Série des Documents de Travail

n° 2019-09

Mixed Causal-Noncausal Autoregressions: Bimodality Issues in Estimation and Unit Root Testing

Frédérique BEC[1]
Heino BOHN NIELSEN[2]
Sarra SAÏDI[3]

[1] CREST; Thema, University of Cergy-Pontoise. E-mail: bec@ensae.fr
[2] Department of Economics, University of Copenhagen. E-mail : heino.bohn.nielsen@econ.ku.dk
[3] Thema, University of Cergy-Pontoise. E-mail : sarra.saidi@u-cergy.fr

Mixed Causal-Noncausal Autoregressions: Bimodality Issues in Estimation and Unit Root Testing *

Frédérique Bec† Heino Bohn Nielsen‡ Sarra Saïdi§

June 16, 2019

Abstract

This paper stresses the bimodality of the widely used Student's t likelihood function applied in modelling Mixed causal-noncausal AutoRegressions (MAR). It first shows that a local maximum is very often to be found in addition to the global Maximum Likelihood Estimator (MLE), and that standard estimation algorithms could end up in this local maximum. It then shows that the issue becomes more salient as the causal root of the process approaches unity from below. The consequences are important as the local maximum estimated roots are typically interchanged, attributing the noncausal one to the causal component and *vice-versa*, which severely changes the interpretation of the results. The properties of unit root tests based on this Student's t MLE of the backward root are obviously affected as well. To circumvent this issues, this paper proposes an estimation strategy which *i)* increases noticeably the probability to end up in the global MLE and *ii)* retains the maximum relevant for the unit root test against a MAR stationary alternative. An application to Brent crude oil price illustrates the relevance of the proposed approach.

JEL Classification: C12, C22, Q41

Keywords: Mixed autoregression, non-causal autoregression, maximum likelihood estimation, unit root test, Brent crude oil price.

*This work has benefited from discussions with Christian Francq, Alain Guay, Alain Hecq, Rasmus Søndergaard Pedersen, Anders Rahbek, and Jean-Michel Zakoian and from remarks by participants at the annual congress of the Canadian Society for Economic Science, Quebec 2019, and Quantitative Finance and Financial Econometrics conference, Marseille 2019. F. Bec and S. Saïdi acknowledge financial support from the Labex MME-DII. H. B. Nielsen acknowledges financial support from the Institute for Advanced Studies at the University of Cergy-Pontoise and the Danish Council for Independent Research, DSF Grant 015-00028B.

†Thema, University of Cergy-Pontoise, and CREST, France, bec@ensae.fr.

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

§Thema, University of Cergy-Pontoise, France, sarra.saidi@u-cergy.fr.

1 Introduction

As emphasized by Hanfelt [2000] in his comment on Small, Wang and Yang [2000]'s paper, most statisticians doubt the proposition that multiple roots pose a serious problem in data analyses. Also, very few published data analyses in scientific journals mention the existence of multiple roots or describe what methods are used to select among them. Our paper is devoted to fill this gap for the case of mixed causal-noncausal autoregressive (MAR) models, and as a by-product we propose an improvement to the test for a unit root against MAR stationary alternatives. Basically, a maximum likelihood estimation strategy is proposed, which has the desirable feature that it selects the relevant maximum for the topic at hand.

Introduced decades ago in the literature of statistics (see for instance Weiss [1975], Findley [1986], Lawrance [1991], Breidt and Davis [1991], Breidt, Davis, Lh and Rosenblatt [1991], Breidt, Davis and Dunsmuir [1992], Rosenblatt [1993], Cambanis and Fakhre-Zakeri [1994], Cambanis and Fakhre-Zakeri [1996] or Rosenblatt [2000]), mixed causal-noncausal autoregressions have recently known a revival of interest amongst researchers in economics and econometrics (see e.g. 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] or Fries and Zakoian [2019]).

To fix ideas, let us introduce the MAR(r,s) model as formulated by Lanne and Saikkonen [2011],

$$\phi(B)\varphi(B^{-1})y_t = \epsilon_t, \tag{1}$$

where B is the backward shift operator ($B^k y_t = y_{t-k}$ for $k = 0, \pm1, ...$) and $\phi(B) = 1 - \phi_1 B - ... - \phi_r B^r$, $\varphi(B^{-1}) = 1 - \varphi_1 B^{-1} - ... - \varphi_s B^{-s}$. Finally, ϵ_t is a sequence of non-Gaussian independent, identically distributed random variables with mean zero, density function $f(\epsilon_t \mid \lambda)$ where λ is a set of parameters to be specialized later, and $E(\epsilon_t^2) < \infty$ unless otherwise mentioned. Indeed, if the error terms were Gaussian distributed, the model could be written indifferently as a backward or a forward autoregression, as these two representations are observationally equivalent asymptotically in this case.[1]

Under the assumption that the polynomials $\phi(z)$ and $\varphi(z)$ ($z \in \mathbb{C}$) have roots outside the unit circle, it is well-known that y_t has a stationary solution in terms of the two-sided moving average representation:

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}. \tag{2}$$

As can be seen from equation (2), MAR models allow for dependence on both the past and the future, by contrast with the well-known backward-looking autoregression which rules out dependence on future observations. If $r > 0$ and $s = 0$, the process defined in equation (1) becomes a *purely causal* process, namely the MAR(r,0) or AR(r):

$$\phi(B)y_t = \epsilon_t. \tag{3}$$

If $s > 0$ and $r = 0$ in equation (1), one obtains the following *purely noncausal* process MAR(0,s):

$$\varphi(B^{-1})y_t = \epsilon_t. \tag{4}$$

[1] See e.g. the discussion in Cambanis and Fakhre-Zakeri [1996].

Papers by Hencic and Gouriéroux [2015], Gouriéroux and Zakoian [2015], Gouriéroux and Jasiak [2016], Gouriéroux and Zakoian [2017] or Fries and Zakoian [2019], assume Cauchy distributed disturbances in (1), *i.e.* very fat tailed distributions needed to capture bubble-like dynamics. For other macroeconomic variables such as the inflation or interest rates, a popular choice is the Student's t-distribution with scale parameter $\sigma > 0$ and $\nu > 2$ degrees of freedom, see *inter alia* Lanne and Saikkonen [2011], Nyberg, Lanne and Saarinen [2012], Lanne and Saikkonen [2013], Lof [2013] or Lof and Nyberg [2017]. With these Student's t-disturbances, the approximate maximum likelihood estimation (MLE) approach has been advocated by Breidt et al. [1991], Andrews, Davis and Breidt [2006] and further promoted by Lanne and Saikkonen [2011]. The last two papers have derived conditions under which the MLE is consistent and asymptotically normal. The least absolute deviation (LAD) estimator has been proposed and its properties studied by Huang and Pawitan [2000] and Wu and Davis [2010] for the cases of unknown density function of the ϵ_t's in (1): they have characterized the properties of the Quasi-MLE when the objective function is the Laplace likelihood, which gives the LAD estimator. Alternative possible distributions include the Gaussian mixture or the generalized error distribution.

More recently, Hecq, Lieb and Telg [2016] have compared the MLE and LAD estimators in finite sample. There are signs of bimodality in their simulated illustration of finite sample likelihood estimation, see Figures 2 and 3[2]. For sample sizes of 200 and 800, two peaks can be seen in the distribution of both LAD and ML estimates when the underlying disturbances are Student's t-distributed with 10 degrees of freedom. To our knowledge, this kind of issue has hardly been discussed so far in the literature. Yet, as noticed by Small et al. [2000], estimating functions such as the first order conditions of the likelihood function or likelihood equations can have more than one root. Unfortunately, in practice it is often not obvious to determine which root is appropriate as a parameter estimate.

