Mixed Causal and Noncausal MAR(p, q) Processes and the modelling of explosive bubbles

Sébastien Fries* and Jean-Michel Zakoïan†

December 1, 2016

Preliminary version

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

†Jean-Michel Zakoïan, University Lille 3 and CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France. E-mail: zakoian@ensae.fr, Phone number: 33.1.41.17.77.25.

Mixed causal and noncausal MAR(p,q) processes and the modelling of explosive bubbles

Sebastien Fries and Jean-Michel Zakoian

Abstract

The adjunction of a noncausal component to standard causal linear autoregressive processes often yields a better fit to economic and financial time series. The general framework of mixed causal/noncausal MAR(p,q) processes with α-stable errors is investigated. The causal dynamics is derived and shown to display quadratic GARCH effects in direct time. The existence of a unit root in this causal dynamics is of particular interest as it allows to exhibit linear noncausal processes which are stationary, and even positive, martingales. Finally, under the broader assumption that the errors belong to the domain of attraction of a stable distribution, the property of OLS estimators are studied and illustrated on simulated data.

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

1 Introduction

The recent linear time series literature has focused on the adjunction of a noncausal component to classical causal autoregressive (AR) processes (see e.g. Andrews, Calder and Davis (2009), Lanne et al. (2012), Lanne, Saikkonen (2011), Davis, Song (2012), Chen et al. (2012), Hecq, Lieb and Telg (2015)). A mixed causal/noncausal AR(p,q) process (X_t), denoted MAR(p,q), is the strictly stationary solution of the equation

$$\psi(F)\phi(B)X_t = \varepsilon_t, \tag{1.1}$$

where B and F are the usual lag and forward operators, respectively, (ε_t) is an independent and identically distributed (iid) sequence, and $\psi(z) = 1 - \sum_{i=1}^{p} \psi_i z^i$ and $\phi(z) = 1 - \sum_{i=1}^{q} \phi_i z^i$ are polynomials of degrees p and q respectively (i.e. $\psi_p \neq 0$ and $\phi_q \neq 0$), with all roots outside the unit circle. A standard assumption is that ε_t is centered and has finite variance but, in order to work with fat tailed distributions, we do not make this assumption. When $p = 0$ (resp. $q = 0$), the model is called purely noncausal (resp. causal). Recently, Gouriéroux and Zakoian (2016) studied Model (1.1) in the case $p = 1$ and $q = 0$. In particular, they showed that this noncausal AR(1) model, combined with with fat tailed errors, is suitable to capture the so-called "bubble phenomenon", that is a local explosive trend followed by a sudden crash. The first aim of this paper is to provide probabilistic results regarding a natural extension of the noncausal AR(1) yielding richer dynamics. For instance, a limitation of the first-order noncausal AR is that the bubbles occur at a constant rate of explosion followed by an instant crash. It is thus natural to study how the noncausal AR(p) can accommodate more complex bubble patterns. Another interesting issue is whether a causal representation of the MAR process exists and can be exhibited. This property is well-known to hold in the L^2 framework, and has been established by Gouriéroux and Zakoian (2016) for the the noncausal AR(1) with α-stable errors. The derivation of such causal representations will be deduced through studying the existence of moments for the marginal and conditional distributions.

The second aim of this article is to provide statistical procedures for estimating the AR parameters in Model (1.1) and testing the validity of the specification, under appropriate assumptions on the tail behaviour of the error terms.

The paper is organised as follows. Section 2 considers Mixed AR(p,q) processes with stable errors for which we derive the marginal distribution and present examples of trajectories to illustrate the role of the main parameters of the model. Section 3 focuses on pure noncausal AR(p) processes. We establish the existence of conditional moments, obtain a closed expression for the point prediction in the symmetric α-stable and also in the asymmetric case under an additional assumption on the coefficients of the moving average representation, which yields a semi-strong linear causal representations of the process. When the errors are Cauchy distributed the conditional second order moment can be derived and a quadratic GARCH representation can be exhibited. Section 4 transposes the results obtained for the pure noncausal AR to Mixed AR by splitting the process into two components, one purely causal and one purely noncausal. We derive the new causal representation and notice that the noncausal component is transformed in a similar fashion as in the previous section while the causal component remains unchanged. Section 5 first examines the existence of a unit root in the linear causal representation derived for pure noncausal processes. A necessary and sufficient condition for the presence of such unit root is derived, which involves the tail parameter α and the AR structure. We also show that such unit root is necessarily simple. All the properties obtained in this section for pure noncausal AR are then immediately adapted for MAR processes by analysing their pure noncausal component. Section 6 is devoted to statistical inference on the model: two least-squares approaches are proposed to estimate the autoregressive parameters under the broader assumption that the errors belong to the domain of attraction of stable law. The first approach attempts to estimate the parameters of the expanded Model (1.1), which amounts to a classical linear regression of X_t on lags and leads of the process. It is shown that the estimators are generally not consistent in this case, except if the process is either purely causal or purely noncausal. The second approach, which is developed for the simpler MAR(1,1) case, attempts to directly estimate Model (1.1) which is nonlinear in the parameters. In this case, the parameters are consistent, but the causal and noncausal components are not identifiable. Both approaches are illustrated by a Monte-Carlo study. Proofs are collected in the Appendix.

2 Stable MAR(p, q) processes

In this section, we study the probabilistic properties of Model (1.1), where the errors ε_t follow a stable non-Gaussian distribution. The generality and convenience of this class of distributions is now well established.[1] Stable laws are generally defined through their characteristic function. In what follows, ε_t is said to follow a stable distribution with parameters $\alpha \in (0,2), \beta \in [-1,1], \sigma > 0,$

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

$\mu \in \mathbb{R}$ denoted $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, if

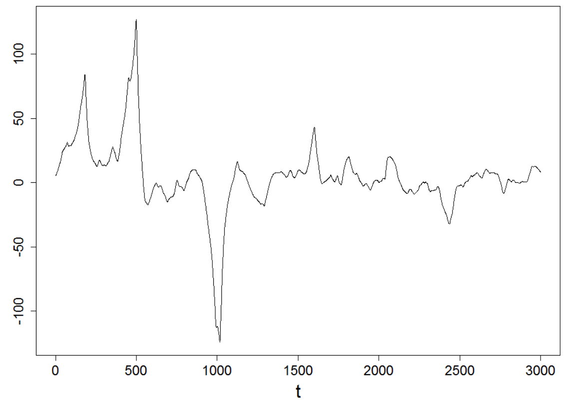
$$\forall s \in \mathbb{R}, \quad E(e^{is\varepsilon_t}) = \exp\Big\{ -\sigma^{\alpha}|s|^{\alpha}\,(1 - i\,\beta\,\mathrm{sign}(s)w(\alpha, s)) + is\mu\Big\},$$

where $w(\alpha, s) = \mathrm{tg}\left(\frac{\pi\alpha}{2}\right)$, if $\alpha \neq 1$, and $w(1, s) = -\frac{2}{\pi}\ln|s|$, otherwise. Recall that an α-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 $c_\alpha > 0$.
We denote $X_t = \sum_{k=-\infty}^{+\infty} d_k \varepsilon_{t+k}$ the MA(∞) representation of process (X_t). A simple index change highlights that (X_t) can be seen as a linear combination of baseline paths determined by the coefficients sequence (d_k) and weighted by the (random) sequence (ε_t):
$X_t = \sum_{\tau\in\mathbb{Z}} \varepsilon_\tau d_{\tau-t}$, and process (X_t) is now a combination of deterministic baseline paths, $Z_\tau(t) = d_{\tau-t}$, weighted by stochastic i.i.d coefficients ε_τ. Examples of trajectories for different noncausal MAR processes are presented on Figure 1.

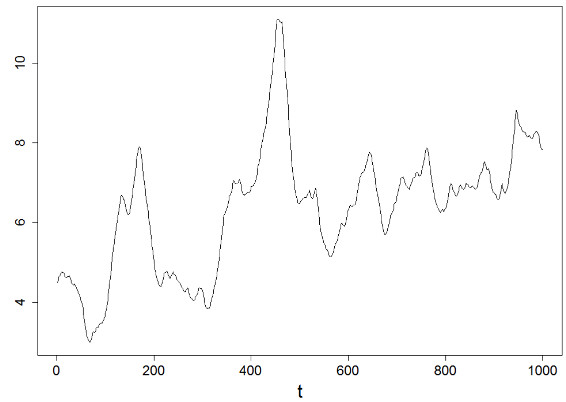
Various patterns can arise. For the particular four cases that are proposed on Figure 1, it can be seen that the trajectories feature locally explosive trends which are suited for the modelling of so-called bubbles and positive feedback loops phenomena. These bubbles can be either upward or downward trending. The tuning of the asymmetry parameter β allows to control for this feature: Figures 1b and 1c display trajectories with $\beta = 1$ and solely upward trending bubbles whereas Figure 1a and 1d with $\beta = 0$ display both types of trends. Also, the bubbles are vanishing at different paces: on Figures 1a, 1b and 1c, the return to central values seems to be as fast as the rate of increase of the bubbles, whereas the explosive trends on Figure 1d are followed by instant crashes. However, it can be also seen on the latter trajectory that instant bursts upwards can be followed by a plateau and finally collapse. This property is probably linked to the persistence of the process which has a root close to unit. Also, the trajectory of Figure 1c has been generated with a tail parameter of $\alpha = 0.98$ far smaller than that of the three other trajectories ($\alpha = 1.5$). The influence of the tail parameter is clearly visible as the trajectory of Figure 1c features several extreme values rendering bubbles and spikes. Last, even though the trajectory on Figure 1b appears positive on this sample, and indeed the probability of reaching the negative territory is very small, it can be shown that the support of its density is however the whole real line, whereas the trajectory of Figure 1c is positive almost surely.

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

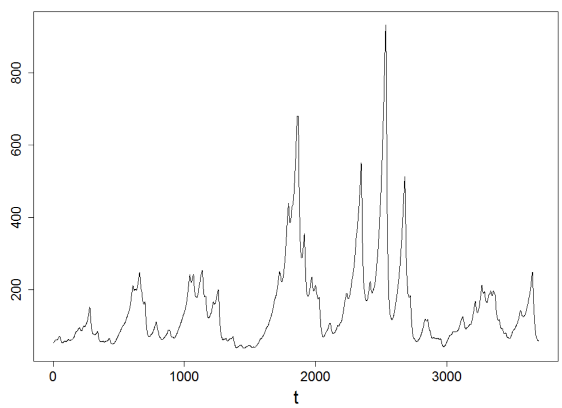
Proposition 2.1 *The strictly stationary MAR(p, q) process (X_t) solution of* (1.1) *with stable er-*



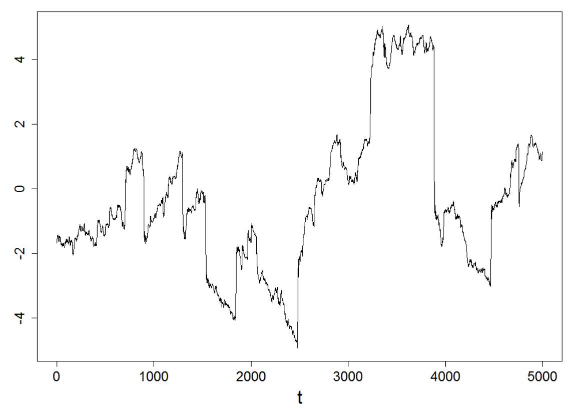
(a) $\alpha = 1.5$, $\beta = 0$, inverses of the causal roots: 0.95, 0.4, inverses of the noncausal roots: 0.98, 0.6



(b) $\alpha = 1.5$, $\beta = 1$, inverses of the causal roots: 0.94, 0.3, inverses of the noncausal roots: 0.98, -0.4



(c) $\alpha = 0.98$, $\beta = 1$, inverses of the causal roots: 0.9, 0.6, inverses of the noncausal roots: 0.98, 0.5



(d) $\alpha = 1.5$, $\beta = 0$, inverses of the causal roots: 0.7, -0.9, inverses of the noncausal roots: 0.9985, 0.2

Figure 1: Examples of trajectories of MAR(2,2) processes with different parameters.

rors, $\varepsilon_t \sim \mathcal{S}(\alpha, \beta, \sigma, \mu)$, has a stable stationary distribution given by $X_t \sim \mathcal{S}(\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}, \tilde{\mu})$ with and

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

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

It is worth noting that the tail index α of X_t is that of the error term. In particular, $E|X_t|^s < \infty$ for $s < \alpha$ and $E|X_t|^\alpha = \infty$. In the following example, we provide an example of a more explicit characterization of the stationary distribution.

Example 2.1 (Noncausal AR(2) process with Cauchy errors) *Consider* $X_t = \psi_1 X_{t+1} + \psi_2 X_{t+2} + \varepsilon_t$ *with* $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(1,0,\sigma,0)$. *Denote by* λ_1 *and* λ_2 *the inverse of the roots of the characteristic polynomial* $1-\psi_1 z - \psi_2 z^2$, *which verify the relations* $\psi_1 = \lambda_1 + \lambda_2$ *and* $\psi_2 = -\lambda_1\lambda_2$. *It can be shown that the coefficients of the MA(∞) representation read for any* $k \geq 0$, $d_k = \dfrac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2}$ *in the case* $\lambda_1 \neq \lambda_2$ *and* $d_k = (1+k)\lambda_1^k$ *if* $\lambda_1 = \lambda_2$. *Without loss of generality, assume* $0 \leq |\lambda_2| \leq |\lambda_1| < 1$. *Then* $X_t \sim \mathcal{S}\left(1, 0, \dfrac{\sigma}{1 - sign(\lambda_1)\psi_1 - \psi_2}, 0\right)$. *Note that the denominator never cancels out under the assumption that the roots are outside the unit circle.*

Example 2.2 (MAR(1,1) process) *Let* (X_t) *be the strictly stationary solution of* $(1-\lambda_1 F)(1-\lambda_2 B)X_t = \varepsilon_t$ *with* $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha,\beta,\sigma,\mu)$. *It can be shown that the coefficients of the MA(∞) representation read* $d_k = \dfrac{\lambda_1^k}{1-\lambda_1\lambda_2}$, *for any* $k \geq 0$, *and* $d_k = \dfrac{\lambda_2^{-k}}{1-\lambda_1\lambda_2}$, *for any* $k \leq 0$. *Then in the general case* $X_t \sim \mathcal{S}\left(\alpha, \tilde{\beta}, \tilde{\sigma}, \tilde{\mu}\right)$, *with*

$$
\begin{aligned}
\tilde{\beta} &= \beta \frac{1 - sign(\lambda_1\lambda_2)|\lambda_1\lambda_2|^\alpha}{1 - |\lambda_1\lambda_2|^\alpha} \frac{1 - sign(\lambda_1)|\lambda_1|^\alpha}{1 - |\lambda_1|^\alpha} \frac{1 - sign(\lambda_2)|\lambda_2|^\alpha}{1 - |\lambda_2|^\alpha}, \\
\tilde{\sigma} &= \frac{\sigma}{1-\lambda_1\lambda_2} \left(\frac{1 - |\lambda_1\lambda_2|^\alpha}{(1-|\lambda_1|^\alpha)(1-|\lambda_2|^\alpha)} \right)^{\frac{1}{\alpha}}, \\
\tilde{\mu} &= \frac{\mu}{(1-\lambda_1)(1-\lambda_2)} - \mathbb{1}_{\{\alpha=1\}} \frac{2\beta\sigma}{\pi(1-\lambda_1\lambda_2)} \left[\frac{\lambda_1 \ln|\lambda_1|}{(1-\lambda)^2} + \frac{\lambda_2 \ln|\lambda_2|}{(1-\lambda_2)^2} - \frac{(1-\lambda_1\lambda_2)\ln|1-\lambda_1\lambda_2|}{(1-\lambda_1)(1-\lambda_2)} \right].
\end{aligned}
$$

In particular, when $\lambda_1 > 0$ *and* $\lambda_2 > 0$ *and the errors are Cauchy distributed, that is* $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(1,0,\sigma,0)$, *then the above formulae simplify and* $X_t \sim \mathcal{S}\left(1, 0, \dfrac{\sigma}{(1-\lambda_1)(1-\lambda_2)}, 0\right)$.

We now turn to the dependence properties of process (X_t), starting with the pure noncausal case.

3 Pure Noncausal AR(p)

In this section, we assume that (X_t) is purely noncausal, namely that $q = 0$ and $p > 0$. We will show later on that all the results extend to the general MAR case. In the pure noncausal case, process (X_t) is the strictly stationary solution of the autoregressive equation

$$X_t = \psi_1 X_{t+1} + \ldots + \psi_p X_{t+p} + \varepsilon_t.$$

The Markov property still holds whatever the errors distribution.

Proposition 3.1 *The noncausal AR(p) process* (X_t) *is an homogeneous Markov chain of order p.*

The existence of moments for the causal conditional distribution of the purely noncausal process is established in the following Theorem, generalising the result obtained for $p = 1$ by Gouriéroux and Zakoïan (2016).

Theorem 3.1 *Let (X_t) be a pure noncausal AR(p) process with $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu)$. Then,*

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

Moreover, if either $|\beta| < 1$ or $\alpha \geq 1$, we have $\mathbb{E}\left[|X_{t+h}|^{\gamma}|\, X_{t-1}, \ldots, X_{t-p}\right] = \infty$, for $\gamma \geq 2\alpha + 1$.

The conditional distribution with respect to past observations has lighter tails than the marginal and the conditional distributions with respect to future values, which only admit moments up to order α. It is interesting to note that the condition for moments existence is independent of the AR order p. An appealing feature of the class of stable distributions is that it contains purely asymmetric distributions, either completely skewed to positive or negative values. They can thus be used to model processes like prices in finance, river flows in hydrology (Camacho, Mcleod and Hipel (1987), Mohammadi, Eslami and Kahawita (2006)) or light curves in astronomy (see Vaughan (2013), McLeod, Ch L., et al. (2010)). Dealing with positive processes using linear time series usually implied to work on log-transformed data, forcing to assume a multiplicative dependence structure of the initial process instead of an additive one, impairing interpretation. This could be avoided in some cases using linear time series with purely asymmetric stable processes, which, even though they have very heavy tails, can feature conditional moments at any order.

Proposition 3.2 *If*

$$|\beta| = 1, \quad \alpha < 1, \quad d_k \geq 0 \quad \text{for all } k \geq 0, \quad \text{and} \quad \beta\psi_p P_i(0) \geq 0 \quad \text{for all } i \in \{0, \ldots, h-1\},$$

then,

$$\mathbb{E}\left[|X_{t+h}|^{\gamma}|\, X_{t-1}, \ldots, X_{t-p}\right] < \infty, \quad \text{a.s., for all } \gamma > 0$$

Now we show that, in the symmetric case, the first-order conditional moment can be explicitly derived at any horizon.

Proposition 3.3 *When $\beta = 0$, we have for any $h \geq 0$,*

$$\mathbb{E}\left[X_{t+h}|\, X_{t-1}, \ldots, X_{t-p}\right] = \tilde{\mu} + \frac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha}_{\epsilon,h} + \tilde{\sigma}^{\alpha}} \left\{P_h(B)X_{t-1} - (\tilde{\mu} + \mu_{\epsilon,h})\right\},$$

where $\mu_{\epsilon,h}$ *and* $\sigma_{\epsilon,h}$ *are, respectively, the drift and scale parameters of the stable distribution of* $Q_h(B)\varepsilon_t$*. In particular, if* $\mu = 0$ *we have*

$$\mathbb{E}\left[X_{t+h} \middle| X_{t-1}, \ldots, X_{t-p} \right] = \frac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha}_{\epsilon,h} + \tilde{\sigma}^{\alpha}} P_h(B) X_{t-1}.$$

The asymptotic behaviour of the conditional expectation -when the horizon tends to infinity- is highly dependent on the tail index α. Proposition 3.3 allows us to distinguish three types of behaviours summarized in the following Corollary.

