Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles

Fries, Sébastien and Zakoian, Jean-Michel

CREST, Paris-Saclay University

September 2017

Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles

Sébastien Fries* and Jean-Michel Zakoïan†

Abstract

Noncausal autoregressive models with heavy-tailed errors generate locally explosive processes and therefore provide a natural framework for modelling bubbles in economic and financial time series. We investigate the probability properties of mixed causal-noncausal autoregressive processes, assuming the errors follow a stable non-Gaussian distribution. We show that the tails of the conditional distribution are lighter than those of the errors, and we emphasize the presence of ARCH effects and unit roots in a causal representation of the process. Under the assumption that the errors belong to the domain of attraction of a stable distribution, we show that a weak AR causal representation of the process can be consistently estimated by classical least-squares. We derive a Monte Carlo Portmanteau test to check the validity of the weak AR representation and propose a method based on extreme residuals clustering to determine whether the AR generating process is causal, noncausal or mixed. An empirical study on simulated and real data illustrates the potential usefulness of the results.

Keywords: Noncausal process, Stable process, Extreme clustering, Explosive bubble, Portmanteau test.

*CREST and Paris-Saclay University. E-Mail: sebastien.fries@gmail.com

†Jean-Michel Zakoïan, CREST and University Lille 3. Address: CREST, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France. E-mail: zakoian@ensae.fr.

1 Introduction

In the analysis of prices of financial assets such as stocks, it is common to observe phases of locally explosive behaviours, together with heavy-tailed marginal distributions and volatility clustering. Such features seem incompatible with classical *linear* models (namely the class of autoregressive-moving average (ARMA) models) which rely on the second-order properties of a time series. On the other hand, nonlinear models such as ARCH or stochastic volatility models are designed to capture volatility clustering, not to produce locally explosive sample paths mimicking bubbles in financial markets. However, the dynamic limitations of ARMA models are reduced if noncausal components (i.e. AR or MA polynomials with roots inside the unit disk) are introduced. For instance, all-pass models[1] are linear time series with nonlinear behaviours, in particular ARCH effects [see Breidt, Davis and Trindade (2001) and the references therein]. More recently, Gouriéroux and Zakoian (2017, GZ hereafter) showed that a simple noncausal AR(1) process with heavy-tailed errors is able to produce the typical nonlinear behaviours observed for the prices of financial assets.

Noncausal processes or random fields have been thoroughly studied in the statistical literature [Rosenblatt (2000), Andrews, Calder and Davis (2009)], and have been applied in various areas, including deconvolution of seismic signals [Wiggins (1978), Donoho (1981), Hsueh and Mendel (1985)], and analysis of astronomical data [Scargle (1981)]. Recent years have witnessed the emergence of a significant line of research on noncausal models in the econometric literature [see e.g., Lanne, Nyberg and Saarinen (2012), Lanne, Saikkonen (2011), Davis and Song (2012), Chen, Choi and Escanciano (2012), Hencic and Gouriéroux (2015), Velasco and Lobato (2015), Hecq, Lieb and Telg (2016a, 2016b, 2017), Cavaliere, Nielsen and Rahbek (2017)]. The distinction between causal and noncausal processes is only meaningful in a non-Gaussian framework, and the increasing interest in Mixed causal-noncausal AR processes (MAR) parallels the widespread use of non-Gaussian heavy-tailed processes in economic or financial applications. Besides, the introduction of noncausal components in univariate or multivariate time series models has received pertinent economic justifications (see Gouriéroux, Jasiak and Monfort (2016)).

One important reason for introducing noncausal components in AR processes is to provide a mechanism for generating financial bubbles. GZ showed that the sample paths of a stationary noncausal AR(1) process with heavy-tailed errors may have locally explosive phases. Other recent researches have focused on data generating processes that are able to produce explosive behaviours

[1] All-pass are ARMA models in which all roots of the AR polynomial are reciprocal of the roots of the MA polynomial.

and model bubbles in financial markets. For example Phillips , Wu and Yu (2011), Phillips , Shi and Yu (2015) and more recently, in a continuous time framework, Chen, Phillips and Yu (2017) investigated mildly explosive processes. Apart from the generation of bubbles, noncausal AR(1) processes with stable distributed errors exhibit surprising features such as a predictive distribution with lighter tails than the marginal distribution, a martingale property in the causal representation when the errors follow a Cauchy distribution, or the presence of GARCH effects. It is of interest to know whether these structural properties extend to higher-order models. Indeed, first-order models are clearly not sufficient to capture complex behaviours of economic series, such as the occurrence of locally explosive behaviours with different rates of explosion, or different types of asymmetries in the growth and downturn phases of the bubbles.

The aim of this paper is to analyze the class of mixed causal-noncausal AR processes with heavy-tailed errors. The probability structure is studied under the assumption that the errors follow stable non-Gaussian distributions. Properties of the Least-Squares (LS) estimator are derived under the less stringent assumption that the noise distribution is in the domain of attraction of a non-Gaussian stable law. The paper is organised as follows. Section 2 studies the sample paths and the marginal distribution of MAR processes with stable errors. Sections 3 analyzes the conditional distributions through conditional moments. Conditional heteroscedasticity effects are depicted and causal representations are exhibited. Section 4 derives the asymptotic properties of LS estimator, deduces a Portmanteau test, and studies identification of the strong representation based on the analysis of extreme residuals clustering. Section 5 and 6 propose numerical illustrations based on simulated and real data, respectively. Section 7 concludes. Proofs and complementary results are collected in two Appendixes.

2 Stable MAR(p, q) processes

A MAR(p, q) process (X_t) is the strictly stationary solution of the difference equation

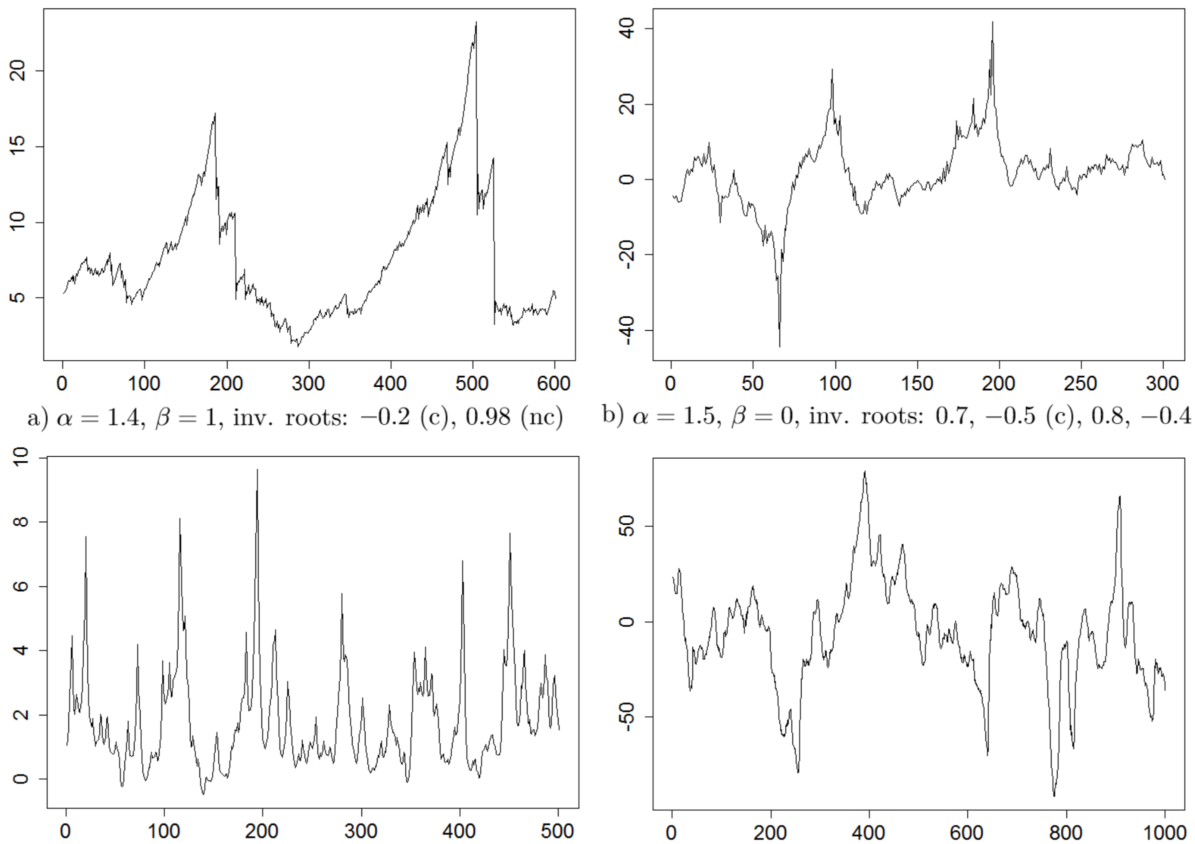
$$\psi(F)\phi(B)X_t = \varepsilon_t, \tag{2.1}$$

where B and F are the usual lag and forward operators ($B^k X_t = X_{t-k}, F^k X_t = X_{t+k}, k \in \mathbb{Z}$), (ε_t) is an independent and identically distributed (i.i.d.) sequence, $\psi(z) = 1 - \sum_{i=1}^{p} \psi_i z^i$ and $\phi(z) = 1 - \sum_{i=1}^{q} \phi_i z^i$ are real polynomials of degrees p and q respectively (i.e. $\psi_p \neq 0$ and $\phi_q \neq 0$), with all roots outside the unit circle. When $q = 0$ (resp. $p = 0$), the model is called purely noncausal (resp. causal).

We assume that the errors ε_t follow a stable non-Gaussian distribution but the assumption will be relaxed for the statistical inference. The generality and convenience of this class of distributions is now well established.[2] Stable laws are easily characterised through their characteristic function: ε_t is said to follow a stable distribution with parameters $\alpha \in]0,2[$, $\beta \in [-1,1]$, $\sigma > 0$, $\mu \in \mathbb{R}$, denoted $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, if

$$\forall s \in \mathbb{R}, \quad \mathbb{E}(e^{is\varepsilon_t}) = \exp\Big\{ -\sigma^{\alpha}|s|^{\alpha}\left(1 - i\,\beta\,\mathrm{sign}(s)w(\alpha, s)\right) + is\mu\Big\},$$

where $w(\alpha, s) = \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, and $w(1, s) = -\frac{2}{\pi}\ln|s|$, otherwise. Recall that a stable random variable X has regularly varying tails in the sense that $\mathbb{P}(X < -x) \sim c_\alpha(1-\beta)x^{-\alpha}$ and $\mathbb{P}(X > x) \sim c_\alpha(1+\beta)x^{-\alpha}$ as $x \to +\infty$, with $c_\alpha > 0$ and $\beta \in (-1, 1)$.



a) $\alpha = 1.4$, $\beta = 1$, inv. roots: -0.2 (c), 0.98 (nc)

b) $\alpha = 1.5$, $\beta = 0$, inv. roots: 0.7, -0.5 (c), 0.8, -0.4 (nc)

c) $\alpha = 1.3$, $\beta = 1$, inv. roots: 0.65 (c), 0.7 (nc)

d) $\alpha = 1.6$, $\beta = 0$, inv. roots: -0.8, 0.7 (c), 0.95, 0.5 (nc)

Figure 1: Examples of trajectories of MAR(1,1) (left panel) and MAR(2,2) (right panel) processes with different parameters (nc: inverse of noncausal roots; c: inverse of causal root).

[2] See for instance Embrechts, Klüppelberg, and Mikosch (1997), Samorodnitsky and Taqqu (1994) for the main properties of stable distributions.

2.1 Sample paths

Examples of trajectories of four noncausal MAR processes are displayed in Figure 1. It can be seen that the trajectories feature locally explosive trends which are suited for the modelling of bubbles and positive feedback loop phenomena. Bubbles can be trending either upward or downward depending on the value of β. When $\beta = 1$, the density of the errors is maximally skewed towards positive values, yielding trajectories like (a) and (c) which could be suited to model prices or volatilities. In particular, trajectory (a) displays bubble patterns similar to those of real prices (see for instance Figure 5 below). The influence of a smaller tail parameter α is visible when comparing trajectories (c) and (d): the extreme events of the former ($\alpha = 1.3$) are more recurrent and further away from the central values than those of the latter ($\alpha = 1.6$).

Under the assumptions made on the AR polynomial, (X_t) admits an MA(∞) representation[3]

$$X_t = \sum_{k=-\infty}^{+\infty} d_k \varepsilon_{t+k}. \tag{2.2}$$

A simple index change $X_t = \sum_{\tau \in \mathbb{Z}} \varepsilon_\tau d_{\tau-t}$ allows to interpret the sample path of X_t as a linear combination of *baseline paths*, $t \longmapsto d_{\tau-t}$, weighted by stochastic i.i.d. coefficients ε_τ. Figure 2 depicts such baseline paths for four different MAR processes. The first panel illustrates the well-known impulse response function of a classical causal AR(1). The second panel displays an explosive exponential trend followed by a downward, faster decay and corresponds to the baseline path of a MAR(1,1) process. The remaining panels show more complex trajectories: the third one depicts the baseline path of a MAR(2,2) with dented upward and downward trends whereas the last one, corresponding to a noncausal AR(4) with two real and two conjugated complex roots, shows an upward trend with oscillations of increasing amplitudes and fixed pseudo-periods.

2.2 Marginal distribution

Our first result characterises the marginal distribution of the stable MAR(p, q).

Proposition 2.1 *The MAR(p, q) solution* (X_t) *of Model* (2.1) *with* $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$*, has a stable*

[3] It follows from Proposition 13.3.1 in Brockwell and Davis (1991) that the infinite sum in (2.2) is well defined under the stable law assumption, ensuring the existence of $\mathbb{E}|\varepsilon_t|^s$ for $s < \alpha$.

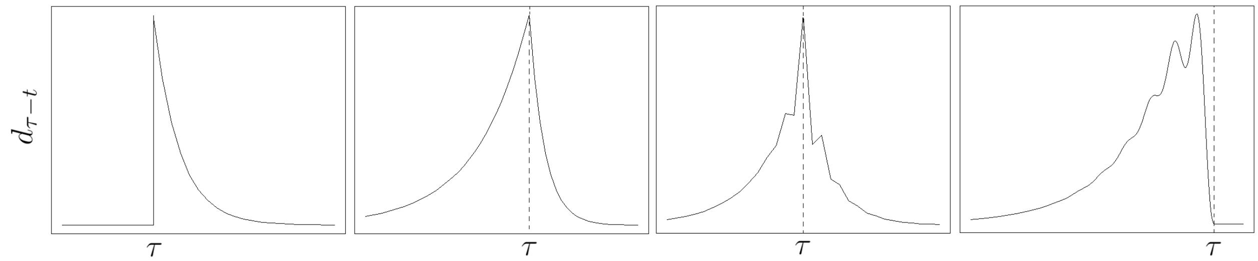


Figure 2: Four examples of baseline paths $t \longmapsto d_{\tau-t}$ of MAR processes with characteristic polynomials, from left to right: $1-0.7B$; $(1-0.7F)(1-0.9B)$; $(1-0.8F)(1+0.4F)(1-0.7B)(1+0.5B)$; $(1-0.99F)(1-965F)(1-0.98e^{i0.045\pi}F)(1-0.98e^{-i0.045\pi}F)$.

stationary distribution, $X_t \sim \mathcal{S}(\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}, \tilde{\mu})$ *where*

$$\tilde{\alpha} = \alpha, \qquad \tilde{\beta} = \beta \frac{\sum_{k=-\infty}^{+\infty} |d_k|^\alpha sign(d_k)}{\sum_{k=-\infty}^{+\infty} |d_k|^\alpha},$$

$$\tilde{\sigma} = \sigma \left(\sum_{k=-\infty}^{+\infty} |d_k|^\alpha \right)^{\frac{1}{\alpha}}, \qquad \tilde{\mu} = \frac{\mu}{\phi(1)\psi(1)} - \mathbb{1}_{\{\alpha=1\}} \frac{2}{\pi} \beta\sigma \sum_{k=-\infty}^{+\infty} d_k \ln |d_k|.$$

It is worth noting that the tail index α of X_t is that of the error term. In particular, $\mathbb{E}|X_t|^s < +\infty$ for $s < \alpha$ and $\mathbb{E}|X_t|^\alpha = +\infty$.

Example 2.1 (MAR$(1,1)$ process) Let the model

$$(1-\psi F)(1-\phi B)X_t = \varepsilon_t, \quad \text{with} \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu), \tag{2.3}$$

with $|\phi| < 1$ and $|\psi| < 1$. We have $d_k = \dfrac{\psi^k}{1-\phi\psi}$, for any $k \geq 0$, and $d_k = \dfrac{\phi^{-k}}{1-\phi\psi}$, for any $k \leq 0$. Then $X_t \sim \mathcal{S}\left(\alpha, \tilde{\beta}, \tilde{\sigma}, \tilde{\mu}\right)$ with

$$\tilde{\beta} = \beta \left(\frac{1-\text{sign}(\phi)|\phi\psi|^\alpha}{1-|\phi\psi|^\alpha} \right) \left(\frac{1-\text{sign}(\psi)|\psi|^\alpha}{1-|\psi|^\alpha} \right) \left(\frac{1-\text{sign}(\phi)|\phi|^\alpha}{1-|\phi|^\alpha} \right),$$

$$\tilde{\sigma} = \frac{\sigma}{1-\phi\psi} \left(\frac{1-|\phi\psi|^\alpha}{(1-|\psi|^\alpha)(1-|\phi|^\alpha)} \right)^{\frac{1}{\alpha}},$$

$$\tilde{\mu} = \frac{\mu}{(1-\psi)(1-\phi)} - \mathbb{1}_{\{\alpha=1\}} \frac{2\beta\sigma}{\pi(1-\phi\psi)} \left[\frac{\psi \ln|\psi|}{(1-\psi)^2} + \frac{\phi \ln|\phi|}{(1-\phi)^2} - \frac{(1-\phi\psi)\ln|1-\phi\psi|}{(1-\psi)(1-\phi)} \right].$$

In particular, when $\psi > 0$, $\phi > 0$ and the errors are Cauchy distributed, that is when $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1, 0, \sigma, 0)$, then the above formulae simplify and $X_t \sim \mathcal{S}\left(1, 0, \dfrac{\sigma}{(1-\psi)(1-\phi)}, 0\right)$.

3 Predictive distributions

In the presence of a noncausal component in the AR polynomial, the predictive density of a future observation given a sample of consecutive observations is generally not available in closed form.

We will derive in this section some features of such predictive distributions. We start by showing that the Markov property holds whatever the errors distribution.

Proposition 3.1 *The MAR(p, q) process (X_t) is an homogeneous Markov chain of order $p+q$.*

3.1 Existence of moments of the conditional distribution

It follows from Proposition 2.1 that $\mathbb{E}|X_t|^s = \infty$ for $s \geq \alpha$. The next result shows a different behaviour for the conditional moments.

Theorem 3.1 *If (X_t) is the MAR(p, q) solution of Model (2.1) with $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, we have*

$$\mathbb{E}\left[\, |X_t|^\gamma |\, X_{t-1}, X_{t-2}, \ldots\right] < \infty, \quad a.s., \quad whenever \quad 0 < \gamma < 2\alpha + 1.$$

The conditional distribution with respect to past observations has lighter tails than the marginal[4] which only admits moments up to the order α. In particular, it follows that, whatever the heaviness of the tails of ε_t, the conditional expectation of X_t always exists. Furthermore, it is sufficient to have $\alpha \geq 1/2$ for X_t to have a conditional variance.

3.2 Prediction of future values for the MAR$(1, q)$ processes.

Prediction at any horizon can be fully characterised for the symmetric MAR$(1, q)$ process $(1 - \psi F)\phi(B)X_t = \varepsilon_t$. Let $\mathscr{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ the canonical filtration of process (X_t).

Proposition 3.2 *When $\beta = 0$, $p = 1$ and $q \geq 0$, there exists for any $h \geq 0$ a polynomial $\mathscr{P}_h$ of degree q such that*

$$\mathbb{E}\left[\, X_{t+h} |\, \mathscr{F}_{t-1}\right] = \mathscr{P}_h(B)X_{t-1}.$$

For $h = 0$, the above formula holds with

$$\mathscr{P}_0(B)X_{t-1} = \psi^{<\alpha-1>}X_{t-1} + (1 - \psi^{<\alpha-1>}B)(\phi_1 X_{t-1} + \ldots + \phi_q X_{t-q}),$$

where for any $x \neq 0$ and $r \in \mathbb{R}$, $x^{<r>} = sign(x)|x|^r$.

Details on $\mathscr{P}_h$ are provided in the proof. It is worth noting that the conditional expectation is linear in the past and can be explicitly computed. Proposition 3.2 for $h = 0$ yields a semi-strong causal representation of X_t.

[4] Various studies provided estimates of the tail index of marginal distributions of financial data. For instance, Mandelbrot (1963) found evidence by graphical methods that $\alpha = 1.7$ for variations of cotton prices, whereas Francq and Zakoian (2014), with a quasi marginal maximum likelihood estimator approach assuming stable distributions, estimated tail parameters between 1.5 and 1.72 for nine financial indexes (i.a. SP500, Nikkei and CAC).

Corollary 3.1 *Under the assumptions of Proposition 3.2, there exists a sequence* (η_t) *of random variables such that for any* $t \in \mathbb{Z}$,

$$(1 - \psi^{<\alpha-1>} B)\phi(B) X_t = \eta_t, \tag{3.1}$$

with $\mathbb{E}\left[\eta_t | \mathscr{F}_{t-1}\right] = 0$.

