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

Fries, Sébastien and Zakoian, Jean-Michel

ENSAE, CREST

September 2017

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

Sébastien Fries*and Jean-Michel Zakoian†

**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. Extending the study of the noncausal AR(1) model by Gouriéroux and Zakoian (2017), we show that the conditional distribution in direct time is lighter-tailed than the errors distribution, and we emphasize the presence of ARCH effects 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 causal AR representation with non-i.i.d. errors can be consistently estimated by classical least-squares. We derive a portmanteau test to check the validity of the estimated 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 Zakoian, 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 (2016, 2017a, 2017b), 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, rational expectations models in economics have been shown to admit solutions with noncausal components when departing from the finite variance assumption (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. Sections 5 and 6 propose numerical illustrations based on simulated and real data, respectively. Section 7 concludes. Proofs are collected in the Appendix. Complementary results are provided in a Supplementary file.

## 2 Stable MAR$(p,q)$ processes

MAR processes have been considered, among others, by Lanne and Saikkonen (2011), Gouriéroux and Jasiak (2016), Hecq, Issler, and Telg (2017).[2] A MAR$(p,q)$ process $(X_t)$ is the strictly stationary solution of the difference equation

$$\psi(F)\phi(B)X_t = \varepsilon_t, \qquad \text{where} \quad \psi(F) = 1 - \sum_{i=1}^{p} \psi_i F^i, \quad \phi(B) = 1 - \sum_{i=1}^{q} \phi_i B^i, \tag{2.1}$$

[2] See the latter reference for additional motivations on the use of MAR processes in time series econometrics. The first two references develop forecasting procedures for noncausal MAR processes.

$B$ and $F$ are the usual lag and forward operators ($B^k X_t = X_{t-k}, F^k X_t = X_{t+k}, k \in \mathbb{Z}$), $(\varepsilon_t)$ is an independent and identically distributed (i.i.d.) sequence, the polynomials $\psi$ and $\phi$ have all their roots outside the unit circle and are such that $\psi_p \neq 0$ and $\phi_q \neq 0$. When $q = 0$ (resp. $p = 0$), the model is called purely noncausal (resp. causal).

We assume that the errors $\varepsilon_t$ 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.[3] Stable laws are easily characterised through their characteristic function: $\varepsilon_t$ 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

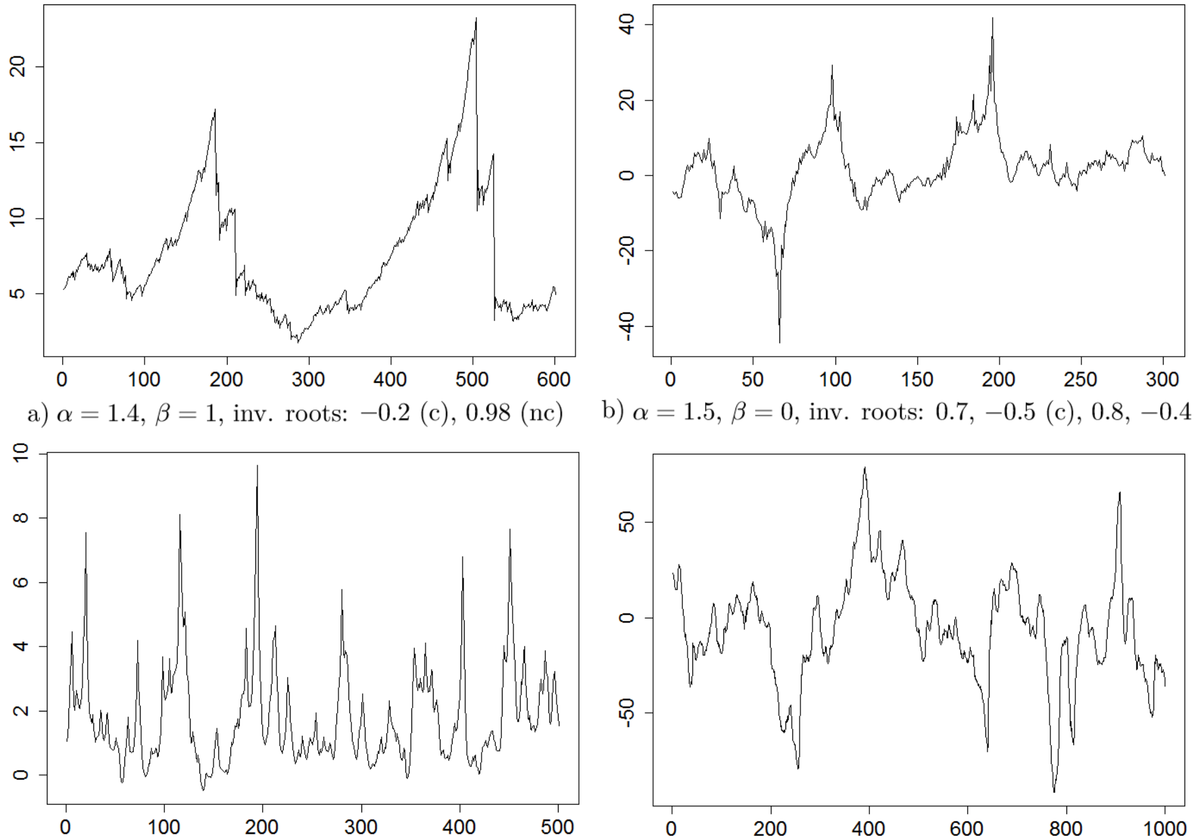
$$\forall s \in \mathbb{R}, \quad \mathbb{E}(e^{is\varepsilon_t}) = \exp\Big\{ -\sigma^\alpha |s|^\alpha \left(1 - i\,\beta \,\mathrm{sign}(s) w(\alpha, s)\right) + is\mu \Big\}, \tag{2.2}$$

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. 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 $\beta \in (-1, 1)$.

## 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 $\beta$. 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 4 below). The influence of a smaller tail parameter $\alpha$ 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 ($\alpha = 1.6$).

[3] See for instance Embrechts, Klüppelberg, and Mikosch (1997), Samorodnitsky and Taqqu (1994) for the main properties of stable distributions. A major justification for using stable distributions rather than other classes of heavy-tailed distributions (such as the Student's $t$, the hyperbolic distributions) is that they are the only possible limit distributions for properly normalized and centered sums of i.i.d. random variables (giving rise to generalized Central Limit Theorems). Moreover, they are sufficiently flexible to accommodate asymmetry as well as fat tails. Finally, moving averages processes based on stable variables also follow stable distributions, as will be detailed below.



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).

Under the assumptions made on the AR polynomial, $(X_t)$ admits an MA($\infty$) representation[4]

$$X_t = \sum_{k=-\infty}^{+\infty} d_k \varepsilon_{t+k}. \tag{2.3}$$

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 $\varepsilon_\tau$. 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

[4] It follows from Proposition 13.3.1 in Brockwell and Davis (1991) that the infinite sum in (2.3) is well defined under the stable law assumption, which ensures the existence of $\mathbb{E}|\varepsilon_t|^s$ for $s < \alpha$.

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.

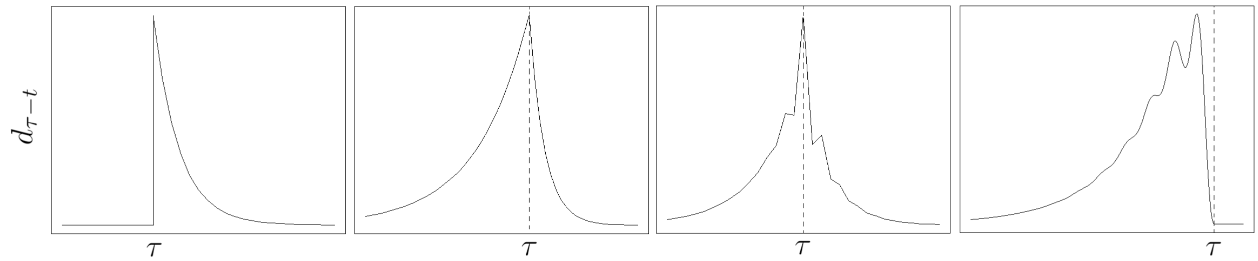



Figure 2: Examples of baseline paths $t \longmapsto d_{\tau-t}$ of MAR processes with characteristic polynomials, from left to right: $1-0.7B$ ; $(1-0.9F)(1-0.7B)$ ; $(1-0.8F)(1+0.4F)(1-0.7B)(1+0.5B)$ ; $(1-0.99F)(1-965F)(1-0.98e^{i0.045\pi}F)(1-0.98e^{-i0.045\pi}F)$.

### 2.2 Marginal distribution

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

**Proposition 2.1** *Let* $(X_t)$ *the strictly stationary solution of the MAR(p, q) Model* (2.1) *where the roots of the polynomials* $\psi$ *and* $\phi$ *are outside the unit disk and* $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$. *Then* $X_t$ *has a stable 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 $\alpha$ of $X_t$ is that of the error term. In particular, $\mathbb{E}|X_t|^s < +\infty$ for $s < \alpha$ and $\mathbb{E}|X_t|^\alpha = +\infty$.

# 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 start by showing that the Markov property holds whatever the errors distribution.

**Proposition 3.1** *Let* $(X_t)$ *the strictly stationary solution of the MAR(p, q) Model* (2.1) *where the roots of the polynomials* $\psi$ *and* $\phi$ *are outside the unit disk and* $(\varepsilon_t)$ *is an i.i.d. sequence (not necessarily stable). Then* $(X_t)$ *is an homogeneous Markov chain of order* $p + q$.

In the rest of the section, we will derive properties of the conditional distribution of $X_t$ in direct time when the errors are stable-distributed. We will focus on (i) the existence of conditional moments; (ii) explicit derivation of predictive formulas for $X_t$; and (iii) the presence of ARCH effects in the case of the MAR(1,$q$) process. More specific results will be detailed for the MAR(1,1) process.

### 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, generalising the result obtained for the AR(1) by GZ.

**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*

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

The conditional distribution in direct time, that is with respect to the past observations, thus has lighter tails than both the marginal distribution and the distribution conditional on the future. In particular, whatever the heaviness of the tails of $\varepsilon_t$, the conditional expectation of $X_t$ always exists. The conditional variance in direct time also exists provided that the tails of the errors distribution are not too fat ($\alpha > 1/2$).[5]

### 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. The next proposition extends in a non trivial way the prediction formula obtained by GZ for the noncausal AR(1), i.e. for the MAR(1,0). Let $\mathscr{F}_t = \sigma(X_t, X_{t-1}, \ldots)$ the canonical filtration of process $(X_t)$. For $x \neq 0$ and $r \in \mathbb{R}$, let $x^{<r>} = \text{sign}(x)|x|^r$.

