Forecasting bubbles with mixed causal-noncausal autoregressive models

Voisin, Elisa and Hecq, Alain

Maastricht University

13 March 2019

Forecasting bubbles with mixed causal-noncausal autoregressive models

Alain Hecq and Elisa Voisin[1]
Maastricht University

February, 2019

Abstract

This paper investigates one-step ahead density forecasts of mixed causal-noncausal models. We compare the sample-based and the simulations-based approaches respectively developed by Gouriéroux and Jasiak (2016) and Lanne, Luoto, and Saikkonen (2012). We focus on explosive episodes and therefore on predicting turning points of bubbles bursts. We suggest the use of both methods to construct investment strategies based on how much probabilities are induced by the assumed model and by past behaviours. We illustrate our analysis on Nickel prices series.

Keywords: noncausal models, forecasting, predictive densities, bubbles, simulations-based forecasts.
JEL. C22, C53

[1]Corresponding author : Elisa Voisin, Maastricht University, Department of Quantitative Economics, School of Business and Economics, P.O.box 616, 6200 MD, Maastricht, The Netherlands. Email: e.voisin@maastrichtuniversity.nl.
The authors would like to thank Sean Telg , Sebastien Fries, Eric Beutner as well as the participants of the CFE 2018, Pisa. All remaining errors are ours.

1 Introduction

Locally explosive episodes have long been observed in financial and economic time series. Such patterns, often observed in stock prices, can be triggered by anticipation or speculation. Given this forward-looking aspect, expectation models have been prevalent for modelling them. As shown for instance by Gouriéroux, Jasiak, and Monfort (2016), equilibrium rational expectation models admit a multiplicity of solutions, and some of them feature such speculative bubble patterns.[2] Models employed to capture them range from simplistic approaches, such as single bubble models with constant probability of crash, to rather complex models depending on numerous parameters. Although those models may a posteriori fit the data well, they are either not informative enough or render predictions uncertain due to their dependence on extensive parameters estimation.

This paper analyses different approaches to perform point and density forecasts from mixed causal-noncausal autoregressive (hereafter MAR) models. MAR models incorporate both lags and leads of the variable of interest with potentially heavy-tailed errors. The most commonly used distributions for such models in the literature are the Cauchy and Student's t-distributions. In spite of their simplicity, MAR models generate non-linear dynamics such as locally explosive episodes in a strictly stationary setting (Fries and Zakoïan, 2019). So far, the focus has mainly been put on identification and estimation. Hecq, Lieb, and Telg (2016), Hencic and Gouriéroux (2015) and Lanne et al. (2012) detect a noncausal component explaining respectively the observed bubbles in the demand of solar panels in Belgium, in Bitcoin and inflation series. Few papers look at the forecasting aspects. Gouriéroux and Zakoïan (2017) demonstrate that the causal conditional distribution possesses more moments than the marginal distribution and suggest this as a cornerstone for forecasting with such models. Some distributions however, like Student's t, do not admit closed-form expressions for conditional moments and distribution. Gouriéroux and Jasiak (2016) or Lanne et al. (2012) developed estimators to approximate them based on past realised values or on simulations respectively. Nonetheless, the literature regarding the ability of MAR models to predict both explosive and stable

[2] In this paper, speculative bubbles, or simply bubbles are referred to as processes characterised by a rapid and persistent increase followed by a crash. Some authors talk about bubbles to denote the deviation from the fundamental solution of a present value type model. Those bubbles might not have the non-linear pattern that we investigate in this paper.

episodes remains scarce (see also Gouriéroux, Hencic, and Jasiak, 2018). The aim of this paper is to analyse and compare in details methods available for forecasting $MAR(r,1)$ models, with unconstrained r number of lags, a unique lead and a positive lead coefficient. Furthermore, the focus is put on positive bubbles since they are prevalent in financial and economic time series. This paper investigates the possibility to predict the turning point of locally explosive episodes. Both statistical and numerical approaches are employed, and in order to rigorously compare them, this paper aims attention at one-step ahead predictions. We find that combining results obtained from two different approaches can help to disentangle how much of the probability of an event is induced by the underlying distribution and by past behaviours of the series. This information could be used for investment strategies, to optimise when leaving an investment before the bubble crash.

The paper is constructed as follows. Section 2 introduces mixed causal-noncausal autoregressive models. Section 3 discusses how they have been used for prediction so far when the conditional moments and densities admit closed-form expressions. In Section 4 are presented the numerical sample-based forecasting approach proposed by Gouriéroux and Jasiak (2016), followed by the simulations-based method proposed by Lanne et al. (2012). The performance of both approaches is compared to closed-form results of an $MAR(0,1)$ processes with a lead coefficient of 0.8 and Cauchy-distributed errors. In Section 5 both approximation methods are illustrated using a detrended Nickel prices series. Section 6 concludes.

2 Mixed causal-noncausal autoregressive models

Consider the univariate $MAR(r,s)$ process defined as follows,

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,$$

where L and L^{-1} are respectively the lag and forward operators; Φ and Ψ are two invertible polynomials of degree r and s respectively. That is, $\Phi(L) = (1-\phi_1 L \cdots - \phi_r L^r)$ and $\Psi(L^{-1}) = (1-\psi_1 L^{-1} \cdots - \psi_s L^{-s})$ with roots strictly outside the unit circle and such that $\Phi(0) = \Psi(0) = 1$. The error term ε_t is i.i.d, following a non-Gaussian distribution. This assumption, not empirically restrictive since non-normality is widely observed in financial and economic time series, is necessary to achieve identification of the model. An $MAR(r,s)$ model can also be expressed as a causal AR model where y_t

depends on its own past and present value of u_t,

$$\Phi(L)y_t = u_t, \tag{1}$$

where u_t is the purely noncausal component of the errors, depending on its own future and on the present value of the error term

$$\Psi(L^{-1})u_t = \varepsilon_t. \tag{2}$$

From this notation it is evident that predicting the variable of interest y requires prediction of its noncausal component u. This is why most prediction methods focus on forecasting purely noncausal processes. Alternatively, we can also filter the process as $\Phi(L)v_t = \varepsilon_t$ with $\Psi(L^{-1})y_t = v_t$ to obtain the backward component of the errors, v_t. The process y_t admits a stationary infinite two-sided MA representation and depends on past, present and future values of ε_t,

$$y_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}.$$

The case in which all coefficients a_i for $-\infty < i \leq 0$ (resp. $0 \leq i < \infty$) are equal to 0, corresponds to a purely causal (resp. noncausal) model.

Despite their apparent simplicity and parsimony, *MAR* models often provide a better fit to economic and financial data as they capture non-linear causal dynamics such as bubbles or asymmetric cycles. The shape of series generated by *MAR*(r,s) processes depends on the presence of leads, lags and the magnitude of their coefficients. Figure 1 displays how the presence of a lag, a lead, or both, affects the shape of *MAR* series. Purely causal (resp. noncausal) processes are only affected by a shock after (resp. before) the impact; this is shown in graph (a) (resp. (b)). Consequently, *MAR* processes are affected both in anticipation and after the shock; the shape of the explosive episode (mostly forward or backward looking) depends on the magnitude of the lag and lead coefficients. When the coefficients are identical (c) the effects of the shock are symmetric around the impact while when the coefficient of the lead is higher (d), the explosive episode is more analogous to what we refer to as a bubble with an asymmetry around the peak.