By comparison with finite variance AR processes, this result is surprising. Indeed, in the L^2 framework, if (X_t) is mixed causal-noncausal satisfying $\psi(F)\phi(B)X_t = \varepsilon_t$, then there exists a causal version of (X_t) given by $\psi(B)\phi(B)X_t = Z_t$, where (Z_t) is uncorrelated with zero mean and finite variance. This representation is obtained by inverting the ill-located roots and leaving the well-located ones unchanged. In our framework, the ill-located root ψ, with $|\psi| < 1$, is transformed into $\psi^{<\alpha-1>}$ which can either be inside, outside, as well as on the unit circle. In the Cauchy case, $(\alpha = 1)$ we get, when $\psi > 0$,

$$\mathbb{E}\Big[X_t \Big| \mathscr{F}_{t-1}\Big] = X_{t-1} + (1 - B)(\phi_1 X_{t-1} + \ldots + \phi_q X_{t-q}), \tag{3.2}$$

with by convention $\phi_1 = \ldots = \phi_q = 0$ when $q = 0$. Hence, the martingale property established by GZ (Proposition 3.3), $\mathbb{E}[X_t | \mathscr{F}_{t-1}] = X_{t-1}$, only holds for the noncausal AR(1) (i.e. when $q = 0$). The asymptotic behaviour of the conditional expectation -when the horizon h tends to infinity- is highly dependent on the tail index α. Proposition 3.2 allows us to distinguish different behaviours summarised in the following Corollary.

Corollary 3.2 *Under the assumptions of Proposition 3.2, we have almost surely*

$$\Big|\mathbb{E}\left[X_{t+h} | \mathscr{F}_{t-1}\right]\Big| \underset{h\to+\infty}{\longrightarrow} \begin{cases} 0 & if \quad \alpha \in (1,2), \\ \ell_{t-1} & if \quad \alpha = 1, \end{cases}$$

where ℓ_{t-1} *is an* $\mathscr{F}_{t-1}$*-measurable random variable. Moreover, when* $\alpha \in (0,1)$ *and* $q = 1$,

$$\Big|\mathbb{E}\left[X_{t+h} | \mathscr{F}_{t-1}\right]\Big| \underset{h\to+\infty}{\longrightarrow} +\infty.$$

If $\alpha \in (1,2)$, that is for lighter tails within the stable family, the conditional expectation always tends to 0 which is the unconditional expectation. This is consistent with the L^2 framework (Brockwell, Davis (1991), p.189). For $\alpha = 1$, the absolute value of the conditional expectation tends to a finite limit whereas the unconditional expectation does not exist. The general case when $\alpha \in (0,1)$ is more intricate and is detailed in Appendix B.

Example 3.1 (Explicit prediction formula for the MAR(1,1) process) Let (X_t) the strictly stationary solution of Model (2.3). Proposition 3.2 yields for any $h \geq 0$,

$$\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big] = \phi^{h+1}\left[X_{t-1} + (X_{t-1} - \phi X_{t-2})\sum_{i=1}^{h+1}\left(\psi^{<1-\alpha>}\phi\right)^{-i}\right],$$
$$= \begin{cases} \phi^{h+1}\left[X_{t-1} - \dfrac{1-(\psi^{<1-\alpha>}\phi)^{-(h+1)}}{1-\psi^{<1-\alpha>}\phi}(X_{t-1}-\phi X_{t-2})\right], & \text{if} \quad \psi^{<1-\alpha>}\phi \neq 1, \\ \phi^{h+1}\left[X_{t-1} + (h+1)(X_{t-1} - \phi X_{t-2})\right], & \text{if} \quad \psi^{<1-\alpha>}\phi = 1 \end{cases}$$

Furthermore, Corollary 3.1 yields a semi-strong causal representation of (X_t):

$$(1 - \psi^{<\alpha-1>}B)(1-\phi B)X_t = \eta_t,$$

with $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$. In particular if $\psi > 0$ and $\alpha = 1$, this causal representation simplifies to

$$(1-B)(1-\phi B)X_t = \eta_t.$$

3.3 Unit root property

The equality $\mathbb{E}\Big[X_t\Big|X_{t-1}\Big] = X_{t-1}$ for the noncausal Cauchy AR(1) with positive AR coefficient shows the existence of a unit root. Indeed, we have $X_t = X_{t-1} + \epsilon_t$ where $\mathbb{E}\Big[\epsilon_t\Big|X_{t-1}\Big] = 0$. The next result actually shows that this property extends to more general MAR processes.

Proposition 3.3 *Let (X_t) be an α-stable MAR(p,q) process admitting an MA(∞) representation with a coefficients sequence (d_k). Then, if $\sum_{\ell\in\mathbb{Z}} d_\ell \left(d_{\ell+1}\right)^{<\alpha-1>} = \sum_{\ell\in\mathbb{Z}} |d_{\ell+1}|^\alpha$, we have,*

$$\mathbb{E}\left[X_t \middle| X_{t-1}\right] = X_{t-1}.$$

In particular this property holds in the Cauchy case ($\alpha = 1$) if and only if $d_k \geq 0$ for all $k \in \mathbb{Z}$.

The proof is available in Appendix B. Note that $\mathbb{E}\big[X_t\big|X_{t-1}\big]$ is in general different from $\mathbb{E}\big[X_t\big|\mathscr{F}_{t-1}\big]$ for $p + q > 1$, which explains the seeming contradiction with (3.2).

3.4 Conditional heteroskedasticity of the Cauchy MAR$(1, q)$

All-pass models are well known examples of strong linear models displaying ARCH effects (namely the correlation of the squares). However, such effects are difficult to characterise without an explicit specification of the errors specification. The following result provides an explicit characterization of ARCH effects through the conditional variance of MAR processes with Cauchy innovation.

Proposition 3.4 *Let X_t be a MAR(1, q) process* $(1-\psi F)\phi(B)X_t = \varepsilon_t$ *with* $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1,0,\sigma,0)$. *Then, for any $h \geq 0$, there exists a polynomial $Q_h(z) = \sum_{i=0}^{h} q_{i,h} z^i$ such that*

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \Big((\phi(B)X_{t-1})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\Big)\Big(c_h - \Big(Q_h(sign\ \psi)\Big)^2\Big),$$

with $c_h = \sum_{i=0}^{h}\sum_{j=0}^{h} q_{i,h} q_{j,h} (sign\ \psi)^{i+j} |\psi|^{-\min(i,j)-1}$.

Polynomials $Q_h(z)$ are defined in Lemma D.1 of the Appendix A, and details regarding their coefficients are given in Lemma E.1. The complete proof is available in Appendix B. The causal representation (3.1) can then be completed and reveals quadratic ARCH effects in the Cauchy MAR$(1,q)$ process.

Corollary 3.3 *Under the assumptions of Proposition 3.4, there exists a sequence (η_t) of random variables such that,*

$$(1 - sign(\psi)B)\phi(B)X_t = \sigma_t \eta_t,$$
$$\sigma_t^2 = \left(\frac{1}{|\psi|} - 1\right)(X_{t-1} - \phi_1 X_{t-2} - \ldots - \phi_q X_{t-q-1})^2 + \frac{\sigma^2}{|\psi|(1-|\psi|)}.$$

where $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$, $\mathbb{E}\left[\eta_t^2 \middle| \mathscr{F}_{t-1}\right] = 1$.

The process $e_t = \sigma_t \eta_t$ is however not a ARCH in the strict sense: first because the errors η_t are not i.i.d., and second because the volatility is a function of the X_{t-i} (not of the e_{t-i}). This representation is actually closer to the Double Autoregressive model studied by Ling (2007) (see also Nielsen and Rahbek (2014) for a multivariate extension).

Example 3.2 (Semi-strong representation of the MAR(1,1) process) When $q = 1$ and $\psi > 0$, Corollary 3.3 yields

$$(1-B)(1-\phi B)X_t = \eta_t \sqrt{(\psi^{-1} - 1)(X_{t-1} - \phi X_{t-2})^2 + \frac{\sigma^2}{(1-\psi)^2}},$$

where $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$ and $\mathbb{E}\left[\eta_t^2 \middle| \mathscr{F}_{t-1}\right] = 1$, and the conditional variance at horizon h in Proposition 3.4 takes the more explicit form, for any $h \geq 0$,

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \Big((X_{t-1} - \phi X_{t-2})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\Big)$$
$$\times \Big[\psi^{-1}\phi^{2h} - \frac{1}{\psi(1-\phi\psi)}\Big(2\psi\phi^{h+1}\frac{1-\phi^h}{1-\phi} - (1+\phi\psi)\frac{\psi^{-h} - \phi^{2h}}{1-\phi^2\psi}\Big)\Big].$$

4 Statistical Inference

In this section, we will relax the assumption that (ε_t) is an α-stable sequence but rather assume that the law of ε_t belongs to the domain of attraction of a stable distribution. More specifically we assume that we observe $X_1, \dots, X_n$ generated by the MAR(p, q) model

$$\psi_0(F)\phi_0(B)X_t = \varepsilon_t, \tag{4.1}$$

where $\psi_0(z) = 1 - \sum_{i=1}^{p} \psi_{0i} z^i, \phi_0(z) = 1 - \sum_{i=1}^{p} \phi_{0i} z^i$, with $\psi_0(z) \neq 0$ and $\phi_0(z) \neq 0$ for $|z| \leq 1$, and (ε_t) is an i.i.d. sequence of symmetric (for simplicity) random variables such that:

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha} L(x), \tag{4.2}$$

where L is a slowly varying function at infinity and $0 < \alpha < 2$.[5] Another representation (see Andrews and Davis (2013)) is given by

$$\eta_0(B)X_t = \zeta_t, \qquad \text{where} \quad \eta_0(B) = \psi_0(B)\phi_0(B) = 1 - \sum_{i=1}^{p+q} \eta_{0i} B^i, \tag{4.3}$$

and (ζ_t) is an all-pass process.[6] In the sequel, we call (4.3) the all-pass causal representation of (X_t). Let $\rho(h) = \left(\sum_{k=-\infty}^{\infty} d_k d_{k-h}\right) / \left(\sum_{k=-\infty}^{\infty} d_k^2\right)$ for $h \in \mathbb{Z}$, where the d_k's are the MA(∞) coefficients in (2.2).

Proposition 4.1 *Let (X_t) be the strictly stationary solution of model* (4.1)-(4.2)*. Then, the $\rho(h)$'s satisfy the recursion*

$$\rho(h) = \sum_{i=1}^{p+q} \eta_{0i} \rho(h - i), \quad \forall h > 0, \tag{4.4}$$

where the coefficients η_{0i} are obtained from (4.3).

It is worth noting that, although the autocorrelations of X_t do not exist, the empirical autocorrelations can be computed and converge to the coefficients $\rho(h)$, which satisfy the usual Yule-Walker equations.

[5] that is, $\lim_{x\to\infty} L(tx)/L(x) = 1, \forall t > 0$.

[6] When the second-order moments are finite, all-pass processes are uncorrelated. Andrews and Davis (2013) showed that this property continues to hold "empirically" in the infinite variance case, in the sense that the sample autocorrelations converge to zero as the sample size goes to infinity.

4.1 Least-squares estimation

We consider least-squares parameter estimation of the all-pass causal representation (4.3), which does not require specifying the errors distribution in the MAR(p,q) model. A LS estimator of $\boldsymbol{\eta}_0 = (\eta_{01}, \dots, \eta_{0,p+q})$ in Model (4.3) is

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}\in\mathbb{R}^{p+q}} \mathcal{L}_n^*(\boldsymbol{\eta}), \tag{4.5}$$

where

$$\mathcal{L}_n^*(\boldsymbol{\eta}) = \sum_{t=p+q+1}^{n} \left(X_t - \sum_{i=1}^{p+q} \eta_i X_{t-i} \right)^2. \tag{4.6}$$

Proposition 4.2 *Let (X_t) be the strictly stationary solution of model (4.1)-(4.2). Then the LS estimator $\hat{\boldsymbol{\eta}}$ is consistent: $\hat{\boldsymbol{\eta}} \to \boldsymbol{\eta}_0$ in probability, as $n\to\infty$.*

To derive the asymptotic distribution of the LS estimator of $\boldsymbol{\eta}_0$, we introduce the sequences

$$a_n = \inf\{x : \mathbb{P}(|\varepsilon_0| > x) \le n^{-1}\}, \quad \text{and} \quad \tilde{a}_n = \inf\{x : \mathbb{P}(|\varepsilon_0\varepsilon_1| > x) \le n^{-1}\}. \tag{4.7}$$

Let $\boldsymbol{J}$ the $(p+q)\times(p+q)$ shift matrix, with ones on the superdiagonal and zeros elsewhere. For $\ell = 1, \dots, p+q$ let $\boldsymbol{K}^{(\ell)} = \boldsymbol{J}^\ell + {}^{\mathrm{t}}\boldsymbol{J}^\ell$ (with $\boldsymbol{K}^{(p+q)} = \boldsymbol{0}$). Let $\boldsymbol{L} = [\boldsymbol{K} \quad \boldsymbol{K}^{(2)} \quad \dots \ \boldsymbol{K}^{(p+q)}]$

Proposition 4.3 *Let (X_t) be the strictly stationary solution of Model (4.1)-(4.2) with $E|\varepsilon_t|^\alpha = \infty$. Then, letting $\boldsymbol{\rho} = [\rho(i)]_{i=1,\dots,p+q}, \boldsymbol{R} = [\rho(i-j)]_{i,j=1,\dots,p+q}$,*

$$\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{d} \boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\mathbf{Z}, \quad \textit{where} \quad \boldsymbol{Z} = (Z_1, \dots, Z_{p+q})', \tag{4.8}$$

$Z_k = \sum_{l=1}^{+\infty} \{\rho(k+l) + \rho(k-l) - 2\rho(l)\rho(k)\} S_l/S_0$, for $k = 1, \dots, p+q$, and $S_0, S_1, S_2, \dots$ are independent stable random variables; S_0 is positive with index $\alpha/2$ and S_j, for $j \ge 1$, has index α. If the law of $|\varepsilon_t|$ is asymptotically equivalent to a Pareto, (4.8) holds with $a_n^2/\tilde{a}_n = (n/\ln n)^{1/\alpha}$.

Example 4.1 (MAR(1,1) process (continued)) For the MAR(1,1) process, Proposition 4.3 allows to compute the asymptotic distribution of the LS estimator of $(\phi_0 + \psi_0, \phi_0\psi_0)$, using

$$\boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\} = \frac{1+\phi_0\psi_0}{(1-\psi_0^2)(1-\phi_0^2)} \begin{pmatrix} (1+\psi_0\phi_0)^2 + (\psi_0+\phi_0)^2 & -(\psi_0+\phi_0) \\ -2(\psi_0+\phi_0)(1+\psi_0\phi_0) & 1+\psi_0\phi_0 \end{pmatrix}.$$

This matrix can be straightforwardly estimated by plugging LS estimators of $\phi_0 + \psi_0$ and $\phi_0\psi_0$. Deriving the asymptotic distribution of an LS estimator of (ϕ_0, ψ_0) by the delta method requires an additional identifiability condition, for instance $\phi_0 > \psi_0$. Assuming $\phi_0 > \psi_0$, we find

$$\frac{a_n^2}{\tilde{a}_n} \begin{pmatrix} \hat{\psi} - \psi_0 \\ \hat{\phi} - \phi_0 \end{pmatrix} \xrightarrow{d} \boldsymbol{\nabla f}(\eta_{01}, \eta_{02}) \boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\mathbf{Z},$$

where $(\psi_0, \phi_0) := f(\eta_{01}, \eta_{02}) = \left(\eta_{01} - \sqrt{\eta_{01}^2 + 4\eta_{02}}, \eta_{01} + \sqrt{\eta_{01}^2 + 4\eta_{02}}\right)/2$.

4.2 Portmanteau test

Validity of the estimated model can be assessed by studying the sample autocorrelations of the residuals. Once the parameters of Model (4.3) have been estimated by LS, with $\hat{\boldsymbol{\eta}} = (\hat{\eta}_i)_{i=1,\dots,p+q}$, the corresponding residuals are defined by

$$\hat{\zeta}_t = X_t - \sum_{i=1}^{p+q} \hat{\eta}_i X_{t-i}, \qquad t = p+q+1, \dots, n. \tag{4.9}$$

Let, for $h \geq 0$, $\hat{\rho}_{\hat{\zeta}}(h) = \hat{\rho}_{\hat{\zeta}}(-h) = \frac{\hat{\gamma}_{\hat{\zeta}}(h)}{\hat{\gamma}_{\hat{\zeta}}(0)}$ where $\hat{\gamma}_{\hat{\zeta}}(h) = \sum_{t=h+p+q+1}^{n} \hat{\zeta}_t \hat{\zeta}_{t-h}$ and $\hat{\gamma}_{\hat{\zeta}}(-h) = \hat{\gamma}_{\hat{\zeta}}(h)$. For a fixed integer $H \geq 1$, let $\hat{\boldsymbol{\rho}}_{\hat{\zeta}} = [\hat{\rho}_{\hat{\zeta}}(1), \dots, \hat{\rho}_{\hat{\zeta}}(H)]'$.

Proposition 4.4 *Under the assumptions of Proposition 4.3, the vector of residuals empirical autocorrelations satisfies*

$$\frac{a_n^2}{\tilde{a}_n} \hat{\boldsymbol{\rho}}_{\hat{\zeta}} \xrightarrow{d} \gamma(0) \boldsymbol{A_H} \boldsymbol{Z}, \quad \textit{where} \quad \boldsymbol{Z} = (Z_1, \dots, Z_{H+p+q})',$$

where $\gamma(0) = \left(\sum_{k=-\infty}^{\infty} d_k^2\right)$, (a_n) *and* $(\tilde{a}_n)$ *are defined in* (4.7)*, the* Z_i*'s are as in Proposition 4.3, and* A_H *is a non random* $H \times (p+q+H)$ *matrix function of the sole AR coefficients (not on the errors distribution).*

Details regarding matrix A_H are available in the proof. It is now possible to propose a Portmanteau test to check for residuals autocorrelations based for instance on the statistic

$$T_H = a_n^2 \tilde{a}_n^{-1} \sum_{i=1}^{H} |\hat{\rho}_{\hat{\zeta}}(i)| \underset{n \to +\infty}{\xrightarrow{d}} \|\gamma(0) \boldsymbol{A_H} \boldsymbol{Z}\|_1, \tag{4.10}$$

with $\|\boldsymbol{x}\|_1 = \sum |x_i|$ for any vector $\boldsymbol{x} = (x_i)$.

The knowledge of index α is not required for the computation of the sum of absolute empirical correlations in the left-hand side, but the asymptotic distribution, as well as the normalizing constants a_n and $\tilde{a}_n$, depend on α. Having estimated the AR coefficients, a standard estimator can be used for the tail index α (for instance, the Hill estimator; see Embrechts et al. (1997), Theorem 6.4.6, for its main properties under various assumptions). Practical implementation of the test finally requires simulating the estimated asymptotic distribution in (4.10).

4.3 Model selection based on extremes clustering

A difficulty in the inference of mixed causal-noncausal AR processes, is that many representations with seemingly uncorrelated errors hold. Breidt, Davis and Trindade (Section 4.3, 2001) showed that if (X_t) is the strictly stationary solution of Model (4.1)-(4.2), then for any polynomial $\eta_0^*(z)$ obtained from $\eta_0(z)$ by inverting one or several roots, we have

$$\eta_0^*(B)X_t = \zeta_t^*, \tag{4.11}$$

where (ζ_t^*) is an all-pass process. Such representations (4.11) will be called all-pass in the following. The strong representation of a MAR process (i.e. with i.i.d. errors) cannot be distinguished from other competing all-pass representations by Portmanteau approaches based on residuals autocorrelations. However, the errors ζ_t^* of all-pass representations are serially dependent.

4.3.1 Point process of exceedances

This dependence materialises in an important feature known as extreme clustering (see e.g. Hsing, Hüsler, Leadbetter (1988), Markovich (2014) and Chavez-Demoulin and Davison (2012) for a literature review) which yields a way to identify the strong representation among the all-pass alternatives. Let us introduce a linear process (Y_t) with two-sided MA(∞) representation $Y_t = \sum_{k\in\mathbb{Z}} c_k \varepsilon_{t+k}$, where (ε_t) is an i.i.d. sequence satisfying (4.2), $\sum_{k\in\mathbb{Z}} |c_k|^s < +\infty$ for some $0 < s < \alpha$, $s \leq 1$, and assume $\max |c_k| = 1$ for convenience. Adapting Section 3.D by Davis and Resnick (1985), we can study the time indices for which $a_n^{-1} Y_k$ falls outside the interval $(-x, x)$, for $x > 0$. The corresponding point process converges as the number of observations n grows to infinity:

$$\sum_{k=1}^{n} \delta_{\left(k/n, a_n^{-1} Y_k\right)}\Big(\ \cdot \ \cap B_x \Big) \xrightarrow{d} \sum_{k=1}^{+\infty} \xi_k \delta_{\Gamma_k}, \tag{4.12}$$

where δ is the Dirac measure, $B_x = (0, +\infty) \times \Big((-\infty, -x) \cup (x, +\infty) \Big)$, $\{\Gamma_k, k \geq 1\}$ are the points of a homogeneous Poisson Random Measure (PRM) on $(0, +\infty)$ with rate $x^{-\alpha}$,[7] and $\xi_k = \mathrm{Card}\{i \in$

[7] See Daley and Vere-Jones (2007): $\{\Gamma_k, k \geq 1\}$ are the points of a homogeneous PRM on $(0, +\infty)$ with rate $x^{-\alpha}$ if and only if, for any $\ell \geq 1$, nonnegative integers $a_1, \dots, a_\ell$ and $b_1, \dots, b_\ell$ such that $a_i < b_i \leq a_{i+1}$, $i = 1, \dots, \ell$, and any nonnegative integers $n_1, \dots, n_\ell$:

$$\mathbb{P}\Big(N(a_i, b_i] = n_i, i = 1, \dots, \ell \Big) = \prod_{i=1}^{\ell} \frac{[x^{-\alpha}(b_i - a_i)]^{n_i}}{n_i!} \exp\left\{ -x^{-\alpha}(b_i - a_i) \right\},$$

where $N(a_i, b_i]$ denotes the number of terms of $\{\Gamma_k, k \geq 1\}$ falling in the half-open interval $(a_i, b_i]$, $i = 1, \dots, \ell$.