Corollary 3.1 *Under the conditions of Proposition 3.3 (*$\mu = 0$*), we have almost surely*

$$\left| \mathbb{E}\left[X_{t+h} \middle| X_{t-1}, \ldots, X_{t-p} \right] \right| \underset{h \to +\infty}{\longrightarrow} \begin{cases} 0 & if \quad \alpha \in (1,2), \\ \ell_{t-1} & if \quad \alpha = 1, \quad \textit{for some random variable } \ell_{t-1} \in (0, +\infty), \\ +\infty & if \quad \alpha \in (0,1). \end{cases}$$

Actually, we even have an equivalent of the conditional expectation as h tends towards infinity. Denote λ the root of ψ of smallest modulus and m its associated multiplicity order. Then for $\alpha \in (0,2)$, $\beta = 0$ (and $\mu = 0$),

$$\mathbb{E}\left[X_{t+h} \middle| X_{t-1}, \ldots, X_{t-p} \right] \underset{h \to +\infty}{\sim} \ell_{t-1} \left(\frac{\lambda}{|\lambda|} \right)^{h} \left(h^{m-1} |\lambda|^{-h} \right)^{1-\alpha}, \quad \text{a.s.} \quad ,$$

where ℓ_{t-1} is a constant (with respect to h) different from 0 almost surely depending on the p past observations. If $\alpha \in (1,2)$, that is for lighter tails, the conditional expectation always tends to 0, that is, to the unconditional expectation. This is consistent with the L^2 framework (Brockwell, Davis (1991), p.189). On the contrary, if $\alpha \in (0,1)$, hence for heavier tails, the conditional expectation tends to $+\infty$ in modulus. Last, $\alpha = 1$ is a tipping point: the term driving the shrinkage or the expansion disappears and the absolute value of the conditional expectation tends to a finite limit. Note that in the two latter cases, the unconditional expectation does not exist. Assume for simplicity that $\mu = 0$. When the errors follow a symmetric α-stable distribution, Proposition 3.3 at horizon $h = 0$ yields a semi-strong causal representation of X_t.

Corollary 3.2 *Under the assumption of Proposition 3.3 with* $\mu = 0$*, the noncausal process* X_t *satisfies*

$$\mathscr{P}_{\alpha}(B) X_t := X_t - \frac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha} |\psi_p|^{-\alpha} + \tilde{\sigma}^{\alpha}} P_0(B) X_{t-1} = \eta_t, \tag{3.1}$$

where $P_0(B) = \psi_p^{-1}(-\psi_{p-1} - \ldots - \psi_1 B^{p-1} + B^p)$*,* (η_t) *is a sequence satisfying* $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1} \right] = 0$ *with* $\mathscr{F}_t := \sigma\left(X_t, X_{t-1}, \ldots \right)$ *being the natural filtration of process* (X_t)*. Furthermore, this representation* *still holds when $\beta \in [-1,1]$ if $d_k \geq 0$ for all k, and in particular if all the roots of the AR polynomial are either positive or complex, up to a nonzero drift term when $\alpha = 1$.*

Example 3.1 (Cauchy AR(2) process (continued)) *Applying Proposition 3.3 at horizon $h = 0$, where $\epsilon_{t-2,0} := -\psi_2^{-1}\epsilon_{t-2} \sim \mathcal{S}\left(1, 0, \sigma_{\epsilon,0} := \sigma|\psi_2|^{-1}, 0\right)$ and $\psi_2 \neq 0$, yields:*

$$\mathbb{E}\left[X_t|\, \mathscr{F}_{t-1}\right] = \frac{\psi_2^{-1}\left(-\psi_1 X_{t-1} + X_{t-2}\right)}{1 + |\psi_2|^{-1}(1 - sign(\lambda_1)\psi_1 - \psi_2)}.$$

This expression implies the existence of a semi-strong causal representation of process (X_t) as:

$$X_t = \frac{\psi_2^{-1}\left(-\psi_1 X_{t-1} + X_{t-2}\right)}{1 + |\psi_2|^{-1}(1 - sign(\lambda_1)\psi_1 - \psi_2)} + \eta_t.$$

We can in particular notice that under the assumption $\lambda_1 > 0$ and $\psi_2 > 0$ (i.e. $\lambda_1 > 0$, $\lambda_2 < 0$ and $|\lambda_1| > |\lambda_2|$), we have:

$$(1 - B)\left(1 + \frac{1}{1-\psi_1}B\right) X_t = \eta_t,$$

which yields a unit root in the causal representation.

By comparison with AR processes with finite variance, the result of Corollary 3.2 is surprising. Indeed in the L^2 framework, if (X_t) is noncausal satisfying $\psi(F)X_t = \varepsilon_t$, then there exists a causal version of (X_t) given by $\psi(B)X_t = Z_t$, where (Z_t) is uncorrelated with mean 0 and finite variance. By contrast, in our stable framework, the lag polynomial $\mathscr{P}$ in (3.1) is very different from ψ, even if both are of order p.

We now turn to the second-order conditional moment at horizon h which, by Theorem 3.1, exists when $\alpha = 1$.

Proposition 3.4 *In the Cauchy case ($\alpha = 1$, $\beta = 0$), we have for any $h \geq 0$,*

$$\mathbb{E}\left[X_{t+h}^2\middle|\, \mathscr{F}_{t-1}\right] = \frac{1}{\sigma_{\epsilon,h} + \tilde{\sigma}}\left[\tilde{\sigma}\left(P_h(B)X_{t-1} - \mu_{\epsilon,h}\right)^2 + \sigma_{\epsilon,h}\tilde{\mu}^2\right] + \sigma_{\epsilon,h}\tilde{\sigma}.$$

When $\mu = 0$, it can be shown that

$$\mathbb{V}\left(X_{t+h}\,|\mathscr{F}_{t-1}\right) = \frac{\sigma_{\varepsilon,h}}{\tilde{\sigma}}\left(\mathbb{E}\left[X_{t+h}|\, X_{t-1}, \ldots\right]\right)^2 + \sigma_{\varepsilon,h}\tilde{\sigma}.$$

Thus the causal representation (3.1) can be refined and reveals quadratic GARCH effects in the noncausal Cauchy AR(p) process.

$$X_t = \frac{\tilde{\sigma}}{\sigma_{\varepsilon,0} + \tilde{\sigma}} P_0(B)X_{t-1} + \sigma_t\eta_t,$$
$$\sigma_t^2 = \frac{\tilde{\sigma}}{\sigma_{\varepsilon,0}\left(\tilde{\sigma} + \sigma_{\varepsilon,0}\right)^2}\left(P_0(B)X_{t-1}\right)^2 + \sigma_{\varepsilon,0}\tilde{\sigma},$$

where $\mathbb{E}\left[\eta_t \middle| \mathscr{F}_{t-1}\right] = 0$, $\mathbb{E}\left[\eta_t^2 \middle| \mathscr{F}_{t-1}\right] = 1$. The process is however not a GARCH in the strict sense: first because the "errors" are not i.i.d., and second because the volatility is a function of the past observed variables X_t (not of ε_t). Alternatively, this representation is also very close to the Double Autoregressive model proposed by Ling (2007), except, again, that the second order moment contains cross-product terms of past observations in our case.

4 MAR(p, q) processes

In this section we transpose the main results derived for pure noncausal processes to MAR processes, using the decomposition of Corollary A.1. Let (X_t) be a MAR(p, q) process with $p > 0$ and $q > 0$ satisfying

$$\psi(F)\phi(B)X_t = \varepsilon_t.$$

We begin by stating the extensions of the Propositions regarding the Markov property and the existence of conditional moments:

Proposition 4.1 *The MAR(p, q) process (X_t) is an homogeneous Markov chain of order $p + q$.*

Proposition 4.2 *Let (X_t) be a MAR(p, q) process with $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu)$.*

$$\textit{If} \quad -1 < \gamma < 2\alpha + 1, \quad \textit{then}, \quad \mathbb{E}\left[|X_{t+h}|^{\gamma}|\, X_{t-1}, \ldots, X_{t-p-q}\right] < +\infty, \quad \textit{for any} \quad h \geq 0, \quad \textit{a.s.}.$$

Next, we can derive from the pure noncausal case a semi-strong linear causal representation of the MAR(p, q) process:

Corollary 4.1 *When $\beta = 0$, the result of Corollary 3.2 still holds for the MAR(p, q) process (X_t) with $\mathscr{P}\alpha(B)$ is replaced by $\mathscr{P}_\alpha(B)\phi(B)$.*

Interestingly, it appears that the causal component of (X_t) remains basically unchanged in this representation whereas the noncausal component is distorted exactly as in the pure noncausal case.

Example 4.1 (MAR(1,1) process (continued)) Let (X_t) be the strictly stationary solution of

$$(1 - \lambda_1 F)(1 - \lambda_2 B)X_t = \varepsilon_t,$$

with $\varepsilon \overset{iid}{\sim} \mathcal{S}(\alpha, 0, \sigma, 0)$. Let also (u_t) and (v_t) be the pure noncausal and causal AR(1) components respectively (see Lanne and Saikkonen(2011) and Gouriéroux and Jasiak (2014)):

$$u_t = (1 - \lambda_2 B)X_t \iff (1 - \lambda_1 F)u_t = \varepsilon_t,$$
$$v_t = (1 - \lambda_1 F)X_t \iff (1 - \lambda_2 B)v_t = \varepsilon_t.$$

Then

$$X_t = \left(\frac{1}{\lambda_1} + \lambda_2\right) X_{t-1} - \frac{\lambda_2}{\lambda_1} X_{t-2} - \frac{\varepsilon_{t-1}}{\lambda_1},$$

and thus

$$X_t = \lambda_2 X_{t-1} + u_t.$$

Taking the expectation conditional on the past of (X_t), using the equivalence between the information sets (X_{t-1}, X_{t-2}) and (u_{t-1}, v_{t-2}) and invoking the independence between u_{t-1} and v_{t-2}, we obtain

$$\begin{aligned}\mathbb{E}\left[X_t | X_{t-1}, X_{t-2}\right] &= \lambda_2 X_{t-1} + \mathbb{E}\left[u_t | u_{t-1}\right],\\ &= \lambda_2 X_{t-1} + \text{sign}(\lambda_1)|\lambda_1|^{\alpha-1}(X_{t-1} - \lambda_2 X_{t-2}),\\ &= \left(\lambda_2 + \text{sign}(\lambda_1)|\lambda_1|^{\alpha-1}\right) X_{t-1} - \lambda_2 \text{sign}(\lambda_1)|\lambda_1|^{\alpha-1} X_{t-2}.\end{aligned}$$

This yields a causal representation as formulated in Corollary 4.1:

$$(1 - \text{sign}(\lambda_1)|\lambda_1|^{\alpha-1}B)(1 - \lambda_2 B)X_t = \eta_t,$$

with $\mathbb{E}\left[\eta_t | X_{t-1}, X_{t-2}\right] = 0$. In particular if $\lambda_1 > 0$ and $\alpha = 1$, the causal representation of (X_t) reads:

$$(1 - B)(1 - \lambda_2 B)X_t = \eta_t.$$

5 Unit roots and Martingales

In the case of noncausal Cauchy AR(1) with a positive AR coefficient, the causal semi-strong representation displays a unit root (see Gouriéroux and Zakoïan (2016)). A question of interest is to determine whether, or under which condition, this property can be extended in the AR(p) case. Note that stationary processes displaying a unit root, that is, such that $\mathbb{E}\left[X_t | \mathscr{F}_{t-1}\right] = X_{t-1}$, have been found in other contexts (see the DAR process of Ling (2007) and its multivariate extension by Nielsen and Rahbek (2014)). Such unit root indicate that a linear combination of present and past values of the process may behave like a martingale. Indeed, assuming that $\mathscr{P}_\alpha(1) = 0$ allows us to write $(1 - B)\Upsilon(B)X_t = \eta_t$, for some polynomial Υ of degree $p - 1$, and it appears clearly that $Z_t := \Upsilon(B)X_t$ is a martingale with respect to the filtration $\mathscr{F}_t = \sigma(X_t, X_{t-1}\ldots)$. Such processes, which are interesting *per se* can also turn out to be useful in the economic and financial fields as stationary solutions of rational expectations models. We will exhibit in particular noncausal processes which are positive stationary martingales and hence suited to describe positive quantities like prices. As illustrated in Example 3.1, the semi-strong causal representation of the noncausal process X_t may admit a unit root when $\alpha = 1$. In fact, when $p > 1$, the causal representation can feature unit roots for other values of $\alpha \in]0, 2[$ as well. We provide a necessary and sufficient condition for the existence of such a unit root for a given tail parameter and also show that if it exists, it is necessarily simple.

Proposition 5.1 *Assume the conditions for the causal representation* (3.1) *are verified. Then, the following hold*

ι) $\mathscr{P}_\alpha(1) = 0$ *if and only if* $|\psi_p|^\alpha \sum_{k=0}^{+\infty} |d_k|^\alpha = \psi_p \sum_{k=0}^{+\infty} d_k$.

$\iota\iota$) $\mathscr{P}_1(1) = 0$ *if and only if* $\psi_p > 0$ *and* $d_k \geq 0$ *for all* $k \geq 0$.

$\iota\iota\iota$) *For any given values of* $\psi_1, \dots, \psi_p$, *there exist up to two values of* $\alpha \in]0, 2[$ *such that* $\mathscr{P}_\alpha(1) = 0$.

$\iota\nu$) *If* $\mathscr{P}_\alpha(1) = 0$, *then* 1 *is a simple root of* $\mathscr{P}_\alpha$.

Remark 5.1 *Since* $-\psi_p = \lambda_1 \dots \lambda_p$, *it can be noticed that the condition* $\psi_p > 0$ *is equivalent to imposing an odd number of positive real roots.*

Finally, thanks to the causal representation derived at Proposition 4.1, all the previous results regarding the existence of a unit root in the causal representation of purely noncausal processes can be straightforwardly transposed to MAR(p, q) simply by analysing its noncausal latent component $u_t = \phi(B)X_t$.

Proposition 5.2 *The causal representation of the MAR(p, q) process admits a unit root if and only if the causal representation of its noncausal AR(p) underlying component admits a unit root.*

Therefore, the causal component of the MAR process is neutral with respect to the existence of unit root in the causal representation.

Example 5.1 (Positive stationary martingale) Consider the linear process defined by

$$\psi(F)Z_t = \Upsilon(B)\varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha, 1, \sigma, 0)$$

where $\psi(z) := 1 - \psi_1 z - \psi_2 z^2 - \psi_3 z^3 := (1 - \lambda_1 z)(1 - \lambda_2 z)(1 - \lambda_3 z)$, with $0 < \lambda_1 < \lambda_2 < \lambda_3 < 1$ and $\Upsilon(B) = 1 - \frac{1-\psi_1}{1-\psi_1-\psi_2} B + \frac{1}{1-\psi_1-\psi_2} B^2$. For a proper parametrisation, process (Z_t) can be a positive stationary martingale, as stated in the next result.

Proposition 5.3 *For any* λ_1, λ_2 *and* λ_3 *sufficiently close, but strictly smaller than unity, there exists a tail parameter* $\alpha_0 \in]0, 1[$ *such that process* (Z_t) *is a positive stationary martingale.*

Figures 2 present trajectories of (X_t) and (Z_t) for a particular choice of (inverses of the) roots: $\lambda_1 = 0{,}6475$, $\lambda_2 = 0{,}7$, $\lambda_3 = 0{,}86$ and the appropriate α_0 which has been numerically approached so that the necessary and sufficient condition of Proposition 5.1.ι is satisfied[2].

[2] With $\lambda_1 = 0{,}6475$, $\lambda_2 = 0{,}7$, $\lambda_3 = 0{,}86$ and $\alpha_0 = 0{,}94919454766929$, the equality of Proposition 5.1.ι, namely $\sum_{l=0}^{+\infty} |\psi_p d_l|^\alpha = \sum_{l=0}^{+\infty} \psi_p d_l$, is satisfied at a precision of $\pm 10^{-14}$. The infinite sum on the left-hand side has been truncated at the smallest integer k_0 such that for any $k \geq k_0$, $d_k = 0$ at machine precision. $k_0 = 4940$ and $d_{4939} \lesssim 1{,}5 \cdot 10^{-322}$.

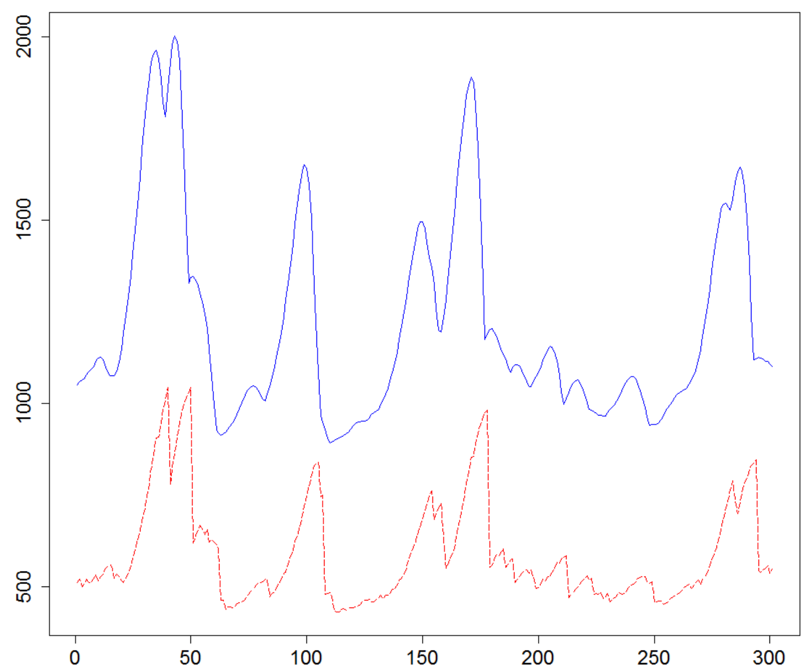


Figure 2: Example of trajectories of processes (X_t) (blue solid line) and (Z_t) (red dashed line) defined by the equations $(1 - 0.6475F)(1 - 0.7F)(1 - 0.86F)X_t = \varepsilon_t$ and $\Upsilon(B)Z_t = X_t$, where $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha_0, 1, 1, 0)$ (see footnote[2] for the choice of α_0).

Figures 3 and 4 illustrate the positivity and martingale properties. 500 trajectories of process (Z_t) with 100 000 observations have been generated and a nonparametric kernel regression (with a bandwidth of 50) of Z_t on Z_{t-1} has been performed in the range $[430, 3000]$ where on average more than 97% of the observations are located. Since (Z_t) is bounded below at 0, the kernel estimation is difficult for low values (roughly 1,5% of observations in the range $[0, 430]$). The average of the kernel estimates has been computed alongside 1,97 times standard deviation bands across the 500 simulations. It can be noticed on figure 3 that the first bisector is either inside of, or not distinguishable from the standard deviations bands, except for very low values where an edge effect is still slightly noticeable. Figure 4 displays similar results for 250 simulations of 1 000 000 observations of (Z_t) but focuses on the very large values (up to 10^6). In particular, a single bandwidth for the whole range was not appropriate and the kernel estimation was performed on consecutive segments with increasing bandwidth chosen to yield a smooth estimation while limiting the edge effect on low values.