The usual practice for estimating and identifying *MAR* models is as follows. Methods based on first and second moments (e.g. OLS) are unable to distinguish between purely causal, noncausal or mixed processes as their

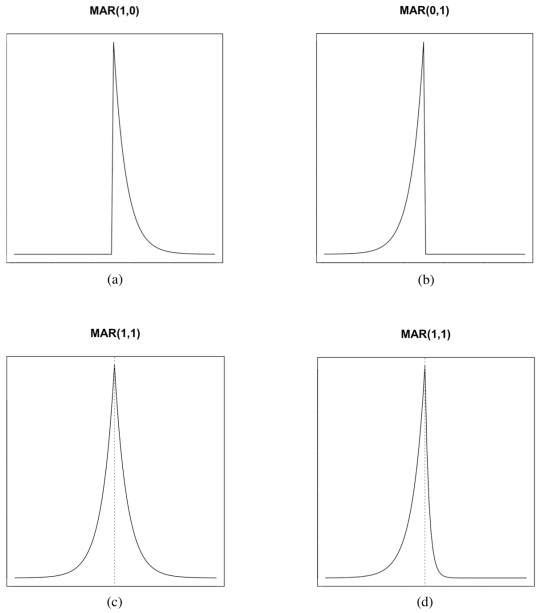

Figure 1: Effects of lags and leads on a $MAR(1,1)$ series (a) $\phi = 0.8$ and $\psi = 0$, (b) $\phi = 0$ and $\psi = 0.8$, (c) $\phi = 0.8$ and $\psi = 0.8$, (d) $\phi = 0.3$ and $\psi = 0.8$

autocovariance functions are identical. Fitting an autoregressive model by OLS allows however to estimate the sum of leads and lags, p.[3] Subsequently, the respective numbers of lags (r) and leads (s), such that $r + s = p$, can be estimated by an approximate maximum likelihood (hereafter AML) approach (Lanne and Saikkonen, 2011). The selected model is the one maximising the AML with respect to r, s and all parameters $\Omega = (\Phi, \Psi, \Theta)$, where $\Phi = (\phi_1, \ldots, \phi_r)$, $\Psi = (\psi_1, \ldots, \psi_s)$ and Θ is the errors distribution parameters, such as the scale or location for instance. The AML estimator is defined as follows,

$$\left(\hat{\Phi}, \hat{\Psi}, \hat{\Theta}\right) = argmax_{\Phi,\Psi,\Theta} \sum_{t=r+1}^{T-s} \ln\left[g\Big(\Phi(L)\Psi(L^{-1})y_t; \Theta\Big)\right],$$

where g denotes the *pdf* of the error term, satisfying the regularity conditions (Andrews, Davis, and Breidt, 2006). Lanne and Saikkonen (2011)

[3] A non-Gaussian MLE can give misleading results in a misspecified model (Gouriéroux and Jasiak, 2018).

show that the resulting (local) maximum likelihood estimator is consistent, asymptotically normal and that $(\hat{\Psi}, \hat{\Phi})$ and $\hat{\Theta}$ are asymptotically independent. Since analytic solution of the maximisation problem at hand is not directly available, numerical gradient-based procedures can be employed. Hecq et al. (2016) indicate that in theory, estimating *MAR* models is easier for more volatile series since the convergence of the estimator is faster for distributions with fatter tails. They propose an alternative way to obtain the standard errors, a method implemented in the R package MARX.

While with a unique lead and a positive coefficient ψ, explosive episodes increase at a fixed rate ψ^{-1} until the crash, other specifications induce complex patterns not resembling the bubble pattern that this paper focuses on. As shown in Figure 2, a negative coefficient (upper graphs) creates increasing oscillations around zero until the crash and multiple leads (bottom graphs) create oscillations along the explosion. Because the presence of multiple leads renders derivations rather intricate, this paper focuses on $MAR(r,1)$ processes with a positive lead coefficient. Except for the empirical analysis of Section 5, the data generating process will be assumed correctly identified throughout the paper to disregard estimation uncertainty.

3 Predictions using closed-form expressions

When it comes to forecasting *MAR* models, different approaches are available. One can predict the next points of the series based on conditional expectations. Alternatively, one can forecast densities, with for instance the aim to visually analyse probabilities of potential future paths or to predict probabilities of turning point in an explosive episode. However, the anticipative aspect of *MAR* models complicates their use for predictions. Results are not as straightforward as they could be with purely backward-looking *ARMA* models. While in some cases mean or density forecasts can be directly obtained from the assumed errors distribution, they sometimes need to be approximated. For this section, let us assume that the data generating process is an $MAR(r,1)$ process $\Phi(L)(1-\psi L^{-1})y_t = \varepsilon_t$, where $\psi > 0$, ε_t is i.i.d. non-Gaussian and $u_t = \Phi(L)y_t$ is the purely non-causal component of the process.

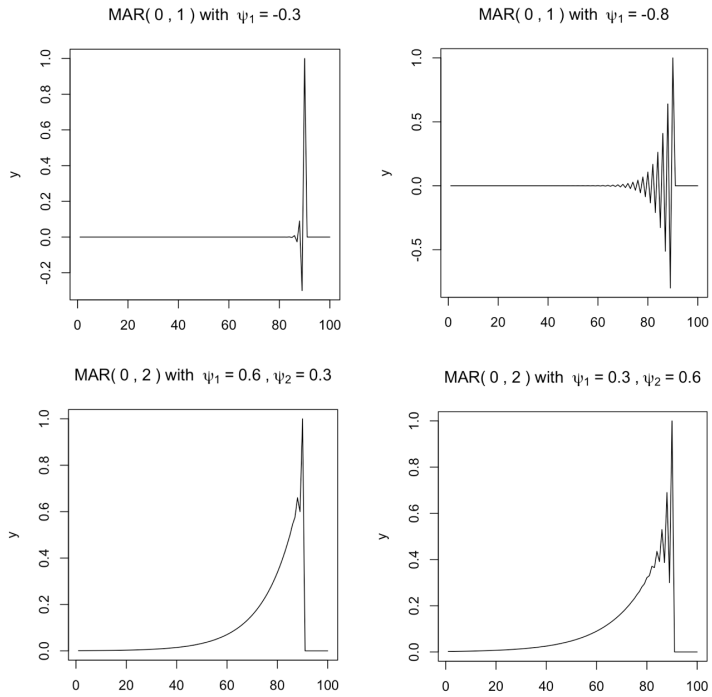

Figure 2: Complex dynamics induced in $MAR(0,s)$ processes by negative coefficients or multiple leads

3.1 Point predictions

Gouriéroux and Zakoïan (2017) derive the first two conditional moments of $MAR(0,1)$ processes,[4] here denoted as u_t, and show that for Cauchy processes, with $\psi > 0$, the conditional expectation of u_{T+1} is

$$\mathbb{E}\big[u_{T+1}|u_T\big] = u_T. \tag{3}$$

This result is puzzling since the conditional expectation of a noncausal process has a unit root even though the process is stationary. More generally, the conditional expectation of $MAR(r,1)$ processes,

$$\mathbb{E}\big[y_{T+1}|y_T\big] = \phi_1 y_T + \cdots + \phi_r y_{T-r+1} + u_T,$$

[4]Note that the linear projection on the past does not correspond, in this context, to the conditional expectation (Gouriéroux and Jasiak, 2018).

and using Equation (1), is equivalent to

$$\begin{aligned}\mathbb{E}\big[y_{T+1}|y_T\big] &= \phi_1 y_T + \cdots + \phi_r y_{T-r+1} + y_T + \phi_1 y_{T-1} + \cdots + \phi_r y_{T-r} \\ &= y_T + (1-L)(\phi_1 y_T + \cdots + \phi_r y_{T-r+1}).\end{aligned}$$

The last equality corroborates the findings of Fries and Zakoïan (2019) for $MAR(r,1)$ Cauchy models. They show that the conditional expectation at any forecast horizon for any symmetric $\alpha-$stable distributed $MAR(r,1)$ process can be expressed as a lag polynomial of the last observed value (see Proposition 3.2 (Fries and Zakoïan, 2019)),

$$\mathbb{E}\big[y_{T+h}|y_T\big] = \mathcal{P}_h(L)y_T,$$

with $h \geq 1$ and where $\mathcal{P}_h(L)$ is a polynomial of degree r. Fries (2018) expanded those results to any admissible parametrisation of the tail and asymmetry parameters of α-stable distributions and derives up to the fourth conditional moments. He also derives the limiting distribution of those four moments when the variable of interest diverges. He shows that during an explosive episode, the computation of those moments gets considerably simplified and are characteristic of a weighted Bernoulli distribution charging probability $\psi^{\alpha h}$ to the value $\psi^{-h}u_T$ and $(1-\psi^{\alpha h})$ to value zero, for a tail parameter $0 < \alpha < 2$. Those results indicate that along a bubble, the process can only either keep on increasing with fixed rate or drop to zero. For Cauchy-distributed errors ($\alpha = 1$), the mean forecast during an explosive episode remains equal to Equation (3), yet for other α-stable distributions the conditional expectation may be drastically simplified. Hence, during an explosive episode, the point forecast of an $MAR(0,1)$ process is a weighted average of the crash and further increase (e.g. a random walk for Cauchy-distributed processes), which can be rather misleading. Density forecasts may therefore be more informative as they carry more information.