**Proposition 3.2** *Let the MAR$(1, q)$ process $(1-\psi F)\phi(B)X_t = \varepsilon_t$, under the assumptions of Model (2.1), with $\varepsilon_t \sim \mathcal{S}(\alpha, 0, \sigma, 0)$. Then 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*

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

[5] A discrepancy between conditions of existence for marginal and conditional moments also holds for many nonlinear causal models: for instance GARCH (see e.g. Francq and Zakoian (2011), Chapter 2), or models for time series of counts (Davis and Liu, 2012).

*and we have the semi-strong causal representation*

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

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

The proof is based on: i) disentangling pure causal and noncausal components of the MAR process (in the spirit of Lanne and Saikkonen (2011), Gouriéroux and Jasiak (2016)); ii) using the closed-form expression of the conditional expectation of the pure noncausal component, and (iii) invoking the Markov property.[6]

It is worth noting that the conditional expectation is linear in the past and can be explicitly computed. By comparison with finite variance AR processes, the semi-strong representation (3.1) 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 (see for instance Brockwell and Davis (1991), Section 4.4).[7] In our framework, the noncausal component $(1-\psi F)$, with $|\psi| < 1$, is transformed into the causal component $(1-\psi^{<\alpha-1>}B)$.

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 $\alpha$. Proposition 3.2 allows us to distinguish different behaviours summarised in the following Corollary.

**Corollary 3.1** *Under the assumptions of Proposition 3.2, we have almost surely*

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

[6] The inherent complexity of the pure noncausal component when $p > 1$, for which no such closed-form expression exists, does not allow us to go beyond $p = 1$ for the results of this section.

[7] The equality $\psi(F)Z_t = \psi(B)\varepsilon_t$ indeed implies that $(Z_t)$ has a spectral density given by $f_Z(\lambda) = \frac{\sigma^2}{2\pi}\frac{|\psi(e^{-i\lambda})|^2}{|\psi(e^{i\lambda})|^2} = \frac{\sigma^2}{2\pi}$ where $\sigma^2 = \mathrm{Var}(\varepsilon_t)$. Therefore, $(Z_t)$ is also a white noise with the same variance $\sigma^2$ as $\varepsilon_t$.

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

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

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 and 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 the Supplementary file.

### 3.3 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, extending again the results obtained by GZ for the noncausal AR(1).

**Proposition 3.3** *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*

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

*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 $Q_h(z)$, for $h \geq 0$, are defined in the Appendix. The causal representation (3.1) can then be completed and reveals quadratic ARCH effects in the Cauchy MAR$(1,q)$ process.

**Corollary 3.2** *Under the assumptions of Proposition 3.3, there exists a sequence $(\eta_t)$ 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 $\eta_t$ 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).

### 3.4 The MAR$(1,1)$ process.

The results of this section can be made completely explicit for the MAR(1,1) model defined by

$$(1-\psi F)(1-\phi B)X_t = \varepsilon_t, \quad \text{with} \quad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha,\beta,\sigma,\mu), \tag{3.3}$$

with $|\phi| < 1$ and $0 < |\psi| < 1$. The coefficients of the MA($\infty$) representation (2.3) are given by: $d_k = \dfrac{\psi^k}{1-\phi\psi}$, 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

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

In particular, when $\psi, \phi > 0$ 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 $X_t \sim \mathcal{S}\left(1, 0, \dfrac{\sigma}{(1-\psi)(1-\phi)}, 0\right)$.

We now derive an explicit prediction formula for the MAR(1,1) process when $\beta = \mu = 0$. Proposition 3.2 yields for any $h \geq 0$,

$$\begin{aligned}
\mathbb{E}\Big[X_{t+h}\Big|\mathscr{F}_{t-1}\Big] &= \phi^{h+1}X_{t-1} + (X_{t-1}-\phi X_{t-2})(\psi^{<\alpha-1>})^{h+1}\sum_{i=0}^{h}\left(\phi\psi^{<1-\alpha>}\right)^i,\\
&= \begin{cases} \phi^{h+1}X_{t-1} + \dfrac{(\psi^{<\alpha-1>})^{h+1}-\phi^{h+1}}{1-\phi\psi^{<1-\alpha>}}(X_{t-1}-\phi X_{t-2}), & \text{if} \quad \phi\psi^{<1-\alpha>} \neq 1,\\ \phi^{h+1}\left[X_{t-1} + (h+1)(X_{t-1}-\phi X_{t-2})\right], & \text{if} \quad \phi\psi^{<1-\alpha>} = 1. \end{cases}
\end{aligned}$$

When $\psi > 0$ and $\alpha = 1$, Corollary 3.2 yields

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

where $\mathbb{E}\left[\eta_t|\,\mathscr{F}_{t-1}\right] = 0$ and $\mathbb{E}\left[\eta_t^2|\,\mathscr{F}_{t-1}\right] = 1$. The conditional variance at horizon $h$ in Proposition 3.3 takes the more explicit form, for any $h \geq 0$,

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

where

$$c_h = \frac{(1+\phi\psi)\psi^{-h-1}}{(1-\phi\psi)(1-\phi^2\psi)} - \frac{2\phi^{h+1}}{(1-\phi)(1-\phi\psi)} + \frac{(1+\phi)\phi^{2(h+1)}}{(1-\phi)(1-\phi^2\psi)}.$$

# 4 Statistical Inference

This section is devoted to the LS estimation of the MAR$(p,q)$ model

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

where $\psi_0(z) = 1 - \sum_{i=1}^{p} \psi_{0i} z^i$, $\phi_0(z) = 1 - \sum_{i=1}^{q} \phi_{0i} z^i$, with $\psi_0(z) \neq 0$ and $\phi_0(z) \neq 0$ for $|z| \leq 1$.

Contrary to other estimation methods such as Maximum Likelihood (ML)[8], LS do not require full specification of the errors distribution. We relax the assumption that $(\varepsilon_t)$ is an $\alpha$-stable sequence and rather assume that the law of $\varepsilon_t$ belongs to the domain of attraction of a stable distribution. Specifically, we assume that there exists a function $L$ which is slowly varying at infinity[9] such that

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha} L(x), \qquad \text{and} \qquad \lim_{x\to\infty} \frac{\mathbb{P}(\varepsilon_0 > x)}{\mathbb{P}(|\varepsilon_0| > x)} \to c \in [0,1]. \tag{4.2}$$

This more general assumption on the errors distribution encompasses in particular the fully parametric $\alpha$-stable framework under which the properties of the previous section were derived. Replacing $\alpha$-stable laws by their domain of attraction alleviates the risk of misspecification.[10]

We will first derive the asymptotic properties of estimators of a "all-pass causal representation" of the MAR$(p,q)$ process. Then, we will develop a portmanteau test for checking the validity of the estimated representation. Finally, we will consider selecting the true model, among the different specifications admitting the same all-pass representation, based on properties of extreme clustering.

## 4.1 All-pass causal representation

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 $\psi_0(z)\phi_0(z)$ by replacing one or several roots by their inverses, we have

$$\eta_0^*(B)X_t = \zeta_t^*, \tag{4.3}$$

[8] See Andrews, Calder and Davis (2009) for asymptotic properties of the ML estimator of both causal and noncausal AR processes with non-Gaussian $\alpha$-stable distribution. In the finite variance setting, ML estimation of MAR models based on Student's $t$ distribution was studied by Hecq et al. (2016).

[9] i.e. $\lim_{x\to\infty} L(tx)/L(x) = 1, \forall t > 0$.

[10] The same assumption was considered in the context of causal AR processes for the study least-absolute deviation (LAD) estimators by An and Chen (1982), and for M-estimators by Davis, Knight and Liu (1992).

where $(\zeta_t^*)$ is an all-pass process.[11] Such representations (4.3) will be called all-pass in the following. In the set of all-pass representations, one is characterized by a polynomial $\eta_0$ having all its roots outside the unit disk

$$\eta_0(B)X_t = \zeta_t, \qquad \text{where} \quad \eta_0(B) = \psi_0(B)\phi_0(B) = 1 - \sum_{i=1}^{p+q} \eta_{0i} B^i, \tag{4.4}$$

and $(\zeta_t)$ is an all-pass process. In the sequel, we call (4.4) the *all-pass causal* representation of $(X_t)$.

Now, 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($\infty$) coefficients in (2.3).

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

$$\rho(h) = \sum_{i=1}^{p+q} \eta_{0i}\rho(h-i), \quad \forall h > 0, \tag{4.5}$$

*where the coefficients $\eta_{0i}$ are obtained from* (4.4)*.*

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. Such equations explain why the coefficients of the all-pass causal representation of $(X_t)$ can be consistently estimated by LS.

### 4.2 Least-squares estimation

We consider LS parameter estimation of the all-pass causal representation (4.4), based on observations $X_1, \dots, X_n$ of the MAR$(p,q)$ model (4.1). A LS estimator of $\boldsymbol{\eta}_0 = (\eta_{01}, \dots, \eta_{0,p+q})'$ is

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta} \in \mathbb{R}^{p+q}} \mathcal{L}_n^*(\boldsymbol{\eta}), \tag{4.6}$$

where

$$\mathcal{L}_n^*(\boldsymbol{\eta}) = \sum_{t=p+q+1}^{n} \left( X_t - \sum_{i=1}^{p+q} \eta_i X_{t-i} \right)^2. \tag{4.7}$$

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

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Gamma}}_n^{-1} \hat{\boldsymbol{\gamma}}_n, \quad \hat{\boldsymbol{\Gamma}}_n = [\hat{\gamma}(i-j)]_{i,j=1,\dots,p+q}, \quad \hat{\boldsymbol{\gamma}}_n = [\hat{\gamma}(i)]_{i=1,\dots,p+q}. \tag{4.8}$$

[11] 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.

**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

$$a_n = \inf\{x : \mathbb{P}(|\varepsilon_0| > x) \leq n^{-1}\}, \quad \text{and} \quad \tilde{a}_n = \inf\{x : \mathbb{P}(|\varepsilon_0 \varepsilon_1| > x) \leq n^{-1}\}, \tag{4.9}$$

defined by Davis and Resnik (1986). 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 $\boldsymbol{K}^{(p+q)} = \mathbf{0}$). Let $\boldsymbol{L} = [\boldsymbol{K} \quad \boldsymbol{K}^{(2)} \quad \dots \quad \boldsymbol{K}^{(p+q)}]$. We start by providing the asymptotic behaviour of the LS estimator under the simplifying assumption that the distribution of $\varepsilon_t$ is symmetric. This assumption will be relaxed in the next section. The following result is a consequence of Davis and Resnik (1986).