$\mathbb{Z} : J_k|c_i| > 1\}$ where $\{J_k, k \geq 1\}$ are i.i.d. on $(1, +\infty)$, independent of $\{\Gamma_k\}$, with common density:

$$f(z) = \alpha z^{-\alpha-1}\mathbb{1}_{(1,+\infty)}(z). \tag{4.13}$$

The sequences $\{\Gamma_k\}$ and $\{\xi_k\}$ are interpreted (see for instance Leadbetter and Nandagopalan (1989)) as describing respectively the occurrence dates of clusters of extreme events and the size of these clusters (i.e. the number of co-occurring extreme events).

4.3.2 Analysing the error processes of competing models

For the sake of simplicity, we will focus on the MAR(1,1) case. Let (X_t) be the MAR(1,1) process solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$. There are four competing models yielding the same type of all-pass causal AR(2) representation (4.3):

$$\text{Pure causal AR(2):} \qquad (1 - \psi_0 B)(1 - \phi_0 B)X_t = \zeta_t, \tag{4.14}$$

$$\text{MAR(1,1):} \qquad (1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t, \tag{4.15}$$

$$\text{MAR(1,1):} \qquad (1 - \psi_0 B)(1 - \phi_0 F)X_t = \nu_t, \tag{4.16}$$

$$\text{Pure noncausal AR(2):} \qquad (1 - \psi_0 F)(1 - \phi_0 F)X_t = \omega_t, \tag{4.17}$$

where (ζ_t), (ν_t) and (ω_t) denote the sequences of errors of each all-pass model. According to Equation (4.12), the point processes of exceedances of each sequence of errors converge in distribution and will feature dramatically contrasting sequences $\{\xi_k\}$. For instance, we will focus on the first two ones.

Error process of the strong (i.i.d.) representation Let us first consider Model (4.15) which is the correct one, i.e. (ε_t) is i.i.d. Thus, $c_0 = 1$, $c_k = 0$ for $k \neq 0$ and $\xi_k = \text{Card}\{i \in \mathbb{Z} : J_k|c_i| > 1\} = \mathbb{1}_{\{J_k>1\}} = 1$: the extreme errors tend to appear isolated from each other.

Error process of the pure causal (all-pass) alternative We consider for instance Model (4.14) with non i.i.d. errors. The (rescaled) errors read $\zeta_t = \sum_{k\geq -1} \frac{c_k}{\max_j |c_j|}\varepsilon_{t+k}$ with $c_k = \psi_0^k(1 - \psi_0^2)$ for $k \geq 0$ and $c_{-1} = -\psi_0$. Denote $(c_{(k)})_{k\geq 1}$ the sequence obtained by sorting $(|c_k|)_{k\geq -1}$ in descending order. Then, $\xi_k = \text{Card}\Big\{i \geq 1 : J_k \frac{c_{(i)}}{c_{(1)}} > 1\Big\} = \arg\max_{i\geq 1}\{J_k > c_{(i)}^{-1}c_{(1)}\}$. From this characterization of the law of $\{\xi_k\}$ and the marginal law of the i.i.d. sequence $\{J_k\}$ given by (4.13), we deduce that for any $\ell \geq 1$:

$$\mathbb{P}\Big(\xi_k \geq \ell\Big) = \mathbb{P}\Big(J_k > c_{(\ell)}^{-1}c_{(1)}\Big) = c_{(\ell)}^{\alpha}c_{(1)}^{-\alpha}. \tag{4.18}$$

Thus, the ξ_k's are charging non-zero probabilities on arbitrarily high cluster sizes, yielding as expected that the extreme errors of Model (4.14) tend to cluster. The effects of both α and the AR parameter ψ_0 on this probability are depicted on Figure 3.

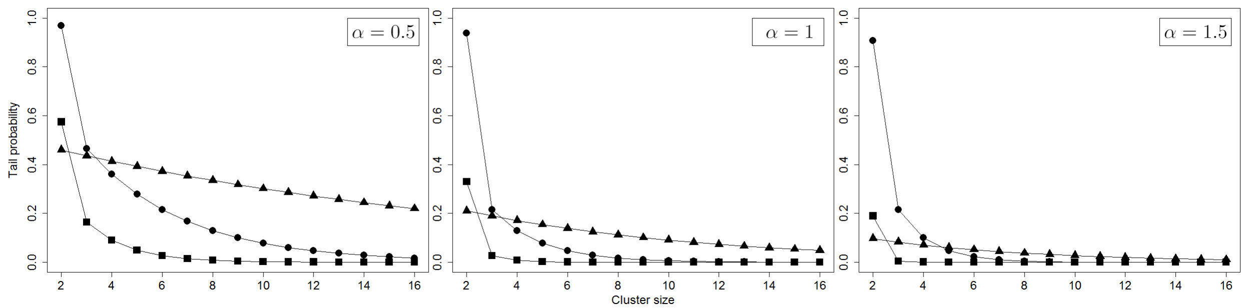


Figure 3: Theoretical tail probability of the cluster size of extreme realisations from errors of Model (4.14) using formula (4.18) for three different values of α. For all panels: $\psi_0 = 0.3$ (squares), $\psi_0 = 0.6$ (points), $\psi_0 = 0.9$ (triangles).

4.3.3 Errors at higher horizons

As can be seen on Figure 3, certain values of α and ψ_0 make large clusters of extremes improbable. For such values, distinguishing strong and all-pass representations from the errors behaviours is difficult. This difficulty can be alleviated by considering the errors of the competing models at further horizons. For simplicity, consider the noncausal AR(1) model. There are two competing models, yielding the same all-pass causal representation (4.3):

$$X_t = \psi_0 X_{t+1} + \varepsilon_t, \quad \text{and} \quad X_t = \psi_0 X_{t-1} + \zeta_t. \tag{4.19}$$

For any $h \geq 1$, expansions of these equations at horizons h read:

$$\varepsilon_{t+h|t} := X_t - \psi_0^h X_{t+h} = \varepsilon_t + \psi_0 \varepsilon_{t+1} + \ldots + \psi_0^{h-1} \varepsilon_{t+h-1}, \tag{4.20}$$

$$\zeta_{t+h|t} := X_{t+h} - \psi_0^h X_t = \psi_0^h \sum_{k=0}^{h-1} \psi_0^k \varepsilon_{t+k-h} + \sum_{k \geq h} \psi_0^k (1 - \psi_0^{2h}) \varepsilon_{t+k}. \tag{4.21}$$

We can deduce that the point processes of excedances of the errors $\varepsilon_{t+h|t}$ and $\zeta_{t+h|t}$ at horizon h will exhibit clusters of random sizes $\xi_k = \text{Card}\left\{ i \in \mathbb{Z} : J_k \frac{|c_i|}{\max_j |c_j|} > 1 \right\}$ where $c_i = \psi_0^i$ if $0 \leq i \leq h-1$ for the strong model, whereas for the all-pass model, the sequence $(|c_i|)$ reads: $|\psi_0|^h, \ldots, |\psi_0|^{2h-1}, 1 - \psi_0^{2h}, |\psi_0|(1-\psi_0^{2h}), |\psi_0|^2(1-\psi_0^{2h}), \ldots$. Thus, the extreme realisations of the errors (4.20) will appear by clusters of at most h consecutive observations, whereas the errors (4.21) will likely appear by larger clusters (see Appendix B for illustration). This analysis can be extended to general MAR processes by disentangling the pure causal and noncausal components of each competitor model (as in the proof of Proposition 3.1).

5 A Monte Carlo study

We conducted three types of experiments in order to gauge the sample properties of the LS procedure applied to the weak all-pass representation. On synthetic data generated from a MAR(1,1) process, we assessed ι) the consistency of the estimators of the roots and the convergence in distribution of the LS estimators of the backward AR(2) specification, $\iota\iota$) the empirical size of the Portmanteau-type statistic $\iota\iota\iota$) the extreme clustering in the residuals of the four competing models that the LS estimation implies.

5.1 LS estimation

We simulated 100,000 paths with lengths 500, 2000 and 5,000 observations of α-stable MAR(1,1) processes solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$ with $\psi_0 = 0.7$, $\phi_0 = 0.9$ and tail indices $\alpha = 1.5$, 1 and 0.5. We computed the LS estimator $(\hat{\eta}_1, \hat{\eta}_2)$ and deduced estimators $(\hat{\psi}, \hat{\phi})$ by taking the inverses of the zeros of $1 - \hat{\eta}_1 X - \hat{\eta}_2 X^2$ (we impose $|\hat{\psi}| \leq |\hat{\phi}|$ for the sake of identifiability). For each model, Table 1 reports the empirical frequencies of estimators that are sufficiently close to the actual values of the roots. As expected, the accuracy increases with n but, more strikingly, it increases sharply as α approaches zero.

		$\alpha = 1.5$			$\alpha = 1$			$\alpha = 0.5$		
n		$a = 0.1$	$a = 0.05$	$a = 0.01$	$a = 0.1$	$a = 0.05$	$a = 0.01$	$a = 0.1$	$a = 0.05$	$a = 0.01$
500	$\hat{p}_a(\phi)$	99.8%	94.6%	33.3%	99.7%	96.4%	48.5%	99.1%	97.5%	71.4%
	$\hat{p}_a(\psi)$	78.2%	55.2%	18.7%	83.8%	69.7%	33.0%	86.2%	79.9%	58.6%
2000	$\hat{p}_a(\phi)$	99.9%	98.9%	54.3%	99.9%	99.2%	74.3%	99.8%	99.4%	90.3%
	$\hat{p}_a(\psi)$	96.3%	87.2%	34.6%	96.0%	91.5%	60.4%	96.4%	94.5%	84.6%
5000	$\hat{p}_a(\phi)$	99.9%	99.8%	74.4%	99.9%	99.7%	88.4%	99.9%	99.7%	95.8%
	$\hat{p}_a(\psi)$	98.7%	96.3%	53.6%	98.5%	96.9%	78.9%	98.6%	97.8%	93.2%

Table 1: Accuracy of the roots-estimation through backward LS: $\hat{p}_a(\theta)$ denotes the frequency of estimations $\hat{\theta}$ belonging to the set $\{|\hat{\theta} - \theta_0| < a\} \cap \{\hat{\theta} \in \mathbb{R}\}$, for $\theta = \phi$ or ψ, for $a = 0.01$, 0.05, 0.1 and over 100,000 simulated paths of the α-stable MAR(1,1) process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, with $\psi_0 = 0.7$ and $\phi_0 = 0.9$.

Turning to the asymptotic distribution of $(\hat{\eta}_1, \hat{\eta}_2)$, results reported in Appendix B show that the finite sample distribution approaches its asymptotic behaviour much slower for lower values of

α. In the same line, a direct implementation of the Portmanteau test using the statistics (4.10) also showed heavy distortions in finite sample. These distortions were expected as they were already reported by Lin and McLeod (2008) in the pure causal AR framework who suggested a Monte Carlo test or parametric bootstrap to alleviate this problem. This approach relies on Monte Carlo simulations of the estimated causal AR model to approximate the distribution of the Portmanteau statistics and could still be used in the present MAR framework. Indeed, it is noticeable that the asymptotic behaviour of our test statistics (4.10) only depends on the all-pass causal AR representation of the process. Therefore, we proceed with the same Monte Carlo test methodology (see Appendix B, Lin and McLeod (2008)). An important difference in our framework is that we estimate by LS an all-pass causal representation in a first step and then use these estimates to simulate paths of the corresponding pure causal AR models, as if it were the strong representation. By Proposition 4.4, we know that the residuals autocorrelations of the latter have the same asymptotic distribution as the residuals autocorrelations of the weak causal representation of the true MAR model.[8] We conducted the experiment on the same MAR(1,1) model as before with $\alpha = 1.5$, 1, 0.5 and $n = 500$. We report in Table 2 the empirical sizes of the 1, 5 and 10% nominal tests for lags $H = 1, \ldots, 10$. It can be seen that using the parametric bootstrap procedure, the Portmanteau test is much better behaved in finite sample, especially for $\alpha = 1.5$, which is the most realistic value for financial series (see Footnote [4]).

5.2 Diagnostic checking of extremal residuals clustering

We now gauge the usefulness of the results of Section 4.3 by simulating paths of the α-stable MAR(1,1) process $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$ with different parameterisations and analysing the empirical residuals of the four competing models (4.14)-(4.17). More specifically for each estimated model, we compute the errors at several horizons h (as in (4.20)-(4.21) for the AR(1)). For a given threshold $x > 0$, we identify the clusters of consecutive "extreme" values (that is, values larger than x in modulus) for such errors series. Let $\hat{\xi}_{k,h}(x)$ denote the number of consecutive exceedances of the threshold x for the k-th cluster. As explained in Section 4.3, for any horizon h, we expect all-pass representations to display larger clusters of extreme errors than for the the strong model, for which clusters larger than h have small probability. We therefore propose an Excess Clustering

[8] Another approach could consist in testing directly whether the coefficients of the all-pass representation are different from zero. This methodology based on bootstrap innovations was developed in the pure noncausal heavy-tailed framework by Cavaliere, Nielsen and Rahbek (2017).

	$\alpha = 1.5$			$\alpha = 1$			$\alpha = 0.5$		
H	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	1.30	5.80	10.5	1.25	5.40	10.4	1.45	4.10	7.35
2	1.55	5.65	10.9	1.60	5.25	9.65	1.35	3.90	7.05
3	1.40	5.35	10.9	1.30	5.05	9.40	1.20	4.45	6.95
4	1.50	5.45	10.5	1.35	5.00	9.90	1.20	4.35	7.00
5	1.25	5.50	9.85	1.20	4.90	9.20	1.10	4.20	7.30
6	1.30	5.00	10.1	1.05	4.70	9.40	1.10	4.25	7.40
7	1.20	5.25	9.75	1.05	4.40	9.15	1.20	4.00	7.50
8	1.10	5.25	9.75	1.15	4.55	8.70	1.05	3.70	7.25
9	1.25	5.10	9.80	1.30	4.30	8.60	1.05	3.75	7.50
10	1.35	5.10	10.1	1.20	4.55	8.70	0.90	3.65	7.15

Table 2: Empirical sizes (%) of the Portmanteau statistics (4.10) implemented by the parametric bootstrap procedure. The empirical size was calculated based on 2000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, with $\psi_0 = 0.7$ and $\phi_0 = 0.9$. Each Monte Carlo test was performed with 1000 simulations.

(EC)[9] indicator defined as:

$$\mathrm{EC}_h = \frac{\sum\limits_{k/\hat{\xi}_{k,h}(x)>h} \left(\hat{\xi}_{k,h}(x) - h\right)}{\mathrm{Card}\{k : \hat{\xi}_{k,h}(x) > h\}}, \quad \text{if} \quad \mathrm{Card}\{k : \hat{\xi}_{k,h}(x) > h\} > 0, \quad \text{else} \quad \mathrm{EC}_h = 0. \tag{5.1}$$

We start by generating 10,000 sample paths of MAR(1,1) processes. For each of these paths, we fit a backward AR(2), estimate the set $\{\phi_0, \psi_0\}$, and for each of the four competing models we estimate a term structure of extreme residuals clustering with respect to the horizon, using the estimator (5.1).

Averaging model-wise across these 10,000 simulations yields a summary of how prominent EC is among the residuals of the competing models. To assess the effects of the multiple parameters, we perform this experiment for two MAR(1,1) processes. The results are displayed in Figure 4. It can be noticed that the all-pass models feature excessively clustering residuals at any horizon whereas the residuals of the strong model are barely deviating from no excess clustering. As expected from (4.18), the heavier the tails the easier it is to identify dependent residuals. This is in line with the

[9] For a given h, EC_h defined at (5.1) corresponds to the average size of clusters larger than h, from which we subtract h, and is 0 if all the clusters are smaller than h. It is related to the Extremal Index, more common in the literature, which is the plain average size of clusters. Also, the choice of clustering scheme, i.e. how the sequence $\left(\hat{\xi}_{k,h}(x)\right)_k$ is constructed, can have an impact on the estimated excess clustering : more elaborate clustering schemes could be considered (see Ferro and Segers (2003)).

findings of Hecq, Lieb and Telg (2015) who are concerned with identification of causal/noncausal models using the LAD estimator. Noticeably, even with very heavy tails ($\alpha = 0.5$), the residuals at any horizons of the strong representation still barely deviate from no excess clustering.

These experiments highlight the usefulness of considering residuals at various horizons, instead of focusing only on basic residuals. Indeed, all the term structures of excess clustering show that the contrast between the competing models does not arise for $h = 1$ but rather tends to peak for intermediate values of h.

Last, we assess how well we can discriminate between the all-pass models and the strong representation by exploiting the excess clustering feature. For each of the 10,000 simulations, we rank the four competing models according to the area under the term structure curve of excess clustering (AUC) and select the candidate with least AUC. Table 3 reports the true positive rates of this procedure. For $\alpha = 1.5$ and $n = 500$, the strong representation was correctly identified in above 88% of the 10,000 simulated paths and this proportion increases with n.

$n = 500$	$n = 2000$	$n = 5000$
88.4%	95.8%	97.5%

Table 3: Correct model selection rates based on least excess clustering across 10,000 simulated paths of the MAR(1,1) process (X_t) solution of $(1 - 0.7F)(1 - 0.9B)X_t = \varepsilon_t$ with i.i.d. 1.5-stable noise.

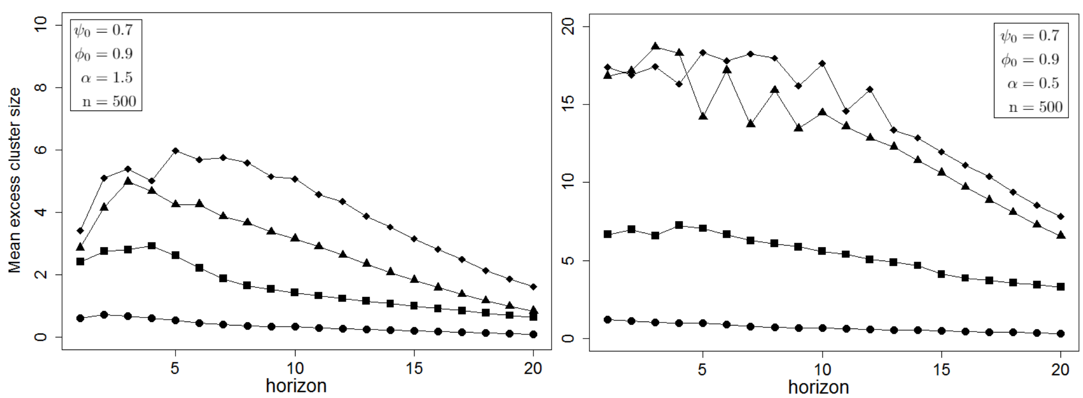


Figure 4: Across 10,000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$, average of the term structure of excess clustering of the linear residuals of the four competing models (4.14) (squares), the strong representation (4.15) (points), (4.16) (triangles) and (4.17) (diamond). The parameterisations and path lengths are indicated on each panel.

6 An application to financial series

In this section, we illustrate the adequacy of MAR models for real economic series. We fitted MAR models on six financial series of monthly prices: stock prices of Coca-Cola (January 1978 to June 2017), Boeing (February 1962 to December 2012), Hong Kong' stock market index (HSI) (December 1986 to April 2017), Walmart (September 1979 to June 2017), Exxon (February 1970 to June 2017), and on the quarterly Shiller Price/Earning ratio (1881 to 2017)[10]. All the series, pictured on Figure 5, have been centered and a linear deterministic trend has been fitted and subtracted. We can see that all the series feature multiple bubble events followed by a sharp drop.

6.1 AR estimation and validation using parametric bootstrap Portmanteau

We start by investigating the appropriate total AR order for each series using the Monte Carlo Portmanteau test of Section 5.1. For each series, starting from total AR order 1, we estimate weak causal representations of increasing order by LS and perform the Portmanteau test for lags $H = 1, \dots, 10$ using the parametric bootstrap Portmanteau procedure of Lin and McLeod (10 000 paths were simulated for each test). At a given total AR order r, if the Monte Carlo test has a P-value below 10% for any lag H, we reject the null hypothesis that a MAR(p, q) model, $r = p + q$, would suitably describe the series and repeat the procedure for total AR order $r + 1$. In the case the P-values of the tests are above 10% for all lags $H = 1, \dots, 10$, we accept the total AR order r. The results[11] are reported in Table 4. The total AR orders retained for each series are: Boeing: 4; Exxon: 1; Coca-Cola: 1; Walmart: 1; HSI: 3; Shiller P/E: 8.

6.2 MAR selection based on extreme clustering

For each of the mentioned series, we apply the methodology of Section 5.2: we fit all possible MAR models of total order $r = 1, \dots, 8$, compute the term structure of excess clustering of the residuals of each competing model and the associated term structure of EC and we then rank the competing models according to the area under the term structure curve. In Table 5, we report for each total AR order $r = p + q$ the MAR(p, q) specification which displays the lowest AUC of excess clustering

[10] A measure of exuberance on financial markets proposed by Economics Nobel Prize Robert Shiller capturing how expensive stocks are compared to the earnings they bring. Data available on https://www.quandl.com.

[11] This procedure yields as a by-product the McCulloch quantile estimates of the tail index α (see McCulloch (1986)) for the six financial series. Values of $\hat{\alpha}$: Boeing: 1.73; Exxon: 1.69; Coca: 1.64; Walmart: 1.67; HSI: 1.38; Shiller P/E: 1.50. In all the cases, the infinite variance hypothesis is plausible.

