Forecast Performance of Noncausal Autoregressions and the Importance of Unit Root Pretesting

Frédérique Bec, Heino Bohn Nielsen

Forecast Performance of Noncausal Autoregressions and the Importance of Unit Root Pretesting*

Frédérique Bec† Heino Bohn Nielsen‡

Abstract

Based on large a simulation study, this paper investigates which strategy to adopt in order to choose the most accurate forecasting model for Mixed causal-noncausal AutoRegressions (MAR) data generating processes: always differencing (D), never differencing (L) or unit root pretesting (P). Relying on recent econometric developments regarding forecasting and unit root testing in the MAR framework, the main results suggest that from a practitioner's point of view, the P strategy at the 10%-level is a good compromise. In fact, it never departs too much from the best model in terms of forecast accuracy, unlike the L (respectively D) strategy when the DGP becomes very persistent (respectively less persistent).

Keywords: Mixed causal-noncausal AutoRegressions, Forecasting, Unit root pretest.
JEL Codes: C22, C53.

*We are grateful to participants at the 6th International Workshop on Financial Markets and Nonlinear Dynamics (Paris, 2022) and at the 4th Quantitative Finance and Financial Econometrics conference (Marseille, 2022) for helpful comments. F. Bec acknowledges financial support from the Labex MMEDII (ANR11-LBX-0023-01). H.B. Nielsen acknowledges financial support from the Institute for Advanced Studies at CY Cergy Paris University and the Danish Council for Independent Research, DSF Grant 015-00028B.

†CY Cergy Paris University, CNRS, THEMA, and CREST, France. Email: frederique.bec@cyu.fr. Corresponding author.

‡Department of Economics, University of Copenhagen, Denmark. Email: heino.bohn.nielsen@econ.ku.dk.

1 Introduction

Regarding time series forecasting, the question about using a model for the series in levels or in first differences can be traced back to Box and Jenkins [1976]. In their popular book, these authors recommend to use the model in levels unless the root of the process to forecast is close to unity, in which case the model in first differences would achieve the best forecast accuracy. The introduction of unit root tests by Dickey and Fuller [1979] and their application in an influential paper by Nelson and Plosser [1982] gave rise to a renewed interest in this question. Indeed, from a large set of long historical time series for the U.S. economy, Nelson and Plosser [1982] conclude that the null hypothesis that these series contain a unit root cannot be rejected for most of them. Yet, as emphasized in Dickey, Bell, and Miller [1986], forecasts of a unit root process are very different from forecasts of a stationary process, also when the latter is strongly persistent. Hence, beside "always difference" or "never difference" (denoted respectively D and L hereafter) forecasting strategies, a third option has emerged: unit-root pretesting (denoted P) and difference or not accordingly. From Monte-Carlo simulations, evidence in favour of pretesting strategy was found for the linear autoregressive class of models by e.g. Campbell and Perron [1991], Stock [1996], Stock and Watson [1999] or Diebold and Kilian [2000], especially for roots close to unity.

In this paper, we answer the same question for the class of mixed causal-noncausal autoregressive (MAR) models. These models, introduced decades ago in statistics, have acquired growing popularity in economics and finance over the past few years. Development of their econometric theory as well as applications to macroeconomic or financial time series modelling and forecasting can be found in Lanne and Saikkonen [2011], Lanne, Luoma, and Luoto [2012], Lanne and Saikkonen [2013], Hencic and Gouriéroux [2015], Gouriéroux and Zakoian [2015], Gouriéroux and Jasiak [2016], Gouriéroux and Zakoian [2017], Fries and Zakoian [2019], Cavaliere, Nielsen, and Rahbek [2020] or Hecq and Voisin [2023], among others. The interest in MAR models, which allow for dependence on both the past and the future — unlike the well-known backward-looking autoregression which rules out dependence on future observations — stems mainly from two of their characteristics. First, they are able to capture epochs of bubble build-up and burst. Second, as noticed in a previous paper [Bec, Nielsen, and Saïdi, 2020], mixed causal-noncausal autoregressive models could prove very useful for forecasting as they can be viewed as very parsimonious representations of more complex nonlinear DGPs (see also Gouriéroux and Zakoian [2017] on this point). Due to their dependence on future observations, specific forecasting methods have been developed for MAR models by Lanne, Luoto, and Saikkonen [2012] and

Gouriéroux and Jasiak [2016]. Similarly, a specific test for a unit root in the backward autoregressive polynomial has been proposed by Saikkonen and Sandberg [2016].

The goal of this paper is twofold: *i)* to compare the performance of the forecast methodologies proposed by Lanne, Luoto, and Saikkonen [2012] and Gouriéroux and Jasiak [2016] and *ii)* to evaluate the relevance of unit root pretesting, P, for these MAR models' forecasting performance, compared to the D and L strategies. To this end, the recent developments cited above are used in a large simulation study. Within this methodological framework, a variety of degrees of persistence, of forecast horizons, of sample sizes as well as of levels of the test are under scrutiny.

After a brief presentation of the MAR model in Section 2, Section 3 exposes the amended version of Saikkonen and Sandberg [2016] unit root test proposed in a previous work [Bec, Nielsen, and Saïdi, 2020] in order to circumvent likelihood bimodality issues involved by the estimation of the MAR models. Then, Section 4 describes the forecasting methods of Lanne, Luoto, and Saikkonen [2012] and Gouriéroux and Jasiak [2016]. Section 5 reports results of preliminary simulation exercises which are used to compare the effectiveness of both methods and to fine-tune the settings of the main simulation study reported in Section 6. Section 7 provides an empirical application in terms of the Brent crude oil price, while Section 8 concludes.

2 The MAR Setting

Consider the mixed causal noncausal autoregression, MAR(1,1), as given by

$$\Phi(L)\Psi(L^{-1})y_t = (1-\phi L)(1-\psi L^{-1})y_t = \epsilon_t, \tag{1}$$

where ϵ_t is assumed i.i.d. with p.d.f. given by $g(\cdot;\theta)$ indexed by parameters in θ. For the data generating process, we consider the case $-1 < \phi \leq 1$ and $-1 < \psi < 1$, allowing a unit root in the causal polynomial, $\Phi(L)$, see Saikkonen and Sandberg [2016] for a discussion.

Following Lanne and Saikkonen [2011] and Gouriéroux and Jasiak [2016], we define the unobserved component $u_t = (1-\phi L)y_t$ such that $(1-\psi L^{-1})u_t = \epsilon_t$ and

$$u_t = (1-\psi L^{-1})^{-1}\epsilon_t = \sum_{j=0}^{\infty} \psi^j \epsilon_{t+j}. \tag{2}$$

Using the terminology of Gouriéroux and Jasiak [2016], u_t is $y-$causal, as it depends on current and *past* values of y, and $\epsilon-$noncausal, as it depends on current and *future* values of ϵ. For a specified innovation density $g(\cdot;\theta)$, the

parameters in (1) can be estimated from a sample $\{y_t\}_{t=1}^T$ using approximate maximum likelihood,

$$(\hat{\phi}, \hat{\psi}, \hat{\theta}) = \arg\max_{\phi,\psi,\theta} \sum_{t=2}^{T-1} \log g\left((1-\phi L)(1-\psi L^{-1})y_t; \theta\right),$$

conditional on initial and terminal values, y_1 and y_T, see e.g. Breidt et al. [1991], and Lanne and Saikkonen [2011]. Below, we use a Student's $t(0, \sigma^2, \lambda)$ distribution, such that

$$g(\epsilon_t; \sigma^2, \lambda) = \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda}{2})} \left(\pi\lambda\sigma^2\right)^{-\frac{1}{2}} \left(1 + \frac{\epsilon_t^2}{\sigma^2\lambda}\right)^{-\frac{\lambda+1}{2}}, \tag{3}$$

with $\theta = (\sigma^2, \lambda)$ containing the degrees of freedom parameter, λ, and the scale, σ^2.

