Learning the predictive density of mixed-causal ARMA processes

Elena Dumitrescu* Arthur Thomas†

March 8, 2024

Abstract

Mixed-causal ARMA processes are known to capture the dynamics of locally explosive behavior, such as bubble assets in finance. However, the limited knowledge of the predictive density of mixed-causal processes, especially during explosive bubble events, complicates their forecast and thus limits their use in practical applications. Given the lack of closed-form formulae for the conditional prediction density (except in special cases), simulation-based and sample-based methods have been proposed in the literature. Hovewer, these methods can be computationally expensive, especially for longer forecast horizons, and do not accurately capture the dynamics during explosive episodes. In this paper, we introduce Machine-learning algorithms for forecasting during bubble periods and show that K nearest neighbours and random forest learning methods are promising for this task. These approaches are shown to provide an interesting approximation to the true theoretical predictive densities and exhibit better forecasting abilities than existing simulation-based methods.

*University Paris-Panthéon-Assas, CRED, France, elena.dumitrescu@u-paris2.fr

†Paris Dauphine University - PSL, UMR CNRS 8007, LEDa, France, arthur.thomas@dauphine.psl.eu

We are grateful to Serge Darolles, Christian Francq, Gaëlle le Fol, Daniel Velasquez Gaviria, Christian Gouriéroux, Alain Hecq, Joan Jasiak, Sébastien Laurent, Yannick Le Pen, Gabriele Mingoli, Aryan Manafi Neyazi and Jean-Michel Zakoian. We also thank the seminar participants at Dauphine University - PSL and The Maastricht University School of Business and Economics, as well as the participants at the 17th International Conference on Computational and Financial Econometrics, and the 22nd Conference Développements Récents de l'Econométrie Appliquée à la Finance" (University of Paris Nanterre) for helpful comments and discussions.

1 Introduction

A close look at the dynamics of various asset prices, sometimes called speculative assets, reveals the presence of phases of locally explosive behaviour, i.e. rising patterns followed by a burst. These phenomena, together with the more traditional properties of heavy-tailed marginal distributions and volatility clustering, have been increasingly identified in financial markets around the world. Empirically, standard financial econometric models (ARMA-GARCH) do a poor job in capturing these non-linear characteristics of speculative bubbles. However, noncausal processes have been found to be suitable for modelling this locally explosive phenomena in financial time series (Cavaliere et al., 2020; Fries and Zakoian, 2019; Fries, 2021; Gourieroux and Jasiak, 2016; Gouriéroux and Jasiak, 2016; Gourieroux and Zakoian, 2017; Gourieroux and Jasiak, 2018; Gourieroux et al., 2021; Hecq and Voisin, 2021; Lanne and Saikkonen, 2011). The attractive flexibility of anticipative processes cannot yet be fully leveraged, however, the limited knowledge of the predictive density of noncausal processes, especially during explosive bubble events, hampers the ability to predict them and thus limits their use in practical applications. A remarkable exception is that of the anticipative α-stable AR(1) for which partial results were obtained in Gourieroux and Zakoian (2017) and further completed in Fries (2021). For multivariate, Gourieroux and Jasiak (2022), propose a semi-parametric estimation methods to compute the true predictive density. Two simulation and sample-based methods have been proposed in the noncausal literature to approximate the conditional distribution of noncausal processes (Lanne and Saikkonen, 2011; Gourieroux and Jasiak, 2016). While they provide flexible alternatives for predicting general noncausal processes, Hecq and Voisin (2021) note that these methods can become computationally intensive for larger prediction horizons and that they fail to accurately capture the dynamics during explosive episodes (see e.g. Gourieroux et al., 2021). As a consequence, despite the importance of risk management during such episodes, to our knowledge there is only one paper De Truchis et al. (2023) that proposes a portfolio allocation strategy that explicitly accounts for the distributional characteristics of bubble assets.

This paper proposes new methods for learning the conditional predictive density of

univariate general mixed-causal autoregressive moving average (MARMA) processes based on statistical learning approaches, i.e. K nearest neighbours and Random Forest (Dalmasso et al., 2020). Our contribution to the literature is threefold. This paper is the very first to propose machine learning methods to approximate and forecast the conditional density of noncausal processes, and to propose a numerical approach for forecasting MARMA processes. Second, we show that these methods exhibit better forecasting abilities than standard algorithms available in the literature. As a minor contribution we also extend the Lanne and Saikkonen (2011)'s forecasting procedure for α-stable laws.

In a Monte Carlo simulation exercise, we show that the learning methods outperform the standard approaches available in the literature and that they provide an interesting approximation to the true theoretical prediction densities.

[To be extended]

The rest of the paper is structured as follows. Section 2, summarises the theory of general MARMA models and introduces the closed-form solutions of the conditional moments of the predictive density. Section 3 discusses machine learning conditional forecasting. The Monte Carlo analysis is presented in Sections 4 and 5. Section **??** details the empirical illustration while Section 6 concludes and discusses avenues for future research.

