Econometric Theory, **35**, 2019, 1234–1270.
doi:10.1017/S0266466618000452

MIXED CAUSAL-NONCAUSAL AR PROCESSES AND THE MODELLING OF EXPLOSIVE BUBBLES

SÉBASTIEN FRIES
CREST and Paris-Saclay University

JEAN-MICHEL ZAKOIAN
CREST and University of Lille

Noncausal autoregressive models with heavy-tailed errors generate locally explosive processes and, therefore, provide a convenient 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.

1. INTRODUCTION

In the analysis of prices of financial assets such as stocks, it is common to observe phases of locally explosive behaviors, 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

The authors are grateful to three anonymous referees, the Co-editor, and the Editor for insightful comments, suggestions, and useful criticisms. The second author acknowledges financial support from the Agence Nationale de la Recherche (ANR), which supported this work via the Project MultiRisk (ANR CE26 2016 - CR), the Labex ECODEC and the Chaire ACPR "Régulation et risques systémiques". Address correspondence to Jean-Michel Zakoian, CREST and University of Lille, CREST, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France; e-mail: zakoian@ensae.fr.

 © Cambridge University Press 2019

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

polynomials with roots inside the unit disk) are introduced. For instance, all-pass models[1] are linear time series with nonlinear behaviors, 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 typical nonlinear behavior observed in 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 and 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 research studies have focused on data-generating processes that are able to produce explosive behaviors 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. Contrary to the latter references, noncausal AR processes do not provide a mechanism for real-time analysis, including real-time detection of bubble phenomena. However, apart from the generation of bubbles in a stationary framework, 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 behaviors of economic series, such as the occurrence of locally explosive behaviors with different rates of explosion, or different types of asymmetries in the growth and downturn phases of the bubbles.

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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 organized as follows: Section 2 studies the sample paths and the marginal distribution of MAR processes with stable errors. Section 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 the 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 online in supplementary material associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

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

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

B and F are the usual lag and forward operators ($B^k X_t = X_{t-k}, F^k X_t = X_{t+k}, k \in \mathbb{Z}$), (ε_t) is an independent and identically distributed (i.i.d.) sequence, and the polynomials ψ and ϕ 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 ε_t follow a stable non-Gaussian distribution but the assumption will be relaxed for the statistical inference. The generality and convenience of this class of distributions is now well established.[3] Stable laws are easily characterized through their characteristic function: ε_t is said to follow a

[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.

[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 average processes based on stable variables also follow stable distributions, as will be detailed below.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

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

where $w(\alpha, s) = \operatorname{tg}\left(\frac{\pi\alpha}{2}\right)$ 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 β. When $\beta = 1$, the density of the errors is maximally skewed towards positive values, yielding trajectories like (a) and (c) which could be suited to model prices or volatilities. In particular, trajectory (a) displays bubble patterns similar to those of real prices (see for instance Figure 4 below). The influence of a smaller tail parameter α is visible when comparing trajectories (c) and (d): the extreme events

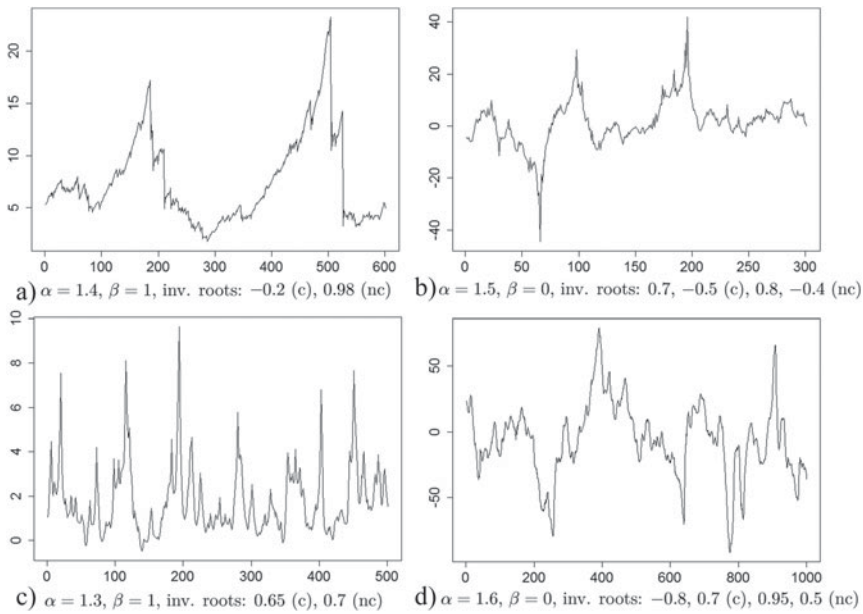


FIGURE 1. Examples of trajectories of MAR(1,1) (left panel) and MAR(2,2) (right panel) processes with different parameters (nc: inverse of noncausal roots; c: inverse of causal root).