In what follows, we will specialize to the MAR(1,1) model to unveil the bimodality issue for ML estimators. Section 2 presents results to evaluate the magnitude of the bimodality phenomenon. Section 3 will evaluate two consequences of the issue, both stemming from the fact that the global and local maxima have approximately the same roots, but their location is interchanged between the causal and noncausal components: First, the interpretation of the processes generated from the global and local maxima are quite different. Second, the size and power of Saikkonen and Sandberg [2016]'s unit root test against a stationary MAR alternative can be strongly impacted by the confusion in the location of the estimated roots. Section 4 proposes an estimation strategy which circumvents this bimodality issue and a modification of Saikkonen and Sandberg [2016]'s unit root test so that its size and power are restored. Section 5 illustrates the benefits of the proposed approach using the Brent crude oil price while Section 6 concludes.

2 The bimodality issue

Let us first present the likelihood function to maximize in the remainder of the paper. We assume that ϵ_t, $t = 1, 2, ..., T$, is a sequence of non Gaussian i.i.d. random variables

[2]See pages 317-318 therein.

with density function $f(\epsilon_t \mid \lambda)$. Lanne and Saikkonen [2011] propose to estimate the MAR model in (1) from the approximate log-likelihood function given by:

$$\log L_T(\theta) = \sum_{t=r+1}^{T-s} \log f(\epsilon_t \mid \lambda), \tag{5}$$

where $\epsilon_t = \phi(B)\varphi(B^{-1})y_t$ and $\theta = (\phi, \varphi, \lambda)$. For the Student's t-case, considered in most of the paper, λ consists of the scale and the degree of freedom parameter, and the density is given by

$$f(\epsilon_t \mid \lambda) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(\pi\nu\sigma^2\right)^{-\frac{1}{2}} \left(1 + \frac{\epsilon_t^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}}, \tag{6}$$

with $\theta = (\phi, \varphi, \sigma^2, \nu)$. The maximisation of $\log L_T(\theta)$ over permissible values of θ, $\theta \in \Theta$, gives the approximate MLE of θ.

Throughout this paper, we focus on a characterization of the shape of the likelihood function and let the parameter space for estimation, Θ, be unrestricted apart from $\sigma^2 > 0$ and $\nu > 0$. An alternative would be to impose the assumptions for the analysis of the stationary case as in Lanne and Saikkonen [2011], $|\phi| < 1$, $|\varphi| < 1$ and $\nu > 2$, or the assumptions for the unit-root case as in Saikkonen and Sandberg [2016], allowing $\phi = 1$. This may be preferable for some empirical analyses, but to illustrate that bimodality often occurs with interchanged roots, in particular when the true backward root is unity, $\phi_0 = 1$, we prefer to maximize the likelihood function for (ϕ, φ) unrestricted.

2.1 Limiting behaviour

First, Figure 1 reports the concentrated log-likelihood function as a function of φ, the noncausal parameter of the MAR(1,1) model, for different cases[3]

The simulated MAR processes are of length $T = 50,000$ or larger, so as to consider the limiting behaviour of the likelihood.

Figure (A)-(D) report results for the Student's t-likelihood for a set of true values $(\varphi_0, \phi_0, \nu_0)$. Figure (A) is characterized by a unit root in the backward lag polynomial, $\phi_0 = 1$, and it shows very clearly that two maxima can be reached by maximizing the likelihood function—no matter the degrees of freedom of the Student's t-distribution. Moreover, the global maximum is reached for the true value of φ, namely $\varphi_0 = 0.6$, while the second local maximum is reached for $\varphi \approx 1$, so that the unit root is wrongly located as the forward component of the MAR model. Finally, it is worth noticing that when ν_0 goes to infinity, corresponding to the Gaussian case, the two maxima reach exactly the same likelihood value. This is the well-known non-identification result, see Cambanis and Fakhre-Zakeri [1994] and Remark 1 below. By contrast, for smaller ν_0 values, the second maximum, interchanging the roots, is a local one only.

Figure (B) repeats the experiments in a stationary case which can still be considered as close to the unit root case: $\phi_0 = 0.9$. The same conclusions hold, but it can be seen

[3]Throughout this paper we consider non-zero mean processes and follow Saikkonen and Sandberg [2016] by demeaning the series before estimation. All results below do not depend on the demeaning or detrending of the data.

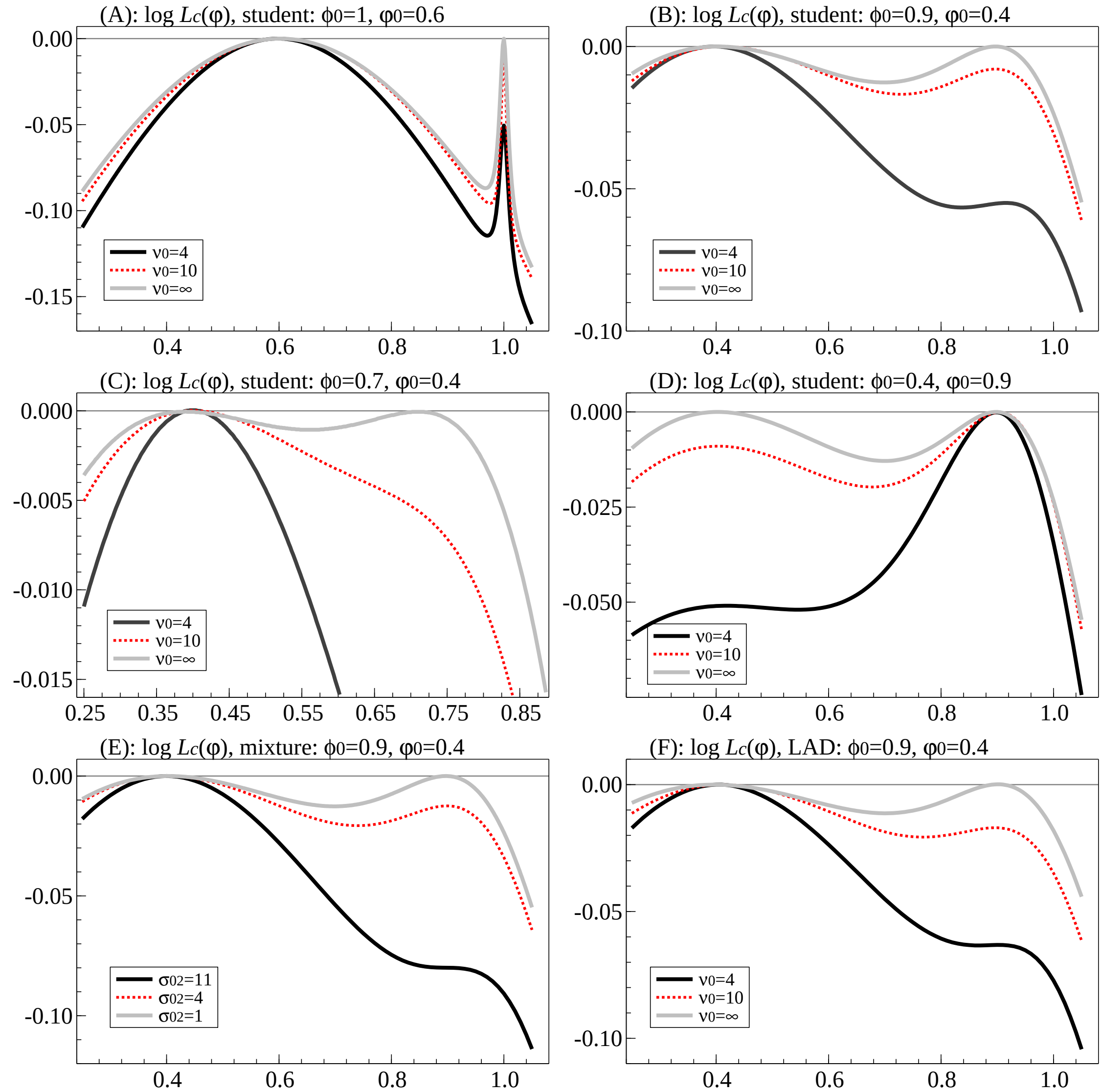

Figure 1: *Profile likelihood function,* $\log L^c(\varphi)$*, for the Student's t, Gaussian mixture, and Laplace likelihood for different values of* ϕ_0 *and* φ_0*. Calculated for T=50,000 or larger. The Gaussian mixture in (E) is given by an N(0,1) with provability 0.9 and* $N(0,\sigma_0^2)$ *with probability 0.1, for different values of* σ_0^2*. To make comparison easy, the plotted likelihood values are the deviations from the maximum close to the true values of the parameters.*