3 Unit Root Testing

The unit root test pertaining to the hypothesis $\mathcal{H}_0 : \phi = 1$ in the MAR model is considered in Saikkonen and Sandberg [2016]. They show that for $g(\cdot; \theta)$ symmetric, the usual one-sided unit-root test statistic,

$$\tau = \frac{\hat{\phi} - 1}{se(\hat{\phi})}, \tag{4}$$

where $se(\hat{\phi})$ is the square root of the relevant entry in the inverse observed information, has a limiting distribution, $D_\tau(\rho)$, that depends on the nuisance parameter ρ. If $g(\cdot; \theta)$ is the Student's t density in (3), with $\lambda > 2$, it holds that ρ is a function of λ:

$$\rho(\lambda)^2 = \frac{(\lambda-2)(\lambda+3)}{\lambda(\lambda+1)}.$$

For mean-zero data, the critical values are given in Saikkonen and Sandberg [2016] in terms of a response surface approximation, as

$$\begin{aligned} cv_{1\%}(\lambda) &= -2.321 - 0.492\rho(\lambda) + 0.251\rho(\lambda)^2 \\ cv_{5\%}(\lambda) &= -1.639 - 0.495\rho(\lambda) + 0.187\rho(\lambda)^2 \\ cv_{10\%}(\lambda) &= -1.276 - 0.480\rho(\lambda) + 0.131\rho(\lambda)^2, \end{aligned}$$

where $cv_{x\%}(\lambda)$ is the critical value at the x percent level as a function of λ, and, for example:

λ	3	4	5	6	8	10	20	∞
$cv_{1\%}(\lambda)$	-2.515	-2.551	-2.557	-2.560	-2.561	-2.562	-2.562	-2.562
$cv_{5\%}(\lambda)$	-1.965	-1.964	-1.959	-1.956	-1.952	-1.950	-1.948	-1.947
$cv_{10\%}(\lambda)$	-1.693	-1.663	-1.649	-1.642	-1.634	-1.631	-1.627	-1.625

However, Bec et al. [2020] show that a bimodality issue occurs in the Student's t distributed errors case, which frequently leads to interchanged roots. Indeed, the backward root can be estimated as the forward root and *vice versa*. Since the unit root test is based on the estimation of the backward root, this can pose a problem. As a consequence, the estimation strategy proposed by Bec et al. [2020] relies on a grid search procedure in order to characterize the entire likelihood surface and list all local maxima. Then, if there are multiple maxima, the maximum with the backward root higher than the forward root ($\hat{\phi} > \hat{\varphi}$) is selected even if there exists a maximum with a higher likelihood but with $\hat{\phi} < \hat{\varphi} \approx 1$. This choice stems from the fact that a unit root in the forward component would lack reasonable economic interpretation: It would mean that agents look into the infinite future without discounting. Beside, the test statistic is similar to Saikkonen and Sandberg [2016]'s. Hence, this modified version of Saikkonen and Sandberg [2016]'s unit root test will be used in the subsequent analysis.

4 Forecasting

To forecast $\{y_{T+h}\}_{h=1}^{H}$ based on the MAR model, we consider the approaches suggested in Lanne, Luoto, and Saikkonen [2012] and Gouriéroux and Jasiak [2016]. Lanne et al. [2012] start from (2) and simulate future paths $\{\epsilon_{T+k}\}_{k=1}^{K}$ to approximate the conditional expectation $E(y_{T+h} \mid y_T)$ and the c.d.f. Gouriéroux and Jasiak [2016], on the other hand, derive the predictive density $p(u_{T+1}, ..., u_{T+H} \mid y_T)$ and use importance sampling to draw from the predictive distibution of $\{y_{T+h}\}_{h=1}^{H}$ given y_T.

4.1 Lanne, Luoto, and Saikkonen [2012] (LLS)

To derive the point forecast, it is used that

$$E(y_{T+h} \mid y_T) = \phi^h y_T + E(\sum_{i=0}^{h-1} \phi^i u_{T+i+1} \mid y_T), \quad h = 1, 2, ..., H.$$

The infinite sum in (2) is then approximated with a truncated version using K terms, and based on N simulated paths, $\{\epsilon_{T+k}^{(i)}\}_{k=1}^{K}$ for $i = 1, 2, ..., N$, Lanne et al. [2012] suggests to estimate the conditional expectation as

$$\begin{aligned} & E(y_{T+h} \mid y_T) \\ = \; & \phi^h y_T + E(\sum\nolimits_{i=0}^{h-1} \phi^i u_{T+i+1} \mid y_T) \\ \approx \; & \phi^h y_T + E\left(\sum\nolimits_{i=0}^{h-1} \phi^i \sum\nolimits_{k=0}^{K-h} \psi^k \epsilon_{t+h+i+k} \mid y_T\right) \\ \approx \; & \phi^h y_T + \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{i=0}^{h-1} \phi^i \sum_{k=0}^{K-h} \psi^k \epsilon_{t+h+i+k}\right) g\left(\epsilon_T(u_T, \epsilon_{T+1}^{(i)}, ..., \epsilon_{T+M}^{(i)}); \theta\right)}{\frac{1}{N}\sum_{i=1}^{N} g\left(\epsilon_T(u_T, \epsilon_{T+1}^{(i)}, ..., \epsilon_{T+M}^{(i)}); \theta\right)} \end{aligned} \tag{5}$$

where ϵ_T is found as

$$\epsilon_T(u_T, \epsilon_{T+1}^{(i)}, ..., \epsilon_{T+M}^{(i)}) = \hat{u}_T - \sum_{k=1}^{K} \psi^k \epsilon_{T+k}^{(i)}. \tag{6}$$

To get also interval forecasts, we evaluate the c.d.f. over a grid $x \in (x_1, ..., x_G)$ using

$$\frac{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{I}\left(\sum_{i=0}^{h-1} \phi^i \sum_{k=0}^{K-h} \psi^k \epsilon_{t+h+i+k} \leq x\right)\right) g\left(\epsilon_T(u_T, \epsilon_{T+1}^{(i)}, ..., \epsilon_{T+M}^{(i)}); \theta\right)}{\frac{1}{N}\sum_{i=1}^{N} g\left(\epsilon_T(u_T, \epsilon_{T+1}^{(i)}, ..., \epsilon_{T+M}^{(i)}); \theta\right)},$$

similar to (5) where $\mathbb{I}(\cdot)$ is the indicator function.

In the implementation below we choose $K \in [K_{\min}, K_{\max}]$, such that $\psi^K < 0.0001$ with $K_{\min} = H+5$ and $K_{\max} = 200$. For the interval forecast, we use a grid of 1024 equally spaced points in the interval between $\min(\mu_u - 12\sigma_u, u_T - \sigma_u)$ and $\max(\mu_u + 12\sigma_u, u_T + \sigma_u)$, where μ_u and σ_u^2 denote the empirical mean and variance of $\{u_t\}_{t=1}^{T}$. These choices ensure that the interval is wide enough to include most probability mass of the forecast distributions for $H = 10$. Specific quantiles are found using linear interpolation.

4.2 Gouriéroux and Jasiak [2016] (GJ)

To obtain density forecasts for $\{y_{T+h}\}_{h=1}^{H}$ conditional on y_T, GJ suggest to forecast the $\epsilon-$noncausal component $\{u_{T+h}\}_{h=1}^{H}$ given y_T, and use $u_t = (1 - \phi L)y_t$ to construct density forecasts for $\{y_{T+h}\}_{h=1}^{H}$ from forecast densities for $\{u_{T+h}\}_{h=1}^{H}$ and y_T.