2 Conditional moments of the predictive density for general MARMA processes

Let us consider the general Mixed ARMA processes (MARMA) —causal, noncausal, invertible or non-invertible— defined as the stationary solution of the following recursive equation:

$$\psi(F)\phi(B)X_t = \theta(F)H(B)\epsilon_t \tag{1}$$

where F (resp. $B = F^{-1}$) denote the forward (resp. backward) operator, and ψ, ϕ, θ and H are polynomials of degrees p, q, r and s with roots outside the unit circle. We assume that $\psi(z) = 1 - \psi_1 z - \cdots - \psi_p z^p$, and $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^q$. Equation (1) admits a unique stationary solution, called a $MARMA(p, q, r, s)$, if $\phi(z) \neq 0$ and $\psi(z) \neq 0$ for all $\mid z \mid \leq 1$, and if ψ (resp. ϕ) has no common root with θ (resp. H). If $H = \theta = 1$, the process simplifies to a MAR(p,q) one. The solution is said to be non-causal if $p \geq 1$.

By the stationarity of the MARMA(p,q,r,s) we can re-write the solution of (1) as an infinite two-sided moving average process (MA):

$$X_t = \sum_{i=-\infty}^{+\infty} a_k \epsilon_{t-k}, \tag{2}$$

where $\sum_{k\in\mathbb{Z}} |a_k| < \infty$ for some $s \in (0, \alpha) \cap [0, 1]$.[1] Besides, the innovation process ϵ_t is i.i.d. non-gaussian, which is a sufficient condition for the identification of the parameters (see Rosenblatt, 2000). The literature considers either α-stable or t-distributed innovations, the choice of one or another conditioning the selection of the forecasting algorithm. According to Fries and Zakoian (2019), in the case of t-distributed errors there are very few theoretical results available in the literature for the predictive density, which prevents a complete characterisation of the forecasting abilities of a given method.

Fortunately, this is not the case of the α-stable distribution. Indeed, following **?**, when $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{S}(\alpha, \beta, \sigma, \mu)$, $\alpha \neq 1$, $\beta \in [-1, 1]$, $\sigma > 0$, one may rely on the theoretical results from Proposition 2 of Fries (2021) and derive closed-form formulas for higher-

[1] In the particular case of $\alpha = 1, \beta = 0$ the condition on the parameters is $\sum_{k\in\mathbb{Z}} |a_k| \, |ln \, |a_k|| < \infty$.

order conditional moments (up to order 4) of two-sided MA(∞) α-stable processes, i.e. $\mathbb{E}[X_{t+h}|X_t = x]$, for $p \in \{1, \dots, 4\}$. They are based on the joint predictive density of $(X_{t+h}, X_t)'$, which has been shown to be a multivariate α-stable process[2]. Proposition 1 below summarizes these results.

Proposition 1. *Let $p \in \{1, 2, 3, 4\}$. $\mathbb{E}[X_{t+h}^p|X_t]$, exist and they depend on $\sigma_1^\alpha, \beta_1, \kappa_p, \lambda_p$ with $\alpha \in (0, 2)\backslash\{1\}$ and a family of functions*

$$\mathcal{H}(n, \theta; x) = \int_0^{+\infty} e^{-\sigma_1^\alpha u^\alpha} u^{n(\alpha-1)} \Big(\theta_1 \cos(ux - \alpha\beta_1\sigma_1^\alpha u^\alpha) + \theta_2 \sin(ux - \alpha\beta_1\sigma_1^\alpha u^\alpha)\Big) du,$$

that can be defined for $n \in \mathbb{N}$, $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$, $x \in \mathbb{R}$ such that, based on Fries (2021) and Samorodnitsky et al. (1996), for $a = tg(\pi\alpha/2)$, one can obtain

$$\begin{aligned}
\mathbb{E}[X_{t+h}|X_t = x] &= \kappa_1 x + \frac{a(\lambda_1 - \beta_1\kappa_1)}{1 + a^2\beta_1^2} f_1(x, a, \beta_1, \mathcal{H}) \\
\mathbb{E}[X_{t+h}^2|X_t = x] &= \kappa_2 x^2 + \frac{ax(\lambda_2 - \beta_1\kappa_2)}{1 + a^2\beta_1^2} f_1(x, a, \beta_1, \mathcal{H}) - f_2(x, a, \sigma_1^\alpha, \mathcal{H}) \\
\mathbb{E}[X_{t+h}^3|X_t = x] &= \kappa_3 x^3 + \frac{ax^2(\lambda_3 - \beta_1\kappa_3)}{1 + a^2\beta_1^2} f_1(x, a, \beta_1, \mathcal{H}) - f_2(x, \alpha, \sigma_1^\alpha, \mathcal{H}) - f_3(x, \alpha, \sigma_1^\alpha, \mathcal{H}) \\
\mathbb{E}[X_{t+h}^4|X_t = x] &= \kappa_4 x^4 + \frac{a(\lambda_4 - \beta_1\kappa_4)}{1 + a^2\beta_1^2} f_1(x, a, \beta_1, \mathcal{H}) - f_2(x, \alpha, \sigma_1^\alpha, \mathcal{H}) \\
&\quad - f_3(x, \alpha, \sigma_1^\alpha, \mathcal{H}) - f_4(x, \alpha, \sigma_1^\alpha, \mathcal{H})
\end{aligned}$$

where the functions $f_p(.)$ are detailed in Fries (2021), $\sigma_1^\alpha = \sigma^\alpha \sum_{k\in\mathbb{Z}} |a_k|^\alpha$, $\beta_1 = \beta \frac{\sum_{k\in\mathbb{Z}} |a_k|^{<\alpha>}}{\sum_{k\in\mathbb{Z}} |a_k|^\alpha}$, $\kappa_p = \frac{\sum_{k\in\mathbb{Z}} |a_k|^\alpha (\frac{a_{k-h}}{a_k})^p}{\sum_{k\in\mathbb{Z}} |a_k|^\alpha}$ and $\lambda_p = \beta \frac{\sum_{k\in\mathbb{Z}} |a_k|^{<\alpha>} (\frac{a_{k-h}}{a_k})^p}{\sum_{k\in\mathbb{Z}} |a_k|^\alpha}$ with $y^{<\alpha>} = sign(y)|y|^\alpha$.

