



Munich Personal RePEc Archive

Ubiquitous multimodality in mixed causal-noncausal processes.

Kindop, Igor

Kiel University

9 July 2021

Online at <https://mpra.ub.uni-muenchen.de/109594/>
MPRA Paper No. 109594, posted 06 Sep 2021 14:44 UTC

Ubiquitous multimodality in mixed causal-noncausal processes

Igor Kindop*
Kiel University

September 4, 2021

Abstract

According to the literature, the bimodality of estimates in mixed causal–non-causal autoregressive processes is due to unlucky starting values and happens only occasionally. This paper shows that a unique and convergent solution is not always the case for models of this class. Instead, the likelihood function is not convex leading to the multimodality of estimated parameters. It can be attributed to the magnitude and sign of the autoregressive coefficients. Simultaneously, the number of local modes grows with the number of autoregressive parameters in the model. This multimodality depends on the parameters of the process and the chosen error distribution. We have to apply grid search methods to extract candidate solutions. The independence of residuals is a necessary hypothesis for the proper identification of the processes. A simple AIC criterion helps to select an independent model. Finally, I sketch a roadmap on estimating mixed causal-noncausal autoregressive models and illustrate the approach with Brent spot oil price returns.

Keywords: non-causal model, non-convex likelihood, non-Gaussian, nonfundamentality, multimodality.

JEL classification: C13, C22, C51, C52, C53, E37

1 Introduction

Non-causal autoregressive models gain popularity in recent years. The term "non-causal" may lead to the idea of the future affecting the current state. Instead, the non-causal

*E-mail: stu210866@mail.uni-kiel.de

component can be seen as a sustained expectational component incorporating our best current knowledge of possible future scenarios. This reasoning follows Hansen and Sargent (1991), who suggested using these models to solve the non-fundamentalness problem of a rational expectations model. Also, the non-causal processes help model speculative bubble phenomena and asymmetric cycles. Finally, Lanne and Saikkonen (2011) suggest comparing non-causal autoregressive with non-invertible moving average processes, which provides a wide range of potential model applications.

The non-causal autoregressions are successful in capturing the effects of omitted variables or mitigating them. Lof (2013) shows that a univariate non-causal process provides a good approximation for a multivariate process with omitted elements. He also finds that non-causal models outperform causal alternatives when data is generated by a nonlinear process. Lof and Nyberg (2017) emphasize that univariate non-causal autoregressions fit the data better when the actual data-generating process is multivariate. By introducing leads of the data, the non-causal autoregression captures causal effects induced by unavailable variables.

In this vein of discussion, it is fruitful to consider the fundamentalness issue. Under the fundamentalness assumption, we believe the econometrician has the same data as agents do. Thus, given all the relevant information, the modeled process can be seen as a sequence of unseen shocks. In other words, the autoregressive structure can be consistently inverted into a moving average process - and this implies causal structure. As part of a large-scale multivariate causal model, a few series do not contain all the information agents possess. Usually, factor models solve the practical non-fundamentalness problem since they extract information from a lot of data. The non-causal models are, in turn, non-fundamental: the econometrician has less information than agents do. Despite the lack of information on actual regressors, leads of the data approximate the effect of unknown relevant predictors. Consequently, non-fundamentalness is modeled by the non-causal framework.

Non-causal models also serve as an indicator of heterogeneous information available for agents. These models can approximate large systems — an excellent engineering example provided in Lu et al. (2019). They compare non-causal models with state-of-the-art systems using high-dimensional gain matrices in a large-scale MIMO framework.

Non-causal autoregressive models are nonlinear and have no closed-form solution. Breidt, Davis, Lii, and Rosenblatt (1991) developed an MLE estimator for the non-causal models. The necessary assumption for estimation is the non-normality of the data, which is often an empirical economic observation. If data is Gaussian, the model

is not identifiable directly (Rosenblatt (2000)). Recently, attempts to identify non-causal models with Gaussian innovations have been made by Lu et al. (2019).

Different fat-tailed parametric assumptions are used in order to support inference on non-causal models. Lanne and Saikkonen (2011) used Student's t -distribution. Gouriéroux and Zakoian (2013) applied Cauchy distribution to their analysis of non-causal processes. Huang and Pawitan (2000), Wu and Davis (2010), as well as Hecq, Lieb, and Telg (2015) advocated the use of Laplace distribution or the LAD estimator. The latter work has shown a critical observation: if the innovations are just slightly non-normal, the parameter estimates have non-symmetric and non-normal distribution around the true values.

Student's t -distribution has a seemingly alluring property - it approximates almost any degree of heavy-tails from Gaussian to Cauchy in both limits. Additionally, it does not belong to the exponential class of distribution, and the logarithmic transformation of the likelihood is not very convenient. At the same time, the Laplacian assumption is parsimonious in parameters and simple in logarithmic transformation. Besides MLE and LAD estimators, Gouriéroux and Jassiak (2017) designed another semiparametric method for estimating non-causal VAR. It aims at getting linearly and nonlinearly uncorrelated residuals associated with the true parameter set. The decision on whether the residuals are correlated or not relies on a combination of Ljung-Box-type test statistics.

Wu and Davis (2010) mention the objective function to be non-convex, but they do not consider how seriously this fact affects the parameter estimates. Recently, Bec, Nielsen, and Saidi (2020) discuss a relevant problem within the context of near-unit root processes and a simple case with positive parameter values.

This paper aims to show that unimodality is instead a unique feature for mixed causal-noncausal processes. In some cases, the number of modes may be as large as the number of distinct roots in the process. Multimodality implies that optimization algorithms crucially depend on starting values. By estimating the model only a few times, we may not find a global optimum.

In this paper, I focus on the multimodality of the objective function. I illustrate how the phenomenon is associated with autoregressive parameters: their magnitude and sign. It is not guaranteed that the majority of estimates are located near the true parameters. I suggest an independence condition is necessary to determine the global optimum. The optimal model always has the maximized likelihood and minimized AIC of the residuals. A parametric assumption plays a significant role as well. In the appendix, I discuss the relevance of the actual distribution of innovations for estimation efficiency.

The rest of the paper consists of five sections. The second section discusses non-

causal models, associated peculiarities, and, briefly, the estimation approach. The third chapter illustrates the issue of multimodality with respect to different parameter settings of mixed causal-noncausal processes. I show when the multimodality of the objective function occurs and demonstrate why it is a potential problem. The fourth section considers an empirical application for the Brent and WTI oil spot price returns. Finally, the last section concludes the paper.

2 A General Model

Noncausal autoregressive processes comprise leading and lagging components in the structure. Breidt et al. (1991) and Lanne and Saikkonen (2011) use the following representation of the lag polynomial:

$$\phi(L^{-1})\psi(L)y_t = \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2), \quad (1)$$

where

$$\psi(L) = (1 - \psi_1 L - \psi_2 L^2 - \dots - \psi_r L^r) = (1 - \lambda_1^\psi L)(1 - \lambda_2^\psi L) \dots (1 - \lambda_r^\psi L)$$

is the lag polynomial with r possibly distinct roots $|\lambda_j^\psi| < 1$, and

$$\phi(L^{-1}) = (1 - \phi_1 L^{-1} - \phi_2 L^{-2} - \dots - \phi_s L^{-s}) = (1 - \lambda_1^\phi L^{-1})(1 - \lambda_2^\phi L^{-1}) \dots (1 - \lambda_s^\phi L^{-1})$$

is the lead polynomial with s distinct roots $|\lambda_j^\phi| < 1$.

A purely causal process has $\lambda_j^\phi = 0, \forall j = 1 \dots s$. Similarly, we get a purely noncausal process if $\lambda_j^\psi = 0, \forall j = 1 \dots r$. Otherwise, it is a mixed causal-noncausal (MAR) process.

We can rewrite Eq. 1 and invert the lead polynomial:

$$\begin{aligned} \phi(L^{-1})\psi(L)y_t &= \epsilon_t \\ \psi(L)y_t &= \phi(L^{-1})^{-1}\epsilon_t \end{aligned}$$

Which (for MAR(1,1)) has a noncausal MA(∞) representation, since the lead polynomial is invertible (in future):

$$\psi(L)y_t = \epsilon_t + \phi\epsilon_{t+1} + \phi^2\epsilon_{t+2} + \dots + \phi^n\epsilon_{t+n} + \dots \quad (2)$$

$\psi(L)$ is an invertible lag polynomial (in past) with a regular MA(∞) representation, then y_t has a double-sided MA representation. Hence, the model from eq. 2 is nonfundamental: the space spanned by current and past values of y_t is not contained in the

space spanned by current and past values of ϵ_t . Following the Definition 1 in Alessi, Barigozzi, and Capasso (2011), the lag-lead polynomial $\psi(L)\phi(L^{-1})$ has r eigenvalues inside the unit disk and s eigenvalues outside the disk.

The double-sided MA representation for the mixed causal-noncausal processes is therefore:

$$y_t = \sum_{i=-\infty}^{\infty} \pi_i \epsilon_{t-i}, \quad (3)$$

where the sequence $\{\pi_i\}$ are the coefficients in front of z_i in the Laurent series expansion of $(\phi(z^{-1})\psi(z))^{-1}$, which exists in an annulus $d < |z| < d^{-1}$ for some $d < 1$, see Brockwell and Davis (1991). Causal autoregressive roots correspond to the right arm, while non-causal roots associated with the left arm of the MA. This MA representation is a natural way to justify and interpret non-causal autoregressions, see Figure 2.

A classical one-sided MA representation, or an impulse response function, can be seen as a particular case of the double-sided one - with zero coefficients in the left arm. In this sense, MAR allows a more general solution for the impulse response coefficients making a classical autoregressive process a special case of the MAR.

Lanne and Saikkonen (2011) suggested a connection between noncausal autoregressive and noninvertible moving average processes. Indeed, consider a noninvertible MA(1) with $|\theta| > 1$: $z_t = (1 + \theta L)e_t$. It can be rewritten as $z_t = \theta L(1 + \theta^{-1}L^{-1})e_t$, which is invertible in future.

Following Brockwell and Davis (1991) and Lanne and Saikkonen (2011) we can see, how the noncausal model MAR(r,s) can be estimated. Consider two auxillary processes: $u_t = \phi(L^{-1})y_t$ and $v_t = \psi(L)y_t$, such that $\phi(L^{-1})v_t = \epsilon_t$. Then,

$$\begin{bmatrix} u_1 \\ \vdots \\ u_{T-s} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} y_1 - \phi_1 y_2 - \cdots - \phi_s y_{s+1} \\ \vdots \\ y_{T-s} - \phi_1 y_{T-s+1} - \cdots - \phi_s y_T \\ y_{T-s+1} - \psi_1 y_{T-s} - \cdots - \psi_r y_{T-s+1-r} \\ \vdots \\ y_T - \psi_1 y_{T-1} - \cdots - \psi_r y_{T-r} \end{bmatrix} = A \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix}$$

Next,

$$\begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} - \psi_1 y_r - \cdots - \psi_r y_1 \\ \vdots \\ u_{T-s} - \psi_1 y_{T-s-1} - \cdots - \psi_r y_{T-s-r} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = B \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} \\ \vdots \\ u_{T-s} \\ v_{T-s+1} \\ \vdots \\ v_T \end{bmatrix}$$

In matrix notation, we have

$$\begin{aligned}
x &= Ay \\
z &= Cx \quad \text{or} \\
z &= CAy
\end{aligned}$$

The processes v_t and u_t are independent since they are built with (non) causal polynomial operators. The joint pdf of z under true parameters may be expressed as

$$h_U(u_1, \dots, u_r) \left(\prod_{t=r+1}^{T-s} f(\epsilon_t; \sigma, \lambda) \right) h_V(v_{T-s+1}, \dots, v_T).$$

Where, h_U and h_V denote the joint pdf functions of first and last components of the residuals (u_1, \dots, u_r) and (v_{T-s}, \dots, v_T) . In terms of the data y_1, \dots, y_T we have:

$$\begin{aligned}
& h_U(\phi(L^{-1})y_1, \dots, \phi(L^{-1})y_r) \left(\prod_{t=r+1}^{T-s} f(\phi(L^{-1})\psi(L)y_t; \sigma, \lambda) \right) \times \\
& h_V(\psi(L)y_{T-s+1}, \dots, \psi(L)y_T) |\det(A)|
\end{aligned}$$

The determinant is independent of the size T , and because r and s are small relative to T , we can approximate the joint pdf by central component and run the Approximate ML estimator. The details of the method can be seen in Lanne and Saikkonen (2011).

The final question regarding the estimation method is what parametric assumption to choose for the likelihood function. The mixed causal-noncausal processes are somewhat similar to the infinite variance autoregressive processes and should have heavy-tailed innovations. In this paper, I proceed in a direction of Wu and Davis (2010) and Hecq et

al. (2015). However, instead of LAD estimator I use Laplacian MLE. An application of plain absolute deviation minimizer worsenes the multimodality situation:

$$\{\hat{\psi}, \hat{\phi}, \hat{\sigma}\} = \arg \min_{\psi, \phi, \sigma} -L(y_t | \psi, \phi, \sigma) \quad (4)$$

$$L(y_t | \psi, \phi, \sigma) = -(n \ln 2\sigma + \frac{1}{\sigma} \sum_{t=s+1}^{n-r-1} |y_t - a(\hat{y}_t | \hat{\psi}, \hat{\phi})|) \quad (5)$$

where $a(\hat{y}_t | \hat{\psi}, \hat{\phi}) = \hat{\phi}(L^{-1})\hat{\psi}(L)y_t$. Laplacian parametric distribution assumes the LAD estimator, but the scale parameter participates in weighing the observations picking up a better combination of parameters. The difference can be seen by comparing Table 3 and Table 4, where the first one reports LAD minimization and the second - Laplacian likelihood maximization. There might be a link to quantile regression as well, but the mixed causal-noncausal processes are nonlinear and are not a good match for linear quantile regression.