To forecast $\{u_{T+h}\}_{h=1}^{H}$, Gouriéroux and Jasiak [2016] rewrite the joint predictive density,

$$\begin{aligned} p(u_{T+1},...,u_{T+H} \mid y_T) &= p(u_{T+1},...,u_{T+H} \mid u_T) \\ &= \frac{p(u_T,u_{T+1},...,u_{T+H})}{p(u_T)} \\ &= \frac{p(u_T,u_{T+1},...,u_{T+H-1} \mid u_{T+H})p(u_{T+H})}{p(u_T)}. \end{aligned}$$

Using estimated versions for $\{u_t\}_{t=1}^{T}$, ψ, and θ and estimating the stationary distributions by sample averages, an estimate of the predictive density is given in closed form

$$\begin{aligned} & \hat{p}(u_{T+1},u_{T+2},...,u_{T+H} \mid \hat{u}_T) \\ = \; & g(u_T - \hat{\psi}u_{T+1};\hat{\theta}) \cdot g(u_{T+1} - \hat{\psi}u_{T+2};\hat{\theta}) \cdots g(u_{T+H-1} - \hat{\psi}u_{T+H};\hat{\theta}) \\ & \times \frac{\sum_{t=1}^{T} g(u_{T+H} - \hat{\psi}\hat{u}_t;\hat{\theta})}{\sum_{t=1}^{T} g(u_T - \hat{\psi}\hat{u}_t;\hat{\theta})}. \end{aligned} \tag{7}$$

The closed form for the joint predictive density in (7) can be used to simulate paths for $\{u_{T+h}\}_{h=1}^{H}$ e.g. using Sampling-Importance-Resampling (SIR), see Rubin [1988].

Below, we implement SIR based on an importance density, defined by a causal AR(1) with Student's $t(0,\sigma^{*2},\lambda^*)$ innovations[1]. We generate candidate forecast paths $\{u_{T+h}^{*(i)}\}_{h=1}^{H}$, for $i = 1,2,...,M$, conditional on $\hat{u}_T$, using the causal AR(1) model

$$u_{T+h}^{*(i)} = \gamma^* u_{T+h-1}^{*(i)} + \sigma^* \eta_{T+h}^{*(i)}, \quad \eta_{T+h}^{*(i)} \sim t(0,1,\lambda^*), \tag{8}$$

with $u_T^{*(i)} = \hat{u}_T$, and use the importance density

$$\hat{q}(u_{T+1}^{*(i)},...,u_{T+H}^{*(i)} \mid \hat{u}_T) = g^*(u_{T+H}^{*(i)} \mid u_{T+H-1}^{*(i)}) \cdots g^*(u_{T+2}^{*(i)} \mid u_{T+1}^{*(i)}) g^*(u_{T+1}^{*(i)} \mid \hat{u}_T),$$

where $g^*(\cdot)$ is the conditional density for the causal AR(1) model depending on $(\gamma^*,\sigma^*,\lambda^*)$.

[1]We have also considered a recursive version of the SIR algorithm, where we resample the pairs $\{y_{T+h-1}^{(j)}, u_{T+h}^{(j)}\}$ at each forecast horizon, $h = 1,2,...,H$, but otherwise use the same approach as below. This is computationally more demanding, but could in principle improve the precision because it does not rely on the fit of entire forecast paths. However, 10,000 Monte Carlo replications of a MAR(1,1) DGP with $\phi = 0.9$, $\psi = 0.6$, $\sigma = 1$, $\lambda = 7$ and $T = 200$ revealed no improvement of the forecasts accuracy compared to the SIR approach of GJ or the LLS method. These additional results are available upon request.

For each candidate path, $\{u_{T+h}^{*(i)}\}_{h=1}^{H}$, we calculate the corresponding importance weight

$$w^{(i)} = \frac{\hat{p}(u_{T+1}^{*(i)}, ..., u_{T+H}^{*(i)} \mid \hat{u}_T)}{\hat{q}(u_{T+1}^{*(i)}, ..., u_{T+H}^{*(i)} \mid \hat{u}_T)}.$$

In the resampling step of the SIR algorithm, we draw N forecast paths from the M candidate paths $\{u_{T+h}^{*(i)}\}_{h=1}^{H}$ $i = 1, 2, ..., M$, with probabilities given by the normalized importance weights

$$p^{(i)} = \frac{w^{(i)}}{\sum_{i=1}^{M} w^{(i)}}.$$

For M and N large, the final sample paths, $\{u_{T+h}^{(j)}\}_{h=1}^{H}$ $j = 1, ..., N$, are draws from the predictive distribution with p.d.f. given by (7).

For the h-step forecast, $h = 1, 2, 3, ..., H$, the forecast paths are updated recursively from

$$y_{T+h}^{(j)} = u_{T+h}^{(j)} + \phi y_{T+h-1}^{(j)}.$$

Point forecasts, denoted $\{\hat{y}_{T+h}\}_{h=1}^{H}$, can be derived using a location measure for the simulated predictive distribution. GJ suggest the pointwise *mode* as the most likely outcome; an alternative would be the *median* or the *mean*.

The effectiveness of the SIR algorithm depends on the chosen importance density, and before discussing the effect of unit root pretesting we consider a small pilot simulation, where we vary $(\gamma^*, \sigma^{*2}, \lambda^*)$ to ensure a sufficiently large variation of the candidate paths.

5 Preliminary Simulations

This Section presents the pilot simulations study which aims to compare the approaches of Lanne et al. [2012] and Gouriéroux and Jasiak [2016] and to guide the choice of *(i)* the importance density of the SIR algorithm, *(ii)* the location measure for the simulated predictive distribution and *(iii)* the forecasting approach[2].

In the simulation below we use B simulated MAR(1,1) time series, $\{y_t^{(b)}\}_{t=1}^{T+H}$ $b = 1, 2, ..., B$, with sample length $T = 200$, and parameters given by $\phi = 0.9$ and $\psi = 0.6$.

[2]We have also explored the influence of the number of simulations. To this end, we have compared forecasts obtained for M and N taking values in $\{50000, 20000, 10000, 5000, 1000\}$. Point forecasts appear to be reliable for quite low number of simulations. For the chosen setting, interval forecasts appear reliable with values of N and M larger than 5000-10000. Accordingly, the subsequent simulation study is conducted for $M = N = 10000$. These additional results are available upon request.

To compare point forecasts, we use the mean absolute deviation[3] from the true value of different implementations for each forecast horizon $h \in \{1, 2, 5, 10\}$, defined as

$$MAD(h) = \frac{1}{B} \sum_{b=1}^{B} \left| \hat{y}_{T+h}^{(b)} - y_{T+h}^{(b)} \right|.$$

We also consider the 90 percent coverage of the density forecasts, i.e. the proportion of cases where the actual value of the time series, y_{T+h}, is included in the 90 percent confidence interval of the forecast. For the pilot simulations, we use $M = N = 10000$ and $B = 10000$ replications unless otherwise mentioned.

In order to choose the importance density of the SIR algorithm, we conduct a simulation study for various values of $(\gamma^*, \sigma^{*2}, \lambda^*)$. First, we consider $\gamma^* \in \{\tilde{\gamma}, 1\}$, where $\tilde{\gamma}$ is the OLS estimate obtained from a causal AR(1) model for $\{\hat{u}_t\}_{t=1}^T$. For the scale we consider $\sigma^{*2} = c\tilde{\sigma}^2$, where $\tilde{\sigma}^2$ is the OLS estimate and where $c \in \{1, 4\}$ potentially inflates the variance. Finally, we take $\lambda^* \in \{\hat{\lambda}, 3, 100\}$, where $\hat{\lambda}$ is the MLE from the MAR model[4]. The forecast accuracy performances of these various calibrations, as measured by the MAD(h) and the 90%-coverage, are reported in Table 1 (columns (1) to (8)), and the following preliminary conclusions emerge:

1. Regarding the location measure, it appears that the mode is always dominated by the mean and the median, the last two ones giving very similar MAD's. This could be due to the well-known noise of the mode estimates. We have considered different versions of mode estimates, and for the present case the maximum of the estimated kernel density (using a Gaussian kernel with 128 steps to 2 percent trimmed data) performs better than the half-sample median, see e.g. Bickel and Frühwirth [2006]. In the subsequent simulation study, we will focus on the *median* which may be more robust for low values of λ.