Most importantly, Fries (2021) shows that the asymptotic bubble behaviour of the process is much simpler to characterize. In fact, during severe bubble periods, i.e. (conditional on an extreme X_t), the $\mathcal{H}$ functions in Prop. 1 vanish and the closed-form formulas for the true predictive density of two-sided MA(∞) α-stable processes not only exist, but they are free of any numerical approximation. Proposition 2 below characterises this asymptotic bubble behavior.

Proposition 2. *The asymptotic bubble behavior (extreme quantiles) of the conditional*

[2] see Fries (2021); Samorodnitsky et al. (1996)

moments of the predictive density for a two-sided MA(∞) process is given by:

$$x^{-p}\mathbb{E}[X_{t+h}^{p}|X_t = x] \underset{x\to\infty}{\to} \frac{\kappa_p + \lambda_p}{1+\beta_1} \text{ and } x^{-p}\mathbb{E}[X_{t+h}^{p}|X_t = x] \underset{x\to-\infty}{\to} \frac{\kappa_p - \lambda_p}{1-\beta_1} \tag{3}$$

where $\sigma_1^{\alpha}, \beta_1, \kappa_p, \lambda_p$ are given in Prop.1.

The α-stable framework, and in particular its tails, appears hence to be the most favorable setup to accurately evaluate the forecasting abilities of competing forecasting algorithms. Note that this distributional assumption is of little importance in practice for our approach as the statistical learning algorithms that we introduce in section 3 make no assumption about the probabilistic properties of the process. At the same time, the forecasting algorithms proposed by Lanne and Saikkonen (2011) and Gourieroux and Jasiak (2016) are semi-parametric, and most importantly they have been found to perform least well in the tails of the distribution than in the center.

3 Conditional density forecasting with machine-learning algorithms

Our contribution to this literature consists mainly in adapting machine-learning algorithms to approximate well the true theoretical predictive densities of MARMA processes. Two main methods are considered: K nearest neighbors (hereafter KNN) and random forest (hereafter RF) for their well-known good forecasting abilities in a variety of classification or regression applications and especially because they can provide nonparametric conditional density estimates in a simple way through kernel density estimation.

Let us define by z_j, $j = \{1, 2, \ldots, k\}$ a sample of observations corresponding to the one-period-ahead values of the K-nearest neighbours of an observation x from the random variable X with bounded continuous density $f(x)$ and let $f(. \mid .)$ be the conditional density of the stochastic process x_t. Then the K-nearest neighbour estimate of the conditional

density $f(z|x)$ is given by

$$\hat{f}^k_{r_k}(z \mid x) = \frac{1}{kr_k} \sum_{i=1}^{k} \tilde{K}(\frac{z - z_i}{r_k}),$$

where the bandwidth $r_k = r_k(x) = \min(k, \{|x - Xj|, \text{ for } j = 1, \ldots, K\})$ is an Euclidian distance between x and the $k - th$ nearest neighbor of x among all X_j for k satisfying $k \to \infty, \frac{k}{T} \to 0$ as the sample size $T \to \infty$. Besides $\tilde{K}$ is a kernel function which satisfies $\int \tilde{K}(x)\mathrm{d}x = 1$.

At the same time, the random forest conditional density estimate is given by the average of the conditional densities of each of the N trees in the forest

$$\hat{f}_{r_k}(z \mid x) = \frac{1}{N} \sum_{j=1}^{N} \hat{f}_{r_k,j}(z \mid x),$$

with

$$\hat{f}_{r_k,j}(z \mid x) = \frac{1}{n_{F_i} r_k} \sum_{i=1}^{n_{F_i}} \tilde{K}(\frac{z - z_i}{r_k}),$$

where F_i denotes the leaf to which a new observation x is assigned in tree i and n_{F_i} is the number of observations in leaf F_i. Similar to the KNN case, z is the vector of one-period-ahead values of the observations x that belong to a specific leaf F_i.

These algorithms will then be trained to learn the full probability density $f(x_{t+h} \mid x_t)$ by following the approach of Dalmasso et al. (2020) and by relying on non-parametric resampling techniques to handle the tails of the distribution.