The parameter estimates are asymptotically Normal, as shown in Wu and Davis (2010) for the LAD estimator and Andrews, Davis, and Breidt (2006) for a general ML estimator. In the non-causal setting, a consistent estimate of a lag order is obtained by minimizing Akaike's information criterion (AIC). At the same time, BIC usually underestimates the number of non-causal autoregressive components, as in Andrews and Davis (2013). AIC for purely causal structures suggests $r + s$ lags and leads. For example, a lag order of 2 obtained with the AIC approach suggests not only AR(2) but two non-causal models (MAR(1,1) and MAR(0,2)) – 3 models in total.

By definition, a mixed causal-noncausal process consists of two autoregressive polynomial parts. We may expect the process to drastic changes if there is a non-causal component. However, it is not always the case. Instead, the behavior of a dominant autoregressive arm is affected by the behavior of a latent one. What is dominant or latent depends on the magnitude of each polynomial's autoregressive roots. A positive non-causal term smoothes the process and increases the amplitude of the causal part. This property is useful in simulating speculative bubbles. A negative non-causal root acts as a frequency modifier of the process and induces clusters of volatility. Figure 1 illustrates this behavior in detail. Non-causal processes with negative roots have properties similar to volatile processes – additional variability is coupled with clusters of volatility, for example, various asset returns data.

3 Multimodality of Estimates

In this chapter, I illustrate multimodality of the objective function in simulations and work with the following model:

$$\begin{aligned}\phi(L^{-1})\psi(L)y_t &= \epsilon_t \\ \epsilon_t &\sim \text{Laplace}(0, \sigma)\end{aligned}\tag{6}$$

Parameter σ is a scale parameter for the Laplacian distribution. It is not comparable to a regular variance or Gaussian scale parameter.

Before the discussion, let's define what is unimodal and what is not. Strictly speaking, there is multimodality if an optimization converges to several optimal parameters. However, we may allow some non-identical results to cluster around the true value due to the rounding error from a practical perspective.

Proposition 1 (Different cases of Multimodality). *Estimated parameters are strictly unimodal if they form a bell shape. The parameters are mildly multimodal if no more than 20% of them are distributed significantly far from the true value. Parameters are heavily multimodal if more than 20% of realizations are located in different parts of the hyperplane.*

Next, I show that strictly unimodal distributions are not general results for mixed causal-noncausal processes. Non-convexity of the likelihood function depends on coefficients of the autoregressive polynomials and their sign. Thus, multimodality is not a special case but a general problem related to non-causal models.

In the following, we develop an auxiliary tool that helps us understand some patterns of multimodality.

First, we can rewrite the lag and lead polynomials as a vector product:

$$\psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \dots = \begin{bmatrix} 1 & 1 & \dots \end{bmatrix} \begin{bmatrix} 1 \\ -\psi_1 L \\ -\psi_2 L^2 \\ \dots \end{bmatrix} = V_1 \Psi,\tag{7}$$

and

$$\phi(L^{-1}) = 1 - \phi_1 L^{-1} - \phi_2 L^{-2} - \dots = \begin{bmatrix} 1 & 1 & \dots \end{bmatrix} \begin{bmatrix} 1 \\ -\phi_1 L^{-1} \\ -\phi_2 L^{-2} \\ \dots \end{bmatrix} = V_1 \Phi,\tag{8}$$

where V_1 is a vector of ones. Next, we can rewrite the lag and lead polynomial as a product of the representations above

$$\begin{aligned} \phi(L^{-1})\psi(L) &= (V_1\Phi)(V_1\Psi)' = V_1\Phi\Psi'V_1' = V_1CV_1' \\ C &= \begin{bmatrix} 1 \\ -\phi_1L^{-1} \\ -\phi_2L^{-2} \\ \dots \end{bmatrix} \times \begin{bmatrix} 1 & -\psi_1L^1 & -\psi_2L^2 & \dots \end{bmatrix} = \begin{bmatrix} 1 & -\psi_1L^1 & -\psi_2L^2 & \dots \\ -\phi_1L^{-1} & \psi_1\phi_1 & \phi_1\psi_2L^1 & \dots \\ -\phi_2L^{-2} & \phi_2\psi_1L^{-1} & \psi_2\phi_2 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \end{aligned} \quad (9)$$

Matrix C may help to describe mixed causal-noncausal processes and to simulate them. We can establish a sufficient condition for the multimodality of estimates.

3.1 A Symmetric Model

Proposition 2 (Sufficient condition for Multimodality). *Let $\psi(L)$ be a stable lag polynomial with r distinct roots and $\phi(L^{-1})$ be a stable lead polynomial with s distinct roots, and $r = s$. Suppose neither of eigenvalues of $\psi(L)$ or $\phi(L^{-1})$ is near unity. Then, a mixed causal-noncausal process shows multimodality of estimates if*

$$T_C = \text{tr}(C) = 1 + \sum_{i=1}^r \psi_i\phi_i < 1.$$

This result holds if ϕ and ψ are sufficiently large.

Consider MAR(1,1) model:

$$\begin{aligned} (1 - \psi L)(1 - \phi L^{-1})y_t &= e_t \\ (1 - \psi L - \phi L^{-1} + \phi\psi)y_t &= e_t \\ (1 + \phi\psi)y_t - \psi y_{t-1} - \phi y_{t+1} &= e_t \\ y_t - \frac{(\psi y_{t-1} + \phi y_{t+1})}{1 + \phi\psi} &= \frac{1}{1 + \phi\psi} e_t \end{aligned}$$

The trace of matrix C enters the reduced form of the mixed causal non-causal process directly and affects the variance of the residual component. Processes with uniform signs have a trace greater than one and processes with alternating signs - smaller than one.

Next, consider simulations for MAR(1,1) model with absolute parameter values $\{\psi_0; \phi_0\} = [0.8; 0.3]$. The particular numbers are close to the values reported in Lanne and Saikkonen (2011). There are four sign combinations possible, and the corresponding

contours of the likelihood functions are plotted in Figure 3. The diagonal plots are for the parameters with uniform signs, and the off-diagonal plots are for the parameters with alternating signs.

The likelihood hypersurface associated with uniform processes is unimodal, but there is a local optimum in other cases. There is a unimodal distribution for diagonal and multimodal for the off-diagonal plots. In latter cases, the modes are somewhat mirrored with respect to the 45° line. If we choose parameters that are close to the dissecting line, the multimodality disappears. Unfortunately, hypersurface plots are not available for other MAR processes.

For the case of multimodal MAR(1,1), there are two centers of attraction for the estimator, i.e., for the top right panel of Figure 3, they are: $\{\psi_A; \phi_A\} = \{0.3; -0.8\}$ and $\{\psi_B; \phi_B\} = \{-0.8; 0.3\}$. If we find an optimum, the second candidate is close to the reflection of the first one. This is usually sufficient to reach both modes of distribution. The resulting distribution for MAR(1,1) is illustrated in Figures 3 and 4.

For processes of a different order, we can track multimodality through the distribution of an autoregressive parameter. The estimator converges to a cluster of similar vectors from different starting values. Thus, we can get a potential optima list by grid search procedure and extract the correspondent clusters.

For a more complex MAR(2,2) model, we have two stable model setups. First, all autoregressive coefficients are inside the unit disk, i.e., their sum in each arm does not exceed one in absolute value. Under this condition, all sign combinations of parameters yield stable and stationary models. For this scenario ("wide case"), we have 16 different stable and stationary models.

Next, some autoregressive parameters may be larger than one. The number of stable and stationary models is less than the number of all possible sign combinations. If only one autoregressive coefficient is larger than one - there are eight combinations. If coefficients in both arms exceed unity - there are just four stationary processes. I will call this scenario a "narrow case."

In simulations, I use the following scheme to visit all regions of the admissible parameter space. I estimate the MAR(r, s) model on a grid of starting values inside a $(r + s)$ -dimensional hypercube with a side $[-0.95; 0.95]$. I divide the parameter space into an equally spaced grid of n intervals to get $n^{(r+s)}$ sub-hypercubes. Then I randomly draw a point from each sub-hypercube and use it as a starting value for the MLE. Next, I get Laplacian ML estimates and obtain a grid of optimal parameters. Such an algorithm visits many regions of the hyperplane with relatively few attempts. A researcher quickly

gets a wrong set of estimated parameters by choosing unlucky starting values. For some parameters, the local mode(s) may be visited more often than the global one.

If all 16 models are stationary, the multimodality of the objective function is related to autoregressive parameter values. The smaller are the parameters in magnitude, the more often estimates are around the true value.

First, consider the processes with $\psi_0 = (0.4, 0.1)$ and $\phi_0 = (0.3, 0.2)$. All sign combinations result in stable and stationary processes. Optimization yields either unimodal or (depending on the particular innovation sequence) mildly multimodal distribution. The illustration is in Figure 5. Perhaps, this set of parameters is sufficiently close to the imaginable 45° symmetry line.

If we increase the parameter values in causal and non-causal parts of the process, there is a gradual decrease in the number of unimodal or mildly multimodal solutions. The general pattern reminds the chessboard pattern of matrix C .

For MAR(2,2) we have the following reduced form:

$$y_t - \frac{(\psi_1 - \phi_1\psi_2)y_{t-1} + (\phi_1 - \psi_1\phi_2)y_{t+1}}{1 + \phi_1\psi_1 + \phi_2\psi_2} = \frac{1}{1 + \phi_1\psi_1 + \phi_2\psi_2}e_t.$$

Consider MAR(2,2) model with absolute parameter values $\psi_0 = (0.5, 0.3)$ and $\phi_0 = (0.6, 0.2)$. The trace of matrix C is $\text{tr}(C) = 1 + \phi_1\psi_1 + \phi_2\psi_2$ and can be found from the Table 1. Here, main diagonal setups have unimodal or mildly multimodal hypersurface, and models on the main antidiagonal represent multimodal cases.

For autoregressive parameters with moderate magnitude, unimodal distributions occur in 50% cases according to the chessboard pattern of matrix C . You can compare Figure 6 with Table 1. In Figure 6, the horizontal line plots have the same ordering concerning the sign pattern as the vertical plots with respect to the non-causal polynomial: $((+,+), (-,+), (+,-), (-,-))$.

The chessboard pattern is not a unique one. With different autoregressive parameters, one can reach another matrix C . However, if parameter values are large enough, the chessboard pattern breaks down, and more elements of matrix C do not reflect empirical multimodality. Especially this situation is relevant when a root of a polynomial approaches unity. Figure 7 demonstrates this situation with MAR parameter values: $\psi_0 = (0.8, 0.7)$ and $\phi_0 = (0.7, 0.2)$. But if the parameter values are small enough, the results are uniformly distributed everywhere.

If we increase the parameters further, we have some unstable sign combinations. Now, consider the situation, when $\psi_0 = (1.2, 0.3)$ and $\phi_0 = (0.5, 0.3)$. The process is

stable if ψ_0 has an alternating sign pattern: $((+,-), (-,-))$, while the sign pattern in non-causal coefficients may vary. The analog of Table 1, in this case, consists of two last columns.

The particular distribution of estimated parameters can be seen in Figure 8. For a model with a large autoregressive parameter, the resulting likelihood function of a stable process is almost always multimodal. Only in the first case with $\psi_0 = (1.2, -0.3)$ and $\phi_0 = (0.5, 0.3)$ the parameter estimates are mildly multimodal, and a share of correct hits exceeds 80%.

Next, consider Figure 9, which reports the distribution of the first autoregressive parameter estimates of a process with $\psi_0 = (0.5, 0.3)$ and $\phi_0 = (1.2, 0.3)$. The non-causal part has a parameter greater than one, and causal component parameters are inside the unit disk.

Finally, consider the process with $\psi_0 = (1.7, 0.8)$ and $\phi_0 = (1.5, 0.6)$. In this case, we end up with four possible sign combinations. An example of such a simulation is in Figure 10.

In this case, the pattern may depend on a difference between autoregressive coefficients of the corresponding polynomial or the autoregressive polynomial roots. The antidiagonal plots in Figure 10 have very few correct hits and show a minimal probability of picking up the correct parameters. When roots of an autoregressive polynomial approach unity, the multimodality gets wilder.

The multimodality considered in this block results from inappropriate starting values. By construction, stochastic search uses points from the unit cube. If the true values do not belong to the unit cube, local maxima are chosen too often. However, in practice, we do not know true parameters. Hence, such an approach is an appropriate standardization.

3.2 An Asymmetric Model

The degree of multimodality in asymmetric models is attributed to signs and magnitudes of autoregressive coefficient of causal and non-causal parts of the process.

If the true autoregressive parameters are small in magnitude, the parameter estimates are uniformly distributed or mildly multimodal. Here, the modes of likelihood function merge and form a unimodal hypersurface. An example of such a result is Figure 12. There are no sets of parameters with a multimodal distribution.

If we increase the parameters to some moderate level, more multimodal outcomes

occur. Eventually, the cases appear according to the following proposition.

Proposition 3 (Sufficient condition for Multimodality). *Let $\psi(L)$ be a stable lag polynomial with r distinct roots and $\phi(L^{-1})$ be a stable lead polynomial with s distinct roots. Suppose neither of eigenvalues of $\psi(L)$ or $\phi(L^{-1})$ is near unity and ϕ and ψ are sufficiently large. Then, a mixed causal-noncausal process shows multimodality of estimates if*

$$T_C = 1 + \sum_{i=1}^{\min(r,s)} \psi_i \phi_i < 1,$$

else the process is unimodal.

This is equivalent to the trace condition for symmetric processes, despite there is no trace for rectangular matrices. In practice, an asymmetric model with one of the polynomial having more roots than the other is more relevant. Consider, MAR(2, 1) model or eq. (3) with $r = 2; s = 1$, and absolute parameter values $\{\psi_0; \phi_0\} = [(0.6; 0.3); (0.8)]$.

We have seen that the multimodality of the objective function is related to the sign and magnitude of the autoregressive parameters. In this simplest asymmetric case, the multimodality problem is also a significant issue. For instance, Table 6 provides the clusters of parameter estimates of a standardized MAR(2, 1) process. Figure 11 illustrates the distribution of the first autoregressive parameter.