2. The coverage rates, given in the last four lines of Table 1, are generally excellent indicating that the density forecast may not be too sensitive to the precise implementation.

[3]The MAD is preferred to the more popular (R)MSE because the latter may be sensitive to a few very large realizations of the $|\epsilon|$'s, which can occur when working with fat-tailed distributed errors. The tail dependence of the forecasts distribution is also the reason why Gouriéroux and Jasiak [2016] prefer to measure the central tendency of the forecasts by the "mode" rather than by the more conventional "mean" forecast.

[4]It is well-known that the effectiveness of importance sampling requires that the effective support of the proposal distributions is at least as big as the target distribution.

Table 1: Performance of Implementations of Forecasts.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Gourieroux and Jasiak (2016)								LLS
γ^*		$\tilde{\gamma}$	$\tilde{\gamma}$	$\tilde{\gamma}$	$\tilde{\gamma}$	1	1	1	1	
c		1	4	1	1	1	4	1	1	
λ^*		$\hat{\lambda}$	$\hat{\lambda}$	3	100	$\hat{\lambda}$	$\hat{\lambda}$	3	100	
MAD(1)	Mean	0.935	0.935	0.933	0.934	0.935	0.935	0.934	0.935	0.934
MAD(2)	Mean	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70	1.70
MAD(5)	Mean	3.36	3.36	3.36	3.36	3.35	3.36	3.36	3.36	3.36
MAD(10)	Mean	4.48	4.48	4.48	4.48	4.48	4.48	4.48	4.48	4.47
MAD(1)	Mode	0.948	0.962	0.946	0.949	0.952	0.976	0.957	0.960	
MAD(2)	Mode	1.72	1.74	1.72	1.72	1.73	1.77	1.74	1.75	
MAD(5)	Mode	3.39	3.44	3.40	3.40	3.41	3.48	3.43	3.43	
MAD(10)	Mode	4.53	4.61	4.54	4.54	4.55	4.64	4.59	4.58	
MAD(1)	Median	0.936	0.937	0.934	0.936	0.936	0.937	0.936	0.936	0.935
MAD(2)	Median	1.70	1.70	1.70	1.70	1.70	1.71	1.70	1.71	1.70
MAD(5)	Median	3.36	3.36	3.36	3.36	3.36	3.36	3.36	3.36	3.36
MAD(10)	Median	4.48	4.49	4.48	4.48	4.48	4.49	4.48	4.48	4.48
Coverage(1)		89.3	89.2	89.4	89.3	89.1	89.1	89.2	88.9	89.3
Coverage(2)		88.9	88.9	88.9	88.7	88.9	88.6	89.0	88.6	88.9
Coverage(5)		87.6	87.8	87.7	87.5	87.7	87.6	87.6	87.5	87.8
Coverage(10)		87.6	87.5	87.5	87.3	87.5	87.4	87.4	87.3	87.7

Note: *The data generating process has* $\phi = 0.9$, $\psi = 0.6$, $\sigma = 1$, $\lambda = 7$ *and* $T = 200$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. Local maxima with* $\hat{\phi} > \hat{\psi}$ *were chosen in 10.75 percent of the cases, while 6.65 percent still has* $\hat{\phi} < \hat{\psi}$.

3. The differences between MAD based on different importance densities are generally small for the present case. Overall, the preferred is case (5) based on a unit root in the candidate model given in Equation (8): this is the one retained for the subsequent simulation study.

4. The last column of Table 1 reports the forecast performance based on LLS approach. It is worth noticing that the differences between the MAD's and 90%-coverages obtained from LLS and GJ methods are generally very small. Since the latter *i)* gives the full distribution of the forecasts conditional on the past and *ii)* does not require to forecast the ϵ's far in the future as the persistence gets strong, it is retained in the subsequent work.

6 Main Simulation Study

Tables 2 to 3 below, and A1 to A2 in the appendix, compare the forecast accuracy reached for $T \in \{100, 200, 400\}$ by five different strategies for the choice of the forecasting model: the one called L refers to the model in levels; the next three ones correspond to the pretest P strategy at the 10%-, 5%- and 1%-levels respectively; the last one, D, corresponds to the model in first differences. In this baseline MAR(1,1) DGP, the forward root parameter is fixed to $\psi = 0.6$ while the degrees of freedom of the Student distribution is set to $\lambda = 7$. Finally, the backward root takes value in $\{0.9, 0.95, 0.975, 0.99, 1\}$, reported in columns (1) to (5).

Table 2 gives the levels of the MADs measures for $T = 100$. Unsurprisingly, the forecast accuracy decreases as the forecast horizon h increases[5]: the MAD values obtained for $h = 1$ range from 0.936 to 0.953 whereas the ones found for $h = 10$ lie between 4.50 and 6.96.

Table 3 (respectively A1 and A2 in the appendix) reports the percent deviation of the MADs from the one of the best model, for $T = 100$ (respectively 200 and 400). It can be seen from Tables 2 and 3 that the best strategy to choose the forecasting model depends heavily on the size of the backward root ϕ. Indeed, for values up to 0.95, the L strategy outperforms all other ones, all the more so as h increases. For this range of ϕ values, the second best strategy is P at the 10%-level: the percent deviation from the L strategy in terms of MAD ranges from 0.6% for $h = 1$ to 6.9% for $h = 10$ when $T = 100$. It ranges from 0% ($h = 1$) to 2.4% ($h = 10$) when

[5]This finding is confirmed for $T = 200$ and 400, as can be seen from Tables A1 and A2 in the appendix. The tables giving the levels of the MAD for these sample sizes are not reported to save space, but are available upon request.

TABLE 2: FORECASTING RESULTS FOR T=100.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.941	0.942	0.943	0.943	0.945
MAD(2)	L	1.70	1.75	1.77	1.79	1.80
MAD(5)	L	3.32	3.66	3.85	3.97	4.07
MAD(10)	L	4.50	5.48	6.11	6.58	6.96
MAD(1)	$P_{10\%}$	0.947	0.947	0.945	0.942	0.938
MAD(2)	$P_{10\%}$	1.72	1.77	1.78	1.78	1.78
MAD(5)	$P_{10\%}$	3.44	3.77	3.89	3.93	3.94
MAD(10)	$P_{10\%}$	4.81	5.80	6.24	6.48	6.57
MAD(1)	$P_{5\%}$	0.948	0.947	0.945	0.942	0.938
MAD(2)	$P_{5\%}$	1.73	1.77	1.78	1.78	1.78
MAD(5)	$P_{5\%}$	3.49	3.79	3.89	3.92	3.94
MAD(10)	$P_{5\%}$	4.98	5.86	6.25	6.46	6.56
MAD(1)	$P_{1\%}$	0.952	0.948	0.943	0.940	0.937
MAD(2)	$P_{1\%}$	1.75	1.77	1.77	1.77	1.77
MAD(5)	$P_{1\%}$	3.56	3.80	3.88	3.92	3.94
MAD(10)	$P_{1\%}$	5.14	5.91	6.25	6.45	6.54
MAD(1)	D	0.953	0.945	0.941	0.939	0.936
MAD(2)	D	1.76	1.77	1.77	1.77	1.77
MAD(5)	D	3.64	3.81	3.88	3.91	3.92
MAD(10)	D	5.39	5.98	6.26	6.43	6.52
Unit root test	ERF 10%	64.6	40.3	24.2	14.9	7.9
Unit root test	ERF 5%	43.7	23.4	13.2	7.7	4.1
Unit root test	ERF 1%	18.7	9.0	4.7	2.8	1.7

NOTE: *The data generating process has* $\psi = 0.6$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

