^{h-k} \varepsilon_{\rho,t+h}\right] d\pi(\rho)\right\}\right].
\end{aligned}
$$

Note that, for any integrable function H (with respect to π), the random variable $\int \varepsilon_{\rho,t} H(\rho) d\pi(\rho)$ follows a Cauchy distribution with scale parameter $E_\pi|H(\rho)|$, provided the latter expectation is finite. Thus we have,

$$
E\left[\exp\left\{i\int \varepsilon_{\rho,t} H(\rho)\, d\pi(\rho)\right\}\right] = \exp\left[-E_\pi\, |H(\rho)|\right].
$$

By the independence of the sequence $(\varepsilon_{\cdot,t})_{t\in\mathbb{Z}}$, the conclusion follows.

A.9. Proof of Proposition 4.3

We have, for $u \in (0, 1)$,

$$
\begin{aligned}
\tilde{\Psi}(u) &= -c\int_{-1}^{-u}\left(1 - \frac{\rho+u}{1+\rho}\right)\pi(\rho)d\rho - c\int_{-u}^{0}\left(1 + \frac{\rho+u}{1+\rho}\right)\pi(\rho)d\rho\\
&\quad -c\int_{0}^{1}\left(1 + \frac{\rho+u}{1-\rho}\right)\pi(\rho)d\rho\\
&= -c(1-u)\int_{-1}^{-u}\frac{1}{1+\rho}\pi(\rho)d\rho - c\int_{-u}^{0}\left(\frac{1+2\rho+u}{1+\rho}\right)\pi(\rho)d\rho\\
&\quad -c(1+u)\int_{0}^{1}\frac{1}{1-\rho}\pi(\rho)d\rho,
\end{aligned}
$$

and for $u \in (-1, 0)$,

$$\begin{aligned} \tilde{\Psi}(u) &= -c(1-u)\int_{-1}^{0} \frac{1}{1+\rho}\pi(\rho)d\rho - c\int_{0}^{-u}\left(\frac{1-2\rho-u}{1-\rho}\right)\pi(\rho)d\rho \\ &\quad -c(1+u)\int_{-u}^{1}\frac{1}{1-\rho}\pi(\rho)d\rho. \end{aligned}$$

Thus for $u \in (0, 1)$,

$$\begin{aligned} \frac{\partial\tilde{\Psi}(u)}{\partial u} &= c\int_{-1}^{-u}\frac{1}{1+\rho}\pi(\rho)d\rho - c\int_{-u}^{0}\frac{1}{1+\rho}\pi(\rho)d\rho - c\int_{0}^{1}\frac{1}{1-\rho}\pi(\rho)d\rho, \\ \frac{\partial^2\tilde{\Psi}(u)}{\partial u^2} &= -\frac{2c}{1-u}\pi(-u), \end{aligned}$$

and for $u \in (-1, 0)$,

$$\begin{aligned} \frac{\partial\tilde{\Psi}(u)}{\partial u} &= c\int_{-1}^{0}\frac{1}{1+\rho}\pi(\rho)d\rho + c\int_{0}^{-u}\frac{1}{1-\rho}\pi(\rho)d\rho - c\int_{-u}^{1}\frac{1}{1-\rho}\pi(\rho)d\rho, \\ \frac{\partial^2\tilde{\Psi}(u)}{\partial u^2} &= -\frac{2c}{1+u}\pi(-u). \end{aligned}$$

The formulas for π and c follow.

A.10. Proof of Proposition 5.1

The result is a consequence of Davis and Resnick (1985, Theorem 4.2). Indeed, the AR process (Y_t) admits an infinite moving average representation (3.3), in which the sequence (ε_t) is i.i.d. with regularly varying tail probabilities. We have $Y_t = \sum_{h=0}^{\infty} c_h\varepsilon_{t+h}$ with $c_h = \rho^h$ and $\rho < 1$; so the condition (2.6) in Davis and Resnick (1985), $\sum_{h=0}^{\infty}|c_h|^\delta < \infty$, is satisfied for any $\delta > 0$.

A.11. Tail behaviours in the Cauchy and Pareto cases

i) First suppose that, in (5.1), ε_t/σ has a standard Cauchy distribution. It follows from the definition of a_n (see Proposition 5.2), that a_n/σ is the quantile of order $1/2n$ of the standard Cauchy distribution:

$$a_n = \sigma\left[\tan\left\{\frac{\pi}{2n}\right\}\right]^{-1} \sim \frac{2\sigma}{\pi}n.$$

Elementary computation shows that, for $x > 0$ and $u_t = \varepsilon_t/\sigma$,

$$P(|u_1u_2| > x) = \left(\frac{2}{\pi}\right)^2\int_0^\infty \tan^{-1}\left(\frac{y}{x}\right)\frac{1}{1+y^2}dy := I_1(x) + I_2(x) + I_3(x)$$

where

$$I_1(x) = \left(\frac{2}{\pi}\right)^2 \int_0^x \frac{y}{x} \frac{1}{1+y^2} dy \sim \left(\frac{2}{\pi}\right)^2 \frac{\log x}{x} \quad \text{as } x \to \infty,$$

$$I_2(x) = \left(\frac{2}{\pi}\right)^2 \int_0^x \left(\frac{y}{x}\right)^3 g\left(\frac{y}{x}\right) \frac{1}{1+y^2} dy,$$

where g is a function which is continuous on $[0,1]$, from which it follows that $I_2(x) = O(1/x)$ as $x \to \infty$, and, for some constant $K > 0$,

$$I_3(x) = \left(\frac{2}{\pi}\right)^2 \int_x^\infty \tan^{-1}\left(\frac{y}{x}\right) \frac{1}{1+y^2} dy \leq K \tan^{-1}\left(\frac{1}{x}\right) = O\left(\frac{1}{x}\right) \quad \text{as } x \to \infty.$$

It follows that

$$P(|\varepsilon_1 \varepsilon_2| > x) \sim \left(\frac{2\sigma}{\pi}\right)^2 \frac{\log x}{x} \quad \text{as } x \to \infty.$$

Thus

$$\tilde{a}_n \sim \left(\frac{2\sigma}{\pi}\right)^2 n \log n.$$

ii) Now suppose that, in (5.1), $|u_t| = |\varepsilon_t|/\sigma$ is Pareto distributed. Thus $P(|u_1| > x) \sim (x/x_0)^{-\alpha}$ for some $x_0 > 0$ as $x \to \infty$. Thus $a_n = x_0 n^{1/\alpha}$. Elementary computation shows that

$$P(|u_1 u_2| > x) \sim x_0^{2\alpha} \frac{\log x^\alpha}{x^\alpha} \quad \text{as } x \to \infty.$$