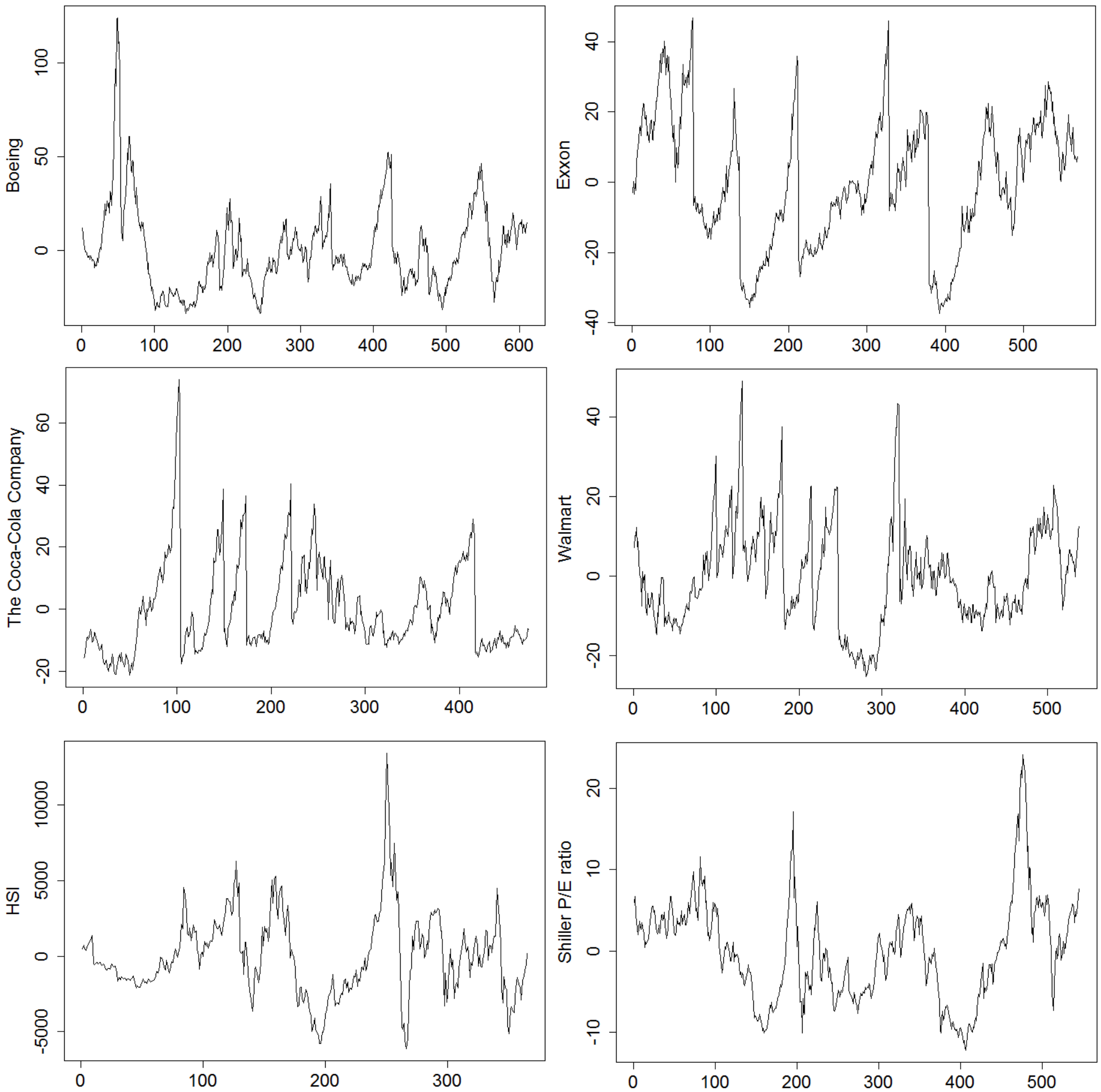


Figure 5: Financial series paths: Boeing (2/1962 to 12/2012), Exxon (2/1970 to 06/2017), Coca-Cola (1/1978 to 06/2017), Walmart (9/1979 to 6/2016), Hang Seng Index (HSI, 12/1986 to 05/2017) and the Shiller P/E ratio (Q1/1881 to Q2/2017). All series are monthly except the latter which is quarterly. Each are centered and a linear trend has been fitted and subtracted.

and the median AUC of its competitors. The favoured specification of some series feature very low AUC excess clustering even for total AR order $r = 1$ (e.g. Walmart, Coca-Cola), whereas others display high excess clustering for low total AR order (e.g. Boeing, Shiller P/E). We can notice that for the latter, excess clustering rapidly decreases as r increases. Besides, we can also see a general decreasing trend for median excess clustering of competing models as r increases. Based on the total AR orders determined at the previous section, we could select a final MAR specification

Total AR order r	Boeing	Exxon	Coca-Cola	Walmart	HSI	Shiller P/E
1	0.01	**63.8**	**14.3**	**27.4**	3.24	0.01
2	0.20				5.41	0.03
3	0.07				**36.3**	0.86
4	**10.4**					0.85
5						0.96
6						2.28
7						3.26
8						**13.2**

Table 4: P-values (%) of the portmanteau tests based on bootstrap for increasing AR order r.

for each of the series. This selection is reported in Table 6 which shows the causal and noncausal orders as well as the (inverted) roots of the corresponding polynomials. Three series have been identified as pure noncausal AR(1) (Exxon, Coca-Cola, Walmart) whereas 3 others display more complex dynamics (Boeing, HSI, Shiller P/E ratio).

7 Concluding remarks

Noncausal models may provide better understanding of the dynamic features of a time series that are not perceived via causal models. We showed that even the adjunction of a very simple noncausal component to an arbitrarily complex classical causal AR is sufficient to profoundly alter its motion and this in turn impacts the way we infer about its future. Moreover, we established that fitting a backward AR of proper order on data generated by a MAR process will yield a model that will pass the Portmanteau tests, although misspecifying at the same time all the noncausal roots as causal. Because of this common fitting procedure, noncausal components in economic and financial time series may have remained concealed and implicitly been considered as causal, rendering predictions potentially critically suboptimal. We therefore laid the bases for a full identification scheme relying 1) on parametric bootstrap Portmanteau tests to determine the most suitable total AR order 2) on extreme residuals clustering to select the causal and noncausal orders and roots. Our estimation results on six monthly/quaterly financial series revealed the plausibility of infinite variance and noncausal AR components. For three of the series purely noncausal processes were obtained, while the remaining three ones featured non-trivial MAR structures.

Total AR order		Boeing	Exxon	Coca-Cola	Walmart	HSI	Shiller P/E
	Favoured specification	(1,0)	(1,0)	(1,0)	(1,0)	(0,1)	(1,0)
1	AUC Excess Clustering	8.67	3.25	2.67	1	7.75	15.6
	Median of competitors	31.1	25.7	51.7	29.3	17	51.2
	Favoured specification	(1,1)	(1,1)	(2,0)	(2,0)	(1,1)	(1,1)
2	AUC Excess Clustering	5.33	2.75	1.67	1	5.75	9.58
	Median of competitors	14.2	13.58	27.8	15.0	9.33	25.5
	Favoured specification	(2,1)	(1,2)	(1,2)	(3,0)	(1,2)	(1,2)
3	AUC Excess Clustering	5.15	2.25	1.33	1	0.5	1.64
	Median of competitors	11.5	13.1	27.7	10.5	6.17	8
	Favoured specification	(2,2)	(3,1)	(4,0)	(4,0)	(1,3)	(1,3)
4	AUC Excess Clustering	7.05	3	1	1	0.5	2.13
	Median of competitors	15.2	6.63	11.7	9.57	5.66	5.95
	Favoured specification	(4,1)	(5,0)	(5,0)	(5,0)	(3,2)	(3,2)
5	AUC Excess Clustering	3.5	0	1	1	2	2
	Median of competitors	12	8.64	12.9	6.74	5.67	4.88
	Favoured specification	(6,0)	(5,1)	(6,0)	(6,0)	(1,5)	(2,4)
6	AUC Excess Clustering	3	0	0	0	0.5	1.33
	Median of competitors	10.3	5.5	11.4	5.73	4.2	5
	Favoured specification	(6,1)	(7,0)	(7,0)	(7,0)	(2,5)	(3,4)
7	AUC Excess Clustering	2	0	1	0	1	1.83
	Median of competitors	10.3	7.25	10.1	6.04	3.25	4
	Favoured specification	(6,2)	(8,0)	(8,0)	(8,0)	(2,6)	(4,4)
8	AUC Excess Clustering	3.75	0	0	0	1.5	1.75
	Median of competitors	9.11	6.5	10.3	5.12	6.2	4.63

Table 5: Selection based on extreme clustering.

Series	Final specification	Noncausal (inverted) roots	Causal (inverted) roots
Boeing	MAR(2,2)	0.66, 0.91	$-0.23 \pm 0.46i$
Exxon	MAR(1,0)	0.95	–
Coca-Cola	MAR(1,0)	0.90	–
Walmart	MAR(1,0)	0.91	–
HSI	MAR(1,2)	0.37	-0.27, 0.89
Shiller P/E	MAR(4,4)	$-0.40 \pm 0.63i$, $0.66 \pm 0.27i$	$0.18 \pm 0.58i$, -0.83, 0.96

Table 6: Selection of the MAR specification for each financial series among the favoured ones of Table 5 based on the total AR order determined in Table 4. The MAR(p, q) specifications indicate the noncausal p and causal q orders as well as the (inverted) roots of the corresponding polynomials.

Appendix A: Proofs

A Proof of Proposition 2.1

Using the MA(∞) representation (2.2) of X_t and the assumption that $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,\mu)$, it follows that

$$\begin{aligned}\forall s\in\mathbb{R},\quad \psi_{X_t}(s) := \mathbb{E}\left[e^{isX_t}\right] = \mathbb{E}\left[e^{is\sum\limits_{k=-\infty}^{+\infty} d_k\varepsilon_{t+k}}\right] &= \prod_{k=-\infty}^{+\infty}\mathbb{E}\left[e^{isd_k\varepsilon_{t+k}}\right]\\ &= \prod_{k=-\infty}^{+\infty}\exp\left\{-\sigma^\alpha|d_ks|^\alpha\left(1-i\beta\text{sign}(d_ks)w(\alpha,d_ks)\right)+id_ks\mu\right\}.\end{aligned}$$

If $\alpha\neq1$, then,

$$\begin{aligned}\forall s\in\mathbb{R},\quad \ln\psi_{X_t}(s) &= \sum_{k=-\infty}^{+\infty} -\sigma^\alpha|d_ks|^\alpha\left(1-i\beta\text{sign}(d_ks)\text{tg}\left(\frac{\pi\alpha}{2}\right)\right)+id_ks\mu\\ &= -\tilde{\sigma}^\alpha|s|^\alpha\left(1-i\beta\frac{\sum\limits_{k=-\infty}^{+\infty}|d_k|^\alpha\text{sign}(d_k)}{\sum\limits_{k=-\infty}^{+\infty}|d_k|^\alpha}\text{sign}(s)\text{tg}\left(\frac{\pi\alpha}{2}\right)\right)+is\mu\sum_{k=-\infty}^{+\infty}d_k.\end{aligned}$$

Whereas if $\alpha=1$, then,

$$\begin{aligned}\forall s\in\mathbb{R},\quad \ln\psi_{X_t}(s) &= \sum_{k=-\infty}^{+\infty} -\sigma|d_ks|\left(1+i\frac{2}{\pi}\beta\text{sign}(d_ks)\ln|d_ks|\right)+id_ks\mu\\ &= -|s|\sigma\left(\sum_{k=-\infty}^{+\infty}|d_k|\right)\left(1+i\frac{2}{\pi}\beta\frac{\sum\limits_{k=-\infty}^{+\infty}d_k\ln|d_k|}{\sum\limits_{k=-\infty}^{+\infty}|d_k|}\text{sign}(s)\ln|s|\right)\\ &\quad+is\left(\sum_{k=-\infty}^{+\infty}d_k\mu-\frac{2}{\pi}\sigma\beta\sum_{k=-\infty}^{+\infty}d_k\ln|d_k|\right).\end{aligned}$$

The conclusion follows.

B Proof of Proposition 3.1

We decompose (X_t) into its pure causal AR(q) and noncausal AR(p) components (see Lanne, Saikkonen (2011) and Gouriéroux, Jasiak (2016)) respectively (v_t) and (u_t), defined by

$$u_t=\phi(B)X_t \iff \psi(F)u_t=\varepsilon_t, \tag{B.1}$$

$$v_t=\psi(F)X_t \iff \phi(B)v_t=\varepsilon_t. \tag{B.2}$$

We first show that (u_t) is a Markov process of order p. When there is no risk of ambiguity we denote by f a generic density, whose definition can change along the proof. Since by (B.1), $u_t = \psi_1 u_{t+1} + \ldots + \psi_p u_{t+p} + \varepsilon_t$, for $k > p$, the conditional density of u_t given its k past values is

$$\begin{aligned} f\left(u_t \middle| u_{t-1}, \ldots, u_{t-k}\right) &= \frac{f\left(u_t, \ldots, u_{t-k}\right)}{f\left(u_{t-1}, \ldots, u_{t-k}\right)} \\ &= \frac{f\left(u_{t-k} \middle| u_{t-k+1}, \ldots, u_{t-k+p}\right) f\left(u_t, \ldots, u_{t+1-k}\right)}{f\left(u_{t-k} \middle| u_{t-k+1}, \ldots, u_{t-k+p}\right) f\left(u_{t-1}, \ldots, u_{t+1-k}\right)} \\ &= \frac{f\left(u_t, \ldots, u_{t-p}\right)}{f\left(u_{t-1}, \ldots, u_{t-p}\right)} = f\left(u_t \middle| u_{t-1}, \ldots, u_{t-p}\right), \end{aligned}$$

where the second equality follows from the Bayes formula and (B.1), that is $u_t = \phi(B)X_t$, and the third equality is obtained by decreasing induction on k. We now turn to the MAR process (X_t). From $u_t = \phi(B)X_t$, we have $X_t = \sum_{i=1}^{q} \phi_i X_{t-i} + u_t$. Thus, with obvious notation, for any $x, x_1, \ldots, x_{p+q} \in \mathbb{R}$,

$$\begin{aligned} & f_{X_t}\left(x \middle| X_{t-1} = x_1, X_{t-2} = x_2, \ldots\right) \\ &= f_{u_t + \sum_{i=1}^{q} \phi_i x_i}\left(x \middle| X_{t-1} = x_1, \ldots\right) \\ &= f_{u_t}\left(x - \sum_{i=1}^{q} \phi_i x_i \middle| u_{t-1} = x_1 - \sum_{i=1}^{q} \phi_i x_{1+i}, u_{t-2} = x_2 - \sum_{i=1}^{q} \phi_i x_{2+i}, \ldots\right) \\ &= f_{u_t}\left(x - \sum_{i=1}^{q} \phi_i x_i \middle| u_{t-1} = x_1 - \sum_{i=1}^{q} \phi_i x_{1+i}, \ldots, u_{t-p} = x_p - \sum_{i=1}^{q} \phi_i x_{p+i}\right) \end{aligned}$$

using the Markov property of (u_t). The latter quantity is a function of $(x, x_1, \ldots, x_{p+q})$, showing that process (X_t) is Markov of order $p + q$.

C Proof of Theorem 3.1

We first show that the theorem holds for $q = 0$ and we then extend it to general MAR(p, q) processes.

Lemma C.1 *Let (X_t) be an α-stable pure noncausal AR(p) process solution of $X_t = \psi_1 X_{t+1} + \ldots + \psi_p X_{t+p} + \varepsilon_t$, where the roots of $\psi(z)$ are outside the unit circle. Then,*

$$\mathbb{E}\left[|X_t|^{\gamma} \middle| X_{t-1}, \ldots, X_{t-p}\right] < +\infty, \quad \textit{for any} \quad \gamma \in (0, 2\alpha + 1).$$

Proof. Suppose $p > 1$ (the result is already known from GZ for $p = 1$). For any $(x_0, \ldots, x_p) \in \mathbb{R}^{p+1}$,

$$f_{X_t|(X_{t+1},\ldots,X_{t+p})=(x_1,\ldots,x_p)}(x_0) = f_{\varepsilon}(x_0 - \psi_1 x_1 - \ldots - \psi_p x_p),$$

because ε_t is independent from $X_{t+1}, \ldots, X_{t+p}$. By the Bayes formula,

$$f_{X_t|(X_{t+1},\ldots,X_{t+p})=(x_1,\ldots,x_p)}(x_0) = \frac{f_{X_{t+p}|(X_t,\ldots,X_{t+p-1})=(x_0,\ldots,x_{p-1})}(x_p)}{f_{X_{t+1},\ldots,X_{t+p}}(x_1,\ldots,x_p)} f_{X_t,\ldots,X_{t+p-1}}(x_0,\ldots,x_{p-1}).$$

Thus,

$$f_{X_{t+p}|(X_t,\ldots,X_{t+p-1})=(x_0,\ldots,x_{p-1})}(x_p) = \frac{f_\varepsilon(x_0 - \psi_1 x_1 - \ldots - \psi_p x_p) f_X(x_p) f_{X_{t+1},\ldots,X_{t+p-1}|X_{t+p}=x_p}(x_1,\ldots,x_{p-1})}{f_{X_t,\ldots,X_{t+p-1}}(x_0,\ldots,x_{p-1})}. \tag{C.1}$$

On the one hand, when $x_p \to \pm\infty$,

$$f_X(x_p) \sim C(x_p)|x_p|^{-\alpha-1}, \tag{C.2}$$

$$f_\varepsilon(x_0 - \psi_1 x_1 - \ldots - \psi_p x_p) \sim C^*(x_p)|x_p|^{-\alpha-1}, \tag{C.3}$$

where $C(x_p)$ and $C^*(x_p)$ are constants depending on x_p, which may change according to whether $x_p \to +\infty$ or $x \to -\infty$. On the other hand, we show that

$$f_{X_{t+1},\ldots,X_{t+p-1}|X_{t+p}=x_p}(x_1,\ldots,x_{p-1}) \underset{|x_p|\to+\infty}{\longrightarrow} 0. \tag{C.4}$$

Let $Z_t = X_t - \psi_1 X_{t+1} - \cdots - \psi_{p-1} X_{t+p-1}$. Conditionally on $X_{t+p} = x_p$, we have $Z_t = \psi_p x_p + \varepsilon_t$. Since X_{t+p} and ε_t are independent and $\psi_p \neq 0$, we have $|Z_t| \to +\infty$ a.s. as $|x_p| \to +\infty$. Therefore, for any $z_0 \in \mathbb{R}$ and any neighbourhood V_{z_0} of z_0, when $|x_p| \to +\infty$,

$$\mathbb{P}\Big(Z_t \in V_{z_0} \Big| X_{t+p} = x_p\Big) \longrightarrow 0, \quad \text{which implies,} \quad \mathbb{P}\Big((X_t,\ldots,X_{t+p-1}) \in V_{\boldsymbol{x}} \Big| X_{t+p} = x_p\Big) \longrightarrow 0,$$

for any point $\boldsymbol{x} \in \mathbb{R}^p$ and neighbourhood $V_{\boldsymbol{x}}$ around this point. Hence the convergence in Equation (C.4). Combining Equations (C.1), (C.2), (C.3) and (C.4), we obtain, for $|x_p|$ large enough,

$$f_{X_{t+p}|(X_t,\ldots,X_{t+p-1})=(x_0,\ldots,x_{p-1})}(x_p) = o\Big(|x_p|^{-2(\alpha+1)}\Big).$$

The conclusion follows.

Let us now prove Theorem 3.1. Let $\gamma \in (0, 2\alpha+1)$. Decomposing (X_t) into its pure causal and noncausal components (v_t) and (u_t), defined in (B.2) and (B.1), we have the equivalence between the information sets

$$(X_{t-1},\ldots,X_{t-p-q}) \quad \text{and} \quad (u_{t-1},\ldots,u_{t-p},v_{t-p-1},\ldots,v_{t-p-q}),$$

and the independence between $(u_{t-1},\ldots,u_{t-p})$ and $(v_{t-p-1},\ldots,v_{t-p-q})$ (see Lanne and Saikkonen (2011), Gouriéroux and Jasiak (2016)). From Equation (B.1), we have for $\gamma \geq 1$ by the triangle inequality,

$$\Big(\mathbb{E}\Big[|X_t|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\Big)^{1/\gamma} \tag{C.5}$$

$$\begin{aligned}
&= \Big(\mathbb{E}\Big[|u_t-\phi_1X_{t-1}-\dots-\phi_qX_{t-q}|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\Big)^{1/\gamma}\\
&\le |\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|+\Big(\mathbb{E}\Big[|u_t|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\Big)^{1/\gamma}\\
&= |\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|+\Big(\mathbb{E}\Big[|u_t|^\gamma\Big|u_{t-1},\dots,u_{t-p},v_{t-p-1},\dots,v_{t-p-q}\Big]\Big)^{1/\gamma}\\
&= |\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|+\Big(\mathbb{E}\Big[|u_t|^\gamma\Big|u_{t-1},\dots,u_{t-p}\Big]\Big)^{1/\gamma},
\end{aligned} \tag{C.6}$$

which is finite almost surely by Lemma C.1 since (u_t) is an α-stable pure noncausal AR(p) process. If $\gamma\in(0,1)$, by the inequality $(a+b)^\gamma\le a^\gamma+b^\gamma$ for any $a,b\ge0$, we have that $|a+b|^\gamma\le\Big(|a|+|b|\Big)^\gamma\le|a|^\gamma+|b|^\gamma$, for any $(a,b)\in\mathbb{R}$. Thus, similarly to (C.6), we show that

$$\mathbb{E}\Big[|X_t|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\le|\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|^\gamma+\mathbb{E}\Big[|u_t|^\gamma\Big|u_{t-1},\dots,u_{t-p}\Big],$$

which completes the proof.

D Proof of Proposition 3.2

The following Lemma will be useful for the proofs of Proposition 3.2 and Corollary 3.2.

Lemma D.1 *Let (X_t) be a MAR(p, q) process. For any $h\ge0$, there exist polynomials P_h and Q_h with $d^\circ(P_h)=q-1$ and $d^\circ(Q_h)=h$, such that for any $t\in\mathbb{Z}$,*

$$X_{t+h}=P_h(B)X_{t-1}+Q_h(F)u_t, \tag{D.1}$$

where (u_t) is defined in (B.1).

