# Bayesian Mixed Autoregressive Models

**Yasmin Ferreira Cavaliere**


Universidade Federal do Rio de Janeiro

Instituto de Matemática

Departamento de Métodos Estatísticos

2019

# Bayesian Mixed Autoregressive Models

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

**Yasmin Ferreira Cavaliere**

Orientador: Hélio S. Migon
Coorientador: Mariane Branco Alves

Rio de Janeiro, RJ - Brasil

2019

CIP - Catalogação na Publicação

F377b

Ferreira Cavaliere, Yasmin
Bayesian Mixed Autoregressive Models / Yasmin Ferreira Cavaliere. -- Rio de Janeiro, 2019.
61 f.

Orientador: Helio dos Santos Migon.
Coorientador: Mariane Branco Alves.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto de Matemática, Programa de Pós-Graduação em Estatística, 2019.

1. Noncausal Models. 2. Bubbles. 3. Non-Gaussian distribution. 4. Non linear prediction. I. dos Santos Migon, Helio , orient. II. Branco Alves, Mariane, coorient. III. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

# Bayesian Mixed Autoregressive Models

**Yasmin Ferreira Cavaliere**

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Mestre em Estatística.

Aprovada por:

---

**Prof. Helio S. Migon**
PhD - IM - UFRJ - Orientador.

---

**Prof. Dani Gamerman**
PhD - IM - UFRJ

---

**Prof. Eduardo Fonseca Mendes**
PhD - EMAp - FGV

Rio de Janeiro, RJ - Brasil

2019

# Acknowledgements

A very special word of thanks and appreciation I would like to address to my family, in particular my parents Lucia and Jose and my sisters Bianca and Priscila. You have always supported me in every possible way and I know that I can always count on you. I hope you realize how important your role has been in me achieving this degree. I am truly blessed to have you close in my life.

I am eternally grateful to my supervisors Prof. Hélio Migon and Prof. Mariane Alves. It has been a pleasure working together with you over the last year. I have not only learnt a lot from both of you, but additionally, our meetings have always been extremely enjoyable.

I would like to thank three friends, in particular. Thaylla, Victor and Daniela: you have been with me on this journey from the very beginning, always supporting me and present in all the special moments of my life like this. We spent fun time together which probably kept me from working on this dissertation, which is also extremely important.

A warm word of thanks to my colleagues from the Statistic Department of UFRJ. Afonso, Fábio and Igor, the department is blessed with the three of you, as you are always happy to help and besides always in for a nice talk.

I thank capes and faperj for the financial support during the first and second year respectively.

Last, but definitely not least, I am also grateful to the members of my committee Professors Dani Gamerman and Eduardo Mendes, for the acceptance of the invitation and for sharing their knowledge with me. I had the pleasure of taking classes with Dani and working with Eduardo.

# Abstract

In the recent literature, there is a growing number of applications involving economic time-series models with causal and noncausal components. These models are able to replicate features that previously could only be obtained by highly nonlinear and complex models, when non-gaussian error distribution is assumed. However, mixed processes depend on unobserved components that lead to inferential procedures based on approximate likelihood. Alternatively, we assume prior distributions for these components and formally estimate these quantities as any other unobserved quantity, using the Bayesian paradigm. Thus, the main objective of this work is to use Bayesian mechanisms that provide a full joint density, in a practical way. We compare the performance of our proposal with the Approximated Maximum Likelihood Estimation (AML) and the Econometrics Bayesian formulation, assuming a t-distribution error. An empirical study on simulated and real data illustrates the potential usefulness of the results.

Key words: Noncausal models, bubbles, non-Gaussian distribution, nonlinear prediction.

# List of Tables

# List of Figures

# 1 Introduction

In the analyses of economic and financial time series such as assets, real estate, commodity and others, it is common to observe the presence of local explosive trends, depicted as bubble effect. However, classical linear autoregressive models that restrict the temporal dependence to the past have often been found insufficient to capture the behaviour of series that increase and then suddenly burst. This feature was the initial motivation to work with models that can generate certain dynamic patterns that cannot be created using conventional causal autoregressions.