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

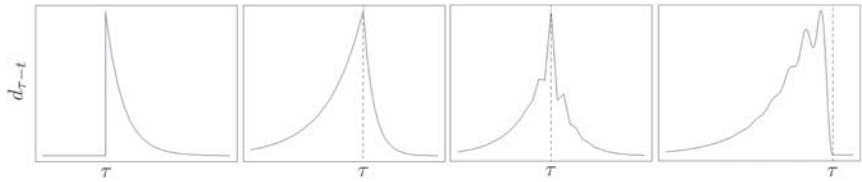


FIGURE 2. Examples of baseline paths $t \longmapsto d_{\tau-t}$ of MAR processes with characteristic polynomials, from left to right: $1-0.7B$; $(1-0.9F)(1-0.7B)$; $(1-0.8F)(1+0.4F)(1-0.7B)(1+0.5B)$; $(1-0.99F)(1-965F)(1-0.98e^{i0.045\pi}F)(1-0.98e^{-i0.045\pi}F)$.

of the former ($\alpha = 1.3$) are more recurrent and further away from the central values than those of the latter ($\alpha = 1.6$).

Under the assumptions made on the AR polynomial, (X_t) admits a two-sided MA(∞) 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 ε_τ. Figure 2 depicts such baseline paths for four different MAR processes. The first panel illustrates the well-known impulse response function of a classical causal AR(1). The second panel displays an explosive exponential trend followed by a downward, faster decay and corresponds to the baseline path of a MAR(1,1) process. The remaining panels show more complex trajectories: the third one depicts the baseline path of a MAR(2,2) with dented upward and downward trends, whereas the last one, corresponding to a noncausal AR(4) with two real and two conjugated complex roots, shows an upward trend with oscillations of increasing amplitudes and fixed pseudo-periods.

2.2. Marginal Distribution

Our first result characterizes 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* ψ *and* ϕ *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*

[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 $s < \alpha$.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

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

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

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 error 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* ψ *and* ϕ *are outside the unit disk and* (ε_t) *is an i.i.d. sequence* (*not necessarily stable*).

Then, (X_t) *is a 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 behavior for the conditional moments, generalizing 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* $\varepsilon_t \sim \mathcal{S}(\alpha,\beta,\sigma,\mu)$ *and* $p \geq 1$, *we have*

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

The conditional distribution in direct time, that is with respect to the past observations, thus has lighter tails than the marginal distribution which is α-stable. In particular, whatever the heaviness of the tails of ε_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]

[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), Chap. 2), or models for time series of counts (Davis and Liu, 2012).

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

3.2. Prediction of Future Values for the MAR$(1,q)$ Processes

Prediction at any horizon can be fully characterized for the symmetric MAR$(1,q)$ process. The next proposition extends in a nontrivial 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

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

and we have the semistrong causal representation

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

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

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 semistrong 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), Sect. 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 $\psi > 0$,

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

[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 $\sigma^2 = \text{Var}(\varepsilon_t)$. Therefore, (Z_t) is also a white noise with the same variance σ^2 as ε_t.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

with by convention $\phi_1 = \cdots = \phi_q = 0$ when $q = 0$. Hence, the martingale property established by GZ (Proposition 3.3), $\mathbb{E}\left[X_t \middle| \mathscr{F}_{t-1}\right] = X_{t-1}$, only holds for the noncausal AR(1) (i.e., when $q = 0$).

The asymptotic behavior of the conditional expectation—when the horizon h tends to infinity—is highly dependent on the tail index α. Proposition 3.2 allows us to distinguish different behaviors summarized in the following Corollary.

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

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

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

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

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 online Supplement.

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 characterize 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*

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

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.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

COROLLARY 3.2. *Under the assumptions of Proposition 3.3, there exists a sequence* (η_t) *of random variables such that,*

$$(1-sign(\psi)B)\phi(B)X_t=\sigma_t\eta_t,$$
$$\sigma_t^2=\left(\frac{1}{|\psi|}-1\right)(X_{t-1}-\phi_1X_{t-2}-\cdots-\phi_qX_{t-q-1})^2+\frac{\sigma^2}{|\psi|(1-|\psi|)},$$

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

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

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(∞) 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

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

In particular, when $\psi,\phi>0$ 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 $h\geq 0$,

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

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

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 (ε_t) is an α-stable sequence and rather assume that the law of ε_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 α-stable framework under which the properties of the previous section were derived. Replacing α-stable laws by their domain of attraction alleviates the risk of misspecification.[10]

[8] See Andrews et al. (2009) for asymptotic properties of the ML estimator of both causal and noncausal AR processes with non-Gaussian α-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 of least-absolute deviation (LAD) estimators by An and Chen (1982), and for M-estimators by Davis, Knight, and Liu (1992).

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

We will first derive the asymptotic properties of estimators of an "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 et al. (2001, Sect. 4.3) 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

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

where (ζ_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 η_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 (ζ_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_ks are the MA(∞) 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 η_{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.

[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.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

4.2. Least-Squares Estimation

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

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

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

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

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 $\ell = 1, \ldots, p+q$, let $\boldsymbol{K}^{(\ell)} = \boldsymbol{J}^\ell + {}^t\boldsymbol{J}^\ell$ (with $\boldsymbol{K}^{(p+q)} = \boldsymbol{0}$). Let $\boldsymbol{L} = [\boldsymbol{K} \quad \boldsymbol{K}^{(2)} \quad \ldots \quad \boldsymbol{K}^{(p+q)}]$. We start by providing the asymptotic behavior of the LS estimator under the simplifying assumption that the distribution of ε_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 (ε_t) satisfying* (4.2) *and $\mathbb{E}|\varepsilon_t|^\alpha = \infty$.*
Then, letting $\boldsymbol{\rho} = [\rho(i)]_{i=1,\ldots,p+q}$, $\boldsymbol{R} = [\rho(i-j)]_{i,j=1,\ldots,p+q}$,