Proof. We prove (D.1) by induction on h. The model can obviously be written under the form $X_t=\sum_{i=1}^q\phi_iX_{t-i}+u_t$ from (B.1). Thus (D.1) holds for $h=0$, with $P_0(B)=\sum_{i=0}^{q-1}\phi_{i+1}B^i$ and $Q_0(F)=I$. Assume that the property holds up to the order $h-1$, for $h\ge1$. For $r=\min(h,q)$, $X_{t+h}=\sum_{i=1}^r\phi_{i+1}X_{t+h-i}+\sum_{i=r+1}^q\phi_{i+1}X_{t+h-i}+u_{t+h}$ where, by convention, the second sum vanishes if $r=q$. Thus

$$X_{t+h}=\sum_{i=1}^r\phi_{i+1}P_{h-i}(B)X_{t-1}+\sum_{i=r+1}^q\phi_{i+1}X_{t+h-i}+u_{t+h}+\sum_{i=1}^rQ_{h-i}(F)u_t,$$

which is of the form (D.1) with

$$P_h(B)=\sum_{i=1}^r\phi_{i+1}P_{h-i}(B)+\sum_{i=r+1}^q\phi_{i+1}B^{i-h-1},\quad Q_h(B)=F^h+\sum_{i=1}^rQ_{h-i}(F). \tag{D.2}$$

Therefore, (D.1) is established.

We now extend Proposition 4.3 by showing that for any $h \geq 0$, $\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_{t-1}, \ldots, X_{t-q-1}\Big] < +\infty$ whenever $0 < \gamma < 2\alpha + 1$. By Lemma D.1 we have, proceeding as for Equation (C.6) and letting $Q_h(z) = \sum_{i=0}^h q_{i,h} z^i$,

$$\Big(\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_{t-1}, \ldots, X_{t-q-1}\Big]\Big)^{1/\gamma} \leq |P_h(B)X_{t-1}| + \sum_{i=0}^h |q_{i,h}| \Big(\mathbb{E}\Big[|u_{t+h}|^\gamma \Big| u_{t-1}\Big]\Big)^{1/\gamma},$$

which is finite almost surely for any $h \geq 0$ whenever $1 \leq \gamma < 2\alpha + 1$ by GZ (Proposition 3.2) since (u_t) is a noncausal AR(1). For $\gamma \in (0, 1)$, we proceed similarly using the inequality $|a + b|^\gamma \leq |a|^\gamma + |b|^\gamma$, for any $(a, b) \in \mathbb{R}$. We now turn to the conditional expectation of X_{t+h}. We have by the independence between u_{t-1} and $(v_{t-2}, \ldots, v_{t-q-1})$

$$\begin{aligned}
\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \ldots, X_{t-q-1}\Big] &= P_h(B)X_{t-1} + \sum_{i=0}^h q_{i,h} \mathbb{E}\Big[u_{t+i} \Big| X_{t-1}, \ldots, X_{t-q-1}\Big] \\
&= P_h(B)X_{t-1} + \sum_{i=0}^h q_{i,h} \mathbb{E}\Big[u_{t+i} \Big| u_{t-1}, v_{t-2}, \ldots, v_{t-q-1}\Big] \\
&= P_h(B)X_{t-1} + \sum_{i=0}^h q_{i,h} \mathbb{E}\Big[u_{t+i} \Big| u_{t-1}\Big]. \qquad \text{(D.3)}
\end{aligned}$$

By GZ (Proposition 3.3), we have for any $i \geq 0$,

$$\mathbb{E}\Big[u_{t+i} \Big| u_{t-1}\Big] = \Big(\psi^{<\alpha-1>}\Big)^{i+1} u_{t-1},$$

and therefore,

$$\begin{aligned}
\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \ldots, X_{t-q-1}\Big] &= P_h(B)X_{t-1} + \psi^{<\alpha-1>} u_{t-1} \sum_{i=0}^h q_{i,h} \Big(\psi^{<\alpha-1>}\Big)^i \\
&= \Big(P_h(B) + \psi^{<\alpha-1>} Q_h(\psi^{<\alpha-1>})\phi(B)\Big) X_{t-1} \\
&:= \mathscr{P}_h(B) X_{t-1}.
\end{aligned}$$

To conclude, we invoke the fact that (X_t) is a Markov chain of order $q+1$, which gives the equality $\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \ldots, X_{t-q-1}\Big] = \mathbb{E}\Big[X_{t+h} \Big| \mathscr{F}_{t-1}\Big]$. The formula for $h = 0$ is obtained by noting that $P_0(B) = \sum_{i=0}^{q-1} \phi_{i+1} B^i$ and $Q_0(F) = I$.

E Proof of Corollary 3.2

We will derive the asymptotic behaviour of $\mathscr{P}_h(B)X_{t-1} = \Big(P_h(B) + \psi^{<\alpha-1>} Q_h(\psi^{<\alpha-1>})\phi(B)\Big) X_{t-1}$ when $1 \leq \alpha < 2$. We start by a result giving details about the behaviours of the coefficients of the polynomials P_h and Q_h defined in Lemma D.1. Denote $P_h(z) := \sum_{i=0}^{q-1} a_{i,h} z^i$ and $Q_h(z) := \sum_{i=0}^{h} b_{i,h} z^i$.

Lemma E.1 *For $h \geq q$, the coefficients of polynomial P_h and Q_h verify:*

$$a_{0,h} = C_1(h)\lambda_1^h + \ldots + C_s(h)\lambda_s^h, \qquad a_{i,h} = \sum_{j=0}^{q-i-1} a_{0,h-j-1}\phi_{i+1+j}, \quad \textit{for} \quad 0 \leq i \leq q-1,$$

$$b_{i,h} = a_{0,h-i-1}, \quad \textit{for} \quad 0 \leq i \leq h, \qquad a_{0,-1} := 1,$$

where the $\lambda_1, \ldots, \lambda_s$ are the distinct (inverse of the) roots with multiplicities $m_1, \ldots, m_s$ of ϕ and $C_1, \ldots, C_s$ are polynomials with degrees $m_1 - 1, \ldots, m_s - 1$.

The proof is relegated to Appendix B.

The proof of Corollary 3.2 involves several steps.

i) Equivalent of $a_{0,h}$

Without loss of generality we can assume that the (inverses of the) roots of $\phi(z)$ are ordered: $0 < |\lambda_s| < \cdots < |\lambda_1| < 1$. For ease of notation, we drop the indexes of the largest root (in modulus) λ_1 and m_1 and we will denote also by C the coefficient associated to the monomial of highest degree of C_1. We thus have

$$a_{0,h} \underset{h\to+\infty}{\sim} Ch^{m-1}\lambda^h, \quad \text{and} \quad |a_{0,h}| \underset{h\to+\infty}{\longrightarrow} 0. \tag{E.1}$$

ii) Limit of $P_h(B)X_{t-1}$

From Lemma E.1, it appears that $P_h(B)X_{t-1} = \sum_{i=0}^{q-1} a_{i,h} X_{t-i-1} \underset{h\to+\infty}{\overset{a.s.}{\longrightarrow}} 0$.

iii) Limit of $Q_h(\psi^{<\alpha-1>})$

$$Q_h(\psi^{<\alpha-1>}) = \sum_{i=0}^{h} a_{0,h-i-1}(\psi^{<\alpha-1>})^i = \left(\psi^{<\alpha-1>}\right)^{h-1}\Big[\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\Big].$$

Let us study the general term of the above series. We have

$$a_{0,i}(\psi^{<1-\alpha>})^i \underset{i\to+\infty}{\sim} Ci^{m-1}\lambda^i(\psi^{<1-\alpha>})^i = C\text{sign}(\lambda\psi)^i i^{m-1}(|\lambda||\psi|^{1-\alpha})^i. \tag{E.2}$$

Different cases arise.

ι) Assume $\alpha = 1$. According to Equation (E.2) for $\alpha = 1$, $|a_{0,i}(\psi^{<1-\alpha>})^i| \sim |C|i^{m-1}|\lambda|^i$ which is the general term of an absolutely convergent series. Thus, $|Q_h(\psi^{<\alpha-1>})| = |Q_h(\text{sign}(\psi))| = \left|\text{sign}(\psi) + \sum_{i=0}^{h-1} a_{0,i}\text{sign}(\psi)^i\right| \underset{i\to+\infty}{\longrightarrow} D$, for some $D \geq 0$.

$\iota\iota$) Assume $1 < \alpha < 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$.

Then $|Q_h(\psi^{<\alpha-1>})| = |\psi|^{(\alpha-1)(h-1)} \left||\psi|^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| \underset{i\to+\infty}{\longrightarrow} 0 \cdot D = 0.$

$\iota\iota\iota$) Assume $\alpha = 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$.

For $i \geq q$, there exists a positive constant A such that

$$|a_{0,i}| = \left|\sum_{j=1}^{q} C_j(i)\lambda_j^i\right| \leq A i^m |\lambda|^i. \tag{E.3}$$

Thus, since $|\lambda||\psi|^{1-\alpha} = 1$,

$$\begin{aligned}
|\psi|^{(\alpha-1)(h-1)} \left|\sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| &\leq A|\psi|^{(\alpha-1)(h-1)} \sum_{i=0}^{h-1} i^m |\lambda|^i |\psi|^{(1-\alpha)i} \\
&\leq A|\psi|^{(\alpha-1)(h-1)} h^{m+1} \underset{h\to+\infty}{\longrightarrow} 0.
\end{aligned}$$

ν) Assume $\alpha > 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$.

From Equation (E.3),

$$\begin{aligned}
|\psi|^{(\alpha-1)(h-1)} \left|\sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| &\leq A|\psi|^{(\alpha-1)(h-1)} \sum_{i=0}^{h-1} i^m |\lambda|^i |\psi|^{(1-\alpha)i} \\
&\leq A|\psi|^{(\alpha-1)(h-1)} h^m \frac{1 - |\lambda|^h |\psi|^{(1-\alpha)h}}{1 - |\lambda||\psi|^{1-\alpha}} \\
&\leq \frac{A h^m |\psi|^{1-\alpha}}{1 - |\lambda||\psi|^{1-\alpha}} \left(|\psi|^{(\alpha-1)h} - |\lambda|^h\right) \qquad \underset{h\to+\infty}{\longrightarrow} 0.
\end{aligned}$$

The proof of the diverging conditional expectation in the MAR$(1,q)$ case with $\alpha \in (0,1)$ is provided in Appendix B.

F Proof of Proposition 4.1

The $\rho(h)$'s are only function of the AR coefficients and coincide with the theoretical autocorrelations of the process $\sum_{k=-\infty}^{\infty} d_k Z_{t-k}$, where (Z_t) is an i.i.d. noise (with finite variance). Thus, the $\rho(h)$'s are the theoretical autocorrelations of the stationary solution (Y_t) of the AR model $\psi_0(F)\phi_0(B)Y_t = Z_t$. We know from Brockwell and Davis (1991, Proposition 3.5.1) that (Y_t) satisfies the causal AR model $\psi_0(B)\phi_0(B)Y_t = Z_t^*$, for some white noise sequence (Z_t^*), from which the recursion on the coefficients $\rho(h)$ is deduced. The conclusion follows.

G Proof of Proposition 4.2

For $h \geq 0$, let $\hat{\gamma}(h) = \sum_{t=0}^{n-h} X_t X_{t+h}$ denote the (mean-unadjusted) sample autocovariance of order h and let $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$ denote the corresponding sample autocorrelation. The LS estimator of $\boldsymbol{\eta}_0$ coincides, up to negligible terms, with the Yule-Walker estimator and is given by

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Gamma}}_n^{-1} \hat{\boldsymbol{\gamma}}_n, \quad \hat{\boldsymbol{\Gamma}}_n = [\hat{\gamma}(i-j)]_{i,j=1,\ldots,p+q}, \quad \hat{\boldsymbol{\gamma}}_n = [\hat{\gamma}(i)]_{i=1,\ldots,p+q}. \tag{G.1}$$

The consistency of $\hat{\boldsymbol{\eta}}$ follows from Davis and Resnick (1986, Section 5.4).

H Proof of Proposition 4.3

Let $\hat{\boldsymbol{\rho}} = [\hat{\rho}(i)]_{i=1,\ldots,p+q}$, $\hat{\boldsymbol{R}} = [\hat{\rho}(i-j)]_{i,j=1,\ldots,p+q}$. In view of (G.1) and (4.4), we have $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{R}}^{-1} \hat{\boldsymbol{\rho}}$ and $\boldsymbol{\eta}_0 = \boldsymbol{R}^{-1}\boldsymbol{\rho}$. We have

$$\begin{aligned} \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 &= \hat{\boldsymbol{R}}^{-1}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) + (\hat{\boldsymbol{R}}^{-1} - \boldsymbol{R}^{-1})\boldsymbol{\rho} \\ &= \hat{\boldsymbol{R}}^{-1} \left\{ (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) + (\boldsymbol{R} - \hat{\boldsymbol{R}})\boldsymbol{R}^{-1}\boldsymbol{\rho} \right\}. \end{aligned} \tag{H.1}$$

We have $\boldsymbol{R} - \hat{\boldsymbol{R}} = \sum_{i=1}^{p+q} \{\rho(i) - \hat{\rho}(i)\} \boldsymbol{K}^{(i)}$. It follows that

$$(\boldsymbol{R} - \hat{\boldsymbol{R}})\boldsymbol{R}^{-1}\boldsymbol{\rho} = -\boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}). \tag{H.2}$$

Thus, since $\hat{\boldsymbol{R}}^{-1} \to \boldsymbol{R}^{-1}$ in probability as $n \to \infty$, $\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has the same asymptotic distribution as $\boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\} \frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$. The convergence in distribution in (4.8) is a direct consequence of Davis and Resnick (1986) who showed that $\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \xrightarrow{d} \mathbf{Z}$.

I Proof of Proposition 4.4

Write, for $t = p+q+1, \ldots, n$,

$$\hat{\zeta}_t = -\sum_{i=0}^{p+q} \eta_{0i} X_{t-i} - \sum_{i=1}^{p+q} (\hat{\eta}_i - \eta_{0i}) X_{t-i} = -\sum_{i=0}^{p+q} \eta_{0i} X_{t-i} - (\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)' \boldsymbol{X}_{t-1},$$

with $\eta_{00} = -1$ and $\boldsymbol{X}_{t-1} = (X_{t-1}, \ldots, X_{t-p-q})'$. Hence

$$\begin{aligned} \tilde{a}_n^{-1} a_n^2 \hat{\rho}_{\hat{\zeta}}(h) &= \frac{\tilde{a}_n^{-1} a_n^2}{\hat{\gamma}_{\hat{\zeta}}(0)} \sum_{t=p+q+1}^{n} \left\{ \sum_{i,j=0}^{p+q} \eta_{0i}\eta_{0j} X_{t-i} X_{t-h-j} \right. \\ &\quad \left. + (\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)' \sum_{i=0}^{p+q} \eta_{0i} (X_{t-i} \boldsymbol{X}_{t-h-1} + X_{t-h-i} \boldsymbol{X}_{t-1}) \right\} + o_P(1), \end{aligned}$$

with by convention $X_s = 0$ for $s \leq 0$. Let the $(p+q+1)\times(p+q+1)$ matrices $\hat{\boldsymbol{R}}_{\boldsymbol{h}} = [\hat{\rho}(h+i-j)]_{i,j=0,\dots,p+q}$, $\boldsymbol{R}_{\boldsymbol{h}} = [\rho(h+i-j)]_{i,j=0,\dots,p+q}$, and for any strictly positive integers, m, m' such that $m \leq m'$, let $\hat{\boldsymbol{\rho}}_{m:m'} = [\hat{\rho}(i)]_{i=m,\dots,m'}$ and $\boldsymbol{\rho}_{m:m'} = [\rho(i)]_{i=m,\dots,m'}$. Then,

$$
\begin{aligned}
\sum_{t=p+q+1}^{n}\sum_{i,j=0}^{p+q}\eta_{0i}\eta_{0j}X_{t-i}X_{t-h-j} &= \hat{\gamma}(0)\sum_{i,j=0}^{p+q}\eta_{0i}\eta_{0j}\hat{\rho}(h+j-i)+o_P(1)\\
&= \hat{\gamma}(0)\sum_{i,j=0}^{p+q}\eta_{0i}\eta_{0j}\{\hat{\rho}(h+j-i)-\rho(h+j-i)\}+o_P(1)\\
&= \hat{\gamma}(0)\boldsymbol{\eta}_0'\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}}-\boldsymbol{R}_{\boldsymbol{h}}\Big)\boldsymbol{\eta}_0+o_P(1)\\
&= \hat{\gamma}(0)\Big(\boldsymbol{\eta}_0'\otimes\boldsymbol{\eta}_0'\Big)\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}}-\boldsymbol{R}_{\boldsymbol{h}}\Big),
\end{aligned}
$$

where the second equality follows from Proposition 4.4. Moreover,

$$
\begin{aligned}
&\sum_{t=p+q+1}^{n}(\hat{\boldsymbol{\eta}}_n-\boldsymbol{\eta}_0)'\sum_{i=0}^{p+q}\eta_{0i}(X_{t-i}\boldsymbol{X}_{t-h-1}+X_{t-h-i}\boldsymbol{X}_{t-1})\\
&\qquad= \hat{\gamma}(0)\sum_{i=0}^{p+q}\sum_{j=1}^{p+q}(\hat{\eta}_{nj}-\eta_{0j})\eta_{0i}\Big(\hat{\rho}(h+j-i)+\hat{\rho}(h+i-j)\Big)+o_P(1)\\
&\qquad= \hat{\gamma}(0)\boldsymbol{\eta}_0'\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}}+\hat{\boldsymbol{R}}_{\boldsymbol{h}}'\Big)(\hat{\boldsymbol{\eta}}_n-\boldsymbol{\eta}_0)+o_P(1).
\end{aligned}
$$

Let the $(p+q+1)\times(p+q+1)$ matrices $\boldsymbol{D}_i = J^i$ and $\boldsymbol{D}_{-i} = {}^{\mathbf{t}}\boldsymbol{J}^{\boldsymbol{i}}$ for $i \geq 0$. We have:

$$
\begin{aligned}
\hat{\boldsymbol{R}}_{\boldsymbol{h}}-\boldsymbol{R}_{\boldsymbol{h}} &= \sum_{i=1}^{p+q-h}\Big(\hat{\rho}(i)-\rho(i)\Big)\Big(\boldsymbol{D}_{h-i}+\boldsymbol{D}_{h+i}\Big)\\
&\qquad + \sum_{i=p+q-h+1}^{h+p+q}\Big(\hat{\rho}(i)-\rho(i)\Big)\boldsymbol{D}_{h-i}, \qquad \text{if} \quad 1\leq h\leq p+q-1,\\
\hat{\boldsymbol{R}}_{\boldsymbol{h}}-\boldsymbol{R}_{\boldsymbol{h}} &= \sum_{i=h-p-q}^{h+p+q}\Big(\hat{\rho}(i)-\rho(i)\Big)\boldsymbol{D}_{h-i}, \qquad \text{if} \quad h\geq p+q.
\end{aligned}
$$

Thus, with

$$
\begin{aligned}
\boldsymbol{L}_{\boldsymbol{h}} &= \Big[\mathrm{vec}(\boldsymbol{D}_{h-1}+\boldsymbol{D}_{h+1})\dots\mathrm{vec}(\boldsymbol{D}_{2h-p-q}+\boldsymbol{D}_{p+q})\ \ \mathrm{vec}(\boldsymbol{D}_{2h-p-q-1})\dots\mathrm{vec}(\boldsymbol{D}_{-p-q})\Big], \quad \text{if} \quad 1\leq h\leq p+q,\\
\boldsymbol{L}_{\boldsymbol{h}} &= \Big[\mathrm{vec}(\boldsymbol{D}_{p+q})\dots\mathrm{vec}(\boldsymbol{D}_{-p-q})\Big], \quad \text{if} \quad h\geq p+q,
\end{aligned}
$$

we can write

$$
\begin{aligned}
\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}}-\boldsymbol{R}_{\boldsymbol{h}}\Big) &= \boldsymbol{L}_{\boldsymbol{h}}\Big(\hat{\boldsymbol{\rho}}_{1:h+p+q}-\boldsymbol{\rho}_{1:h+p+q}\Big), \qquad \text{if} \quad 1\leq h\leq p+q,\\
\mathrm{vec}\Big(\hat{\boldsymbol{R}}_{\boldsymbol{h}}-\boldsymbol{R}_{\boldsymbol{h}}\Big) &= \boldsymbol{L}_{\boldsymbol{h}}\Big(\hat{\boldsymbol{\rho}}_{h-p-q:h+p+q}-\boldsymbol{\rho}_{h-p-q:h+p+q}\Big), \qquad \text{if} \quad h\geq p+q+1.
\end{aligned}
$$

The two last expressions point to the fact that $\left(\hat{\rho}_{\hat{\zeta}}(h)\right)_{h=1,\ldots,H}$ will depend on $\left(\hat{\boldsymbol{\rho}}(i) - \boldsymbol{\rho}(i)\right)_{i=1,\ldots,H+p+q}$. We therefore rewrite $\mathrm{vec}\left(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}}\right)$ as

$$\mathrm{vec}\left(\hat{\boldsymbol{R}}_{\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}}\right) = \boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}}\left(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\right),$$

with $\boldsymbol{M}_{\boldsymbol{h}}$ being the matrix of size $(h+p+q)\times(H+p+q)$ if $0 \le h \le p+q$ and $(2(p+q))\times(H+p+q)$ if $h \ge p+q+1$ picking the appropriate components of $\left(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\right)$. More explicitly,

$$\begin{aligned}
\boldsymbol{M}_{\boldsymbol{h}} &= \left(\boldsymbol{I}_{h+p+q} \;\; \boldsymbol{0}_{h+p+q\times H-h}\right), && \text{if} \quad 0 \le h \le p+q,\\
\boldsymbol{M}_{\boldsymbol{h}} &= \left(\boldsymbol{0}_{2(p+q)+1\times h-p-q-1} \;\; \boldsymbol{I}_{2(p+q)+1} \;\; \boldsymbol{0}_{2(p+q)+1\times H-h}\right), && \text{if} \quad h \ge p+q+1.
\end{aligned}$$

Thus, using equations (H.1) and (H.2),

$$\tilde{a}_n^{-1}a_n^2\hat{\rho}_{\hat{\zeta}}(h) = \tilde{a}_n^{-1}a_n^2\frac{\hat{\gamma}(0)}{\hat{\gamma}_{\hat{\zeta}}(0)}\left[\left(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\right)\boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}} + \boldsymbol{\eta}_0'\left(\hat{\boldsymbol{R}}_{\boldsymbol{h}} + \hat{\boldsymbol{R}}_{\boldsymbol{h}}'\right)\hat{\boldsymbol{P}}\right]\left(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\right) + o_P(1),$$

with $\hat{\boldsymbol{P}} := \begin{pmatrix}\boldsymbol{0}_{1\times p+q}\\ \boldsymbol{I}_{p+q}\end{pmatrix}\hat{\boldsymbol{R}}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q}\otimes\boldsymbol{R}^{-1}\boldsymbol{\rho})\}\boldsymbol{M}_{\boldsymbol{0}}$.