that the width of the set of starting values which will lead to the wrong optimum is much larger than in the unit root case[4].

The simulation experiments reported in Figure (C) reveal that the bimodality issue may disappear for particular parameter combinations. Actually, with an even lower backward root, here set to $\phi_0 = 0.7$, fat-tailed Student's t-distributions make the profile likelihood unimodal. Nevertheless, the bimodality reappears as ν_0 increases.

Figure (D) illustrate the likelihood surface for the opposite case, $\varphi > \phi$. The profile

[4]This stems from the convergence of the estimator at rate T if $\phi_0 = 1$ and at rate $\sqrt{T}$ if $|\phi_0| < 1$.

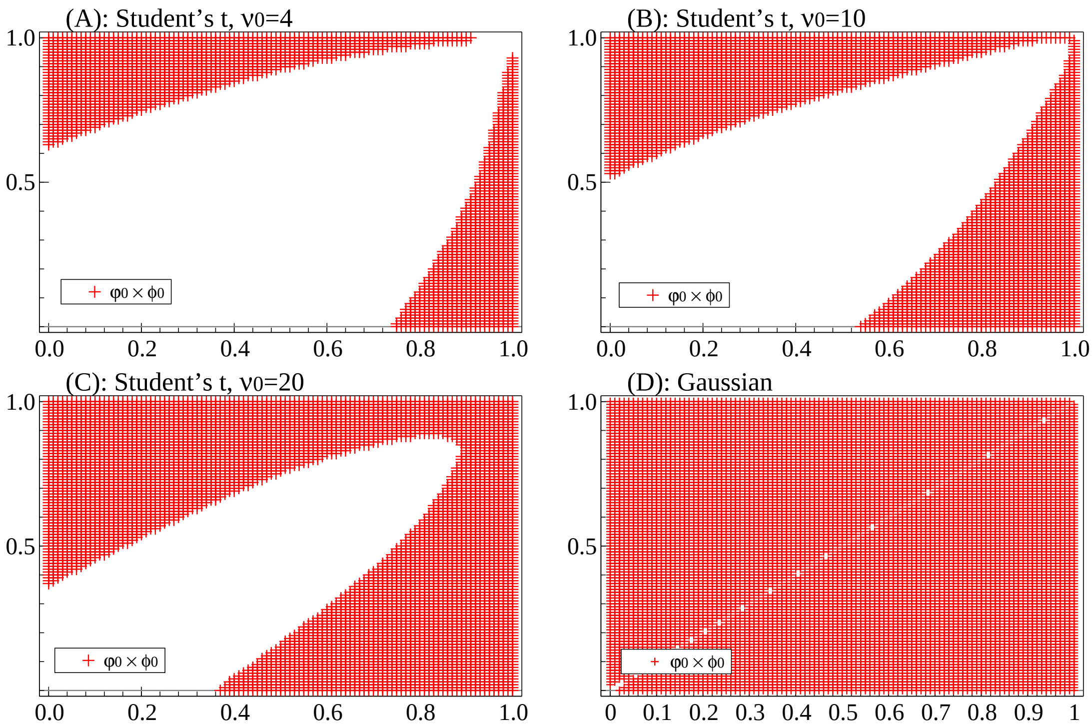

Figure 2: *Bimodality occurence as a function of the parameters values, T=5,000.*

is close to being a mirror image of Figure (B).

To emphasize that the bimodality issue is related to the MAR model *per se*, and not to the particular choice of a Student's t-distribution, Figure (E) shows the profile likelihood for the case where $f(\epsilon_t \mid \lambda)$ is chosen as a mixture of Gaussian distributions (which is also guaranteed to be leptokurtic). The three cases have been chosen to mimic the shape of the tails of the Student's t-distribution with $\nu_0 \in \{4, 10, \infty\}$, and the results are very similar to the results for the Student's t-likelihood.

Finally, Figure (F) shows the profile likelihood when $f(\epsilon_t \mid \lambda)$ is chosen as a Laplace distribution, leading to the LAD estimator. We generate the data using the Student's t-distribution to emphasize the role of LAD estimation as a robust QML estimator when the true distribution of the errors is unknown. Again we observe a pronounced bimodality.

The two following remarks sum up these simulation experiments and give conditions on the parameters for bimodality to occur. We first establish analytically the bimodality of the likelihood function in the Gaussian case. Here the bimodality is a reflection of ϕ and φ being unidentified as $T \to \infty$, which has already been stressed from the simulations experiments reported in Figure 1 above. We next emphasize some conditions on the parameters of the MAR(1,1) DGP which increase the probability of bimodality occurence.

Let us begin with the following remark, whose derivations are given in the Appendix:

Remark 1 *Consider the MAR(1,1) model in (1) with Gaussian errors and* $|\varphi_0| < 1$, $|\phi_0| < 1$. *Let* $\log L_\infty(\phi, \varphi)$ *denote the limit of the Gaussian likelihood function as* $T \to \infty$. *If* $(\varphi, \phi) = (\varphi_0, \phi_0)$ *is a local maximum of* $\log L_\infty(\phi, \varphi)$*, then the interchanged param-*

eter point, $(\varphi, \phi) = (\phi_0, \varphi_0)$*, is also a maximum, and* $\log L_\infty(\varphi_0, \phi_0) = \log L_\infty(\phi_0, \varphi_0)$.

Let us now figure out what are the conditions of the MAR(1,1) model parameters which increase the bimodality case probability. Figure 2 plots in red, for $\nu_0 \in \{4, 10, 50, \infty\}$, all the couples ($\phi_0$,$\varphi_0$) $\in \{0.0, 0.01, \cdots, 0.99, 1.00\}^2$ for which bimodality with interchanged roots occurs. It amounts to 10,000 gridpoints (ϕ_0,φ_0) for each ν_0. This is done with T=5,000 in order to approximate the limiting likelihood function.

Remark 2 *As can be seen from Figure 2, the frequency of bimodality occurence increases if:*

i) the fatness of the tails decreases,

ii) the causal and/or noncausal roots approach one from below,

iii) the difference between the causal and non causal roots increases.

2.2 Finite sample behaviour

Let us now have a look at the finite-sample properties of the likelihood function and the MLE. First, Figure 3 shows profile likelihoods for smaller samples, $T \in \{100, 250, 1000\}$, corresponding to some of the limiting cases in Figure 1. Unlike the limiting case, here approximated by T=50,000, the profile likelihood for finite T depends on the specific realization of the simulated MAR time series. As such these are only illustrative examples.