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

$Z_k = \sum_{l=1}^{+\infty} \{\rho(k+l) + \rho(k-l) - 2\rho(l)\rho(k)\} S_l/S_0$, for $k = 1, \ldots, p+q$, and $S_0, S_1, S_2, \ldots$ are independent stable random variables; S_0 is positive with index $\alpha/2$ and S_j, for $j \geq 1$, has index α. If the law of $|\varepsilon_t|$ is asymptotically equivalent to a Pareto, (4.10) *holds with $a_n^2/\tilde{a}_n = (n/\ln n)^{1/\alpha}$.*

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

The α-stable domain of attraction assumption on (ε_t) impacts the asymptotic behavior of the LS estimator in two important aspects: $\imath$) the limiting distribution depends on α and $\imath\imath$) 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 α-stable domain of attraction (see Cline (1983), Thm. 3.3iv) 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 $\mathcal{R}_{n,\underline{\alpha}}$ be 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 $\mathcal{R}^*_{n,\underline{\alpha}}$ be the symmetric set of $\mathcal{R}_{n,\underline{\alpha}}$ around the line $r_1 = r_2$, we get an asymptotic confidence region for (ϕ_0, ψ_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 α 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 α. 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 α. 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 ith order statistic in decreasing order $(X_{(1)} \geq X_{(2)} \geq \cdots X_{(n)})$. 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 ℓ-dependence, β-mixing, ARCH—were established by various authors (see e.g., Hill (2010), De Hann, Mercadier, and Zhou (2016) and the references therein). An alternative to the estimation of the asymptotic distribution is to base inference on the bootstrap. Recently, Cavaliere, Nielsen, and Rahbek (2018) proposed bootstrap schemes for noncausal AR models with infinite variance, and showed their

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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 behavior of the LS estimator of η_0 assuming the errors (ε_t) were symmetrically distributed. We here relax the symmetry assumption and only require (ε_t) to satisfy (4.2). The asymptotic behavior 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 $\tilde{\rho}(h) = \tilde{\gamma}(h)/\tilde{\gamma}(0)$ the mean-adjusted sample autocorrelation of order h. Similarly to (4.8), define the mean-adjusted Yule-Walker estimator $\tilde{\boldsymbol{\eta}}$ by

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

PROPOSITION 4.4. *Let (X_t) be the strictly stationary solution of Model (4.1), where (ε_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 $\hat{\boldsymbol{\eta}} = (\hat{\eta}_i)_{i=1,\ldots,p+q}$, the corresponding residuals are defined by

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

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 $\hat{\boldsymbol{\rho}}_{\hat{\zeta}} = [\hat{\rho}_{\hat{\zeta}}(1),\ldots,\hat{\rho}_{\hat{\zeta}}(H)]'$.

PROPOSITION 4.5. *Under the assumptions of Proposition 4.3, the vector of residual empirical autocorrelations satisfies*

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

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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 error 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

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

with $\|\boldsymbol{x}\|_1 = \sum |x_i|$ for any vector $\boldsymbol{x} = (x_i)$. Practical implementation of the test finally requires simulating/bootstrapping 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 note11). Having estimated the coefficients of the polynomial η_0, a natural strategy for assessing the validity of a candidate model, with polynomial η_0^*, is to test the independence of the ζ_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 materializes 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(∞) representation $Y_t = \sum_{k\in\mathbb{Z}} c_k \varepsilon_{t+k}$, where (ε_t) is an i.i.d. sequence satisfying (4.2), $\sum_{k\in\mathbb{Z}} |c_k|^s < +\infty$ for some $0 < s < \alpha$, $s \le 1$, and assume $\max |c_k| = 1$ for convenience. In our context, Y_t will typically be substituted for the errors ε_t of the strong representation and the errors ζ_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 (1985), Sect. 3.D.):

[13] Alternative approaches based on nonparametric 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 $\sqrt{n}$-consistent estimators of the model parameters, which is not the case in the α-domain of attraction framework as shown in Proposition 4.3.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

where δ is the Dirac measure, $B_x=(0,+\infty)\times\Big((-\infty,-x)\cup(x,+\infty)\Big)$, $\{\Gamma_k,k\geq 1\}$ are the points of a homogeneous Poisson Random Measure (PRM) on $(0,+\infty)$ with rate $x^{-\alpha}$,[14] and $\xi_k=\text{Card}\{i\in\mathbb{Z}:J_k|c_i|>1\}$ 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 now outline some reasons for analyzing extreme event clustering and the potential for model selection amongst competing representations.

4.5.2. Analyzing the Error Processes of Competing Models. Let (X_t) be the MAR(p,q) strictly stationary solution of Model (4.1)–(4.2) and assume the order $p+q$ and the roots of η_0 are known. There is a finite number of competing representations among which the strong one lies. Denoting (ε_t) of the errors of the strong representation and generically (ζ_t^*) the errors of any specific all-pass representation, we can analyze their extreme clustering behavior using (4.14).

1. Error process of the strong representation. The i.i.d. errors (ε_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 ε_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 (ξ_k) describing the size of the cluster of extremes are degenerate in this case: the extreme errors (ε_t) tend to appear isolated from each other.