Thus

$$\tilde{a}_n \sim x_0^{-2} (n \log n)^{1/\alpha}.$$

A.12. Proof of Proposition 5.2

The distribution of ε_t being symmetric, it follows from Davis and Resnick (1986, Theorem 4.4, ii)), that:

$$\frac{a_n^2}{\tilde{a}_n} \{\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}\} \xrightarrow{d} \left(\sum_{j=1}^{\infty} \{\rho(\ell+j) + \rho(\ell-j) - 2\rho(\ell)\rho(j)\} S_j / S_0 \right)_{\ell=1,\ldots,M}$$

where, in view of the MA(∞) representation (3.3), for $h \geq 0$,

$$\rho(h) = \rho(-h) = \frac{\sum_{j=0}^\infty \rho^j \rho^{j+h}}{\sum_{j=0}^\infty \rho^{2j}} = \rho^h.$$

The convergence in (5.5) follows. Results for the particular cases follow from Appendix A.11. For the Cauchy distribution, i) shows that $a_n^2/\tilde{a}_n \sim n/\log n$. For the Pareto distribution, ii) shows that $a_n^2/\tilde{a}_n \sim (n/\log n)^{1/\alpha}$.

A.13. Proof of Proposition 5.3

Let us denote, for $i \geq 0$,

$$C_n(i) = \sum_{t=1}^{n} Y_t Y_{t-i}, \quad D_n(i) = \sum_{t=1}^{n} \varepsilon_t \varepsilon_{t-i},$$

with by convention $Y_t = \varepsilon_t = 0$ for $t \leq 0$. By Davis and Resnick (1986, Theorem 3.3), we have the following weak convergences for the partial sums of the i.i.d. process:

$$(a_n^{-2} D_n(0), \tilde{a}_n^{-1} D_n(1), \ldots, \tilde{a}_n^{-1} D_n(h)) \overset{d}{\to} (S_0, S_1, \ldots, S_h),$$

where the variables S_i are as in Proposition 5.2. Up to some negligible terms, we have,

$$\begin{aligned}
\sum_{t=1}^{n-1} \hat{\varepsilon}_t^2 &= \sum_{t=1}^{n-1} \{\varepsilon_t + (\rho - \hat{\rho}_n) Y_{t+1}\}^2 \\
&= D_n(0) + 2(\rho - \hat{\rho}_n)\{C_n(1) - \rho C_n(0)\} + (\rho - \hat{\rho}_n)^2 C_n(0) \\
&= D_n(0) - (\rho - \hat{\rho}_n)^2 C_n(0).
\end{aligned}$$

Since $a_n^{-2} C_n(0)$ converges in distribution (Theorem 4.2 in Davis and Resnick, 1985), we have:

$$a_n^{-2} \sum_{t=1}^{n-1} \hat{\varepsilon}_t^2 = a_n^{-2} D_n(0) + O_P(a_n^{-2} \tilde{a}_n)^2. \tag{A.10}$$

We also have:

$$\begin{aligned}
\sum_{t=1}^{n-1} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} &= \sum_{t=1}^{n-1} \{\varepsilon_t + (\rho - \hat{\rho}_n) Y_{t+1}\}\{\varepsilon_{t-1} + (\rho - \hat{\rho}_n) Y_t\} \\
&= D_n(1) + (\rho - \hat{\rho}_n)\{C_n(2) - 2\rho C_n(1) + C_n(0)\} + (\rho - \hat{\rho}_n)^2 C_n(1).
\end{aligned}$$

Now, by using the MA(∞) representation (3.3), for $h \geq 0$,

$$\begin{aligned}
C_n(h) &= \sum_{t=1}^{n} \left(\sum_{i \geq 0} \rho^i \varepsilon_{t+i} \right) \left(\sum_{i \geq 0} \rho^i \varepsilon_{t+i-h} \right) \\
&= \sum_{t=1}^{n} \sum_{i \geq 0} \rho^{2i+h} (\varepsilon_{t+i})^2 + \sum_{t=1}^{n} \sum_{i,j \geq 0, i \neq j} \rho^{i+j+h} \varepsilon_{t+i} \varepsilon_{t+j}.
\end{aligned}$$

By Davis and Resnick (1986, Propositions 4.2 and 4.3), we have

$$a_n^{-2} \sum_{t=1}^{n} \sum_{i \geq 0, i \neq j} \rho^{i+j+h} \varepsilon_{t+i} \varepsilon_{t+j} \to 0 \qquad \text{in probability,}$$

$$a_n^{-2} \left(\sum_{t=1}^{n} \sum_{i \geq 0} \rho^{2i+h} (\varepsilon_{t+i})^2 - \sum_{i \geq 0} \rho^{2i+h} D_n(0) \right) \to 0 \qquad \text{in probability,}$$

and thus

$$a_n^{-2}\left(C_n(h) - \frac{\rho^h}{1-\rho^2}D_n(0)\right) \quad \to 0 \qquad \text{in probability.}$$

Moreover, we have

$$\hat{\rho}_n - \rho \quad = \quad \{C_n(0)\}^{-1}\sum_{t=1}^{n} Y_t\varepsilon_{t-1} = \{C_n(0)\}^{-1}\sum_{i\geq 0}\rho^i D_n(i+1).$$

Therefore

$$\begin{aligned}
\tilde{a}_n^{-1}\sum_{t=1}^{n-1}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} &= \tilde{a}_n^{-1}D_n(1) - \tilde{a}_n^{-1}\sum_{i\geq 0}\rho^i D_n(i+1)\frac{C_n(2)-2\rho C_n(1)+C_n(0)}{C_n(0)} + o_P(1)\\
&= \tilde{a}_n^{-1}D_n(1) - \tilde{a}_n^{-1}\{1-\rho^2\}\sum_{i\geq 0}\rho^i D_n(i+1) + o_P(1)\\
&= \rho^2\tilde{a}_n^{-1}D_n(1) - \tilde{a}_n^{-1}\{1-\rho^2\}\sum_{i\geq 1}\rho^i D_n(i+1) + o_P(1).
\end{aligned}$$

Together with (A.10), this establishes (5.7).