It is worth noting in Figure 3, that in small samples the global maximum might correspond to the case of interchanged roots. This can be seen from the examples in panels (A), (B), and (C), where the likelihood for $T = 100$ is larger for the interchanged parameter points than for points close to the true values.

Also observe that even if the limiting likelihood function is bimodal, the finite sample likelihood may be unimodal. This happens for one realization for $T = 100$ in panel (B). Panel (C) illustrates the reverse case, where the limiting likelihood is unimodal, while the small sample likelihoods for $T = 100$ and $T = 250$ are bimodal.

To illustrate the occurrence of these issues, Table 1 reports the results of simulation experiments for $\nu_0 = 4$ or $\nu = 10$. The same DGPs as the ones considered in Figure 1 have been simulated 5,000 times for smaller sample sizes: T $\in \{100, 250, 500, 1000, 10000\}$. The global maximum is found using a grid search over initial values, which will be described in Section 4 below. The columns labelled "% Bimodal" report the percentage of the corresponding model estimates leading to two or more maxima. The columns labelled "% Roots interchanged" report the percentage of global maxima which yields interchanged roots, $\hat{\varphi} > \hat{\phi}$.

We observe that the finite sample likelihood function is frequently bimodal, also in cases where the limiting likelihood is unimodel. Even fat tailed disturbances associated with clearly stationary roots ($\phi_0 = 0.7$ and $\varphi_0 = 0.4$, or $\phi_0 = 0.4$ and $\varphi_0 = 0.7$) raise bimodality and interchanged roots issues in small samples, but the issue disappears as $T \to \infty$. For instance, when $T = 100$ and $\nu_0 = 4$, 20.4% of the DGPs display two maxima, and the roots are interchanged in around 15% of the global MLE. With $\nu_0 = 10$, these percentages are much higher, even for $T = 500$ where bimodality is found in about 42% of the cases and still around 15% of the global maxima interchange the roots location. As $\phi_0 \uparrow 1$, the bimodality occurrences strongly increase — it reaches

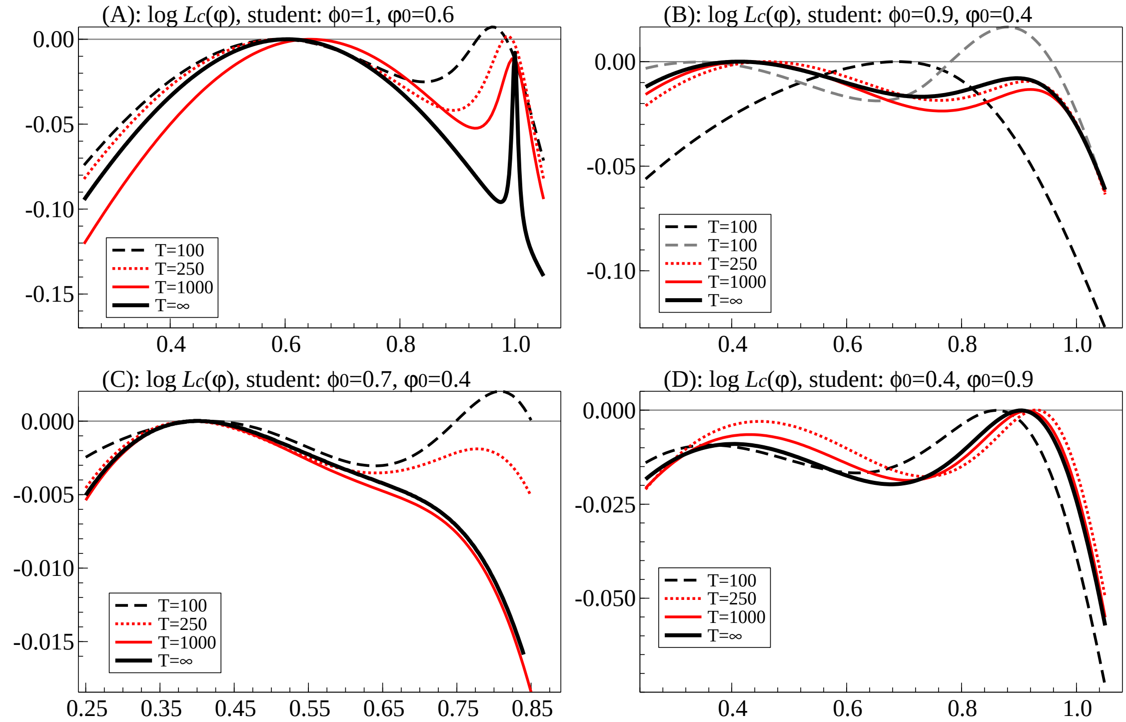

Figure 3: *Profile likelihood function,* $\log L^c(\varphi)$*, for the Student's t likelihood for finite samples, with* $\nu_0 = 10$*. To make comparison easy, the plotted likelihood values are the deviations from the maximum close to the true values of the parameters.*

100% or so as T increases — while the percentage of the interchanged roots is not so much affected. Besides, the latter is a decreasing function of the sample size.

3 Its consequences

As discussed earlier, the roots obtained from global and local maxima are often interchanged between backward and forward components. Interpretation is strongly affected and so is Saikkonen and Sandberg [2016]'s unit root test.

3.1 Impact on interpretation

As already mentioned, an obvious consequence of the roots interchange concerns the interpretation of the series at hand. To illustrate this very simply, let us consider two MAR(1,1) processes with T=200, the scale parameter $\sigma_0^2 = 1$, and $\nu_0 = 4$ degrees of freedom. In Model 1 we set $(\phi_0, \varphi_0) = (0.9, 0.4)$ while for Model 2 we interchange the roots, $(\phi_0, \varphi_0) = (0.4, 0.9)$. These simulated processes, submitted to the exact same sequence of t-distributed shocks, are plotted in Figure 4.

Figure (A) represents the process with a large (respectively small) backward (resp. forward) root and as such, one may recognize the usual behaviour of AR processes. Following the strong negative shock which occurs a little bit before the 70th realization, it plunges deeply from a slightly positive value to around -30, and then progressively

Table 1: The extent of the bimodality issue in finite sample

			$\nu_0 = 4$		$\nu_0 = 10$	
T	ϕ_0	φ_0	% Bimodal	% Roots interchanged	% Bimodal	% Roots interchanged
100	1	0.6	60.6	14.6	74.3	37.7
250	1	0.6	85.1	2.6	97.8	27.5
500	1	0.6	97.4	0.1	99.9	15.8
1000	1	0.6	99.9	0.0	100.0	4.7
10000	1	0.6	100.0	0.0	100.0	0.0
100	0.9	0.4	58.3	12.1	71.8	37.4
250	0.9	0.4	61.2	1.6	90.0	24.4
500	0.9	0.4	63.0	0.1	97.2	12.1
1000	0.9	0.4	65.9	0.0	99.7	3.5
10000	0.9	0.4	82.2	0.0	100.0	0.0
100	0.7	0.4	20.4	15.0	42.9	39.1
250	0.7	0.4	6.4	3.2	45.6	27.7
500	0.7	0.4	1.0	0.3	41.7	15.0
1000	0.7	0.4	0.0	0.0	32.7	6.0
10000	0.7	0.4	0.0	0.0	0.6	0.0
100	0.4	0.7	20.4	14.6	42.6	38.0
250	0.4	0.7	6.7	2.8	45.8	26.9
500	0.4	0.7	0.9	0.4	41.6	15.4
1000	0.4	0.7	0.0	0.0	32.4	6.0
10000	0.4	0.7	0.0	0.0	0.7	0.0

Notes: "% Bimodal" refers to the percentage of the model estimates leading to at least two maxima. "% Roots interchanged" reports the percentage of global maxima which yields interchanged roots, $\hat{\varphi} > \hat{\phi}$. All figures are computed from 5,000 drawings of the Student's t MAR DGPs.

absorbs the shock over the thirty next periods.