3.2 Density predictions

The equivalence of information sets $(y_1, \ldots, y_T, y^*_{T+1}, \ldots, y^*_{T+h})$ and $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-1}, u_T, u^*_{T+1}, \ldots, u^*_{T+h})$, where $v_t = \Phi(L)^{-1}\varepsilon_t$ and $u_t = (1-\psi L^{-1})^{-1}\varepsilon_t$, allows to predict future values of y from predictions of the forward-looking component of ε, namely u. The asterisk indicates unrealised values of the random variables. Most prediction methods hence aim attention at purely noncausal processes – here u_t. Predictions of the $MAR(0,1)$ component can then be converted into predictions of the $MAR(r,1)$ process. The conditional predictive density (as named by

Gouriéroux and Jasiak, 2016) or the causal transition distribution (as named by Gouriéroux and Zakoïan, 2017) of the h future values, $(u^*_{T+1}, \ldots, u^*_{T+h})$, given the information known at time T is as follows,

$$\begin{aligned} & l(u^*_{T+1}, \ldots, u^*_{T+h} | y_1, \ldots, y_T) \\ & = l(u^*_{T+1}, \ldots, u^*_{T+h} | v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-1}, u_T) \\ & = l(u^*_{T+1}, \ldots, u^*_{T+h} | u_T), \end{aligned} \tag{4}$$

where l denotes densities associated with the noncausal process u_t. The reduction of the conditional information set stems from information sets equivalence and the independence between error components. While the interest is on predicting the future given present and past information, it is only possible, by the model definition, to derive the density of a point conditional on its future point. Bayes' rule is first used to get rid of the conditioning on the present point and a second time to condition on the last point of the forecast. Then, the theorem is applied repeatedly on the first term until the density of all points is conditional on future ones. The conditional *pdf* in (4) can thus be expressed as follows,

$$\begin{aligned} & l(u^*_{T+1}, \ldots, u^*_{T+h} | u_T) \\ & = \frac{l(u_T, u^*_{T+1}, \ldots, u^*_{T+h-1}, u^*_{T+h})}{l(u_T)} \\ & = l(u_T, u^*_{T+1}, \ldots, u^*_{T+h-1} | u^*_{T+h}) \times \frac{l(u^*_{T+h})}{l(u_T)} \\ & = \left\{ l(u_T | u^*_{T+1}, \ldots, u^*_{T+h}) l(u^*_{T+1} | u^*_{T+2}, \ldots, u^*_{T+h}) \ldots l(u^*_{T+h-1} | u^*_{T+h}) \right\} \\ & \qquad \times \frac{l(u^*_{T+h})}{l(u_T)}. \end{aligned}$$

Equation (2) states that $\varepsilon_t = u_t - \psi u_{t+1}$, hence, for all t, only u_{t+1} is necessary to derive u_t. Furthermore, given u_{t+1}, the conditional density of u_t (which we do not know) is equivalent to the density of ε_t (which we know) evaluated at the point $u_t - \psi u_{t+1}$.[5] That is, for any assumed errors distribution g we have,

[5]Since $u_\tau = \psi u_{\tau+1} + \varepsilon_\tau$ for any time point $1 \leq \tau \leq T$, $l_{u_\tau | u_{\tau+1}}(u|x) = g_{\varepsilon_\tau}(u - \psi x)$ for all time point τ and values u and x. For simplicity, the density distributions related to u_t (resp. ε_t) are just denoted by l (resp. g).

$$
\begin{aligned}
&l(u^*_{T+1},\dots,u^*_{T+h}|u_T)\\
&=\Big\{l(u_T|u^*_{T+1})l(u^*_{T+1}|u^*_{T+2})\dots l(u^*_{T+h-1}|u^*_{T+h})\Big\}\times\frac{l(u^*_{T+h})}{l(u_T)}\\
&=g(u_T-\psi u^*_{T+1})\dots g(u^*_{T+h-1}-\psi u^*_{T+h})\times\frac{l(u^*_{T+h})}{l(u_T)}.
\end{aligned}
$$

Problems may however arise with the two remaining marginal *pdf* $l(u_T)$ and $l(u^*_{T+h})$. We know that $u_t=\psi u_{t+1}+\varepsilon_t=\sum_{i=0}^{\infty}\psi^i\varepsilon_{t+i}$ but the *pdf* of a linear combinations of errors may not admit closed-form expressions for some distributions.

For instance, Gouriéroux and Zakoïan (2013) present closed-form solutions for the predictive conditional density of purely noncausal $MAR(0,1)$ processes with Cauchy-distributed errors. They show that the characteristic function of the infinite sum corresponds to that of a Cauchy with scale parameter $\frac{\gamma}{(1-|\psi|)}$, where γ is the scale of the distribution of the errors ε_t. Hence, in the $MAR(r,1)$ case with Cauchy errors, $u_t\sim Cauchy\Big(0,\frac{\gamma}{(1-|\psi|)}\Big)$. The predictive density of the purely noncausal process (u_t) can thus be computed as such,

$$
\begin{aligned}
&l(u^*_{T+1},\dots,u^*_{T+h}|u_T)\\
&=g(u_T-\psi u^*_{T+1})\dots g(u^*_{T+h-1}-\psi u^*_{T+h})\times\frac{l(u^*_{T+h})}{l(u_T)}\\
&=\frac{1}{(\pi\gamma)^h}\left(\frac{1}{1+\frac{(u_T-\psi u^*_{T+1})^2}{\gamma^2}}\cdots\frac{1}{1+\frac{(u^*_{T+h-1}-\psi u^*_{T+h})^2}{\gamma^2}}\right)\\
&\times\frac{\gamma^2+(1-|\psi|)^2u_T^2}{\gamma^2+(1-|\psi|)^2(u^*_{T+h})^2}.
\end{aligned}
$$

Analogously, since $u_T=\psi^h u^*_{T+h}+\sum_{i=0}^{h-1}\psi^i\varepsilon_{T+i}$, it follows that $u_T|u^*_{T+h}\sim Cauchy\Big(0,\gamma_h\Big)$, where $\gamma_h=\sum_{i=0}^{h-1}|\psi^i|\gamma$. Hence, for Cauchy distributed errors, instead of the h-dimensional conditional joint density, the conditional density of an h-step ahead point forecast can be obtain as such,

$$
\begin{aligned}
l(u^*_{T+h}|u_T)&=l(u_T|u^*_{T+h})\times\frac{l(u^*_{T+h})}{l(u_T)}\\
&=\frac{1}{\pi\gamma_h}\frac{\gamma_h^2}{(u_T-\psi^h u^*_{T+h})^2+\gamma_h^2}\times\frac{\gamma^2+(1-|\psi|)^2u_T^2}{\gamma^2+(1-|\psi|)^2(u^*_{T+h})^2}.
\end{aligned}
$$

To illustrate how the predictive density evolves as the series diverges, Figure 3 shows one-step ahead forecasts for different levels ($y_T = \{0.50, 10, 50\}$)) of a purely noncausal process with a lead coefficient of 0.8 and standard Cauchy-distributed errors. While the predictive distribution is uni-modal for low levels (close to zero), it splits and becomes a bi-modal distribution as the level of the series increases, and the more it diverges the more evident is the bi-modality of the distribution. The two modes correspond to a drop to 0 and a continuous increase to $(1/0.8)u_T$; each event has probability 0.2 and 0.8 respectively. For instance, when the series attains 50, the probability that it will keep on increasing to a point close to 62.5 is 0.8. Those results corroborate what Fries (2018) shows for diverging Cauchy-distributed $MAR(0,1)$ series. Note that results are analogous for any parameters.

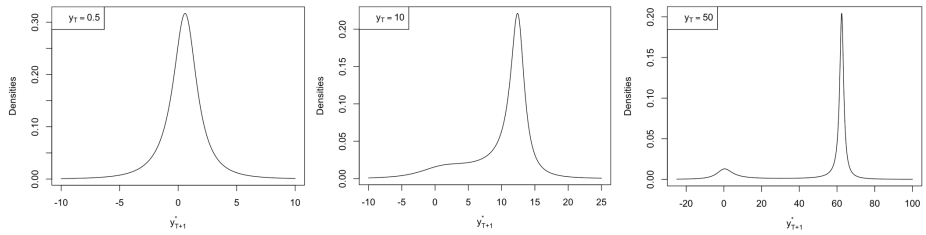

Figure 3: Evolution of the 1-step ahead predictive density as the level of the series increases for an $MAR(0,1)$ with $\psi = 0.8$.