Finally, letting $\hat{\boldsymbol{A}}_{\boldsymbol{H}} = \left[\left(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\right)\boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}} + \boldsymbol{\eta}_0'\left(\hat{\boldsymbol{R}}_{\boldsymbol{h}} + \hat{\boldsymbol{R}}_{\boldsymbol{h}}'\right)\hat{\boldsymbol{P}}\right]_{h=1,\ldots,H}$ denote the matrix resulting from the vertical piling of vectors, we have

$$\frac{a_n^2}{\tilde{a}_n}\hat{\boldsymbol{\rho}}_{\hat{\zeta}} = \hat{\boldsymbol{A}}_{\boldsymbol{H}}\frac{a_n^{-2}\hat{\gamma}(0)}{a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0)}\tilde{a}_n^{-1}a_n^2\left(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\right) + o_P(1).$$

By Theorem 4.2 by Davis and Resnick (1985), Theorem 4.4 by Davis and Resnick (1986) and Lemma I.2 below, $\hat{\boldsymbol{P}} \xrightarrow{p} \boldsymbol{P} := \begin{pmatrix}\boldsymbol{0}_{1\times p+q}\\ \boldsymbol{I}_{p+q}\end{pmatrix}\boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q}\otimes\boldsymbol{R}^{-1}\boldsymbol{\rho})\}\boldsymbol{M}_{\boldsymbol{0}}$,

$\hat{\boldsymbol{A}}_{\boldsymbol{H}} \xrightarrow{p} \left[\left(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\right)\boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}} + \boldsymbol{\eta}_0'\boldsymbol{R}_{\boldsymbol{h}}'\boldsymbol{P}\right]_{h=1,\ldots,H} := \boldsymbol{A}_{\boldsymbol{H}}$ and $\hat{\boldsymbol{\rho}}_{\hat{\zeta}} \xrightarrow{d} \gamma(0)\boldsymbol{A}_{\boldsymbol{H}}\boldsymbol{Z}$ where $\boldsymbol{Z} = (Z_1,\ldots,Z_{H+p+q})$, and where the (Z_i) are defined at Proposition 4.3.

Lemma I.1 *Under the assumptions of Proposition 4.4,* $a_n^{-2}\left(\hat{\gamma}(h) - \gamma(h)\hat{\gamma}_{\zeta}(0)\right) \xrightarrow{p} 0$.

Lemma I.2 *Under the assumptions of Proposition 4.4,* $a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0) = a_n^{-2}\dfrac{\hat{\gamma}(0)}{\gamma(0)} + o_P(1)$.

I.1 Proof of Lemma I.1

We have

$$\hat{\gamma}(h) = \sum_{t=1}^{n} X_tX_{t-h} = \sum_{t=1}^{n}\sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} d_id_j\varepsilon_{t+i}\varepsilon_{t+j-h} = \sum_{t=1}^{n}\sum_{i\in\mathbb{Z}}\sum_{j\neq i} d_id_{j+h}\varepsilon_{t+i}\varepsilon_{t+j} + \sum_{t=1}^{n}\sum_{i\in\mathbb{Z}} d_id_{i+h}\varepsilon_{t+i}^2.$$

From Proposition 4.2 by Davis and Resnick (1986), we have

$$a_n^{-2}\sum_{t=1}^{n}\sum_{i\in\mathbb{Z}}\sum_{j\neq i} d_i d_{j+h}\varepsilon_{t+i}\varepsilon_{t+j} \xrightarrow{p} 0. \tag{I.1}$$

A direct extension of Proposition 4.3.ii by Davis and Resnick (1986) (see also the proof of Proposition 4.3 by GZ in the AR(1) case) yields

$$a_n^{-2}\left(\sum_{t=1}^{n}\sum_{i\in\mathbb{Z}} d_i d_{i+h}\varepsilon_{t+i}^2 - \gamma(h)\sum_{t=1}^{n}\varepsilon_t^2\right) \xrightarrow{p} 0. \tag{I.2}$$

Combining equations (I.1) and (I.2), we get $a_n^{-2}\Big(\hat{\gamma}(h)-\gamma(h)\hat{\gamma}_\zeta(0)\Big) \xrightarrow{p} 0$.

I.2 Proof of Lemma I.2

$$\begin{aligned}
a_n^{-2}\sum_{t=1}^{n}\hat{\zeta}_t^2 &= a_n^{-2}\sum_{t=1}^{n}\left(X_t-\sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}X_{t-i}+\sum_{i=1}^{p+q}\Big(\hat{\boldsymbol{\eta}}_i-\boldsymbol{\eta}_{0i}\Big)X_{t-i}\right)^2\\
&= a_n^{-2}\sum_{t=1}^{n}\Bigg[\left(X_t-\sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}X_{t-i}\right)^2+2\sum_{i=1}^{p+q}\Big(\hat{\boldsymbol{\eta}}_i-\boldsymbol{\eta}_{0i}\Big)\left(X_tX_{t-i}-\sum_{j=1}^{p+q}\boldsymbol{\eta}_{0j}X_{t-i}X_{t-j}\right)\\
&\qquad+\sum_{i=1}^{p+q}\sum_{i=1}^{p+q}\Big(\hat{\boldsymbol{\eta}}_i-\boldsymbol{\eta}_{0i}\Big)\Big(\hat{\boldsymbol{\eta}}_j-\boldsymbol{\eta}_{0j}\Big)X_{t-i}X_{t-j}\Bigg]\\
&= a_n^{-2}\Bigg[\hat{\gamma}(0)-\sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}\hat{\gamma}(-j)-\sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}\left(\hat{\gamma}(i)-\sum_{j=1}^{p+q}\boldsymbol{\eta}_{0j}\hat{\gamma}(i-j)\right)\\
&\qquad+\sum_{i=1}^{p+q}\sum_{i=1}^{p+q}\Big(\hat{\boldsymbol{\eta}}_i-\boldsymbol{\eta}_{0i}\Big)\Big(\hat{\boldsymbol{\eta}}_j-\boldsymbol{\eta}_{0j}\Big)\hat{\gamma}(i-j)\Bigg].
\end{aligned}$$

Using Lemma I.1, the fact that $\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0 \longrightarrow 0$ in probability and the convergence in distribution of the vector $a_n^{-2}\Big(\hat{\gamma}(i),\quad 0\leq i\leq L\Big)$ for any integer L, we get:

$$a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0)=a_n^{-2}\hat{\gamma}_\zeta(0)\left[\gamma(0)-\sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}\gamma(-j)-\sum_{i=1}^{p+q}\boldsymbol{\eta}_{0i}\left(\gamma(i)-\sum_{j=1}^{p+q}\boldsymbol{\eta}_{0j}\gamma(i-j)\right)\right]+o_P(1).$$

From Proposition 4.1, we have that $\boldsymbol{\eta}_0(B)\gamma(i)=0$ for any $i\geq 1$ and $\boldsymbol{\eta}_0(B)\gamma(0)=1$. Thus

$$a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0)=a_n^{-2}\hat{\gamma}_\zeta(0)+o_P(1)=a_n^{-2}\frac{\hat{\gamma}(0)}{\gamma(0)}+o_P(1).$$

Appendix B: Complementary results

This Appendix consists of six sections of additional results: J) asymptotic prediction of the MAR$(1, q)$ when $\alpha \in (0,1)$ and an explicit example in the MAR$(1,1)$ case; K) expectation of MAR(p, q) processes conditionally on a linear combination of past values and proof of the unit root property; L) conditional correlation structure of noncausal AR(1) processes and proofs of Proposition 3.4 and Example 3.2; M) proof of Lemma E.1; N) recursion over polynomials P_h and Q_h; O) tail probability of cluster sizes of extreme errors (4.20) and (4.21) at various horizons; P) complementary results on the empirical study and details about the estimation of excess clustering term structures.

J A complement to Corollary 3.2 in the case $\alpha \in (0,1)$ and $q > 1$

Under the conditions of Proposition 3.2, when $\alpha \in (0,1)$, we have almost surely

$$\left|\mathbb{E}\left[X_{t+h}\middle|\mathscr{F}_{t-1}\right]\right| \underset{h\to+\infty}{\longrightarrow} \begin{cases} 0 & if \quad \psi^{<\alpha-1>} + \sum_{i=0}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i = 0, \\ +\infty & else, \end{cases}$$

where the $a_{0,i}$'s are defined in Lemma E.1.

Proof.

To complete the proof of Corollary 3.2 in this case, we will derive the limit of $Q_h(\psi^{<\alpha-1>}) = \left(\psi^{<\alpha-1>}\right)^{h-1}\left[\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right]$ when $\alpha < 1$. Recall that we have shown $a_{0,h} \underset{h\to+\infty}{\sim} Ch^{m-1}\lambda^h$.

In this case, we have $|\lambda||\psi|^{1-\alpha} < 1$, thus $\left|\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\right| \underset{i\to+\infty}{\longrightarrow} D$, where D is a nonnegative constant.

- Assume $D > 0$. Then $|Q_h(\psi^{<\alpha-1>})| \to +\infty$ as h tends to infinity, since $|\psi|^{(\alpha-1)(h-1)} \to +\infty$.
- Assume $D = 0$. We will show that $|Q_h(\psi^{<\alpha-1>})| \longrightarrow 0$.

  Indeed, we have

$$\psi^{<\alpha-1>} + \sum_{i=0}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i = 0$$
$$\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i = -\sum_{i=h}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i.$$

Thus,

$$\begin{aligned}|Q_h(\psi^{<\alpha-1>})| &= |\psi|^{(\alpha-1)(h-1)}\Big|\psi^{<\alpha-1>} + \sum_{i=0}^{h-1} a_{0,i}(\psi^{<1-\alpha>})^i\Big| \\ &= |\psi|^{(\alpha-1)(h-1)}\Big|\sum_{i=h}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i\Big| \\ &\leq |\psi|^{(\alpha-1)(h-1)}\sum_{i=h}^{+\infty} |a_{0,i}||\psi|^{i(1-\alpha)},\end{aligned}$$

and

$$\sum_{i=h}^{+\infty} |a_{0,i}||\psi|^{i(1-\alpha)} \underset{h\to+\infty}{\sim} |C|\sum_{i=h}^{+\infty} i^{m-1}(|\lambda||\psi|^{1-\alpha})^i.$$

We will show that for any $x \in (0,1)$, and any integer $r \geq 0$,

$$\sum_{i=h}^{+\infty} i^r x^i \underset{h\to+\infty}{\sim} h^r x^h (1-x)^{-1}, \tag{J.1}$$

which will imply

$$|\psi|^{(\alpha-1)(h-1)}\sum_{i=h}^{+\infty} |a_{0,i}||\psi|^{i(1-\alpha)} \underset{h\to+\infty}{=} O(h^{m-1}|\lambda|^h),$$

and thus $|Q_h(\psi^{<\alpha-1>})| \longrightarrow 0$, yielding the conclusion.

Let us now prove Equation (J.1). Notice that for $x \in (0,1)$, the sequences $\Big(i^r x^i\Big)_i$ and $\Big(i(i-1)\dots(i-r+1)x^i\Big)_i$ are equivalent as i tends to infinity and are both general terms of absolutely convergent series. Thus,

$$\sum_{i=h}^{+\infty} i^r x^i \underset{h\to+\infty}{\sim} \sum_{i=h}^{+\infty} i(i-1)\dots(i-r+1)x^i = x^r g^{(r)}(x),$$

where $g(x) := \sum_{i=h}^{+\infty} x^i = x^h(1-x)^{-1}$.

By Leibniz formula, we obtain

$$g^{(r)}(x) = \sum_{j=0}^{r} \frac{h!(r-j)!}{(h-j)!}\frac{x^{h-j}}{(1-x)^{r-j+1}} \underset{h\to+\infty}{\sim} \frac{h^r x^{h-r}}{1-x},$$

and thus,

$$\sum_{i=h}^{+\infty} i^r x^i \underset{h\to+\infty}{\sim} x^r\frac{h^r x^{h-r}}{1-x} = \frac{h^r x^h}{1-x}.$$

Substituting x by $|\lambda||\psi|^{1-\alpha}$ concludes the proof.

In the case $\alpha \in (0,1)$, i.e. for the heavier tails within the stable family, the absolute conditional expectation tends to $+\infty$ in modulus whenever the quantity $\psi^{<\alpha-1>} + \sum_{i=0}^{+\infty} a_{0,i}(\psi^{<1-\alpha>})^i$ does not vanish. This divergence is coherent with the fact that the unconditional expectation of (X_t)

does not exist when $\alpha < 1$. It would be striking to have a case for which the above quantity is exactly zero, which would imply that the conditional expectation vanishes even for this class of particularly extreme processes. However, as the following example shows, all MAR(1,1) feature diverging conditional expectation when $\alpha < 1$.

Example J.1 (Asymptotic predictions of the MAR(1,1) process) Let (X_t) be defined by Equation (2.3). From the explicit predictions formulated in Example 3.1, we deduce the asymptotic equivalents as the horizon h tends to infinity:

$$\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big] \underset{h\to+\infty}{\overset{a.s.}{\sim}} \begin{cases} \dfrac{(\psi^{<\alpha-1>})^{h+1}}{1-\psi^{<1-\alpha>}\phi}\left(X_{t-1}-\phi X_{t-2}\right), & \text{if} \quad |\phi| \quad < |\psi|^{\alpha-1}, \\ \dfrac{\phi^{h+2}}{\phi-\psi^{<\alpha-1>}}\Big(X_{t-1}-\psi^{<\alpha-1>}X_{t-2}\Big), & \text{if} \quad |\phi| \quad > |\psi|^{\alpha-1}, \\ \phi^{h+1}\left(X_{t-1}-\dfrac{1+(-1)^h}{2}(X_{t-1}-\phi X_{t-2})\right), & \text{if} \quad \phi \quad = -\psi^{<\alpha-1>}, \\ (h+1)\phi^{h+1}(X_{t-1}-\phi X_{t-2}), & \text{if} \quad \phi \quad = \psi^{<\alpha-1>}. \end{cases}$$

Noticing that the condition $|\phi| < |\psi|^{\alpha-1}$ is equivalent to $\alpha < 1 + \frac{\ln|\phi|}{\ln|\psi|}$, with $\frac{\ln|\phi|}{\ln|\psi|} > 0$, it can be seen that the three asymptotic limits of Corollary 3.2 are consistent with these equivalents. In particular, when $\alpha = 1$, we always have $|\phi| < 1 = |\psi|^{\alpha-1}$ and we get that, almost surely,

$$\Big|\mathbb{E}\left[X_{t+h}\right|X_{t-1},X_{t-2}]\Big| \underset{h\to+\infty}{\longrightarrow} \ell_{t-1} = \left|\frac{X_{t-1}-\phi X_{t-2}}{1-\mathrm{sign}(\psi)\phi}\right|.$$

K Extension of the unit root property

In the following result, we derive the expectation of X_t conditionally on any linear combination of the past. The following Proposition will also yield the unit root property of Proposition 3.3.

Proposition K.1 *Let $\alpha > 1$ and let X_t be a MAR(p, q) process with MA(∞) representation given by* (2.2). *Then for any $h \geq 0$, $k \geq 1$, and $a_1, \dots, a_k$ such that there exists $\ell \in \mathbb{Z}$, $a_1 d_{\ell+1} + \dots + a_k d_{\ell+k} \neq 0$, we have*

$$\mathbb{E}\Big[X_{t+h}\Big|\sum_{j=1}^k a_j X_{t-j}\Big] = \frac{\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left(\sum_{j=1}^k a_j d_{\ell+j}\right)^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^k a_j d_{\ell+j}\right|^{\alpha}}(a_1X_{t-1}+\dots+a_kX_{t-k}). \tag{K.1}$$

Proposition 3.3 is obtained for $k=1$, $a_1 = 1$.

Proof.

Let us introduce $Y_{t-1,k} = a_1X_{t-1} + \dots + a_kX_{t-k}$. Let $\varphi(u,v) = \mathbb{E}\left[e^{iuY_{t-1,k}+ivX_{t+h}}\right]$. If $\alpha > 1$, for any $(u,v) \in \mathbb{R}^2$ we have,

$$\begin{aligned}\varphi(u,v) &= \mathbb{E}\left[\exp\left\{iu\sum_{j=1}^{k} a_j \sum_{\ell\in\mathbb{Z}} d_\ell \varepsilon_{t+\ell-j} + v\sum_{\ell\in\mathbb{Z}} d_\ell \varepsilon_{t+\ell+h}\right\}\right] \\ &= \mathbb{E}\left[\exp\left\{iu\sum_{\ell\in\mathbb{Z}}\left(u\sum_{j=1}^{k} a_j d_{\ell+j} + v d_{\ell-h}\right)\varepsilon_{t+\ell}\right\}\right] \\ &= \exp\left\{-\sigma^\alpha \sum_{\ell\in\mathbb{Z}}\left|u\sum_{j=1}^{k} a_j d_{\ell+j} + v d_{\ell-h}\right|^\alpha\right\}.\end{aligned}$$

Thus,

$$\frac{\partial\varphi}{\partial u}(u,v) = -\alpha\sigma^\alpha\varphi(u,v)\sum_{\ell\in\mathbb{Z}}\left(\sum_{j=1}^{k} a_j d_{\ell+j}\right)\left|u\sum_{j=1}^{k} a_j d_{\ell+j} + v d_{\ell-h}\right|^{<\alpha-1>},$$

and

$$\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} = -\alpha\sigma^\alpha |u|^{<\alpha-1>}\varphi(u,0)\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^\alpha.$$

We also have

$$\begin{aligned}\frac{\partial\varphi}{\partial v}(u,v) &= -\alpha\sigma^\alpha\varphi(u,v)\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left|u\sum_{j=1}^{k} a_j d_{\ell+j} + v d_{\ell-h}\right|^{<\alpha-1>}, \\ \left.\frac{\partial\varphi}{\partial v}\right|_{v=0} &= -\alpha\sigma^\alpha |u|^{<\alpha-1>}\varphi(u,0)\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^{<\alpha-1>}.\end{aligned}$$

Therefore,

$$\left.\frac{\partial\varphi}{\partial v}\right|_{v=0} = \frac{\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^\alpha}\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} \tag{K.2}$$

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

$$\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} = i\mathbb{E}\left[Y_{t-1,k}e^{iuY_{t-1,k}}\right], \quad \left.\frac{\partial\varphi}{\partial v}\right|_{v=0} = i\mathbb{E}\left[X_{t+h}e^{iuY_{t-1,k}}\right].$$

Therefore, for $u \in \mathbb{R}^*$:

$$\mathbb{E}\left[\left(X_{t+h} - \frac{\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^\alpha}Y_{t-1,k}\right)e^{iuY_{t-1,k}}\right] = 0. \tag{K.3}$$

Hence, from Bierens (Theorem 1, 1982): Thus

$$\mathbb{E}\left[X_{t+h}\,|\,Y_{t-1,k}\right] = \frac{\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^\alpha}Y_{t-1,k}.$$

The case $\alpha \leq 1$ is more intricate because the expectation on the left-hand side of (K.1) might not exist. However, the conditions for existence can be established using Theorem 2.13 of Samorodnistky, Taqqu (1994). This is left for further research.

Example K.1 (Cauchy AR(2) process (continued)) Consider the noncausal AR(2) solution of $(1-\lambda_1 F)(1-\lambda_2 F)X_t = \varepsilon_t$ with $0 < |\lambda_2| < \lambda_1 < 1$ and $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1,0,\sigma,0)$. Then $X_t = \sum_{k\geq 0}^{+\infty} d_k \varepsilon_t$ where $d_k = \frac{\lambda_1^{k+1}-\lambda_2^{k+1}}{\lambda_1-\lambda_2} > 0$ for all $k \geq 0$. Applying Proposition K.1 to X_{t+h}, $h \geq 0$, with $(a_1,\dots,a_k) \in (\mathbb{R}^+)^k$, $k \geq 1$, such that $a_1+\ldots+a_k \neq 0$, it can be shown that

$$\mathbb{E}\left[X_{t+h} \middle| a_1 X_{t-1} + \ldots + a_k X_{t-k}\right] = \frac{a_1 X_{t-1} + \ldots + a_k X_{t-k}}{a_1 + \ldots + a_k}.$$

L Conditional heteroscedasticity of the MAR$(1,q)$ process

In order to prove Proposition 3.4, we need to show some preliminary results about the conditional covariance of noncausal AR(1) processes. We will then turn to the conditional covariance of MAR$(1,q)$ process from which the conditional variance will be a obtainable.