To get the asymptotic distribution of R_n under the assumptions of Proposition 5.3, write $R_n = U_n/V_n$ where $U_n = \sum_{t=3}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}$ and $V_n = \sum_{t=2}^{n}\hat{\varepsilon}_t^2$. Similarly, let $R_n^* = U_n^*/V_n^*$ where $U_n^* = \sum_{t=2}^{n-1}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}$ and $V_n^* = \sum_{t=1}^{n-1}(\hat{\varepsilon}_t)^2$. We note that

$$\begin{aligned}
V_n - V_n^* &= \sum_{t=2}^{n}(Y_t - \hat{\rho}_n Y_{t-1})^2 - \sum_{t=1}^{n-1}(Y_t - \hat{\rho}_n Y_{t+1})^2 = (1-\hat{\rho}_n^{*2})(Y_n^2 - Y_1^2),\\
U_n - U_n^* &= \sum_{t=3}^{n}(Y_t - \hat{\rho}_n Y_{t-1})(Y_{t-1} - \hat{\rho}_n Y_{t-2}) - \sum_{t=2}^{n-1}(Y_t - \hat{\rho}_n Y_{t+1})(Y_{t-1} - \hat{\rho}_n Y_t)\\
&= (1-\hat{\rho}_n^2)(Y_nY_{n-1} - Y_2Y_1).
\end{aligned}$$

By the stationarity of (Y_nY_{n-1}), we have $a_nY_nY_{n-1} \to 0$ in probability as $n \to \infty$, for any deterministic sequence (a_n) converging to zero. Because $V_n^* = O_P(n)$, it follows that $(n/\log n)^{1/\alpha}Y_nY_{n-1}/V_n^*$ tends to zero in probability when $\alpha \geq 1$. By the same argument we find that

$$\frac{n}{\log n}(R_n^* - R_n) \quad = \quad \frac{n}{\log n}\frac{U_n^* - U_n}{V_n^*} + \frac{n}{\log n}\frac{U_n}{V_nV_n^*}(V_n^* - V_n) \to 0,$$

in probability as $n \to \infty$. Therefore, the asymptotic distributions of $\frac{n}{\log n}R_n$ and $\frac{n}{\log n}R_n^*$ are the same when $\alpha \geq 1$.

Now we turn to the Cauchy case. In view of the independence between S_1 and the S_j for $j \geq 2$, the right hand side has the same distribution as the variable $\rho(1+2\rho)S_1/S_0$. By the arguments used to obtain (5.6), the last part of Proposition 5.3 is established.

A.14. Proof of Proposition 5.4

We have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=3}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} &= \frac{1}{\sqrt{n}}\sum_{t=3}^{n}\{\xi_t+(1-\hat{\rho}_n)Y_{n,t-1}\}\{\xi_{t-1}+(\hat{\rho}_n-1)Y_{n,t-2}\}\\ &= \frac{1}{\sqrt{n}}\sum_{t=3}^{n}\xi_t\xi_{t-1}+\frac{1}{\sqrt{n}}X_{1n}+\frac{1}{\sqrt{n}}X_{2n},\end{aligned}$$

where

$$\begin{aligned}X_{1n} &= n(1-\hat{\rho}_n)\left(\frac{1}{n}\sum_{t=3}^{n}Y_{n,t-1}\xi_{t-1}+\frac{1}{n}\sum_{t=3}^{n}Y_{n,t-2}\xi_t\right),\\ X_{2n} &= \{n(1-\hat{\rho}_n)\}^2\frac{1}{n^2}\sum_{t=3}^{n}Y_{n,t-1}Y_{n,t-2}.\end{aligned}$$

Phillips (1987) established the following weak convergence for the near unit root process (5.10):

$$(Z_{1n},Z_{2n})=\left(\frac{1}{n^2}\sum_{t=1}^{n}Y_{n,t}^2,\frac{1}{n}\sum_{t=1}^{n}Y_{n,t-1}\xi_t\right)\overset{d}{\to}(Z_1,Z_2)$$

for some real random variables Z_1 and Z_2 depending on c. It follows that, by the continuous mapping theorem,

$$X_{1n}=Z_{1n}^{-1}Z_{2n}\{a_nZ_{2n}+\sigma^2+a_n^{-1}Z_{2n}+o_P(1)\}\overset{d}{\to}Z_1^{-1}Z_2(2Z_2+\sigma^2),$$

and thus $\frac{1}{\sqrt{n}}X_{1n}\to 0$ in probability. Similarly, $\frac{1}{\sqrt{n}}X_{2n}\to 0$ in probability. It follows that

$$\frac{1}{\sqrt{n}}\sum_{t=3}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1}\overset{d}{=}\frac{1}{\sqrt{n}}\sum_{t=3}^{n}\xi_t\xi_{t-1}\overset{d}{\to}\mathcal{N}(0,\sigma_\xi^4).$$

We also have

$$\frac{1}{n}\sum_{t=3}^{n}\hat{\varepsilon}_t^2 = \frac{1}{n}\sum_{t=3}^{n}\{\xi_t+(1-\hat{\rho}_n)Y_{n,t-1}\}^2=\frac{1}{n}\sum_{t=3}^{n}\xi_t^2+\frac{1}{n}X_{3n}+\frac{1}{n}X_{4n},$$

where

$$X_{3n} = 2(1-\hat{\rho}_n)\sum_{t=3}^{n}Y_{n,t-1}\xi_t,\quad X_{4n}=\{n(1-\hat{\rho}_n)\}^2\frac{1}{n^2}\sum_{t=3}^{n}Y_{n,t-1}^2.$$

By arguments already used,

$$\frac{1}{n}\sum_{t=3}^{n}\hat{\varepsilon}_t^2 = \frac{1}{n}\sum_{t=3}^{n}\xi_t^2+o_P(1)=\sigma_\xi^2+o_P(1).$$

The convergence in probability of $n|R_n|/\log n$ to infinity follows. The convergence of $n|R_n^*|/\log n$ is established similarly.

B. Causal conditional cdf

Tedious computation in the Cauchy case, using the conditional density function derived in Proposition 3.4, allows to obtain the conditional cdf.