The predictive density of the purely noncausal filtration of the errors, u, needs to be transformed into that of the variable of interest, y. As explained before, due to equivalence of information sets, the density for $(u^*_{T+1}, \ldots, u^*_{T+h})$ can directly be converted into the predictive density of $(y^*_{T+1}, \ldots, y^*_{T+h})$. In case y_t is a purely noncausal process, it is equal to the process u_t, and forecasting one is equivalent to forecasting the other. However, in the $MAR(r,1)$, since $y^*_{T+1} = \phi_1 y_T + ... + \phi_r y_{T-r+1} + u^*_{T+1}$, the density of the purely noncausal process is shifted by $\phi_1 y_T + ... + \phi_r y_{T-r+1}$. For $h = 2$, y^*_{T+2} depends on u^*_{T+2} and y^*_{T+1}, which itself depends on u^*_{T+1}. Overall, the predictive density of y^*_{T+h} (or of the future path of length h) is determined by the h-dimensional conditional joint density of $(u^*_{T+1}, \ldots, u^*_{T+h})$. Another way of approaching this is to directly write the predictive density in terms of y^*_{T+k}, with $1 \leq k \leq h$, in the conditional joint density of $(u^*_{T+1}, \ldots, u^*_{T+h})$. For an $MAR(1,1)$ process with lag coefficient ϕ for instance, the conditional

predictive density of h future y's could be obtained as follows,

$$l(y^*_{T+1},\dots,y^*_{T+h}|y_T) = \frac{1}{(\pi\gamma)^h} \times \frac{1}{1+\frac{(u_T-\psi(y^*_{T+1}-\phi y_T))^2}{\gamma^2}} \cdots \frac{1}{1+\frac{(y^*_{T+h-1}-\phi y^*_{T+h-2})-\psi(y^*_{T+h}-\phi y^*_{T+h-1}))^2}{\gamma^2}} \times \frac{\gamma^2+(1-|\psi|)^2u_T^2}{\gamma^2+(1-|\psi|)^2(y^*_{T+h}-\phi y^*_{T+h-1})^2}.$$

Figure 4 shows the evolution of two-step ahead forecasts of a purely noncausal process with lead coefficient 0.8 and Cauchy-distributed errors as the variable increases. For high levels of the series, the split of the distribution is evident; at each step the series can either keep on increasing or drop to zero, where the latter corresponds to an absorbing state. The interpretation of each area of the graph is explained in Figure 5, showing which region of the graph corresponds to which potential future shape of the series given the last observed point.

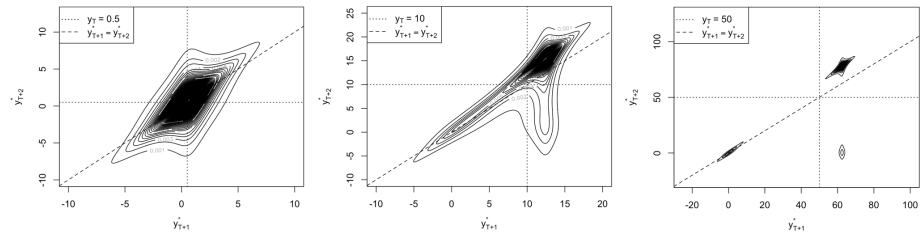

Figure 4: Evolution of the 2-step ahead joint predictive density as the level of the series increases for an $MAR(0,1)$ with $\psi = 0.8$ and Cauchy-distributed errors

Overall, density predictions yield a more complete forecast as they carry more information regarding potential future patterns of the series. They cannot be easily graphically displayed for forecast horizons larger than 2, yet results can be used to compute probabilities regarding future patterns. For instance, when the variable follows an explosive path, probabilities of a crash can be computed from the densities by choosing a threshold, such as the last observed value or its half for instance. Nonetheless, as indicated by Fries (2018) for α-stable distributions, explosive episodes seem to be memoryless and as the series diverges, probabilities of a crash tend to the

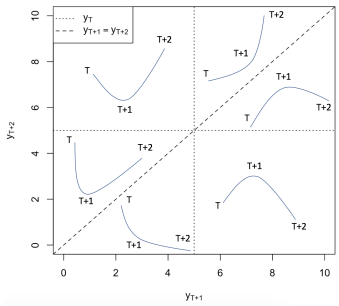

Figure 5: Future patterns based on joint predictive density. The dotted lines correspond to the last observed point and the diagonal dashed line to the line $y^*_{T+1} = y^*_{T+2}$.

constant $|\psi|^{\alpha h}$ for given a given horizon $h \geq 1$. This may however not be very realistic when it comes to real life data. We might expect probabilities of a crash in stock prices for instant to increase with the level of prices since a bubble cannot go on forever. This reason and the fact that the use of other fat-tail distributions (e.g. Student's t) may lead to the absence of closed-form expressions for the conditional moments and densities led to the elaboration of approximation methods. The next Section presents two approaches to approximate the conditional densities; the first one uses sample-counterparts (Gouriéroux and Jasiak, 2016) and the second is based on simulations (Lanne et al., 2012).

4 Predictions using approximation methods

4.1 Predictions using sample-based approximations

This section is based on the approach proposed by Gouriéroux and Jasiak (2016). They derive a sample-based estimator of the predictive densities based on past values of the series and this method can be applied to any non-Gaussian distribution.

Recall that the predictive density of h future values $(u^*_{T+1}, \ldots, u^*_{T+h})$ of an $MAR(0,1)$ process is as follows,

$$\begin{aligned} &l(u^*_{T+1}, \ldots, u^*_{T+h}|u_T) = \\ &g(u_T - \psi u^*_{T+1}) \ldots g(u^*_{T+h-1} - \psi u^*_{T+h}) \times \frac{l(u^*_{T+h})}{l(u_T)}. \end{aligned} \tag{5}$$

One of the reason leading to the derivation of this sample-based estimator is that for some distributions, the marginal distribution of u_T and u^*_{T+h} do not admit closed-form expressions. They can however, based on the iterated expectation theorem, be expressed as follows,

$$l(u_\tau) = \mathbb{E}_{\tau+1}\big[l(u_\tau|u_{\tau+1})\big],$$

with $\tau = \{T, T+h\}$. Once again the noncausal relationship described in Equation (2) is used to evaluate the conditional distribution of $l(u_\tau|u_{\tau+1})$ with the distribution of the errors, $g(u_\tau - \psi u_{\tau+1})$. Subsequently, the expected value of the latter can be approximated by its sample-based counterparts as the average obtained using all points from the sample for the conditional variable,

$$l(u_\tau) = \mathbb{E}_{\tau+1}\big[g(u_\tau - \psi u_{\tau+1})\big] \approx \frac{1}{T}\sum_{t=1}^{T}\Big\{g(u_\tau - \psi u_t)\Big\}. \tag{6}$$

Hence, the predictive density for the $MAR(0,1)$ process u_t can be approximated by plugging the sample counterparts (6) in (5),

$$\begin{aligned} &l(u^*_{T+1}, \ldots, u^*_{T+h}|u_T) \\ &\approx g(u_T - \psi u^*_{T+1}) \ldots g(u^*_{T+h-1} - \psi u^*_{T+h}) \frac{\sum_{t=1}^{T} g(u^*_{T+h} - \psi u_t)}{\sum_{t=1}^{T} g(u_T - \psi u_t)}. \end{aligned}$$

For centred Cauchy- or Student's t-distributed errors for instance, the density $g(u_\tau - \psi u_t)$, with $\tau = \{T, T+h\}$, is maximised when $u_\tau = \psi u_t$, for some $1 \leq t \leq T$. That is, as u_τ departs from all past realised values of the series, $\frac{1}{T}\sum_{t=1}^{T} g(u_\tau - \psi u_t)$ will tend to zero. Hence, the estimated ratio may significantly differ from $\frac{l(u^*_{T+h})}{l(u_T)}$ due to approximation errors and since the denominator approaches zero when the last observed point departs from all past values, approximations errors of the whole ratio will be amplified for high values of u_T. Furthermore, since at time T u_T is known and its realised value is used in the estimation, the ratio varies as a function of u^*_{T+h}. Hence, the closer will u^*_{T+h} be to past values of the sample, the higher will the ratio be. We can therefore expect to obtain significant discrepancies between closed-form and sample-based conditional densities for point forecasts that are of the similar magnitude as values already observed in the past.