Subsequently, we evaluate the performance of the proposed approaches. On one hand, we rely on absolute evaluation criteria to assess the forecasting abilities of machine learning algorithms. As discussed in section 2, the MARMA framework with α-stable errors is a natural benchmark for this analysis as the first four moments of the conditional distribution are available in closed form and they completely characterize this distribution. If these moments are well forecasted, then the entire conditional density is. In practice, the

predictive conditional moments are estimated from $\hat{f}$ using Simpson integral algorithm

$$E\left[X_{t+h}^{p} \mid X_t = x\right] = \int_{-\infty}^{+\infty} x^p \hat{f}_{X_{t+h}|X_t}(x \mid x_t)dx.$$

For its simplicity, the average Euclidian distance (weighted or not by computation time) is retained as the evaluation criterion. On the other hand, the forecasting abilities of the machine-learning algorithms considered are compared by relying on Giacomini and White (2006)'s test of conditional predictive abilities.

4 Monte Carlo experiments

In this section we analyse the forecasting abilities of the ML algorithms based on simulated data. We use Lanne and Saikkonen (2011)'s decomposition method to design two Monte-Carlo experiments each including N=100 trajectories of M=10^8 observations. The first one relies on a general MARMA(1,1,1,1) data-generating process defined by

$$(1 - 0.9F)(1 + 0.3B)X_t = (1 + 0.4F)(1 - 0.3B)\epsilon_t, \tag{4}$$

while the second one uses a simple MAR(0,1) process defined as

$$(1 - 0.9F)X_t = \epsilon_t. \tag{5}$$

$\epsilon_t \sim S(1.8, 0.5, 0.2, 10)$ in both cases. Besides, the MAR(0,1) in Eq 5 is a restricted model that allows one to compare the forecasting performance of our general algorithms to that of less general existing forecasting approaches proposed by Lanne and Saikkonen (2011) and Gourieroux and Jasiak (2016).

The two model specifications are realistic, (see Fries, 2021; Hecq and Voisin, 2021; De Truchis et al., 2023) and the parameters verify the necessary conditions for the existence of moments (see Prop. 1). Within this setup, the theoretical distribution of X_t is defined by 69.85 as the quantile at 0.05% and 82.12 as the quantile at 99.95% when it follows a

MARMA(1,1,1,1), and if X_t following Eq 5, and by 94.50 as the quantile at 0.05% and 108.25 as the quantile at 99.95% when it is generated by a MAR(0,1) process.

To choose the hyper-parameters of the machine-learning algorithms, we rely on 50-fold cross-validation and test several bandwidths, $r_k = \{0.01, 0.05, 0.1, 0.5\}$, for the kernel density function. In the RF case we consider a number of trees between 10 and 1000 (with 50, 100 and 500 as intermediary values) of depth 10, 50, 100, 500, 1000 and 5000, that was measured by the maximum number of observations per leaf. At the same time, the KNN algorithm was tested with 10, 20, 50, 100, 150, 200 and 500 neighbors.

5 Simulation results

In this section we discuss the KNN and RF Monte-Carlo results when the hyperparamers are chosen by cross-validation among the set of competitors described in the previous Section. To be more precise, r_k is set to 0.1 in both cases, KNN relies on 100 neighbours and RF uses 500 trees of maximal depth 100.

5.1 Resampling problem

As a preliminary step, we study the ability of the two algorithms to learn the pattern of the predictive density of mixed causal models.

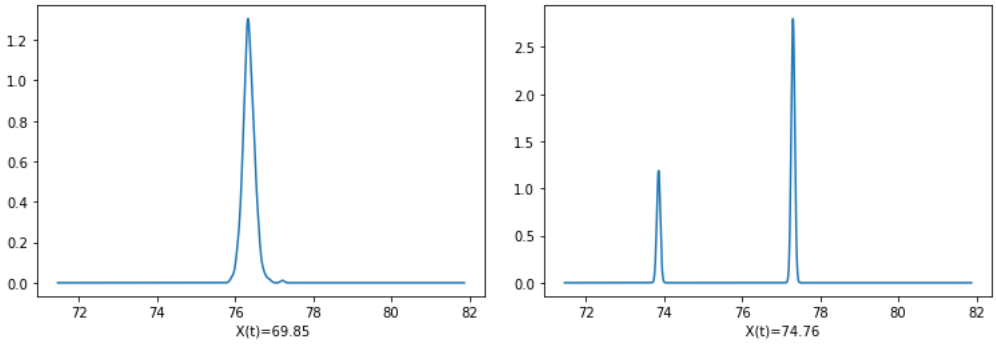
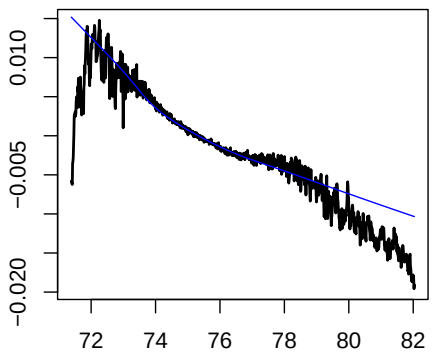


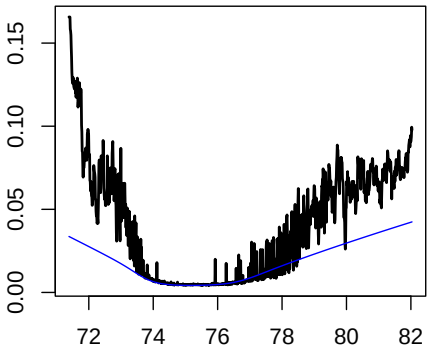
Figure 1: RF learned densities for MARMA(1,1,1,1) for a forecasting horizon $h = 1$