PROPOSITION B.1. *The causal conditional cdf of the noncausal Cauchy linear AR process is given, for $\sigma = 1$, by:*

$$\begin{aligned} F(Y_t|Y_{t-1}) &= \frac{\alpha(Y_{t-1},\rho^*)}{\pi} \log\left\{\frac{1+(1-|\rho^*|)^2 Y_t^2}{1+(Y_{t-1}-\rho^* Y_t)^2}\frac{\rho^{*2}}{(1-|\rho^*|)^2}\right\} \\ &+ \frac{\beta(Y_{t-1},\rho^*)}{\pi}\left\{\frac{\pi}{2} - \text{sign}(\rho^*)\tan^{-1}(Y_{t-1}-\rho^* Y_t)\right\} \\ &+ \frac{1-\beta(Y_{t-1},\rho^*)}{\pi}\left\{\tan^{-1}[(1-|\rho^*|)Y_t] + \frac{\pi}{2}\right\}, \end{aligned}$$

where

$$\begin{aligned} \alpha(Y_{t-1},\rho^*) &= \frac{\rho^*(1-|\rho^*|)^2 Y_{t-1}}{(1-2|\rho^*|)^2 + (1-|\rho^*|)^2 Y_{t-1}^2}, \\ \beta(Y_{t-1},\rho^*) &= \frac{|\rho^*|\{(1-|\rho^*|)^2 Y_{t-1}^2 - (1-2|\rho^*|)\}}{(1-2|\rho^*|)^2 + (1-|\rho^*|)^2 Y_{t-1}^2}. \end{aligned}$$

Examples of conditional cdf are displayed in Figure 6, for $\rho = \pm 0.9$ and different values of Y_{t-1}. For the same values, examples of functions $\Phi^{-1}[F(\cdot \mid Y_{t-1})]$ are displayed in Figure 7.

C. Causal strong nonlinear AR representation

We consider the nonlinear (or generalized) causal innovations of the process [see e.g. Rosenblatt (2000), Corollary 5.4.2, and Gouriéroux and Jasiak (2005)]. For a Markov process, the nonlinear innovations are defined in a unique way if they are standard Gaussian, i.i.d., and in an increasing relationship with Y_t conditional on Y_{t-1}.

Denote by $F(\cdot|y)$ the conditional cumulative distribution function (cdf) of Y_t given $Y_{t-1} = y$. The Gaussian nonlinear innovations are defined by:

$$v_t = \Phi^{-1}[F(Y_t \mid Y_{t-1})], \tag{C.1}$$

where Φ is the cdf of the standard normal. Processes (Y_t) and (v_t) generate the same information set at any date t. Relationship (C.1) can be inverted to derive the causal

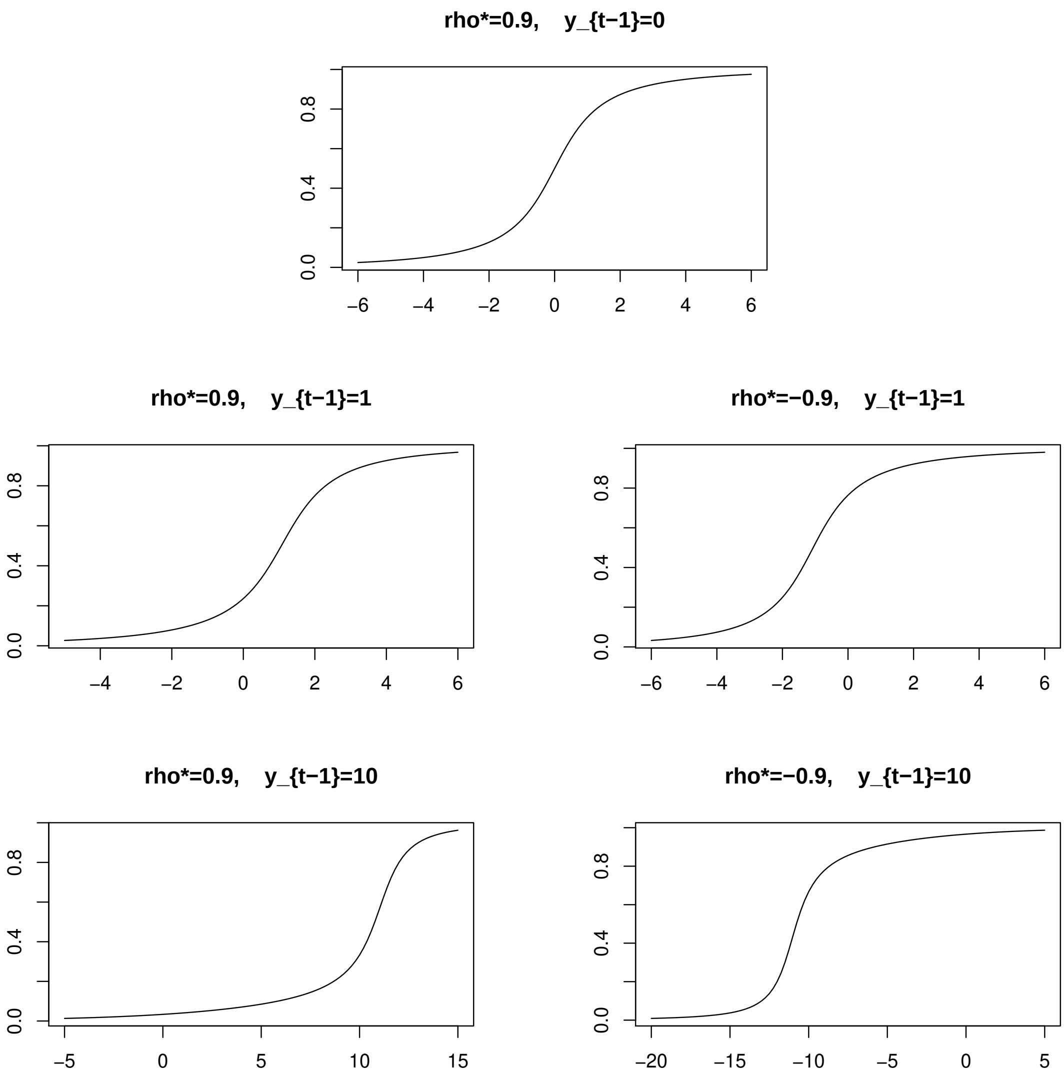

Figure 6. Examples of conditional cdf $F(\cdot \mid Y_{t-1})$ of Proposition B.1, for different values of (ρ, Y_{t-1}).