Table 3: Results for T=100. Deviation from Best Forecast in Percent.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.2	0.4	1.0
MAD(2)	L	0.0	0.0	0.0	1.1	1.7
MAD(5)	L	0.0	0.0	0.0	1.5	3.8
MAD(10)	L	0.0	0.0	0.0	2.3	6.7
MAD(1)	$P_{10\%}$	0.6	0.5	0.4	0.3	0.2
MAD(2)	$P_{10\%}$	1.2	1.1	0.6	0.6	0.6
MAD(5)	$P_{10\%}$	3.6	3.0	1.0	0.5	0.5
MAD(10)	$P_{10\%}$	6.9	5.8	2.1	0.8	0.8
MAD(1)	$P_{5\%}$	0.7	0.5	0.4	0.3	0.2
MAD(2)	$P_{5\%}$	1.8	1.1	0.6	0.6	0.6
MAD(5)	$P_{5\%}$	5.1	3.6	1.0	0.3	0.5
MAD(10)	$P_{5\%}$	10.7	6.9	2.3	0.5	0.6
MAD(1)	$P_{1\%}$	1.2	0.6	0.2	0.1	0.1
MAD(2)	$P_{1\%}$	2.9	1.1	0.0	0.0	0.0
MAD(5)	$P_{1\%}$	7.2	3.8	0.8	0.3	0.5
MAD(10)	$P_{1\%}$	14.2	7.8	2.3	0.3	0.3
MAD(1)	D	1.3	0.3	0.0	0.0	0.0
MAD(2)	D	3.5	1.1	0.0	0.0	0.0
MAD(5)	D	9.6	4.1	0.8	0.0	0.0
MAD(10)	D	19.8	9.1	2.5	0.0	0.0
Unit root test	ERF 10%	64.6	40.3	24.2	14.9	7.9
Unit root test	ERF 5%	43.7	23.4	13.2	7.7	4.1
Unit root test	ERF 1%	18.8	9.0	4.7	2.8	1.7

Note: *The data generating process has* $\psi = 0.6$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

$T = 200$ and is 0% for all horizons when $T = 400$. For backward root values up to 0.95, the worst strategy is D. For instance, Table A1 in the appendix indicates that this strategy's percent deviation from the L strategy in terms of MAD ranges from 0.9% for $h = 1$ to 20.3% for $h = 10$ when $T = 200$.

By contrast, when ϕ approaches 1 from below, i.e. local alternatives for the unit root pretest, the D strategy is the best one, the strategy P being the second best again at the 10%-level. These conclusions hold for larger sample sizes, see Tables A1 and A2 in the Appendix, and provide support to Box and Jenkins [1976]'s recommendation. Furthermore, a closer look at Tables 3, A1 and A2, also reveals that the forecast performance deterioration of the L strategy when ϕ approaches 1 (first four lines of columns (3) to (5)) is much weaker than the one of the D strategy when ϕ is well below unity (last four lines of columns (1) to (2)), all the more so as the sample size increases[6]. This finding would support the L strategy. Nevertheless, from a practitioner's point of view, as values of $\phi > 0.95$ are typically found for quarterly and monthly macroeconomic data, see Diebold and Kilian [2000], the P strategy at the 10%-level seems to be a good compromise. In fact, as stressed earlier, its performance is similar to the one of the L strategy for backward root values far from unity — especially for $T > 100$ — while it deteriorates in general less than the one of the L strategy for backward root values as ϕ approaches 1 from below. Indeed, looking at columns (3) to (5) of e.g. Table 3, it can be seen that the percent deviations of the L strategy from the best model lie between 0% ($h = 2$) and 6.7% ($h = 10$) whereas the ones of the $P_{10\%}$ range from 0.2% ($h = 1$) to 2.1% ($h = 10$).

The same conclusions hold for a smaller value of the forward root, *i.e.* $\psi = 0.3$ as reported in Table A3 in the appendix, or for fatter tails in the errors Student's t distribution, namely $\lambda = 3$ as reported in Table A3. As a matter of fact, the relative performance of the $P_{10\%}$ is even improved.

Results obtained from demeaned and detrended data — reported in Tables A5 and A6 in the appendix — are more mitigated. They also support the choice of the P strategy at the 10%-level, for the same reasons as above, but for horizons $h = 1, 2$ only. Indeed, for longer forecast horizons, the deterioration of its performance is more sizeable than the one of the L strategy. This stems from the fact that in presence of a deterministic component, the P strategy suffers from higher critical values as well as from one or two more parameters to estimate for demeaning or detrending. As a consequence, the power of the unit root test decreases noticeably: For $\phi = 0.9$, the empirical rejection frequency (ERF) at the 5%-level drops from 87.2% (Table A1) to 34.3% in the demeaned case (Table A5) and 18% in the detrended case (Table

[6]This outcome emphasizes the importance of the power of the unit root test.

A6). It is worth noticing that in these demeaned and detrended cases, the unit root test is undersized, as can be seen from the last two lines, column (5), of the corresponding tables. This suggests that for forecast horizons greater than 2, a modified pretest strategy should be considered such as one relying on a higher nominal level than 10%, e.g. 20%, or one where the deterministic component coefficients would be estimated following the lines of Elliott et al. [1996], namely using local-to-unity GLS instead of OLS. We leave this analysis for further research.

7 The Case of Brent Crude Oil Prices

In this section, we focus on the oil price data for two reasons. The first one is that most empirical work focusing on oil price forecasting compare various approaches — relying on e.g. AR, ARIMA, VAR, Time Varying Parameters models, on oil futures markets or survey forecasts — with conventional no-change forecasts. Indeed, the latter are considered as a benchmark since they are difficult to beat in out-of-sample oil price forecasting exercises[7]. The no-change forecasting approach assumes that the oil price follows a random walk process. Basically, it implies that the forecasts for the next few months or quarters are simply set to the last observed value. Hence, it amounts to always difference the series. The second reason is that oil prices have been found to be well represented by a MAR model in e.g. Hecq and Voisin [2023].

Here, the monthly Brent crude oil price data are studied from January 2004 to April 2023. This series comes from the Federal Reserve Bank of Saint-Louis FRED database and is plotted in Figure 1 (A). All estimations are performed until June 2022 so as to keep ten observations for the out-of-sample forecasting exercise.

It turns out that two lags in a linear AR model are enough to eliminate any serial correlation in the residuals up to order 24. Moreover, the Jarque and Bera test statistics strongly rejects the null of Gaussian-distributed errors with a p-value lower than 0.01%, which is important as non-Gaussianity is required for the identification of MAR models. Using the Bec et al. [2020]'s unit root test against a MAR(1,1) stationary alternative, a value of -2.92 is obtained: as the 5 and 10% critical values are -2.74 and -2.43 respectively, the unit root null is rejected. Following the strategy stemming from the simulation study presented in the previous section, the forecasts should be done from the MAR model in levels. For comparison purpose, a MAR in first

[7]See, amongst others, Alquist et al. [2013], Baumeister and Kilian [2015], Baumeister et al. [2022], Ellwanger and Snudden [2023] and the numerous references therein.

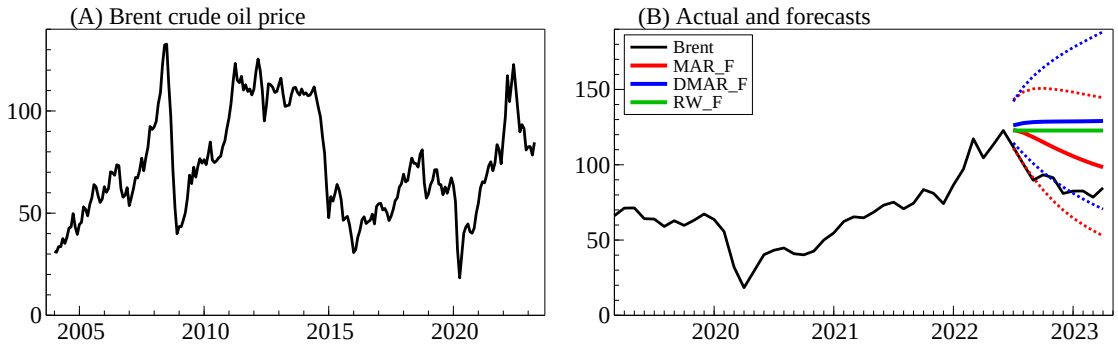

FIGURE 1: *Brent crude oil prices January 2004 to April 2023 (US Dollars per Barrel) in panel (A), and a comparison of forecasts from MAR, DMAR and Random Walk model in panel (B).*

differences — denoted DMAR hereafter — is also considered. Both models' estimates are reported in Table 4.