Figure 1, displays a typical learning density obtained by RF for a forecasting horizon $h = 1$. The left panel of the figure corresponds to a conditioning value $X_t = 69.85$, which

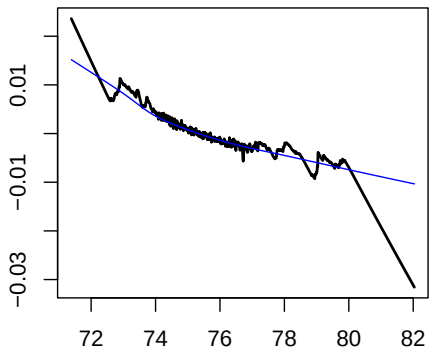
is very far in the left tail of the distribution, while the right panel considers a conditioning value that is in the middle of the distribution, i.e. $X_t = 74.76$. A bimodal predictive density arises in the second case, which is well known in the literature of mixed causal models (see e.g. Gourieroux and Jasiak, 2016; Fries, 2021). This is not the case of the tail conditioning value, but still, the maximum probability is obtained for a value of X_{t+1} which is around 76 and lies in the middle of the distribution just as in the right panel. The results suggest that RF algorithms are able to learn this type of pattern and that they are good candidates for our forecasting problem.



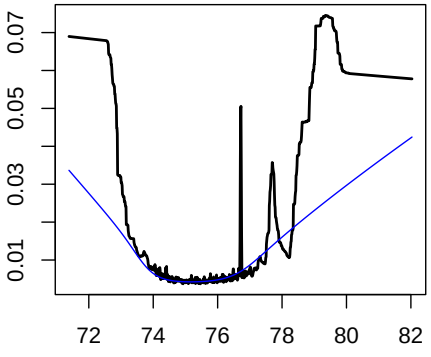
(a) KNN: ($\mathbb{E}\left[R_{t+h} \mid X_t\right]$)



(b) KNN: ($\mathbb{E}\left[R_{t+h}^2 \mid X_t\right]$)



(c) RF: ($\mathbb{E}\left[R_{t+h} \mid X_t\right]$)



(d) RF: ($\mathbb{E}\left[R_{t+h}^2 \mid X_t\right]$)

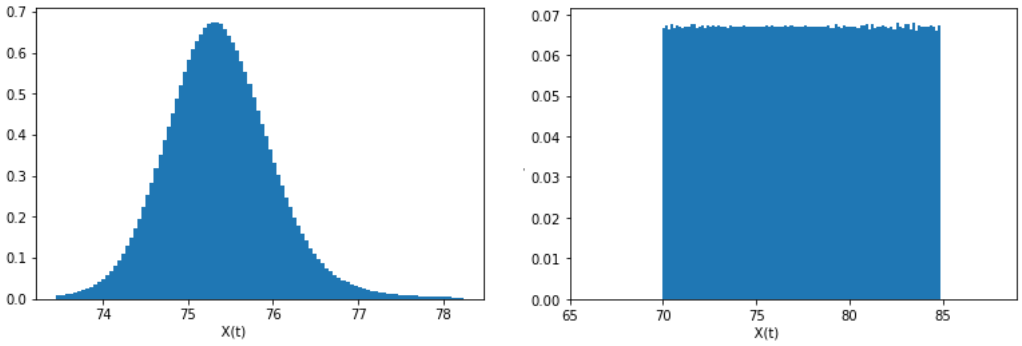
Figure 2: Conditional expectation and standard deviation of R_{t+h}

Notes: The true conditional moments are in blue, while the KNN and RF moment estimations are in orange. The conditioning values are $X_t = x \in (70, 82)$ (99.9% of the probability mass of the marginal distribution of X_t is supported on (70,82)).

To simplify the presentation, in the following, the results will focus on the first two moments of the growth-rate $R_{t+h} = (X_{t+h} - X_t)/X_t$. Figure 2 shows that the estimated moments are very close to the theoretical ones in the center of the conditional distribution but their departures are significant in the tails regardless of the algorithm used. Similarly to the numerical algorithms available in the literature, the machine-learning techniques hence seem to do a poor job during bubble events (see Hecq and Voisin, 2021).

In fact, one easily notices that this problem comes from the fact that despite the simulated trajectory includes a large number of observations (10^8), very few are in the tails of the distribution (see the left panel of Figure 3) and correctly learning the joint density $f(X_{t+h} \mid X_t)$ in this case is impossible without appropriate adjustments. We solve this issue

Figure 3: Histograms of simulated data without and with resampling



by relying on a nonparametric resampling method which is based on a simple principle: we discretize our sample by splitting it into N intervals from quantile 0.0005 to quantile 0.9995 with a step of 0.01.[3] Let us denote by q_n the inferior bound of the n^{th} intervals, with $n \in \{0, 1, \ldots, N\}$. For each interval $[q_n, q_{n+1}]$, we sample with replacement $M = 1000$ times the observed values of the couple (X_t, X_{t+h}) by conditioning on $X_t \in [q_n, q_{n+1}]$. The resampled conditional distribution is displayed in the right panel of Figure 3. Hereafter, we label "resampled random forest (R-RF)" and "resampled K-near neighbour (R-KNN)" the RF and KNN algorithms trained on a resampled trajectory.