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

$$\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{d} \boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\mathbf{Z}, \quad \textit{where} \quad \boldsymbol{Z} = (Z_1, \dots, Z_{p+q})', \tag{4.10}$$

*$Z_k = \sum_{l=1}^{+\infty} \{\rho(k+l) + \rho(k-l) - 2\rho(l)\rho(k)\} S_l/S_0$, 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 $j \geq 1$, has index $\alpha$. If the law of $|\varepsilon_t|$ is asymptotically equivalent to a Pareto, (4.10) holds with $a_n^2/\tilde{a}_n = (n/\ln n)^{1/\alpha}$.*

The $\alpha$-stable domain of attraction assumption on $(\varepsilon_t)$ impacts the asymptotic behaviour of the LS estimator in two important aspects: $\iota$) the limiting distribution depends on $\alpha$ and $\iota\iota$) the convergence rate is $\frac{a_n^2}{\tilde{a}_n} \sim n^{1/\alpha}\tilde{L}(n)$, for some slowly varying function $\tilde{L}$ (see p.551, Davis and Resnick, 1986). Requiring $\mathbb{E}|\varepsilon_t|^{\alpha} = \infty$ ensures that the law of $\varepsilon_0\varepsilon_1$ belongs to the $\alpha$-stable domain of attraction (see Cline (1983), Theorem 3.3 iv) p. 80).

**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

$$\boldsymbol{R}^{-1}\{\boldsymbol{I}_2 - \boldsymbol{L}(\boldsymbol{I}_2 \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\} = \frac{1 + \phi_0\psi_0}{(1-\psi_0^2)(1-\phi_0^2)} \begin{pmatrix} (1+\psi_0\phi_0)^2 + (\psi_0+\phi_0)^2 & -(\psi_0+\phi_0) \\ -2(\psi_0+\phi_0)(1+\psi_0\phi_0) & 1+\psi_0\phi_0 \end{pmatrix}.$$

This matrix can be straightforwardly estimated by plugging LS estimators of $\phi_0 + \psi_0$ and $\phi_0\psi_0$. From the estimated asymptotic distribution, at a given confidence level $\underline{\alpha} \in (0,1)$, we can deduce an asymptotic confidence region, $\mathcal{A}_{n,\underline{\alpha}}$ say, such that $P[(\phi_0 + \psi_0, \phi_0\psi_0) \in \mathcal{A}_{n,\underline{\alpha}}] = 1 - \underline{\alpha}$. Denote by $r_1 < r_2$ the inverses of the roots of $\eta_0(z)$ (thus $r_i \in \{\phi_0, \psi_0\}$ for $i = 1, 2$), that is $(r_1, r_2) = \left(\eta_{01} - \sqrt{\eta_{01}^2 + 4\eta_{02}}, \eta_{01} + \sqrt{\eta_{01}^2 + 4\eta_{02}}\right)/2 := f(\eta_{01}, \eta_{02})$. By the delta method, an asymptotic confidence region can be deduced for $(r_1, r_2)$: let $\mathcal{R}_{n,\underline{\alpha}}$ such that $P[(r_1, r_2) \in \mathcal{R}_{n,\underline{\alpha}}] = 1 - \underline{\alpha}$. Thus, $P[\{(\phi_0, \psi_0), (\psi_0, \phi_0)\} \subset \mathcal{R}_{n,\alpha}] = 1 - \underline{\alpha}$. Finally, letting $\mathcal{R}^*_{n,\underline{\alpha}}$ the symmetric of $\mathcal{R}_{n,\underline{\alpha}}$ around the line $r_1 = r_2$, we get an asymptotic confidence region for $(\phi_0, \psi_0)$: $P[(\phi_0, \psi_0) \in \mathcal{R}_{n,\underline{\alpha}} \cup \mathcal{R}^*_{n,\underline{\alpha}}] \geq 1 - \underline{\alpha}$.

The knowledge of index $\alpha$ is not required for the computation of the LS estimator, but the asymptotic distribution, as well as the normalizing constants $a_n$ and $\tilde{a}_n$, depend on $\alpha$. The presence of this nuisance parameter renders inference difficult for this class of model. Having estimated the AR coefficients, one could overcome this hurdle by using a standard estimator for the tail index $\alpha$. For instance, for a random sample $(X_1, \ldots, X_n)$, the so-called Hill (1975) estimator of $1/\alpha$ based on $m + 1$ upper order statistics is defined as:

$$\hat{\alpha}_m^{-1} = \frac{1}{m} \sum_{i=1}^{m} \log\left(\frac{X_{(i)}}{X_{(m+1)}}\right),$$

where $X_{(i)} > 0$ is the $i$th order statistic in decreasing order $(X_{(1)} \geq X_{(2)} \geq \ldots X_{(n)})$. Mason (1982) proved that the Hill estimator is a consistent estimator of $1/\alpha$, provided $n \to \infty$, $m \to \infty$ and $m/n \to 0$, in the case of i.i.d. variables. Consistency and asymptotic normality under serial dependence conditions - including $\ell$-dependence, $\beta$-mixing, ARCH - were established by various authors (see e.g. Hill (2010), De Hann et al. (2016) and the references therein). An alternative to the estimation of the asymptotic distribution is to base inference on bootstrap. Recently Cavaliere, Nielsen and Rahbek (2018) proposed bootstrap schemes for noncausal AR models with infinite variance, and showed their usefulness for hypothesis testing. Extension of this approach to mixed AR models remains an open issue.

### 4.3 Relaxing the symmetry assumption

In the previous section, we derived the asymptotic behaviour of the LS estimator of $\eta_0$ assuming the errors $(\varepsilon_t)$ were symmetrically distributed. We here relax the symmetry assumption and only require $(\varepsilon_t)$ to satisfy (4.2). The asymptotic behaviour of the LS estimator remains unchanged in the case $0 < \alpha < 1$, and holds for $1 < \alpha < 2$ after a mean-adjustment.[12]

Let $\tilde{\gamma}(h) = \sum_{t=0}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})$ where $\bar{X} = 1/n \sum_{t=0}^{n} X_t$, and denote $\tilde{\rho}(h) = \tilde{\gamma}(h)/\tilde{\gamma}(0)$

[12] A bias term appears in the case $\alpha = 1$ when departing from the symmetry assumption (See Davis and Resnik (1986), Theorem 4.4).

the mean-adjusted sample autocorrelation of order $h$. Similarly to (4.8), define the mean-adjusted Yule-Walker estimator $\tilde{\boldsymbol{\eta}}$ by

$$\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\Gamma}}_n^{-1} \tilde{\boldsymbol{\gamma}}_n, \quad \tilde{\boldsymbol{\Gamma}}_n = [\tilde{\gamma}(i-j)]_{i,j=1,\dots,p+q}, \quad \tilde{\boldsymbol{\gamma}}_n = [\tilde{\gamma}(i)]_{i=1,\dots,p+q}. \tag{4.11}$$

**Proposition 4.4** *Let $(X_t)$ be the strictly stationary solution of Model (4.1), where $(\varepsilon_t)$ is an i.i.d. sequence satisfying (4.2) and $\mathbb{E}|\varepsilon_t|^\alpha = \infty$.*

- *If $0 < \alpha < 1$, then (4.10) holds.*
- *If $1 < \alpha < 2$, then (4.10) holds with $\hat{\boldsymbol{\eta}}$ replaced by the mean-adjusted estimator $\tilde{\boldsymbol{\eta}}$.*

## 4.4 Diagnostic checking

Validity of the estimated model can be assessed by studying the sample autocorrelations of the residuals. Once the parameters of the all-pass representation (4.4) have been estimated by LS, with $\hat{\boldsymbol{\eta}} = (\hat{\eta}_i)_{i=1,\dots,p+q}$, the corresponding residuals are defined by

$$\hat{\zeta}_t = X_t - \sum_{i=1}^{p+q} \hat{\eta}_i X_{t-i}, \qquad t = p+q+1, \dots, n. \tag{4.12}$$

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 4.5** *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 $\gamma(0) = \sum_{k=-\infty}^{\infty} d_k^2$, 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 of the errors distribution).*

Details regarding matrix $A_H$ are available in the proof. Note that the symmetry assumption can be relaxed as in Section 4.3. It is now possible to propose a portmanteau test to check for residuals autocorrelations based for instance on the statistic

$$T_H = a_n^2 \tilde{a}_n^{-1} \sum_{i=1}^{H} |\hat{\rho}_{\hat{\zeta}}(i)| \underset{n \to +\infty}{\xrightarrow{d}} \|\gamma(0) \boldsymbol{A_H} \boldsymbol{Z}\|_1, \tag{4.13}$$

with $\|\boldsymbol{x}\|_1 = \sum |x_i|$ for any vector $\boldsymbol{x} = (x_i)$. Practical implementation of the test finally requires simulating/bootstraping the estimated asymptotic distribution in (4.13) (see Section 4.2).

### 4.5 Model selection based on extremes clustering

The all-pass causal representation (4.4) is compatible with all MAR($p', q'$) models of the form (4.3) (with $p' + q' = p + q$). Such models could have generated the observations, and it is important to detect which one is the true model. A distinctive feature of the latter is that the errors are i.i.d., not only "asymptotically empirically non-autocorrelated" as in (4.3) (see footnote [11]). Having estimated the coefficients of the polynomial $\eta_0$, a natural strategy for assessing the validity of a candidate model, with polynomial $\eta_0^*$, is to test the independence of the $\zeta_t^*$ in (4.3). We propose an approach based on extreme clustering of the residuals.[13]

#### 4.5.1 Point process of exceedances

This dependence materialises here 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($\infty$) representation $Y_t = \sum_{k\in\mathbb{Z}} c_k \varepsilon_{t+k}$, where $(\varepsilon_t)$ 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. In our context, $Y_t$ will typically be substituted for the errors $\varepsilon_t$ of the strong representation and the errors $\zeta_t^*$ of competing all-pass representations.

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 (see Davis and Resnick (Section 3.D, 1985)):

$$\sum_{k=1}^{n} \delta_{\left(k/n, a_n^{-1} Y_k\right)}\Big( \ \cdot \ \cap B_x\Big) \xrightarrow{d} \sum_{k=1}^{+\infty} \xi_k \delta_{\Gamma_k}, \tag{4.14}$$

where $\delta$ 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 $x^{-\alpha}$,[14] and $\xi_k = \mathrm{Card}\{i \in$

[13] Alternative approaches based on non-parametric and rank-based tests have been proposed for testing iid-ness of innovations (see for instance Brock, Dechert, LeBaron and Scheinkman (1996), and Duchesne, Ghoudi and Rémillard (2012)). However, the validity of these tests requires to have $\sqrt{n}$-consistent estimators of the model parameters, which is not the case in the $\alpha$-domain of attraction framework as shown in Proposition 4.3.