Computations for $MAR(r,1)$ processes depend on the last r observed values of y and on u_T, and this dependence makes it more difficult to generalise results. Hence, for the sake of simplicity and comparison, we consider $MAR(0,1)$ processes with a lead coefficient of 0.8 and standard Cauchy-distributed errors. For low levels of the series, results are similar between closed-form and sample-based predictions, regardless of past behaviours. As shown in Figure 6, when the series is close to zero, the sample-based estimator fully recovers closed-form results but as expected, when the series departs from central values distortions appear. Nonetheless, the estimator captures the split of the density and thus the potential outset of an explosive episode.

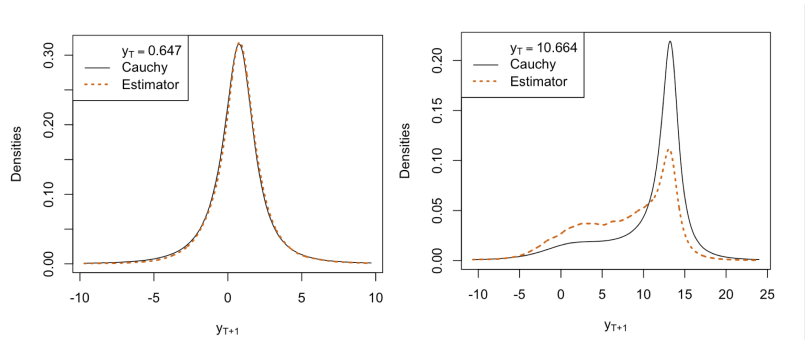

Figure 6: 1-step ahead forecasts of an $MAR(0,1)$ series with lead coefficient $\psi = 0.8$ at 2 arbitrary low levels (0.65 and 10.66).

Figure 7 shows one-step ahead forecasts of two $MAR(0,1)$ series – one with no bigger previous bubbles (left) and the other with two larger explosive episodes – evaluated at similar levels close to 50. As explained above, approximations errors are more considerable for higher levels. The bottom left graphs depicts how, when the level of the series has exceeded all past values, distortions mostly happen around the crash (spanning the range of past values). The bottom right graph on the other hand shows how previous explosive paths may influence the probabilities of a further increase. Distortions are now mostly apparent on the increasing part of the predictive distribution, induced by the past explosive episodes. Indeed, the process already attained such level before and kept on increasing, and the values to which it then increased are assigned significant probabilities due to the sample-based approximation. Subsequently, the empirical conditional cumulative density function (hereafter *cdf*) can be derived from

the predictive density. To compute cumulative probabilities, we arbitrarily chose the last observed value as the threshold to obtain probabilities of a decrease. For the two series, probabilities of a decrease are respectively 0.57 and 0.33 compared to the closed-form Cauchy-derived result of approximately 0.23.[6] Note that the choice of threshold may significantly affect the outcome as probabilities of a decrease may be significantly larger than probabilities of a drop of at least 20% for instance. This is due to large probabilities assigned to points of similar magnitude to the last observed value. The difference between the sample-based probabilities of the two processes arise from the learning mechanism of this approach. On the other hand, differences between closed-form and sample-based results stem from realised values inducing higher probabilities of following similar paths as before in the approximation method. Those approximation errors can however be considered as updates of probabilities based on what previously happened in the series. If statistically they are approximations errors, in real life, series may tend to behave similarly as in the past and in such case, using only past values instead of all potential values in the estimation of the densities may capture this phenomenon.

Results are similar for two-step ahead predictions; the estimated densities depend both on past behaviours and level of the series. However, predictions with this approach are significantly computationally heavier and if the variable follows an explosive path, precise forecasts of more than two steps ahead are laboriously obtained. Gouriéroux and Jasiak (2016) propose a method to tackle this issue by elaborating a Sampling Importance Resampling (SIR) algorithm. The algorithm aims at recovering a predictive density based on simulations from a misspecified instrumental model from which it is easier to simulate. They suggest using a Gaussian AR model of order s (here an $AR(1)$) to simulate the process u_t. This approach recovers the intended densities for low levels of the series but fails to recover both the parts corresponding to the crash and to the increase when the variable exceeds some threshold. This threshold depends on past behaviours and on the underlying distribution, but this failure of the algorithm for high levels of the series stems from the intention to recover a bi-modal distribution from a uni-modal distribution. If the variance of the uni-modal instrumental distribution is not large enough

[6] The closed-form result slightly departs from the limiting results of Fries (2018) due to the choice of threshold used to compute the cumulative probabilities (u_T). It may be located too close to the left tail of the increasing part of the density, amplifying the total probabilities of a decrease.

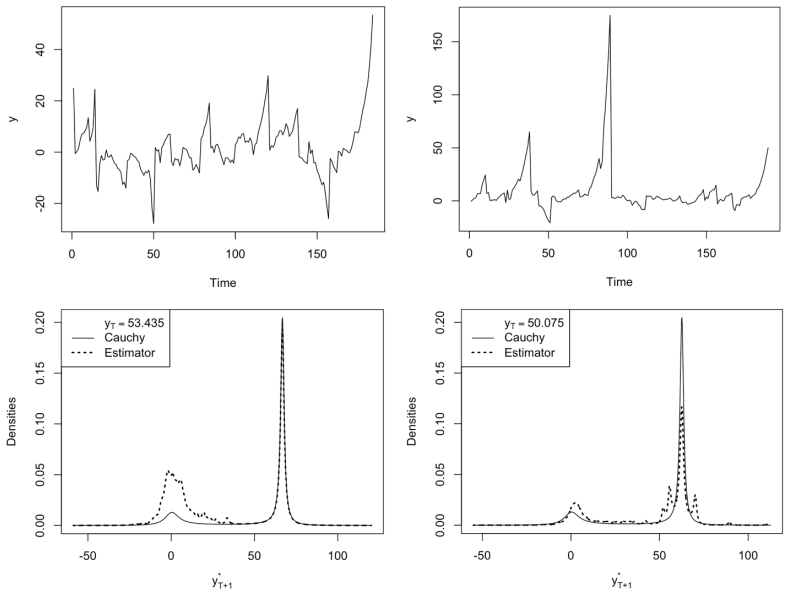

Figure 7: 1-step ahead forecasts of two $MAR(0,1)$ series with lead coefficient $\psi = 0.8$ at an arbitrary level (around 50).

to cover both modes of the sample-based density, the algorithm will not be able to recover the whole conditional distribution. The shape of the Normal distribution significantly depends on past behaviours of the series since the variance is estimated as the variance of the residuals of the *MAR* model. Hence, for more volatile series, the variance of the instrumental Normal distribution will be larger, yet, as the variable increases and the two modes diverge, there will always be a point from which the SIR algorithm does not succeed in recovering the density anymore.

Gouriéroux et al. (2018) find that the quality of forecasts diminishes when the series follows an explosive episode. Indeed, approximations errors amplify with the level of the series, and there is a point from which the SIR algorithm does not recover the whole density anymore. Yet, we find that the sample-based estimator captures the split of the conditional density as the series departs from central values and always comprises both the crash and increase parts of the predictive density. Furthermore, it yields varying probabilities of events based on its learning mechanism. While sample-based predictive densities based on Student's t-distributions cannot be compared to closed-form predictions, results corroborate the conclusions drawn with

Cauchy. A limitation is that when closed-form results are not available, we cannot disentangle how much of the derived probabilities are induced by the underlying distribution and how much by past behaviours. An alternative approximation method was proposed by Lanne et al. (2012) and it is based on simulations rather than past points.

4.2 Predictions using simulations-based approximations

Lanne et al. (2012) base their methodology on the fact that the noncausal component of the errors, u, can be expressed as an infinite sum of future errors, which in the $MAR(r,1)$ case is easily obtained as such,

$$u_t = \Psi(L^{-1})^{-1}\varepsilon_t = \sum_{i=0}^{\infty} \psi^i \varepsilon_{t+i}.$$

Since stationarity is assumed, the sequence (ψ^i) is converging to zero. Hence, they assumed that there exists an integer M large enough so that any future point of the noncausal component of the errors can be approximated as the following finite sum,

$$u^*_{T+h} \approx \sum_{i=0}^{M-h} \psi^i \varepsilon^*_{T+h+i}, \tag{7}$$

for any $h \geq 1$.