The last ten observations in our sample (July 2022-April 2023) are then used to compare the forecasting accuracy of the MAR and DMAR models. Panel (B) of Figure 1 reports median forecasts obtained from the MAR (MAR_F, red line) and the DMAR (DMAR_F, blue line) models together with the actual observations (black line). As a benchmark, the forecasts from a simple random walk (RW_F, green line) are also considered.

Obviously, the forecasts computed from the MAR model in levels outperform the ones obtained from the MAR in first differences and from the random walk. It is also worth noticing that the 95% confidence band associated to DMAR forecasts (dotted blue lines) is wider than the one of the MAR forecasts (dotted red lines) and includes the actual value of the oil price only at horizons 8 to 10, namely from February to April 2023. The median forecast of the MAR in levels is closer to the actual value than the median forecast of the DMAR. The Diebold-Mariano test confirms the superiority of the MAR median forecasts compared to the DMAR median forecasts with a t-statistic of -3.94 and a p-value of 0.003, and to the random walk forecasts with a t-statistic of -2.88 and a p-value of 0.016. Hence, in this particular case, the pretesting strategy suggested above would have led to the best forecasts for the period considered.

Finally Figure 2 presents the forecast densities of MAR and DMAR for forecast horizons $h \in \{1, 2, 5, 10\}$. For short forecasting horizons, the difference is minor in this example, but for the larger horizons, in particular for the last observed oil price, April 2023, the differences in density forecasts are

Table 4: Estimation Results, Crude Oil Price.

	MAR			DMAR		
	Estimate	s.e.	t-ratio	Estimate	s.e.	t-ratio
ψ	0.433	0.070	6.102	0.363	0.065	5.623
ϕ	0.915	0.028	31.607	—	—	—
λ	4.423	1.488	2.972	5.051	1.795	2.814
σ^2	21.421	3.879	5.522	23.917	4.089	5.846
Log-likelihood	-701.347			-706.771		

Note: *ψ, ϕ, λ and σ^2 denote the forward root, the backward root, the number of degrees of freedom and the scale of the Student-t distribution, respectively. The estimation sample is January 2004 to June 2022.*

pronounced. Also note that the dispersion of the density forecast is markedly lower for the MAR forecast for larger forecast horizons, by construction.

8 Concluding Remarks

Our paper's goal was to investigate which strategy to adopt in order to choose the best forecasting model — in terms of accuracy — for MAR(1,1) DGPs: always differencing (D), never differencing (L) or unit root pretesting (P).

As a by-product of our analysis, a preliminary simulations study has revealed that for the MAR(1,1) models considered here, *(i)* the importance density of the SIR algorithm should retain a unit root in the candidate model, *(ii)* the median should be the location measure for the simulated predictive distribution and *(iii)* the Lanne et al. [2012] and Gouriéroux and Jasiak [2016] forecasting approaches produce very similar forecasts accuracy.

The main results obtained here from a large simulation study support Box and Jenkins [1976]'s recommendation to use the model in levels unless the root of the process to forecast is close to unity. Moreover, they confirm the ones obtained in favour of the pretesting strategy by e.g. Campbell and Perron [1991], Stock [1996], Stock and Watson [1999] or Diebold and Kilian [2000] for the linear autoregressive class of models. Extending these works by considering a mixed causal-noncausal autoregressive DGP as well as various levels of the unit root pretest strategy, it turns out that from a practitioner's point of view, the P strategy at the 10%-level seems to be a good compromise for MAR(1,1) models. Indeed, it never departs too much from the best model, unlike the L (respectively D) strategy when the DGP becomes very persistent (respectively with little persistence).

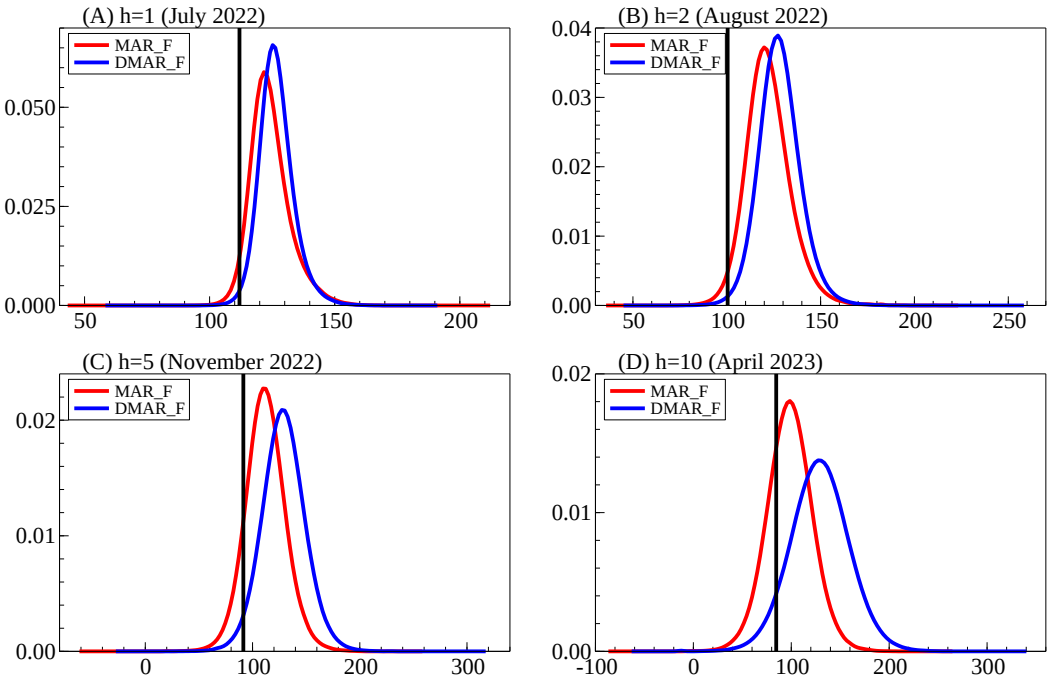

FIGURE 2: *Density forecasts for crude oil prices by MAR and DMAR models for forecasting horizons* $h \in \{1, 2, 5, 10\}$. *The vertical lines indicate the observed oil price.*

For the Brent crude oil price series, an out-of-sample forecasting exercise conducted from July 2022 until April 2023 illustrates that the P strategy suggested in this paper would have led to the best forecasts.

References

Alquist, R., L. Kilian, and R. Vigfusson (2013). Forecasting the price of oil. In *Handbook of Economic Forecasting*, Volume 2, Chapter Chapter 8, pp. 427–507. Elsevier.

Baumeister, C. and L. Kilian (2015). Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business and Economic Statistics 33*, 338–351.

Baumeister, C., D. Korobilis, and T. Lee (2022). Energy Markets and Global

Economic Conditions. *The Review of Economics and Statistics 104*, 828–844.

Bec, F., H. B. Nielsen, and S. Saïdi (2020). Mixed Causal–Noncausal Autoregressions: Bimodality Issues in Estimation and Unit Root Testing. *Oxford Bulletin of Economics and Statistics 82*, 1413–1428.

Bickel, D. and R. Frühwirth (2006). On a fast, robust estimator of the mode: Comparisons to other robust estimators with applications. *Computational Statistics and Data Analysis 50*, 3500–3530.