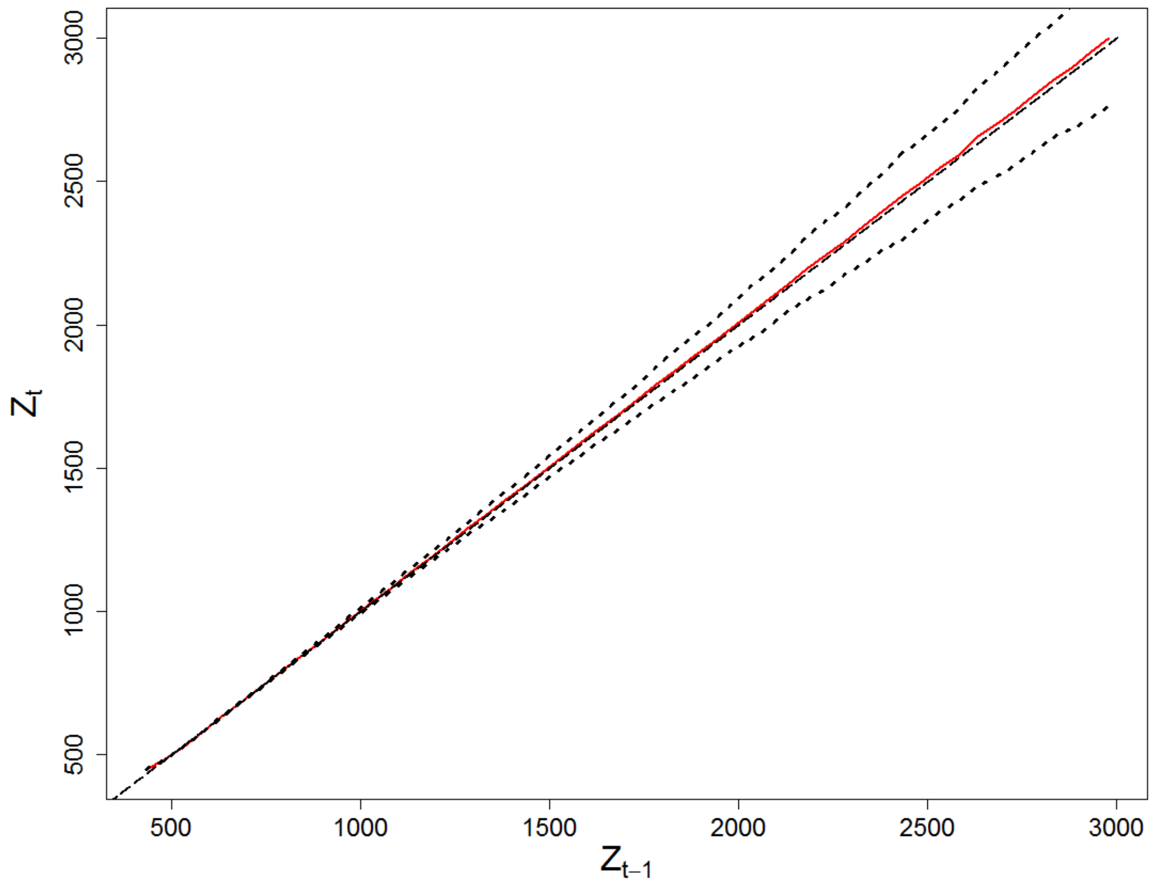


Figure 3: Nonparametric kernel regression of Z_t on Z_{t-1} where $Z_t = \Upsilon(B)X_t$, $(1-0.6475F)(1-0.7F)(1-0.86F)X_t = \varepsilon_t$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha_0, 1, 1, 0)$. The roots have been chosen so that the coefficients of the moving average of Z_t are all positive, and the tail parameter α_0 has been numerically approached so that Proposition 5.1.ι holds (see footnote [2]). The straight black line indicates the first bisector, the red line is the average of kernel estimates. 500 simulations of process (Z_t) have been performed with 100 000 observations. The two black curves indicate 1.97 times the sample standard deviation of kernel estimates at each point. Note that on average more than 97,1% of observations are in the range $[430, 3000]$ chosen here. Since Z_t is bounded below, it is difficult to estimate the kernel for low values because of an edge effect. Bandwidth = 50.

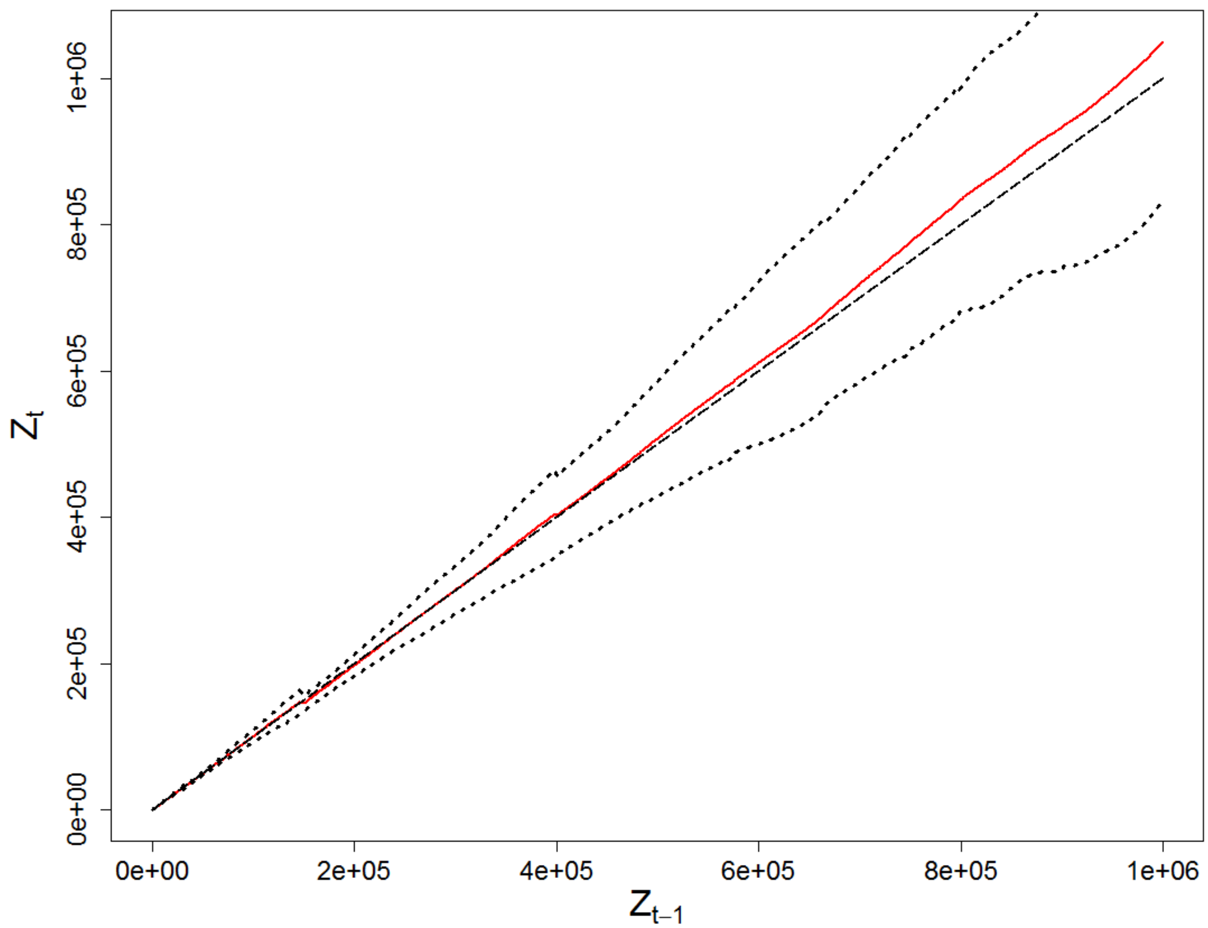


Figure 4: Nonparametric kernel regression of Z_t on Z_{t-1} where $Z_t = \Upsilon(B)X_t$, $(1-0.6475F)(1-0.7F)(1-0.86F)X_t = \varepsilon_t$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha_0, 1, 1, 0)$. The roots have been chosen so that the coefficients of the moving average of Z_t are all positive, and the tail parameter α_0 has been numerically approached so that Proposition 5.1.ι holds (see footnote [2]). The straight black line indicates the first bisector, the red line is the kernel estimate. 250 simulations of process (Z_t) have been performed with 1 000 000 observations. The two black curves indicate 1.97 times the sample standard deviation at each point. Note that on average, around 1,1% of observations are above the value 3000.

6 Statistical Inference

In the what follow, we will relax the assumption that (ε_t) is an iid sequence of α-stable noise but rather assume that it belongs to the domain of attraction of a stable distribution. More specifically we assume that (ε_t) is an iid sequence of random variables such that:

$$\mathbb{P}(|\varepsilon_0| > x) = x^{-\alpha} L(x), \tag{6.1}$$

where L is a slowly varying function at infinity and $0 < \alpha < 2$ and

$$\frac{\mathbb{P}(\varepsilon_t > x)}{\mathbb{P}(|\varepsilon_t| > x)} \to p, \quad \frac{\mathbb{P}(\varepsilon_t \leq x)}{\mathbb{P}(|\varepsilon_t| > x)} \to 1 - p, \tag{6.2}$$

as $x \to +\infty$ and $p \in [0, 1]$. Define also $a_n := \inf\{x : \mathbb{P}(|\varepsilon_0| > x) \leq n^{-1}\}$.

6.1 Linear Least Squares Estimation of a MAR(p, q) process

The idea of the Linear Least Square estimator is to consider a MAR(p, q) process written after expanding the product $\psi(F)\phi(B)$ and isolating the term in X_t (assuming the associated coefficient is nonzero)

$$X_t = \sum_{j=-q, j\neq 0}^{p} \theta_{0j} X_{t+j} + \epsilon_t, \tag{6.3}$$

and to perform a classical regression by minimising the squared error:

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{t=0}^{n} \Big(X_t - \theta_{-q} X_{t-q} - \ldots - \theta_{-1} X_{t-1} - \theta_1 X_{t+1} - \ldots - \theta_p X_{t+p} \Big)^2,$$

with $n + p + q + 1$ observations $X_{-q}, \ldots, X_{n+p}$ of process (6.3) and $\boldsymbol{\theta} := {}^{\mathrm{t}}(\theta_{-q}, \ldots, \theta_{-1}, \theta_1, \ldots, \theta_p)$. The least squares estimator $\hat{\boldsymbol{\theta}} := {}^{\mathrm{t}}(\hat{\theta}_{-q}, \ldots, \hat{\theta}_{-1}, \hat{\theta}_1, \ldots, \hat{\theta}_p)$ verifies:

$$\hat{\boldsymbol{\theta}} = ({}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{X})^{-1}\, {}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{Y},$$

with $\boldsymbol{X} = (X_{t+j})_{\substack{0\leq t\leq n \\ -q\leq j\leq p \\ j\neq 0}}$ and ${}^{\mathrm{t}}\boldsymbol{Y} = (X_0, \ldots, X_n)$. Let us also introduce

$$\forall h \in \mathbb{Z}, \quad \rho(h) := \frac{\sum_{k\in\mathbb{Z}} d_k d_{k+h}}{\sum_{k\in\mathbb{Z}} d_k^2}, \qquad \forall h \in \mathbb{Z}, \forall n \in \mathbb{N}^*, \quad \hat{\rho}(h) := \frac{\sum_{t=0}^{n} X_t X_{t+h}}{\sum_{t=0}^{n} X_t^2}.$$

and $\boldsymbol{R} := \Big(\rho(|i-j|)\Big)_{\left\{\substack{-q\leq i,j\leq p \\ i,j\neq 0}\right\}}$ and $\boldsymbol{\rho} := {}^{\mathrm{t}}(\rho(q), \ldots, \rho(1), \rho(1), \ldots, \rho(p))$, where (d_k) are the coefficients of the moving average representation of process (X_t). Note that, even if the theoretical autocorrelations of (X_t) do not exist, the function $\rho(\cdot)$ is well defined.

Proposition 6.1 *Under the above assumptions, we have*

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{R}^{-1}\boldsymbol{\rho} \tag{6.4}$$

Remark 6.1 *Given that* $\boldsymbol{\rho} := {}^{t}\Big(\rho(q), \ldots, \rho(1), \rho(1), \ldots, \rho(p)\Big)$ *and since* $\boldsymbol{R}^{-1}$ *is symmetric, it is remarkable that* $\hat{\theta}_{-1} = \hat{\theta}_1, \ldots, \hat{\theta}_{-\min(p,q)} = \hat{\theta}_{\min(p,q)}$*. Obviously, the OLS estimator does not converge in general towards the true hyperparameter* $\boldsymbol{\theta}$*.*

Example 6.1 Considering the case $p = q = 1$, we have

$$\boldsymbol{\rho} = \begin{pmatrix} \rho(1) \\ \rho(1) \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 1 & \rho(2) \\ \rho(2) & 1 \end{pmatrix} \quad \text{and therefore} \quad \boldsymbol{R}^{-1} = \frac{1}{1-\rho(2)^2} \begin{pmatrix} 1 & -\rho(2) \\ -\rho(2) & 1 \end{pmatrix}.$$

Hence, after direct computation, we get that:

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{R}^{-1}\boldsymbol{\rho} = \begin{pmatrix} \dfrac{\rho(1)}{1+\rho(2)} \\ \dfrac{\rho(1)}{1+\rho(2)} \end{pmatrix}. \tag{6.5}$$

Besides, in this case, the coefficients $\rho(1)$ and $\rho(2)$ can be easily computed. Assume the MAR(1,1) process we are studying here verifies $(1 - \psi_0 F)(1 - \phi_0 B)X_t = \varepsilon_t$. Then, it can be shown that

$$\begin{aligned}
\gamma(0) := \sum_{k\in\mathbb{Z}} d_k^2 \qquad &= \frac{1+\psi_0\phi_0}{(1-\psi_0\phi_0)(1-\psi_0^2)(1-\phi_0^2)}, \\
\gamma(1) := \sum_{k\in\mathbb{Z}} d_k d_{k+1} &= \frac{1}{(1-\psi_0\phi_0)^2}\left[\frac{l}{1-\psi_0^2} + \frac{\phi_0}{1-\phi_0^2}\right], \\
\gamma(2) := \sum_{k\in\mathbb{Z}} d_k d_{k+2} &= \frac{1}{(1-\psi_0\phi_0)^2}\left[\frac{\psi_0^2}{1-\psi_0^2} + \frac{\phi_0^2}{1-\phi_0^2} + \psi_0\phi_0\right],
\end{aligned}$$

which yields

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} \quad \text{and} \quad \rho(2) = \frac{\gamma(2)}{\gamma(0)}.$$

The limit obtained at 6.5 does not correspond in general to the true values of the autoregressive parameters which are $\dfrac{\psi_0}{1+\psi_0\phi_0}$ and $\dfrac{\phi_0}{1+\psi_0\phi_0}$.

Numerical simulations were performed to gauge empirically this theoretical results and are presented on tables 1, 2 and 3. For sample sizes of 200, 500, 1000, 2000 and 5000 observations, 2000 simulations of pure noncausal AR(2) and MAR(1,1) with different dispositions of the roots were generated and the Least Square Estimate was computed. It can be noticed that in the case of the pure noncausal AR(2), the parameters converge indeed towards their true values, whereas in the case of the MAR(1,1), both parameters tend as expected to the same limit which does not correspond to the true values.

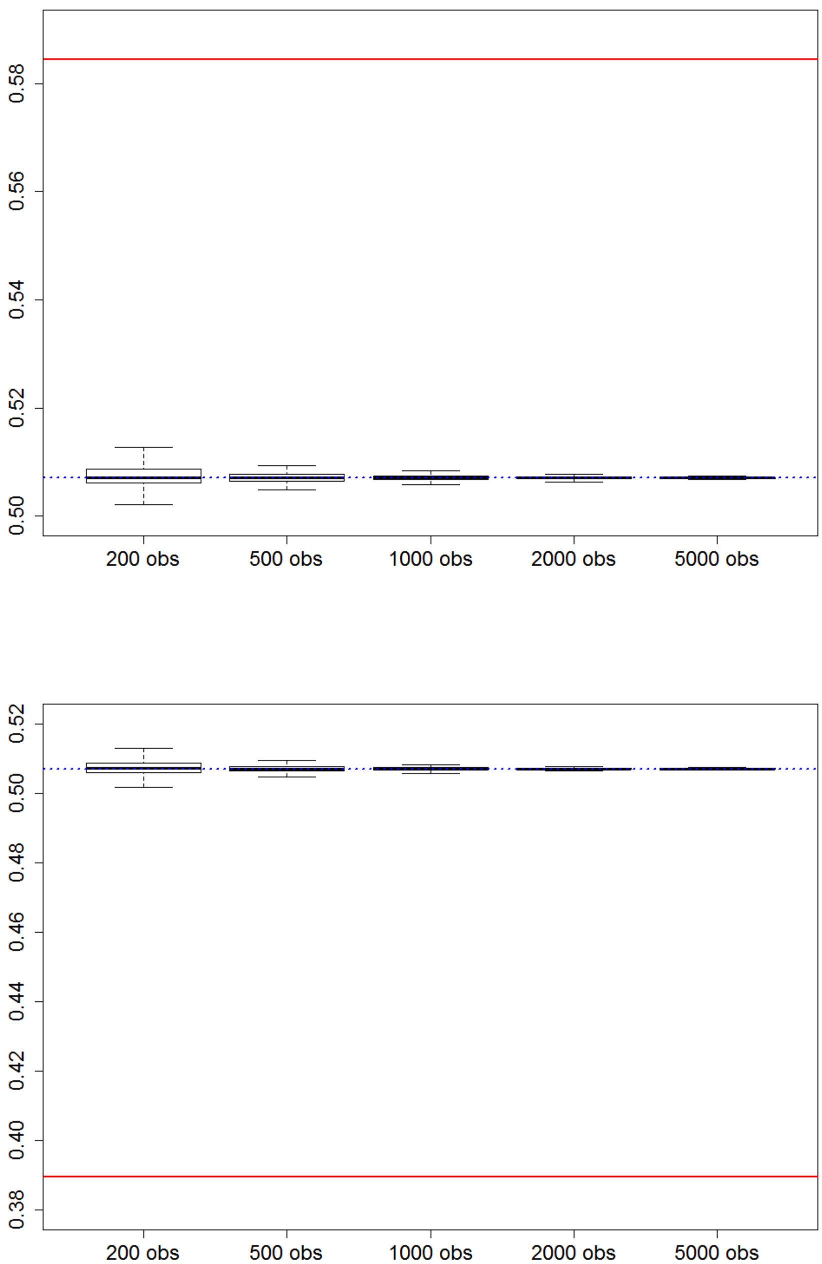


Table 1: Monte Carlo experiment: Least Square estimation of the noncausal MAR(1,1) process $X_t = \dfrac{\psi_0}{1+\psi_0\phi_0}X_{t+1} + \dfrac{\phi_0}{1+\psi_0\phi_0}X_{t-1} + \dfrac{\varepsilon_t}{1+\psi_0\phi_0}$, where (ε_t) is an iid sequence of standard Cauchy noise, $\psi_0 = 0{,}9$ and $\phi_0 = 0{,}6$. 2000 simulations for each sample sizes of 200, 500, 1000, 2000 and 5000 observations. Upper (resp. lower) panel summarises the values taken by the estimator of the coefficient associated to the forward term X_{t+1} (resp. backward term X_{t-1}). Red solid line: theoretical value of the parameter. Blue dotted line: theoretical limit of the estimator.