As shown before, any point forecast y^*_{T+h} depends on the sequence forecast $(u^*_{T+1}, \ldots, u^*_{T+h})$. Using the companion form of an $MAR(r,1)$ model, y^*_{T+h} can be expressed as the sum of a known component and the h future values of u_t, where the latter, based on Equation (7), can be approximated as a linear combination of M future errors

$$\begin{aligned} y^*_{T+h} &= \iota' \mathbf{\Phi}^h \mathbf{y}_T + \sum_{i=0}^{h-1} \iota' \mathbf{\Phi}^i \iota u^*_{T+h-i} \\ &\approx \iota' \mathbf{\Phi}^h \mathbf{y}_T + \sum_{i=0}^{h-1} \iota' \mathbf{\Phi}^i \iota \sum_{j=0}^{M-h+i} \psi^j \varepsilon^*_{T+h-i+j}, \end{aligned} \tag{8}$$

where

$$\mathbf{y}_T = \begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-r+1} \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \dots & \phi_r \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} (r\times r) \quad \text{and} \quad \iota = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (r\times 1).$$

Thus, forecasting any future point y^*_{T+h} or the path $(y^*_{T+1}, \dots, y^*_{T+h})$, with $h \geq 1$, requires forecasting the sequence of M future errors $(\varepsilon^*_{T+1}, \dots, \varepsilon^*_{T+M})$ which we will denote as ε^*_+. The issue is that the M-dimensional conditional distribution of ε^*_+ is almost impossible to obtain. Instead, Lanne et al. (2012) propose a way to obtain point and cumulative density forecasts. While the estimation approach they propose (Lanne and Saikkonen, 2011) requires finite variance for the errors distribution, this restriction is not necessary for their forecasting method.

Let $g(\varepsilon^*_+|u_T)$ be the conditional joint distribution of the M future errors, which, using Bayes' rule can be expressed as follows,

$$g(\varepsilon^*_+|u_T) = \frac{l(u_T|\varepsilon^*_+)}{l(u_T)} g(\varepsilon^*_+).$$

Thus, for any function q,

$$\begin{aligned} \mathbb{E}\Big[q(\varepsilon^*_+)\Big|u_T\Big] &= \int q(\varepsilon^*_+) g(\varepsilon^*_+|u_T) d\varepsilon^*_+ \\ &= \frac{1}{l(u_T)} \int q(\varepsilon^*_+) l(u_T|\varepsilon^*_+) g(\varepsilon^*_+) d\varepsilon^*_+ \\ &= \frac{\mathbb{E}\Big[q(\varepsilon^*_+) l(u_T|\varepsilon^*_+)\Big]}{l(u_T)}. \end{aligned} \tag{9}$$

Similarly as before, $l(u_T|\varepsilon^*_+)$ can be obtained from the errors distribution g. Yet, since it is conditional on ε^*_+ instead of u^*_{T+1}, we can only obtain the following approximation,

$$l\big(u_T|\varepsilon^*_+\big) \approx g\left(u_T - \sum_{i=1}^{M} \psi^i \varepsilon^*_{T+i}\right).$$

Using this approximation and the Iterated Expectation theorem, the marginal distribution of u_T can be approximated as follows,

$$l(u_T) = \mathbb{E}_{T+1}\big[l(u_T|\varepsilon^*_+)\big] \approx \mathbb{E}\left[g\left(u_T - \sum_{i=1}^{M} \psi^i \varepsilon^*_{T+i}\right)\right].$$

Overall, by plugging the aforementioned approximations in (9), we obtain

$$\mathbb{E}\Big[q(\varepsilon_+^*)\big|u_T\Big] \approx \frac{\mathbb{E}\Big[q(\varepsilon_+^*)g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\Big)\Big]}{\mathbb{E}\Big[g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\Big)\Big]}.$$

While Gouriéroux and Jasiak (2016) use past sample to estimate the entity of interest, Lanne et al. (2012) make use of simulations. Let $\varepsilon_+^{*(j)} = \Big(\varepsilon_{T+1}^{*(j)}, \ldots, \varepsilon_{T+M}^{*(j)}\Big)$, with $1 \leq j \leq N$, be the j-th simulated series of M independent errors. Assuming that the number of simulations N is large enough, the conditional expectation of interest can be approximated as follows,

$$\mathbb{E}\Big[q(\varepsilon_+^*)\big|u_T\Big] \approx \frac{N^{-1}\sum_{j=1}^{N} q\Big(\varepsilon_+^{*(j)}\Big)g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^{*(j)}\Big)}{N^{-1}\sum_{j=1}^{N} g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^{*(j)}\Big)}. \tag{10}$$

From (8), we can obtain conditional cumulative probabilities as follows,

$$\begin{aligned}\mathbb{P}\Big(y_{T+h}^* \leq x\big|u_T\Big) &= \mathbb{E}\Big[\mathbf{1}\big(y_{T+h}^* \leq x\big)\big|u_T\Big] \\ &\approx \mathbb{E}\left[\mathbf{1}\left(\iota'\mathbf{\Phi}^h\mathbf{y}_T + \sum_{i=0}^{h-1}\iota'\mathbf{\Phi}^i\iota\sum_{j=0}^{M-h+i}\psi^j\varepsilon_{T+h-i+j}^* \leq x\right)\Big|u_T\right]\end{aligned}$$

That is, for any $MAR(r,1)$ process and for any forecast horizon $h \geq 1$, choosing $q(\varepsilon_+^*) = \mathbf{1}\big(\sum_{i=0}^{h-1}\iota'\mathbf{\Phi}^i\iota\sum_{j=0}^{M-h+i}\psi^j\varepsilon_{T+h-i+j}^* \leq x_u\big)$ in (10), where $x_u = x - \iota'\mathbf{\Phi}^h\mathbf{y}_T$, will provide an approximation of $\mathbb{P}\big(y_{T+h}^* \leq x|u_T\big)$. By computing its value for all possible x we can obtain the whole conditional *cdf* of y_{T+h}^*.

For the sake of simplicity and comparison, we again consider $MAR(0,1)$ processes with a lead coefficient of 0.8 and Cauchy-distributed errors. Probability forecasts are performed for different levels of u_T; the levels were arbitrarily chosen to represent different stages of the process, from stable to explosive. The threshold x in the indicator function was chosen as the last observed value to compute probabilities of a decrease. While augmenting the truncation parameter M has no notable consequences, computation time significantly increases with the number of simulations N. We hence computed the probabilities with a fixed truncation parameter $M = 100$ and

u_T	**Cauchy**	**Simulations-based predictive probabilities**		
		10,000 simulations	50,000 simulations	100,000 simulations
10	0.320	0.322 *(0.011)*	0.322 *(0.005)*	0.322 *(0.003)*
25	0.259	0.260 *(0.019)*	0.259 *(0.009)*	0.259 *(0.006)*
50	0.231	0.235 *(0.036)*	0.231 *(0.015)*	0.231 *(0.011)*
100	0.215	0.238 *(0.075)*	0.221 *(0.028)*	0.218 *(0.019)*
200	0.208	0.278 *(0.168)*	0.226 *(0.062)*	0.217 *(0.040)*

Table 1: Probabilities of a decrease computed for different levels of an $MAR(0,1)$ series with lead coefficient $\psi = 0.8$, derived from closed-form results and from simulations-based approximations. For the latter, mean and standard deviation (in brackets) over 1,000 iterations are reported.

different number of simulations $N = \{10{,}000,\ 50{,}000,\ 100{,}000\}$, including the 10,000 suggested by Lanne et al. (2012).