[14] 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 $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_\ell$ such that $a_i < b_i \leq a_{i+1}$, $i = 1, \ldots, \ell$, and

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

$$f(z) = \alpha z^{-\alpha-1}\mathbb{1}_{(1,+\infty)}(z). \tag{4.15}$$

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). We will now outline the interest of analysing extreme events for model selection.

### 4.5.2 Analysing the error processes of competing models

Let $(X_t)$ be the MAR$(p, q)$ strictly stationary solution of Model (4.2)-(4.1) and assume the order $p+q$ and the roots of $\eta_0$ are known. There is a finite number of competing representations among which the strong one lies. Denoting $(\varepsilon_t)$ of the errors of the strong representation and generically $(\zeta_t^*)$ the errors of any specific all-pass representation, we can analyse their extreme clustering behaviour using (4.14).

**Error process of the strong representation** The i.i.d. errors $(\varepsilon_t)$ of the strong representation admit the trivial MA form $\varepsilon_t = \sum_{k\in\mathbb{Z}} c_k \varepsilon_{t+k}$ with $c_0 = 1$ and $c_k = 0$ for $k \neq 0$. Thus, substituting $Y_t$ for $\varepsilon_t$, (4.14) holds with $\xi_k = \text{Card}\{i \in \mathbb{Z} : J_k|c_i| > 1\} = \mathbb{1}_{\{J_k>1\}} = 1$. The random variables $(\xi_k)$ describing the size of the cluster of extremes are degenerate in this case: the extreme errors $(\varepsilon_t)$ tend to appear isolated from each other.

**Error process of an all-pass representation** The (rescaled) errors $\zeta_t^*$ of an all-pass representation always admit an infinite one-sided or two-sided MA form, say $\zeta_t^* = \sum\limits_{k\in\mathbb{Z}} \frac{c_k}{\max_j |c_j|}\varepsilon_{t+k}$. Denote $(c_{(k)})_{k\geq 1}$ the sequence obtained by sorting $(|c_k|)_{k\in\mathbb{Z}}$ in descending order. Substituting $Y_t$ for $\zeta_t^*$, (4.14) holds with $\xi_k = \text{Card}\left\{i \geq 1 : J_k \frac{c_{(i)}}{c_{(1)}} > 1\right\} = \arg\max_{i\geq 1}\{J_k > c_{(i)}^{-1} c_{(1)}\}$ and from (4.15), we deduce that for any $\ell \geq 1$:

$$\mathbb{P}\Big(\xi_k \geq \ell\Big) = \mathbb{P}\Big(J_k > c_{(\ell)}^{-1} c_{(1)}\Big) = c_{(\ell)}^{\alpha} c_{(1)}^{-\alpha}. \tag{4.16}$$

In this case, the $\xi_k$'s can take arbitrarily high values with non-zero probability, indicating as expected that the extremes of $(\zeta_t^*)$ tend to cluster.

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 $(a_i, b_i]$, $i = 1, \dots, \ell$.

**Errors at higher horizons** Considering the extreme clustering of errors at further horizons can provide additional discriminating information. For simplicity, consider the noncausal AR(1) model. There are two competing models, yielding the same all-pass causal representation (4.4):

$$X_t = \psi_0 X_{t+1} + \varepsilon_t, \quad \text{and} \quad X_t = \psi_0 X_{t-1} + \zeta_t. \tag{4.17}$$

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

$$\varepsilon_{t+h|t} := X_t - \psi_0^h X_{t+h} = \varepsilon_t + \psi_0 \varepsilon_{t+1} + \ldots + \psi_0^{h-1} \varepsilon_{t+h-1}, \tag{4.18}$$

$$\zeta_{t+h|t} := X_{t+h} - \psi_0^h X_t = (\psi_0^{-h} - \psi_0^h) \sum_{k \geq h} \psi_0^k \varepsilon_{t+k} - \psi_0^h \sum_{k=0}^{h-1} \psi_0^k \varepsilon_{t+k}. \tag{4.19}$$

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 (4.18) will appear by clusters of at most $h$ consecutive observations, whereas the errors (4.19) will likely appear by larger clusters (see the Supplementary file for illustration). This analysis can be extended to general MAR processes by disentangling the pure causal and noncausal components of each competiting model (as in the proof of Proposition 3.1).

### 4.5.3 Application to model selection

The previous section highlights that the extreme errors of all-pass representations are likely to appear in large clusters, contrarily to the extreme errors of the strong representation that tend to appear isolated. Selecting the strong MAR$(p, q)$ representation, assuming only $p + q$ known, can thus be achieved by looking for evidence of extreme clustering in the errors of all competing representations. In principle, such evidence shall be found in the errors of all representations but the strong one.

## 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 all-pass causal representation. On synthetic data generated from a MAR(1,1) process, we assessed $\iota$) 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, and $\iota\iota\iota$) 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 $\alpha$-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 imposed $|\hat{\psi}| \leq |\hat{\phi}|$ for the sake of identifiability when the roots are real). 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 $\alpha$ 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 $\left\{\left|\hat{\theta} - \theta_0\right| < a\right\} \cap \left\{\hat{\theta} \in \mathbb{R}\right\}$, for $\theta = \phi$ or $\psi$, for $a = 0.01$, 0.05, 0.1 and over 100,000 simulated paths of the $\alpha$-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 the Supplementary file show that the finite sample distribution approaches its asymptotic behaviour much slower for lower values of $\alpha$. In the same line, a direct implementation of the portmanteau test using the statistics (4.13) 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. To alleviate the problem, they suggested a Monte Carlo test which relies on simulations of the estimated causal AR model to approximate the distribution of the portmanteau statistics under the null hypothesis of correct specification. An important difference in our framework is that, under the null, the estimated causal AR is only an all-pass representation of the process and we use its estimated coefficients to simulate paths of the corresponding pure causal AR as if it were the strong representation. Given that the residuals autocorrelations of this pure causal AR have the same asymptotic distribution as those of the all-pass causal representation, the procedure remains valid.[15] We therefore proceeded with this methodology (see the Supplementary file). The empirical sizes of the 1, 5 and 10% nominal tests for lags $H = 1, \dots, 10$ are reported in Table 2. It can be seen that using the Monte Carlo procedure, the portmanteau test is much better behaved in finite sample, especially for $\alpha = 1.5$, which is a realistic value for financial series.

| | $\alpha = 1.5$ | | | $\alpha = 1$ | | | $\alpha = 0.5$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $H$ | 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 (4.13) implemented by the Monte Carlo test procedure. The empirical size was calculated based on 2000 simulations of the $\alpha$-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.

### 5.2 Selection based on extreme residuals clustering

We now gauge the usefulness of the results of Section 4.5 by simulating paths of the $\alpha$-stable MAR(1,1) process $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$ with different parameterisations and analysing the residuals of the competing representations. There are four competing models yielding the same

[15] It can indeed be noticed that the asymptotic distributions of the LS estimator and of the residuals autocorrelations remain unchanged whether $X_t$ is defined as the solution of $\psi_0(F)\phi_0(B)X_t = \varepsilon_t$ or $\psi_0(B)\phi_0(B)X_t = \varepsilon_t$ in (4.1).

all-pass causal AR(2) representation:

$$\text{Pure causal AR(2):} \qquad (1-\psi_0 B)(1-\phi_0 B)X_t = \zeta_t, \tag{5.1}$$

$$\text{MAR(1,1):} \qquad (1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t, \tag{5.2}$$

$$\text{MAR(1,1):} \qquad (1-\psi_0 B)(1-\phi_0 F)X_t = \nu_t, \tag{5.3}$$

$$\text{Pure noncausal AR(2):} \qquad (1-\psi_0 F)(1-\phi_0 F)X_t = \omega_t, \tag{5.4}$$

where $(\zeta_t)$, $(\nu_t)$ and $(\omega_t)$ denote the sequences of errors of each all-pass representations. More specifically for each estimated model, we compute the errors at several horizons $h$ (as in (4.18)-(4.19) for the AR(1). For each MAR(1,1) alternatives, we disentangle the causal and noncausal components and compute their respective error series). For each errors series and a given threshold $x > 0$, we identify the clusters of consecutive extreme values, i.e., errors larger than $x$ in modulus. As explained in Section 4.5, for any horizon $h$, we expect all-pass representations to display larger clusters of extreme errors than the strong model, for which clusters larger than $h$ have zero probability. Letting $\hat{\xi}_{k,h}(x)$ denote the number of consecutive exceedances for the $k$-th cluster, we therefore propose an Excess Clustering (EC)[16] indicator defined as:

$$\mathrm{EC}_h = \frac{\sum\limits_{k/\hat{\xi}_{k,h}(x)>h} \left(\hat{\xi}_{k,h}(x) - h\right)}{\mathrm{Card}\{k : \hat{\xi}_{k,h}(x) > h\}}, \quad \text{if} \quad \mathrm{Card}\{k : \hat{\xi}_{k,h}(x) > h\} > 0, \quad \text{else} \quad \mathrm{EC}_h = 0. \tag{5.5}$$

We start by generating 10000 sample paths of MAR(1,1) processes. For each path, 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 residuals excess clustering with respect to the horizon, using the indicator (5.5). Averaging model-wise across the 10000 simulations yields the typical excess clustering behaviours of the residuals of each competing models. We perform this experiment for several MAR(1,1) processes and display the results of two parameterisations in Figure 3 (see the Supplementary file for additional results and details regarding the methodology).

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 we could expect from (4.16), the heavier the tails the easier it is to identify dependent residuals. This

[16] For a given $h$, $\mathrm{EC}_h$ defined at (5.5) 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 reciprocal of the average size of clusters. Also, the choice of clustering scheme, i.e. how the sequence $(\hat{\xi}_{k,h}(x))_k$ is constructed, can have an impact on the estimated excess clustering : more elaborate clustering schemes could be considered (see for instance Ferro and Segers (2003) and Robert et al. (2009)).

is in line with the findings of Hecq, Lieb and Telg (2016) 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 in addition 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.