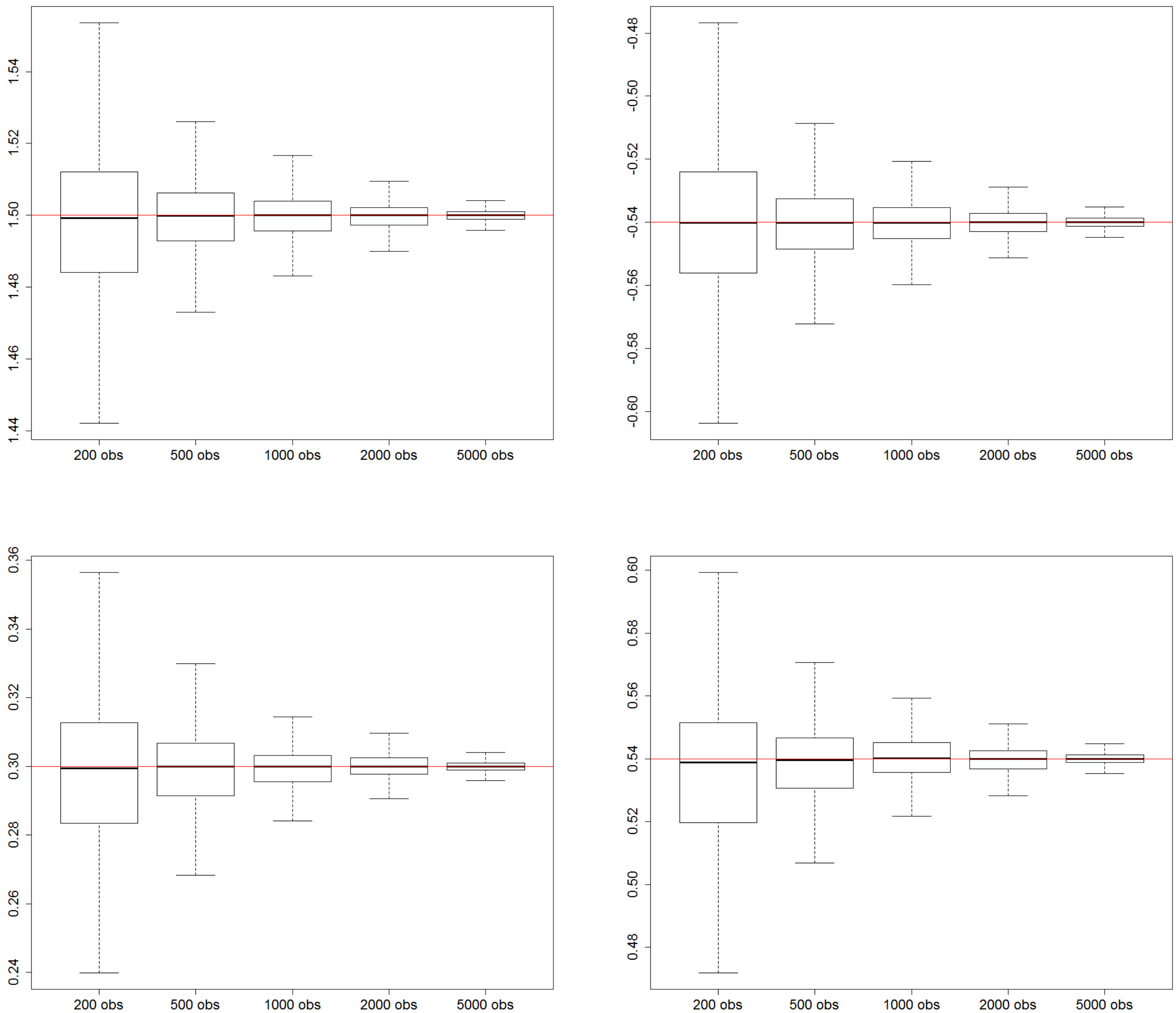


Table 2: Monte Carlo experiment: Least Square estimation of two noncausal AR(2) processes: $X_t = \psi_1 X_{t+1} + \psi_2 X_{t+2} + \varepsilon_t$, where (ε_t) is an iid standard Cauchy sequence. Model of the two upper panels: $\psi_1 = 1{,}5$ and $\psi_2 = -0{,}54$; model of the two lower panels: $\psi_1 = 0{,}3$ and $\psi_2 = 0{,}54$. Boxplots of $\hat{\psi}_1$ are on the left-hand side and boxplots of $\hat{\psi}_2$ are on the right-hand side. 2000 trajectories of length 200, 500, 1000, 2000 and 5000 for each model. Solid red line: true value of the parameter.

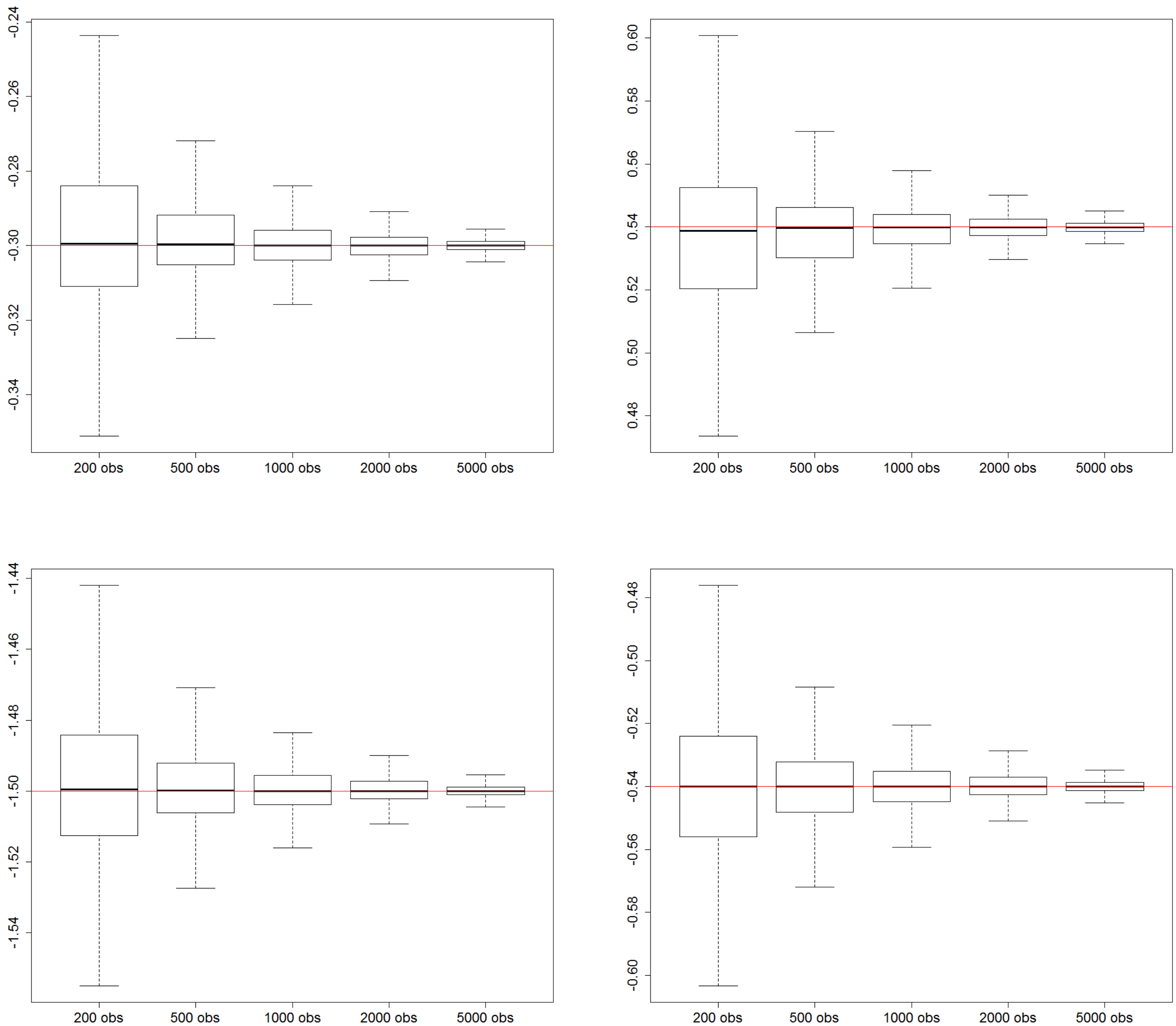


Table 3: Monte Carlo experiment: Least Square estimation of two noncausal AR(2) processes: $X_t = \psi_1 X_{t+1} + \psi_2 X_{t+2} + \varepsilon_t$, where (ε_t) is an iid standard Cauchy sequence. Model of the two upper panels: $\psi_1 = -0{,}3$ and $\psi_2 = 0{,}54$; model of the two lower panels: $\psi_1 = -1{,}5$ and $\psi_2 = -0{,}54$. Boxplots of $\hat{\psi}_1$ are on the left-hand side and boxplots of $\hat{\psi}_2$ are on the right-hand side. 2000 trajectories of length 200, 500, 1000, 2000 and 5000 for each model. Solid red line: true value of the parameter.

6.2 Non-linear Least Squares Estimation of a MAR$(1,1)$ process

Since the classical OLS are not consistent, a better approach is perhaps to estimate the process under its nonexpanded definition. We focus in this section on the simpler MAR(1,1) case. Let (X_t) be the strictly stationary solution of the equation

$$(1-\psi_0 F)(1-\phi_0 B)X_t = \varepsilon_t, \tag{6.6}$$

where $0 < |\psi_0| < 1$, $0 < |\phi_0| < 1$.

Let n be a positive integer and $X_1, \ldots, X_n$ be n observations of process (X_t). Consider the objective function

$$\mathcal{L}_n(\phi,\psi) = \sum_{t=2}^{n-1} \left[(1-\psi F)(1-\phi B)X_t\right]^2, \tag{6.7}$$

and let us consider the Nonlinear Least Square Estimate of (ϕ_0, ψ_0) defined by

$$(\hat{\phi}, \hat{\psi}) = \underset{(\phi,\psi)\in\mathbb{R}^2}{\arg\min}\, \mathcal{L}_n(\phi,\psi).$$

The estimator defined above is consistent, although not able to disentangle the causal from the noncausal component.

Proposition 6.2 *Under the above assumptions*

$$\{\hat{\phi}, \hat{\psi}\} \xrightarrow{p} \{\phi, \psi\}.$$

Figure 5 presents the results of a Monte Carlo experiment to gauge the convergence of the nonlinear LS estimator. 2000 simulations of process (6.6) with $\phi_0 = 0.7$ and $\psi_0 = 0.4$ with sample sizes of 200, 500, 1000 and 2000 observations were performed and the corresponding estimator $(\hat{\phi}, \hat{\psi})$ was computed. For each of the four sample sizes, we pooled together the estimates $\hat{\phi}$ and $\hat{\psi}$ and plotted the density of their location. Two spikes, well centered around the two true values ϕ_0 and ψ_0 and increasingly sharp with the sample size can be recognised: the estimators are indeed converging towards the actual values of the model. Not reported here are the density plots of the estimators of $\hat{\phi}$ and $\hat{\psi}$ not pooled together. As expected, two spikes are still visible, although of different size, which is not surprising since the nonlinear LS estimator is not able to identify causality and noncausality.

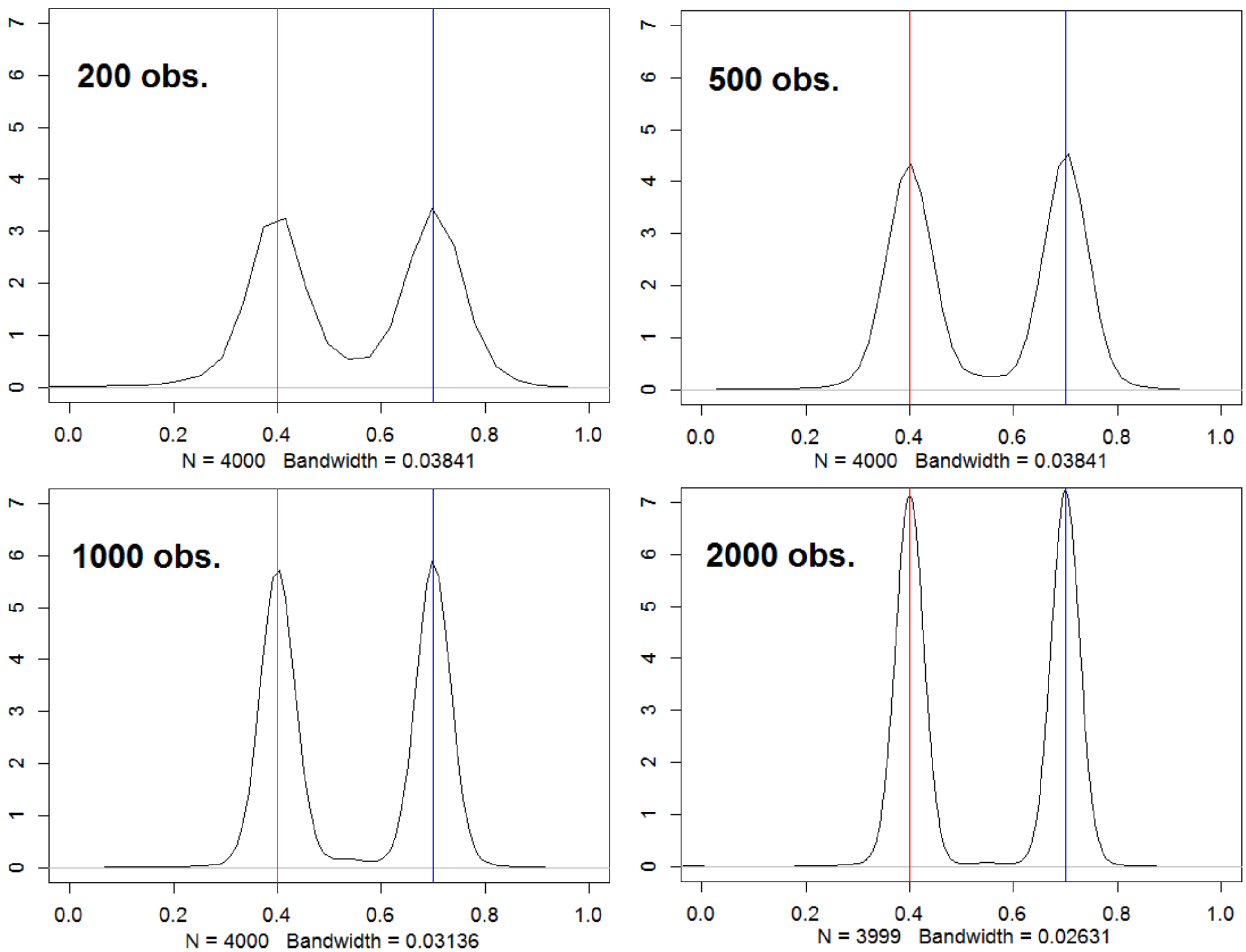


Figure 5: Density plots of the 2000 nonlinear LS estimates of $\{\hat{\phi}, \hat{\psi}\}$ (pooled together) from model (6.6) for four sample sizes : 200, 500, 1000 and 2000 observations. True values: $\phi_0 = 0.7$ (blue vertical line), $\psi_0 = 0.4$ (red vertical line).

A Proofs

A.1 For the results of Section 2

Proof of Proposition 2.1

We use the MA(∞) representation of X_t and the assumption $\varepsilon_t \overset{iid}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu)$:

$$\forall s \in \mathbb{R}, \quad \psi_{X_t}(s) = \mathbb{E}\left[e^{isX_t}\right] = \mathbb{E}\left[e^{is\sum\limits_{l=-\infty}^{+\infty} d_l \varepsilon_{t+l}}\right] = \prod_{l=-\infty}^{+\infty} \mathbb{E}\left[e^{isd_l\varepsilon_{t+l}}\right],$$

$$= \prod_{l=-\infty}^{+\infty} \exp\left\{-\sigma^\alpha |d_l s|^\alpha \left(1 - i\beta \text{sign}(d_l s) w(\alpha, d_l s)\right) + i d_l s \mu\right\}.$$

If $\alpha \neq 1$, then

$$\forall s \in \mathbb{R}, \quad \psi_{X_t}(s) = \prod_{l=-\infty}^{+\infty} \exp\left\{-\sigma^\alpha |d_l s|^\alpha \left(1 - i\beta \text{sign}(d_l s) \text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i d_l s \mu\right\},$$

$$= \exp\left\{-\sigma^\alpha \sum_{l=-\infty}^{+\infty} |d_l|^\alpha |s|^\alpha \left(1 - i\beta \frac{\sum\limits_{l=0}^{+\infty} |d_l|^\alpha \text{sign}(d_l)}{\sum\limits_{l=-\infty}^{+\infty} |d_l|^\alpha} \text{sign}(s) \text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i d_l s \mu\right\}.$$

Whereas if $\alpha = 1$, then

$$\forall s \in \mathbb{R}, \quad \psi_{X_t}(s) = \prod_{l=-\infty}^{+\infty} \exp\left\{-\sigma |d_l s| \left(1 + i\frac{2}{\pi}\beta \text{sign}(d_l s) \ln|d_l s|\right) + i d_l s \mu\right\},$$

$$= \exp\left\{-\sigma \sum_{l=-\infty}^{+\infty} |d_l||s| \left(1 + i\frac{2}{\pi}\beta \frac{\sum\limits_{l=0}^{+\infty} d_l}{\sum\limits_{l=-\infty}^{+\infty} |d_l|} \text{sign}(s) \ln|s|\right) + is\left(\sum_{l=-\infty}^{+\infty} d_l \mu - \frac{2}{\pi}\sigma\beta \sum_{l=-\infty}^{+\infty} d_l \ln|d_l|\right)\right\}.$$

Proof of Example 2.1 From Proposition 2.1, we deduce straightforwardly the tail, asymmetry and location parameters. As for the scale parameter, we need to compute $\sum_{l=0}^{+\infty} |d_l|$. Assume first that $\lambda_1 > 0$. Since we assumed $0 < |\lambda_2| \leq |\lambda_1| < 1$, then $d_l = \dfrac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} \geq 0$ for all l and $\tilde{\sigma} = \sigma \sum_{l=0}^{+\infty} d_l = \sigma \psi(1)^{-1}$ by definition of (d_l), and thus $\tilde{\sigma} = \dfrac{\sigma}{1 - \psi_1 - \psi_2}$.

Assume now that $\lambda_1 < 0$. Then $d_{2l} \geq 0$ whereas $d_{2l+1} \leq 0$ for all l and

$$
\begin{aligned}
\tilde{\sigma} &= \frac{\sigma}{\lambda_1 - \lambda_2} \left[\sum_{l=0}^{+\infty} \left(\lambda_1^{2l+1} - \lambda_2^{2l+1} \right) - \sum_{l=0}^{+\infty} \left(\lambda_1^{2l} - \lambda_2^{2l} \right) \right], \\
&= \frac{\sigma}{\lambda_1 - \lambda_2} \left[\frac{\lambda_1}{1-\lambda_1^2} - \frac{\lambda_2}{1-\lambda_2^2} - \frac{1}{1-\lambda_1^2} + \frac{1}{1-\lambda_2^2} \right], \\
&= \frac{\sigma}{(1+\lambda_1)(1+\lambda_2)}, \\
&= \sigma \psi^{-1}(-1), \\
&= \frac{\sigma}{1+\psi_1-\psi_2},
\end{aligned}
$$

which yields the conclusion.

Proof of Proposition 3.1 and 4.1

Define the pure noncausal AR(p) component (u_t) of the MAR process (X_t) (see Lanne and Saikkonen(2011) and Gouriéroux and Jasiak (2014)):

$$
u_t = \phi(B) X_t \iff \psi(F) u_t = \varepsilon_t,
$$

We first show that the pure noncausal AR(p) process (u_t) is a Markov process of order p. Since $u_t = \psi_1 u_{t+1} + \ldots + \psi_p u_{t+p} + \varepsilon_t$, we have for any lag $k > p$,

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

We now turn to the MAR process (X_t). From $u_t = \phi(B) X_t$, we deduce that $X_t = \sum_{i=1}^{q} \phi_i X_{t-i} + u_t$.

Thus,

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

hence process (X_t) is Markov of order $p+q$.

Proof of Theorem 3.1 The following Lemma and its Corollary, stated in the general MAR(p,q) case, will be useful for the proof:

Lemma A.1 *Let (X_t) be a MAR(p,q) process. For any $h \geq 0$, there exist polynomials P_h and Q_h, of respective degrees $p+q-1$ and h, such that for any $t \in \mathbb{Z}$,*

$$X_{t+h} = P_h(B)X_{t-1} + Q_h(B)\varepsilon_{t-p+h}. \tag{A.1}$$

Besides, polynomials P_h and Q_h verify the following recursive equations:

$$BP_{h+1}(B) = P_h(B) - \frac{P_h(0)}{\psi^*(0)}\psi^*(B)\phi(B), \qquad Q_{h+1}(B) = Q_h(B) + \frac{P_h(0)}{\psi^*(0)}B^{h+1}, \tag{A.2}$$

with initial conditions $Q_0(B) = \frac{1}{\psi^*(0)}$, $P_0(B) = B^{-1}\left(1-\frac{\psi^*(B)\phi(B)}{\psi^*(0)}\right) = \frac{1}{\theta_p}\Big\{-\theta_{p-1}-\theta_{p-2}B-\dots-\theta_{-q}B^{p+q-1}\Big\}$ *and* $\psi^*(B) := B^p\psi(F)$.