[3] A robustness check with respect to the choice of N has been performed.

5.2 Resampling-based ML algorithms for MARMA(1,1,1,1) processes

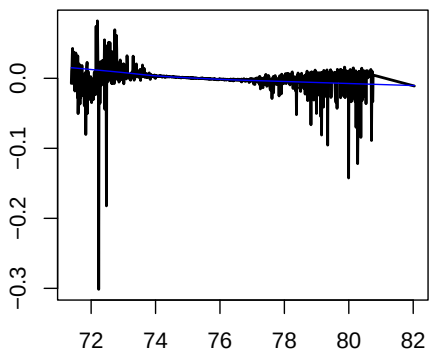
In this subsection we analyze the forecasting abilities of the standard and resample-based ML algorithms in the case of MARMA(1,1,1,1) processes. Figure 4 shows that the resampling method corrects the estimation of first and second order moments when the conditioning is located in the extreme quantiles of the distribution. Note however that the ML algorithms and their resample-based versions are subject to a bias-variance tradeoff. Indded, resampling reduces bias when conditioning on distribution's tails, but this comes with the price of an increase in the variability of these forecasts. This visual insight is confirmed by the Euclidian distance measure between the true density and the learned density reported in Table 1 for the ML algorithms with and without resampling. This goodnes-of-fit measure is computed for the first four conditional moments over all conditioning values of X_t and highlights the superiority of the resample-based techniques. Note also that, as expected, for a given algorithm, the estimates of the higher order moments are noisier than the first two.

To check whether the differences between the moments estimated with the ML algorithms and the true moments in Proposition 1 are significantly different, we perform Giacomini and White (2006) test of conditional predictive ability for each conditional moment. Table 2 displays the results.

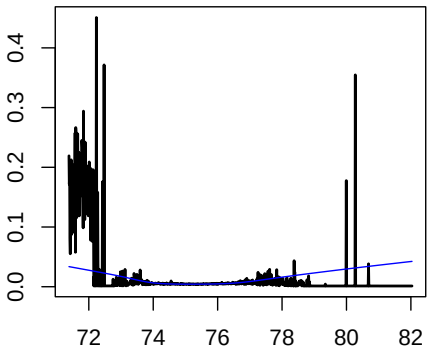
Table 1: Euclidean distance between theoretical and estimated conditional moments of the growth-rate of X_t

	Expected Value	Variance	Skewness	Kurtosis
KNN	0.0020	0.0222	10.2826	292.7744
RF	0.0031	0.0168	9.1577	237.3197
R-KNN	0.0091	0.0247	3.1012	29.3427
R-RF	0.0035	0.0086	1.4765	21.3860

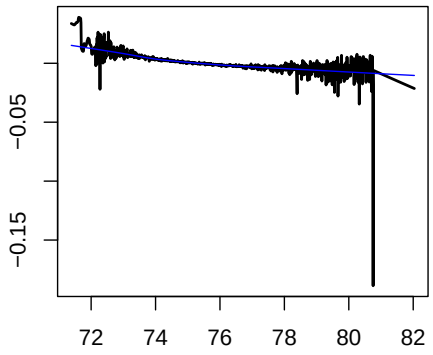
The KNN seems to be a good algorithm to forecast the conditional expected value, whereas R-RF algorithm appears to have the best predictive abilities for the other three moments according to Giacomini and White (2006) test. However, table 3 shows that the latter is also the most time-consuming. A simple tradeoff between parcimony (in terms



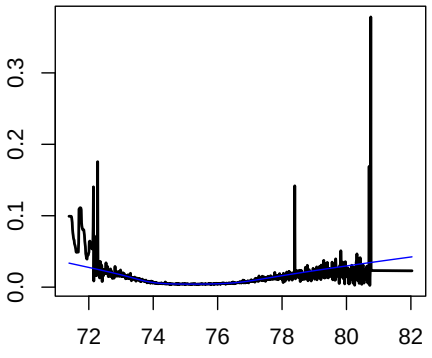
(a) R-KNN: ($\mathbb{E}\left[R_{t+h} \mid X_t\right]$)



(b) R-KNN: ($\mathbb{E}\left[R_{t+h}^2 \mid X_t\right]$)



(c) R-RF: ($\mathbb{E}\left[R_{t+h} \mid X_t\right]$)



(d) R-RF: ($\mathbb{E}\left[R_{t+h}^2 \mid X_t\right]$)

Figure 4: Resample-based conditional expectation and standard deviation of R_{t+h}

Notes: The true conditional moments are in blue, while the R-KNN and R-RF moment estimations are in black. The conditioning values are $X_t = x \in (70, 82)$ (99.9% of the probability mass of the marginal distribution of X_t is supported on (70,82)).