L.1 Conditional correlation structure of the MAR$(1,q)$

Lemma L.1 *Let X_t be a noncausal AR(1) process satisfying $X_t = \psi X_{t+1} + \varepsilon_t$, with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1,0,\sigma,0)$. Then, for any nonnegative integers h and τ:*

$$\mathbb{E}\Big[X_{t+h}X_{t+h+\tau}\Big|X_{t-1}\Big] = (sign\ \psi)^\tau \left[|\psi|^{-h-1}\left(X_{t-1}^2 + \frac{\sigma^2}{(1-|\psi|)^2}\right) - \frac{\sigma^2}{(1-|\psi|)^2}\right].$$

Remark L.1 From the previous result, it is possible to derive the whole conditional correlation structure of (X_t). It can be shown that for any $t \in \mathbb{Z}$, and any positive integers h and τ:

$$\frac{\mathrm{Cov}\Big(X_{t+h}, X_{t+h+\tau}\Big|X_{t-1}\Big)}{\sqrt{\mathbb{V}\Big(X_{t+h}\Big|X_{t-1}\Big)}\sqrt{\mathbb{V}\Big(X_{t+h+\tau}\Big|X_{t-1}\Big)}} = (\text{sign } \psi)^\tau \sqrt{\frac{|\psi|^{-h-1}-1}{|\psi|^{-h-\tau-1}-1}},$$

which, when $\tau \to +\infty$, is asymptotically equivalent to $(\text{sign } \psi)^\tau |\psi|^{\tau/2}\sqrt{1-|\psi|^{h+1}}$ for any $h \geq 0$, and to $(\text{sign } \psi)^\tau |\psi|^{\tau/2}$ when h becomes large. Although in our infinite variance framework, the unconditional correlation is not defined, empirical correlations can always be computed. We know from Davis and Resnick (1985,1986) that they converge in probability towards the theoretical autocorrelations that would prevail in the L^2 framework. Given n observations of process (X_t), we have for any $\tau \geq 0$,

$$\frac{\sum_{t=1}^{n-\tau+1} X_t X_{t+\tau}}{\sum_{t=1}^{n} X_t^2} \xrightarrow[n\to+\infty]{p} \psi^\tau.$$

Surprisingly, the "unconditional" autocorrelations of (X_t) do not converge to the conditional ones when $n \to +\infty$, and are vanishing at a much slower rate ($|\psi|^{\tau/2}$ instead of $|\psi|^{\tau}$).

We now turn to the MAR$(1, q)$ process.

Proposition L.1 *Let X_t be a MAR$(1,q)$ process, $q \geq 0$, solution of Equation (2.1) with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(1, 0, \sigma, 0)$. Then, for any positive integers h and τ, there exist polynomials P_h, $P_{h+\tau}$, both of degrees $q-1$, and Q_h, $Q_{h+\tau}$ of respective degrees h and $h+\tau$ such that*

$$\begin{aligned}\mathbb{E}\left[X_{t+h}X_{t+h+\tau} \,|\, \mathscr{F}_{t-1}\right] &= (P_h(B)X_{t-1})(P_{h+\tau}(B)X_{t-1}) \\ &\quad + sign(\psi)(\phi(B)X_{t-1})\Big[(P_h(B)X_{t-1})Q_{h+\tau}(sign\ \psi) + (P_{h+\tau}(B)X_{t-1})Q_h(sign\ \psi)\Big] \\ &\quad + c_{h,\tau}\left((\phi(B)X_{t-1})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\right) - \frac{\sigma^2}{(1-|\psi|)^2}Q_h(sign\ \psi)Q_{h+\tau}(sign\ \psi),\end{aligned}$$

with $c_{h,\tau} = \sum_{i=0}^{h+\tau}\sum_{j=0}^{h} q_{i,h+\tau}q_{j,h}(sign\ \psi)^{i+j}|\psi|^{-\min(i,j)-1}$ and $Q_k(z) = \sum_{i=0}^{k} q_{i,k}z^i$, for any $k \geq 0$.

This result yields Proposition 3.3 by taking $h = \tau = 0$, with $P_0(B) = \phi_1 + \phi_2 B + \ldots + \phi_q B^q$ and $Q_0(B) = 1$.

L.2 Proof of Lemma L.1

Consider $\varphi(x,y,z) := \mathbb{E}\Big(e^{ixX_{t+k}+iyX_{t+\ell}+izX_{t-1}}\Big)$, with $0 \leq \ell \leq k$, $X_t = \psi X_{t+1} + \varepsilon_t$ and $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, 0, \sigma, 0)$. We have

$$\varphi(x,y,z) = \mathbb{E}\left(e^{i\sum_{n\in\mathbb{Z}}(xd_{n-k}+yd_{n-\ell}+zd_{n+1})\varepsilon_{t+n}}\right) = \exp\left\{-\sigma^\alpha \sum_{n\in\mathbb{Z}} |xd_{n-k} + yd_{n-\ell} + zd_{n+1}|^\alpha\right\}.$$

Thus, on the one hand,

$$\begin{aligned}\frac{\partial \varphi}{\partial z} &= -\alpha\sigma^\alpha \sum_{n\in\mathbb{Z}} d_{n+1}|xd_{n-k} + yd_{n-\ell} + zd_{n+1}|^{<\alpha-1>}\varphi(x,y,z), \\ \frac{\partial^2 \varphi}{\partial z^2} &= (\alpha\sigma^\alpha)^2\left(\sum_{n\in\mathbb{Z}} d_{n+1}|xd_{n-k} + yd_{n-\ell} + zd_{n+1}|^{<\alpha-1>}\right)^2\varphi(x,y,z) \\ &\qquad - \alpha(\alpha-1)\sum_{n\in\mathbb{Z}} d_{n+1}^2|xd_{n-k} + yd_{n-\ell} + zd_{n+1}|^{\alpha-2}\varphi(x,y,z),\end{aligned}$$

$$\left.\frac{\partial^2 \varphi}{\partial z^2}\right|_{\substack{x=0\\y=0}} = (\alpha\sigma^\alpha)^2|z|^{2(\alpha-1)}\left(\sum_{n\in\mathbb{Z}}|d_{n+1}|^\alpha\right)^2\varphi(0,0,z) - \alpha(\alpha-1)|z|^{\alpha-2}\sum_{n\in\mathbb{Z}}|d_{n+1}|^\alpha\varphi(0,0,z).$$

And on the other hand,

$$\begin{aligned}
\frac{\partial \varphi}{\partial y} &= -\alpha\sigma^\alpha \sum_{n\in\mathbb{Z}} d_{n-\ell}|xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^{<\alpha-1>}\varphi(x,y,z),\\
\frac{\partial^2 \varphi}{\partial x \partial y} &= (\alpha\sigma^\alpha)^2 \left(\sum_{n\in\mathbb{Z}} d_{n-\ell}|xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^{<\alpha-1>}\right)\\
&\qquad\qquad \times \left(\sum_{n\in\mathbb{Z}} d_{n-k}|xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^{<\alpha-1>}\right)\varphi(x,y,z)\\
&\quad - \alpha(\alpha-1)\sum_{n\in\mathbb{Z}} d_{n-\ell}d_{n-k}|xd_{n-k}+yd_{n-\ell}+zd_{n+1}|^{\alpha-2}\varphi(x,y,z),\\
\left.\frac{\partial^2 \varphi}{\partial x \partial y}\right|_{\substack{x=0\\y=0}} &= (\alpha\sigma^\alpha)^2|z|^{2(\alpha-1)}\left(\sum_{n\in\mathbb{Z}} d_{n-\ell}|d_{n+1}|^{<\alpha-1>}\right)\left(\sum_{n\in\mathbb{Z}} d_{n-k}|d_{n+1}|^{<\alpha-1>}\right)\varphi(0,0,z)\\
&\quad - \alpha(\alpha-1)|z|^{\alpha-2}\sum_{n\in\mathbb{Z}} d_{n-\ell}d_{n-k}|d_{n+1}|^{\alpha-2}\varphi(0,0,z).
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{1}{A_2}\left[\left.\frac{\partial^2 \varphi}{\partial x \partial y}\right|_{\substack{x=0\\y=0}} - (\alpha\sigma^\alpha)^2 A_1 |z|^{2(\alpha-1)}\varphi(0,0,z)\right] &= -\alpha(\alpha-1)|z|^{\alpha-2}\varphi(0,0,z),\\
\frac{1}{B}\left[\frac{\partial^2 \varphi}{\partial z^2} - (\alpha\sigma^\alpha)^2 B^2 |z|^{2(\alpha-1)}\varphi(0,0,z)\right] &= -\alpha(\alpha-1)|z|^{\alpha-2}\varphi(0,0,z),
\end{aligned}$$

with

$$\begin{aligned}
A_1 &= \left(\sum_{n\in\mathbb{Z}} d_{n-\ell}|d_{n+1}|^{<\alpha-1>}\right)\left(\sum_{n\in\mathbb{Z}} d_{n-k}|d_{n+1}|^{<\alpha-1>}\right),\\
A_2 &= \sum_{n\in\mathbb{Z}} d_{n-\ell}d_{n-k}|d_{n+1}|^{\alpha-2},\\
A_3 &= \sum_{n\in\mathbb{Z}} |d_{n+1}|^{\alpha}.
\end{aligned}$$

Therefore,

$$\frac{1}{A_2}\left[\left.\frac{\partial^2 \varphi}{\partial x \partial y}\right|_{\substack{x=0\\y=0}} - (\alpha\sigma^\alpha)^2 A_1 |z|^{2(\alpha-1)}\varphi(0,0,z)\right] = \frac{1}{A_3}\left[\frac{\partial^2 \varphi}{\partial z^2} - (\alpha\sigma^\alpha)^2 A_3^2 |z|^{2(\alpha-1)}\varphi(0,0,z)\right],$$

This yields for $\alpha = 1$,

$$\frac{1}{A_2}\left[\left.\frac{\partial^2 \varphi}{\partial x \partial y}\right|_{\substack{x=0\\y=0}} - \sigma^2 A_1 \varphi(0,0,z)\right] = \frac{1}{A_3}\left[\frac{\partial^2 \varphi}{\partial z^2} - \sigma^2 A_3^2 \varphi(0,0,z)\right].$$

Taking into account that $d_n = \psi^n \mathbb{1}_{\{n \geq 0\}}$ for the noncausal AR(1) and noticing that

$$\frac{\partial^2 \varphi}{\partial x \partial y} = -\mathbb{E}\left[X_{t+k}X_{t+\ell}e^{izX_{t-1}}\right],$$
$$\frac{\partial^2 \varphi}{\partial z^2} = -\mathbb{E}\left[X_{t-1}^2 e^{izX_{t-1}}\right],$$

we get for any $z \in \mathbb{R}^*$:

$$\mathbb{E}\left[\left\{X_{t+kX_{t+\ell}} - (\text{sign } \psi)^{k+\ell}\left(|\psi|^{-\ell-1}(X_{t-1}^2 + \tilde{\sigma}^2) - \tilde{\sigma}^2\right)\right\} e^{izX_{t-1}}\right] = 0,$$

with $\tilde{\sigma} = \dfrac{\sigma}{1-|\psi|}$. From Bierens (Theorem 1, 1982):

$$\mathbb{E}\left[X_{t+k}X_{t+\ell}\middle|X_{t-1}\right] = (\text{sign } \psi)^{k+\ell}\left(|\psi|^{-\ell-1}(X_{t-1}^2 + \tilde{\sigma}^2) - \tilde{\sigma}^2\right),$$

which concludes the proof.

L.3 Proof of Proposition L.1

Let k and ℓ be two positive integers such that $\ell \leq k$. From Lemma D.1, we know that for any $h \geq 0$, there exist two polynomials P_h and Q_h of respective degrees $q-1$ and h such that:

$$X_{t+h} = P_h(B)X_{t-1} + Q_h(F)u_t.$$

Thus, using the same device as in the Proof of Proposition 3.2,

$$\begin{aligned}
\mathbb{E}\Big[X_{t+k}X_{t+\ell}\Big|X_{t-1},\dots,X_{t-q-1}\Big] &= \mathbb{E}\left[\Big(P_k(B)X_{t-1} + Q_k(F)u_t\Big)\Big(P_\ell(B)X_{t-1} + Q_\ell(F)u_t\Big)\middle|X_{t-1},\dots,X_{t-q-1}\right], \\
&= \Big(P_k(B)X_{t-1}\Big)\Big(P_\ell(B)X_{t-1}\Big) \\
&\quad + \Big(P_k(B)X_{t-1}\Big)\mathbb{E}\Big[Q_\ell(F)u_t\Big|u_{t-1}\Big] + \Big(P_\ell(B)X_{t-1}\Big)\mathbb{E}\Big[Q_k(F)u_t\Big|u_{t-1}\Big] \\
&\quad + \sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j \mathbb{E}\Big[u_{t+i}u_{t+j}\Big|u_{t-1}\Big].
\end{aligned}$$

The second and third terms can be expressed as:

$$\begin{aligned}
&\Big(P_k(B)X_{t-1}\Big)\mathbb{E}\Big[Q_\ell(F)u_t\Big|u_{t-1}\Big] + \Big(P_\ell(B)X_{t-1}\Big)\mathbb{E}\Big[Q_k(F)u_t\Big|u_{t-1}\Big] = \\
&\qquad \text{sign}(\psi)\Big(\phi(B)X_{t-1}\Big)\Big[Q_\ell(\text{sign } \psi)\Big(P_k(B)X_{t-1}\Big) + Q_k(\text{sign } \psi)\Big(P_\ell(B)X_{t-1}\Big)\Big],
\end{aligned}$$

whereas the fourth term can be rewritten using Lemma L.1:

$$\begin{aligned}\sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j \mathbb{E}\Big[u_{t+i}u_{t+j}\Big|u_{t-1}\Big] &= \sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j (\text{sign } \psi)^{i+j}\left[|\psi|^{-\min(i,j)-1}\Big((\phi(B)X_{t-1})^2+\tilde{\sigma}^2\Big)-\tilde{\sigma}^2\right],\\ &= -\tilde{\sigma}^2 Q_k(\text{sign } \psi)Q_\ell(\text{sign } \psi)\\ &\quad + \Big((\phi(B)X_{t-1})^2+\tilde{\sigma}^2\Big)\sum_{i=0}^{k}\sum_{j=0}^{\ell} q_i q_j (\text{sign } \psi)^{i+j}|\psi|^{-\min(i,j)-1}.\end{aligned}$$

L.4 Proof of Proposition 3.4

The result of Proposition 3.4 is obtained by substituting $\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big]$ and $\mathbb{E}\Big[X_{t+h}^2\Big|\mathscr{F}_{t-1}\Big]$ in

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \mathbb{E}\Big[X_{t+h}^2\Big|\mathscr{F}_{t-1}\Big] - \Big(\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big]\Big)^2,$$

using the formulas of Propositions 3.2 and L.1.

L.5 Details on Example 3.2

By Lemma E.1, the polynomial Q_h intervening in Proposition 3.4 reads in the case of the MAR(1,1)

$$Q_h(z) = \sum_{i=0}^{h} \phi^{h-i} z^i.$$

Applying Proposition 3.4, we know that

$$\mathbb{V}\Big(X_{t+h}\Big|\mathscr{F}_{t-1}\Big) = \left((X_{t-1}-\phi X_{t-2})^2 + \frac{\sigma^2}{(1-|\psi|)^2}\right)\left(c_h - \Big(Q_h(\text{sign } \psi)\Big)^2\right),$$

with $c_h = \sum_{i=0}^{h}\sum_{j=0}^{h} q_{i,h}q_{j,h}(\text{sign } \psi)^{i+j}|\psi|^{-\min(i,j)-1}$. Using the explicit form of the $q_{i,k}$'s, the coefficients of polynomial Q_h, we can deduce that for $\psi > 0$

$$c_h - \Big(Q_h(\text{sign } \psi)\Big)^2 = \phi^{2h}\sum_{i=0}^{h}\sum_{j=0}^{h}\phi^{-i-j}(\psi^{-\min(i,j)-1}-1).$$

which can be simplified by elementary calculations after splitting the sums according to whether $i \geq j$ or $j > i$.

M Proof of Lemma E.1

For $h = 0$, Equation (D.1) holds with $P_0(B) = \phi_1 + \phi_2 B^2 \ldots + \phi_q B^{q-1}$ and $Q_0(B) = 1$. We have

$$\begin{aligned}
X_{t+h} &= a_{0,h} X_{t-1} + \sum_{i=1}^{q-1} a_{i,h} X_{t-i-1} + \sum_{i=0}^{h} b_{i,h} u_{t+i} \\
&= a_{0,h} \left(\sum_{i=0}^{q-1} \phi_{i+1} X_{t-i-2} + u_{t-1} \right) + \sum_{i=1}^{q-1} a_{i,h} X_{t-i-1} + \sum_{i=0}^{h} b_{i,h} u_{t+i} \\
&= \sum_{i=0}^{q-2} \Big(a_{i+1,h} + a_{0,h}\phi_{i+1} \Big) X_{t-i-2} + a_{0,h}\phi_q X_{t-q-1} + a_{0,h} u_{t-1} + \sum_{i=0}^{h} b_{i,h} u_{t+i}.
\end{aligned}$$

Since this last formula holds at any $t \in \mathbb{Z}$, this last equation yields

$$X_{t+h+1} = \sum_{i=0}^{q-2} \Big(a_{i+1,h} + a_{0,h}\phi_{i+1} \Big) X_{t-i-1} + a_{0,h}\phi_q X_{t-q} + a_{0,h} u_t + \sum_{i=1}^{h+1} b_{i-1,h} u_{t+i}.$$

However, we also have by definition

$$X_{t+h+1} = P_{h+1}(B) X_{t-1} + Q_{h+1}(F) u_t = \sum_{i=0}^{q-1} a_{i,h+1} X_{t-i-1} + \sum_{i=0}^{h+1} b_{i,h+1} u_{t+i}.$$

Thus, by identification,

$$\begin{aligned}
a_{q-1,h+1} &= a_{0,h}\phi_q, \\
a_{i,h+1} &= a_{i+1,h} + a_{0,h}\phi_{i+1}, \quad \text{for} \quad 0 \le i \le q-2, \\
a_{0,h} &= b_{0,h+1}, \\
b_{i,h+1} &= b_{i-1,h}, \quad \text{for} \quad 1 \le i \le h+1.
\end{aligned}$$

We deduce from these equations that for any $h \ge 0$,

$$\begin{aligned}
b_{i,h+1} &= a_{0,h-i}, \quad \text{for} \quad 0 \le i \le h+1, \\
a_{i,h+1} &= \sum_{j=0}^{\min(q-i-1,h)} a_{0,h-j}\phi_{i+1+j}, \quad \text{for} \quad 0 \le i \le q-1,
\end{aligned}$$

with the convention $a_{0,-1} = 1$. We obtain that $(a_{0,h})$ is the solution of the linear recurrent equation of order q

$$a_{0,h+q} = \phi_1 a_{0,h+q-1} + \ldots + \phi_q a_{0,h}, \quad \text{for} \quad h \ge 0, \tag{M.1}$$

with initial values $(a_{0,0}, \ldots, a_{0,q-1})$ that could be expressed as functions of $\phi_1, \ldots, \phi_q$. Denote $\lambda_1, \ldots, \lambda_s$ the distinct roots of the polynomial $F^q\phi(B)$ with respective multiplicities $m_1, \ldots, m_s$, with $s \leq q$, $m_1 + \ldots + m_s = q$. Since ϕ has all its roots outside the unit circle, we know that $|\lambda_i| < 1$ for all i. Therefore, there exist polynomials $C_1, \ldots, C_q$ of respective degrees $m_1, \ldots, m_s$ such that for any $h \geq q$,

$$a_{0,h} = C_1(h)\lambda_1^h + \ldots + C_s(h)\lambda_s^h.$$

N A recursive scheme for computing polynomials P_h and Q_h of Lemma D.1

Lemma N.1 *Polynomials P_h and Q_h of Lemma D.1 satisfy the following recursive equations:*

$$BP_{h+1}(B) = P_h(B) - P_h(0)\phi(B), \qquad Q_{h+1}(F) = FQ_h(F) + P_h(0), \tag{N.1}$$

with initial conditions $Q_0(B) = 1$, $P_0(B) = \phi_1 + \phi_2 B + \ldots + \phi_q B^{q-1}$.

Proof. By applying polynomial $\phi(B)$ to (D.1), we get by (B.1)

$$\phi(B)X_{t+h} = P_h(B)\phi(B)X_{t-1} + Q_h(F)\phi(B)u_t,$$
$$B^{-h}u_t = BP_h(B)u_t + Q_h(F)\phi(B)u_t,$$

which implies $B^{h+1}P_h(B) + B^h Q_h(F)\phi(B) = 1$. The same holds at rank $h+1$. Thus, denoting $Q_h(F) = \sum_{i=0}^h q_{i,h}F^i$ and $Q_h^*(B) := B^h Q_h(F) = \sum_{i=0}^h q_{h-i,h}B^i$, we also have: $B^{h+2}P_{h+1}(B) + Q_{h+1}(B)\psi^*(B)\phi(B) = 1$. Subtracting the expressions at ranks h and $h+1$ yields:

$$B^{h+1}\Big(BP_{h+1}(B) - P_h(B)\Big) + \phi(B)\Big(Q_{h+1}^*(B) - Q_h^*(B)\Big) = 0. \tag{N.2}$$

We can notice that the term of degree zero in this expression is: $\phi(0)\Big(Q_{h+1}^*(0) - Q_h^*(0)\Big) = 0$, hence $q_{h+1,h+1} = q_{h,h}$. Focusing on the next terms of degrees $i = 1, \ldots, h$, we can iteratively show that $q_{h+1-i,h+1} = q_{h-i,h}$. Finally, focusing on the term of degree $h+1$, we now deduce that $-P_h(0) + q_{1,h+1} - q_{0,h} = 0$. This leads us to the equality

$$Q_{h+1}^*(B) = Q_h^*(B) + B^{h+1}P_h(0), \tag{N.3}$$

or equivalently $Q_{h+1}(F) = FQ_h(F) + P_h(0)$, which establishes the right-hand side equation of (N.1). Finally, replacing (N.3) in (N.2) concludes the proof of Lemma N.1.