Corollary A.1 *Let (X_t) be a MAR(p,q) process. For any $h \geq 0$, there exist polynomials P_h, Q_{1h} and Q_{2h} of respective degrees $p+q-1$, $p-1$ and h such that for any $t \in \mathbb{Z}$,*

$$X_{t+h} = \Big(P_h(B) + Q_{1,h}(B)\phi(B)\Big)X_{t-1} + Q_{2,h}(F)u_t, \tag{A.3}$$

where $BQ_{1,h}(B) + Q_{2,h}(F) = B^{p-h}Q_h(B)\psi(F)$*, and P_h and Q_h are defined as in Lemma A.1.*

Let us now prove Theorem 3.1. Let (X_t) be a pure noncausal AR(p) process. From Lemma A.1 with $p>0$ and $q=0$, we have that for any $h \geq 0$, there exist polynomials P_h and Q_h of respective degree $p-1$ and h such that:

$$X_{t+h} - P_h(B)X_{t-1} = Q_h(B)\epsilon_{t-p}.$$

Denote $Y_{t-1} = P_h(B)X_{t-1}$ and $\varepsilon^*_{t-p,h} = Q_h(B)\epsilon_t$. The backward innovation $\varepsilon^*_{t-p,h}$ and Y_{t-1}, which could written as one-sided moving averages, follow stable distributions with tail exponents α. Letting $f_{\epsilon^*,h}$ denote the pdf of $\epsilon^*_{t-p,h}$, the pdf of Y_{t-1} given X_{t+h} is thus the function $y \longmapsto f_{\epsilon^*,h}(X_{t+h} - y)$. By the Bayes formula, the pdf of X_{t+h} given $Y_{t-1} = y$ is thus the function

$$g : x \longmapsto f_{\epsilon^*,h}(x - y)f_X(x)/f_Y(y), \tag{A.4}$$

where f_X denotes the marginal pdf of X_t and f_Y the pdf of Y_t. If $\alpha \geq 1$ or $|\beta| \neq 1$, we can conclude by the same argument as Gouriéroux and Zakoïan (2016) that process (X_t) admits moment up to order $2\alpha + 1$.

Proof of Proposition 3.2

If $\alpha < 1$ and $|\beta| = 1$, then the errors ε_t are either positive or negative almost surely. Assume $\alpha < 1$. When $\beta = 1$, then $\varepsilon_t > 0$ for all $t \in \mathbb{Z}$ almost surely. Then, the support of the density f_X is either $\mathbb{R}$ if at least one of the coefficient of the MA representation is strictly negative or is bounded below if all the coefficients are greater or equal to zero. If the support is $\mathbb{R}$, then the density g cannot have compact support and hence we conclude that process (X_t) admits conditional moment up to order $2\alpha+1$. In the other case where the support of f_X is bounded below, then the support of g is compact if and only if the support of $f_{\epsilon^*,h}$ is bounded above, which is equivalent to $\psi_p^{-1}b_{1,i} > 0$ for all $i \leq h$. If such is the case, g has compact support and X_t admits conditional moments at any order, otherwise it admits moments up to order $2\alpha + 1$.
Symmetrically, for $\beta = -1$, the support of f_X is bounded above if $d_l \geq 0$ for all l and is $\mathbb{R}$ otherwise. If it is $\mathbb{R}$, then (X_t) admits conditional moment up to order $2\alpha + 1$. If it is bounded above, then g has compact support if and only if the support of $f_{\epsilon^*,h}$ is bounded below, which is equivalent to $\psi_p^{-1}b_{1,i} < 0$ for all $i \leq h$. The conclusion follows.

Proof of Lemma A.1 and Corollary A.1

For simplicity, we will show the existence of polynomials $P_h(B)$ and $Q_h(B)$ for $p > 0$ and $q = 0$, that is, for a pure noncausal process. We will however show that the recursive equations (A.2) hold in the general case. We proceed by induction. For $h = 0$, the property holds by definition with $P_0(B) := \sum_{j=1}^{p} b_{j,0}B^{j-1}$ with $b_{j,0} = -\psi_p^{-1}\psi_{p-j}$ for $1 \leq j \leq p-1$ and $Q_0(B) := -\psi_p^{-1}$, where we introduced $\psi_0 = -1$. Let us now assume that for some $h \geq 0$, equation (A.1) holds. Let us denote

$P_h(B) := \sum_{j=1}^{p} b_{j,h} B^{j-1}$ and $Q_h(B) := \sum_{i=0}^{h} c_{i,h} B^i$. Equation (A.1) now reads:

$$X_{t+h} = b_{1,h} X_{t-1} + \sum_{j=2}^{p} b_{j,h} X_{t-j} + \sum_{i=0}^{h} c_{i,h} \epsilon_{t-p+h-i}.$$

Using equation (1.1) at time $t-p-1$, we replace X_{t-1} to get:

$$\begin{aligned} X_{t+h} &= b_{1,h} \left(\sum_{j=1}^{p} -\psi_p^{-1} \psi_{p-j} X_{t-j-1} - \psi_p^{-1} \epsilon_{t-p-1} \right) + \sum_{j=2}^{p} b_{j,h} X_{t-j} + \sum_{i=0}^{h} c_{i,h} \epsilon_{t-p+h-i}, \\ &= \sum_{j=1}^{p-1} \Big(-b_{1,h} \psi_p^{-1} \psi_{p-j} + b_{j+1,h} \Big) X_{t-j-1} - b_{1,h} \psi_p^{-1} \psi_0 X_{t-p-1} - b_{1,h} \psi_p^{-1} \epsilon_{t-p-1} + \sum_{i=0}^{h} c_{i,h} \epsilon_{t-p+h-i}. \end{aligned}$$

We now define

$$\forall 1 \leq j \leq p-1, \quad b_{j,h+1} = -b_{1,h} \psi_p^{-1} \psi_{p-j} + b_{j+1,h}, \tag{A.5}$$

$$b_{p,h+1} = -b_{1,h} \psi_p^{-1} \psi_0, \tag{A.6}$$

$$c_{h+1,h+1} = -b_{1,h} \psi_p^{-1}, \tag{A.7}$$

$$\forall 0 \leq i \leq h, \quad c_{i,h+1} = c_{i,h}. \tag{A.8}$$

Hence we obtain:

$$X_{(t-1)+(h+1)} = \sum_{j=1}^{p} b_{j,h+1} X_{(t-1)-j} - \psi_p^{-1} \sum_{i=0}^{h+1} b_{1,i-1} \epsilon_{(t-1)-p+(h+1)-i}.$$

with $b_{1,-1} = 1$. Since this relation holds for any $t \in \mathbb{Z}$, we can change the time index by $t' = t-1$ and obtain that equation (A.1) holds at horizon $h+1$. We have proven the existence of polynomials $P_h(B)$ and $Q_h(B)$ for any $h \geq 0$.

We now derive the recursive equations (A.2) in the MAR(p,q). Let h be a positive integer. We apply polynomial $\psi(F)\phi(B)$ to equation (A.1):

$$\begin{aligned} \psi(F)\phi(B) X_{t+h} &= P_h(B) \psi(F)\phi(B) X_{t-1} + Q_h(B) \psi(F)\phi(B) \varepsilon_{t-p+h}, \\ B^{-h} \varepsilon_t &= B P_h(B) \varepsilon_t + Q_h(B) \psi(B^{-1}) \phi(B) B^{p-h} \varepsilon_t, \end{aligned}$$

which implies:

$$B^{h+1} P_h(B) + Q_h(B) \psi^*(B) \phi(B) = 1,$$

where $\psi^*(B) = \psi(B^{-1}) B^p = -\psi_p - \psi_{p-1} B - \ldots - \psi_1 B^{p-1} + B^p$. The same holds at rank $h+1$: $B^{h+2} P_{h+1}(B) + Q_{h+1}(B) \psi^*(B) \phi(B) = 1$. Substracting both expression at ranks h and $h+1$ yields:

$$B^{h+1} \Big(B P_{h+1}(B) - P_h(B) \Big) + \psi^*(B) \phi(B) \Big(Q_{h+1}(B) - Q_h(B) \Big) = 0. \tag{A.9}$$

Since $\deg\Big(Q_{h+1}(B) - Q_h(B)\Big) \leq h+1$ and $\phi(0) = 1$, we can notice that:

$$-B^{h+1}P_h(0) + \psi^*(0)\Big(Q_{h+1}(B) - Q_h(B)\Big) = 0,$$

thus

$$Q_{h+1}(B) = Q_h(B) + \frac{P_h(0)}{\psi^*(0)}B^{h+1}. \tag{A.10}$$

Finally, replacing (A.10) in (A.9) concludes the proof of Lemma A.1.

Substituting ε_{t-p+h} by $\psi(F)u_{t-p+h}$ in Equation A.1 yields the Corollary.

□

We investigate further the coefficients of the polynomials P_h and Q_h in the pure noncausal framework $p > 0$ and $q = 0$ to derive results which will be useful in the coming proofs. Equations (A.5)-(A.8) define deterministic series of coefficients $(b_{j,h})_{\substack{h\geq 0\\ 1\leq j\leq p}}$, verifying for any j and h:

$$b_{j,h+1} = -\sum_{k=0}^{\min(p-j,h)} b_{1,h-k}\psi_p^{-1}\psi_{p-j-k}, \tag{A.11}$$

with initial terms:

$$\forall 1 \leq j \leq p, \quad b_{j,0} = -\psi_p^{-1}\psi_{p-j}.$$

Solving for $b_{1,h}$ is thus sufficient to reach the general terms of all the other coefficient series. From equation (A.5) and (A.6), we deduce or any $h \geq p-1$, $(b_{1,h})$ verifies:

$$\psi_p b_{1,h+1} = -\sum_{k=0}^{p-1} b_{1,h-k}\psi_{p-k-1},$$

which can be reformulated as:

$$b_{1,h-p} = \psi_1 b_{1,h-p+1} + \ldots + \psi_p b_{1,h}.$$

We recognise the characteristic polynomial ψ of equation (1.1) which was assumed to have s distinct roots $\lambda_1, \ldots, \lambda_s$ with respective multiplicities $m_1, \ldots, m_s$. Hence, there exist polynomials $\Pi_j, 1 \leq j \leq s$ such that for any j, $\deg(\Pi_j) = m_j - 1$ and for any $h \geq p-1$:

$$b_{1,h} = \sum_{j=1}^{s} \Pi_j(h)\lambda_j^{-h}$$

Since polynomial Π_j has m_j coefficients, there is a total of p coefficients to be determined in order to have the $(b_{j,h})_{j,h}$ explicitly. For $0 \leq h \leq p-1$, the coefficients of polynomials P_h and Q_h can be easily computed by iterating over the equations (A.5)-(A.8). Given $b_{1,0}, \ldots, b_{1,p-1}$, the coefficients of the polynomials $\Pi_j(h) := \pi_{0,j} + \pi_{1,j}h + \ldots + \pi_{m_j-1,j}h^{m_j-1}$ can be identified for all j. Indeed, for any h, $0 \leq h \leq p-1$, we have:

$$\left(\pi_{0,1} + \pi_{1,1}h + \ldots + \pi_{m_1-1,1}h^{m_1-1}\right)\lambda_1^{-h} + \ldots + \left(\pi_{0,p} + \pi_{1,p}h + \ldots + \pi_{m_p-1,j}h^{m_p-1}\right)\lambda_p^{-h} = b_{1,h}.$$

The system composed of the p equations for $0 \leq h \leq p-1$ can be seen as a linear system in the coefficients $(\pi_{i,j})$ with two intricated Vandermonde matrices which nodes are respectively the integers between 0 and $p-1$ (with the convention $0^0 = 1$) on the one hand, and the roots $\lambda_1, \ldots, \lambda_s$ on the other hand. The nodes within each Vandermonde matrices being clearly distinct, we know that they are invertible and we thus obtain for $1 \leq j \leq s$ the coefficients of the polynomial Π_j, associated with the unique root λ_j. The series of $(b_{1,h})$ is hence fully determined, and the series $(b_{j,h})_{h\geq 0}$ follow.
Note that in particular, if the roots (λ_j) are simple, the polynomials Π_j collapse to constants.

Proof of Proposition 3.3

If $\alpha \neq 1$, for any $(u,v) \in \mathbb{R}^2$:

$$\begin{aligned}
\varphi(u,v) &= \mathbb{E}\left[e^{iuY_{t-1}+ivX_{t+h}}\right] \\
&= \mathbb{E}\left[e^{ivX_{t+h}}\mathbb{E}\left[e^{iuY_{t-1}}\middle| X_{t+h}\right]\right] \\
&= \mathbb{E}\left[e^{i(u+v)X_{t+h}}\right]\mathbb{E}\left[e^{-iu\epsilon^*_{t-p,h}}\right] \\
&= \exp\left\{-\tilde{\sigma}^\alpha|u+v|^\alpha\left(1 - i\tilde{\beta}\,\text{sign}(u+v)\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + iv\tilde{\mu} - \sigma^\alpha_{\epsilon,h}|u|^\alpha\left(1 - i\beta_{\epsilon,h}\,\text{sign}(u)\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + iu(\tilde{\mu} + \mu_{\epsilon,h})\right.
\end{aligned}$$

whereas if $\alpha = 1$:

$$\varphi(u,v) = \exp\left\{-\tilde{\sigma}|u+v|\left(1 + \frac{2}{\pi}i\tilde{\beta}\,\text{sign}(u+v)\ln|u+v|\right) + iv\tilde{\mu} - \sigma_{\epsilon,h}|u|\left(1 + \frac{2}{\pi}i\beta_{\epsilon,h}\,\text{sign}(u)\ln|u|\right) + iu(\tilde{\mu} + \mu_{\epsilon,h})\right\}$$

Thus, if $\alpha \neq 1$, for $u > 0$, in the vicinity of $v = 0$,

$$\begin{aligned}
\frac{\partial\varphi}{\partial u} &= \left[-\alpha\tilde{\sigma}^\alpha(u+v)^{\alpha-1}\left(1 - i\tilde{\beta}\,\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) - \alpha\sigma^\alpha_{\epsilon,h}u^{\alpha-1}\left(1 - i\beta_{\epsilon,h}\,\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i(\tilde{\mu} + \mu_{\epsilon,h})\right]\varphi(u,v) \\
\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} &= \left[-\alpha\left(\tilde{\sigma}^\alpha + \sigma^\alpha_{\epsilon,h}\right)u^{\alpha-1}\left(1 - i\frac{\tilde{\beta}\tilde{\sigma}^\alpha + \beta_{\epsilon,h}\sigma^\alpha_{\epsilon,h}}{\tilde{\sigma}^\alpha + \sigma^\alpha_{\epsilon,h}}\,\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i(\tilde{\mu} + \mu_{\epsilon,h})\right]\varphi(u,0)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial\varphi}{\partial v} &= \left[-\alpha\tilde{\sigma}^\alpha(u+v)^{\alpha-1}\left(1 - i\tilde{\beta}\,\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i\tilde{\mu}\right]\varphi(u,v) \\
\left.\frac{\partial\varphi}{\partial v}\right|_{v=0} &= \left[-\alpha\tilde{\sigma}^\alpha u^{\alpha-1}\left(1 - i\tilde{\beta}\,\text{tg}\left(\frac{\pi\alpha}{2}\right)\right) + i\tilde{\mu}\right]\varphi(u,0)
\end{aligned}$$

Therefore:

$$\left.\frac{\partial\varphi}{\partial v}\right|_{v=0} - i\tilde{\mu}\varphi(u,0) = \frac{\tilde{\sigma}^\alpha}{\sigma^\alpha_{\epsilon,h}+\tilde{\sigma}^\alpha} \frac{1 - i\tilde{\beta}\,\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)}{1 - i\frac{\tilde{\beta}\tilde{\sigma}^\alpha + \beta_{\epsilon,h}\sigma^\alpha_{\epsilon,h}}{\tilde{\sigma}^\alpha + \sigma^\alpha_{\epsilon,h}}\,\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)} \left(\left.\frac{\partial\varphi}{\partial u}\right|_{v=0} - i(\tilde{\mu}+\mu_{\epsilon,h})\varphi(u,0)\right) \tag{A.12}$$

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

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

Therefore, for $u > 0$:

$$\mathbb{E}\left[\left[(X_{t+h}-\tilde{\mu}) - \frac{\tilde{\sigma}^\alpha}{\sigma^\alpha_{\epsilon,h}+\tilde{\sigma}^\alpha} \frac{1 - i\tilde{\beta}\,\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)}{1 - i\frac{\tilde{\beta}\tilde{\sigma}^\alpha + \beta_{\epsilon,h}\sigma^\alpha_{\epsilon,h}}{\tilde{\sigma}^\alpha + \sigma^\alpha_{\epsilon,h}}\,\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)} \left(Y_{t-1} - (\tilde{\mu}+\mu_{\epsilon,h})\right)\right] e^{iuY_{t-1}}\right] = 0 \tag{A.13}$$

It can be checked that for $u < 0$:

$$\mathbb{E}\left[\left[(X_{t+h}-\tilde{\mu}) - \frac{\tilde{\sigma}^\alpha}{\sigma^\alpha_{\epsilon,h}+\tilde{\sigma}^\alpha} \frac{1 + i\tilde{\beta}\,\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)}{1 + i\frac{\tilde{\beta}\tilde{\sigma}^\alpha + \beta_{\epsilon,h}\sigma^\alpha_{\epsilon,h}}{\tilde{\sigma}^\alpha + \sigma^\alpha_{\epsilon,h}}\,\mathrm{tg}\left(\frac{\pi\alpha}{2}\right)} \left(Y_{t-1} - (\tilde{\mu}+\mu_{\epsilon,h})\right)\right] e^{iuY_{t-1}}\right] = 0 \tag{A.14}$$

Hence, in the symmetric case ($\beta = 0$), from Bierens (Theorem 1, 1982):

$$\mathbb{E}\left[(X_{t+h}-\tilde{\mu}) - \frac{\tilde{\sigma}^\alpha}{\sigma^\alpha_{\epsilon,h}+\tilde{\sigma}^\alpha}\left(Y_{t-1} - (\tilde{\mu}+\mu_{\epsilon,h})\right)\middle| Y_{t-1}\right] = 0, \quad for \quad \alpha \neq 1$$

It can be checked that in the symmetric case, this formula still holds for $\alpha = 1$.

thus:

$$\mathbb{E}\left[X_{t+h} | Y_{t-1}\right] = \tilde{\mu} + \frac{\tilde{\sigma}^\alpha}{\sigma^\alpha_{\epsilon,h}+\tilde{\sigma}^\alpha}\left(Y_{t-1} - (\tilde{\mu}+\mu_{\epsilon,h})\right)$$

and since: $Y_{t-1} = \sum_{j=1}^{p} b_{j,h}X_{t-j}$,

$$\mathbb{E}\left[\mathbb{E}\left[X_{t+h}|\, Y_{t-1}\right]|\, X_{t-1},\ldots,X_{t-p}\right] = \mathbb{E}\left[X_{t+h}|\, X_{t-1},\ldots,X_{t-p}\right]$$

Finally:

$$\mathbb{E}\left[X_{t+h}|\, X_{t-1},\ldots,X_{t-p}\right] = \tilde{\mu} + \frac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha}_{\epsilon,h} + \tilde{\sigma}^{\alpha}}\left(\sum_{j=1}^{p} b_{j,h}X_{t-j} - (\tilde{\mu} + \mu_{\epsilon,h})\right)$$

Proof of Corollary 3.1

We will derive the asymptotic behaviour of $P_h(B)X_{t-1} = \sum_{j=1}^{p} b_{j,h}X_{t-j}$, conditional on $X_{t-1},\ldots,X_{t-p}$, and $\sigma^{\alpha}_{\varepsilon,h} = \sum_{i=0}^{h}|b_{1,h}/\psi_p|^{\alpha}$ with respect to h. We will first exhibit an equivalent of $b_{1,h}$, from which equivalents of $b_{j,h}$ can be easily deduced for any $2 \le j \le p$. This will yield an equivalent of $P_h(B)X_{t-1}$. Finally, we will use our previous result on the asymptotic behaviour of $b_{1,h}$ to find an equivalent of $\sigma^{\alpha}_{\varepsilon,h}$ and we will be able to conclude.