Table 1 shows the probabilities of a decrease of $MAR(0,1)$ series, evaluated at arbitrary levels. Closed-form results are presented in column 2, and as the level of the series increases, probabilities of a decrease approaches the limiting results presented by Fries (2018). The reason for which probabilities are higher than 0.20 even for high levels was explained before; the threshold (last observed value) used to compute probabilities may be too close to the left tail of the upper part of the distribution, slightly amplifying probabilities of a decrease. As the two modes of the conditional distribution diverges when the level increases, the impact of the choice of threshold becomes negligible. Simulations-based approximations are presented in columns 3 to 5 where the mean and standard deviation of 1,000 iterations are reported for the three different number of simulations within the computations. As the level of the series increases, results between iterations become more volatile. For instance, for a level of 200 with 10,000 simulations, predicted probabilities of a decrease range from 0.06 to 0.92; that is from almost certainty of a further increase to almost certainty of a crash. This means that the same inquiry repeated twice

may give completely opposite results. Recall that a bubble is triggered by an extreme value, and the date at which this extreme value is attained is the date of the crash. Consequently, if we are investigating an explosive episode there must be an extreme value in future points triggering the current explosion. However, as we have seen before, only extreme values at ε^*_{T+i} inducing the natural rate of increase $(1/\psi)^i$ are assigned significant probabilities. The simulations-based approach will therefore tend to indicate high probabilities of a crash, even with a significantly large number of simulations, if no such values are simulated. However, increasing the number of simulations reduces this discrepancy and brings the mean closer to closed-form results for all levels. Yet, even with 100,000 simulations, standard deviations remain quite large for high levels of the series and this indicates that more simulations are necessary. Overall, the higher the level of the series, the higher should the number of simulations be.

Repeating computations for all potential threshold x yields the whole conditional *cdf*. By choosing the appropriate number of simulations based on the the level of the series, an accurate approximation of the true conditional *cdf* can be obtained. Figure 8 shows the convergence of predictions towards the true predictive *cdf* (Cauchy-derived distribution) as the number of simulations in each iteration increases. Each point of the predictive probabilities is based on a different set of simulations. It represents one-step ahead predictive *cdf*'s of a series whose last observed value is 100 with 10,000, 50,000 and 100,000 simulations in the computations. The higher the number of simulations, the less noisy approximations are and the closer the results get to the closed-form distribution. It is important to note that the bi-modality of the conditional distribution is always captured. We illustrate this convergence of results with a maximum of 100,000 simulations due to time and computation constraints; yet, increasing it even more will reduce further the deviation from closed-form results. Once the predictive *cdf* is obtained, the empirical *pdf* can be derived from it; however, it is rather computationally demanding to obtain the whole conditional *cdf* with a large number of simulations. Hence, it can be quite restrictive and may lead to noisy approximations if the number of simulations is lowered. Instead, computing probabilities of fewer events with an even larger number of simulations may be more adequate. We can see from Figure 8 that the choice of threshold to compute probabilities of a drop, whether it is of a decrease or of a crash of 50%, will not significantly affect results. Indeed, probabilities assigned to events between the crash and the further increase are almost zero, thus cumulative probabilities do not significantly vary in

this interval.

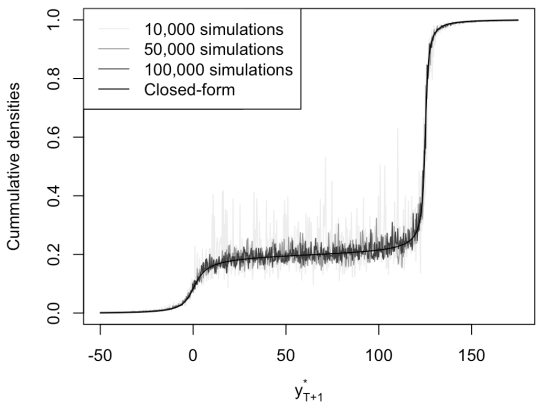

Figure 8: Simulations-based predictive *cdf*'s evaluated at $y_T = 100$ with different number of simulations compared to closed-form results.

For Student's t-distributions, while predictions cannot be compared to closed-form results, they are analogous; probabilities converge as the number of simulations is increased. The stability between iterations depends both on the distribution and amplitude of the series but with a large enough number of simulations, they seem to be a good approximation of closed-form results. Overall, once the series follows an explosive path, the number of simulations within the estimator needs to be increased.

5 Empirical Analysis

We now emiricallt evaluate the performance of the different approaches presented in Section 4. We forecast the bubble pattern in commodity prices and in particular in the monthly Global price of Nickel. The series is obtained from the International Monetary Fund and spans the period from January 1980 to June 2017. There seems to be a positive trend in the data but making the series stationary is far from obvious. Indeed, usual unit root tests do not perform well for this type of variable with very large spikes. For instance ADF tests would reject the null of a unit root against both a mean and a trend reverting alternative. A conclusion that does not

seem satisfactory from the graphs of the data. It might also well be that the series is stationary around a shift in mean. Hencic and Gouriéroux (2015) use a cubic deterministic trend for isolating the bubble in the Bitcoin. In order to preserve the bubble features of the data and to obtain a stationary series with locally explosive episodes (that would disappear by taking the returns) we have instead considered the Hodrick-Prescott filtering approach. The detrended series is reported in Figure 9. We are of course aware that this first step might alter the dynamics of the series, probably in the same manner that a X-11 seasonal filter modifies *MAR* models (see Hecq, Telg, and Lieb, 2017). We leave this important issue for further research. We first estimate an autoregressive model by OLS on the whole HP-detrended Nickel price series. Information criteria (AIC,BIC and HQ) all pick up a pseudo lag length of $p = 2$. The three possible $MAR(r,s)$ specifications are consequently a $MAR(2,0)$, a $MAR(1,1)$ or a $MAR(0,2)$. Using the MARX package a $MAR(1,1)$ with a t-distribution with a degree of freedom of 1.34 and a scale parameter of 356.147 is favoured. The value of the causal and the noncausal parameters are respectively 0.60 and 0.74. We are consequently in the situation in which the predictive density does not admit closed-form expressions (although not very far from the Cauchy), hence the sample- and simulations-based approaches can be used.

We aim attention at the main explosive episode, which crashed in June 2007. To investigate the evolution of predictions along the bubble like in a real-time setting, we estimate the model with expanding window at every step. Note that at each point, even if parameters differed, the model identified was always an $MAR(1,1)$. The points at which we perform predictions are represented by the diamonds on the series in Figure 9. We investigate five points along the main explosive episode and one after, to capture the effects on the inclusion of the crash in the estimation and prediction. Each point is assigned an index between 1 and 6 indicating their order of arrival. At each point, we compute the sample-based predictive density and the corresponding probability of a decrease (hence the presence of a turning point), as well as the probability of a decrease derived from the simulations-based approach with 3,000,000 simulations and a truncation parameter of 100. Results are presented in Table 2 where the model identified at each point is reported. Note that until now we assumed the model was correctly specified, hence, only the sample-based approach depended on past values in the prediction procedure. Here past realised values influence the model estimated at each point for the two approaches. Hence, only the sample-based method incorporates past values

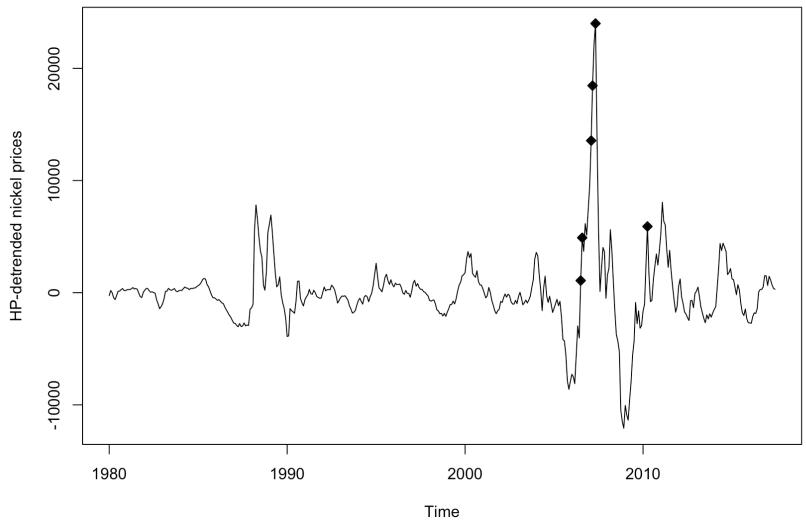

Figure 9: HP-detrended nickel prices series. The diamonds represent points from which one-step ahead forecasts are performed in this analysis.

in the predictions but both methods use the information carried by the series until the point of the forecast in the estimation.