For Model 2, in Figure 4 (B), where the roots are interchanged, the process anticipates the shock by decreasing steadily almost from 20 periods before on, and then recovers immediately after the adverse shock occurs. For the positive large shock which occurs around the 130th realization, Model 2 again anticipates it and build up a bubble like dynamics since the 110th realization approximately, and then burst right after the shock happens. Again, a mirroring dynamics is generated by model one which is "surprised" by the occurrence of the favourable shock and increases abruptly instantaneously, before the shock effect slowly vanishes with the time passing by.

This difference has already been widely commented on when comparing pure causal and non causal dynamics. What we want to emphasize here, is that the same can be

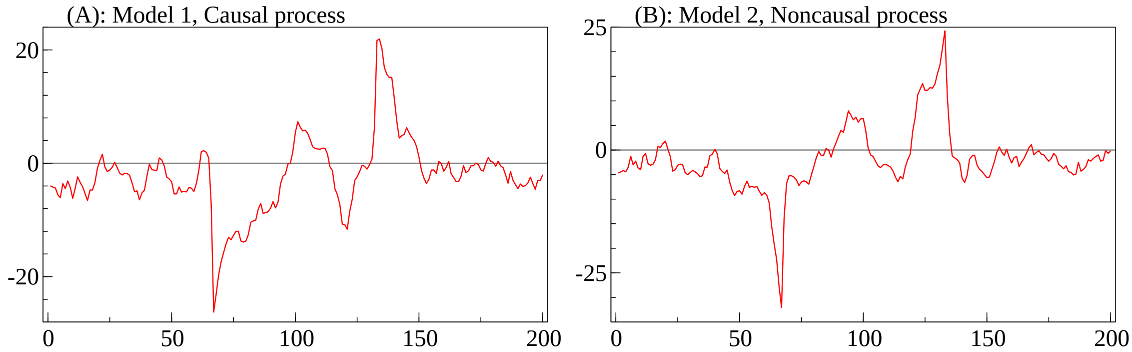

Figure 4: *Simulated MAR(1,1) processes. Model 1 has* $(\phi_0, \varphi_0) = (0.9, 0.4)$ *while model 2 has* $(\phi_0, \varphi_0) = (0.4, 0.9)$.

observed in mixed causal-noncausal processes.

The uncertainty in the location of the forward and backward roots implied by the bimodality issue does not only affect the direct interpretation of the dynamics, e.g. the likelihood of bubble-like dynamics occuring, but may also have severe consequences for forecasts based on the estimated MAR models.

3.2 Impact on unit root testing

Another severe consequence is related to the size and power of Saikkonen and Sandberg [2016]'s unit root test. Their test statistic is given by:

$$\tau_T = \frac{T(\hat{\phi} - 1)}{\sqrt{G_T^{\phi}(\hat{\theta})}},$$

where $G_T^{\phi}(\hat{\theta})$ is the element of the inverse standardized Hessian matrix corresponding to $\hat{\phi}$. Their Proposition 3, p.107, derives the limiting distribution of this statistics under the unit root null $\phi_0 = 1$ and shows that it is not nuisance parameters free. For the Student's t-case, the limiting distribution depends on

$$\mathcal{J} = \frac{\nu_0(\nu_0 + 1)}{(\nu_0 - 2)(\nu_0 + 3)},$$

where $\nu_0 > 2$ is the degrees of freedom parameter.

With this in mind, the issue raised by bimodality and interchanged roots can be easily explained. Intuitively, every time the largest root is wrongly located in the forward component, such that $\hat{\varphi} \approx 1$ and $\hat{\phi} \ll 1$, the unit root test rejects the null $\phi_0 = 1$. Hence, the test's size and power will be spuriously increased.

This is illustrated in Table 2, which reports in the empirical rejection rates of the null by the τ_T statistics for Saikkonen and Sandberg [2016]'s unit root test. The authors consider three cases depending on the deterministic specification, and we focus here on case 2, for which the series is demeaned before estimation and unit root testing is performed. The critical values used in this Table have been computed using Table 1, p.110, in Saikkonen and Sandberg [2016].

Table 2: Impact of bimodality on Saikkonen-Sandberg's unit root test size

T	Global maximum				Preferred maximum			
	% Roots	τ_T ERF in %			% Roots	τ_T ERF in %		
	interchanged	1%	5%	10%	interchanged	1%	5%	10%
				$(\phi_0, \varphi_0, \nu_0)$ = (1,0.6,4)				
100	14.6	11.8	16.3	20.5	3.8	2.5	6.0	9.9
250	2.6	3.5	6.2	11.1	0.0	0.9	3.6	8.5
500	0.1	1.1	4.5	9.0	0.0	1.0	4.4	9.0
1000	0.0	0.8	4.6	9.6	0.0	0.8	4.6	9.6
				$(\phi_0, \varphi_0, \nu_0)$ = (1,0.6,10)				
100	37.7	28.0	33.5	37.2	11.3	4.8	8.5	11.4
250	27.5	27.5	28.8	31.5	0.4	0.5	1.8	4.5
500	15.8	16.0	18.5	22.9	0.1	0.3	2.8	7.3
1000	4.7	5.2	8.2	12.8	0.0	0.5	3.6	8.4

Notes: ERF stands for empirical rejection frequency. All figures are computed from 5,000 simulated MAR models. The critical values for the unit root test have been calculated from Table 1 in Saikkonen and Sandberg [2016] using the estimated degrees of freedom in each replication, $\hat{\nu}$.

Top and bottom panels of Table 2 show the results obtained when $\nu_0 = 4$ and $\nu_0 = 10$, respectively. The left panel reports the empirical rejection frequency (ERF) when the global maximum estimates are used to perform the τ_T unit root test.

The right panel labelled "Preferred maximum" performs the unit root test after the maximum with $\hat{\varphi}$ close to one is eliminated, when there is more than one maximum. Indeed, this case doesn't make sense from an economic point of view, as will be further discussed in Section 4.2. below. Again, the MLE is obtained using a grid search over initial values which will be described in the next section. The impact of interchanged roots on this unit root test size is of course bigger when the percentage of reversed roots is higher, which can be seen from the fourth columns under the "Global maximum" label: In the case where $\nu_0 = 4$, with T=100, 14.6% of the global maximum interchange the roots which yields ERFs well above their nominal levels. For instance, considering the τ_T unit root test at the 5% level gives an ERF of 16.3% instead of 5%. The issue becomes worse in the bottom panel when $\nu_0 = 10$, since at the 5% level and for T=100, one obtains an ERF of 33.5%. With this ν_0, the issue persists for larger samples too. For instance, when T=500, the ERF corresponding to the nominal 5% level of the test is still 18.5%.

Regarding power, Table 3 compares the empirical rejection rates of the null obtained from our estimation method — column labeled "Preferred maximum ERF" — to the ones given in Saikkonen and Sandberg [2016], Fig.2, p.114 therein. As expected, it can be seen from the last two columns that for small sample sizes, the Saikkonen and Sandberg [2016]'s ERFs are larger than the ones of our modified version of the test by 10 to 15 percentage points. Nevertheless, this power issue becomes smaller and vanishes

Table 3: Impact of bimodality on Saikkonen-Sandberg's unit root test power

T	Preferred maximum, ERF in %	SS, ERF in %
	$(\phi_0, \varphi_0, \nu_0) = (0.85,0.5,4)$	
100	58	73
250	98	100
500	100	100
	$(\phi_0, \varphi_0, \nu_0) = (0.9,0.5,4)$	
100	39	54
250	96	99
500	100	100
	$(\phi_0, \varphi_0, \nu_0) = (0.95,0.5,4)$	
100	18	27
250	68	80
500	99	100

Notes: ERF stands for empirical rejection frequency. All figures are computed from 5,000 simulated MAR models. The ERFs for Saikkonen and Sandberg [2016] unit root test are the ones reported in Fig.2, p.114 therein.

when the sample size increases, as does the bimodality issue. In Section 5, we illustrate the benefits from implementing our approach by applying it to the Brent crude oil price series.