2. Error process of an all-pass representation. The (rescaled) errors ζ_t^* of an all-pass representation always admit an infinite one-sided or two-sided MA form, say $\zeta_t^*=\sum_{k\in\mathbb{Z}}\frac{c_k}{\max_j|c_j|}\varepsilon_{t+k}$. Denote $(c_{(k)})_{k\geq 1}$ the sequence obtained by sorting $(|c_k|)_{k\in\mathbb{Z}}$ in descending order. Substituting Y_t for ζ_t^*, (4.14) holds with

[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 any nonnegative integers $n_1,\ldots,n_\ell$:

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

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

$\xi_k = \text{Card}\left\{i \geq 1 : J_k \frac{c_{(i)}}{c_{(1)}} > 1\right\} = \arg\max_{i\geq 1}\{J_k > c_{(i)}^{-1} c_{(1)}\}$, 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 ξ_k's can take arbitrarily high values with nonzero probability, indicating as expected that the extremes of (ζ_t^*) tend to cluster.

3. 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:

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

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

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 realizations 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 online Supplement for illustration). This analysis can be extended to general MAR processes by disentangling the pure causal and noncausal components of each competing 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, contrary 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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

data generated from a MAR(1,1) process, we assessed *ι*) the consistency of the estimators of the roots and the convergence in distribution of the LS estimators of the backward AR(2) specification, *ιι*) the empirical size of the portmanteau-type statistic, and *ιιι*) 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, 2,000, and 5,000 observations of α-stable MAR(1,1) processes solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$ with $\psi_0 = 0.7$, $\phi_0 = 0.9$, and tail indices $\alpha = 1.5$, 1, and 0.5. We computed the LS estimator $(\hat{\eta}_1, \hat{\eta}_2)$ and deduced estimators $(\hat{\psi}, \hat{\phi})$ by taking the inverses of the zeros of $1 - \hat{\eta}_1 X - \hat{\eta}_2 X^2$ (we 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 α approaches zero.

Turning to the asymptotic distribution of $(\hat{\eta}_1, \hat{\eta}_2)$, results reported in the online Supplement show that the finite sample distribution approaches its asymptotic behavior much slower for lower values of α. 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

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 ψ, for $a = 0.01$, 0.05, 0.1 and over 100,000 simulated paths of the α-stable MAR(1,1) process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, with $\psi_0 = 0.7$ and $\phi_0 = 0.9$

		$\alpha = 1.5$			$\alpha = 1$			$\alpha = 0.5$		
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%
2,000	$\hat{p}_a(\phi)$	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%
5,000	$\hat{p}_a(\phi)$	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%

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

TABLE 2. Empirical sizes (%) of the portmanteau statistics (4.13) implemented by the Monte Carlo test procedure. The empirical size was calculated based on 2,000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$, with $\psi_0 = 0.7$ and $\phi_0 = 0.9$. Each Monte Carlo test was performed with 1,000 simulations

	$\alpha = 1.5$			$\alpha = 1$			$\alpha = 0.5$		
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

procedure remains valid.[15] We therefore proceeded with this methodology (see the online Supplement). The empirical sizes of the 1, 5, and 10% nominal tests for lags $H = 1, \ldots, 10$ 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.

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 α-stable MAR(1,1) process $(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t$ with different parameterizations and analyzing the residuals of the competing representations. There are four competing models yielding the same all-pass causal AR(2) representation:

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

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

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

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

where (ζ_t), (ν_t), and (ω_t) denote the sequences of errors of each all-pass representation. 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) alternative, 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,

[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).

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

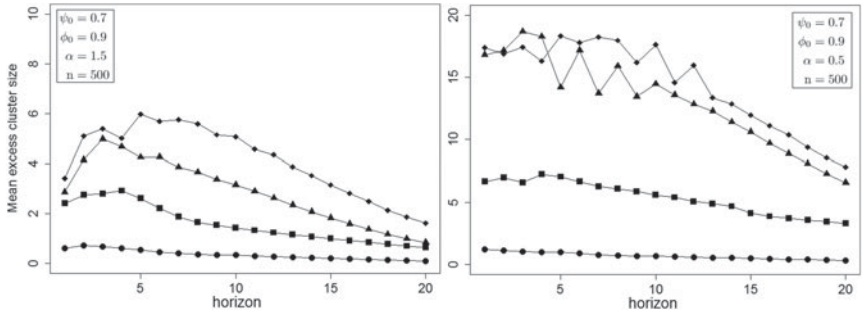


FIGURE 3. Across 10,000 simulations of the α-stable MAR(1,1) process (X_t) solution of $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$, the plots show the average of the term structure of excess clustering of the linear residuals of the four competing models (5.1) (squares), the strong representation (5.2) (points), (5.3) (triangles), and (5.4) (diamonds). The parameterizations and path lengths are indicated on each panel.

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:

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

We start by generating 10,000 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 compute the indicator (5.5) for $h = 1, \ldots, 20$. We refer to vectors $(EC_1, \ldots, EC_{20})$ as estimators of the *term structure* of residuals excess clustering. Averaging model-wise across the 10,000 simulations yields the typical excess clustering behaviors of the residuals of each competing model. We perform this experiment for several MAR(1,1) processes and display the results of two parameterizations in Figure 3 (see the online Supplement 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 barely deviate from no excess clustering. As we could expect from (4.16), the heavier the tails the easier it is to identify dependent residuals. This is in line with the findings of Hecq et al. (2016) who are concerned with identification of causal/noncausal models

[16] For a given h, 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, Segers, and Ferro (2009)).