Table 2: Conditional moments forecast comparison (Giacomini and White (2006) test)

Expected value			
	RF	R-KNN	R-RF
KNN	-9.1572^{***}	-14.0274^{***}	-7.5829^{***}
RF		-11.3622^{***}	-2.0368^{*}
R-KNN			10.7459^{***}
Variance			
	RF	R-KNN	R-RF
KNN	10.4680^{***}	-2.5082^{***}	21.3040^{***}
RF		-7.0247^{***}	13.7922^{***}
R-KNN			14.2770^{***}
Skewness			
	RF	R-KNN	R-RF
KNN	3.6997^{***}	15.3169^{***}	29.4514^{***}
RF		13.5900^{***}	29.6380^{***}
R-KNN			4.1561^{***}
Kurtosis			
	RF	R-KNN	R-RF
KNN	2.9370^{***}	-0.2866	16.2250^{***}
RF		13.9211^{***}	16.1859^{***}
R-KNN			1.9248

The test is carried out for each pair of methods. Asterisks ***, **, and * indicate significance at the 99.9%, 99% and 95% level, respectively. A positive (negative) and significant test-statistic leads to the rejection of the null of equal forecasting abilities in favor of the algorithm in column (row).

Table 3: Execution time

Methods	Time (in seconds)
KNN	40.19
RF	187.34
R-KNN	508.01
R-RF	653.89

of computational time in this context) and performance, would consist in considering an evaluation criterion given by the Euclidean distance weighted by the calculation time. This simple exercise designates the KNN method as the best equally-weighted arbitrage between performance and execution time.

Table 4: Euclidean distance weighted by Execution time

	Expected Value	Variance	Skewness	Kurtosis
KNN	8.04	89.22	413.15	11 766.60
RF	58.06	314.73	1 715.6	44 459.47
R-KNN	46.63	193.04	9 347.38	14 906.38
R-RF	22.29	562.34	9 654.68	13 980.17

5.3 Resampling-based ML algorithms for MAR(0,1) processes

The case of MAR(0,1) processes is particularly interesting to study as it facilitates the comparison of the ML approaches with Lanne and Saikkonen (2011)'s simulation-based forecasting method in terms of conditional expectation forecast.

The Euclidian distance measure between each conditional forecast and the theoretical moment reported in Table 5 increases for the higher-order moments but it is at least partially compensated by the use of the resampling technique. RF registers the smallest goodness-of-fit for the conditional expectation, while Lanne and Saikkonen (2011)'s is among the highest ones. This goes along the lines of the visual intuition in Figure 6 which shows that Lanne and Saikkonen (2011)'s conditional expectation forecast is much more volatile than that of RF and KNN. The same bias-variance tradeoff in quatiles can be noticed for the ML algorithms and their resample-based versions as in the previous sections. Indded, resampling allows one to better forecast when conditioning on tail values, but with the price of an increase in the variability of these forecasts.

Most importantly, Giacomini and White (2006)'s test in Table 6 further supports that RF (R-RF) registers the best forecasting abilities for low(high)-order moments and accounting for execution time does not change this outcome (see Table 7). Furthermore, the comparison with Lanne and Saikkonen (2011)'s approach in Tables 6 and 7 clearly shows that the forecasting abilities of their simulation-based algorithm are well below those of all the ML approaches while being much more time-consuming than the latter methods.

Table 5: Euclidean distance between theoretical and estimated conditional moments of the growth-rate of X_t

	Expected Value	Variance	Skewness	Kurtosis
KNN	0.0048	0.0111	16.5257	1 691.6394
RF	0.0040	0.0168	1.7530	66.4564
R-KNN	0.0107	0.0218	1.9337	501.6693
R-RF	0.0077	0.0181	1.2535	9.0693
Lanne and Saikkonen (2011)'s	0.0111	x	x	x

Table 6: Conditional moments forecast comparison (Giacomini and White (2006) test)

Expected value			
	RF	R-KNN	R-RF
KNN	8.2694^{***}	-23.7875^{***}	-7.8287^{***}
RF		-24.7093^{***}	-9.3367^{*}
R-KNN			8.1813^{***}
Variance			
	RF	R-KNN	R-RF
KNN	0.9808	-11.0256^{***}	-4.9146^{***}
RF		-10.8941^{***}	-4.8536^{***}
R-KNN			4.5387^{***}
Skewness			
	RF	R-KNN	R-RF
KNN	18.8893^{***}	17.7174^{***}	20.3599^{***}
RF		-0.41806	2.8850^{***}
R-KNN			1.7094^{*}
Kurtosis			
	RF	R-KNN	R-RF
KNN	13.1257^{***}	4.4948^{***}	13.8827^{***}
RF		-1.8071^{*}	2.5866^{***}
R-KNN			2.0551^{***}

The test is carried out for each pair of methods. Asterisks ***, **, and * indicate significance at the 99.9%, 99% and 95% level, respectively. A positive (negative) and significant test-statistic leads to the rejection of the null of equal forecasting abilities in favor of the algorithm in column (row).