The model estimated at the outset of the bubble (point 1) is mostly backward looking, the lag coefficient is 0.201 larger than the lead coefficient. However, as we move along the bubble, and therefore add higher points in the estimation, identification tends to favour more forward looking models. There seems to be a structural change in the series between the consecutive points 3 and 4 (between February and March 2007) where the scale suddenly increases by more than 20 and where magnitudes of lag and lead coefficients invert so that the series becomes mostly forward-looking. Furthermore, the degrees of freedom of the t-distribution vary between 1.41 and 1.54 before the crash and decrease to 1.18 once the crash is included in the estimation, implying higher probabilities of extreme events.

As mentioned before, a number of simulations large enough, the method proposed by Lanne et al. (2012) seems to be a good approximation of closed-form results. Simulations-based predictions only depend on the estimated underlying distribution and level of the series as opposed to sample-based results, computed with the method of Gouriéroux and Jasiak

Point	Model estimated			Probability of a decrease		
	ϕ	ψ	t(λ, σ)	Sample-based	Simulations-based	Difference
1	0.763	0.562	$t(1.54, 259.2)$	0.444	0.341	0.103
2	0.758	0.572	$t(1.43, 258.7)$	0.607	0.490	0.117
3	0.748	0.606	$t(1.41, 257.6)$	0.645	0.487	0.158
4	0.573	0.756	$t(1.53, 281.1)$	0.593	0.347	0.246
5	0.571	0.765	$t(1.49, 280.4)$	0.600	0.336	0.264
6	0.677	0.689	$t(1.18, 289.5)$	0.275	0.305	-0.030

Table 2: Models estimated at the 6 distinct points in calendar order, where ϕ and ψ are the causal and noncausal coefficients respectively and λ and σ the degrees of freedom and scale of the distribution. The corresponding probabilities of decrease at an horizon of 1 computed from the sample- and simulations-based approaches and the difference between them is reported in the last column.

(2016), which furthermore incorporates all past realised values of the series. The difference between them, reported in the last column, represents how much of the sample-based probabilities was induced by the learning mechanism of this approach. We see that as we move along the explosive episode, the difference in probabilities between the two methods increases; this is mostly due to the fact that we get further away from past maximum levels of the series and the sample-based approach, based on its learning mechanism, predict higher probabilities of decrease than the underlying distribution suggests. Probabilities of a crash are the highest at points 2 and 3 due to the identified models, a lower lead coefficient imply higher probabilities of a decrease. At point 1, the low probabilities are due to the relatively low level of the series and for the sample-based approach also to the fact that it is still close to numerous past values. Once predictions are performed after the bubble (point 6), probabilities of a drop with both methods significantly decrease. This is firstly due to the inclusion of the crash in estimation, which alters parameters (lower degrees of freedom and higher lead coefficient), which induces lower probabilities of a decrease, but also to the learning mechanism of the sample-based approach. Indeed, the series already reached three times this level and it once kept on increasing, thus, probabilities that it drops are now lowered by the main bubble and the probability from the sample-based approach is even 3% lower than the simulation-based prediction.

Overall, once the level of the series attains past maxima, probabilities of

a decrease significantly increase (from less than 0.45 to around 0.6 along the bubble) and the proportion of the probabilities induced by the learning mechanism also increases with the level of the series (from 0.1 to 0.26 difference). We investigated probabilities of a decrease but any threshold can be chosen. From an investor's perspective, the probabilities that the series will drop under some fixed threshold or by a certain percentage could also be derived, as well as values at risk or expected shortfall. Figure 10 shows how the choice of threshold may affect final probabilities. Cumulative probabilities are computed as the area under the curve on the left of the threshold. In the scenario depicted in Figure 10, probability of a decrease at the top of the bubble (point 5) is equal to 0.6 while probability of a drop of at least 25% is 0.484. The simulations-based probabilities of a decrease of at least 25% is equal to 0.322 compared to 0.336 for the probability of a decrease. That is, as explained before, the sample-based approach is much more sensitive to the choice of threshold and this may significantly alter the conclusions drawn from the results. In this scenario the difference between the two methods is reduced to 0.162. Furthermore, note that the detrending method applied to the series may significantly alter results interpretation. A combination of both the sample- and simulations-based predictive probabilities could also be employed, relying on the beliefs regarding how likely is the series to follow similar paths as before. Furthermore, farther horizon could be investigated but this is limited by computationally demanding sample-based approximations. An adaptation of the SIR algorithm proposed by Gouriéroux and Jasiak (2016) to bi-modal distribution could be considered.

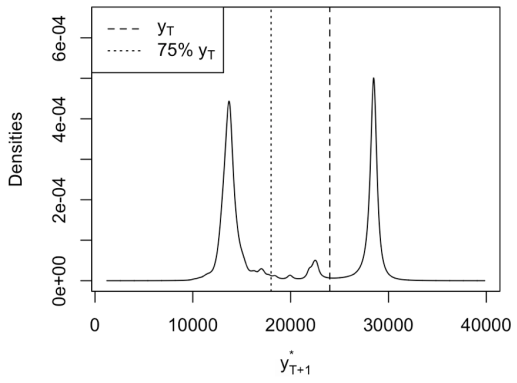

Figure 10: Sample-based predictive densities and thresholds to compute probability of decrease and probability of a decrease of at least 25%

6 Conclusion

This paper analyses and compares in details two approximation methods developed to forecast mixed causal-noncausal autoregressive processes. It focuses on $MAR(r,1)$ processes and aims attention at predictive densities rather than point forecasts as they are more informative, especially in the case of explosive episodes.
The sample-based (Gouriéroux and Jasiak, 2016) and simulations-based (Lanne et al., 2012) methods are compared to closed-form results using $MAR(0,1)$ processes with a lead coefficient of 0.8 and Cauchy-distributed errors. We focus on one-step ahead forecasts to give a rigorous analysis of how and why they may differ from closed-form results. We find that closed-form and sample-based predictive densities start to differ as the series departs from central values, and the discrepancies increase with the level of the series. The sample-based approach gives time-varying probabilities and depends on how similar the event under investigation is to past events. This approach yields results that are a mixture of probabilities ensuing from the underlying distribution and from past behaviours of the series. On the other hand, simulations-based predictive probabilities are a good approximation of closed-form results obtained with Cauchy-distributed errors, as long as the number of simulations in the approximations is large enough relative to the level of the series. Both methods capture the bi-modality of the conditional distribution as the series diverges from central values, which is an indicator of a potential bubble outset.

We illustrate the different methods with a detrended Nickel prices series. When the underlying distribution does not admit closed-form expressions for the predictive densities, the only way to disentangle how much was induced by past behaviours in the sample-based approach is to compute the difference with the probability computed with the simulations-based method. This information can be used in investment decisions, depending on how much the series is assumed to follow past behaviours.

References

Andrews, B., Davis, R., and Breidt, J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, *97*(7), 1638–1659.

Fries, S. (2018). Conditional moments of anticipative α-stable markov processes. *arXiv preprint arXiv:1805.05397*.

Fries, S., and Zakoïan, J.-M. (2019). Mixed causal-noncausal AR processes and the modelling of explosive bubbles. *Econometric Theory*, 1–37.

Gouriéroux, C., Hencic, A., and Jasiak, J. (2018). Forecast performance in noncausal MAR(1, 1) processes.

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

Gouriéroux, C., and Jasiak, J. (2018). Misspecification of noncausal order in autoregressive processes. *Journal of Econometrics*, *205*(1), 226–248.

Gouriéroux, C., Jasiak, J., and Monfort, A. (2016). Stationary bubble equilibria in rational expectation models. *CREST Working Paper. Paris, France: Centre de Recherche en Economie et Statistique*.

Gouriéroux, C., and Zakoïan, J.-M. (2013). Explosive bubble modelling by noncausal process. *CREST. Paris, France: Centre de Recherche en Economie et Statistique*.

Gouriéroux, C., and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(3), 737–756.

Hecq, A., Lieb, L., and Telg, S. (2016). Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*(123/124), 307–331.

Hecq, A., Telg, S., and Lieb, L. (2017). Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, *5*(4), 48.

Hencic, A., and Gouriéroux, C. (2015). Noncausal autoregressive model in application to bitcoin/ exchange rates. In *Econometrics of risk* (pp. 17–40). Springer.

Lanne, M., Luoto, J., and Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, *28*(3), 623–631.

Lanne, M., and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, *3*(3).