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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 estimators of the term structure 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. For $\alpha = 1.5$, the true positive rates of this procedure read 88.4% for $n = 500$, 95.8% for $n = 2,000$, and 97.5% for $n = 5,000$.

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 and economic series: cotton price (monthly, USD/pound, August 1972 to July 2017),[17] soybean price (monthly, USD/bushel, January 1973 to May 2006),[18] sugar price (monthly, USD/pound, November 1962 to August 2018),[19] coffee price (monthly, USD/pound, April 1976 to May 2018),[20] Hang Seng Index (HSI) (monthly, HKD, November 1986 to Mars 2017),[21] and the quarterly Shiller Price/Earning ratio (Q1 1881 to Q2 2017).[22]

In the following study, a linear trend is fitted and subtracted from the HSI series and the other series are centered. These series are depicted on Figure 4.

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 (1,000 paths were simulated for each test). The results of the portmanteau test, reported in Table 3, allow to discard nonadmissible low order models at the level 5%. We retain for the following the

[17] https://www.macrotrends.net/2533/cotton-prices-historical-chart-data.
[18] https://www.macrotrends.net/2531/soybean-prices-historical-chart-data.
[19] https://www.macrotrends.net/2537/sugar-prices-historical-chart-data.
[20] https://www.macrotrends.net/2535/coffee-prices-historical-chart-data.
[21] https://finance.yahoo.com/quote/%5EHSI/history?p=%5EHSI.
[22] https://www.quandl.com/data/multpl/shiller_pe_ratio_month-shiller-pe-ratio-by-month.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

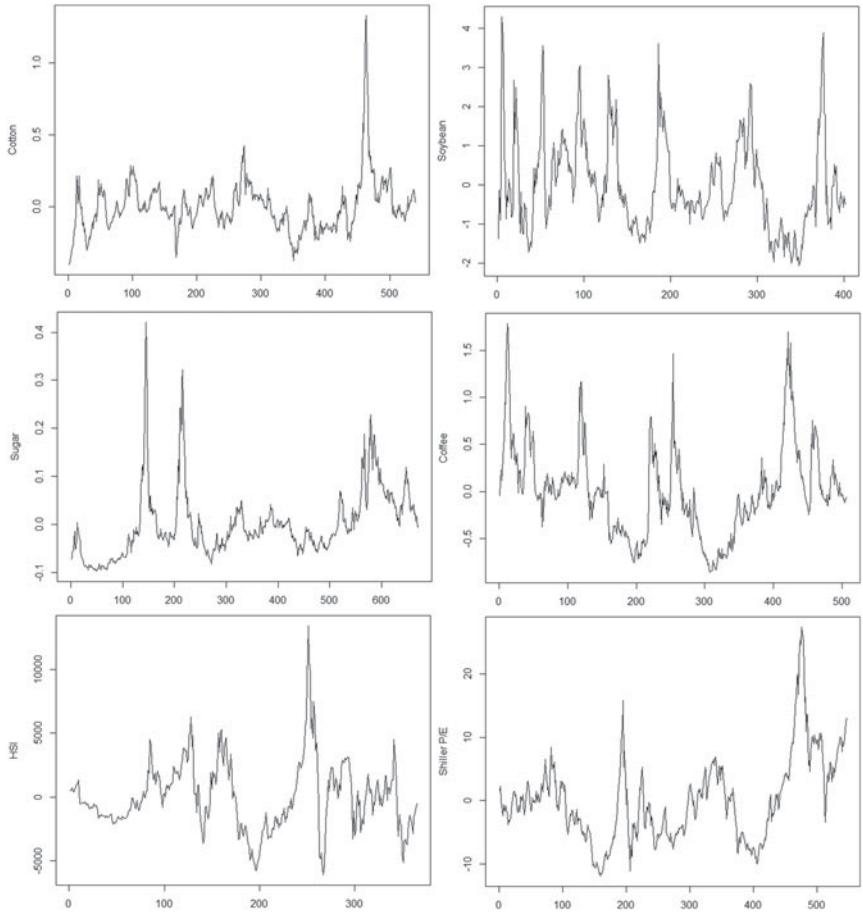


FIGURE 4. Financial series paths: cotton (8/1972 to 7/2017), soybean (1/1973 to 5/2006), sugar (11/1962 to 8/2018), coffee (4/1976 to 5/2018), Hang Seng index (HSI, 11/1986 to 3/2017), and Shiller's P/E ratio (Q1/1881 to Q2/2017). All series are monthly except the latter which is quarterly. A linear trend is fitted and subtracted from the HSI series and the other series are centered.

lowest orders (indicated in bold) which pass the portmanteau procedure: cotton: 2; soybean: 5; sugar: 4; coffee: 4; HSI: 3; Shiller P/E: 6.[23]

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

[23] This procedure yields as a by-product the McCulloch quantile estimates of the tail index α (see McCulloch (1986)) for the six series. Values of $\hat{\alpha}$: cotton: 1.53; soybean: 1.41; sugar: 1.35; coffee: 1.39; HSI: 1.34; and Shiller P/E: 1.49. In all the cases, the infinite variance hypothesis is plausible.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

TABLE 3. P-values (%) of the Monte Carlo portmanteau tests with $H = 50$ lags for increasing AR order r. Rejection if P-value $< 5\%$

Total AR order r	Cotton	Soybean	Sugar	Coffee	HSI	Shiller P/E
1	0.30	1.50	0.30	0.70	4.50	0.10
2	**6.60**	0.90	3.30	0.80	4.50	0.20
3		1.60	4.50	4.00	**7.10**	0.20
4		2.60	**8.20**	**7.99**		0.40
5		**15.4**				1.00
6						**5.90**