O Diagnostic test of extreme residuals clustering

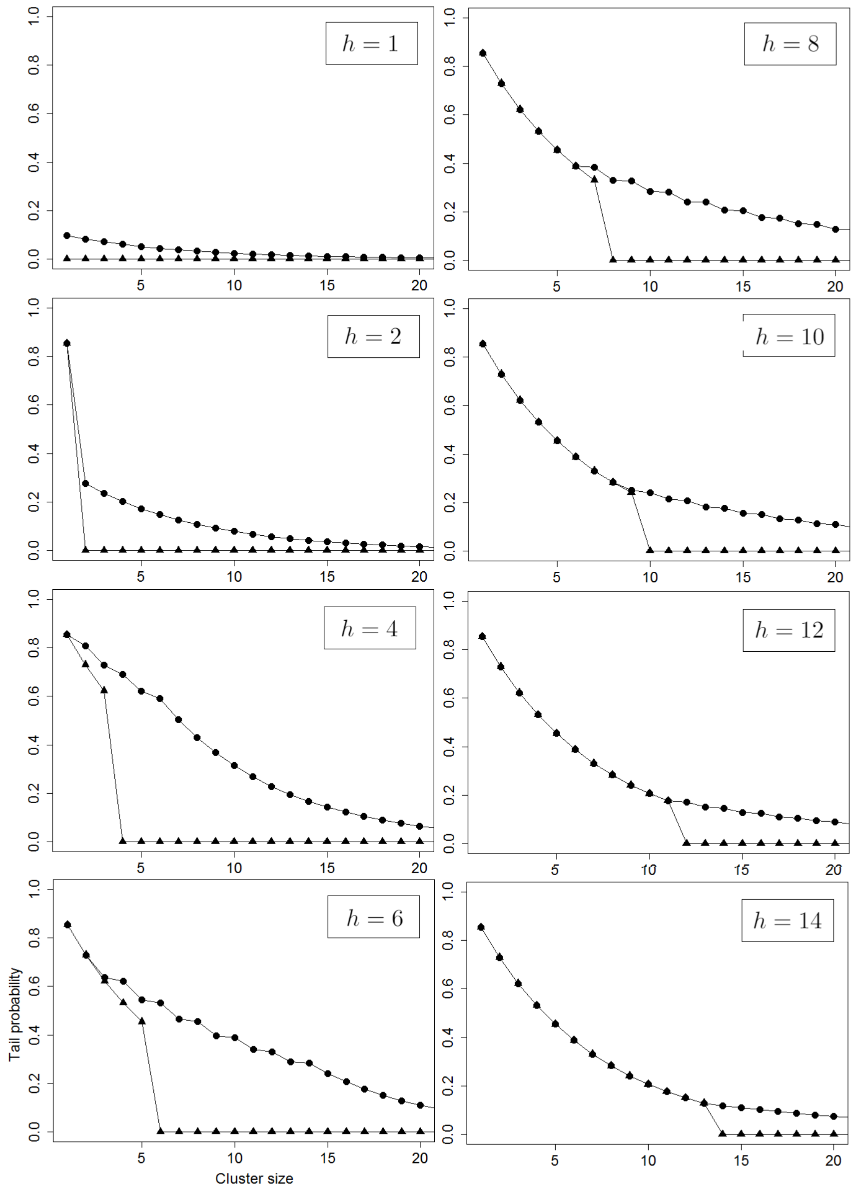


Figure 6: Theoretical tail probability (given by Equation (4.18)) of cluster sizes of extreme errors (4.20) (true model, triangles) and (4.21) (misspecified model, points) for $\alpha = 1.5$, $\psi_0 = 0.9$ at different horizons h.

We illustrate the extreme clustering behaviours of the two error sequences considered in Section 4.3.3 for horizons $h = 1, \dots, 20$ and parameter values $\alpha = 1.5$, $\psi_0 = 0.9$. From equations (4.20) and (4.21), we deduce the sequence $(c_{(k)})$ and compute the tail probability distributions of the cluster size using (4.18). As depicted on Figure 6, the errors of the all-pass alternative are rather unlikely to appear by large clusters for small horizons (e.g. $h = 1$), which makes them difficult to distinguish from the errors of the strong representation. However, for intermediate values of h, the error sequence of the all-pass displays a much higher probability of large clusters of extremes whereas the clusters sizes of the errors of the strong representation have probability zero of exceeding h.

P Monte Carlo study: complementary results and methodology

P.1 Asymptotic distribution of the LS estimator

			$\alpha = 1.5$	$\psi = 0.7$	$\phi = 0.9$			$\alpha = 1$	$\psi = 0.7$	$\phi = 0.9$	
n		$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$	$q_{0.1}$	$q_{0.25}$	Median	$q_{0.75}$	$q_{0.9}$
500	$\hat{\delta}_1$	-2.759	-1.338	-0.527	-0.061	0.231	-12.69	-3.569	-0.731	0.012	0.691
	$\hat{\delta}_2$	-0.265	0.038	0.495	1.284	2.653	-0.873	-0.049	0.694	3.430	12.13
2000	$\hat{\delta}_1$	-1.558	-0.746	-0.226	0.086	0.417	-6.321	-1.732	-0.221	0.247	1.382
	$\hat{\delta}_2$	-0.448	-0.105	0.214	0.730	1.521	-0.662	-0.320	0.001	0.322	0.655
5000	$\hat{\delta}_1$	-1.188	-0.565	-0.132	0.156	0.513	-4.564	-1.269	-0.097	0.387	1.824
	$\hat{\delta}_2$	-0.536	-0.172	0.125	0.561	1.177	-2.098	-0.469	0.096	1.357	4.749
∞	$\hat{\delta}_1$	-0.726	-0.252	0.000	0.246	0.719	-5.470	-0.856	0.000	0.954	5.686
	$\hat{\delta}_2$	-0.762	-0.264	0.000	0.268	0.768	-6.687	-1.110	0.000	1.006	6.503
			$\alpha = 0.5$	$\psi = 0.7$	$\phi = 0.9$			$\alpha = 1.7$	$\psi = 0.3$	$\phi = 0.4$	
500	$\hat{\delta}_1$	-1307	-114.6	-5.247	0.157	14.06	-1.003	-0.513	-0.042	0.408	0.870
	$\hat{\delta}_2$	-21.31	-0.412	5.176	114.8	1239	-0.958	-0.484	-0.008	0.466	0.956
2000	$\hat{\delta}_1$	-524.3	-40.97	-0.493	2.804	54.63	-0.662	-0.328	-0.016	0.290	0.618
	$\hat{\delta}_2$	-74.37	-4.171	0.506	46.28	563.9	-0.662	-0.320	0.001	0.322	0.655
5000	$\hat{\delta}_1$	-385.3	-28.11	-0.109	5.402	96.34	-0.641	-0.313	-0.008	0.292	0.608
	$\hat{\delta}_2$	-127.1	-7.493	0.111	33.07	445.0	-0.647	-0.318	-0.001	0.316	0.648
∞	$\hat{\delta}_1$	-1546	-31.43	0.000	32.34	1614	-0.555	-0.235	0.000	0.231	0.554
	$\hat{\delta}_2$	-2129	-42.88	0.000	41.63	2068	-0.614	-0.257	0.001	0.261	0.621

Table P.1: Characteristics of the empirical distribution of $\hat{\delta}_i = \left(\frac{n}{\ln n}\right)^{1/\alpha}(\hat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_{0i})$, for $i = 1, 2$ over 100,000 simulated paths of α-stable MAR(1,1) processes (X_t) solution of $(1-\psi F)(1-\phi B)X_t = \varepsilon_t$ with four different parametrisations $(\alpha, \psi_0, \phi_0) \in \{(1.5, 0.7, 0.9), (1, 0.7, 0.9), (0.5, 0.7, 0.9), (1.7, 0.3, 0.4)\}$. The empirical a-quantile is denoted q_a. The results for $n = \infty$ are obtained by simulations of the asymptotic distribution in (4.8). [See Example 4.1]

P.2 Direct implementation of the Portmanteau test

We conducted an experiment to assess the direct implementation of the Portmanteau test (without bootstrap) and focused on $\alpha = 1.5$, which seems more realistic for financial series (see Footnote

[4]). We computed the residuals of the 100,000 simulated paths based on the weak causal AR(2) fits, evaluate the statistic (4.10) for $h = 1, \ldots, 10$ and simulate its asymptotic distribution. For each path, we performed the test at three different different nominal sizes 1%, 5% and 10% by comparing the statistic to the appropriate quantile of the asymptotic distribution. The empirical sizes are reported in Table P.2. The test suffers heavy distortions, especially in finite sample, which was expected from the results by Lin and McLeod (2008) in the pure causal AR framework. It is generally oversized for small lags and progressively becomes undersized as more lags are included. The empirical sizes slowly approach the nominal sizes as the number of observations increases and the discrepancy between few and more lags also gets smaller.

	$n = 500$			$n = 2000$			$n = 5000$		
H	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	6.69	21.2	31.7	3.08	9.42	17.0	1.92	6.28	12.5
2	4.54	16.4	27.1	2.40	7.80	14.7	1.60	5.77	11.6
3	3.40	13.4	22.8	1.96	6.41	12.4	1.36	4.84	10.1
4	2.65	10.7	19.0	1.64	5.38	10.3	1.17	4.17	8.74
5	2.11	8.96	16.2	1.37	4.58	8.96	1.04	3.59	7.61
6	1.61	7.58	13.8	1.16	3.93	7.94	0.91	3.20	6.84
7	1.24	6.49	12.1	1.01	3.51	7.17	0.80	2.86	6.22
8	0.96	5.66	10.6	0.89	3.19	6.58	0.70	2.62	5.73
9	0.74	5.08	9.62	0.81	2.94	5.99	0.64	2.42	5.30
10	0.57	4.55	8.74	0.75	2.70	5.50	0.60	2.26	5.00

Table P.2: Empirical sizes of Portmanteau tests with nominal sizes 1%, 5% and 10% using the first H lags, $H = 1, \ldots, 10$ of the residuals' autocorrelations of 100,000 simulated paths of process (X_t) solution of $(1 - 0.7F)(1 - 0.9B)X_t = \varepsilon_t$, with 1.5-stable noise.

P.3 Extreme residuals clustering

P.3.1 Estimating the term structure of excess clustering

In practice, for one simulated path of the MAR(1,1) process (X_t) and one horizon h, we have six series of residuals $(\hat{\zeta}^i_{t+h|t})_t$, $i = 1, \ldots, 6$, one each for the pure causal and noncausal AR(2) competitors, and two each for the two MAR(1,1) competitors. To compute for each of them the empirical cluster sizes sequence $\left(\hat{\xi}^i_{k,h}(x)\right)_k$ as defined in Section 5.2, we need to choose a threshold $x > 0$ which will determine whether a residual is extreme or not. It is desirable to use thresholds such that we can harmoniously the clustering behaviours of the six series of residuals. For the experiment detailed below, instead of directly using the series of residuals, we worked with the autostandardised series $\hat{v}^i_{t+h|t} := \left(\frac{\hat{\zeta}^i_{t+h|t}}{\max_s |\hat{\zeta}^i_{s+h|s}|}\right)_t$ which all lie between 0 and 1, and for each horizon h, we used the threshold

$$x_h := \max_{i=1,\dots,6} q_a\Big(|\hat{v}^i_{t+h|t}|\Big), \tag{P.1}$$

where $q_a(\cdot)$ the a-percent quantile. In our experiments, $a = 0.9$ was used. We now detail the procedure of the experiment. For a given parameterisation (α, ψ_0, ϕ_0) and path length n, we simulate 10,000 paths of process (X_t) solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$ and conducted the experiment as follows. For each simulated path of (X_t) and a given horizon $h \geq 1$:

ι) Estimate the regression $X_t = \hat{\eta}_1 X_{t-1} + \hat{\eta}_2 X_{t-2} + \hat{\zeta}_t$ invoking Proposition 4.3.

$\iota\iota$) Obtain the (inverted) roots $(\hat{\psi}, \hat{\phi})$ from the estimated characteristic polynomial $1-\hat{\eta}_1 B-\hat{\eta}_2 B^2$ (we impose $|\hat{\psi}| \leq |\hat{\phi}|$ for the sake of identifiability).

$\iota\iota\iota$) For each of the four competing models (4.14)-(4.17), decompose the process into pure causal and noncausal components and compute $(\hat{v}^i_{t+h|t})$, the series of standardised errors at horizons h.

$\iota\nu$) Compute x_h, and for each series $(\hat{v}^i_{t+h|t})$, $i = 1, \dots, 6$, compute the cluster sizes sequence $\left(\hat{\xi}^i_{k,h}(x_h)\right)_k$ and the associated Excess Clustering indicator according to equation (P.1) and (5.1). For a given parameterisation, this procedure yields for each six series of residuals, 10,000 term structures of excess clustering. Averaging model-wise across these term structures -recall that the MAR(1,1) competitors have each two series of residuals whereas the pure causal and noncausal AR(2) have only one each- allows to gauge the excess clustering of each competing model.

P.3.2 Excess clustering for other parameterisations

If excess residuals clustering is clearly present for the all-pass representations already for small sample size, it can be seen that the contrast sharply increases as the sample length grows (see the two upper panels of Figure P.1). Also, even with a much smaller noncausal parameter $\psi = 0.2$ (lower right panel of Figure P.1), the strong representation still displays the least excess clustering compared to the three other competitors. We can notice in this case that the pure causal AR(2) alternative is not far from the strong representation (points). This is coherent with the fact that the noncausal parameter ψ is relatively small, especially compared to the causal parameter ϕ, yielding much weaker dependence across the residuals of the misspecified pure causal AR(2).

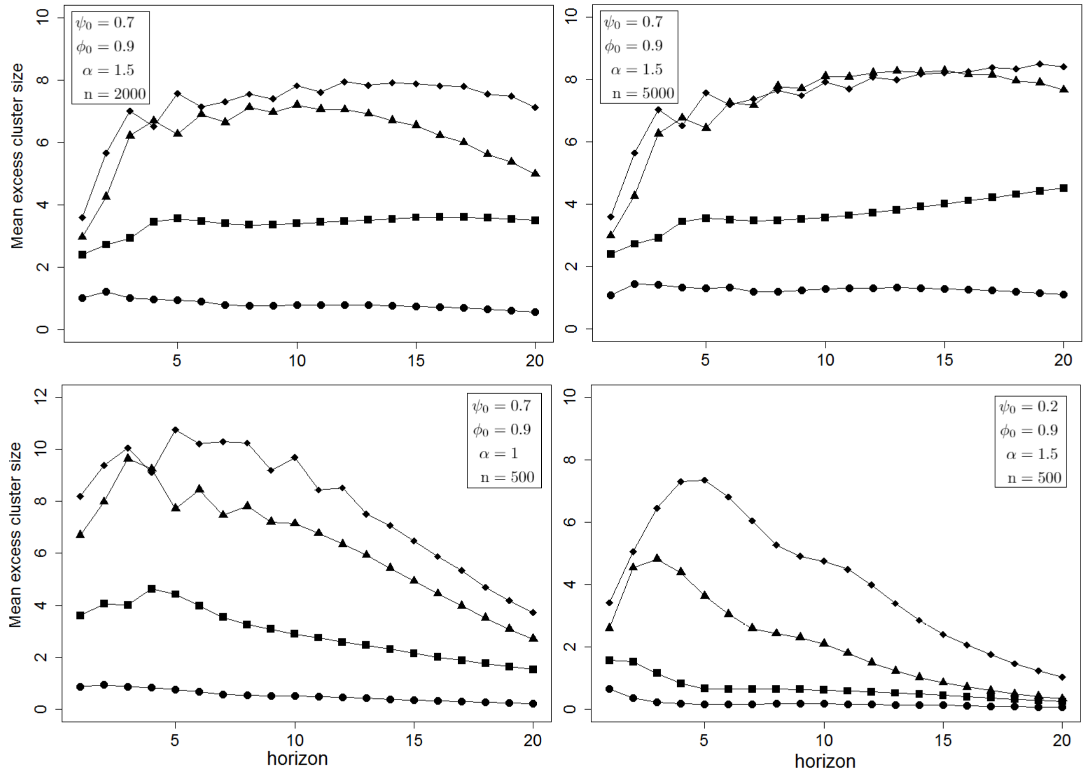


Figure P.1: Across 10,000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, average of the term structure of excess clustering of the linear prediction residuals of the four competing models (4.14) (squares), the strong representation (4.15) (points), (4.16) (triangles) and (4.17) (diamond). The parameterisations and path lengths are indicated on each panel.

P.3.3 Procedure of EC diagnosis for real data

ι) For increasing total AR order r, fit a backward AR(r) and compute the roots of the characteristic polynomial.

$\iota\iota$) For all the competing MAR(p, q) models possible, disentangle the causal and the noncausal components and compute the residuals at multiple horizons.

$\iota\iota\iota$) Using the procedure of section 5.2, for each competing model, compute the term structure of excess clustering.

$\iota\nu$) Select the specification according to some criterion, a natural criterion for instance: minimal area under the curve (AUC) of the term structure of excess clustering.

The last step of this procedure might be highly dependent on the chosen selection criterion. In particular, the natural criterion we mentioned might favour more complex models with higher total AR order, leading to overfitting. To avoid this, it could be augmented by a penalisation term.

Alternatively, for each competing model, we could derive the exact asymptotic distribution of the EC indicator of Equation (5.1) under the null hypothesis that the model is correctly specified to test whether there is statistically significant excess clustering. These questions are left for further research.

References

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

Andrews, B. and R. A., Davis (2013): Model Identification for Infinite Variance Autoregressive Processes. *Journal of Econometrics*, 172(2), 222-234.

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

Breidt, F.J., Davis, R.A. and A.A. Trindade (2001): Least Absolute Deviation Estimation for All-Pass Time Series Models. *Annals of Statistics*, 29, 919-946.

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

Cavaliere, G., Nielsen, H.B, and A. Rahbek (2017): Bootstrapping Non-Causal Autoregressions: with Applications to Explosive Bubble Modelling. Unpublished document, University of Copenhagen.

Chavez-Demoulin, V., and A. C., Davison (2012): Modelling Time Series Extremes. *REVSTAT-Statistical Journal*, 10, 109-133.

Chen, Y., Phillips, P.C.B. and J. Yu (2017): Inference in Continuous Systems with Mildly Explosive Regressors. Unpublished document, Singapore Management University.

Daley, D. J. and D., Vere-Jones (2007): *An introduction to the Theory of Point Processes*, Springer, New York, vol. 1.

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

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

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

Donoho, D. (1981): On Minimum Entropy Deconvolution. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 565-608. Academic Press, New York.

Embrechts, P., Klüppelberg, C. and T. Mikosch (1997): *Modelling Extremal Events. Applications of Mathematics*, Springer-Verlag, Berlin.

Ferro, C. A., and J., Segers (2003): Inference for Clusters of Extreme Values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2), 545-556.

Francq, C., and J.-M., Zakoian (2013): Estimating the Marginal Law of a Time Series with Applications to Heavy-Tailed Distributions. *Journal of Business & Economic Statistics*, 31(4), 412-425.

Gourieroux, C., and J., Jasiak (2016): Filtering, Prediction and Simulation Methods for Noncausal Processes. *Journal of Time Series Analysis*, 37(3), 405-430.

Gourieroux, C., Jasiak, J. and A., Monfort (2016): Stationary Bubble Equilibria in Rational Expectation Models. Working Papers from Centre de Recherche en Economie et Statistique (CREST), No 2016-31.

Gouriéroux, C., and J.-M., Zakoian (2015): On Uniqueness of Moving Average Representations of Heavy-tailed Stationary Processes. *Journal of Time Series Analysis*, 36(6), 876-887.

Gouriéroux, C. and J.-M., Zakoian (2017): Local Explosion Modelling by Non-Causal Process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 79, 737-756.

Hecq, A., Lieb, L., and S. M., Telg (2016a): Identification of Mixed Causal-Noncausal Models in Finite Samples. *Annals of Economics and Statistics*, (123/124), 307-331. doi:10.15609/annaeconstat2009.123-124.0307

Hecq, A., Telg, S., and L., Lieb (2016b): Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates? MPRA Paper 74922, University Library of Munich, Germany.

Hecq, A., Telg, S., and L, Lieb (2017): Simulation, Estimation and Selection of Mixed Causal-Noncausal Autoregressive Models: The MARX Package. Available at SSRN: https://ssrn.com/abstract=3015797.

Hencic, A., and C., Gouriéroux (2015): Noncausal autoregressive model in application to bitcoin/usd exchange rates. *Econometrics of Risk*, Springer International Publishing, p. 17-40.

Hsing, T., Hüsler, J., and M. R., Leadbetter (1988): On the Exceedance Point Process for a Stationary Sequence. *Probability Theory and Related Fields*, 78(1), 97-112.

Hsueh, A.C., and J. M. Mendel (1985): Minimum-Variance and Maximum-Likelihood Deconvolution For Noncausal Channel Models. IEEE Trans. *Geoscience and Remote Sensing*, 23, 797-808.

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. *Journal of Time Series Econometrics*, 3(3).

Leadbetter, M. R., and S., Nandagopalan (1989). On Exceedance Point Processes for Stationary Sequences Under Mild Oscillation Restrictions. *Lecture Notes in Statistics*, 51, 69-80.

Lin, J. W., and A. I., McLeod (2008): Portmanteau tests for ARMA models with infinite variance. *Journal of Time Series Analysis*, 29(3), 600-617.

Ling, S. (2007): A Double AR(p) Model: Structure and Estimation. *Statistica Sinica*, 17, 161.

Markovich, N. M. (2014). Modelling Clusters of Extreme Values. *Extremes*, 17(1), 97-125.

McCulloch, J. H. (1986): Simple consistent estimators of stable distribution parameters. *Communication in Statistics–Computation and Simulation* 15, 1109–36.

Nielsen, H. B., and A., Rahbek (2014): Unit Root Vector Autoregression with Volatility Induced Stationarity. *Journal of Empirical Finance*, 29, 144-167.

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

Phillips, P.C.B., Wu, Y. and J. Yu (2011): Explosive Behaviour in the 1990's Nasdaq: When Did Exuberance Escalate Asset Values? *International Economic Review*, 52, 201-226.

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.

Scargle, J. D. (1981): Phase-Sensitive Deconvolution to Model Random Processes, with Special Reference to Astronomical Data. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 549-564. Academic Press, New York.

Wiggins, R. A. (1978): Minimum Entropy Deconvolution. *Geoexploration*, 16, 21-35.