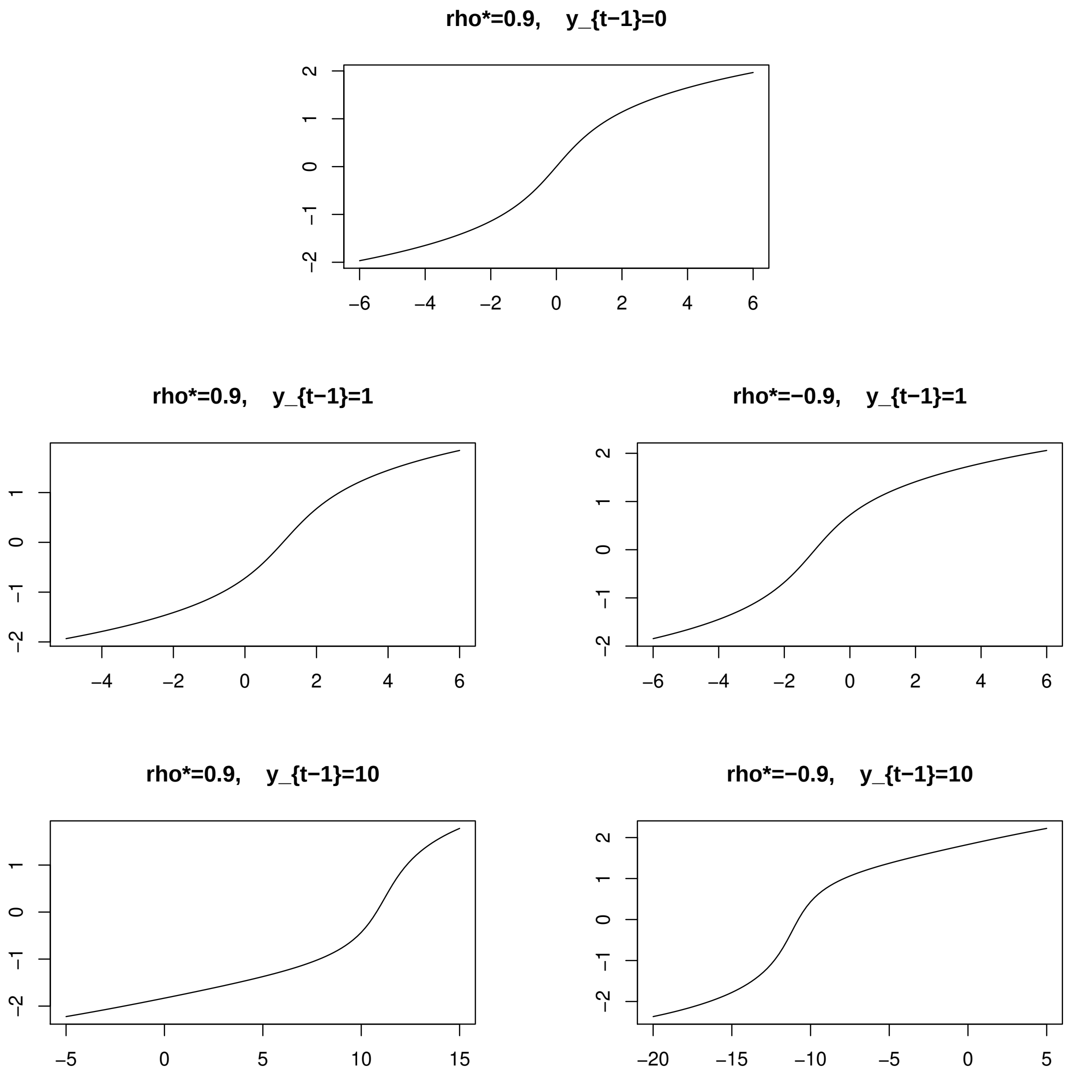

Figure 7. Examples of functions $\Phi^{-1}[F(\cdot \mid Y_{t-1})]$ of Proposition B.1, for different values of (ρ, Y_{t-1}).

strong nonlinear AR representation of process (Y_t):

$$\begin{aligned} Y_t &= G(Y_{t-1}, v_t), \qquad (v_t) \overset{i.i.d.}{\sim} \mathcal{N}(0,1), &\text{(C.2)}\\ \text{with } \; G(Y_{t-1}, .) &= F^{-1}[\Phi(.) \mid Y_{t-1}]. &\text{(C.3)}\end{aligned}$$

Therefore, process (Y_t) admits in reverse time a strong linear AR representation, (3.2), in which the backward innovation ε_t is independent of the future $Y_{t+1}, Y_{t+2}, \ldots$, and a causal strong nonlinear representation given by (C.2)-(C.3). The forward strong nonlinear innovation v_t is independent of the past $Y_{t-1}, Y_{t-2}, \ldots$ The causal nonlinear Gaussian innovations can be used to simulate paths of the noncausal AR process, but these paths can also be deduced from the noncausal strong linear representation. In the first case, the simulated path is deduced in direct time from an initial value Y_0; in the second case, it is deduced in reverse time from a terminal value Y_T, say. Note that simulating paths through the nonlinear Gaussian innovations requires knowing explicitly the conditional cdf. The explicit conditional cdf in the Cauchy case was derived in Proposition B.1.

C.1. Simulations of nonlinear Gaussian innovations

For a simulated sequence (Y_t^s) of the noncausal Cauchy AR(1) process, let us consider the associated simulated nonlinear causal Gaussian innovation as $v_t^s = \Phi^{-1}[F(Y_t^s | Y_{t-1}^s)]$, and noncausal Gaussian innovations as:

$$w_t^s = \Phi^{-1}[F_\varepsilon(\varepsilon_t^s)] = \Phi^{-1}[F_\varepsilon(Y_t^s - \rho Y_{t+1}^s)],$$

where $F_\varepsilon(\varepsilon) = \frac{1}{\pi}[\tan^{-1}(\varepsilon) + \frac{\pi}{2}]$ is the cdf of the Cauchy distribution. In view of Equation (3.6), we can also consider, for $\rho \neq 0$, the standardized causal innovations

$$\eta_t^s = \sqrt{|\rho|(1-|\rho|)} \frac{Y_t^s - \text{sign}(\rho) Y_{t-1}^s}{\sqrt{(1-|\rho|)^2 \, (y_{t-1}^s)^2 + \sigma^{*2}}}.$$

We provide in Figure 8 and 9 the associated plots of v_t^s, η_t^s, for $\rho = 0.1, \rho = 0.5, \rho = \pm 0.9$, and w_t^s. While the graphs of the nonlinear Gaussian innovations v_t^s appear similar to simulations of independent standard Gaussian variables, the graphs of the standardized causal innovations η_t^s display much more extreme values (of both signs). We report in Table 7 some descriptive statistics for the two series, in the case $\rho = 0.9$, showing that the distribution of the η_t^s's is asymmetric and strongly leptokurtic. Results not reported here obtained for a larger sample size ($n = 20000$) show that the empirical variance of η_t^s converges to 1, as expected. The empirical autocorrelation functions of $v_t^s, \eta_t^s, (v_t^s)^2, (\eta_t^s)^2$,

	mean	stand. dev.	skewness	exc. kurtosis
v_t^s	0.026	0.963	0.133	0.094
η_t^s	0.018	0.792	0.392	8.164

Table 7: Descriptive statistics for the nonlinear Gaussian innovations v_t^s and the standardized causal innovations η_t^s of Figure 8 for $\rho = 0.9$.

displayed in Figure 10 for $\rho = 0.9$, confirm the absence of autocorrelation of the nonlinear Gaussian innovations, the standardized causal innovations and their square. Similar graphs were obtained for the other values of ρ.