Perhaps multimodality depends on an appropriate choice of the likelihood function. Consider the case with $\psi_0 = (0.8, 0.1)$ and $\phi_0 = (0.8)$. The distribution of parameter estimates can be seen in Figure 13. The two stochastic grid runs are reported in Table 3 and 4. The first one relies on LAD minimization and comprises $6^3 = 216$ draws from the stochastic grid with only 5% hit probability. The second uses Laplacian MLE and contains $10^3 = 1000$ simulations with a significantly higher correct hit probability.

At the same time, the true mode is not necessarily the most frequently visited, as illustrated in Table 5. Sometimes, the local optimum may even have all-positive autoregressive parameter estimates, even if the true parameters are not. In this case, we may be lured to a trap by a nicely looking parameter vector.

It may seem evident that the magnitude of a non-causal component is solely responsible for the multimodality problem. However, it is not always true. Consider the process with parameter values $\psi_0 = (0.8, 0.1)$ and $\phi_0 = (0.1)$. Here, the non-causal root is small in magnitude, but the likelihood function is not unimodal. Figure 14 indicates a multimodal distribution of estimates. Albeit the non-causal component is small, it

significantly affects parameter estimates distribution. The resulting set of histograms is multimodal in each case.

These are sound arguments in favor of the stochastic grid search approach for MAR estimation. The multimodality observation per se and situation when the true model is not always dominant show that grid search is the only available tool to support our choice of an appropriate solution.

3.3 Simulation Results

We can summarize the results for symmetric and asymmetric models in the following way. First, if true parameters (both causal and non-causal) are small in magnitude, we have either unimodal or mildly multimodal distribution of parameter values around the proper set.

Next, if the true parameters are moderately large, we have a chessboard pattern for multimodality of estimates. In this case, matrix C is a helpful tool to explore the cases of multimodality.

Finally, if autoregressive parameters are large, parameter estimates are almost always either mildly or significantly multimodal. In the extreme case of near-unit roots, the distribution of parameter estimates is often multimodal and concentrated around false values with only a few correct hits.

Technically, any estimation method relies on the analysis of correctly identified residuals. However, in mixed causal-noncausal models, residuals are a nonlinear function in autoregressive parameters. Therefore, it might be challenging to come up with correct residuals first of all. There may be a connection between multimodality and eigenvalues of causal and non-causal parts. However, a direct comparison of eigenvalues or distances between the largest causal and non-causal eigenvalues does not help to guess the multimodality pattern.

Yet, there is a valuable observation about a possible number of modes in the likelihood function. For the $MAR(r,s)$ process, there are usually not more than $r + s$ different converging local and global modes. The main contribution of this paper is an emphasis on the multimodal structure of the parameter estimates from mixed causal-noncausal processes.

There are many competing models and no prominent tool to discriminate between them (especially when data is not fat-tailed enough). Lanne and Saikkonen (2011) suggested using the Ljung-Box test on autocorrelation for regular(LBT) and squared

(McLeodLi test or absolute (LBT2)) residuals and pick a model with independent residuals. This approach requires the independence of residuals as our identifying assumption.

The distribution of the Ljung-Box test statistic alone is bimodal and has a hyper-surface similar to the likelihood function. Thus, linear independence is not a sufficient criterion for model discrimination. The distribution of the Ljung-Box test of absolute transformations is unimodal, but the test is almost prohibitively restrictive.

Fortunately, we can extract possible modes of the likelihood function and get associated parameter sets. Next, we need to check the residuals associated with the parameter sets for independence. AIC is a good pointer on such independence.

In principle, AIC checks the remaining autoregressive structure in residuals associated with the parameter sets. Since there is a discrepancy between how the AIC is calculated (Gaussian likelihood, causal structure) and how we get the residuals (Laplacian likelihood, non-causal structure), some information may still exist in the residuals. However, if AIC yields optimal lag length different to 0, we conclude the residuals to remain dependent.

Proposition 4 (AIC for model identification (post-AIC)). *Suppose we have a set of competing models: $\psi(L)\phi(L^{-1})y_t^{(i)} = \epsilon_t^{(i)}$. We calculate AIC with respect to the autoregressive structure for $i = 1, \dots$, possible models. We consider the models with the smallest estimated lag length of the residuals.*

This idea is conceptually similar to the LBT criterion, but it works better in the considered problem. The post-AIC may drop unsatisfactory models. However, it cannot be a sole criterion due to its discrete nature. Although post-AIC is almost always zero in globally optimal cases, it may yield zeros in suboptimal and misspecified models (see Table 7).

If there are several models with minimal post-AIC value, the conclusion has to be made with the objective function criterion. Ideally, the three criteria should coincide: best objective function value, the least AIC and the best LBT (LBT2) statistic. For example, table 7 shows that the globally optimal parameter set has minimal objective function value, smallest possible post-AIC value, and a highly significant Ljung-Box test statistic. This rule is the best among many I have considered and works well for different objective functions.

4 Empirical Application

Lof (2013), Lof and Nyberg (2017) They claim that noncausality can often be found in financial data. The true price process may be incredibly complex, but we can approximate it with a simple non-causal process. The purpose of this chapter is to show multimodality in real applications.

For the empirical part, I use monthly crude oil price data ("Brent" and "WTI") obtained from the US. Energy Information Administration. The data covers 1M1990 - 10M2020. The prices are "calculated by EIA from daily data by taking an unweighted average of the daily closing spot prices for a given product over the monthly time period."

Next, I approximate growth rates by taking the first differences of price logarithms, see Figure 15. We estimate the optimal lag length by minimizing the Akaike Information Criteria (AIC), which is four for both series. Thus, we can choose between 5 models: one purely causal AR(4), one purely non-causal MAR(0,4), and three mixed causal-noncausal MAR(4-k, k), where $k = 1, 2, 3$.

To validate the application of non-causal models, we need to test the underlying probability distribution. If our data is Gaussian – we apply causal methods and dispose of non-causal regressions (see Degression for details). If data is fat-tailed - the non-causal options are allowed. In the Degression and Table 11 I illustrate the non-causal inference is one of the most efficient for Laplacian data. Unfortunately, testing Laplacian distribution is not a standard procedure in econometrics. There is only a limited number of tools developed, in contrast to testing Gaussianity. I use the following two methods to infer probability distribution.

First, we can assume the data follows Exponential Power Distribution, which nests Gaussian and Laplacian as special cases. We can distinguish them by estimating a specific shape parameter. This EPD nests a continuum of leptokurtic and platykurtic cases. Thus, we can rule out Gaussianity and other platykurtic possibilities.

$$g(x; \mu, \sigma, b) = \frac{1}{2\sigma b^{1/b}\Gamma(1 + 1/b)} \exp \left[\frac{-1}{b} \left| \frac{x - \mu}{\sigma} \right|^b \right] \quad (10)$$

The shape parameter can be estimated either by MLE or by the method developed by Mineo (2003) and used in Franke (2014). If the parameter $\hat{b} = 2$, the data is Gaussian. If it is smaller - the data are leptokurtic. If $\hat{b} = 1$ - the data is Laplacian. Finally, if $\hat{b} > 2$, the distribution is platykurtic.

Second, we can refer to the ratio of maximized likelihoods (RML test). This test was described in Balakrishnan, Kannan, and Nagaraja (2004, pp. 65-79) to discriminate

between Laplacian or Gaussian distribution. Within this procedure, we can have a glimpse on Type I and Type II errors and analyze both hypotheses.

$$Z_n(\theta_0, \theta_1) = \prod_{i=1}^n \frac{p(X_i, \theta_0)}{p(X_i, \theta_1)} \quad (11)$$

Define a ratio of maximized likelihoods and construct a test on Normality (or Laplacianity) ¹. According to Balakrishnan et.al. (2005), the test problem can be written in two useful ways:

Problem 1 : H_0 : Normal *vs.* H_1 : Laplace

Problem 2 : H_0 : Laplace *vs.* H_1 : Normal

For each problem, there is a different rejection rule. Critical values for the Gaussian null hypothesis are 9.46 and 5.98 under 95% and 99% significance level, respectively. The critical values for Laplacian null hypothesis are -10.90 and -4.36 under 95% and 99% significance level, respectively. The minimum sample sizes for the test are 130 and 259 observations correspondingly. Table 8 provides various Normality tests. Generally, Normality is rejected, and the Laplacian hypothesis is not.

Next, I estimate candidate models by Laplacian MLE from a grid of starting values. For each model suggested by $AIC = 4$, we have a 4-dimensional parameter hypercube. Then, I split each side of the hypercube into six intervals and get a grid of $6^4 = 1296$ regions to explore. Finally, I pick a random starting point from each region and use it as a grid search element.

The purely causal and non-causal models are unimodal by construction. However, mixed causal-noncausal models are multimodal. For each MAR, we have a table of 1296 estimates, and I report the distribution of the first autoregressive parameter.

There are four possible autoregressive parameters, and four candidate modes are possible. Thus, I pick the four largest bins of the first causal parameter histogram. Then, I extract the best parameter set from each bin with the maximal objective function value. At this moment, I have four candidate modes per each mixed causal-noncausal model plus two unimodal solutions for pure (non)causal models - 14 candidate solutions in total for each oil price index. These solutions are presented in Tables 10 and 9.

For WTI and Brent, the best model according to the log-likelihood function criterion is MAR(2,2). On the other hand, most models reject autocorrelation in residuals by

¹Here I use the following sample statistics and rewrite the ratio of likelihoods in logarithmic form as:
 $T = \frac{n}{2} \ln 2 - \frac{n}{2} \ln \pi + n \ln \hat{\theta} - n \ln \hat{\sigma} + \frac{n}{2}$

means of the standard Ljung-Box test(at 8 lags). Albeit, the opposite is applicable to the same test with absolute transformations of residuals.

For Brent oil, no model satisfies the Ljung-Box test on independence for absolute residuals. This test is very restrictive, even for simulations. MAR(2,2) is the only model with zero post-AIC. Although most models show independence according to the standard LBT, they still have some autoregressive structure according to AIC.

For WTI oil price, the models (1) and (2) in Table 10 are almost identical with parameter values and log-likelihood. There are more competitive models with respect to the post-AIC criterion. Here, four models produce independent residuals: a purely causal AR(4), MAR(2,2), and two MAR(1,3) models (# 2 and # 3 are almost identical). Each represents a version with independent residuals. Finally, we pick up the model with the best objective function value - the mixed causal-noncausal model. Here we see, that the post-AIC argument supports the choice taken based on multimodality analysis.

Despite the empirical multimodality, we have tools to discriminate the best model out of the set of candidate solutions. However, there may be situations when the difference between candidate models is minuscule. In this case, the candidate solutions should be reported.

Generally, we can interpret the results using impulse response curves. The MAR(2,2) model has a two-sided moving average representation. The right arm approximates the causal polynomial with negative roots and oscillates around zero. The left-arm decays exponentially in the past and describes the non-causal polynomial. The impulse response curves can be seen on the right part of Figure 16 and roughly correspond to the one depicted in the second plot of the third panel in Figure 2.

There are several layers of interpretation of the non-causal regression. First, the non-causal component captures possible nonlinearities in the data and suggests speculative behavior specific to the spot oil market. The left arm of the impulse response curve sets up some kind of anticipation phase. The causal part of the impulse response describes the behavior attributed to a correction phase.

Next, the non-causal part of the univariate impulse response may be generated by omitted variables. Following Lof(2013), we can suggest the true price is a multivariate process with possibly unobserved determinants. Those determinants are responsible for the anticipation phase. It is important that market participants observe this information.

Finally, the anticipation phase indicates that oil price alone does not contain all relevant information. Therefore, there is a non-fundamentality problem for univariate series. In the case of Brent oil, classical methods produce suboptimal results in terms of

fit and residual autocorrelation.

Albeit MAR(2,2) is the best WTI model according to the three chosen criteria jointly, it is not optimal in terms of model parsimony. Classical AR(4) is worse in likelihood value, the coefficients are not statistically significant, but it is a linear model, in contrast to MAR(2,2) or MAR(1,3). Lastly, causal AR(4) is inferior relative to the purely non-causal MAR(0,4) that has better fit to the data, and similarly insignificant parameters.

Another interesting observation is the minuscule difference between MAR(2,2) and the best of MAR(1,3) models. Appealing to the argument of statistical significance, MAR(1,3) has sharper estimates than MAR(2,2) model. However, this argument has to be discarded by the log-likelihood function value criterion.

Both WTI and Brent have very similar realizations and behave slightly differently only in the past ten years. Thus they are governed mainly by the same economic (and stochastic) laws. The identical results show that non-causal models can be successfully augmented to the econometrician's toolbox.

Empirically, the oil market is often driven by expectations of production increases or cuts, i.e., OPEC decisions or US oil reserves changes. Directions of these movements are usually predictable but not a magnitude (i.e., a cut or an increase). The anticipation phase represents this predictability. Moreover, a unit shock does not have a magnitude of 1 at the impact. Due to the predictability, the shock is spread among few periods indicating how predictability affects the efficiency of political action.

5 Conclusion

In this paper, I discuss parameter estimation and identification problems in mixed causal-noncausal autoregressive processes. Under the correct parametric assumption, the likelihood function is often non-convex and leads to multimodality of parameter estimates. Appropriate starting points are crucial in attaining the global optimum. Unfortunately, there is no ubiquitous starting point sufficient to achieve the true optimum. Instead, the grid search procedure is implemented to find possible optima. A researcher has to choose a model with the highest objective function value, given the independence of residuals.

The multimodality occurs when the autoregressive parameters are moderately large and differ in sign. Usually, the likelihood function is unimodal for purely non-causal models and mixed causal-noncausal with uniform parameter signs. However, the sign does not matter if the largest autoregressive coefficient is small in magnitude, leaving the likelihood unimodal. For moderate autoregressive cases, the trace of matrix C matches

the observed multimodality sufficiently well. Finally, if autoregressive parameters are large, the parameter estimates distribution becomes multimodal irrespective of the sign.

Under certain circumstances (coefficients close to the unit root, certain sign combinations), parameter estimates are concentrated around false mode. In this case, only a few starting points hit the global maximum correctly. This fact requires a grid search approach to visit various regions of the parameter space.