The mixed model or noncausal model allows for dependence on future as well as past values of the variable in question. Hence, this model is a mixture of a purely causal AR model where the structure forces the variable to depend only on its past, with a purely noncausal AR model that allows dependence only on the future. The presence of a lag, a lead or both, affects the shape of realizations of mixed process. Purely causal models are only affected by a shock after impact, i.e., a jump is followed by an exponential decrease. On the other hand, purely noncausal models are only affected by a shock before impact, in this case, the shape is similar to a mirror of the causal process. As the mixed processes are determined by these two components, the shape of the explosive episode depends on the magnitude of the lag and lead coefficients.

Lanne e Saikkonen (2008) suggest that expanding the set of univariate AR models in the noncausal direction can indeed be worthwhile in empirical economic research. There are many advantages to using noncausal models. These models make explicit how expectations of future error terms of the model affect both the current value and expected future values of the variable of interest. Besides the agents are able to forecast a part of the future values of the economic variable in question by information unknown to the econometricians, i.e. agent's information sets being greater than that of the econometrician, who is estimating only a univariate AR model. Lastly, causal-noncausal models are able to replicate features that previously could only be obtained by highly

nonlinear and complex models (Hecq, Lieb e Telg (2015)), in special, producing locally explosive sample paths mimicking bubbles in financial markets.

An important hypothesis on noncausal autoregressions is that a non-Gaussian error term is required to achieve identification, because causality and noncausality are not distinguishable under Gaussian errors, i.e. causal and noncausal models of the same order also produce the same value of the likelihood function (Breidt et al. (1991)).This characterization stems from the properties that the Gaussian distribution is fully determined by second-order properties and due to the symmetry of autocovariance functions and spectral densities it is impossible to distinguish between causal and noncausal models.

Breidt et al. (1991), Lii e Rosenblatt (1996), Rosenblatt (2000), Huang e Pawitan (2000) and Davis (2006) have studied the literature on noncausal AR models and autoregressive moving average, which is not voluminous, and so far very few economic applications exist. However, from the econometric perspective, noncausal AR models have been considered by Lanne e Saikkonen (2008). In this context, the use of prior distribution to unobserved components distinguishes our method from the existing methods based on approximated maximum likelihood estimation (MLE) that excludes observations from the extremes which depend on unobserved components (Breidt et al. (1991) and Lanne e Saikkonen (2008)) and Bayesian estimation proposed by Lanne, Luoma e Luoto (2012).

The main objective of this work is to provide Bayesian analyses of noncausal AR models with an easier and faster algorithm than the method proposed by Lanne, Luoma e Luoto (2012). Indeed, a useful implication of our formulation is that full likelihood obtainment is facilitated and it becomes, for example, straightforward to perform statistical inference on autoregressive parameters. This objective is important to make these models accessible to practitioners who wish to work with them.

## 1.1 Motivation: The bubble effect

An economic explanation of a bubble or asset bubble (sometimes also referred to as a speculative bubble, a market bubble, a price bubble, a financial bubble, a speculative mania, or a balloon) is as follows: a bubble is trade in an asset at a price or price range that strongly exceeds the asset's intrinsic value. However, there exist several definitions of a bubble in economic literature. Gouriéroux, Zakoïan et al. (2013), for example, propose a different approach and assume that the bubbles are rather short-lived explosive patterns

caused by extreme valued shocks in a noncausal, stationary process. Formally, that process is a noncausal AR(1) model with Cauchy distributed errors. The approach in reverse time, based on a noncausal model allows for accommodating the asymmetric pattern of the bubble.

In general, as mentioned by Hencic e Gouriéroux (2015), a bubble has two phases: (1) a phase of fast upward (or downward) departure from the stationary path that resembles an explosive pattern and displays an exponential rate of growth, followed by (2) a phase of sudden almost vertical drop (or upspring) back to the underlying fundamental path.

As a real example of the bubble effect, Figure 1 shows Bitcoin electronic currency over the US Dollar.