- Equivalent of $b_{1,h}$

  We know from the proof of Lemma A.1 that

  $$b_{1,h} = \sum_{j=1}^{s}\Pi_j(h)\lambda_j^{-h},$$

  where the $\lambda_1,\ldots,\lambda_s$ are the distinct roots with multiplicities $m_1,\ldots,m_s$ of the characteristic equation of process X_t and $\Pi_1,\ldots,\Pi_s$ are polynomials with degrees $m_1-1,\ldots,m_s-1$. Without loss of generality we can assume that the roots are ordered: $0 < |\lambda_1| < \cdots < |\lambda_s| < 1$. Denoting λ_1 and m_1 respectively by λ and m, and denoting by C the coefficient associated to the monomial of highest degree of Π_1, we have

  $$b_{1,h} \underset{h\to+\infty}{\sim} Ch^{m-1}\lambda^{-h}, \quad \text{and} \quad |b_{1,h}| \underset{h\to+\infty}{\longrightarrow} +\infty. \tag{A.15}$$

- Equivalent of $b_{j,h}/b_{1,h}$, $2 \le j \le p$

  Still from Lemma A.1, equation (A.11), we have for any $1 \le j \le p$

  $$\forall h \ge p-j, \quad b_{j,h+1} = -\sum_{k=0}^{p-j} b_{1,h-k}\psi_p^{-1}\psi_{p-j-k}.$$

  and since for any $k \ge 0$, $b_{1,h-k}/b_{1,h+1} \longrightarrow \lambda^{-(k+1)}$

  $$\frac{b_{j,h+1}}{b_{1,h+1}} \underset{h\to+\infty}{\longrightarrow} -\sum_{k=0}^{p-j}\lambda^{k+1}\psi_p^{-1}\psi_{p-j-k} := l_j.$$

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

  Therefore,

$$\frac{P_h(B)X_{t-1}}{b_{1,h}} = \sum_{j=1}^{p} \frac{b_{j,h}}{b_{1,h}} X_{t-j} \underset{h\to+\infty}{\longrightarrow} X_{t-1} + \sum_{j=2}^{p} l_j X_{t-j}.$$

  Since $\mathbb{P}\left(X_{t-1} + \sum_{j=2}^{p} l_j X_{t-j} = 0\right) = 0$, we have almost surely

$$P_h(B)X_{t-1} \underset{h\to+\infty}{\sim} b_{1,h}\left(X_{t-1} + \sum_{j=2}^{p} l_j X_{t-j}\right).$$

- Equivalent of $\sigma_{\varepsilon,h}^{\alpha}$

  Recall that $\sigma_{\varepsilon,h}$ is the scale parameter of $Q_h(B)\varepsilon_{t-p+h} = -\psi_p^{-1}\sum_{i=0}^{h} b_{1,i-1}\varepsilon_{t-p+h-i}$ which follows an α-stable distribution. It can be shown, following the proof of of Proposition 2.1 that $\sigma_{\varepsilon,h}^{\alpha} = \sigma^{\alpha}\sum_{i=0}^{h} |b_{1,i-1}/\psi_p|^{\alpha}$. We know from (A.15) that $b_{1,h}$ tends to infinity with h. Since $\sum_{i=0}^{h} |Ci^{m-1}\lambda^{-i}|^{\alpha}$ diverges for any $\alpha \in (0,2)$, it follows that

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

  If $m = 1$, it follows directly that

$$\sum_{i=0}^{h} |b_{1,i-1}/\psi_p|^{\alpha} \underset{h\to+\infty}{\sim} |C/\psi_p| \frac{|\lambda|^{-h\alpha}}{1-|\lambda|^{\alpha}}.$$

  However if $m \geq 2$, factorising by $|\lambda|^{-h\alpha}$ and developing $\left(1-\frac{i}{h}\right)^{\alpha(m-1)}$ into a power series:

$$\begin{aligned}\sum_{i=0}^{h} i^{\alpha(m-1)}|\lambda|^{-i\alpha} &= |\lambda|^{-h\alpha} h^{\alpha(m-1)} \sum_{i=0}^{h} \left(1-\frac{i}{h}\right)^{\alpha(m-1)} |\lambda|^{i\alpha},\\ &= |\lambda|^{-h\alpha} h^{\alpha(m-1)} \sum_{k=0}^{+\infty} (-1)^k \mathrm{C}_{\alpha(m-1)}^{k} h^{-k} \sum_{i=0}^{h} i^k |\lambda|^{i\alpha},\end{aligned}$$

  where for any $r \in \mathbb{R}$, $\mathrm{C}_r^k := \frac{(r)_k}{k!} = \frac{r(r-1)\cdots(r-k+1)}{k!}$. We show now that $\sum_{k=0}^{+\infty}(-1)^k \mathrm{C}_{\alpha(m-1)}^{k} h^{-k} \sum_{i=0}^{h} i^k|\lambda|^{i\alpha}$ tends to a finite limit as h tends to infinity. Define the series of functions:

$$\forall x \geq 1, \quad s_n(x) = \sum_{k=0}^{n} (-1)^k \mathrm{C}^k_{\alpha(m-1)} x^{-k} \sum_{i=0}^{\lfloor x \rfloor} i^k |\lambda|^{i\alpha}.$$

(s_n) converges uniformly on $[1, +\infty)$, indeed for $x > 1$:

$$\begin{aligned} |s_n(x)| &\leq \sum_{k=0}^{n} \left| \mathrm{C}^k_{\alpha(m-1)} \right| x^{-k} \sum_{i=0}^{\lfloor x \rfloor} \left(\lfloor x \rfloor \right)^k |\lambda|^{i\alpha}, \\ &\leq \frac{1}{1 - |\lambda|^\alpha} \sum_{k=0}^{n} \left| \mathrm{C}^k_{\alpha(m-1)} \right|. \end{aligned} \tag{A.16}$$

A characterisation of Euler's Gamma function by Gauss (see Krantz (2012)) reads:

$$\forall z \in \mathbb{C}, \quad \Gamma(z) = \lim_{k \to +\infty} \frac{k! k^z}{z(z+1)\dots(z+k)}.$$

Therefore, for $z = -\alpha(m-1)$:

$$\Gamma\Big(- \alpha(m-1) \Big) \underset{k \to +\infty}{\sim} \frac{1}{\mathrm{C}^k_{\alpha(m-1)} \Big(\alpha(m-1) - k \Big)} (-1)^{k+1} k^{-\alpha(m-1)},$$

which yields the following equivalent for the general term of the series appearing on the righthandside of inequality (A.16):

$$\left| \mathrm{C}^k_{\alpha(m-1)} \right| \underset{k \to +\infty}{\sim} \frac{M}{k^{1+\alpha(m-1)}}.$$

where $M > 0$ is a constant. Since $\alpha > 0$, and we assumed $m \geq 2$, we have shown the uniform convergence of (s_n) on $[1, +\infty)$. Therefore:

$$\begin{aligned} \lim_{h \to +\infty} \left[\sum_{k=0}^{+\infty} (-1)^k \mathrm{C}^k_{\alpha(m-1)} h^{-k} \sum_{i=0}^{h} i^k |\lambda|^{i\alpha} \right] &= \sum_{k=0}^{+\infty} \lim_{h \to +\infty} \left[(-1)^k \mathrm{C}^k_{\alpha(m-1)} h^{-k} \sum_{i=0}^{h} i^k |\lambda|^{i\alpha} \right], \\ &= \frac{1}{1 - |\lambda|^\alpha}. \end{aligned}$$

Which gives us the following equivalent:

$$\sigma^\alpha_{\varepsilon,h} = \sigma^\alpha \sum_{i=0}^{h} |b_{1,i-1}/\psi_p|^\alpha \underset{h \to +\infty}{\sim} \sigma^\alpha |C/\psi_p| \frac{|\lambda|^{-h\alpha} h^{\alpha(m-1)}}{1 - |\lambda|^\alpha}.$$

- Conclusion

We deduce the asymptotic behaviour of $\mathbb{E}\left[X_{t+h} \middle| X_{t-1}, \dots, X_{t-p} \right]$ from the previous results:

$$\frac{\tilde{\sigma}^\alpha}{\sigma^\alpha_{\epsilon,h}+\tilde{\sigma}^\alpha}P_h(B)X_{t-1} \underset{h\to+\infty}{\sim} \left(\sum_{l=0}^{+\infty}|d_l|^\alpha\right)\left(\frac{\lambda}{|\lambda|}\right)^h \text{sign}\,(C/\psi_p)\,(1-|\lambda|^\alpha)\left(X_{t-1}+\sum_{j=2}^{p}l_jX_{t-j}\right)\left(h^{m-1}|\lambda|^{-h}\right)^{1-\alpha}. \tag{A.17}$$

Proof of Proposition 3.2

From the proof of Proposition 3.3, it can be noticed that the closed formula obtained for the first conditional moment can be extended to errors with asymmetric stable distributions at least in some cases. Indeed, to apply Bierens Theorem to equations (A.13) and (A.14), a weaker condition than $\beta = 0$ is sufficient, namely that $\tilde{\beta} = \beta_{\varepsilon,h}$. At horizon $h = 0$, we have

$$\beta_{\varepsilon,0} = \beta\text{sign}(\psi_p),$$
$$\tilde{\beta} = \beta\,\frac{\sum_{l=0}^{+\infty}|d_l|^\alpha\text{sign}(d_l)}{\sum_{l=0}^{+\infty}|d_l|^\alpha}$$

Therefore when $\beta \neq 0$, $\tilde{\beta} = \beta_{\varepsilon,0}$ is equivalent to

$$\sum_{l=0}^{+\infty}|d_l|^\alpha\,(\text{sign}(\psi_p)-\text{sign}(d_l)) = 0.$$

Noticing that the general term of the above series is of constant sign, *i.e.* positive if $\psi_p > 0$, negative if $\psi_p < 0$, the last equation holds true if and only if d_l is of the sign of ψ_p for all l. The following lemma gives insights on the sign of (d_l).

Lemma A.2 *Let $-1 < \lambda_1 < \ldots < \lambda_p < 1$ be p distinct ordered roots different from 0 of the AR polynomial. Then for any $l \geq 0$, there exists $\xi_l \in [\lambda_1, \lambda_p]$ such that*

$$d_l = C^{p-1}_{l+p-1}\xi_l^l$$

This characterisation implies that for even indexes l, d_l is always positive. Furthermore, if the roots are all positive (resp. negative), then d_l is positive (resp. negative) for all odd l indexes. This result, though partial, allows us to conclude that the condition $\tilde{\beta} = \beta_{\varepsilon,0}$ can never been met whenever $d_{l_0} < 0$ for some integer l_0, but it is automatically met in the case where all the roots are positive.

Proof of Lemma A.2

Let $p \geq 2$, $l \geq 0$ and assume $-1 < \lambda_1 < \ldots < \lambda_p < 1$. We reformulate d_l to show that it is equal (up to a positive multiplicative constant) to the interpolation error between the polynomial

X^{l+p-1} and the Lagrangian interpolation polynomial of the nodes $(\lambda_1, \lambda_1^{l+p-1}), \ldots, (\lambda_{p-1}, \lambda_{p-1}^{l+p-1})$, evaluated at point $X = \lambda_p$. We begin by exhibiting an analytical expression of d_l as a function of the roots of the autoregressive polynomial. The unique strictly stationary solution of X_t verifying equation (1.1) reads:

$$X_t = \frac{1}{(1-\lambda_1 F)\ldots(1-\lambda_p F)}\varepsilon_t = \Psi(F)^{-1}\varepsilon_t.$$

Decomposing $\Psi(F)^{-1}$ into partial elements, it can be shown that it satisfies:

$$\Psi(F)^{-1} = \sum_{j=1}^{p} \frac{a_j}{1-\lambda_j F} = \sum_{l=0}^{+\infty} F^l \sum_{j=1}^{p} a_j \lambda_j^l,$$

with

$$a_j = \frac{\lambda_j^{p-1}}{\prod\limits_{\substack{i=1\\i\neq j}}^{p} (\lambda_j - \lambda_i)}. \tag{A.18}$$

We identify the coefficients d_l of the MA(∞) representation of X_t:

$$\forall l \geq 0, \quad d_l = \sum_{j=1}^{p} \frac{\lambda_j^{l+p-1}}{\prod\limits_{\substack{i=1\\i\neq j}}^{p} (\lambda_j - \lambda_i)}$$

We now show that these coefficients can be seen as an Lagrange interpolation error:

$$\begin{aligned}
d_l &= \sum_{j=1}^{p} \frac{\lambda_j^{l+p-1}}{\prod\limits_{\substack{i=1\\i\neq j}}^{p} (\lambda_j - \lambda_i)} \\
&= \sum_{j=1}^{p-1} \frac{\lambda_j^{l+p-1}}{(\lambda_j - \lambda_p)\prod\limits_{\substack{i=1\\i\neq j}}^{p-1} (\lambda_j - \lambda_i)} + \frac{\lambda_p^{l+p-1}}{\prod\limits_{i=1}^{p-1} (\lambda_p - \lambda_i)} \\
&= \frac{1}{\prod\limits_{i=1}^{p-1} (\lambda_p - \lambda_i)} \left[\lambda_p^{l+p-1} - \sum_{j=1}^{p-1} \lambda_j^{l+p-1} \prod_{\substack{i=1\\i\neq j}}^{p-1} \frac{\lambda_p - \lambda_i}{\lambda_j - \lambda_i} \right].
\end{aligned}$$

We know from Abramowitz and Stegun (1964) that the interpolation error between a function f, p times differentiable, and the Lagrangian interpolation polynomial of nodes $(x_0, f(x_0)), \ldots, (x_{p-1}, f(x_{p-1}))$, denoted L, evaluated at point x can be expressed as follows:

$$\exists \xi \in [\min(x_0, \ldots, x_{p-1}, x), \max(x_0, \ldots, x_{p-1}, x)], \quad f(x) - L(x) = \frac{f^{(p)}(\xi)}{p!} \prod_{i=0}^{p-1} (x - x_i).$$

In our case, f is the monomial X^{l+p-1}. Hence, there exists $\xi_l \in [\lambda_1, \lambda_p]$ such that:

$$d_l = \mathrm{C}_{l+p-1}^{p-1} \xi_l^l,$$

which gives us straightforwardly that $d_{2l} \geq 0$ for any integer l. Furthermore, if $\lambda_1 > 0$ (resp. $\lambda_p < 0$), that is if all roots are positive (resp. negative), we also have $d_{2l+1} > 0$ (resp. $d_{2l+1} < 0$).

Proof of Proposition 3.4

$\beta = 0$. Let us denote $\tilde{Y}_{t-1} = Y_{t-1} - (\tilde{\mu} + \mu_{\epsilon,h})$ and $\tilde{X}_{t+h} = X_{t+h} - \tilde{\mu}$.

$$\begin{aligned}
\varphi(u,v) &:= \ln \mathbb{E}\left[e^{iu\tilde{Y}_{t-1} + iv\tilde{X}_{t+h}}\right] \\
&= \ln\left(\mathbb{E}\left[e^{iuY_{t-1} + ivX_{t+h}}\right] e^{-iu(\tilde{\mu}+\mu_{\epsilon,h}) - iv\tilde{\mu}}\right) \\
&= \ln\left(\exp\left\{-\tilde{\sigma}^\alpha |u+v|^\alpha + iv\tilde{\mu} - \sigma_{\epsilon,h}^\alpha |u|^\alpha + iu(\tilde{\mu} + \mu_{\epsilon,h})\right\} \exp\left\{-iu(\tilde{\mu}+\mu_{\epsilon,h}) - iv\tilde{\mu}\right\}\right) \\
&= -\tilde{\sigma}^\alpha |u+v|^\alpha - \sigma_{\epsilon,h}^\alpha |u|^\alpha
\end{aligned}$$

Thus, for $u > 0$:

$$\begin{aligned}
\left.\frac{\partial \varphi}{\partial u}\right|_{v=0} &= -\alpha \left(\tilde{\sigma} + \sigma_{\epsilon,h}\right) |u|^{\alpha-1} \\
\left.\frac{\partial^2 \varphi}{\partial u^2}\right|_{v=0} &= \alpha(1-\alpha) \left(\tilde{\sigma} + \sigma_{\epsilon,h}\right) |u|^{\alpha-2}
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial \varphi}{\partial v}\right|_{v=0} &= -\alpha\tilde{\sigma}|u|^{\alpha-1} \\
\left.\frac{\partial^2 \varphi}{\partial v^2}\right|_{v=0} &= \alpha(1-\alpha)\tilde{\sigma}|u|^{\alpha-2}
\end{aligned}$$

Therefore:

$$\left.\frac{\partial^2 \varphi}{\partial v^2}\right|_{v=0} = \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}} \left.\frac{\partial^2 \varphi}{\partial u^2}\right|_{v=0} \qquad \left.\frac{\partial \varphi}{\partial v}\right|_{v=0} = \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}} \left.\frac{\partial \varphi}{\partial u}\right|_{v=0} \tag{A.19}$$

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

$$\begin{aligned}
\left.\frac{\partial^2 \varphi}{\partial u^2}\right|_{v=0} &= \frac{-\mathbb{E}\left[\tilde{Y}_{t-1}^2 e^{iu\tilde{Y}_{t-1}}\right]}{\mathbb{E}\left[e^{iu\tilde{Y}_{t-1}}\right]} - \left(\left.\frac{\partial \varphi}{\partial u}\right|_{v=0}\right)^2 \\
\left.\frac{\partial^2 \varphi}{\partial v^2}\right|_{v=0} &= \frac{-\mathbb{E}\left[\tilde{X}_{t+h}^2 e^{iu\tilde{Y}_{t-1}}\right]}{\mathbb{E}\left[e^{iu\tilde{Y}_{t-1}}\right]} - \left(\left.\frac{\partial \varphi}{\partial v}\right|_{v=0}\right)^2
\end{aligned}$$

Therefore, for $\alpha = 1$ and $u > 0$:

$$\mathbb{E}\left[\left(\tilde{X}_{t+h}^2 - \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}}\tilde{Y}_{t-1}^2 - \sigma_{\epsilon,h}\tilde{\sigma}\right) e^{iu\tilde{Y}_{t-1}}\right] = 0$$

It can be checked that this formula still holds for $u < 0$. Hence, from Bierens (Theorem 1, 1982):

$$\mathbb{E}\left[\tilde{X}_{t+h}^2 - \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}}\tilde{Y}_{t-1}^2 - \sigma_{\epsilon,h}\tilde{\sigma}\middle| Y_{t-1}\right] = 0$$

thus:

$$\mathbb{E}\left[X_{t+h}^2 - 2\tilde{\mu}X_{t+h} + \tilde{\mu}^2\middle| Y_{t-1}\right] = \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}}\left(Y_{t-1} - (\tilde{\mu} + \mu_{\epsilon,h})\right)^2 + \sigma_{\epsilon,h}\tilde{\sigma}$$

which implies that

$$\begin{aligned}\mathbb{E}\left[X_{t+h}^2\middle| Y_{t-1}\right] &= 2\tilde{\mu}\left(\tilde{\mu} + \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}}\left(Y_{t-1} - (\tilde{\mu} + \mu_{\epsilon,h})\right)\right) - \tilde{\mu}^2 + \frac{\tilde{\sigma}}{\sigma_{\epsilon,h} + \tilde{\sigma}}\left(Y_{t-1} - (\tilde{\mu} + \mu_{\epsilon,h})\right)^2 + \sigma_{\epsilon,h}\tilde{\sigma} \\ &= \frac{1}{\sigma_{\epsilon,h} + \tilde{\sigma}}\left[\tilde{\sigma}\left(Y_{t-1} - \mu_{\epsilon,h}\right)^2 + \sigma_{\epsilon,h}\tilde{\mu}^2\right] + \sigma_{\epsilon,h}\tilde{\sigma}\end{aligned}$$

and since: $Y_{t-1} = P_h(B)X_{t-1}$,

$$\mathbb{E}\left[\mathbb{E}\left[X_{t+h}^2\middle| Y_{t-1}\right]\middle| X_{t-1}, \dots, X_{t-p}\right] = \mathbb{E}\left[X_{t+h}^2\middle| X_{t-1}, \dots, X_{t-p}\right]$$

Finally:

$$\mathbb{E}\left[X_{t+h}^2\middle| X_{t-1}, \dots, X_{t-p}\right] = \frac{1}{\sigma_{\epsilon,h} + \tilde{\sigma}}\left[\tilde{\sigma}\left(P_h(B)X_{t-1} - \mu_{\epsilon,h}\right)^2 + \sigma_{\epsilon,h}\tilde{\mu}^2\right] + \sigma_{\epsilon,h}\tilde{\sigma}$$