Lii and Rosenblatt (1996) discuss grid search to estimate non-gaussian non-minimum phase ARMA sequences and illustrate multimodality of an ARMA(1, 1) model with a causal AR component and non-invertible MA part. Following Lanne and Saikkonen (2011) we may relate non-causal autoregressive processes to non-invertible moving average processes. In this case, the source of multimodality may come from the non-invertible moving average representation of non-causal autoregressions.

The inference on the mixed causal-noncausal autoregressive processes may be conducted in the following steps.

1. First, a researcher tests if data follows Laplacian (or any other fat-tailed) distribution;
2. Second, the lag order is estimated consistently by minimizing AIC for classical causal models;
3. Third, the estimated MAR lag length as $(r + s)$ a sum of possible lag orders in causal/non-causal parts. Thus, we have $r + s + 1$ possible autoregressive models.
4. Later, for each candidate model, a grid search approach with various starting values must be considered. It is important to visit all regions of the parameter space.
5. Next, we pick up models with independent residuals (or the least dependent) using post-AIC criterion and choose the model with the maximal objective function value. Optimally, the model residuals should pass the Ljung-Box test on linear and nonlinear autocorrelation jointly.

The multimodality poses a significant challenge for univariate non-causal models. Simultaneously, the problem is magnified for non-causal vector autoregressions. I leave this complication for future research.

Appendix

How important it is to have non-Gaussian data?

Current degression extends the debate on MAR identification and estimation by considering different parametric distributions of the innovation sequences. Multimodality can be traced and analyzed with sufficiently fat-tailed distribution but not caused by a particular distribution. Rather, it is a model's response to a nonlinear autoregressive structure. If the noise component is not sufficiently fat-tailed, the false mode may have a higher likelihood function value in (relatively) small samples. In the extreme case, with Gaussian innovations, there is no mode but a uniform distribution of estimates over some domain.

Again, I consider MAR(1,1) model with parameter absolute values $\psi_0 = 0.8$ and $\phi_0 = 0.3$. I simulate the processes 500 times for several sample sizes $n = (100; 300; 500; 1000)$ and distributions {Gaussian, Student- t ($\nu = 20$), Student- t ($\nu = 9$), Student- t ($\nu = 3$), Laplacian }. Table 11 reports how frequent the first causal parameter falls within the interval $[0.75; 0.85]$. Additionally, the table provides mean estimate of the first autoregressive parameter, obtained from all converging runs from different starting values. This statistic indicates precision of our approach. In each case, I pick the result with globally maximized likelihood function value.

First, Gaussian innovations prevent model identification and estimation in a sensible interval. If innovations are Student t -distributed with 20 degrees of freedom, the percentage of correct hits ranges from 25% in a sample size $n = 100$ to 65% for sample size $n = 1000$. Next, innovations with Student's t ($\nu = 9$) distribution, have a correct hits range from 30% for $n = 100$ to 85% for large sample. Finally, Laplacian or t -distributed with 3 degrees of freedom innovations are similar in performance and have at least 57% of correct hits in small samples. For a large sample, the share of correct hits may exceed 99%.

These results can be generalized in the following way. First, the inference is better for a large sample. We observe the share of correctly identified models grows significantly when the sample size increases from $n = 100$ to $n = 1000$.

Second, a lot of tail-events are necessary to identify and estimate a mixed causal-noncausal model correctly. For $n = 100$, the percentage of correct hits increases from 16% for Gaussian cases to about 61% with Laplacian innovations. For large samples, there are even greater differences.

Thus, MAR estimation is appropriate in small samples if there is a guaranteed fat-

tailedness. A similar conclusion is drawn by Hecq et al. (2015). However, they consider models with all-positive parameters.

We cannot estimate the MAR process with Gaussian innovations. The Nonnormality of innovations is a necessary assumption for the MAR process to be identified. It is essential to avoid the temptation of applying MAR for subtly non-normal residuals. The stronger is Nonnormality, the more often a global optimum is a proper one. Eventually, if the sample is large and Laplacian, the proper estimation frequency converges to 1. Thus, it is crucial to test data on Laplacianity to provide enough justification for MAR application.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. The data that support the findings of this study are openly available at

https://www.eia.gov/dnav/pet/pet_pri_spt_s1.m.htm

or at the following link:

https://osf.io/x9w7p/?view_only=605db51038ce45699fa0d22badd2b210.

Bibliography

References

- Alessi, L., Barigozzi, M., & Capasso, M. (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, 79(1), 16-47.
- Andrews, B., & Davis, R. A. (2013). Model identification for infinite variance autoregressive processes. *Journal of Econometrics*, 172, 222 - 234.
- Andrews, B., Davis, R. A., & Breidt, F. J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, 97, 1638–1659.
- Balakrishnan, N., Kannan, N., & Nagaraja, H. (2004). *Advances in ranking and selection, multiple comparisons, and reliability. statistics for industry and technology*. Birkhauser Boston.

- Bec, F., Nielsen, H., & Saidi, S. (2020). Mixed causal–noncausal autoregressions: Bimodality issues in estimation and unit root testing. *Oxford Bulletin of Economics and Statistics*.
- Breidt, F. J., Davis, R. A., Lii, K.-S., & Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, *36*, 175-198.
- Brockwell, P., & Davis, R. (1991). *Time series: series and methods*. (2nd ed.). Springer series in statistics.
- Franke, R. (2014). How non-normal is us output? *Metroeconomica.*, *66*.
- Gorieroux, C., & Zakoian, J. (2013, April). Explosive bubble modelling by noncausal process. *Working Paper. Center for Research in Economics and Statistics.*
- Gourieroux, C., & Jassiak, J. (2017). Noncausal var: Representation, identification and semi-parametric estimation. *Journal of Econometrics.*
- Hansen, L., & Sargent, T. (1991). Rational expectations econometrics. In (p. 77-119). Westview Press, Inc.
- Hecq, A., Lieb, L., & Telg, S. (2015). Identification of mixed causal-noncausal models. how fat should we go? *Working Paper RM/15/035, Maastricht University*.
- Huang, J., & Pawitan, Y. (2000). Quasi-likelihood estimation of noninvertible moving average processes. *Scandinavian Journal of Statistics*, *27*, 689-710.
- Lanne, M., & Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, *3*.
- Lii, K.-S., & Rosenblatt, M. (1996). Maximum likelihood estimation for nongaussian nonminimum phase arma sequences. *Statistica Sinica*, *6*, 1-22.
- Lof, M. (2013). Noncausality and asset pricing. *Studies in Nonlinear Dynamics & Econometrics*, *17*(2), 211-220.
- Lof, M., & Nyberg, H. (2017). Noncausality and the commodity currency hypothesis. *Energy Economics*, *65*, 424-433.
- Lu, Q., Loewen, P., Gopaluni, R., Forbes, M., Backstroem, J., Dumont, G., & Davies, M. (2019). Identification of symmetric noncausal processes. *Automatica*, *103*, 515-530.

Mineo, A. (2003). On the estimation of the structure parameter of a normal distribution of order p . *Statistica*, 63(1), 109-122.

Rosenblatt, M. (2000). *Gaussian and non-gaussian linear time series and random fields*. Springer-Verlag.

Wu, R., & Davis, R. (2010). Least absolute deviation estimation for general autoregressive moving average time series models. *Journal of Time Series Analysis*, 31(2), 98-112.

Tables

noncausal \ causal	(+,+)	(-,+)	(+,-)	(-,-)
(+,+)	1.36	0.76	1.24	0.64
(-,+)	0.76	1.36	0.64	1.24
(+,-)	1.24	0.64	1.36	0.76
(-,-)	0.64	1.24	0.76	1.36

Table 1: Trace of matrix C for different sign combinations of the MAR(2,2) process with $\psi_0 = (0.5, 0.3)$ and $\phi_0 = (0.6, 0.2)$.

	Causal		NonCausal		σ	AIC		LBT (8 lags)		LLV (minus)	N 1296
	ψ_1	ψ_2	ϕ_1	ϕ_2		MA	AR	Linear	Absolute		
Simulated	-0.500	0.300	-0.600	0.200							
# 1	-1.665 (0.161)	-0.682 (0.163)	0.522 (0.181)	-0.032 (0.188)	30.433 (13.607)	1	2	0.042	0.000	25.544	165
# 2	-1.268 (0.307)	-0.312 (0.293)	0.152 (1.353)	0.300 (0.643)	30.803 (13.689)	4	4	0.000	0.000	25.635	3
# 3	-0.503 (0.166)	0.288 (0.190)	-0.580 (0.182)	0.207 (0.166)	28.027 (12.535)	0	0	0.147	0.111	25.131	940
# 4	0.624 (0.339)	-0.142 (0.257)	-1.753 (0.166)	-0.762 (0.161)	29.762 (13.317)	5	5	0.010	0.001	25.429	188

Table 2: There may be many solutions for the MAR(2,2) model. Distribution of the autoregressive parameter vector estimates according to different clusters in the MAR(2,2) model with $(-,+,-,+)$ sign pattern. (Mode 3 is associated with the true set.)

	ψ_1	ψ_2	ϕ_1	LBT p.val	LAD	AIC	N = 216
Simulated	0.8	0.1	-0.8				
Cluster 1	-0.01	0.83	-0.02	0.03	484.38	4	177
Cluster 2	0.01	0.81	-0.05	0.08	484.48	0	30
Cluster 3	0.81	0.09	-0.81	0.53	450.66	0	9

Table 3: An estimation with LAD minimizer, instead of Laplacian MLE. The result is generally the same. However in some specific cases, the frequency of a correct hit is extremely small. For this process it is less than 5%.

	Causal		Noncausal	σ	AIC		LBT (8 lags)		LLV (minus)	N 1000
	ψ_1	ψ_2	ϕ_1		MA	AR	Linear	Absolute		
Simulated	0.8	0.1	-0.8							
# 1	-1.007 (0.185)	-0.155 (0.170)	0.930 (0.098)	125.087 (62.610)	2	2	0.005	0.459	26.084	245
# 2	0.034 (0.203)	0.764 (0.168)	-0.069 (0.187)	123.199 (61.553)	0	0	0.043	0.109	26.031	168
# 3	0.781 (0.233)	0.094 (0.271)	-0.795 (0.198)	115.021 (57.480)	0	0	0.359	0.277	25.755	587

Table 4: MLE and LAD estimates are different. Distribution of the autoregressive parameter vector estimates according to different clusters in the MAR(2,1) model with (+,+, -) sign pattern. (Mode 3 is associated with the true set.)

	Causal		Noncausal	σ	AIC		LBT (8 lags)		LLV (minus)	N 1000
	ψ_1	ψ_2	ϕ_1		MA	AR	Linear	Absolute		
Simulated	0.8	-0.1	-0.8							
# 1	-0.206 (0.353)	0.595 (0.326)	0.177 (0.230)	152.396 (76.188)	2	2	0.086	0.025	26.879	608
# 2	0.821 (0.188)	-0.115 (0.236)	-0.801 (0.226)	144.412 (72.235)	0	0	0.845	0.133	26.662	392

Table 5: The most frequent mode is not necessarily the true one. Distribution of the autoregressive parameter vector estimates according to different clusters in the MAR(2,1) model with (+,-,-) sign pattern. (Mode 2 is associated with the true set.)

	Causal		Noncausal	σ	AIC		LBT (8 lags)		LLV (minus)	N 1000
	ψ_1	ψ_2	ϕ_1		MA	AR	Linear	Absolute		
Simulated	0.6	-0.300	-0.800							
# 1	-0.559 (0.188)	0.135 (0.163)	0.309 (0.232)	177.468 (88.714)	3	3	0.000	0.005	27.489	277
# 2	-0.303 (0.229)	0.258 (0.301)	0.060 (0.253)	178.527 (89.263)	3	3	0.000	0.001	27.512	32
# 3	0.573 (0.171)	-0.286 (0.170)	-0.766 0.164	161.954 (81.003)	0	0	0.750	0.964	27.121	691

Table 6: Distribution of the autoregressive parameter vector estimates according to different clusters in the MAR(2,1) model with (+,-,-) sign pattern. (Mode 3 is associated with the true set.)

	Causal	NonCausal		σ	AIC		LBT (8 lags)		LLV	N
	ψ_1	ϕ_1	ϕ_2		MA	AR	Linear	Absolute	(minus)	1000
Simulated	0.9	-0.6	0.3							
# 1	-0.966 (0.107)	1.223 (0.281)	-0.266 (0.269)	94.309 (47.094)	2	2	0.203	0.493	24.964	354
# 2	0.197 (0.230)	0.053 (0.184)	0.872 (0.153)	94.139 (47.077)	0	0	0.479	0.037	24.951	285
# 3	0.910 (0.150)	-0.657 (0.143)	0.238 (0.151)	88.879 (44.492)	0	0	0.958	0.959	24.717	361

Table 7: A non near unit root noncausal process may have a near unit root mirroring process. Distribution of the autoregressive parameter vector estimates according to different clusters in the MAR(1,2) model with (+,-,+) sign pattern. (Mode 3 is associated with the true set.)

	Data	AR(4)	MAR(3,1)	MAR(2,2)	MAR(1,3)	MAR(0,4)
J-B test	610.602	461.885	3516.667	3100.786	2936.761	827.530
<i>p</i> -value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Shapiro-Wilk	0.916	0.933	0.886	0.891	0.888	0.924
<i>p</i> -value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EPD: \hat{b}	0.984	1.074	0.895	0.903	0.895	0.981
RML test	-29.516	-21.847	-38.596	-37.169	-37.253	-29.969

Table 8: Strong rejection of Normality hypothesis and evidence in favor of Laplacian hypothesis. The EPD shape parameter and the RML test statistic for the data, and for the various model's residuals.