C.2. The pattern of impulse response

We follow the definition of shocks by means of the causal innovations of the observable process (Y_t), that is, the approach developed by Sims (1980). The paths are also required when we want to analyze causal impulse responses. The procedure is the following one:

i) From any simulated path v_t^s, t varying, of the causal nonlinear Gaussian noise, we deduce a simulated path (Y_t^s) of the process by applying recursively the formula: $Y_t^s = G(Y_{t-1}^s, v_t^s), t = 1, 2 \ldots,$ with some initial value Y_0^s.

ii) The effect of a transitory shock of magnitude δ at time τ is deduced by computing recursively: for $t = 1, 2 \ldots,$

$$Y_t^s(\delta) = G\{Y_{t-1}^s(\delta), v_t^s(\delta)\}, \quad \text{with} \quad v_t^s(\delta) = \begin{cases} v_t^s, & t \neq \tau, \\ v_\tau^s + \delta & t = \tau. \end{cases}$$

The shocks are introduced by means of the causal Gaussian nonlinear innovations (v_t^s).

We provide in Figure 11 the impulse response function corresponding to different magnitudes δ of the transitory shock for one simulation. We observe a transitory effect of the shock when it occurs at a standard period, and a more persistent effect when the shock occurs at the beginning of an explosive bubble. The procedure could be replicated to evaluate the uncertainty in the effect of shocks.

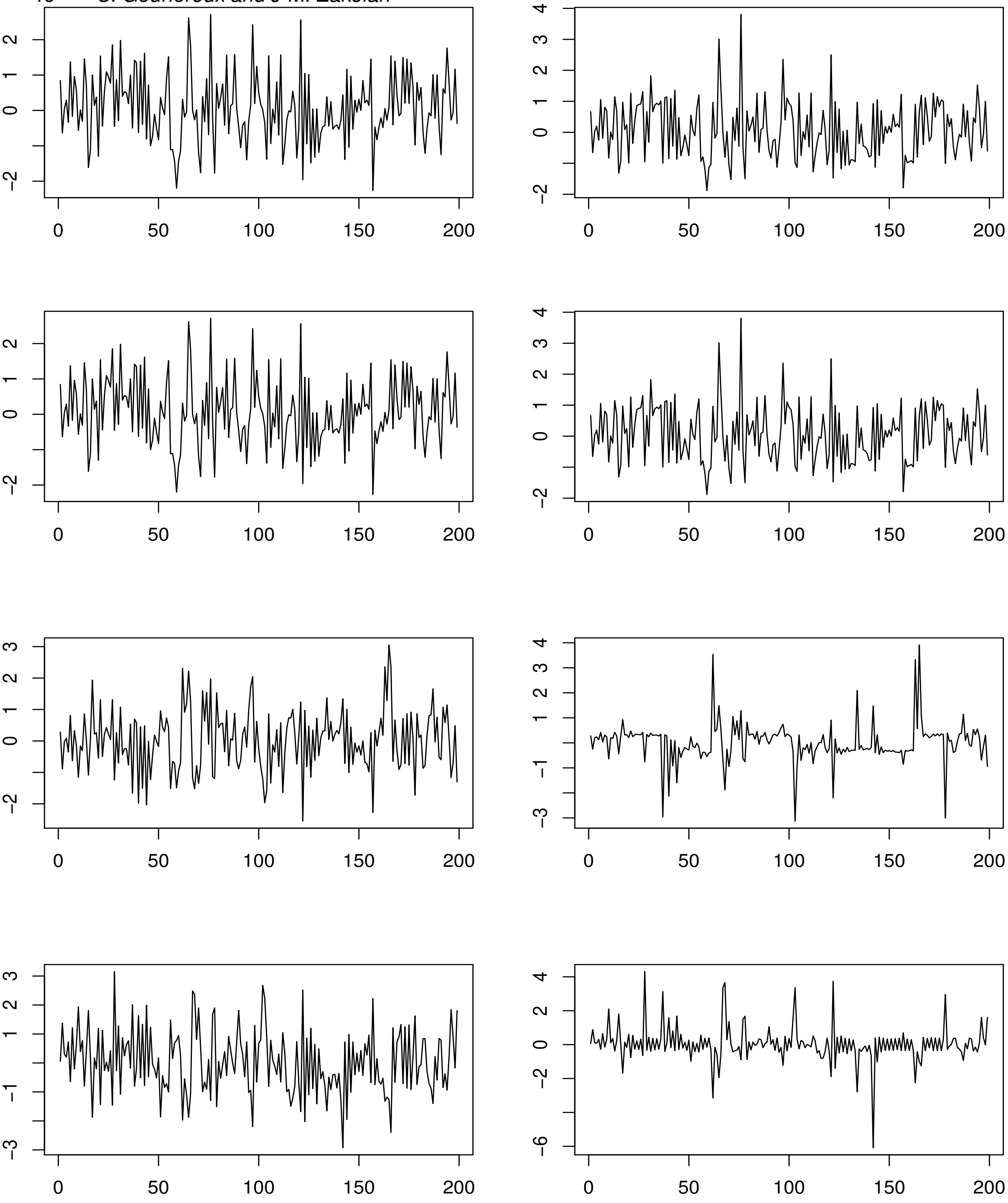

Figure 8. Nonlinear Gaussian innovations $v_t^s = \Phi^{-1}[F(Y_t^s|Y_{t-1}^s)]$ (left panels) and standardized causal innovations η_t^s of equation (3.6) (right panels) for the simulations of Figure 3 ($\rho = 0.1, 0.5, 0.9, -0.9$ from up to bottom).

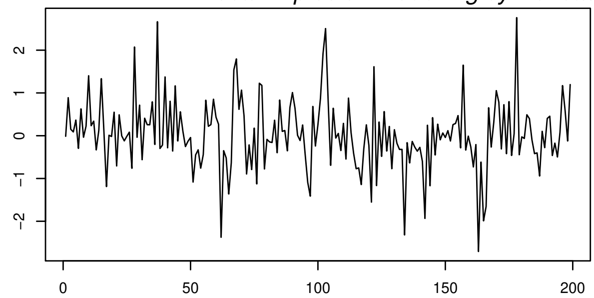

Figure 9. Noncausal innovations $w_t^s = \Phi^{-1}[F_\varepsilon(\varepsilon_t^s)]$ for the simulations of Figure 3.