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 4, 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. One can first notice that except for $r = 1$, the favored specifications always display a mixed structure with both nontrivial causal and noncausal components. One can furthermore see that overall, excess clustering decreases as r increases: as the total AR order increases, the models brought to the data gain in flexibility and can account for more complex patterns, improving the fitting. This is quite striking in the case of the Shiller P/E ratio for which the median AUC of nonfavored models is halved for each incremention of the total AR order from $r = 1$ to $r = 4$. 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,

TABLE 4. Selection based on extreme clustering for each of the six financial series. For a given total AR order $r = p+q$, the table provides the favored MAR(p,q) specification exhibiting the least AUC of excess clustering, its corresponding AUC, and the median AUC of all the other possible specifications

Total AR order		Cotton	Soybean	Sugar	Coffee	HSI	Shiller P/E
1	Favored specification	(0,1)	(0,1)	(1,0)	(0,1)	(0,1)	(1,0)
	AUC Excess Clustering	23.3	6.50	28.2	9.53	7.75	15.7
	Median of competitors	29.7	7.03	41.7	13.6	16.5	51.9
2	Favored specification	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)	(1,1)
	AUC Excess Clustering	8.25	3.9	22.3	5.35	5.75	8.54
	Median of competitors	16.1	6.10	29.1	8.97	9.33	26.4
3	Favored specification	(1,2)	(2,1)	(1,2)	(1,2)	(1,2)	(1,2)
	AUC Excess Clustering	6.20	2.30	21.4	2.83	0.50	2.13
	Median of competitors	10.5	3.45	83.1	9.60	6.17	12.4
4	Favored specification	(1,3)	(3,1)	(3,1)	(1,3)	(1,3)	(1,3)
	AUC Excess Clustering	5.83	2.42	6.67	3.00	0.50	1.63
	Median of competitors	10.1	3.38	14.8	5.06	5.66	6.25
5	Favored specification	(1,4)	(2,3)	(2,3)	(1,4)	(3,2)	(3,2)
	AUC Excess Clustering	3.00	1.75	3.17	3.00	2.00	2.08
	Median of competitors	8.30	3.00	13.3	5.38	5.67	5.17
6	Favored specification	(1,5)	(2,4)	(2,4)	(3,3)	(1,5)	(2,4)
	AUC Excess Clustering	3.00	1.40	4.82	2.50	0.50	1.33
	Median of competitors	8.30	3.61	17.3	6.30	4.20	4.67

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

TABLE 5. Selection of the MAR specification for each financial series among the favored ones of Table 4 based on the total AR order determined in Table 3. The MAR(p,q) specifications indicate the noncausal p and causal q orders as well as the (inverted) roots of the corresponding polynomials

Series	Final specification	Noncausal (inverted) roots	Causal (inverted) roots
Cotton	MAR(1,1)	0.92	0.17
Soybean	MAR(2,3)	$0.37 \pm 0.40i$	0.87, -0.60 , -0.07
Sugar	MAR(3,1)	0.94, -0.08 ± 0.49	0.52
Coffee	MAR(1,3)	0.55	0.92, $-0.29 \pm 0.21i$
HSI	MAR(1,2)	0.37	0.89, -0.26
Shiller P/E	MAR(2,4)	$0.58 \pm 0.29i$	0.97, -0.70, $-0.21 \pm 0.60i$

- Assign the total AR order r validated by the portmanteau test (see Table 3).
- For the assigned order r, select then the least excess clustering competing representation (see the favored specifications in Table 4).

The selection is reported in Table 5 which shows the causal and noncausal orders as well as the (inverted) roots of the corresponding polynomials. For all series, mixed models with nontrivial causal and noncausal components are favored, which is compatible with the upward/downward trends displayed by most extreme events featured in the trajectories.

7. CONCLUDING REMARKS

Noncausal models may provide better understanding of the dynamic features of a time series that are not perceived via causal models. Even the addition 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 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.

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 et 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.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

REFERENCES

An, H.Z. & Z.G. Chen (1982) On convergence of LAD estimates in autoregression with infinite variance. *Journal of Multivariate Analysis* 12, 335–345.

Andrews, B., M. Calder, & R. Davis (2009) Maximum likelihood estimation for α-stable autoregressive process. *Annals of Statistics* 37, 1946–1982.

Andrews, B. & R.A. Davis (2013) Model identification for infinite variance autoregressive processes. *Journal of Econometrics* 172(2), 222–234.

Breidt, F.J., R.A. Davis, & A.A. Trindade (2001) Least absolute deviation estimation for all-pass time series models. *Annals of Statistics* 29, 919–946.

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

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

Cavaliere, G., H.B. Nielsen, & A. Rahbek (2018) Bootstrapping non-causal autoregressions: With applications to explosive bubble modelling. *Journal of Business and Economic Statistics*, forthcoming.

Chavez-Demoulin, V. & A.C. Davison (2012) Modelling time series extremes. *REVSTAT-Statistical Journal* 10, 109–133.

Chen, Y., P.C.B. Phillips, & J. Yu (2017) Inference in continuous systems with mildly explosive regressors. *Journal of Econometrics* 201, 400–416.

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. & D. Vere-Jones (2007) *An Introduction to the Theory of Point Processes*, vol. 1. Springer.