Model	Const	Causal				Noncausal				σ	AIC		LBT(8)		LLV	N
		ψ_1	ψ_2	ψ_3	ψ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ	MA	AR	Linear	Absolute	(minus)	
AR(4)	0.014 (0.027)	0.215 (0.174)	-0.149 (0.335)	0.009 (0.148)	-0.118 (0.312)					4.295 (1.753)	4	4	0.174	0.000	18.903	1296
MAR(3,1)																
1	0.008 (0.023)	-0.318 (0.080)	-0.219 (0.119)	-0.097 (0.088)		0.521 (0.083)				4.218 (1.722)	4	4	0.010	0.000	18.796	327
2	0.010 (0.022)	-0.114 (0.196)	-0.107 (0.213)	-0.033 (0.183)		0.358 (0.197)				4.219 (1.722)	4	4	0.227	0.000	18.796	519
3	0.010 (0.013)	-0.100 (0.431)	-0.105 (0.077)	-0.031 (0.164)		0.342 (0.272)				4.220 (1.723)	4	4	0.240	0.000	18.797	58
4	0.006 (0.024)	0.575 (0.100)	-0.201 (0.098)	0.010 (0.127)		-0.362 (0.112)				4.283 (1.751)	4	4	0.244	0.000	18.879	392
MAR(2,2)																
1	0.006 (0.024)	-0.485 (0.144)	-0.167 (0.153)			0.719 (0.084)	-0.236 (0.127)			4.159 (1.697)	0	0	0.417	0.000	18.711	436
2	0.005 (0.030)	-0.367 (0.264)	-0.080 (0.264)			0.597 (0.156)	-0.187 (0.317)			4.050 (1.598)	4	4	0.340	0.001	18.762	287
3	0.009 (0.024)	-0.057 (0.212)	-0.146 (0.189)			0.307 (0.097)	0.029 (0.175)			4.221 (1.722)	4	4	0.114	0.000	18.801	319
4	0.010 (0.015)	0.007 NA	-0.188 (0.146)			0.216 (0.227)	0.083 (0.092)			4.236 (1.733)	4	4	0.047	0.000	18.807	254
MAR(1,3)																
1	0.003 (0.025)	-0.367 (0.074)				0.631 (0.110)	-0.264 (0.112)	0.083 (0.084)		4.204 (1.717)	4	4	0.125	0.005	18.773	580
2	0.007 (0.014)	-0.220 (0.253)				0.446 (0.133)	-0.149 (0.079)	0.017 NA		4.233 (1.729)	4	4	0.200	0.000	18.815	140
3	0.009 (0.026)	-0.060 (0.090)				0.305 (0.111)	-0.060 (0.099)	-0.023 (0.099)		4.239 (1.733)	4	4	0.306	0.000	18.818	94
4	0.006 (0.021)	0.460 (0.111)				-0.285 (0.110)	-0.110 (0.139)	0.010 (0.061)		4.336 (1.767)	4	4	0.001	0.000	18.972	482
MAR(0,4)																
	0.007 (0.023)					0.274 (0.110)	-0.049 (0.101)	0.022 (0.081)	-0.087 (0.127)	4.199 (1.714)	4	4	0.210	0.000	18.768	1296

Table 9: Classical AR(4) is inferior to MAR(2,2) in explaining Brent spot oil price returns and contains autocorrelated residuals. In each mixed case, 4 best models reported. LBT is a p-value of the Ljung-Box test on autocorrelation with 8 degrees of freedom, LBT.a - same for residuals in absolute form. The columns 11 and 12 are AIC-based optimal lag length estimates with respect to AR and MA structures. The last column is the number of solutions nearby the chosen local minimum.

Model	Const	Causal				Noncausal				σ σ	AIC		LBT(8)		LLV (minus)	N
		ψ_1	ψ_2	ψ_3	ψ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4		MA	AR	Linear	Absolute		
AR(4)	0.007 (0.025)	0.231 (0.116)	-0.056 (0.127)	-0.052 (0.136)	-0.110 (0.117)					4.069 (1.660)	0	0	0.567	0.000	18.585	1296
MAR(3,1)																
1	0.010 (0.021)	-0.367 (0.242)	-0.224 (0.240)	-0.150 (0.168)		0.562 (0.120)				4.003 (1.633)	4	4	0.003	0.000	18.484	615
2	0.012 (0.020)	-0.115 (0.099)	-0.059 (0.079)	-0.125 (0.072)		0.334 (0.106)				4.021 (1.640)	4	4	0.264	0.000	18.514	267
3	0.010 (0.027)	-0.083 (0.278)	-0.057 (0.215)	-0.117 (0.127)		0.319 (0.266)				4.025 (1.643)	4	4	0.367	0.000	18.517	275
4	0.006 (0.021)	0.698 (0.122)	-0.369 (0.097)	0.089 (0.083)		-0.520 (0.106)				4.065 (1.660)	4	4	0.102	0.000	18.573	139
MAR(2,2)																
1	0.006 (0.022)	-0.421 (0.141)	-0.129 (0.171)			0.677 (0.097)	-0.249 (0.128)			3.957 (1.616)	0	0	0.668	0.006	18.411	757
2	0.007 (0.021)	-0.400 (0.021)	-0.148 (0.131)			0.675 (0.213)	-0.201 NA			3.958 (1.616)	0	0	0.486	0.006	18.412	166
3	0.002 (0.027)	0.559 (0.102)	-0.164 (0.077)			-0.375 (0.123)	-0.050 (0.178)			4.056 (1.658)	4	4	0.140	0.000	18.554	187
4	0.005 (0.020)	0.251 (0.124)	-0.214 (0.226)			0.006 (0.123)	0.198 (0.241)			4.066 (1.659)	4	4	0.183	0.000	18.576	186
MAR(1,3)																
1	0.008 (0.025)	-0.418 (0.071)				0.676 (0.060)	-0.302 (0.059)	0.043 (0.068)		3.966 (1.619)	0	0	0.401	0.014	18.424	886
2	0.008 (0.032)	-0.233 (0.203)				0.498 (0.239)	-0.198 (0.226)	0.006 (0.102)		4.010 (1.638)	0	0	0.532	0.000	18.490	33
3	0.012 (0.044)	-0.171 (0.184)				0.413 (0.167)	-0.138 (0.211)	-0.021 (0.236)		4.068 (1.681)	0	0	0.467	0.000	18.507	21
4	-0.001 (0.022)	0.493 (0.103)				-0.306 (0.105)	-0.101 (0.167)	-0.036 (0.105)		4.095 (1.671)	4	4	0.003	0.000	18.623	356
MAR(0,4) 4	0.008 (0.026)					0.237 (0.212)	-0.032 (0.183)	-0.008 (0.086)	-0.073 (0.083)	4.008 (1.635)	4	4	0.197	0.000	18.493	1296

Table 10: Classical AR(4) is inferior to MAR(2,2) in explaining WTI spot oil price returns. However, both produce residuals without autocorrelation. In each mixed case, 4 best models reported. LBT is a p-value of the Ljung-Box test on autocorrelation with 8 degrees of freedom, LBT.a - same for residuals in absolute form. The columns 11 and 12 are AIC-based optimal lag length estimates with respect to AR and MA structures. The last column is the number of solutions nearby the chosen local minimum.

Sign pattern	Distribution	$n = 100$		$n = 300$		$n = 500$		$n = 1000$	
		$\hat{\psi}_1$	share of successes						
(+,+)	Gaussian	0.550	0.202	0.540	0.258	0.535	0.292	0.549	0.328
	Student- $t, \nu = 20$	0.567	0.190	0.613	0.354	0.636	0.436	0.680	0.588
	Student- $t, \nu = 9$	0.628	0.280	0.698	0.466	0.731	0.584	0.751	0.776
	Student- $t, \nu = 3$	0.748	0.586	0.791	0.854	0.798	0.942	0.801	0.996
	Laplacian	0.749	0.586	0.787	0.872	0.794	0.946	0.798	0.996
(-,+)	Gaussian	-0.234	0.190	-0.193	0.284	-0.214	0.350	-0.269	0.446
	Student- $t, \nu = 20$	-0.293	0.216	-0.361	0.402	-0.389	0.496	-0.327	0.518
	Student- $t, \nu = 9$	-0.363	0.292	-0.436	0.480	-0.508	0.610	-0.442	0.636
	Student- $t, \nu = 3$	-0.621	0.562	-0.726	0.842	-0.746	0.918	-0.695	0.898
	Laplacian	-0.607	0.606	-0.656	0.828	-0.688	0.874	-0.639	0.852
(+,-)	Gaussian	0.217	0.202	0.233	0.310	0.259	0.382	0.264	0.460
	Student- $t, \nu = 20$	0.351	0.266	0.323	0.374	0.382	0.494	0.322	0.504
	Student- $t, \nu = 9$	0.385	0.292	0.478	0.486	0.481	0.580	0.449	0.648
	Student- $t, \nu = 3$	0.643	0.546	0.708	0.824	0.730	0.910	0.688	0.892
	Laplacian	0.600	0.586	0.663	0.810	0.682	0.878	0.665	0.874
(-,-)	Gaussian	-0.559	0.200	-0.561	0.260	-0.551	0.298	-0.563	0.396
	Student- $t, \nu = 20$	-0.564	0.234	-0.623	0.352	-0.644	0.442	-0.673	0.586
	Student- $t, \nu = 9$	-0.629	0.284	-0.692	0.490	-0.728	0.638	-0.756	0.792
	Student- $t, \nu = 3$	-0.754	0.594	-0.791	0.854	-0.796	0.942	-0.799	0.994
	Laplacian	-0.749	0.620	-0.796	0.894	-0.798	0.974	-0.800	0.998

Table 11: For the Laplacian data, the global optimum is most likely the true one. simulation study of MAR(1,1) with true parameters $\{\psi_0, \phi_0\} = \{0.8, 0.3\}$, different sign pattern (column 1), different innovations (column 2) and sample sizes. For each case a total of 500 models considered. For each vertical block $\hat{\psi}_1$ is a mean estimate of the first causal autoregressive parameter for the global modes. The share of successes is a frequency of global optima yielding the first causal estimate to be in $\hat{\psi}_1 = \psi_0 \pm 0.05$.

Figures

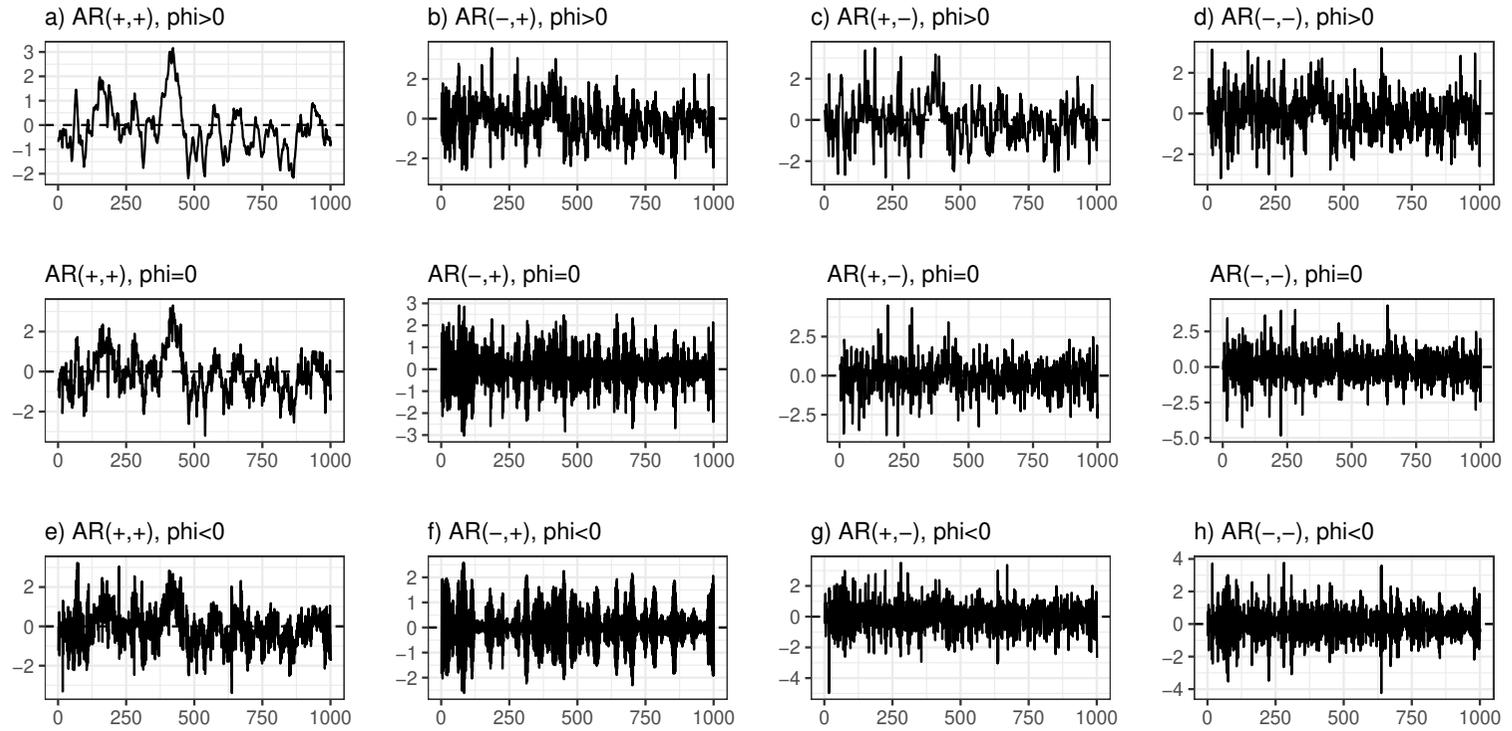


Figure 1: Simulated causal AR(2) and mixed causal-noncausal MAR(2,1) processes with the same innovations and parameters $\psi = (0.6, 0.3)$, $\phi = 0.8$. The first block refers to MAR(2,1) with a positive noncausal root, second - to AR(2), the third - MAR(2,1) with a negative noncausal root. The vertical blocks represent the sign structures of the causal autoregressive polynomial.