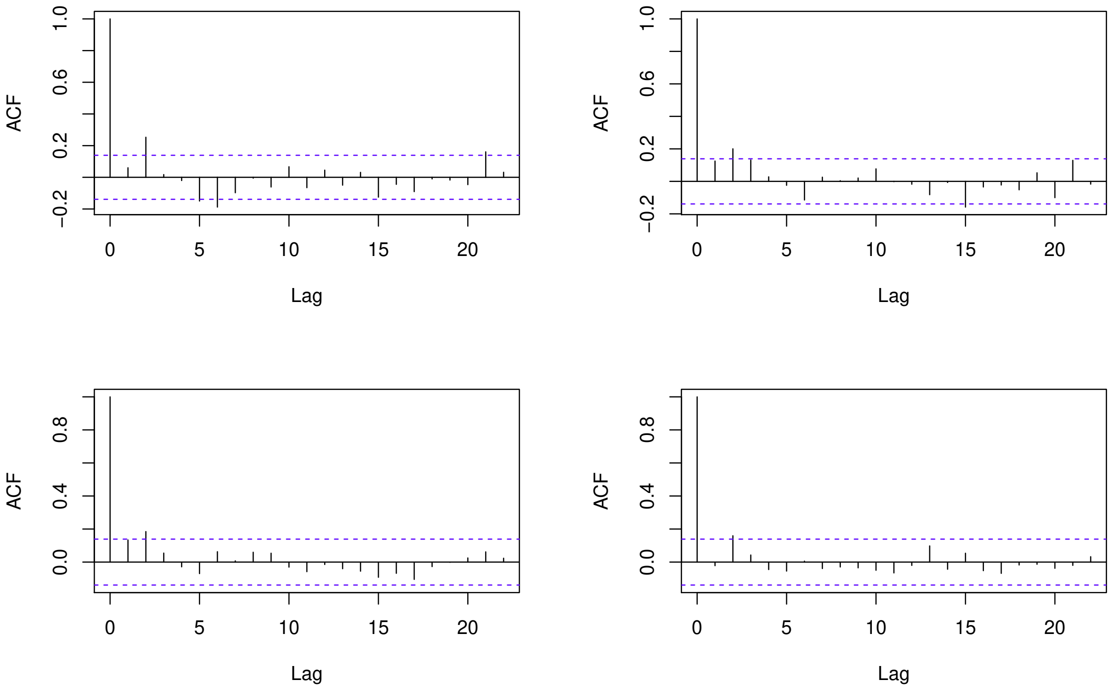

Figure 10. Empirical autocorrelation functions of $v_t^s, \eta_t^s, (v_t^s)^2, (\eta_t^s)^2$ (from left to right and up to bottom) for the simulations of Figure 8 with $\rho = 0.9$).

References

Andrews, B., Calder, M., and R., Davis (2009): Maximum Likelihood Estimation for α-Stable Autoregressive Process. *Ann. Statist.*, 37, 1946–1982.

Bierens, H. (1982) Consistent Model Specification Tests. *Journal of Econometrics*, 20, 105–134.

Breidt, F.J. and R.A., Davis (1992) Time-Reversibility, Identifiability and Independence of Innovations for Stationary Time-Series. *J. Time Series Anal.* 13, 377–390.

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods*, Springer-Verlag, New York, 2nd edition.

Busetti, F., and A.M. Taylor (2004): Tests of Stationarity Against a Change in Persistence. *J. Econometrics,* 123, 33–66.

Cambanis,S., and I., Fakhre-Zakeri (1994): On Predictions of Heavy Tailed Autoregressive Sequences: Forward versus Reversed Time. *Theory Prob. Appl.* 39, 217-233.

Charemza, W., and D. Deadman (1995): Speculative Bubble with Stochastic Explosive Roots: The Failure of Unit Root Testing. *J. Empir. Finance*, 2, 153–163.

Chen, B., Choi, J., and J. C. Escanciano (2012): Testing for Fundamental Vector Moving Average Representations. Discussion Paper, Indiana University.

Cheng, Q. (1992): On the Unique Representation of Non-Gaussian Linear Processes. *Ann. Statist.*, 20, 1143–1145.

Cooper, G. (2008): *The Origin of Financial Crises: Central Banks, Credit Bubbles and the Efficient Market Fallacy.* Vintage Books, New-York.

Davis, R. and S. Resnick (1985): Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities. *Ann. Probab.*, 13, 179–195.

Davis, R. and S. Resnick (1986): Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages. *Ann. Statist.*, 14, 533–558.

Davis, R. and L. Song (2012): Noncausal Vector AR Process with Application to Economic Time Series. Discussion Paper, Columbia University.