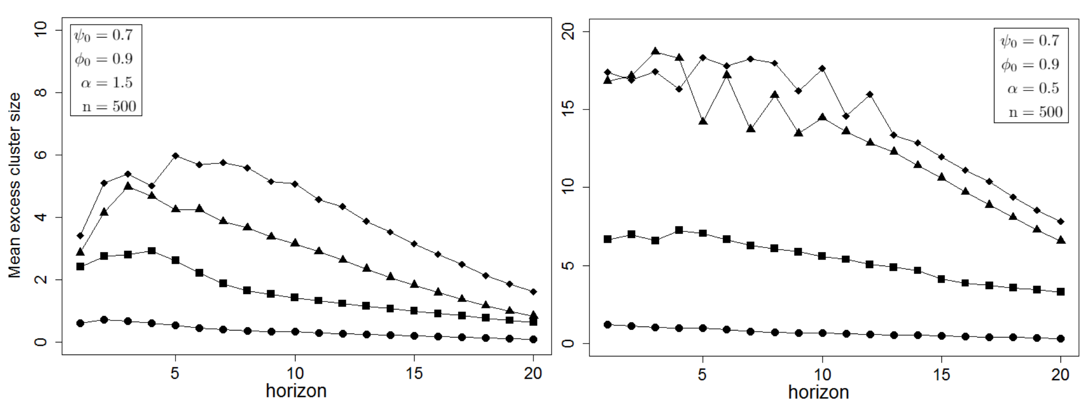



Figure 3: Across 10,000 simulations of the $\alpha$-stable MAR(1,1) process $(X_t)$ solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$, average of the term structure of excess clustering of the linear residuals of the four competing models (5.1) (squares), the strong representation (5.2) (points), (5.3) (triangles) and (5.4) (diamonds). 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 the quarterly Shiller Price/Earning ratio (1881 to 2017). All the series, pictured on Figure 4, have been centered and a linear deterministic trend has been fitted and subtracted.

## 6.1 AR estimation and validation using the Monte Carlo Portmanteau test

We start by investigating the appropriate total AR order ($r = p+q$) for each series using the Monte Carlo portmanteau test of Section 5.1. For each series, starting from total AR order 1, we estimate all-pass causal representations of increasing order by LS and perform the portmanteau test with $H = 50$ lags using the Monte Carlo portmanteau procedure of Lin and McLeod (1000 paths were simulated for each test). The results of the portmanteau test, reported in Table 4, allow to discard non-admissible low order models at the level 5%. We retain for the following the lowest orders (indicated in bold) which pass the portmanteau procedure: Boeing: 2; Exxon: 1; Coca-Cola: 1; Walmart: 1; HSI: 3; Shiller P/E: 6.[17]

| Total AR order $r$ | Boeing | Exxon | Coca-Cola | Walmart | HSI | Shiller P/E |
|---|---|---|---|---|---|---|
| 1 | 0.20 | **88.5** | **64.2** | **6.39** | 3.50 | 0.02 |
| 2 | **13.8** | | | | 3.40 | 0.01 |
| 3 | | | | | **8.19** | 0.90 |
| 4 | | | | | | 0.80 |
| 5 | | | | | | 1.30 |
| 6 | | | | | | **6.69** |

Table 4: P-values (%) of the Monte Carlo portmanteau tests with $H = 50$ lags for increasing AR order $r$. Rejection if P-value $< 5\%$.

[17] This procedure yields as a by-product the McCulloch quantile estimates of the tail index $\alpha$ (see McCulloch (1986)) for the six financial series. Values of $\hat{\alpha}$: Boeing: 1.79; Exxon: 1.69; Coca: 1.64; Walmart: 1.67; HSI: 1.38; Shiller P/E: 1.54. In all the cases, the infinite variance hypothesis is plausible.

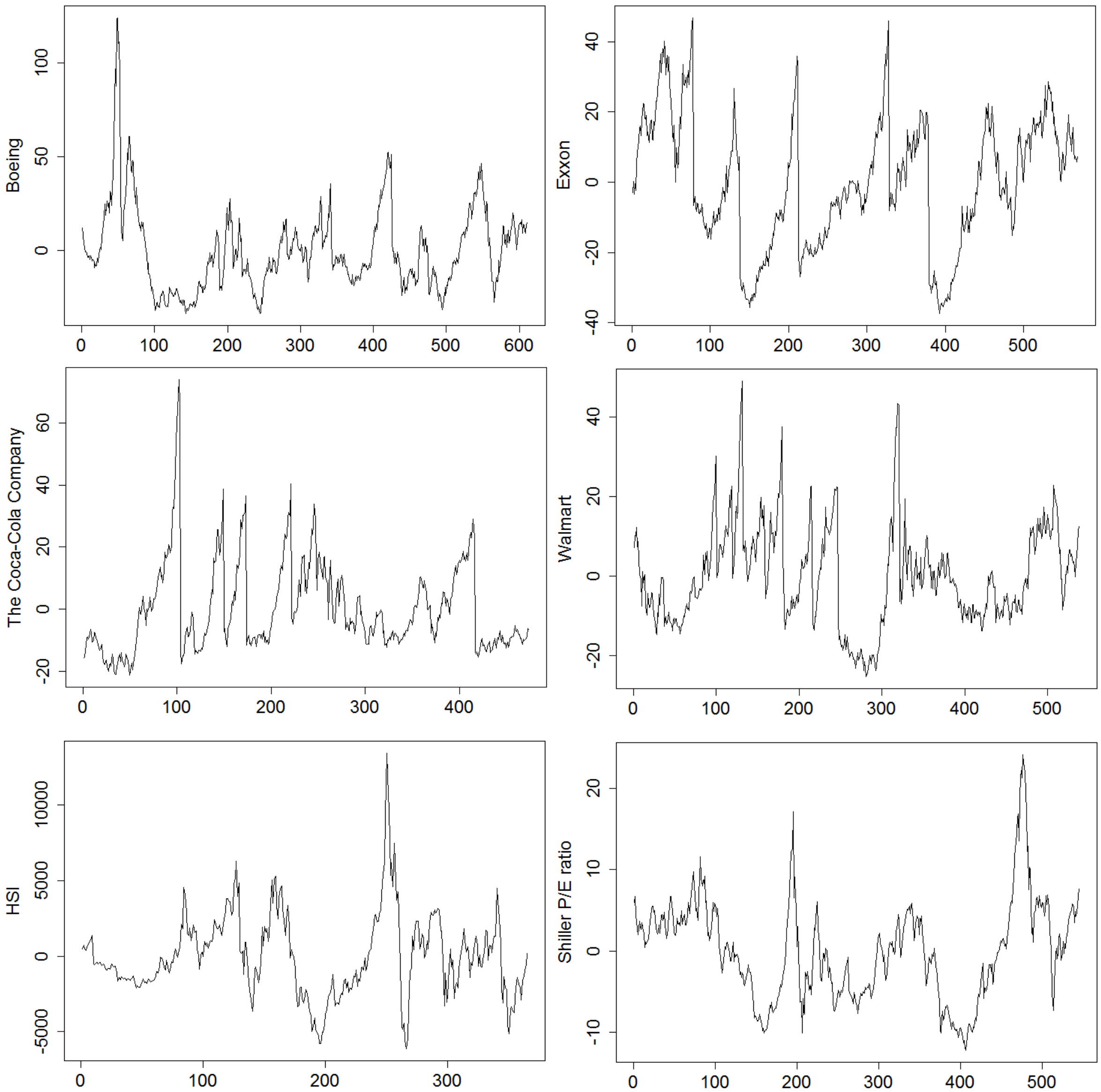



Figure 4: Financial series paths: Boeing (2/1962 to 12/2012), Exxon (2/1970 to 06/2017), Coca-Cola (1/1978 to 06/2017), Walmart (9/1979 to 6/2016), Hang Seng Index (HSI, 12/1986 to 05/2017) and the Shiller P/E ratio (Q1/1881 to Q2/2017). All series are monthly except the latter which is quarterly. Each are centered and a linear trend has been fitted and subtracted.

## 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 $r = 1, \ldots, 6$, 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 $r$ the MAR$(p,q)$ specification which displays the lowest AUC of excess clustering 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.

Combining the results from the portmanteau tests of the previous Section and of the extreme clustering analysis, we select a final specification for each series as follows. For a given series,

- Assign the total AR order $r$ validated by the portmanteau test (see Table 4).
- For the assigned order $r$, select then the least excess clustering competing representation (see the favoured specifications in Table 5).

The selection is reported in Table 6 which shows the causal and noncausal orders as well as the (inverted) roots of the corresponding polynomials. On the one hand, three series have been identified as pure noncausal AR(1): Exxon, Coca-Cola, Walmart, which is compatible we the fact that they all display multiple bubble patterns followed by sharp drops. On the other hand, the remaining series display more complex dynamics (Boeing, HSI, Shiller P/E ratio). Noticeably, the presence of bubbles in the HSI is unclear. Clear-cut drops are not visible which is compatible with the high causal root identified.

## 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. Building on Gouriéroux and Zakoïan (2017), we showed in this paper that several important properties of the pure noncausal AR(1) with stable errors extend to mixed AR models: the Markov property, the existence of a conditional mean whatever the size of the tails of the errors distribution, and the presence of ARCH effects in the Cauchy case. On the other hand, if the unit root property continues to hold when $\alpha = 1$, the martingale property is lost when a causal part is present in the model. A more complete description of the conditional distribution would require deriving higher-order moments, in particular to study the conditional skewness and kurtosis. We leave this issue for further research.

| 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 |

Table 5: Selection based on extreme clustering.

| Series | Final specification | Noncausal (inverted) roots | Causal (inverted) roots |
|---|---|---|---|
| Boeing | MAR(1,1) | 0.95 | 0.18 |
| 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(2,4) | $0.58 \pm 0.29i$ | $-0.21 \pm 0.6i$, 0.96, $-0.70$ |

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.

In the statistical part of the paper, we showed that LS estimation of a causal representation of the process allows to consistently identify the roots of the MAR polynomials, though not to distinguish causal and noncausal roots. Such identification issues were addressed by Hecq at al. (2016) for MAR processes, and by Cavaliere et al. (2018) using bootstrap inference for pure noncausal processes. We proposed an alternative strategy based on extreme clustering and leave its asymptotic properties for further investigations.

# Appendix: Proofs

## A Proof of Proposition 2.1

Using the MA($\infty$) representation (2.3) of $X_t$ and the assumption that $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu)$, it follows that

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

If $\alpha \neq 1$, then,

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

Whereas if $\alpha = 1$, then, for any $s \in \mathbb{R}$

$$\ln \psi_{X_t}(s) = \sum_{k=-\infty}^{+\infty} -\sigma |d_k s| \left(1 + i\frac{2}{\pi}\beta \text{sign}(d_k s) \ln |d_k s|\right) + i d_k s \mu$$
$$= -|s|\sigma \left(\sum_{k=-\infty}^{+\infty} |d_k|\right) \left(1 + i\frac{2}{\pi}\beta \frac{\sum_{k=-\infty}^{+\infty} d_k}{\sum_{k=-\infty}^{+\infty} |d_k|} \text{sign}(s) \ln |s|\right) + is \left(\sum_{k=-\infty}^{+\infty} d_k \mu - \frac{2}{\pi}\sigma\beta \sum_{k=-\infty}^{+\infty} d_k \ln |d_k|\right).$$

The conclusion follows from the characterization of stable laws in (2.2).