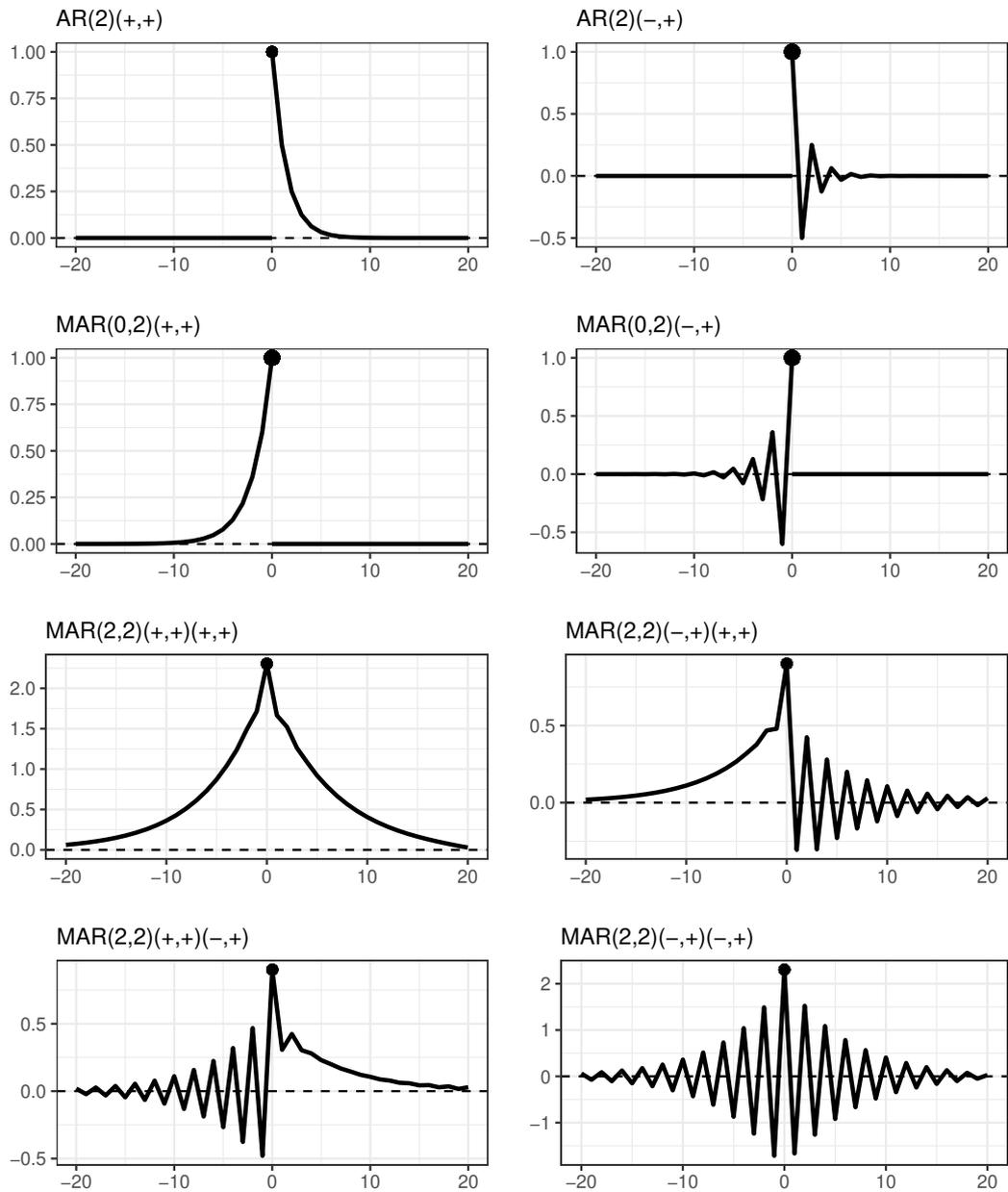


Figure 2: Simulated impulse responses from various autoregressive processes with $\psi = (0.5, 0.3), \phi = (0.6, 0.2)$. The upper row represent causal processes, second row - noncausal processes. The lower 4 rows represent mixed causal-noncausal processes. A dot represents a unit shock at $t = 21$.

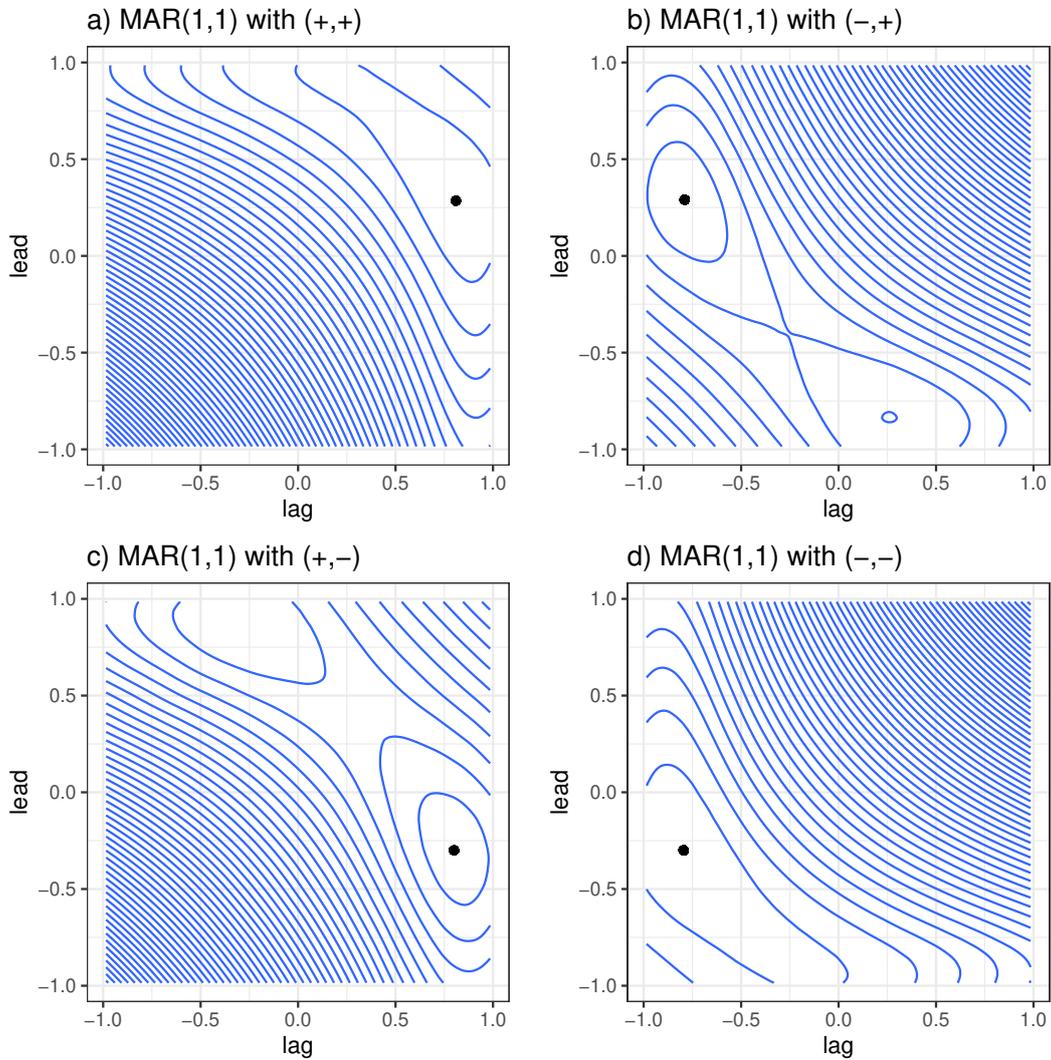


Figure 3: Simulated likelihood function contour plots for the MAR(1,1) process with $\psi = 0.8, \phi = 0.3$. The upper left plot is for all-positive autoregressive roots. The bottom right plot is for all-negative roots. The off-diagonal plots represent processes non-uniform in roots.

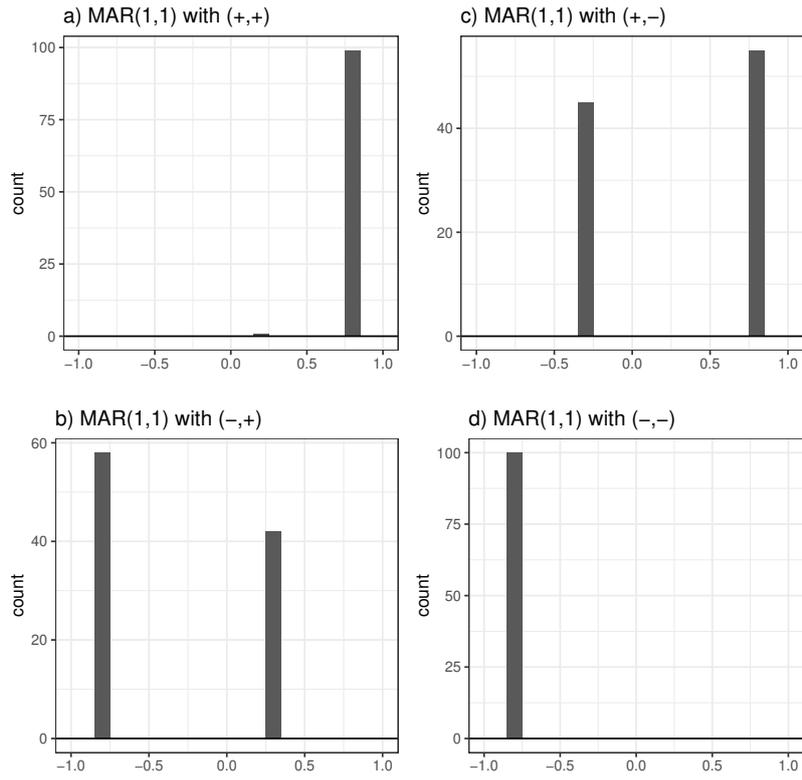


Figure 4: Distributions of the first causal autoregressive parameter of the MAR(1,1) process with $\psi = 0.8, \phi = 0.3$. The upper left plot is for all-positive autoregressive roots. The bottom right plot is for all-negative roots. The off-diagonal plots represent processes non-uniform in roots.

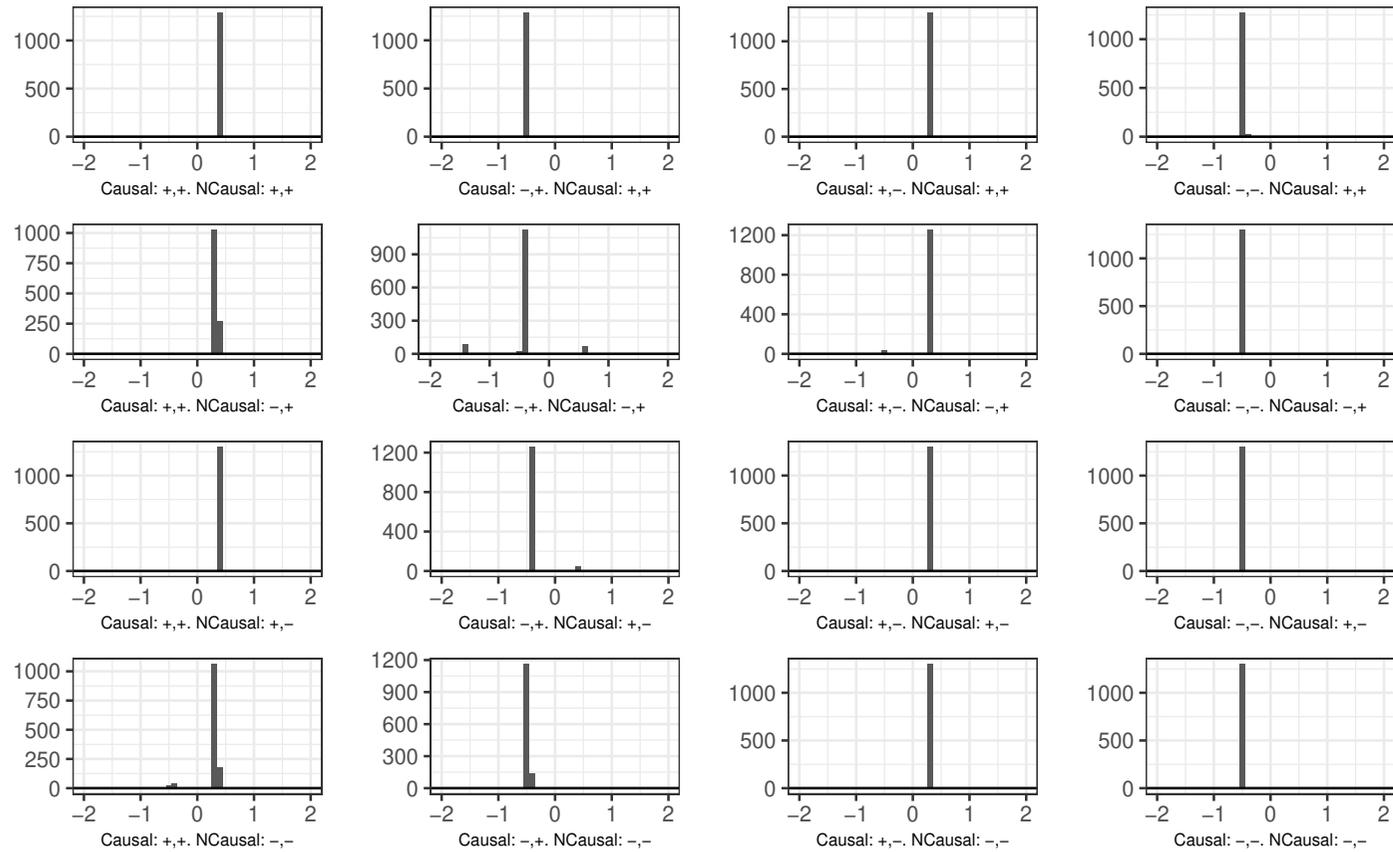


Figure 5: The parameter vector distribution is unimodal when the parameters are small. Distributions of the first causal autoregressive parameter of the MAR(2,2) process with $\psi = (0.4, 0.1)$, $\phi = (0.3, 0.2)$. Diagonal plots are for processes with matching sign pattern for causal and noncausal polynomial. The first diagonal plot is for all-positive parameters, the last one - for all negative parameters. The off-diagonal plots represent processes with non-uniform parameters.