Box, G. and G. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Breidt, F., R. Davis, K.-S. Lii, and M. Rosenblatt (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis 36*, 175–198.

Campbell, J. and P. Perron (1991). Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots. In *NBER Macroeconomics Annual 1991, Volume 6*, pp. 141–220. National Bureau of Economic Research, Inc.

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

Dickey, D. and W. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–431.

Dickey, D. A., W. R. Bell, and R. B. Miller (1986). Unit roots in time series models: Tests and implications. *The American Statistician 40*(1), 12–26.

Diebold, F. and L. Kilian (2000). Unit-root tests are useful for selecting forecasting models. *Journal of Business and Economic Statistics 18*, 265–273.

Elliott, G., T. Rothenberg, and J. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*, 813–836.

Ellwanger, R. and S. Snudden (2023). Forecasts of the real price of oil revisited: Do they beat the random walk? *Journal of Banking and Finance*. Forthcoming.

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

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

Gouriéroux, C. and J.-M. Zakoian (2015). On uniqueness of moving average representation of heavy tailed stationary processes. *Journal of Time Series Analysis 36*, 876–887.

Gouriéroux, C. and J.-M. Zakoian (2017). Local explosion modelling by noncausal process. *J. R. Stat. Soc. B 79*, 737–756.

Hecq, A. and E. Voisin (2023). Predicting crashes in oil prices during the covid-19 pandemic with mixed causal-noncausal models. In *Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications*, Volume 45 of *Advances in Econometrics*, pp. 209–233. Emerald Group Publishing Limited.

Hencic, A. and C. Gouriéroux (2015). Noncausal autoregressive model in application to bitcoin/usd exchange rates. In *Studies in Computational Intelligence: Econometrics of Risk*, pp. 17–40. Heidelberg.

Lanne, M., A. Luoma, and J. Luoto (2012). Bayesian model selection and forecasting in noncausal autoregressive models. *Journal of Applied Econometrics 27*, 812–830.

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

Lanne, M. and P. Saikkonen (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics 3*, 1–32.

Lanne, M. and P. Saikkonen (2013). Noncausal vector autoregression. *Econometric Theory 29*, 447–481.

Nelson, C. and C. Plosser (1982). Trends and random walks in macroeconmic time series : Some evidence and implications. *Journal of Monetary Economics 10*, 139–162.

Rubin, D. (1988). Using the sir algorithm to simulate posterior distributions. In J. Bernardo, M. DeGroot, D. Lindley, and A. Smith (Eds.), *Bayesian Statistics 3*. Oxford University Presss.

Saikkonen, P. and R. Sandberg (2016). Testing for a unit root in noncausal autoregressive models. *Journal of Time Series Analysis 37*, 99–125.

Stock, J. (1996). VAR, Error Correction and Pretest Forecasts at Long Horizons. *Oxford Bulletin of Economics and Statistics 58*, 685–701.

Stock, J. and M. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, pp. 1–44. Oxford: Oxford University Press.

Appendix: Additional Simulation Results

TABLE A1: RESULTS FOR T=200. DEVIATION FROM BEST FORECAST IN PERCENT.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.0	0.3	0.5
MAD(2)	L	0.0	0.0	0.0	0.0	1.1
MAD(5)	L	0.0	0.0	0.0	0.5	2.0
MAD(10)	L	0.0	0.0	0.0	0.0	3.5
MAD(1)	$P_{10\%}$	0.1	0.2	0.3	0.3	0.1
MAD(2)	$P_{10\%}$	0.6	0.6	1.1	0.0	0.6
MAD(5)	$P_{10\%}$	0.6	1.1	1.3	0.8	0.2
MAD(10)	$P_{10\%}$	1.1	2.4	3.0	0.9	0.5
MAD(1)	$P_{5\%}$	0.4	0.5	0.4	0.2	0.1
MAD(2)	$P_{5\%}$	1.2	1.1	1.1	0.0	0.6
MAD(5)	$P_{5\%}$	1.8	2.2	1.6	0.3	0.2
MAD(10)	$P_{5\%}$	3.6	5.3	4.1	0.6	0.2
MAD(1)	$P_{1\%}$	1.2	1.0	0.5	0.1	0.0
MAD(2)	$P_{1\%}$	2.4	1.7	1.1	0.0	0.6
MAD(5)	$P_{1\%}$	5.4	3.8	1.6	0.0	0.0
MAD(10)	$P_{1\%}$	11.4	9.0	4.6	0.3	0.0
MAD(1)	D	1.8	0.9	0.3	0.0	0.0
MAD(2)	D	3.5	2.3	1.1	0.0	0.0
MAD(5)	D	9.2	4.1	1.3	0.0	0.0
MAD(10)	D	20.3	10.3	4.5	0.2	0.0
Unit root test	ERF 10%	96.4	84.4	51.8	23.2	8.0
Unit root test	ERF 5%	87.2	57.3	26.2	10.1	3.4
Unit root test	ERF 1%	42.6	14.0	4.5	1.4	0.5

NOTE: *The data generating process has* $\psi = 0.6$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

Table A2: Results for T=400. Deviation from Best Forecast in Percent.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.0	0.2	0.3
MAD(2)	L	0.0	0.0	0.0	0.0	0.6
MAD(5)	L	0.0	0.0	0.0	0.0	1.0
MAD(10)	L	0.0	0.0	0.0	0.0	2.0
MAD(1)	$P_{10\%}$	0.0	0.0	0.1	0.3	0.1
MAD(2)	$P_{10\%}$	0.0	0.0	0.0	0.6	0.0
MAD(5)	$P_{10\%}$	0.0	0.0	0.3	0.8	0.3
MAD(10)	$P_{10\%}$	0.0	0.0	1.0	1.4	0.3
MAD(1)	$P_{5\%}$	0.0	0.0	0.3	0.2	0.0
MAD(2)	$P_{5\%}$	0.0	0.0	0.0	0.6	0.0
MAD(5)	$P_{5\%}$	0.0	0.3	0.8	0.8	0.0
MAD(10)	$P_{5\%}$	0.2	0.2	2.4	1.6	0.2
MAD(1)	$P_{1\%}$	0.1	0.3	0.3	0.0	0.0
MAD(2)	$P_{1\%}$	0.0	0.6	1.2	0.0	0.0
MAD(5)	$P_{1\%}$	0.6	1.7	2.4	0.8	0.0
MAD(10)	$P_{1\%}$	0.9	3.8	5.4	1.7	0.0
MAD(1)	D	1.7	0.8	0.2	0.0	0.0
MAD(2)	D	4.2	1.8	0.6	0.0	0.0
MAD(5)	D	10.8	5.4	2.4	0.5	0.0
MAD(10)	D	22.1	12.0	5.9	1.6	0.0
Unit root test	ERF 10%	99.8	99.8	91.3	47.7	9.4
Unit root test	ERF 5%	99.3	98.9	75.4	27.2	4.4
Unit root test	ERF 1%	96.9	78.4	27.5	5.0	0.5

Note: *The data generating process has* $\psi = 0.6$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