4 The proposed approach

4.1 Estimation strategy

The issue of parameter estimation based on multimodal criteria functions has a long history in the statistical literature. Barnett [1966] reviews different approaches, and although the computational possibilities have changed enormously over the last fifty years, the main principles of the possible approaches are still the same.

It is well-known that gradient ascent methods, such as the Newton estimation algorithm, will converge to a local maximum of the criteria function — highly depending on the initial values for the estimation, $\theta^{(0)}$ say. One approach in the literature is therefore to use estimation procedures that seek the global maximum by using algorithms allowing also downhill movements. Examples of such algorithms include simulated annealing (e.g. Kirkpatrick, Gelatt and Vecchi [1983]), genetic algorithms (e.g. Goldberg [1989]), or particle swarm methods (e.g. Zhang, Wang and Ji [2015]).

In some cases, however, it is of interest to characterize the entire likelihood surface, and to list the different local maxima. An example of this case is the MAR model with a root close to unity, where a forward root in the vicinity of unity can be excluded by economic reasoning, see e.g. Saikkonen and Sandberg [2016]. Selection between local maxima is also sometimes done in Markov-Switching modelling where a preferred "reasonable" model is chosen among the set of local maxima.

To obtain all local maxima for a given criteria function the standard approaches are either *i*) the use of randomized initial values, e.g. with $\theta^{(0)}$ drawn uniformly over a subset of the parameter space or *ii*) the use of a deterministically specified grid on the parameter space. In this paper we opt for the grid-search, and based on the insight from the characterization of the likelihood function for finite T as well as the limit $T \to \infty$ above, we choose a coarse but wide grid. In particular, we choose[5]

$$(\varphi^{(0)}, \phi^{(0)}, \sigma^{2(0)}, \nu^{(0)}) \in \{0.05, ..., 0.95\}^2 \times \{0.5q, ..., 2q\} \times \{3, ..., 50\}, \quad (7)$$

where q is a robust measure of the scale, here taken to be the inter-quartile range of the residuals obtained from a causal AR(2). For the simulations, we use four points for the grid on φ and ϕ and three points on σ and ν, as it has proven to be a good compromise between a quickly increasing number of starting values and the need to have enough of them to detect the bimodality. Because of the striking tendency of the likelihood estimation to produce interchanged roots, we consider — in addition — a sequence of initial values given by

$$(\varphi^{(0)}, \phi^{(0)}, \sigma^{2(0)}, \nu^{(0)}) = (\hat{\phi}, \hat{\varphi}, \hat{\sigma}^2, s\hat{\nu}), \quad \text{with} \quad s \in \{0.1, 0.5, 1, 2, 5\}, \quad (8)$$

where $(\hat{\phi}, \hat{\varphi}, \hat{\sigma}^2, \hat{\nu})$ are the (possibly local) estimates obtained from our standard initial values: Note that the backward and forward estimates are interchanged on purpose to initialize the new starting values. In total, this design of starting values covers $B = 4^2 3^2 + 6 = 150$ points in the parameter space[6]. For each starting point we apply the quasi-Newton optimization algorithm BFGS, see e.g. Nocedal and Wright [2006], and record the unique maxima.

As documented in Table 1, the grid-search algorithm often finds two maxima corresponding to cases with interchanged roots. In addition, it sometimes identifies local maxima with very large ν, corresponding to the maxima of the Gaussian likelihood function. The Gaussian local maxima have much lower likelihood in all cases, and will not be discussed further in this paper[7].

4.2 A modified unit root test

For estimation of the MAR model in Sections 2 and 3, we used this grid-search over initial values in order to characterize the entire likelihood surface and identify the global maximum. The grid-search procedure yields a list of unique maxima, and the MLE is taken to be the candidate with the highest maximum. As $T \to \infty$, this global maximization would obviously produce a consistent estimator.

[5]None of the models in this paper have negative roots, but if negative roots are likely, the grid could be easily extended.

[6]We have also considered alternative and less computationally expensive grid-designs, e.g. the interchanged roots in (8) alone, but it did not consistently locate all local maxima.

[7]Anyway, remind that in this case ϕ and φ are unidentified as $T \to \infty$.

For very persistent processes, however, where one root is in the vicinity of unity, the situation is slightly different. According to Table 1, there is a large probability of observing interchanged roots in finite samples, meaning that even if the true parameter set induces a causal unit root, $\phi_0 = 1$ and $\varphi_0 \ll 1$, both case $(a) : \hat{\phi} \approx 1$ and $\hat{\varphi} \ll 1$ and case $(b) : \hat{\phi} \ll 1$ and $\hat{\varphi} \approx 1$ are likely to occur. Case (b), however, is not reasonable from an *economic* point of view, because as $\hat{\varphi} \uparrow 1$ it implies that the behavior of the estimated model would be given by

$$y_t = \hat{\phi} y_{t-1} + v_t, \quad \text{with} \quad v_t = \sum_{i=0}^{\infty} \epsilon_{t+i}, \tag{9}$$

with the interpretation that agents would have to look into the infinite future without discounting.

In light of this lack of a reasonable economic interpretation when $\hat{\varphi} \approx 1$ and in light of the tendency towards interchanged roots, we therefore suggest to select amongst the local maxima whenever $\max(\hat{\varphi}, \hat{\phi}) \approx 1$. In this case we prefer models with $\hat{\phi} > \hat{\varphi}$, even if this maximum is only local and there exists a maximum with $\hat{\phi} < \hat{\varphi} \approx 1$ and a higher likelihood. This is similar to the approach sometimes retained in Markov switching models, where a "reasonable" solution is chosen within the set of local maxima.

To be precise, we suggest the following procedure:

Estimation procedure: Do a full search of local maxima for the MAR model, see Section 4.1 above.

i) If there is only one maximum of the likelihood function, it is selected as the preferred.
ii) If there are multiple maxima, and $\max(\hat{\varphi}, \hat{\phi}) \approx 1$, the maxima with $\hat{\phi} > \hat{\varphi}$ is the preferred one.

Based on this *preferred maximum*, we implement the unit root test of Saikkonen and Sandberg [2016] — thereby avoiding the spurious rejections whenever the roots are interchanged by random finite-sample estimation uncertainty. Remark that this procedure is feasible in practice and that the simulated size and power properties for the Saikkonen and Sandberg [2016] unit root test reported above are the ones to be expected in empirical applications. This is unlike the reported results from the simulations in Saikkonen and Sandberg [2016], where the correct maximum is obtained in the majority of replications because the estimation has been initialized at the true values of the DGP, such that the issue of bimodality is unsubstantial. This approach is obviously infeasible in practice.

Right part of Table 2 shows the empirical rejection frequencies for the unit root test based on our chosen maximum with $\hat{\phi} > \hat{\varphi}$. It can be seen that our proposed estimation and test strategy clearly outperforms the results obtained from the global maximum estimation — left panel of the Table. This is especially the case for small sample sizes $T \in \{100, 250\}$ and $\nu_0 = 10$. It is also worth noticing that the size gets close to the nominal as $T \to \infty$.