## 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 $\alpha$-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 $\varepsilon_t$ is independent from $X_{t+1}, \dots, X_{t+p}$. By the Bayes formula,

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

Thus,

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

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

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

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 $\varepsilon_t$ 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 $|x_p| \to +\infty$,

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

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,

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

Thus Lemma C.1 is established.

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

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

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

$$
\begin{aligned}
&\Big(\mathbb{E}\Big[|X_t|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\Big)^{1/\gamma} && \text{(C.5)}\\
&\qquad= \Big(\mathbb{E}\Big[|u_t-\phi_1X_{t-1}-\dots-\phi_qX_{t-q}|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\Big)^{1/\gamma}\\
&\qquad\le |\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|+\Big(\mathbb{E}\Big[|u_t|^\gamma\Big|X_{t-1},\dots,X_{t-p-q}\Big]\Big)^{1/\gamma}\\
&\qquad= |\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|+\Big(\mathbb{E}\Big[|u_t|^\gamma\Big|u_{t-1},\dots,u_{t-p},v_{t-p-1},\dots,v_{t-p-q}\Big]\Big)^{1/\gamma}\\
&\qquad= |\phi_1X_{t-1}+\dots+\phi_qX_{t-q}|+\Big(\mathbb{E}\Big[|u_t|^\gamma\Big|u_{t-1},\dots,u_{t-p}\Big]\Big)^{1/\gamma}, && \text{(C.6)}
\end{aligned}
$$

which is finite almost surely by Lemma C.1 since $(u_t)$ is an $\alpha$-stable pure noncausal AR($p$) process. If $\gamma \in (0,1)$, by the inequality $(a+b)^\gamma \le a^\gamma + b^\gamma$ for any $a, b \ge 0$, 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 3.1.

**Lemma D.1** *Let $(X_t)$ be a MAR(p, q) process with $q > 0$. For any $h \ge 0$, 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 $h$. In view of (B.1) we have $X_t = \sum_{i=1}^q \phi_i X_{t-i} + u_t$ from. 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 $h \ge 1$. 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

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

which is of the form (D.1) with

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

Therefore, (D.1) is established.

We now extend Theorem 3.1 in the case $p = 1$, by showing that for any $h \geq 0$, $\mathbb{E}\Big[|X_{t+h}|^\gamma \Big| X_{t-1}, \dots, X_{t-q-1}\Big] < +\infty$ whenever $0 < \gamma < 2\alpha + 1$. By Lemma D.1 we have, proceeding as for Equation (C.6) and letting $Q_h(z) = \sum_{i=0}^h q_{i,h} z^i$,

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

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 $(v_{t-2}, \dots, v_{t-q-1})$

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

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

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

and therefore,

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

To conclude, we invoke the fact that $(X_t)$ is a Markov chain of order $q+1$, which gives the equality $\mathbb{E}\Big[X_{t+h} \Big| X_{t-1}, \dots, X_{t-q-1}\Big] = \mathbb{E}\Big[X_{t+h} \Big| \mathscr{F}_{t-1}\Big]$. 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 3.1

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 $\phi$ and $C_1, \ldots, C_s$ are polynomials with degrees $m_1 - 1, \ldots, m_s - 1$.*

The proof is relegated to the Supplementary file.

The proof of Corollary 3.1 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) $\lambda_1$ 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 $Q_h(\psi^{<\alpha-1>})$

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

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.

$\iota$) 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

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

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

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

$\nu$) Assume $\alpha > 1 + \dfrac{\ln|\lambda|}{\ln|\psi|}$. From Equation (E.3),

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

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

# 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

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

## H Proof of Proposition 4.3

Let $\hat{\boldsymbol{\rho}} = [\hat{\rho}(i)]_{i=1,\dots,p+q}$, $\hat{\boldsymbol{R}} = [\hat{\rho}(i-j)]_{i,j=1,\dots,p+q}$. In view of (4.5) and (4.8), we have $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\rho}}$ and $\boldsymbol{\eta}_0 = \boldsymbol{R}^{-1}\boldsymbol{\rho}$. We have

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

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 (4.10) 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

The case $0 < \alpha < 1$ is a direct consequence of Theorem 4.4 (i) by Davis and Resnick (1986). Consider the case $1 < \alpha < 2$. From the proof of Corollary 1 p. 553 by Davis and Resnick (1986), we know that $\tilde{\rho}(h) - \hat{\rho}(h) = o_p(\tilde{a}_n a_n^{-2})$ for $h \geq 1$. Given $\hat{\rho}(h) \xrightarrow{p} \rho(h)$, it holds that $\tilde{\rho}(h) \xrightarrow{p} \rho(h)$ for $h \geq 1$. Following the proof of Proposition 4.3 with obvious notations, it can then be shown that $\tilde{a}_n^{-1}a_n^2(\tilde{\eta} - \eta_0)$ has the same asymptotic distribution as $\boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\frac{a_n^2}{\tilde{a}_n}(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho})$. The conclusion follows from Corollary 1 by Davis and Resnick (1986).

## J Proof of Proposition 4.5

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

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

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

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

with by convention $X_s = 0$ for $s \le 0$. 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 $m \le m'$, 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,

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

where the second equality follows from (4.5). Moreover,

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

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 $i \ge 0$. We have:

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

Thus, with

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

we can write

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

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

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

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 $\Big(\hat{\boldsymbol{\rho}}_{1:H+p+q} - \boldsymbol{\rho}_{1:H+p+q}\Big)$. More explicitly,

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

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

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

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

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

By Theorem 4.2 by Davis and Resnick (1985), Theorem 4.4 by Davis and Resnick (1986) and Lemma J.2 below, $\hat{\boldsymbol{P}} \xrightarrow{p} \boldsymbol{P} := \begin{pmatrix} \boldsymbol{0}_{1\times p+q} \\ \boldsymbol{I}_{p+q} \end{pmatrix} \boldsymbol{R}^{-1}\{\boldsymbol{I}_{p+q} - \boldsymbol{L}(\boldsymbol{I}_{p+q} \otimes \boldsymbol{R}^{-1}\boldsymbol{\rho})\}\boldsymbol{M}_{\boldsymbol{0}}$,
$\hat{\boldsymbol{A}}_{\boldsymbol{H}} \xrightarrow{p} \left[\Big(\boldsymbol{\eta}_0' \otimes \boldsymbol{\eta}_0'\Big)\boldsymbol{L}_{\boldsymbol{h}}\boldsymbol{M}_{\boldsymbol{h}} + \boldsymbol{\eta}_0'\boldsymbol{R}_{\boldsymbol{h}}'\boldsymbol{P}\right]_{h=1,\dots,H} := \boldsymbol{A}_{\boldsymbol{H}}$ and $\hat{\boldsymbol{\rho}}_{\hat{\zeta}} \xrightarrow{d} \gamma(0)\boldsymbol{A}_{\boldsymbol{H}}\boldsymbol{Z}$ where $\boldsymbol{Z} = (Z_1, \dots, Z_{H+p+q})$, and where the $(Z_i)$ are defined at Proposition 4.3.

**Lemma J.1** *Under the assumptions of Proposition 4.5,* $a_n^{-2}\Big(\hat{\gamma}(h) - \gamma(h)\hat{\gamma}_{\zeta}(0)\Big) \xrightarrow{p} 0.$

**Lemma J.2** *Under the assumptions of Proposition 4.5,* $a_n^{-2}\hat{\gamma}_{\hat{\zeta}}(0) = a_n^{-2}\dfrac{\hat{\gamma}(0)}{\gamma(0)} + o_P(1).$

### J.1 Proof of Lemma J.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

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

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

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

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

### J.2 Proof of Lemma J.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 J.1, the fact that $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \longrightarrow 0$ in probability and the convergence in distribution of the vector $a_n^{-2}\Big(\hat{\gamma}(i), \quad 0 \le i \le L\Big)$ 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 $i \ge 1$ and $\boldsymbol{\eta}_0(B)\gamma(0) = 1$. Thus

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

## References

**An, H.Z., and Z.G. Chen** (1982): On Convergence of LAD Estimates in Autoregression with Infinite Variance. *Journal of Multivariate Analysis* 12, 335-345.

**Andrews, B., Calder, M., and R., Davis** (2009): Maximum Likelihood Estimation for $\alpha$-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.

**Brock, W. A., Dechert, W. D., LeBaron, B., and J. A. Scheinkman** (1996): A test for independence based on the correlation dimension. *Econometric reviews*, 15(3), 197-235.

**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** (2018): Bootstrapping Non-Causal Autoregressions: with Applications to Explosive Bubble Modelling. Forthcoming in *Journal of Business and Economic Statistics*.

**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.

**Cline, D.** (1983) Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data. Ph.D. Thesis, Department of Statistics, Colorado State University.

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

**Davis, R., Knight, K. and J., Liu** (1992). M-estimation for Autoregressions with Infinite Variance. *Stochastic Processes and their Applications*, 40, 145–180.

**Davis, R., and H, Liu** (2012): Theory and Inference for a Class of Observation Driven Models with Application to Time Series of Counts. ArXiv preprint. arXiv:1204.3915.

**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.

**De Haan, L., Mercadier, C., and C. Zhou** (2016): Adapting Extreme Value Statistics to Financial Time Series: Dealing with Bias and Serial Dependence. *Finance and Stochastics*, 20, 321-354.

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

**Duchesne, P., Ghoudi, K. and B. Rémillard** (2012): On Testing for Independence between the Innovations of Several Time Series. *Canadian Journal of Statistics*, 40(3), 447-479.

**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** (2010): *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

**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 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, 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., Issler, J. V., and S. M., Telg** (2017): Mixed Causal-Noncausal Autoregressions with Strictly Exogenous Regressors. MPRA Paper 80767, Germany.

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

**Hecq, A., Telg, S., and L., Lieb** (2017a): Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates? *Econometrics*, 5, 1–22.