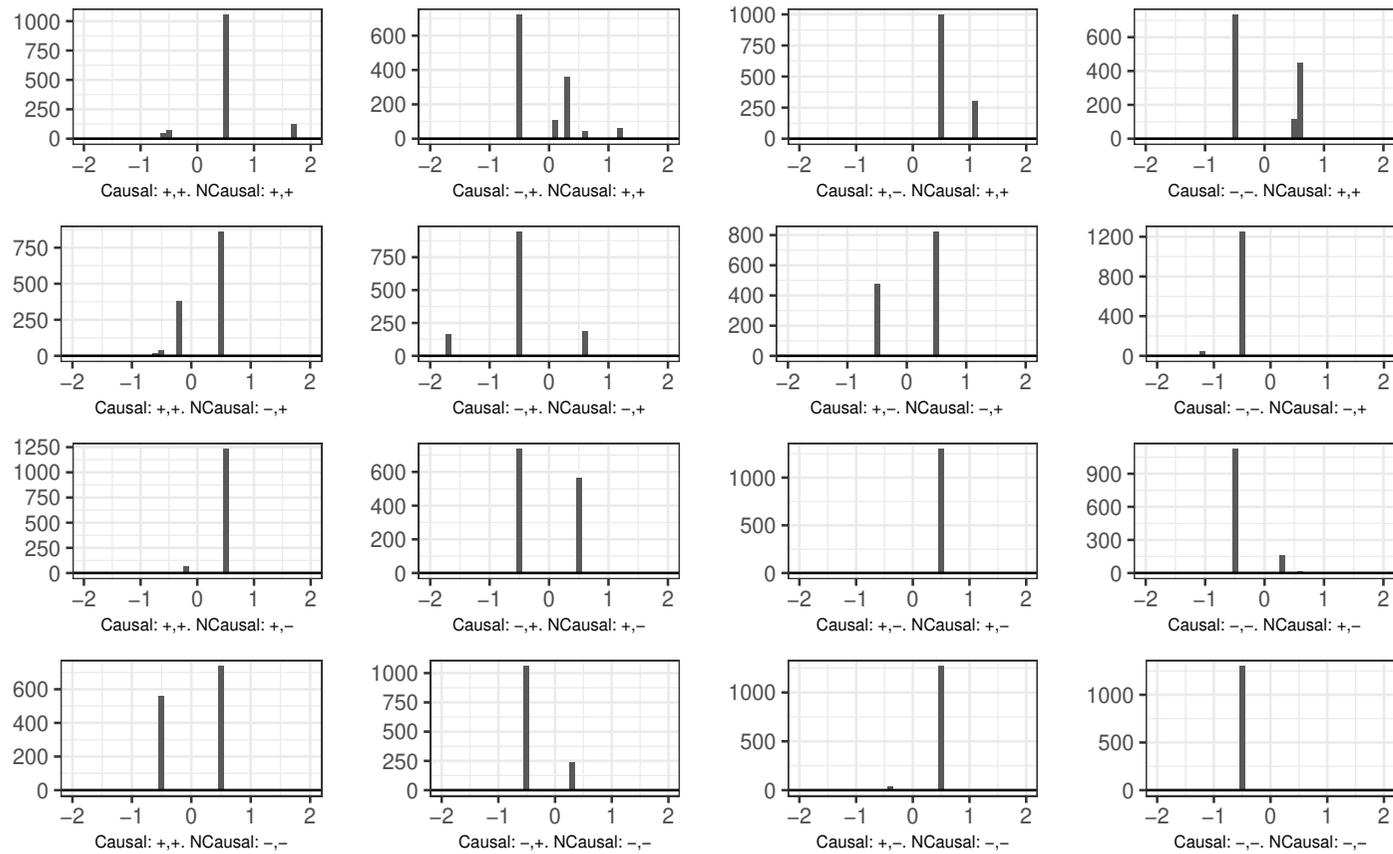


Figure 6: The antidiagonal parameter combinations are multimodal when the parameters are moderately large. Distributions of the first causal autoregressive parameter of the MAR(2,2) process with $\psi = (0.5, 0.3), \phi = (0.6, 0.2)$. Diagonal plots are for processes with matching sign pattern for causal and noncausal polynomial. The off-diagonal plots represent processes with non-uniform parameters.

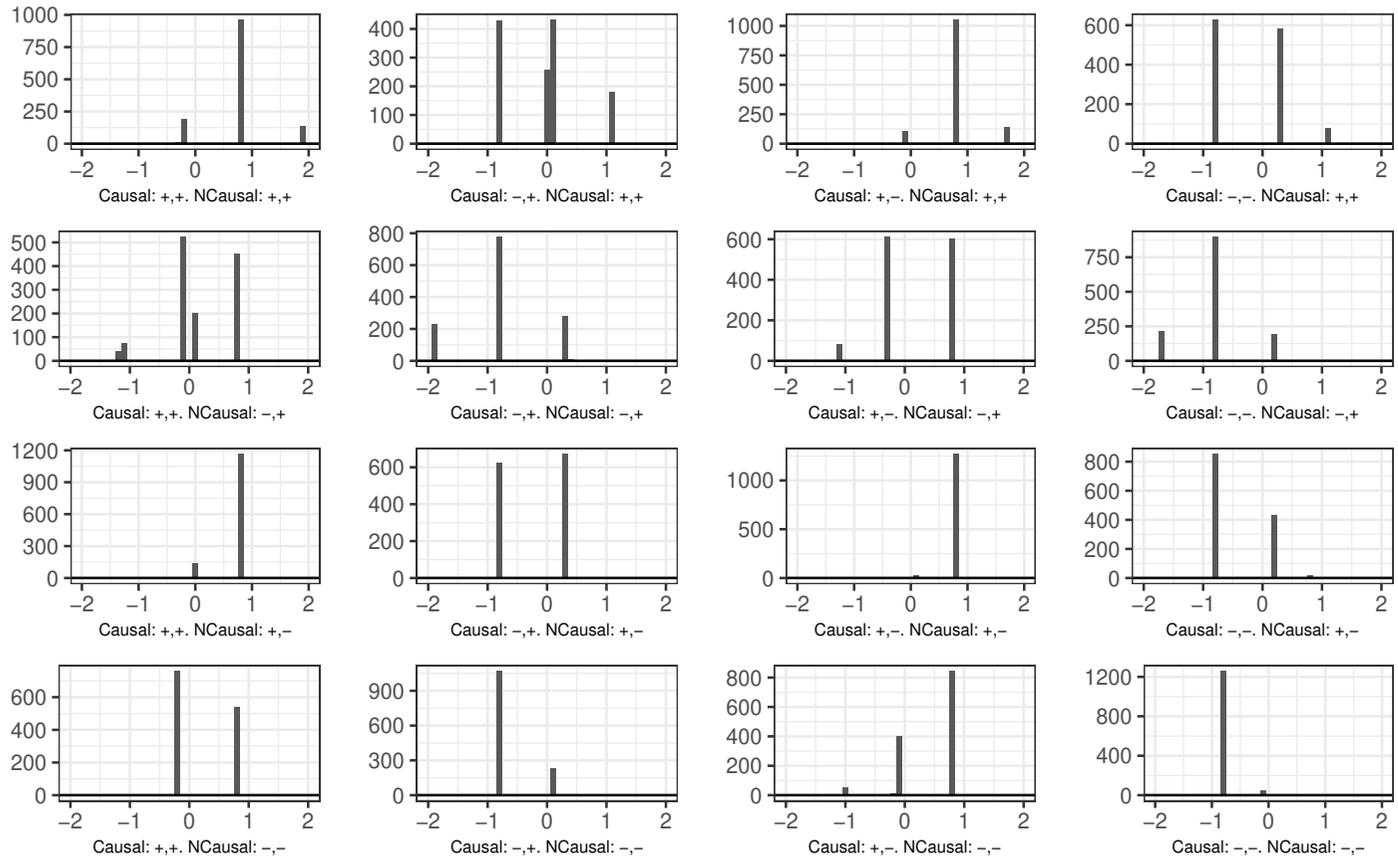


Figure 7: Almost all parameter vector combinations are multimodal when the parameters are large. Distributions of the first causal autoregressive parameter of the MAR(2,2) process with $\psi = (0.8, 0.1), \phi = (0.7, 0.2)$. Diagonal plots are for processes with matching sign pattern for causal and noncausal polynomial. The off-diagonal plots represent processes with non-uniform parameters.

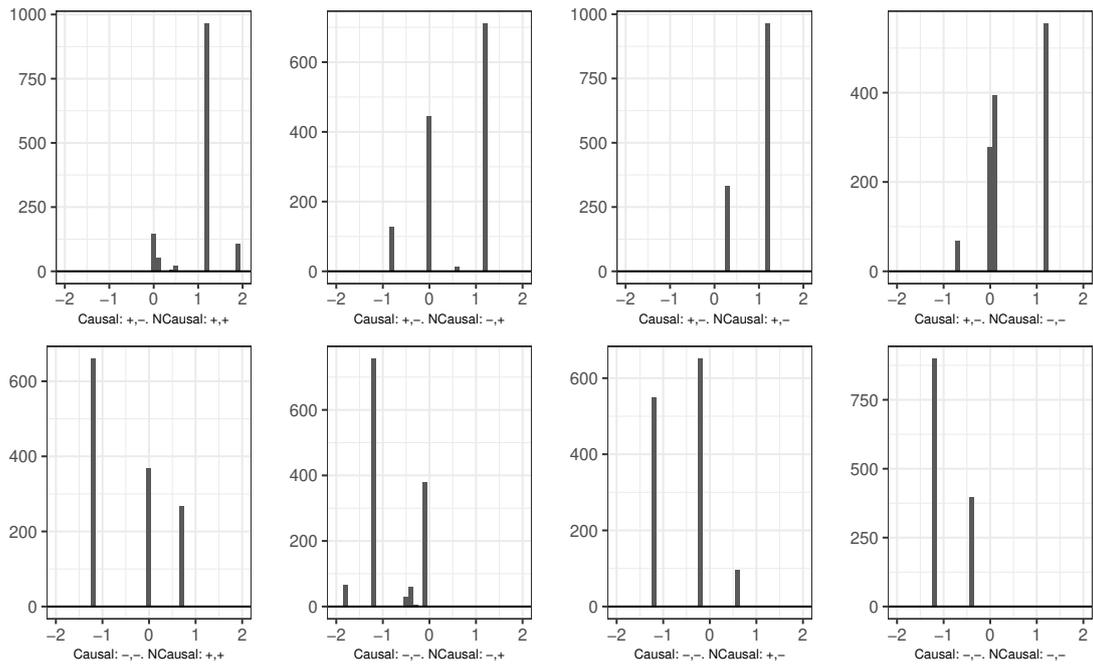


Figure 8: Distributions of the first causal autoregressive parameter of the MAR(2,2) process with $\psi = (1.2, 0.3)$, $\phi = (0.5, 0.3)$. Horizontal blocks represent different sign patterns of the noncausal component. Vertical blocks are for causal patterns.

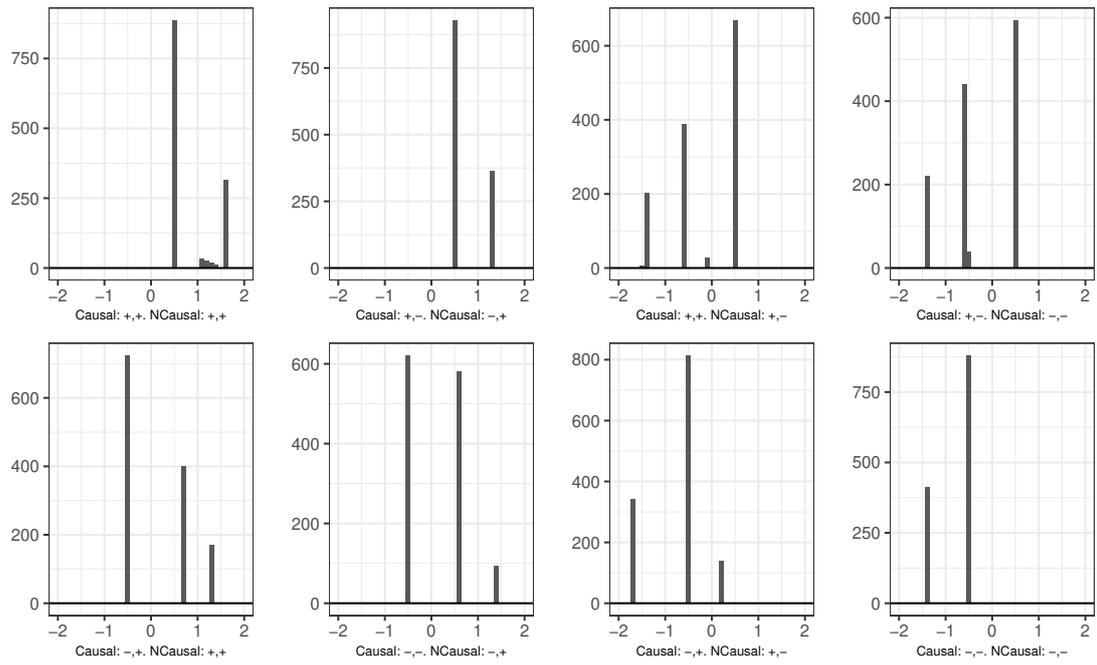


Figure 9: Distributions of the first causal autoregressive parameter of the MAR(2,2) process with $\psi = (0.5, 0.3)$, $\phi = (1.2, 0.3)$.