Table 7: Execution time

Methods	Time (in seconds)
KNN	16.27
RF	79.81
R-KNN	518.18
R-RF	606.84
Lanne and Saikkonen (2011)'s	3501.60

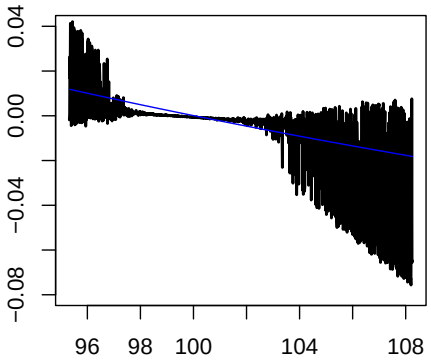
Table 8: Conditional moments forecast comparison (Giacomini and White (2006) test)

	KNN	RF	R-KNN	R-RF
Lanne and Saikkonen (2011)	18.9517***	19.2825***	1.1527*	6.6165***

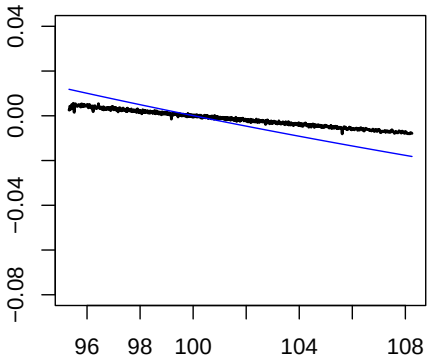
See note to Table 6.

6 Conclusion

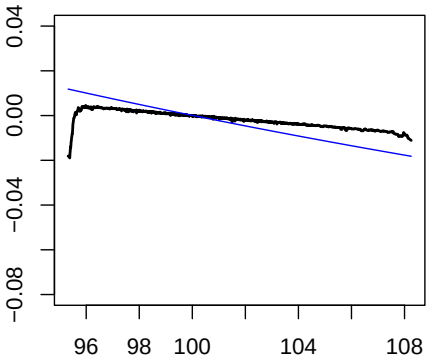
In this paper we propose a new forecasting algorithm based on machine learning techniques to forecast general non-causal and mixed-causal model MARMA. We proposed an extension of the Lanne and Saikkonen (2011), procedure for α-stable law. In a Monte Carlo simulation exercise, we show that our approach outperform the standard approaches available in the literature, in terms of accuracy and computational time, and that they provide an interesting approximation to the true theoretical prediction densities.



(a) Lanne and Saikkonen (2011)

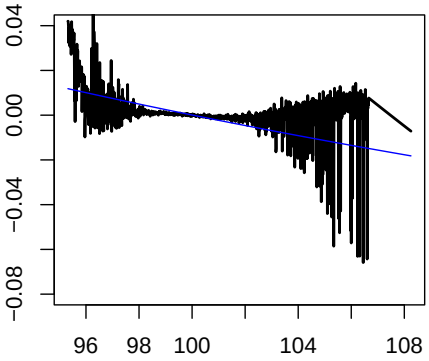


(b) RF

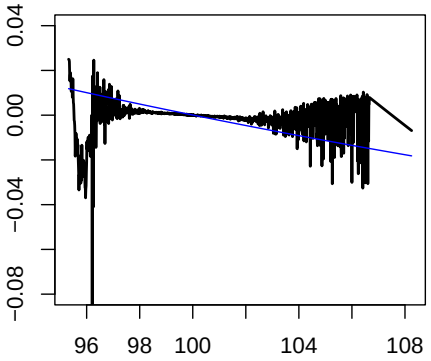


(c) KNN

Figure 5: Conditional expectation of R_{t+h} for a MAR(0,1) process



(a) R RF



(b) R KNN

Figure 6: Conditional expectation of R_{t+h} for a MAR(0,1) process

Notes: The true conditional moments are in blue, while the R-KNN and R-RF moment estimations are in black. The conditioning values are $X_t = x \in (95, 108)$ (99.9% of the probability mass of the marginal distribution of X_t is supported on (95,108)).

References

Cavaliere, G., Nielsen, H.B., Rahbek, A., 2020. Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. Journal of Business & Economic Statistics 38, 55–67.

Dalmasso, N., Pospisil, T., Lee, A., Izbicki, R., Freeman, P., Malz, A., 2020. Conditional density estimation tools in python and r with applications to photometric redshifts and likelihood-free cosmological inference. Astronomy and Computing 30, 100362.

De Truchis, G., Dumitrescu, E.I., Fries, S., Thomas, A., 2023. Bet on a bubble asset ? an optimal portfolio allocation strategy. mimeo .

Fries, S., 2021. Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds. Journal of Business & Economic Statistics 0, 1–21.

Fries, S., Zakoian, J.M., 2019. Mixed causal-noncausal ar processes and the modelling of explosive bubbles. Econometric Theory 35, 1234–1270.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Gourieroux, C., Hencic, A., Jasiak, J., 2021. Forecast performance and bubble analysis in non-causal mar(1,1) processes. Journal of Forecasting 40, 301–326.

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

Gourieroux, C., Jasiak, J., 2018. Misspecification of noncausal order in autoregressive processes. Journal of Econometrics 205, 226–248.

Gourieroux, C., Jasiak, J., 2022. Nonlinear forecasts and impulse responses for causal-noncausal (s)var models. `arXiv:2205.09922`.

Gourieroux, C., Zakoian, J.M., 2017. Local explosion modelling by non-causal process. Journal of the Royal Statistical Society Series B 79, 737–756.

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

Hecq, A., Voisin, E., 2021. Forecasting bubbles with mixed causal-noncausal autoregressive models. Econometrics and Statistics 20, 29–45.

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

Rosenblatt, M., 2000. Gaussian and Non-Gaussian Linear Time Series and Random Fields.

Samorodnitsky, G., Taqqu, M.S., Linde, R., 1996. Stable non-gaussian random processes: stochastic models with infinite variance. Bulletin of the London Mathematical Society 28, 554–555.