**Hecq, A., Telg, S., and L, Lieb** (2017b): 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, 17-40.

**Hill, B.M.** (1975): A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Mathematical Statistics* 3, 1163-1174.

**Hill, J.B.** (2010): On Tail Index Estimation For Dependent, Heterogeneous Data. *Econometric Theory*, 26, 1398-1436.

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

**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, 600-617.

**Ling, S.** (2007): A Double AR($p$) Model: Structure and Estimation. *Statistica Sinica*, 17, 161-175.

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

**Mason, D.** (1982): Laws of Large Numbers for Sums of Extreme Values. *Annals of Probability* 10, 754-764.

**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.

**Robert, C. Y., Segers, J., and Ferro, C. A.** (2009): A Sliding Blocks Estimator for the Extremal Index. *Electronic Journal of Statistics*, 3, 993-1020.

**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.

# Supplementary file: Complementary results and proofs

This Appendix consists of six sections of additional results: K) asymptotic prediction of the MAR$(1,q)$ when $\alpha \in (0,1)$ and an explicit example in the MAR$(1,1)$ case; L) expectation of MAR$(p,q)$ processes conditionally on a linear combination of past values and proof of the unit root property; M) conditional correlation structure of noncausal AR(1) processes, proofs of Proposition 3.3 and of the conditional variance of the MAR(1,1); N) proof of Lemma E.1; O) recursion over polynomials $P_h$ and $Q_h$; P) Cluster size distribution, an illustration with the noncausal AR(1); Q) complementary results on the empirical study and details about the estimation of excess clustering term structures; R) Complementary estimation of the financial series using the R package 'MARX'.

## K A complement to Corollary 3.1 in the case $\boldsymbol{\alpha} \in (0,1)$ and $q>1$

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

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

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

**Proof.**

To complete the proof of Corollary 3.1 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 $r\geq 0$,

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

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 (K.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 K.1 (Asymptotic predictions of the MAR(1,1) process)** Let $(X_t)$ be defined by Equation (3.3). From the explicit predictions formulated in Section 3.4, 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 3.1 are consistent with these equivalents. In particular, when $\alpha = 1$, we always have $|\phi| < 1 = |\psi|^{\alpha-1}$ and we get that, almost surely,

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

# L Unit root property and extension

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

**Proposition L.1** *Let $X_t$ be the MAR$(p,q)$ process solution of* (2.1) *with symmetric $\alpha$-stable errors, $1 < \alpha < 2$. Denoting $(d_k)$ the coefficients sequence of its MA$(\infty)$ representation, we have*

$$\mathbb{E}\left[X_t\right| X_{t-1}] = \frac{\sum\limits_{k\in\mathbb{Z}} d_k\left(d_{k+1}\right)^{<\alpha-1>}}{\sum\limits_{k\in\mathbb{Z}} |d_{k+1}|^{\alpha}} X_{t-1}.$$

The condition for the existence of a unit root is now straightforward.

**Corollary L.1** *Under the assumptions of Proposition L.1,*

$$\mathbb{E}\left[X_t\right| X_{t-1}] = X_{t-1} \qquad \Longleftrightarrow \qquad \sum_{k\in\mathbb{Z}} d_k\left(d_{k+1}\right)^{<\alpha-1>} = \sum_{k\in\mathbb{Z}} |d_{k+1}|^{\alpha}.$$

The case $\alpha \leq 1$ is more intricate because the expectation on the left-hand side of (L.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. Proposition L.1 is a consequence of the more general conditional expectation of $X_t$ given any linear combination of the past that we provide in the next result.

**Proposition L.2** *Let $X_t$ be the MAR$(p,q)$ process solution of* (2.1) *with symmetric $\alpha$-stable errors, $1<\alpha<2$. Denote $(d_k)$ the coefficients sequence of its MA$(\infty)$ representation. 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*

$$\mathbb{E}\Big[X_{t+h}\Big|\sum_{j=1}^{k} a_j X_{t-j}\Big] = \frac{\sum_{\ell\in\mathbb{Z}} d_{\ell-h}\left(\sum_{j=1}^{k} a_j d_{\ell+j}\right)^{<\alpha-1>}}{\sum_{\ell\in\mathbb{Z}}\left|\sum_{j=1}^{k} a_j d_{\ell+j}\right|^{\alpha}}(a_1 X_{t-1} + \dots + a_k X_{t-k}). \tag{L.1}$$

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

**Proof.**

Let us introduce $Y_{t-1,k} = a_1 X_{t-1} + \dots + a_k X_{t-k}$. Let $\varphi(u,v) = \mathbb{E}\left[e^{iuY_{t-1,k}+ivX_{t+h}}\right]$. For any $(u,v) \in \mathbb{R}^2$ we have,

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

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,

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

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 $u \in \mathbb{R}^*$:

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

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

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

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

In order to prove Proposition 3.3, 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.

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

**Lemma M.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 $\tau$:*

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

**Remark M.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 $\tau$:

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

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 $\tau \geq 0$,

$$\frac{\sum_{t=1}^{n-\tau+1} X_t X_{t+\tau}}{\sum_{t=1}^{n} X_t^2} \underset{n\to+\infty}{\overset{p}{\longrightarrow}} \psi^\tau.$$

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 M.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 $\tau$, 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*

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

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

This result yields Proposition 3.2 by taking $h = \tau = 0$, with $P_0(B) = \phi_1 + \phi_2 B + \ldots + \phi_q B^q$ and $Q_0(B) = 1$.

## M.2 Proof of Lemma M.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,

$$\frac{\partial \varphi}{\partial z} = -\alpha\sigma^\alpha \sum_{n\in\mathbb{Z}} d_{n+1}(xd_{n-k} + yd_{n-\ell} + zd_{n+1})^{<\alpha-1>}\varphi(x,y,z),$$

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

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

And on the other hand,

$$\frac{\partial \varphi}{\partial y} = -\alpha\sigma^\alpha \sum_{n\in\mathbb{Z}} d_{n-\ell}(xd_{n-k} + yd_{n-\ell} + zd_{n+1})^{<\alpha-1>}\varphi(x,y,z),$$

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

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

Hence,

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

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,

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

This yields for $\alpha = 1$,

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

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

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

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):

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

which concludes the proof.

## M.3 Proof of Proposition M.1

Let $k$ and $\ell$ 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,

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

The second and third terms can be expressed as:

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

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

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

## M.4 Proof of Proposition 3.3

The result of Proposition 3.3 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 M.1.

## M.5 Details on the conditional variance of the MAR(1,1) of Section 3.4

By Lemma E.1, the polynomial $Q_h$ intervening in Proposition 3.3 reads in the case of the MAR(1,1)

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

Applying Proposition 3.3, 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 $q_{i,h}$'s, the coefficients of polynomial $Q_h$, we can deduce that for $\psi > 0$

$$Q_h(\text{sign }\psi) = \frac{1-\phi^{h+1}}{1-\phi},$$
$$c_h = \psi^{-h-1}\sum_{i=0}^{h}\sum_{j=0}^{h}\phi^i\phi^j\psi^{\max(i,j)},$$

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

# N 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 $q$

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

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 $\phi$ 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 $h \geq q$,

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

# O A recursive scheme for computing polynomials $P_h$ and $Q_h$ of Lemma D.1

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

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

*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:

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

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

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

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

# P Cluster size distribution: the noncausal AR(1) case

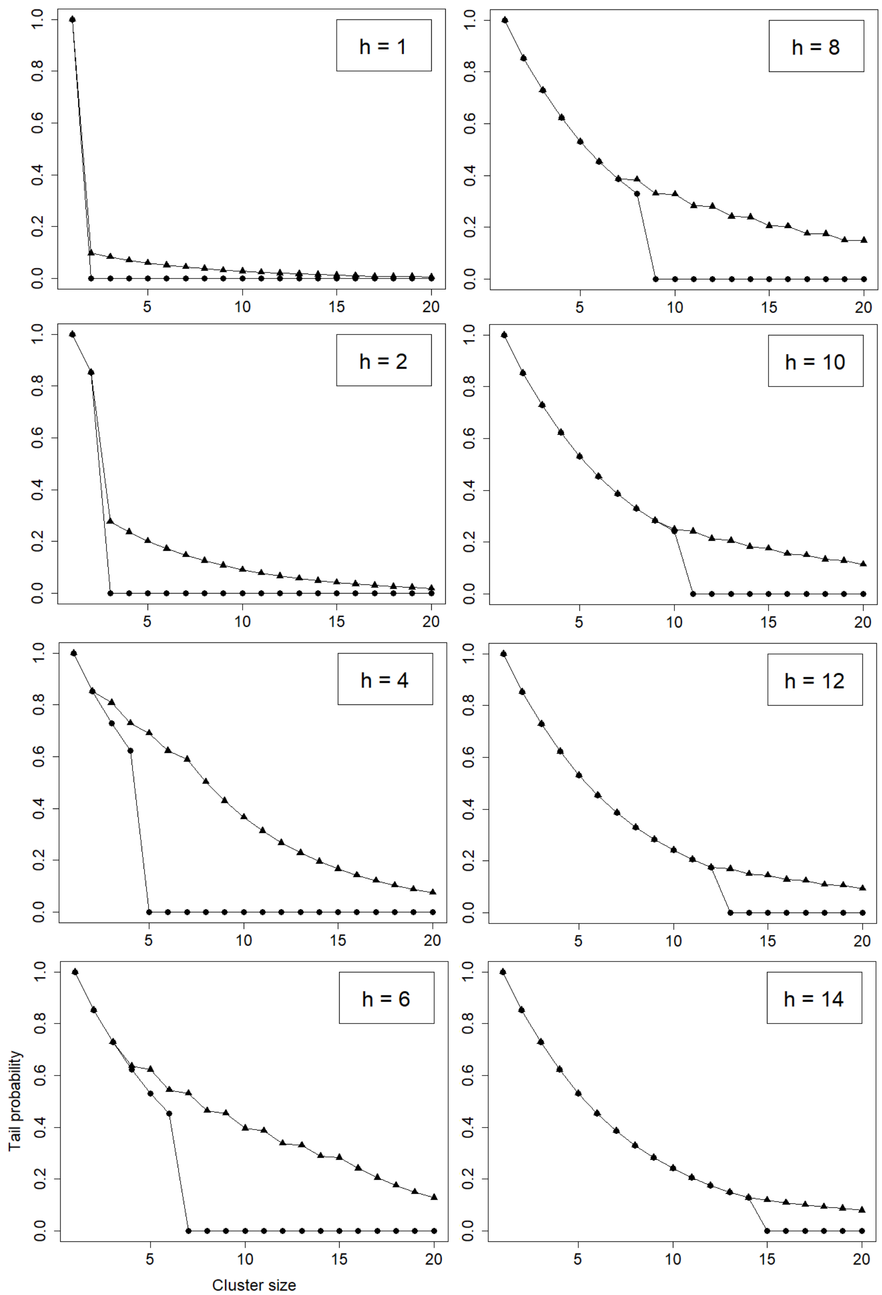



Figure 5: Theoretical tail probability given by Equation (4.16) of cluster sizes of extreme errors (4.18) (strong representation, points) and (4.19) (all-pass representation, triangles) for $\alpha = 1.5$, $\psi_0 = 0.9$ at different horizons $h$.