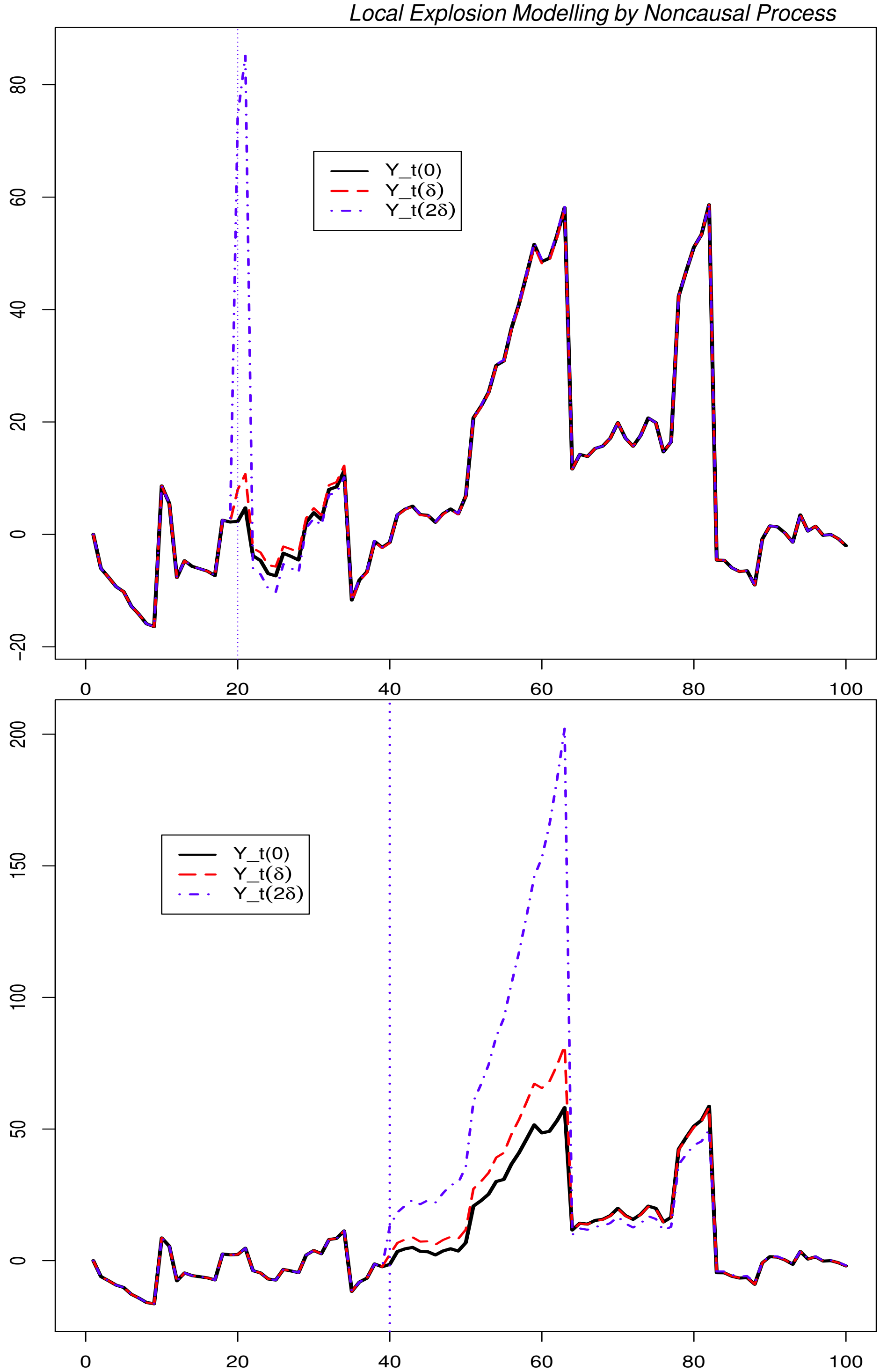

Figure 11. For Model (3.2) with $\alpha = 1, \beta = 0$, $\rho = 0.9$, simulated sample paths of Y_t, $Y_t(\delta)$ and $Y_t(2\delta)$, with shock $\delta = 2$ at time $\tau = 20$ (top panel), with shock $\delta = 1$ at time $\tau = 40$ (bottom panel).

Diba, B. and H. Grossman (1988): Explosive Rational Bubbles in Stock Prices? *The American Economic Review*, 78, 520–530.

Embrechts, P., Klüppelberg, C. and T. Mikosch (1997) *Modelling extremal events.* Springer.

Evans, G. (1991): Pitfalls in Testing for Explosive Bubbles in Asset Prices. *Am. Econ. Rev.*, 81, 922–930.

Flood, R., and P., Garber (1980): Market Fundamentals Versus Price Level Bubbles. The First Tests. *Journal of Political Economy*, 88, 745–770.

Francq, C. and J-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

Gouriéroux, C., and A., Hencic (2015): Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange rate. In *Econometrics of Risk*, Edited by V-N Huynh, V. Kreinovich, S. Sriboonchitta, K. Suriya, Springer.

Gouriéroux, C., and J., Jasiak (2005): Nonlinear Impulse Response Function. *Annales d'Economie et de Statistique*, 78, 1–33.

Gouriéroux, C., and J., Jasiak (2015): Filtering, Prediction and Simulation Methods for Noncausal Processes. *J. Time Series Anal.*, forthcoming.

Gouriéroux, C., and J-M., Zakoïan (2015): On Uniqueness of Moving Average Representations of Heavy-Tailed Stationary Processes. *J. Time Series Anal.*, 36, 876–887.

Heathcote, C.R. (1977) The integrated squared error estimation of parameters. *Biometrika*, 64, 255–264.

Homm, U. and J. Breitung (2012): Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods. *J. Financ. Econometrics*, 10, 198–231.

Krengel, U. (1985) *Ergodic theorems.* De Gruyter, Berlin-New York.

Lanne, M., Nyberg, H., and E., Saarinen (2012): Does Noncausality Help in Forecasting Economic Time Series? *Economics Bulletin*, 32, 2849–2859.

Lanne, M., and P., Saikkonen (2011): NonCausal Autogressions for Economic Time Series. *J. Time Ser. Econom.*, 3, Article 2.

Lin, J-W. and A.I. McLeod (2008) Portmanteau Tests for ARMA Models with Infinite Variance. *J. Time Series Anal.*, 29, 600–617.

Ling, S. and D. Li (2008): Asymptotic Inference For a Nonstationary Double AR(1) Model. *Biometrika*, 95, 257–263.

Paulson, A.S., Holcomb, E.W. and R.A. Leitch (1975): The Estimation of the Parameters of the Stable Laws. *Biometrika*, 62, 163–170.

Pham, D.T. (1986): The Mixing Property of Bilinear And Generalized Random Coefficient Autoregressive Models. *Stochastic Process. Appl.,* 20, 295–306.

Phillips, P.C.B. (1987): Towards a Unified Asymptotic Theory for Autoregression. *Biometrika,* 74, 535–547.

Phillips, P.C.B., Shi, S. and J. Yu (2015): Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P500. *Int. Econ. Rev.*, 56, 1043-1077.

Phillips, P.C.B., Wu, Y. and J. Yu (2011): Explosive Behavior in the 1990's Nasdaq: When Did Exuberance Escalate Asset Values? *Int. Econ. Rev.*, 52, 201–226.

Ramsey, J., and P., Rothman (1996): Time Irreversibility and Business Cycle Asymmetry. *Journal of Money and Banking*, 28, 1–21.

Rosenblatt, M. (2000): *Gaussian and Non-Gaussian Linear Time Series and Random Fields*, Springer Verlag, New-York.

Samorodnitsky G. and M. S. and Taqqu (1994): *Stable Non-Gaussian Random Processes*, Chapman & Hall, London.

Taufer, E. and N. Leonenko (2009): Characteristic Function Estimation of Non-Gaussian Ornstein-Uhlenbeck Processes. *J. Statist. Plann. Inference*, 139, 3050-3063.

Sims, C. (1980): Macroeconomics and Reality. *Econometrica*, 1–48.

Yu, J. (2004): Empirical Characteristic Function Estimation and Its Applications. *Economet. Rev.*, 23, 93–123.