Proof of Proposition 4.2

From Lemma A.1, we know that the process (X_t) verifies the forward recursive equation at horizon $h+1$

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

where P_h and Q_h have respective degrees $p+q-1$ and h. Denoting $Q_h(B) := \sum_{i=0}^{h} q_i B^i$, we have for any $\gamma > 1$,:

$$\begin{aligned}\left|\mathbb{E}\left[|X_{t+h}|^\gamma\right| X_{t-1}, \dots, X_{t-p-q}]\right|^{\frac{1}{\gamma}} &= \left|\mathbb{E}\left[|P_h(B)X_{t-1} + Q_h(B)u_t|^\gamma\right| X_{t-1}, \dots, X_{t-p-q}]\right|^{\frac{1}{\gamma}} \\ &\leq \left|P_h(B)X_{t-1}\right| + \sum_{i=0}^{h}|q_i|\left|\mathbb{E}\left[|u_{t+i}|^\gamma\right| u_{t-1}, \dots, u_{t-p}]\right|^{\frac{1}{\gamma}}\end{aligned}$$

As we have shown in the purely noncausal case (Theorem 3.1), if $-1 < \gamma < 2\alpha + 1$, then $\mathbb{E}\left[|u_{t+h}|^{\gamma}|\, u_{t-1}, \ldots, u_{t-p}\right] < +\infty$ for any $h \geq 0$. Hence for $-1 < \gamma < 2\alpha + 1$, we have

$$\mathbb{E}\left[|X_{t+h}|^{\gamma}|\, X_{t-1}, \ldots, X_{t-p-q}\right] < +\infty$$

Proof of Corollary 4.1

It can be shown that there exist a polynomial $P_0(B)$ of degree $p + q - 1$ such that:

$$X_t = BP_0(B)X_t + u_t. \tag{A.20}$$

Taking the expectation of the previous equation:

$$\mathbb{E}\Big[X_t\Big|X_{t-1}, \ldots, X_{t-p-q}\Big] = BP_0(B)X_t + \mathbb{E}\Big[u_t\Big|u_{t-1}, \ldots, u_{t-p}\Big].$$

Applying the results of Proposition 3.3 to process (u_t), we get on the one hand that:

$$\mathbb{E}\Big[X_t\Big|X_{t-1}, \ldots, X_{t-p-q}\Big] = \Big[BP_0(B) + \Big(1 - \mathscr{P}\alpha(B)\Big)\phi(B)\Big]X_t,$$

which yield the following causal representation

$$X_t = \Big[BP_0(B) + \Big(1 - \mathscr{P}\alpha(B)\Big)\phi(B)\Big]X_t + \eta_t,$$

where (η_t) is a martingale difference sequence. On the other hand however, we have from equation (A.20) that

$$\Big(1 - BP_0(B) - \phi(B)\Big)X_t = 0,$$

Combining both expressions we deduce that the causal representation can be written:

$$\mathscr{P}_{\alpha}(B)\phi(B)X_t = \eta_t.$$

Proof of Proposition 5.1

ι) The causal representation of the noncausal AR(p)

$$X_{t+h} = \frac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha}_{\epsilon,0} + \tilde{\sigma}^{\alpha}} P_0(B)X_{t-1} + \eta_t,$$

admits a unit root if and only if the sum of the coefficients of the polynomial $\dfrac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha}_{\epsilon,0} + \tilde{\sigma}^{\alpha}} P_0(B)$ is equal to one.

$$\frac{\tilde{\sigma}^{\alpha}}{\sigma^{\alpha}|\psi_p|^{-\alpha} + \tilde{\sigma}^{\alpha}} P_0(1) = 1,$$

$$\iff \quad \tilde{\sigma}^{\alpha}(1 - \psi_1 - \ldots - \psi_{p-1})\psi_p^{-1} = \sigma^{\alpha}|\psi_p|^{-\alpha} + \tilde{\sigma}^{\alpha},$$

$$\iff \quad \tilde{\sigma}^{\alpha}(\Psi(1) + \psi_p)\psi_p^{-1} = \sigma^{\alpha}|\psi_p|^{-\alpha} + \tilde{\sigma}^{\alpha},$$

$$\iff \quad |\psi_p|^{\alpha}\tilde{\sigma}^{\alpha} = \psi_p\sigma^{\alpha}\,\Psi(1)^{-1}.$$

Since all the terms involved in the previous equation, except *a priori* ψ_p, are strictly positive (recall that $\Psi(1)^{-1} = \dfrac{1}{(1-\lambda_1)\dots(1-\lambda_p)}$ with $|\lambda_i| < 1$ for any $i = 1, \dots, p$), it implies that ψ_p is itself necessarily strictly positive. Therefore

$$|\psi_p|^\alpha \tilde{\sigma}^\alpha = \psi_p \sigma^\alpha \, \Psi(1)^{-1} \iff |\psi_p|^\alpha \sum_{l=0}^{+\infty} |d_l|^\alpha = \psi_p \sum_{l=0}^{+\infty} d_l,$$

$\iota\iota$) The necessary and sufficient condition of Proposition (5.1) when $\alpha = 1$ boils down to

$$|\psi_p| \sum_{l=0}^{+\infty} |d_l| = \psi_p \sum_{l=0}^{+\infty} d_l.$$

Since we assume $\psi_p > 0$, this yields

$$\sum_{l=0}^{+\infty} (|d_l| - d_l) = 0,$$

and since $|d_l| - d_l \geq 0$ for any l, the conclusion follows.

$\iota\iota\iota$) Given the autoregressive polynomial Ψ, we are looking for values of the tail parameter inducing the presence of unit roots in the characteristic polynomial of the causal representation of process X_t. This is equivalent to studying the zeros of the function:

$$f : a \longmapsto \sum_{l=0}^{+\infty} |\psi_p d_l|^a - \sum_{l=0}^{+\infty} \psi_p d_l.$$

The function f is of class C^∞ on $]0,2[$. Differentiating it twice yields:

$$\forall a \in \left]0,2\right], \quad f''(a) = \sum_{l=0}^{+\infty} \left[\ln(|\psi_p d_l|)\right]^2 |\psi_p d_l|^a > 0.$$

f is thus strictly convex and admits at most two zeros.

$\iota\nu$) Assume $\mathscr{P}(1) = 0$. Then,

$$\frac{\tilde{\sigma}^\alpha}{\tilde{\sigma}^\alpha + \sigma^\alpha |\psi_p|^{-\alpha}} = \frac{\psi_p}{1 - \psi_1 - \dots - \psi_{p-1}},$$

and therefore:

$$\mathscr{P}(B) = 1 - \frac{B^p - \psi_1 B^{p-1} - \dots - \psi_{p-1} B}{1 - \psi_1 - \dots - \psi_{p-1}}.$$

Let $\sum_{j=0}^{p-1} q_j B^j$ be the polynomial such that $\mathscr{P}(B) = (1-B) \sum_{j=0}^{p-1} q_j B^j$. Then

$$1 - \frac{B^p - \psi_1 B^{p-1} - \dots - \psi_{p-1} B}{1 - \psi_1 - \dots - \psi_{p-1}} = q_0 - q_{p-1} B^p + \sum_{j=1}^{p-1} (q_j - q_{j-1}) B^j.$$

By identification, we derive that $q_j = \frac{\sum\limits_{i=0}^{p-j-1} \psi_i}{\sum\limits_{i=0}^{p-1} \psi_i}$ for any $j \in \{0, \ldots, q-1\}$, with $\psi_0 = -1$. Finally, 1 is a root of multiplicity at least 2 if and only if $\sum_{j=0}^{p-1} q_j = 0$. However:

$$\sum_{j=0}^{p-1} q_j = \sum_{j=0}^{p-1} \left[\frac{\sum\limits_{i=0}^{p-j-1} \psi_i}{\sum\limits_{i=0}^{p-1} \psi_i} \right] = \sum_{j=0}^{p-1} \sum_{i=0}^{p-j-1} \psi_i = \sum_{i=0}^{p-1} \psi_i (p-i),$$

and we deduce that 1 is a root of $\mathscr{P}$ of multiplicity at least 2 if and only if $\sum_{i=0}^{p-1} \psi_i (p-i) = 0$. We now show that this is not possible under the assumption that the polynomial $\psi(\cdot)$ has all its root outside the unit circle. Denoting, $\tilde{P}(z) = -1 - \psi_p^{-1} \sum_{i=0}^{p-1} \psi_i z^{p-i}$, we first notice that

$$\tilde{P}'(z) = -\psi_p^{-1} \sum_{i=0}^{p-1} \psi_i \, (p-i) \, z^{p-i-1},$$

and thus $\sum_{i=0}^{p-1} \psi_i (p-i) = 0$ is equivalent to $\tilde{P}'(1) = 0$. Secondly, for any $z \neq 0$,

$$\begin{aligned} \tilde{P}(z) = 0 \iff & \quad -1 - \psi_p^{-1} \left(-z^p + \psi_1 z^{p-1} + \ldots + \psi_{p-1} z \right) = 0, \\ \iff & \quad 1 - \psi_1 z^{-1} - \ldots - \psi_p z^{-p} = 0, \\ \iff & \quad \psi \left(z^{-1} \right) = 0, \end{aligned}$$

and since $\psi \left(z^{-1} \right) = 0$ implies that $|z| < 1$, we have that all the roots of $\tilde{P}$ lie inside the unit circle. Invoking Gauss-Lucas Theorem, we can deduce that the roots of $\tilde{P}'$ lie within the complex hull of the roots of $\tilde{P}$ which is a strict subset of the the unit disc disjoint from the unit circle. Therefore, $\tilde{P}'(1) \neq 0$, which implies that $\sum_{i=0}^{p-1} \psi_i (p-i) \neq 0$ and we conclude that 1 is a simple root of $\mathscr{P}$.

Proof of Proposition 5.3 (Example 5.1) Let (X_t) be the strictly stationary solution of

$$\psi(F) X_t = \varepsilon_t.$$

From Corollary 3.2 we know that (X_t) admits the causal representation

$$\mathscr{P}_\alpha(B) X_t = \eta_t,$$

with

$$\mathscr{P}_\alpha(B) = 1 - \frac{\tilde{\sigma}^\alpha}{\sigma^\alpha |\psi_p|^{-\alpha} + \tilde{\sigma}^\alpha} \psi_p^{-1} \Big(-\psi_{p-1} B - \ldots - \psi_1 B^{p-1} + B^p \Big).$$

Lemma A.3 *For any λ_1, λ_2, λ_3 sufficiently close, but strictly smaller than 1, there exists a tail parameter $\alpha_0 \in]0, 1[$ such that $\mathscr{P}_{\alpha_0}(1) = 0$.*

From Lemma A.3 (see also the Proof of Proposition 5.1.$\iota\nu$) for the derivation of $\Upsilon(B)$), the causal representation can be written

$$(1-B)\Upsilon(B)X_t = \eta_t, \quad \text{with} \quad \mathbb{E}\left[\eta_t | \mathscr{F}_{t-1}\right] = 0.$$

Now we can notice that $Z_t = \Upsilon(B)X_t$. Automatically $\mathbb{E}\left[Z_t | \mathscr{F}_{t-1}\right] = Z_{t-1}$ and process (Z_t) is an $\mathscr{F}_t$-martingale. Furthermore:

Lemma A.4 *For any λ_1, λ_2, λ_3 sufficiently close, but strictly smaller than 1, all the coefficients of the moving average of (Z_t) are positive.*

Therefore, for any λ_1, λ_2, λ_3 sufficiently close, but strictly smaller than 1, there exists a tail parameter $\alpha_0 \in]0,1[$ such that $Z_t = \Upsilon(B)X_t$ is an $\mathscr{F}_t$-martingale. Since we also assume $\beta = 1$, we deduce that (Z_t) is a moving average with positive coefficients of a sequence (ε_t) which is positive almost surely (iid sequence of stable random variables with tail parameter $\alpha_0 < 1$ and asymmetry $\beta = 1$). Hence, (Z_t) is itself positive almost surely.

Proof of Lemma A.3

Recall that the values of α for which $\mathscr{P}_\alpha(1) = 0$ are the zeros of the function

$$f : a \longmapsto \sum_{l=0}^{+\infty} |\psi_p d_l|^a - \sum_{l=0}^{+\infty} \psi_p d_l.$$

Since $\psi_3 = \lambda_1\lambda_2\lambda_3 > 0$ and for any $k \geq 0$, $d_k = \mathrm{C}_{k+2}^2 \xi_k^k > 0$ (see Lemma A.2), we already know from Proposition 5.1 that $\mathscr{P}_1(1) = 0$ and therefore $f(1) = 0$. We will actually exhibit $\lambda_1, \ldots, \lambda_p$, for any integer p odd, such that $f(1) = f(\alpha) = 0$ for some $\alpha \in]0,1[$. For simplicity, we will assume that the (λ_i) are ordered and distinct such that $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_p < 1$. Since function f is strictly convex on $]0,2[$ and $\lim_{a\to 0} f(a) = +\infty$, it is sufficient to find $\lambda_1, \ldots, \lambda_p$ such that $f(2) > 0$ and $f'(1) \neq 0$. On the one hand,

$$\sum_{l=0}^{+\infty} (\psi_p d_l)^2 = \psi_p^2 \sum_{i=1}^{p} a_i \psi(\lambda_i)^{-1},$$

where the coefficients (a_i) are defined at equation (A.18). For any i, we have:

$$\psi_p^2 a_i \psi(\lambda_i)^{-1} = \frac{(\lambda_1 \ldots \lambda_p)^2 \lambda_i^{p-1}}{\prod\limits_{\substack{j=1 \\ j\neq i}}^{p} (\lambda_i - \lambda_j)} \frac{1}{(1-\lambda_i^2) \prod\limits_{\substack{j=1 \\ j\neq i}}^{p} (1-\lambda_i\lambda_j)}.$$

As λ_p tends towards 1, the quantity

$$\psi_p^2 a_p \psi(\lambda_p)^{-1} \sim \frac{(\lambda_1 \dots \lambda_{p-1})^2}{\prod_{j=1}^{p-1}(1-\lambda_j)^2} \frac{1}{(1-\lambda_p^2)},$$

becomes arbitrarily large, whereas for $i < p$, $\psi_p^2 a_i \psi(\lambda_i)^{-1}$ tends to a finite limit. Thus,

$$\psi_p^2 \sum_{i=1}^{p} a_i \psi(\lambda_i)^{-1} \underset{\lambda_p \to 1}{\sim} \frac{(\lambda_1 \dots \lambda_{p-1})^2}{\prod_{j=1}^{p-1}(1-\lambda_j)^2} \frac{1}{(1-\lambda_p^2)}.$$

On the other hand,

$$\sum_{l=0}^{+\infty} \psi_p d_l = \frac{\psi_p}{(1-\lambda_p)\prod_{j=1}^{p-1}(1-\lambda_j)} \underset{\lambda_p \to 1}{\sim} \frac{\lambda_1 \dots \lambda_{p-1}}{(1-\lambda_p)\prod_{j=1}^{p-1}(1-\lambda_j)}.$$

Hence:

$$\frac{\sum_{l=0}^{+\infty}(\psi_p d_l)^2}{\sum_{l=0}^{+\infty}\psi_p d_l} \underset{\lambda_p \to 1}{\sim} \frac{1}{2}\prod_{j=1}^{p-1}\frac{\lambda_j}{1-\lambda_j},$$

which is greater than 1 as soon as $p \geq 2$ and $\lambda_j > \frac{2}{3}$ for all $j < p$. Therefore, for any odd integer $p \geq 3$, there exists $\bar{\lambda} \in (0,1)$ such that for any $\lambda_p \in \left[\bar{\lambda}, 1\right)$, and roots $\lambda_1, \dots, \lambda_{p-1}$ in $\left(\frac{2}{3}, 1\right)$ such that $f(2) > 0$. Since all roots are positive and $\psi_p > 0$, we know that $f(1) = 0$. Provided that $1 \neq \arg\min_{a \in (0,2]} f(a)$, we know from the previous discussion on the shape of the function f that there exists another $\alpha \in (0,2)$, $\alpha \neq 1$, such that the causal representation of X_t admits a unit root. Let us notice that

$$f'(1) = \sum_{l=0}^{+\infty} \ln(\psi_p d_l)\psi_p d_l \geq g(\lambda_1) := \sum_{l=0}^{+\infty} \ln(\psi_p \mathrm{C}_{l+p-1}^{p-1}\lambda_1^l)\psi_p \mathrm{C}_{l+p-1}^{p-1}\lambda_1^l \underset{\lambda_1 \to 1}{\longrightarrow} +\infty,$$

Since the function g defined above is continuous on $(0,1)$, there exists a threshold $\underline{\lambda} \in (0,1)$ such that for any $\lambda_1 \geq \underline{\lambda}$, $f'(1) \geq g(\lambda_1) > 0$, which guarantees that $f(1) \neq \min_{a \in (0,2]} f(a)$. The fact that $f'(1) > 0$ yields that there exists $\alpha_0 \in]0,1[$ such that $f(\alpha_0) = 0$..

Proof of Lemma A.4

Let (Z_t) be the linear process solution of

$$\psi(F)Z_t = \Upsilon(B)\varepsilon_t,$$

where $\Upsilon(B) = 1 + \gamma_1 B + \gamma_2 B^2$, with $\gamma_1 = \frac{1-\psi_1}{1-\psi_1-\psi_2}$ and $\gamma_2 = \frac{1}{1-\psi_1-\psi_2}$. It admits the following moving average representation

$$Z_t = \gamma_2 \varepsilon_{t-2} + (\gamma_1 + \gamma_2 d_1)\varepsilon_{t-1} + \sum_{k=0}^{+\infty}(d_k + \gamma_1 d_{k+1} + \gamma_2 d_{k+2})\varepsilon_t.$$

The coefficients of the MA representation of (Z_t) are positive if and only if

$$\begin{aligned} \gamma_2 &> 0, \\ \gamma_1 + \gamma_2 d_1 &> 0, \\ \forall k \geq 0, \quad d_k + \gamma_1 d_{k+1} + \gamma_2 d_{k+2} &> 0. \end{aligned}$$

On the one hand, using the result of Lemma A.2, we know that for any $k \geq 0$, there exists $\xi_k \in]\lambda_1, \lambda_3[$ such that $\frac{(k+1)(k+2)}{2}\lambda_1^k \leq d_k = \frac{(k+1)(k+2)}{2}\xi_k^k \leq \frac{(k+1)(k+2)}{2}\lambda_3^k$. Since we assume $0 < \lambda_1 < \lambda_2 < \lambda_3 < 1$, we have that $\lambda_2 \to 1$ and $\lambda_3 \to 1$ as $\lambda_1 \to 1$ and thus, $d_k \to \frac{(k+1)(k+2)}{2}$. On the other hand, it can be shown that $\gamma_1 \to -2$ and $\gamma_2 \to 1$ as $\lambda_1 \to 1$. Therefore, the first inequality is automatically verified for roots close enough to 1, whereas:

$$\begin{aligned} &\gamma_1 + \gamma_2 d_1 \to 1 \\ \forall k \geq 0, \quad &d_k + \gamma_1 d_{k+1} + \gamma_2 d_{k+2} \to \frac{(k+1)(k+2) - 2(k+2)(k+3) + (k+3)(k+4)}{2} = 1 \end{aligned}$$

Hence, all inequalities are verified for (inverses of the) roots close enough to unity.