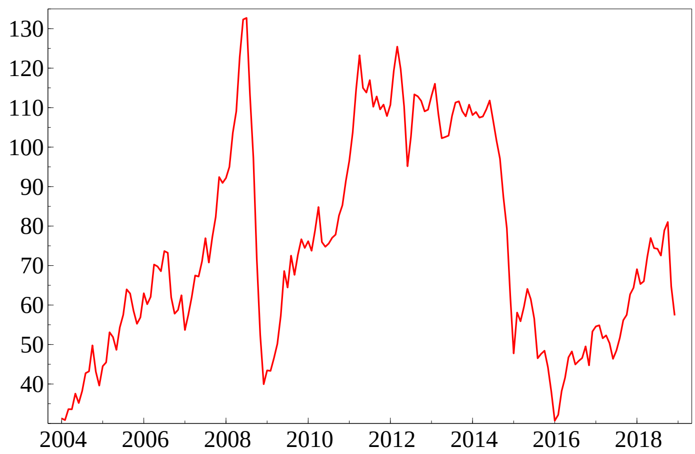

Figure 5: *Brent crude oil prices, US dollars per barrel.*

5 Application to Brent crude oil prices

In what follows, we provide an example which illustrates the bimodality-induced weaknesses of Saikkonen and Sandberg [2016]'s unit root test and the benefits of our proposed approach.

5.1 The data and unit root test

The data we analyse are monthly Brent crude oil prices, expressed in US dollars per barrel, from January 2004 to December 2018. The series[8] is plotted in Figure 5.

First, we find out that an AR(2) model is enough to eliminate any serial correlation in the residuals up to order 24. Then, the normality of the residuals is strongly rejected according to the Jarque and Bera statistics, with a test statistics of 8.44, which rejects the null of normality at the 1% level.[9]

Let us now turn to the unit root tests. From an ADF unit root test with two lags in levels and including an intercept, we find a statistics value of -2.63, corresponding to a $p-$value of 9%. Hence, the unit root is not rejected at the 5% level, but the conclusion is not so clear cut either. Let us now turn to Saikkonen and Sandberg [2016] unit root test. The results reported in Table 4 illustrate the bimodality of the Student's t ML estimation and the misleading conclusions it can induce.

Actually, as can be seen from this Table, two maxima co-exist for this series from the ML estimation. For this application, we have not used the method described earlier first, but have purposely started from the most intuitive (or neutral) values for the two roots, *i.e.* 0.5 for each of them. As can be seen from the first line of the results, this leads to a local maximum, with a lower log-likelihood than the one of the global maximum, and interchanged roots. As a result, Saikkonen and Sandberg [2016]'s unit root test wrongly concludes to a strong rejection of the null. As a matter of fact, the global maximum is reached when the values of ϕ and φ are initialized to 0.05 instead. In this case, reported

[8]The time series is taken from FRED Economic Data website and is named MCOILBRENTEU.

[9]All results regarding normality tests have been double-checked using the robust version of the Jarque and Bera statistic proposed by Gel and Gastwirth [2008] and the classical Anderson-Darling statistic: all conclusions are confirmed.

Table 4: Modified τ_T unit root test for demeaned series

Starting values		Estimation results			Unit root test		
$\phi^{(0)}$	$\varphi^{(0)}$	Log-lik	$\hat{\phi}$	$\hat{\varphi}$	τ_T	5% c.v.	10% c.v.
0.5	0.5	-555.70	0.46	0.92	**-5.26**	-2.83	-2.54
0.05	0.05	**-551.44**	0.94	0.40	-2.26	-2.81	-2.51

Notes: All estimations use the 176 observations between 2004m3 and 2018m10. Starting values for the scale and ν parameters are 2.5 and 10 respectively. The critical values reported here have been computed using case 2 of Table 1, p.110, in Saikkonen and Sandberg [2016].

in the second line of the results, the largest root is located in the backward component which in turn leads the unit root test not to reject the null. This is also the maximum which is found by our modified version of Saikkonen and Sandberg [2016]'s unit root test.

5.2 MAR models estimates

Table 5 reports the estimation results for all Student's t distributed MAR models where $r + s = 2$, namely the purely backward AR(2,0), the purely forward AR(0,2), and both the global and local ML estimates, denoted respectively $\text{MAR}_G(1,1)$ and $\text{MAR}_L(1,1)$. Fries and Zakoian [2019] suggest to select between candidate models based on the independence of residuals from estimated causal and noncausal models, see also Cavaliere, Nielsen and Rahbek [2019]. As an alternative to the maximum likelihood criterion, we have also tried several independence tests applied to estimated residuals in order to select between models with interchanged roots. For larger samples, some independence tests, e.g. the BDS test, see Brock, Dechert, Scheinkman and LeBaron [1996] were found to do rather well, but the overall conclusion was that for small and moderate sample sizes (with less than 500 observations, say) the model selection based on maximizing the likelihood outperforms any independence test. Accordingly, all candidate models have been estimated from the exactly same period so that their estimated likelihoods are perfectly comparable, *i.e.* from 2004m3 to 2018m10.

These results confirm the mixed nature of this autoregressive process, with an estimated backward root smaller than the estimated forward root. Even though smaller, the latter is still significantly different from zero. This noncausal component might be interpreted as capturing the epochs of bubble build-up and burst, as well as nonfundamentalness of shocks. The latter can in turn be seen as evidence that the econometrician uses less information than economic agents do.

6 Conclusion

This paper aimed at emphasizing and circumventing the undesirable consequences of bimodality of the widely used Student's t likelihood function applied in modelling mixed causal-noncausal autoregressions. The main consequence is that the MLE can end up

Table 5: MAR estimation results for demeaned Brent crude oil price series.

	Specifications			
	AR(2,0)	AR(0,2)	$MAR_G(1,1)$	$MAR_L(1,1)$
ϕ_1	**1.36**		**0.94**	**0.46**
	(0.07)		(0.03)	(0.10)
ϕ_2	**-0.39**			
	(0.07)			
φ_1		**1.35**	**0.40**	**0.92**
		(0.07)	(0.08)	(0.04)
φ_2		**-0.38**		
		(0.07)		
σ	**27.74**	**25.17**	**22.46**	**25.83**
	(4.54)	(4.64)	(4.59)	(4.49)
$\sqrt{\nu}$	**4.26**	**2.88**	**2.56**	**3.02**
	(2.39)	(0.89)	(0.71)	(0.94)
Log-lik	-551.98	-555.39	**-551.44**	-555.70
J-B p-val.	0.03	<0.01	<0.01	<0.01

Notes: Estimations use 176 effective observations. MAR_G and MAR_L refer to global and local MLE respectively. Standard errors are given in () and J-B p-val. gives the p-value of the Jarque and Bera normality test.

in a local maximum where the backward and forward roots are interchanged, hence impacting severely the interpretation of the series dynamics. Another important consequence of this bimodality concerns the unit root test procedure developed by Saikkonen and Sandberg [2016], as interchanged roots will mislead the test: any time the largest root is wrongly located in the forward component of the MAR model's estimate, the test on the backward root will wrongly reject the unit root null. This affects significantly both the size and power of the test. For this reason, an estimation strategy relying on grid search is proposed so as to characterize the entire likelihood surface and list all local maxima. In the particular context of unit backward root testing, this can in turn be used to eliminate the meaningless maxima where $\hat{\varphi} \approx 1$. This proposed strategy is shown to correct both size and power of Saikkonen and Sandberg [2016] unit root test very efficiently. The relevance of our proposed approach is illustrated from the Brent crude oil price for which a naive choice of starting values for the roots would have led to a local minimum with interchanged roots, hence leading to reject the unit root null wrongly.