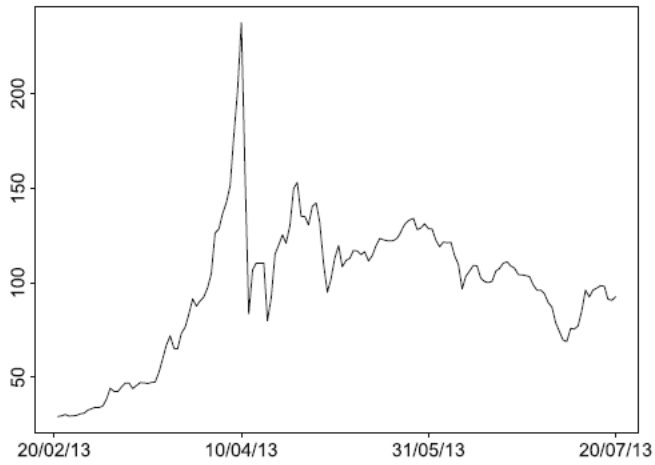



Figure 1: Bitcoin/USD exchange rate, Feb-July 2013.

As it can be seen, the dynamics of the daily Bitcoin/USD exchange rate series displays episodes of local trends, which can be modelled and interpreted as speculative bubbles.

## 1.2 Contribution

In this work, we assume prior distribution to unobserved components and update by MCMC, once the exact likelihood function depends on its past and future values.

Our main contribution is to develop an efficient computing procedure for noncausal models. Therefore, we propose hierarchical mixed models based on Student-t and Slash errors. The introduction of latent variables in the inferential problem brings great benefits as very often the complete conditional distributions have clear explicit forms.

Finally, we investigate the effect of validation sample sizes, we explore the effect of different choices of causal and noncausal polynomial values and different degrees of freedom in Student-t distribution.

## 1.3 Outline of the dissertation

This dissertation is organized as follows: in Chapter 2 the mixed causal-noncausal model is introduced and its properties are discussed. Chapter 3 provides the inferential approaches adopted, that is, the three methods mentioned earlier, highlighting two distributions of error, prediction, selection and comparison of models for our approach. Some empirical study on simulated, first considering all the the methods and then focusing only on our proposal, and a real data are reported in Chapter 4. Finally, Chapter 5 concludes with a discussion based on comparisons between methods described in Chapter 3 and points out directions for future research.

# 2 Mixed autoregressive models

Let $y_t, t = 0, \pm 1, \pm 2, \ldots$, be a stochastic process. The univariate mixed causal-noncausal autoregressive model, denoted by $MAR(r; s)$, is defined as

$$\begin{aligned} \phi(L)\varphi(L^{-1})y_t &= \epsilon_t \qquad (2.1) \\ (1 - \phi_1 L - \ldots - \phi_r L^r)(1 - \varphi_1 L^{-1} - \ldots - \varphi_s L^{-s})y_t &= \epsilon_t, \end{aligned}$$

where $\epsilon_t$ is a sequence of i.i.d. and non-Gaussian random variables with zero mean and scale parameter $\sigma$ with $E(|\epsilon_t|^\delta) < \infty$, for some $\delta \in (0, 1)$[3]. Furthermore, $L$ is the backshift operator, i.e., $L^k y_t = y_{t-k}, k = 0, \pm 1, \pm 2, \ldots$ and $\phi$, $\varphi$ are two polynomials with roots strictly outside the unit circle, that is:

$$\phi(z) \neq 0 \text{ and } \varphi(z) \neq 0, |z| \leq 1. \qquad (2.2)$$

The conditions above imply that the noncausal and mixed process admits a two-sided stationary moving average representation, $y_t = \sum_{i=-\infty}^{\infty} \alpha_i \epsilon_{t-i}$ such that $\alpha_i = 0$ for all $i < 0$ implies a purely causal process and a purely noncausal process results when $i > 0$ (Lanne e Saikkonen (2011)).

Note that the mixed model depends on both past and future values of $y_t$. In this way, if $\phi_1 = \ldots = \phi_r = 0$, we have the purely noncausal model with dependence on future values. If $\varphi_1 = \ldots = \varphi_s = 0$, the process $y_t$ is a purely causal autoregressive model with $y_t$ depending on its past but not future values.

An important feature of noncausal AR models is that causality and noncausality can only be distinguished when the error term follows a non-Gaussian distribution. The reason for this is that Gaussian processes are time-reversible. Weiss (1975) proved that timereversibility is typically a Gaussian property (the derivation of this result is collected in Appendix A). This characterization stems from the properties that strict and covariance

stationarity coincide for a Gaussian process and the fact that the Gaussian distribution is fully determined by its autocovariance structure (which is a symmetric measure).