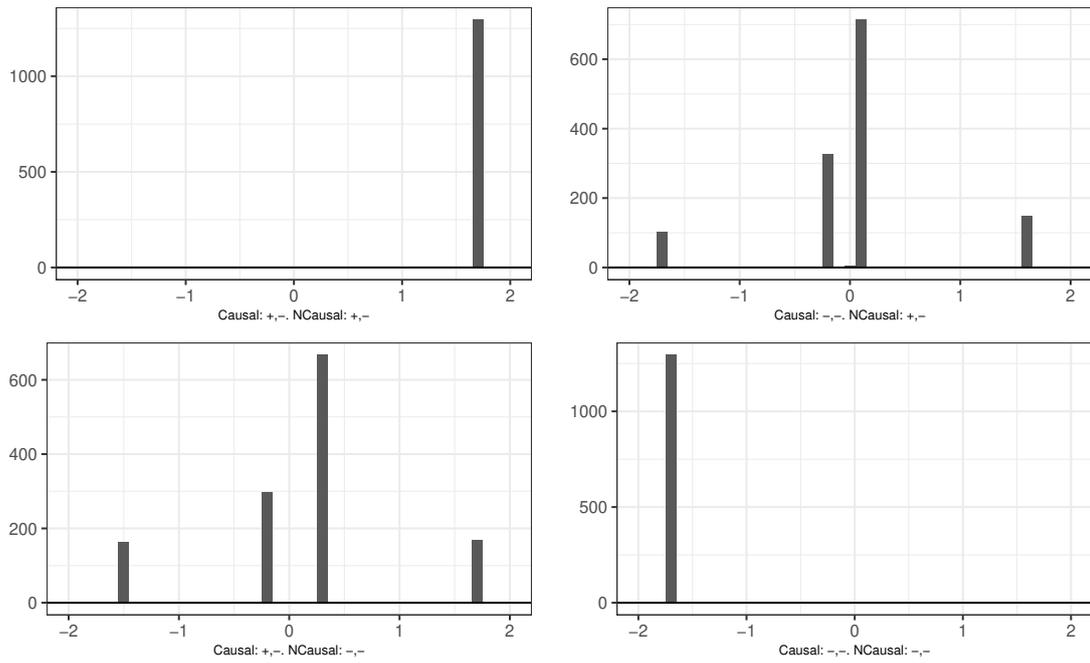


Figure 10: Frequency of a correct choice is very low for antidiagonal cases. Distributions of the first causal autoregressive parameter of the MAR(2,2) process with $\psi = (1.7, 0.8), \phi = (1.5, 0.6)$.

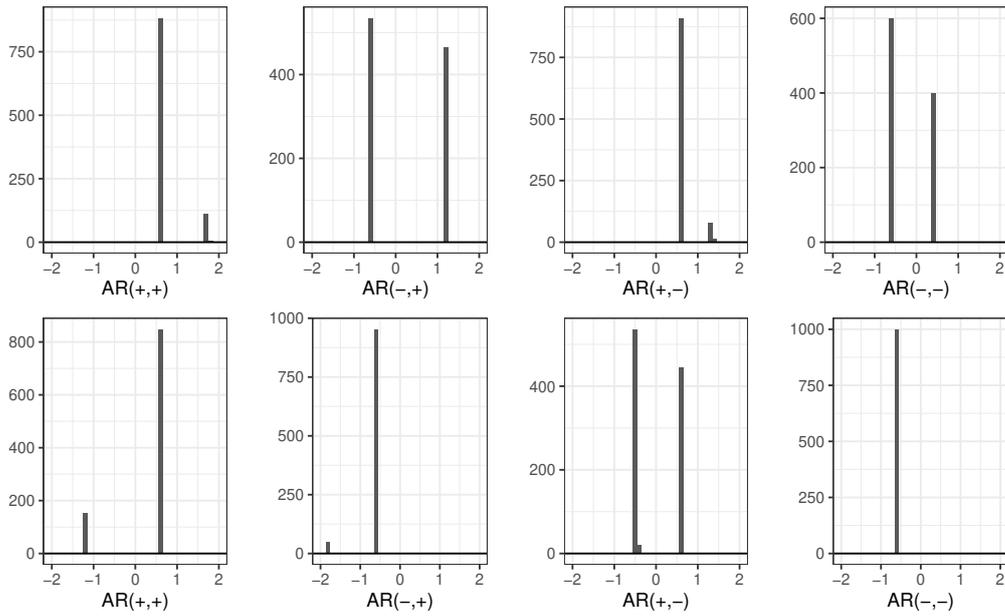


Figure 11: Distributions of the first causal autoregressive parameter of the MAR(2,1) process with $\psi = (0.6, 0.3)$, $\phi = 0.8$. The upper left plot is for all-positive autoregressive roots. The bottom right plot is for all-negative roots. The off-diagonal plots represent processes non-uniform in roots.

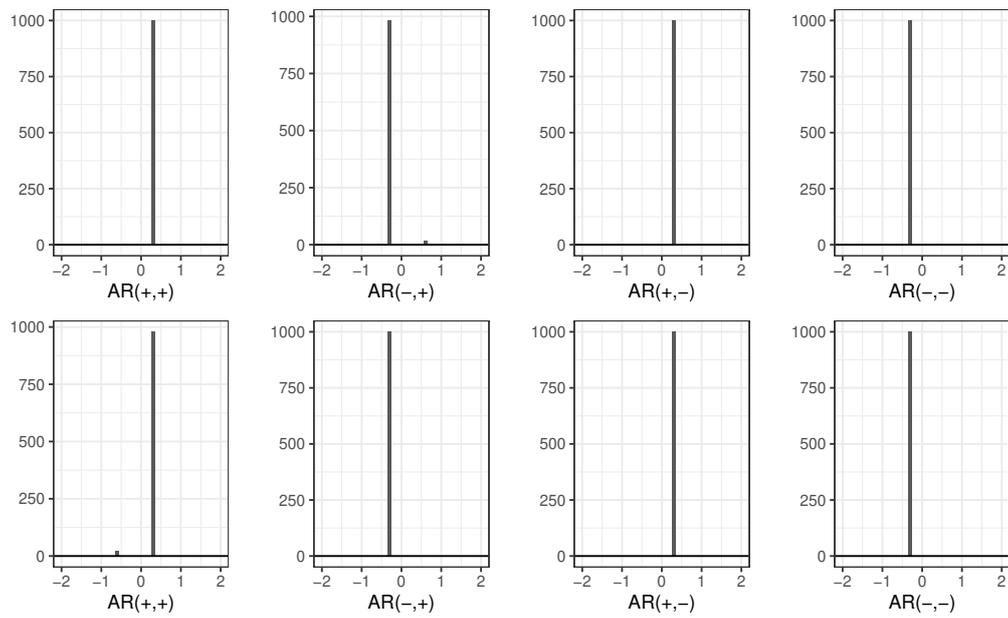


Figure 12: Distributions of the first causal autoregressive parameter of the MAR(2,1) process with $\psi = (0.3, 0.1)$, $\phi = 0.4$. The upper left plot is for all-positive autoregressive roots. The bottom right plot is for all-negative roots.

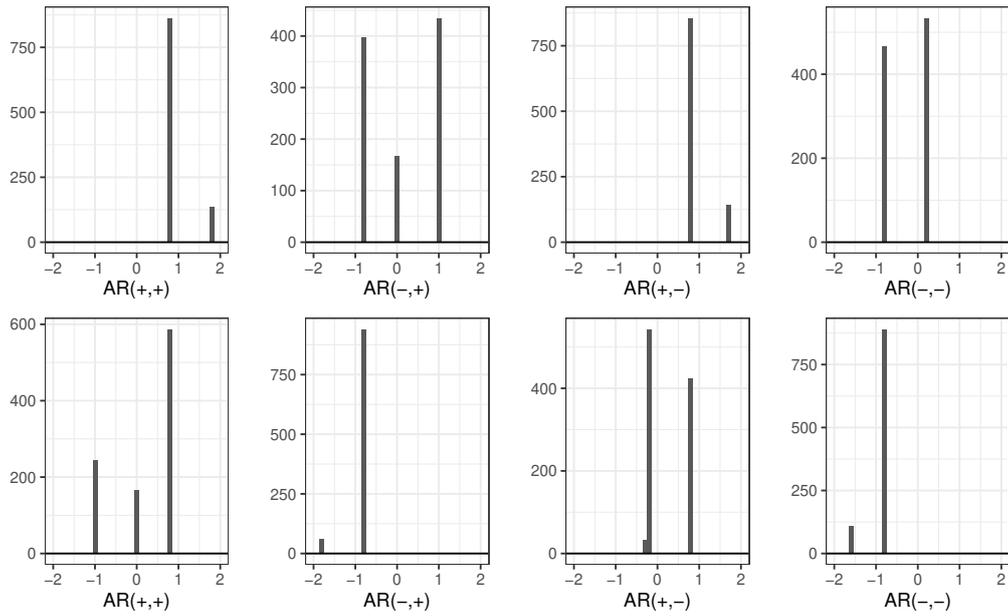


Figure 13: Distributions of the first causal autoregressive parameter of the MAR(2,1) process with $\psi = (0.8, 0.1)$, $\phi = 0.8$. The upper left plot is for all-positive autoregressive roots. The bottom right plot is for all-negative roots. The off-diagonal plots represent processes non-uniform in roots.

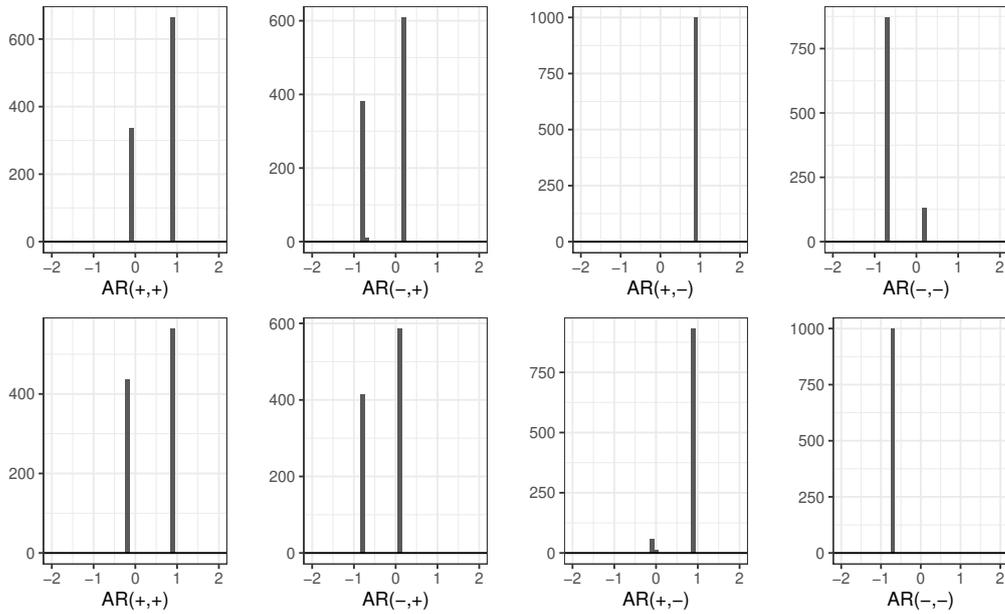


Figure 14: Distributions of the first causal autoregressive parameter of the MAR(2,1) process with $\psi = (0.8, 0.1)$, $\phi = 0.1$. The upper left plot is for all-positive autoregressive roots. The bottom right plot is for all-negative roots. The off-diagonal plots represent processes non-uniform in roots.

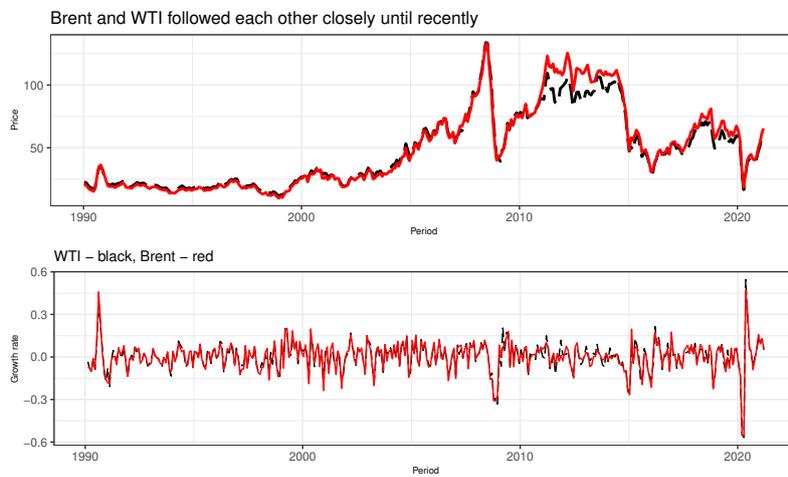


Figure 15: Upper panel: Brent (red, solid) and WTI (black, dashed) oil spot prices 1M1990-10M2020. Lower panel: Approximated Brent (red, solid) and WTI (black, dashed) oil spot price growth rates.

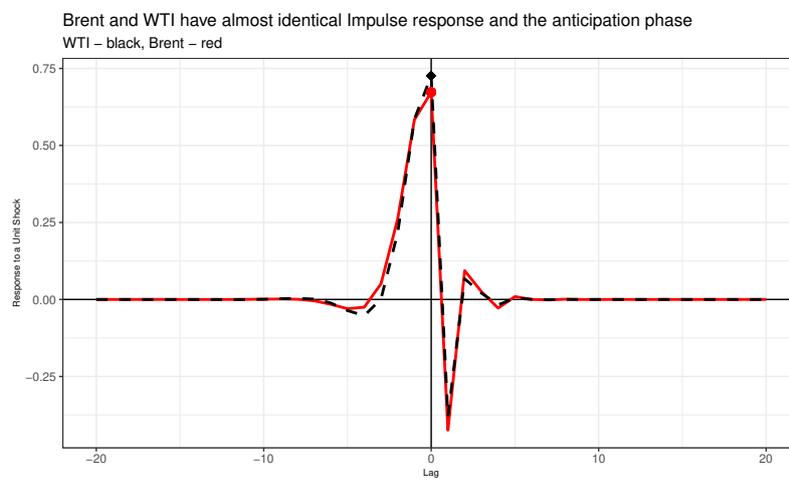


Figure 16: Brent (red, solid) and WTI (black, dashed) oil spot price impulse response curve according to the best MAR(2,2) model.