A.2 For the Propositions of Section 6

Proof of Proposition 6.1

We will denote $\hat{\boldsymbol{R}} := \left(\hat{\rho}(|i-j|)\right)_{\left\{\substack{-q\leq i,j\leq p \\ i,j\neq 0}\right\}}$, $\hat{\boldsymbol{\rho}} := {}^{\mathrm{t}}(\hat{\rho}(q), \ldots, \hat{\rho}(1), \hat{\rho}(1), \ldots, \hat{\rho}(p))$, $\hat{\boldsymbol{r}} := {}^{\mathrm{t}}(\hat{\rho}(1), \ldots, \hat{\rho}(p+q-1))$ and $\boldsymbol{r} := {}^{\mathrm{t}}(\rho(1), \ldots, \rho(p+q))$. We have from Davis and Resnick (1986) Theorem 4.4 (see also Davis and Resnick (1985)) that

$$\hat{\boldsymbol{R}} \xrightarrow{P} \boldsymbol{R}, \tag{A.21}$$

$$\tilde{a}_n^{-1} a_n^2 (\hat{\boldsymbol{r}} - \boldsymbol{r}) \Longrightarrow \boldsymbol{\delta}. \tag{A.22}$$

Noticing that

$$\begin{aligned} ({}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{X}) &= \left(\sum_{t=0}^{n} X_{t+i}X_{t+j}\right)_{\left\{\substack{-q\leq i,j\leq p \\ i,j\neq 0}\right\}}, \\ {}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{Y} &= {}^{\mathrm{t}}\left(\sum_{t=0}^{n} X_t X_{t-q}, \ldots, \sum_{t=0}^{n} X_t X_{t-1}, \sum_{t=0}^{n} X_t X_{t+1}, \ldots, \sum_{t=0}^{n} X_t X_{t+p}\right), \end{aligned}$$

thus

$$\hat{\boldsymbol{\theta}} = ({}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{X})^{-1}\, {}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{Y} = \left(\frac{1}{\sum_{t=0}^{n} X_t^2}\, {}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{X}\right)^{-1} \frac{1}{\sum_{t=0}^{n} X_t^2}\, {}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{Y} = \hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\rho}}.$$

We then have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{R}^{-1}\boldsymbol{\rho} = \hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\rho}} - \boldsymbol{R}^{-1}\boldsymbol{\rho},$$
$$\hat{\boldsymbol{\theta}} - \boldsymbol{R}^{-1}\boldsymbol{\rho} = \hat{\boldsymbol{R}}^{-1}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) + \left(\hat{\boldsymbol{R}}^{-1} - \boldsymbol{R}^{-1}\right)\boldsymbol{\rho}. \tag{A.23}$$

To derive the asymptotic behaviour of the LFS term, let us express it as a function of $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}$.

$$\begin{aligned}\left(\hat{\boldsymbol{R}}^{-1} - \boldsymbol{R}^{-1}\right)\boldsymbol{\rho} &= \operatorname{vec}\left(\left(\hat{\boldsymbol{R}}^{-1} - \boldsymbol{R}^{-1}\right)\boldsymbol{\rho}\right),\\ &= -\operatorname{vec}\left(\hat{\boldsymbol{R}}^{-1}\left(\hat{\boldsymbol{R}} - \boldsymbol{R}\right)\boldsymbol{R}^{-1}\boldsymbol{\rho}\right),\\ &= -{}^{\mathrm{t}}\left(\boldsymbol{R}^{-1}\boldsymbol{\rho}\right)\otimes\hat{\boldsymbol{R}}^{-1}\operatorname{vec}\left(\hat{\boldsymbol{R}} - \boldsymbol{R}\right).\end{aligned} \tag{A.24}$$

We can notice that

$$\begin{aligned}\hat{\boldsymbol{R}} - \boldsymbol{R} = \Big(\tilde{\rho}(|i-j|) := \hat{\rho}(|i-j|) - \rho(|i-j|)\Big)_{\substack{-q\le i,j\le p\\ i,j\neq 0}} &= \sum_{l=-(p+q)}^{p+q}\tilde{\rho}(|l|)D_l,\\ &= \sum_{l=1}^{p+q}\tilde{\rho}(l)\left(D_{-l} + D_l\right),\end{aligned}$$

where for any $l \geq 0$ (resp. $l \leq 0$), D_l is the square matrix of size $p+q$ with ones on the lower (resp. upper) l^{th} diagonal and zeros outside. Hence, by denoting

$$\mathcal{J}_0 = \Big(\operatorname{vec}\left(D_1 + D_{-1}\right)\Big|\cdots\Big|\operatorname{vec}\left(D_{p+q} + D_{-(p+q)}\right)\Big),$$

we obtain that

$$\operatorname{vec}\left(\hat{\boldsymbol{R}} - \boldsymbol{R}\right) = \mathcal{J}_0\left(\hat{\boldsymbol{r}} - \boldsymbol{r}\right). \tag{A.25}$$

Let us introduce a matrix $\mathcal{J}_1$ such that $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} = \mathcal{J}_1\left(\hat{\boldsymbol{r}} - \boldsymbol{r}\right)$, for instance

$$\mathcal{J}_1 := \begin{pmatrix} I_q^A & 0_{q,p}\\ I_p & 0_{p,q}\end{pmatrix},$$

is appropriate, where I_q^A denotes the square matrix of size q with an antidiagonal of ones and zeros elsewhere. Then, substituting this last equality in (A.24) and then injecting in (A.23), we obtain:

$$\begin{aligned}\hat{\boldsymbol{\theta}} - \boldsymbol{R}^{-1}\boldsymbol{\rho} &= \hat{\boldsymbol{R}}^{-1}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) - \left\{{}^{\mathrm{t}}\left(\boldsymbol{R}^{-1}\boldsymbol{\rho}\right)\otimes\hat{\boldsymbol{R}}^{-1}\right\}\operatorname{vec}\left(\hat{\boldsymbol{R}} - \boldsymbol{R}\right)\\ &= \hat{\boldsymbol{R}}^{-1}\mathcal{J}_1\left(\hat{\boldsymbol{r}} - \boldsymbol{r}\right) - \left\{{}^{\mathrm{t}}\left(\boldsymbol{R}^{-1}\boldsymbol{\rho}\right)\otimes\hat{\boldsymbol{R}}^{-1}\right\}\mathcal{J}_0(\hat{\boldsymbol{r}} - \boldsymbol{r})\\ &= \left(\hat{\boldsymbol{R}}^{-1}\mathcal{J}_1 - \left\{{}^{\mathrm{t}}\left(\boldsymbol{R}^{-1}\boldsymbol{\rho}\right)\otimes\hat{\boldsymbol{R}}^{-1}\right\}\mathcal{J}_0\right)\left(\hat{\boldsymbol{r}} - \boldsymbol{r}\right)\end{aligned}$$

Finally, we conclude by invoking (A.21), (A.22) and Slutsky's theorem.

Proof of Proposition 6.2 First of all, let us notice that minimising

$$\sum_{t=2}^{n-1} \left[(1-\psi F)(1-\phi B)X_t\right]^2,$$

is equivalent to minimising

$$\frac{\sum_{t=2}^{n-1} \left[(1-\psi F)(1-\phi B)X_t\right]^2}{\sum_{t=2}^{n-1} X_t^2}.$$

hence in this proof, we will denote $\mathcal{L}_n(\phi,\psi) := \frac{\sum_{t=2}^{n-1} \left[(1-\psi F)(1-\phi B)X_t\right]^2}{\sum_{t=2}^{n-1} X_t^2}$. Let us also denote for any $h \geq 0$

$$\gamma(h) = \sum_{k\in\mathbb{Z}} d_i d_{i+h}, \qquad \hat{\gamma}(h) = \sum_{t=2}^{n-1} X_t X_{t+h}.$$

Derivating equation (6.7) with respect to ψ and ϕ, we get

$$\begin{aligned}
\frac{\partial \mathcal{L}_n}{\partial \phi}(\phi,\psi) &= 2\left[(\psi(1+\phi\psi)+\phi)\,\hat{\gamma}(0) - \left(1+2\phi\psi+\psi^2\right)\hat{\gamma}(1) + \psi\hat{\gamma}(2)\right],\\
\frac{\partial \mathcal{L}_n}{\partial \psi}(\phi,\psi) &= 2\left[(\phi(1+\phi\psi)+\psi)\,\hat{\gamma}(0) - \left(1+2\phi\psi+\phi^2\right)\hat{\gamma}(1) + \phi\hat{\gamma}(2)\right],
\end{aligned}$$

which yield the first order conditions

$$(\psi(1+\phi\psi)+\phi) - \left(1+2\phi\psi+\psi^2\right)\hat{\rho}(1) + \psi\hat{\rho}(2) = 0, \tag{A.26}$$

$$(\phi(1+\phi\psi)+\psi) - \left(1+2\phi\psi+\phi^2\right)\hat{\rho}(1) + \phi\hat{\rho}(2) = 0. \tag{A.27}$$

Taking the difference of these two equations on the one hand, we obtain:

$$(\psi-\phi)\left[\phi\psi - (\phi+\psi)\hat{\rho}(1) + \hat{\rho}(2)\right] = 0.$$

Similarily, multiplying the first equation by ϕ, the second by ψ and taking the difference:

$$(\psi-\phi)\left[-(\phi+\psi) + (1+\phi\psi)\hat{\rho}(1)\right] = 0.$$

Therefore, the local optima necessarily verify

$$\phi = \psi, \quad \text{or} \quad \begin{cases} \phi+\psi &= \dfrac{\hat{\rho}(1)(1-\hat{\rho}(2))}{1-\hat{\rho}^2(1)},\\ \phi\psi &= \dfrac{\hat{\rho}^2(1)-\hat{\rho}(2)}{1-\hat{\rho}^2(1)} \end{cases}$$

As n tends to infinity, using a result from Davis and Resnick (1986) and the formulae of Lemma A.6 we get

$$\begin{aligned}
\frac{\hat{\rho}(1)(1-\hat{\rho}(2))}{1-\hat{\rho}^2(1)} &\xrightarrow{p} \frac{\rho(1)(1-\rho(2))}{1-\rho^2(1)} = \phi_0+\psi_0,\\
\frac{\hat{\rho}^2(1)-\hat{\rho}(2)}{1-\hat{\rho}^2(1)} &\xrightarrow{p} \frac{\rho^2(1)-\rho(2)}{1-\rho^2(1)} = \phi_0\psi_0,
\end{aligned}$$

When $\phi = \psi$, the two first order conditions (A.26) collapse to a single third degree equation:

$$g_n(\psi) := \psi^3 - 3\hat{\rho}(1)\psi^2 + \Big(2 + \hat{\rho}(2)\Big)\psi - \hat{\rho}(1) = 0 \tag{A.28}$$

The derivative of g_n reads:

$$g_n'(\psi) = 3\psi^2 - 6\hat{\rho}(1)\psi + 2 + \hat{\rho}(2),$$

which is always positive provided that n is great enough (the discriminant tends to $3(3\rho^2(1) - \rho(2) - 2)$ which can be shown to be strictly negative using formulae (A.35) and (A.37)). Hence, g_n is stricly increasing and (A.28) has a unique real root ψ_c from which we deduce that (ψ_c, ψ_c) is a potential local extremum of $\mathcal{L}_n$. We now show that it is a local minimum if and only if $\phi_0 = \psi_0$. To do so, we will show that the Hessian of $\mathcal{L}_n$ at this point has a negative eigenvalue whenever $\phi_0 \neq \psi_0$. A direct computation yields the Hessian of $\mathcal{L}_n$ at any point (ψ, ψ) of $\mathbb{R}^2$:

$$H_n(\psi) := 2\hat{\gamma}(0) \begin{pmatrix} 1 - 2\hat{\rho}(1) + \psi^2 & 1 + \hat{\rho}(2) - 4\hat{\rho}(1)\psi + 2\psi^2 \\ 1 + \hat{\rho}(2) - 4\hat{\rho}(1)\psi + 2\psi^2 & 1 - 2\hat{\rho}(1) + \psi^2 \end{pmatrix} \xrightarrow{p} H(\psi),$$

where

$$H(\psi) := 2\gamma(0) \begin{pmatrix} 1 - 2\rho(1) + \psi^2 & 1 + \rho(2) - 4\rho(1)\psi + 2\psi^2 \\ 1 + \rho(2) - 4\rho(1)\psi + 2\psi^2 & 1 - 2\rho(1) + \psi^2 \end{pmatrix}$$

We reach the eigenvalues using the trace and the determinant. Denote $\lambda_1(\psi_c)$ and $\lambda_2(\psi_c)$ the two real eigenvalues of $H(\psi_c)$, which always exist since H is symmetric. Using the result of Lemma A.5, we deduce that $\lambda_1(\psi_c) + \lambda_2(\psi_c) > 0$ for any $(\phi_0, \psi_0) \in]0, 1[^2$ and $\lambda_1(\psi_c)\lambda_2(\psi_c) < 0$ whenever $\phi_0 \neq \psi_0$, implying in this case that $H(\psi_c)$ has exactly one strictly negative and one strictly positive root. If $\phi_0 = \psi_0$, then $H(\psi_c)$ has one strictly positive root and another root equal to zero. In this latter case, $H(\psi_c)$ is positive semi-definite.

Lemma A.5 *For any* $(\phi_0, \psi_0) \in]0, 1[^2$,

$$\operatorname{Tr} H(\psi_c) > 0,$$
$$\det H(\psi_c) \leq 0, \quad \textit{and} \quad \det H(\psi_c) = 0 \quad \textit{if and only if} \quad \phi_0 = \psi_0.$$

Proof.

We have

$$\lambda_1(\psi) + \lambda_2(\psi) = 2(\psi^2 - 2\rho(1)\psi + 1), \tag{A.29}$$

$$\lambda_1(\psi)\lambda_2(\psi) = -4\gamma^2(0)\Big(\psi^2 - 2\rho(1)\psi + \rho(2)\Big)\Big(3\psi^2 - 6\rho(1)\psi + \rho(2) + 2\Big). \tag{A.30}$$

Using (A.35), equation (A.29) implies that for any $\psi \in \mathbb{R}$, $\lambda_1(\psi)+\lambda_2(\psi) > 0$ and hence, at least one of the two eigenvalues is positive. Let us examine the sign of $\lambda_1\lambda_2(\cdot)$. From equation (A.30), it can be seen that the sign of the product of eigenvalues depends on two second order polynomials. The second polynomial we already encountered and can be proven to be positive for any $(\phi_0, \psi_0) \in]0,1[^2$ and $(\phi, \psi) \in \mathbb{R}^2$ using formulae (A.35) and (A.37). The discriminant of the first polynomial reads

$$\Delta = \rho(1)^2 - \rho(2),$$

which is negative whenever $\phi_0\psi_0 < 0$ according to formula (A.36). In this case, we can deduce that $\lambda_1(\psi_c)\lambda_2(\psi_c) < 0$ and thus (ψ_c, ψ_c) is a saddle point of $\mathcal{L}_n$. However, when $\phi_0\psi_0 > 0$, the equation

$$\psi^2 - 2\rho(1)\psi + \rho(2) = 0,$$

has two real roots, $\rho(1) - \sqrt{\rho^2(1) - \rho(2)}$ and $\rho(1) + \sqrt{\rho^2(1) - \rho(2)}$. The last point to show is that ψ_c lies outside the segment $\left[\rho(1) - \sqrt{\rho^2(1) - \rho(2)}, \rho(1) + \sqrt{\rho^2(1) - \rho(2)}\right]$ whenever $\phi_0 \neq \psi_0$ and is on the boundary otherwise. Since $g(\psi) = \psi^3 - 3\rho(1)\psi^2 + \big(2+\rho(2)\big)\psi - \rho(1)$ is strictly increasing, it is equivalent to show that $g\left(\rho(1) - \sqrt{\rho^2(1) - \rho(2)}\right) > 0$ or $g\left(\rho(1) + \sqrt{\rho^2(1) - \rho(2)}\right) < 0$. A direct computation yields

$$g\left(\rho(1) - \sqrt{\rho^2(1) - \rho(2)}\right) > 0 \iff \sqrt{\rho^2(1) - \rho(2)} < \frac{\rho(1)}{2}\left(\frac{1 - 2\rho^2(1) + \rho(2)}{1 - \rho^2(1)}\right), \tag{A.31}$$

$$g\left(\rho(1) + \sqrt{\rho^2(1) - \rho(2)}\right) > 0 \iff \sqrt{\rho^2(1) - \rho(2)} < -\frac{\rho(1)}{2}\left(\frac{1 - 2\rho^2(1) + \rho(2)}{1 - \rho^2(1)}\right). \tag{A.32}$$

The cases $\rho(1) > 0$ and $\rho(1) < 0$ are symmetric. Let us treat the first one. Then $\frac{\rho(1)}{2}\left(\frac{1 - 2\rho^2(1) + \rho(2)}{1 - \rho^2(1)}\right) > 0$. Taking the square of the right-hand side inequality of (A.31) and using the formulae of Lemma A.6 it follows that

$$\psi_c < \rho(1) - \sqrt{\rho^2(1) - \rho(2)} \iff 4 < \frac{(1 - \phi_0\psi_0)^2(\phi_0 + \psi_0)^2}{\phi_0\psi_0(1 - \phi_0^2)(1 - \psi_0^2)}.$$

Using the classical inequality $a^2 + b^2 \geq 2ab$, we get for any $(x, y) \in]0,1[^2$:

$$\frac{(1 - xy)^2(x + y)^2}{xy(1 - x^2)(1 - y^2)} = \left(2 + \frac{x^2 + y^2}{xy}\right)\frac{(1 - xy)^2}{(1 - x^2)(1 - y^2)} \geq 4.$$

Therefore, $\psi_c \leq \rho(1) - \sqrt{\rho^2(1) - \rho(2)}$ and $\psi_c = \rho(1) - \sqrt{\rho^2(1) - \rho(2)}$ if and only if $\phi_0 = \psi_0$. We show by a similar reasoning that in the case $\rho(1) < 0$, $\psi_c \geq \rho(1) + \sqrt{\rho^2(1) - \rho(2)}$ and $\psi_c = \rho(1) + \sqrt{\rho^2(1) - \rho(2)}$ if and only if $\phi_0 = \psi_0$.

Thus, the critical point (ψ_c, ψ_c) is a saddle point of $\mathcal{L}_n$ whenever $\phi_0 \neq \psi_0$ and a local minimum if and if only if $\phi_0 = \psi_0$.

Lemma A.6 *For any* $(\phi_0, \psi_0) \in]0, 1[^2$*:*

$$\rho(1) = \frac{\phi_0 + \psi_0}{1 + \phi_0\psi_0}, \tag{A.33}$$

$$\rho(2) = 1 - \frac{(1 - \phi_0^2)(1 - \psi_0^2)}{1 + \phi_0\psi_0}, \tag{A.34}$$

$$1 - \rho(1)^2 = \frac{(1 - \psi_0^2)(1 - \phi_0^2)}{(1 + \psi_0\phi_0)^2}, \tag{A.35}$$

$$\rho(1)^2 - \rho(2) = \frac{\psi_0\phi_0(1 - \psi_0^2)(1 - \phi_0^2)}{(1 + \psi_0\phi_0)^2}, \tag{A.36}$$

$$1 - 2\rho(1)^2 + \rho(2) = \frac{(1 - \psi_0\phi_0)(1 - \psi_0^2)(1 - \phi_0^2)}{(1 + \psi_0\phi_0)^2}, \tag{A.37}$$

References

Abramowitz, M. and I. A. Stegun (1964): Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Courier Corporation, Vol. 55.

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

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

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

Camacho, F., McLeod, A. I., and K. W., Hipel (1987): Multivariate contemporaneous ARMA model with hydrological applications. Stochastic Hydrology and Hydraulics, 1(2), 141-154.

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

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

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

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

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

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

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

Gouriéroux, C. and J.-M., Zakoïan (2016): Local explosion modelling by non-causal process. Journal of the Royal Statistical Society: Series B (Statistical Methodology). doi: 10.1111/rssb.12193

Hecq, A. W., Lieb, L., and S. M., Telg (2015): Identification of Mixed Causal-Noncausal Models: How Fat Should We Go?. Maastricht University School of Business and Economics, GSBE.

Krantz, S. G. (2012): Optimal Handbook of complex variables. Springer Science & Business Media.

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

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

Ling, S. (2007): A double AR (p) model: structure and estimation. Statistica Sinica, 17(1), 161.

MacLeod, Ch L., et al. (2010): Modeling the time variability of SDSS Stripe 82 quasars as a damped random walk. The Astrophysical Journal 721.2 1014.

Mohammadi, K., Eslami, H. and Kahawita, V. (2006): Parameter Estimation of an ARMA Model for River Flow Forecasting Using Goal Programming. Journal of Hydrology, 331, 293-299.

Nielsen, H. B., and A., Rahbek (2014): Unit root vector autoregression with volatility induced stationarity. Journal of Empirical Finance, 29, 144-167.

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

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

Vaughan, S. (2013): Random time series in astronomy. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 371(1984), 20110549.