Davis, R., K. Knight, & J. Liu (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* 40, 145–180.

Davis, R. & H. Liu (2012) Theory and inference for a class of observation driven models with application to time series of counts. ArXiv e-prints 1204.3915.

Davis, R. & S. Resnick (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* 13, 179–195.

Davis, R. & S. Resnick (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* 14, 533–558.

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

De Haan, L., C. Mercadier, & 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 D.F. Findley (ed.), *Applied Time Series Analysis II*, pp. 565–608. Academic Press.

Duchesne, P., K. Ghoudi, & B. Rémillard (2012) On testing for independence between the innovations of several time series. *Canadian Journal of Statistics* 40(3), 447–479.

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

Ferro, C.A. & 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. & J-M. Zakoian (2010) *GARCH Models*: *Structure, Statistical Inference and Financial Applications.* Wiley.

Gourieroux, C. & J. Jasiak (2016) Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis* 37(3), 405–430.

Gourieroux, C., J. Jasiak, & A. Monfort (2016) Stationary Bubble Equilibria in Rational Expectation Models. CREST Working paper 2016-31.

Gouriéroux, C. & J.-M. Zakoian (2017) Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*) 79, 737–756.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

Hecq, A., J.V. Issler, & S.M. Telg (2017) Mixed Causal-Noncausal Autoregressions with Strictly Exogenous Regressors. MPRA paper 80767.
Hecq, A., L. Lieb, & S.M. Telg (2016) Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics* 123/124, 307–331.
Hecq, A., S. Telg, & L. Lieb (2017a) Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics* 5, 1–22.
Hecq, A., S. Telg, & 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. & C. Gouriéroux (2015) Noncausal autoregressive model in application to bitcoin/USD exchange rates. In V.-N. Huynh, V. Kreinovich, S. Sriboonchitta, K. Suriya (eds.), *Econometrics of Risk*, pp. 17–40. Springer International Publishing.
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., J. Hüsler, & M.R. Leadbetter (1988) On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* 78, 97–112.
Hsueh, A.C. & J.M. Mendel (1985) Minimum-variance and maximum-likelihood deconvolution for noncausal channel models. *IEEE Transactions on Geoscience and Remote Sensing* 23, 797–808.
Lanne, M., H. Nyberg, & E. Saarinen (2012) Does noncausality help in forecasting economic time series? *Economics Bulletin* 32, 2849–2859.
Lanne, M. & P. Saikkonen (2011) Noncausal autogressions for economic time series. *Journal of Time Series Econometrics* 3.
Leadbetter, M.R. & S. Nandagopalan (1989) On exceedance point processes for stationary sequences under mild oscillation restrictions. *Lecture Notes in Statistics* 51, 69–80.
Lin, J.W. & 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, 1109–1136.
Nielsen, H.B. & A. Rahbek (2014) Unit root vector autoregression with volatility induced stationarity. *Journal of Empirical Finance* 29, 144–167.
Phillips, P.C.B., S. Shi, & 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., Y. Wu, & J. Yu (2011) Explosive behavior in the 1990's Nasdaq: When did exuberance escalate asset values? *International Economic Review* 52, 201–226.
Robert, C.Y., J. Segers, & C.A. Ferro (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.
Samorodnitsky, G. & M.S. Taqqu (1994) *Stable Non-Gaussian Random Processes*. Chapman & Hall.
Scargle, J.D. (1981) Phase-sensitive deconvolution to model random processes, with special reference to astronomical data. In D.F. Findley (ed.), *Applied Time Series Analysis II*, pp. 549–564. Academic Press.
Wiggins, R.A. (1978) Minimum entropy deconvolution. *Geoexploration* 16, 21–35.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

APPENDIX A: Proof of Proposition 2.1

Using the MA(∞) 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

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

If $\alpha \neq 1$, then,

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

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

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

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

APPENDIX B: Proof of Proposition 3.1

We decompose (X_t) into its pure causal AR(q) and noncausal AR(p) components (see Lanne and Saikkonen (2011) and Gouriéroux and 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), $u_t = \psi_1 u_{t+1} + \cdots + \psi_p u_{t+p} + \varepsilon_t$, for $k > p$, the conditional density of u_t given its

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

k past values is

$$f\left(u_t\middle|u_{t-1},\dots,u_{t-k}\right)=\frac{f\left(u_t,\dots,u_{t-k}\right)}{f\left(u_{t-1},\dots,u_{t-k}\right)}$$
$$=\frac{f\left(u_{t-k}\middle|u_{t-k+1},\dots,u_{t-k+p}\right)f\left(u_t,\dots,u_{t+1-k}\right)}{f\left(u_{t-k}\middle|u_{t-k+1},\dots,u_{t-k+p}\right)f\left(u_{t-1},\dots,u_{t+1-k}\right)}$$
$$=\frac{f\left(u_t,\dots,u_{t-p}\right)}{f\left(u_{t-1},\dots,u_{t-p}\right)}=f\left(u_t\middle|u_{t-1},\dots,u_{t-p}\right),$$

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 $x,x_1,\dots,x_{p+q}\in\mathbb{R}$,

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

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

APPENDIX C: Proof of Theorem 3.1

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

LEMMA C.1. *Let (X_t) be an α-stable pure noncausal AR(p) process solution of $X_t=\psi_1X_{t+1}+\cdots+\psi_pX_{t+p}+\varepsilon_t$, where the roots of $\psi(z)$ are outside the unit circle. Then,*