We illustrate the extreme clustering behaviours of the two error sequences (4.18) and (4.19) for various horizons and parameter values $\alpha = 1.5$, $\psi_0 = 0.9$. From equations (4.18) and (4.19), we deduce the sequence $(c_{(k)})$ and compute the tail probability distributions of the cluster size using

(4.16). As depicted on Figure 5, the contrast between the errors of the all-pass representations and those of the strong representations is the highest for intermediate values of $h$.

# Q Monte Carlo study: complementary results and methodology

## Q.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 |
| $\infty$ | $\hat{\delta}_1$ | -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 |
| $\infty$ | $\hat{\delta}_1$ | -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 Q.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 $\alpha$-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 (4.10). [See Example 4.1]

## Q.2 Direct implementation of the Portmanteau test

We conducted an experiment to assess the direct implementation of the portmanteau test (without Monte Carlo) and focused on $\alpha = 1.5$. We computed the residuals of the 100,000 simulated paths based on the all-pass causal AR(2) fits, evaluate the statistic (4.13) for $h = 1, \dots, 10$ 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 statistics to the appropriate quantile of the asymptotic distribution. The empirical sizes are reported in Table Q.2. The test suffers heavy distortions, especially in smaller 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 Q.2: Empirical sizes of portmanteau tests with nominal sizes 1%, 5% and 10% using the first $H$ lags, $H = 1, \dots, 10$ 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.

## Q.3 Extreme residuals clustering

### Q.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 $(\hat{\zeta}^i_{t+h|t})_t$, $i = 1, \dots, 6$, one each for the pure causal and noncausal AR(2) competitors, and two each for the two MAR(1,1) competitors (one for the causal component, one for the noncausal component). To compute the cluster sequences $(\hat{\xi}^i_{k,h}(x))_k$ as defined in Section 5.2 for each residuals series, we need to choose a threshold $x > 0$. It would be desirable to use thresholds such that we can harmoniously compare the clustering behaviours of the six series of residuals. For the experiment detailed below, we worked with the autostandardised series of residuals

$$\hat{v}^i_{t+h|t} := \left( \frac{\hat{\zeta}^i_{t+h|t}}{\max_s |\hat{\zeta}^i_{s+h|s}|} \right)_t , \tag{Q.1}$$

which lie between 0 and 1, and for each horizon $h$, we used the threshold

$$x_h := \max_{i=1,\dots,6} q_a\Big(|\hat{v}^i_{t+h|t}|\Big), \tag{Q.2}$$

where $q_a(\cdot)$ the $a$-percent quantile. In our experiments, $a = 0.9$ was used.

**Outline of the experiment**

For a given parameterisation $(\alpha, \psi_0, \phi_0)$ and path length $n$, we simulate 10000 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 $h \geq 1$:

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

$\iota\iota$) Obtain the set of (inverted) roots $\{\hat{\psi}, \hat{\phi}\}$ by solving for the zeros of $\hat{\eta}(z) = 1 - \hat{\eta}_1 z - \hat{\eta}_2 z^2$.

$\iota\iota\iota$) For each of the four competing models (5.1)-(5.4), decompose the process into pure causal and noncausal components and compute $(\hat{v}^i_{t+h|t})$, the series of autostandardised errors at horizons $h$ as in (Q.1).

$\iota\nu$) Compute $x_h$ as in (Q.2) and obtain the cluster sizes sequences $(\hat{\xi}^i_{k,h}(x_h))_k$ for each series $(\hat{v}^i_{t+h|t})$, $i = 1, \dots, 6$.

$\nu$) Compute the Excess Clustering at horizon $h$ of each residuals series as in (5.5).

$\nu\iota$) For the two MAR(1,1) competitors, average the Excess Clustering indicators obtained from the residuals of the causal and noncausal components.

For a given simulated path $(X_t)$, we repeat the above steps for horizons $h = 1, \dots, H$ and obtain four term structures of Excess Clustering, one for each competing models (5.1)-(5.4). Across the 10000 simulated paths of $(X_t)$, one can then either:

(i) average model-wise across the obtained term structures to gauge the typical excess clustering behaviour of each competing model (as in Figures 3 and Q.1), or

(ii) for each of the simulated paths $(X_t)$, compute the area under the four term structures, select the least clustering model and evaluate the rate of correct selections (as in Table 3).

### Q.3.2 Excess clustering for additional parameterisations

We evaluated the residuals excess clustering behaviours of the four alternatives (5.1)-(5.4) for additional parameterisations and sample sizes of the MAR(1,1) data generating process. Excess

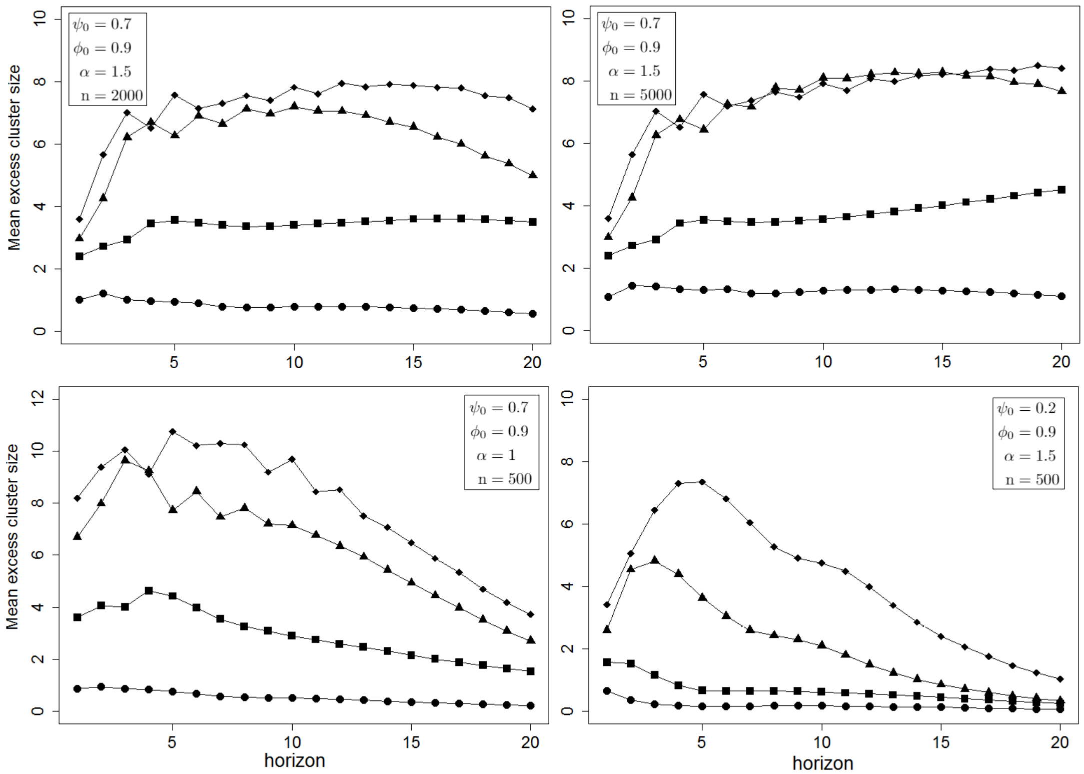



Figure Q.1: Across 10,000 simulations of the $\alpha$-stable MAR(1,1) process $(X_t)$ solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, average of the term structure of excess clustering of the linear residuals of the four competing models (5.1) (squares), the strong representation (5.2) (points), (5.3) (triangles) and (5.4) (diamond). The parameterisations and path lengths are indicated on each panel.

clustering in all-pass residuals is apparent even for small sample sizes. The contrast between the residuals of the strong representation and those of the all-pass increases as the sample size grows (see the left panel of Figure 3 and the two upper panels of Figure Q.1). Also, even with a much smaller noncausal parameter $\psi = 0.2$ (lower right panel of Figure Q.1), the strong representation still clearly displays the least excess clustering compared to the three other competitors. We can nevertheless 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 $\psi$ is relatively small, especially compared to the causal parameter $\phi$, yielding much weaker dependence across the residuals of the misspecified pure causal AR(2).

# R Real data: complementary results using the R package 'MARX'

## R.1 Total AR orders selection by Information Criterion

The portmanteau procedure of Section 6.1 allowed to discard non-admissible low order models for the six financial time series considered. Portmanteau tests are however not designed to select an «optimal» model. To go further, we report in Table Q.3 the orders that minimise Akaike's information criterion (AIC) using the R package 'MARX' available on CRAN (see Hecq, Telg and Lieb (2017b)). The validity of such AIC's for innovations in the domain of attraction of a stable law has been studied by Knight (1989). Except for the HSI, the results of the two procedures are highly compatible.

| | Boeing | Exxon | Coca-Cola | Walmart | HSI | Shiller P/E |
|---|---|---|---|---|---|---|
| Selected total AR order | 4 | 1 | 1 | 1 | 1 | 8 |

Table Q.3: Optimal order minimising the AIC criterion.

## R.2 Identification of causal and noncausal roots

Given the lowest total AR orders validated by the portmanteau procedure (see Table 6.1), we used the routine marx.t of the 'MARX' package to fit MAR models on the six financial series by t-Student ML. The results are presented in Table Q.4. For the MAR(1,0) models, the results appear very similar but we note some discrepancies for higher-order models.

| Series | Final specification | Noncausal (inverted) roots | Causal (inverted) roots |
|---|---|---|---|
| Boeing | MAR(2,0) | 0.96, 0.11 | – |
| Exxon | MAR(1,0) | 0.97 | – |
| Coca-Cola | MAR(1,0) | 0.93 | – |
| Walmart | MAR(1,0) | 0.93 | – |
| HSI | MAR(3,0) | 0.92, 0.28, $-0.21$ | – |
| Shiller P/E | MAR(4,2) | 0.95, $-0.45$, $0.51 \pm 0.23i$ | $-0.21 \pm 0.44i$ |

Table Q.4: Estimation of the MAR$(p, q)$ specification for each financial series by t-Student ML using the routine marx.t of the 'MARX' package. This routine requires as input the total AR order $p+q$, we used the validated orders given by Table 4.

## References

**Knight, K.** (1989): Consistency of Akaike's information criterion for infinite variance autoregressive processes. *The Annals of Statistics*, 824-840.