Of course, this paper is only a first insight into this bi- or multi-modality issue in the MAR estimation and testing literature. The extension of our results to the more general MAR(r,s) model is on our research agenda.

References

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

Barnett, V., Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots, *Biometrika*, 1966, *53*, 151–65.

Breidt, F. and R. Davis, Time-reversibility, identifiability and independence of innovations for stationary time series, *Journal of Time Series Analysis*, 1991, *13*, 377–390.

—— , —— , and W. Dunsmuir, On Backcasting in Linear Time Series Models, in D. et al. Brillinger, editor, *New Directions in Time Series Analysis. Part I*, New-York: Springer-Verlag, 1992, pp. 25–40.

—— , —— , K.-S. Lh, and M. Rosenblatt, Maximum likelihood estimation for noncausal autoregressive processes, *Journal of Multivariate Analysis*, 1991, *36*, 175–198.

Brock, W., W. Dechert, J. Scheinkman, and B. LeBaron, A Test for Independence Based on the Correlation Dimension, *Econometric Reviews*, 1996, *15*, 197–235.

Cambanis, S. and I. Fakhre-Zakeri, On prediction of heavy-tailed autoregressive sequences: forward versus reversed time, *Teoriya Veroyatnostei i ee Primeneniya*, 1994, *39*, 294–312.

—— and —— , Forward and Reversed Time Prediction of Autoregressive Sequences, *Journal of Applied Probability*, 1996, *33*, 1053–1060.

Cavaliere, G., H.B. Nielsen, and A. Rahbek, Bootstrapping Noncausal Autoregressions: With Applications to Explosive Bublle Modeling, *Journal of Business and Economic Statistics*, 2019. Forthcoming.

Findley, D., On bootstrap estimates of forecast mean square errors for autoregressive processes, in D. Allen, editor, *Computer Science and Statistics: The Interface*, Amsterdam: North-Holland, 1986, pp. 11–17.

Fries, S. and J.-M Zakoian, Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles, *Econometric Theory*, 2019. Forthcoming.

Gel, Y.R. and J.L. Gastwirth, A robust modification of the Jarque–Bera test of normality, *Economics Letters*, 2008, *99*, 30–32.

Goldberg, D., *Genetic Algorithms in Search, Optimization and Machine Learning*, Reading, MA: Addison-Wesley Professional 1989.

Gouriéroux, C. and J. Jasiak, Filtering, Prediction and Simulation Methods for Noncausal Processes, *Journal of Time Series Analysis*, 2016, *37*, 405–430.

—— and J.-M. Zakoian, On Uniqueness of Moving Average Representation of Heavy Tailed Stationary Processes, *Journal of Time Series Analysis*, 2015, *36*, 876–887.

——— and ———, Local explosion modelling by noncausal process, *J. R. Stat. Soc. B*, 2017, *79*, 737–756.

Hanfelt, J., Eliminating Multiple Root Problems in Estimation: Comment, *Statistical Science*, 2000, *15*, 333–334.

Hecq, A., L. Lieb, and S. Telg, Identification of Mixed Causal-Noncausal Models in Finite Samples, *Annals of Economics and Statistics*, 2016, *123/124*, 307–331.

Hencic, A. and C. Gouriéroux, Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates, in Studies in Computational Intelligence: Econometrics of Risk 2015, pp. 17–40.

Huang, J. and Y. Pawitan, Quasi-likelihood Estimation of Noninvertible Moving Average Processes, *Scandinavian Journal of Statistics*, 2000, *27*, 689–710.

Kirkpatrick, S., C. Gelatt, and M. Vecchi, Optimization by Simulated Annealing, *Science*, 1983, *220*, 671–80.

Lanne, M., A. Luoma, and J. Luoto, Bayesian Model Selection And Forecasting In Noncausal Autoregressive Models, *Journal of Applied Econometrics*, 2012, *27*, 812–830.

——— and P. Saikkonen, Noncausal Autoregressions for Economic Time Series, *Journal of Time Series Econometrics*, 2011, *3*, 1–32.

——— and ———, Noncausal Vector Autoregression, *Econometric Theory*, 2013, *29*, 447–481.

Lawrance, A., Directionality and reversibility in time series, *International Statistical Review*, 1991, *59*, 67–79.

Lof, M., Noncausality and asset pricing, *Studies in Nonlinear Dynamics and Econometrics*, 2013, *17*, 211–220.

——— and H. Nyberg, Noncausality and the commodity currency hypothesis, *Energy Economics*, 2017, *65*, 424–433.

Nocedal, J. and S. Wright, *Numerical Optimization*, Springer-Verlag, 2006.

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

Rosenblatt, M., A note on prediction and an autoregressive sequence, in S. et al. Cambanis, editor, *Stochastic Processes, A Festschrift in Honour of Gopinath Kallia*, New-York: Springer-Verlag, 1993, pp. 291–295.

———, *Gaussian and Non-Gaussian Linear Time Series and Random Fields*, New-York: Springer-Verlag, 2000.

Saikkonen, P. and R. Sandberg, Testing for a Unit Root in Noncausal Autoregressive Models, *Journal of Time Series Analysis*, 2016, *37*, 99–125.

Small, C., J. Wang, and Z. Yang, Eliminating Multiple Root Problems in Estimation, *Statistical Science*, 2000, *15*, 313–341.

Weiss, G., Time-reversibility of linear stochastic processes, *Journal of Applied Probability*, 1975, *12*, 821–836.

Wu, R. and R. Davis, Least Absolute Deviation Estimation for General Autoregressive Moving Average Time-Series Models, *Journal of Time Series Analysis*, 2010, *31*, 98–112.

Zhang, Y., S. Wang, and G. Ji, A comprehensive survey on particle swarm optimization algorithm and its applications, *Mathematical Problems in Engineering*, 2015, *2015*.

A Derivations for Remark 1

The density of the Gaussian distribution is given by

$$f(\epsilon_t|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right), \tag{A.1}$$

where

$$\epsilon_t = (1-\phi L)(1-\varphi L^{-1})y_t = (1+\phi\varphi)y_t - \phi y_{t-1} - \varphi y_{t+1}, \tag{A.2}$$

and the approximate average likelihood function for the Gaussian MAR(1,1) model is

$$\log L_T(\phi,\varphi,\sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{T}\sum_{t=1}^{T}((1-\phi L)(1-\varphi L^{-1})y_t)^2.$$

For simplicity we focus on (ϕ, φ) and fix $\sigma^2 = \sigma_0^2$, such that

$$\log L_T(\phi,\varphi) = -\frac{1}{2}\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\frac{1}{T}\sum_{t=1}^{T}((1+\phi\varphi)y_t - \phi y_{t-1} - \varphi y_{t+1})^2.$$

Under the stationarity condition, $|\varphi_0| < 1$ and $|\phi_0| < 1$, it holds that

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{p} \gamma_0, \quad \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} \xrightarrow{p} \gamma_1, \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t+1} \xrightarrow{p} \gamma_2, \tag{A.3}$$

and the limit of the likelihood per observation is

$$\begin{aligned}\log L_\infty(\phi,\varphi) \;=\; & -\tfrac{1}{2}\log(2\pi\sigma_0^2) - \left(1+\phi^2\varphi^2+\phi^2+\varphi^2+2\phi\varphi\right)\gamma_0 \\ & + \left(2\phi^2\varphi + 2\phi\varphi^2 - 2\phi - 2\varphi\right)\gamma_1 - 2\phi\varphi\gamma_2. \end{aligned} \tag{A.4}$$

The limiting likelihood function in (A.4) is symmetric in ϕ and φ, which gives the results i Remark 1.