TABLE A3: RESULTS FOR T=200 AND $\psi = 0.3$.
DEVIATION FROM BEST FORECAST IN PERCENT.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.0	0.1	0.6
MAD(2)	L	0.0	0.0	0.0	0.0	1.3
MAD(5)	L	0.0	0.0	0.0	0.0	2.2
MAD(10)	L	0.0	0.0	0.0	0.0	3.6
MAD(1)	$P_{10\%}$	0.0	0.2	0.2	0.1	0.1
MAD(2)	$P_{10\%}$	0.7	0.7	0.7	0.0	0.0
MAD(5)	$P_{10\%}$	0.4	0.8	1.5	0.7	0.7
MAD(10)	$P_{10\%}$	0.4	1.8	2.6	1.0	0.5
MAD(1)	$P_{5\%}$	0.1	0.5	0.4	0.1	0.1
MAD(2)	$P_{5\%}$	0.7	1.4	0.7	0.0	0.0
MAD(5)	$P_{5\%}$	0.4	2.0	1.9	0.4	0.4
MAD(10)	$P_{5\%}$	0.7	4.1	3.9	1.0	0.2
MAD(1)	$P_{1\%}$	0.9	1.0	0.4	0.0	0.1
MAD(2)	$P_{1\%}$	2.1	2.0	1.3	0.0	0.0
MAD(5)	$P_{1\%}$	3.9	4.3	2.2	0.0	0.4
MAD(10)	$P_{1\%}$	6.7	8.8	4.5	0.5	0.0
MAD(1)	D	2.6	1.0	0.2	0.0	0.0
MAD(2)	D	4.9	2.7	1.3	0.0	0.0
MAD(5)	D	10.8	5.1	1.9	0.0	0.0
MAD(10)	D	21.6	10.8	4.5	0.5	0.0
Unit root test	ERF 10%	99.9	91.5	59.6	27.0	9.3
Unit root test	ERF 5%	98.8	75.5	36.7	13.7	4.6
Unit root test	ERF 1%	78.3	29.0	8.2	2.4	0.8

NOTE: *The data generating process has* $\psi = 0.3$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

TABLE A4: RESULTS FOR T=200 AND $\lambda = 3$.
DEVIATION FROM BEST FORECAST IN PERCENT.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.0	0.8	0.0
MAD(2)	L	0.0	0.0	0.0	0.4	0.4
MAD(5)	L	0.0	0.0	0.0	0.0	1.9
MAD(10)	L	0.0	0.0	0.0	0.0	2.9
MAD(1)	$P_{10\%}$	0.0	0.0	0.8	0.8	0.0
MAD(2)	$P_{10\%}$	0.0	0.4	0.4	0.4	0.0
MAD(5)	$P_{10\%}$	0.5	0.4	0.8	0.2	0.4
MAD(10)	$P_{10\%}$	1.0	0.5	1.2	0.5	0.6
MAD(1)	$P_{5\%}$	0.0	0.0	0.0	0.8	0.0
MAD(2)	$P_{5\%}$	0.0	0.4	0.4	0.0	0.0
MAD(5)	$P_{5\%}$	0.5	0.8	1.0	0.2	0.2
MAD(10)	$P_{5\%}$	1.0	1.2	1.9	0.7	0.2
MAD(1)	$P_{1\%}$	0.0	0.0	0.8	0.0	0.0
MAD(2)	$P_{1\%}$	0.5	0.9	0.4	0.0	0.0
MAD(5)	$P_{1\%}$	1.1	2.1	2.0	0.2	0.0
MAD(10)	$P_{1\%}$	2.1	3.5	3.4	0.8	0.0
MAD(1)	D	1.7	0.8	0.8	0.0	0.0
MAD(2)	D	4.6	2.2	0.9	0.0	0.0
MAD(5)	D	12.2	5.6	2.8	0.2	0.0
MAD(10)	D	21.9	10.4	4.9	0.7	0.0
Unit root test	ERF 10%	98.0	95.2	77.4	42.3	9.9
Unit root test	ERF 5%	97.5	89.0	61.3	25.9	4.8
Unit root test	ERF 1%	92.0	63.7	28.8	7.9	0.8

NOTE: *The data generating process has* $\psi = 0.6$, $\lambda = 3$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

Table A5: Results for T=200. Demeaned Data. Deviation from Best Forecast in Percent.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.0	0.5	1.1
MAD(2)	L	0.0	0.0	0.0	0.6	1.7
MAD(5)	L	0.0	0.0	0.0	1.8	4.0
MAD(10)	L	0.0	0.0	0.0	2.3	5.9
MAD(1)	$P_{10\%}$	1.2	0.6	0.3	0.2	0.1
MAD(2)	$P_{10\%}$	1.8	1.1	0.6	0.0	0.6
MAD(5)	$P_{10\%}$	4.8	3.0	0.5	0.5	0.8
MAD(10)	$P_{10\%}$	9.1	6.3	2.3	0.6	0.8
MAD(1)	$P_{5\%}$	1.5	0.9	0.2	0.1	0.0
MAD(2)	$P_{5\%}$	2.4	1.1	0.6	0.0	0.6
MAD(5)	$P_{5\%}$	5.7	3.2	0.3	0.3	0.5
MAD(10)	$P_{5\%}$	11.7	7.4	2.3	0.3	0.5
MAD(1)	$P_{1\%}$	1.7	0.7	0.1	0.1	0.0
MAD(2)	$P_{1\%}$	2.9	1.1	0.6	0.0	0.0
MAD(5)	$P_{1\%}$	7.4	3.5	0.3	0.0	0.3
MAD(10)	$P_{1\%}$	15.2	7.8	2.1	0.2	0.2
MAD(1)	D	2.0	0.7	0.1	0.0	0.0
MAD(2)	D	3.5	1.7	0.6	0.0	0.0
MAD(5)	D	8.9	3.5	0.0	0.0	0.0
MAD(10)	D	18.8	8.3	1.9	0.0	0.0
Unit root test	ERF 10%	51.6	23.9	10.9	6.5	5.7
Unit root test	ERF 5%	34.3	12.1	5.1	3.0	2.7
Unit root test	ERF 1%	12.9	3.5	1.5	1.0	0.9

Note: *The data generating process has* $\psi = 0.6$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*

Table A6: Results for T=200. Detrended data. Deviation from Best Forecast in Percent.

		(1)	(2)	(3)	(4)	(5)
ϕ		0.9	0.95	0.975	0.99	1
MAD(1)	L	0.0	0.0	0.0	0.6	0.9
MAD(2)	L	0.0	0.0	0.0	1.1	1.7
MAD(5)	L	0.0	0.0	0.5	3.0	3.7
MAD(10)	L	0.0	0.0	1.4	5.4	7.1
MAD(1)	$P_{10\%}$	1.8	1.0	0.2	0.1	0.1
MAD(2)	$P_{10\%}$	3.0	1.7	0.0	0.6	0.0
MAD(5)	$P_{10\%}$	6.6	3.2	0.3	0.2	0.2
MAD(10)	$P_{10\%}$	11.4	4.7	0.5	0.5	0.3
MAD(1)	$P_{5\%}$	2.0	1.0	0.2	0.1	0.1
MAD(2)	$P_{5\%}$	3.6	2.3	0.0	0.6	0.0
MAD(5)	$P_{5\%}$	7.5	3.5	0.3	0.0	0.0
MAD(10)	$P_{5\%}$	13.0	5.0	0.2	0.2	0.1
MAD(1)	$P_{1\%}$	2.3	1.1	0.1	0.0	0.0
MAD(2)	$P_{1\%}$	4.1	2.3	0.0	0.0	0.0
MAD(5)	$P_{1\%}$	9.0	3.5	0.0	0.0	0.0
MAD(10)	$P_{1\%}$	15.3	5.2	0.0	0.0	0.1
MAD(1)	D	2.6	1.1	0.1	0.0	0.1
MAD(2)	D	4.7	2.3	0.0	0.6	0.0
MAD(5)	D	10.2	3.8	0.0	0.0	0.0
MAD(10)	D	17.1	5.4	0.0	0.0	0.0
Unit root test	ERF 10%	28.5	11.5	6.2	4.4	4.0
Unit root test	ERF 5%	18.0	6.4	3.3	2.4	2.3
Unit root test	ERF 1%	7.1	2.3	1.4	1.1	1.1

Note: *The data generating process has* $\psi = 0.6$, $\lambda = 7$, *and* $\sigma = 1$. *The forecasts are based on* $M = N = 10000$ *draws and the results are based on* $B = 10000$ *Monte Carlo replications. ERF* $x\%$ *denotes the empirical rejection frequency of the null at the* $x\%$ *level.*