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

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

$$f_{X_t|(X_{t+1},\dots,X_{t+p})=(x_1,\dots,x_p)}(x_0)=f_{\varepsilon}(x_0-\psi_1x_1-\cdots-\psi_px_p),$$

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

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

Thus,

$$f_{X_{t+p}|(X_t,\dots,X_{t+p-1})=(x_0,\dots,x_{p-1})}(x_p) = \frac{f_\varepsilon(x_0-\psi_1x_1-\ldots-\psi_px_p)f_X(x_p)f_{X_{t+1},\dots,X_{t+p-1}|X_{t+p}=x_p}(x_1,\dots,x_{p-1})}{f_{X_t,\dots,X_{t+p-1}}(x_0,\dots,x_{p-1})}. \tag{C.1}$$

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

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

$$f_\varepsilon(x_0-\psi_1x_1-\cdots-\psi_px_p) \sim C^*(x_p)|x_p|^{-\alpha-1}, \tag{C.3}$$

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 ε_t are independent and $\psi_p \neq 0$, we have $|Z_t| \to +\infty$ a.s. as $|x_p| \to +\infty$. Therefore, for any $z_0 \in \mathbb{R}$ and any neighbourhood V_{z_0} of z_0, when $|x_p| \to +\infty$,

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

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,

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

$$\mathbb{E}\Big[|X_t|^\gamma \Big| X_{t-1}, \ldots, X_{t-p-q}\Big] \leq |\phi_1 X_{t-1} + \cdots + \phi_q X_{t-q}|^\gamma + \mathbb{E}\Big[|u_t|^\gamma \Big| u_{t-1}, \ldots, u_{t-p}\Big],$$

which completes the proof.

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

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

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

of X_{t+h}. We have by the independence between u_{t-1} and $(v_{t-2}, \ldots, v_{t-q-1})$

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

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,

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

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

APPENDIX E: Proof of Corollary 3.1

We will derive the asymptotic behavior of $\mathscr{P}_h(B)X_{t-1} = \left(P_h(B) + \psi^{<\alpha-1>}Q_h(\psi^{<\alpha-1>})\phi(B)\right)X_{t-1}$ when $1 \leq \alpha < 2$. We start by a result giving details about the behaviors 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*:

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

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

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

The proof is relegated to the online Supplement.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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) λ_1 and m_1 and we will denote also by C the coefficient associated to the monomial of highest degree of C_1. We thus have

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

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

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

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

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

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

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

Different cases arise.

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

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

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

iii) 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 $|\lambda||\psi|^{1-\alpha} = 1$,

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

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

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

APPENDIX F: Proof of Proposition 4.1

The $\rho(h)$ are only functions 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 $\rho(h)$ 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, Prop. 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.

APPENDIX G: Proof of Proposition 4.2

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

APPENDIX 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 $\frac{a_n^2}{\tilde{a}_n}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \xrightarrow{d} \mathbf{Z}$.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

APPENDIX 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).

APPENDIX J: Proof of Proposition 4.5

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

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

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

Thus, with

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

we can write

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

The two last expressions point to the fact that $\Big(\hat{\rho}_{\hat{\zeta}}(h)\Big)_{h=1,\ldots,H}$ will depend on $\Big(\hat{\boldsymbol{\rho}}(i)-\boldsymbol{\rho}(i)\Big)_{i=1,\ldots,H+p+q}$. 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 $0\leq h\leq p+q$ and $(2(p+q))\times(H+p+q)$ if $h\geq p+q+1$ picking the appropriate components of $\Big(\hat{\boldsymbol{\rho}}_{1:H+p+q}-\boldsymbol{\rho}_{1:H+p+q}\Big)$. More explicitly,

$$\boldsymbol{M}_{\boldsymbol{h}}=\Big(\boldsymbol{I}_{h+p+q}\quad\boldsymbol{0}_{h+p+q\times H-h}\Big),\qquad\text{if}\quad 0\leq h\leq p+q,$$
$$\boldsymbol{M}_{\boldsymbol{h}}=\Big(\boldsymbol{0}_{2(p+q)+1\times h-p-q-1}\quad\boldsymbol{I}_{2(p+q)+1}\quad\boldsymbol{0}_{2(p+q)+1\times H-h}\Big),\quad\text{if}\quad h\geq p+q+1.$$

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

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

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

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

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

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

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

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

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

$$\hat{\gamma}(h) = \sum_{t=1}^{n} X_t X_{t-h} = \sum_{t=1}^{n}\sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} d_i d_j \varepsilon_{t+i}\varepsilon_{t+j-h} = \sum_{t=1}^{n}\sum_{i\in\mathbb{Z}}\sum_{j\neq i} d_i d_{j+h}\varepsilon_{t+i}\varepsilon_{t+j} + \sum_{t=1}^{n}\sum_{i\in\mathbb{Z}} d_i d_{i+h}\varepsilon_{t+i}^2.$$

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 $a_n^{-2}\left(\hat{\gamma}(h) - \gamma(h)\hat{\gamma}_{\zeta}(0)\right) \xrightarrow{p} 0$.

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press

J.2. Proof of Lemma J.2

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

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 $a_n^{-2}\left(\hat{\gamma}(i),\quad 0\le i\le L\right)$ 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,

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

https://doi.org/10.1017/S0266466618000452 Published online by Cambridge University Press