Hence, the error term in 2.1 is assumed to be non-Gaussian. Following Lanne and Saikkonen (2009b) we assume in this paper that the disturbances are Students' t-distributed, since this distribution is found to adequately capture the fat tails of data series (see e.g. Lanne et al, 2010 or Appendix B). In addition, we consider the case where the errors are Slash distributed, as it has been shown that this distribution can mimic processes with explosive bubbles. Besides, the non-Gaussianity assumption is rather broad and can, thus, also encompass distributions with non-existing first and second order moments (in fact, all $\alpha$-stable distributions with $\alpha \in (0, 2]$ are allowed as well).

The $MAR(r, s)$ process can be decomposed into "causal"and "noncausal"components. The polynomials $\phi$ and $\varphi$ can be inverted and the process can be rewritten as:

$$y_t = \frac{1}{\phi(L)} \frac{1}{\varphi(L^{-1})} \epsilon_t. \tag{2.3}$$

Consequently, process $y_t$ admits an infinite two sided moving average representation, which is the unique strictly stationary solution of equation 2.1. We can see that the two-side moving average resulting from mixed models, includes a purely causal component and a purely noncausal. One takes, as an example, the causal AR(1) and noncausal AR(1) given, respectively, by:

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \text{with} |\phi_1| < 1 \tag{2.4}$$

and

$$y_t = \varphi_1 y_{t+1} + \epsilon_t, \text{with} |\varphi_1| < 1 \tag{2.5}$$

with $\epsilon_t$ white noise. The stationary solution of these models is a linear combination of current and past (or future) values of the white noise sequence $\epsilon_t$ given, respectively, by

$$y_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i}, \text{ and } y_t = \sum_{i=0}^{\infty} \beta_i \epsilon_{t+i}. \tag{2.6}$$

Following Gourieroux e Jasiak (2016), we define the unobserved causal and noncausal components of the process as:

$$u_t = \phi(L) y_t \leftrightarrow \varphi(L^{-1}) u_t = \epsilon_t \tag{2.7}$$

and

$$v_t = \varphi(L^{-1})y_t \leftrightarrow \phi(L)v_t = \epsilon_t. \tag{2.8}$$

The specification of these filtered values is very useful in simulating, estimating and forecasting mixed causal-noncausal processes.

In order to explain the bubble phenomenon, consider a causal and noncausal AR(1) as defined in equations 2.4 and 2.5, with (for the sake of simplicity) a deterministic error term that equals 1 at time $t = 30$ and 0 otherwise. The autoregressive parameter equals 0.7 and the starting value $y_0 = 0$ for the causal AR(1) and the terminal value $y_{T+1} = 0$ for the noncausal AR(1). Then we have in the causal case:

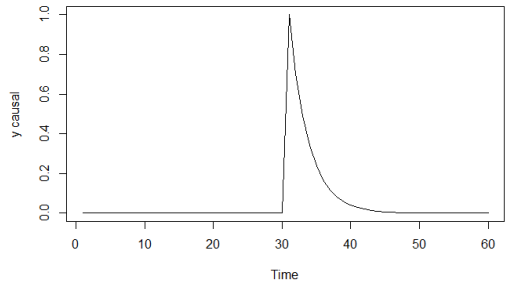
$$\begin{aligned}
y_{28} &= 0.7y_{27} + \epsilon_{28} = 0.7 \cdot 0 + 0 = 0; \\
y_{29} &= 0.7y_{28} + \epsilon_{29} = 0.7 \cdot 0 + 0 = 0; \\
y_{30} &= 0.7y_{29} + \epsilon_{30} = 0.7 \cdot 0 + 1 = 1; \\
y_{31} &= 0.7y_{30} + \epsilon_{31} = 0.7 \cdot 1 + 0 = 0.7; \\
y_{32} &= 0.7y_{31} + \epsilon_{32} = 0.7 \cdot 0.7 + 0 = 0.49, \\
&\vdots
\end{aligned}$$

that is, we see a major increase followed by a slow decay. In the noncausal case, the pattern is as follows (we reverse time order for convenience):

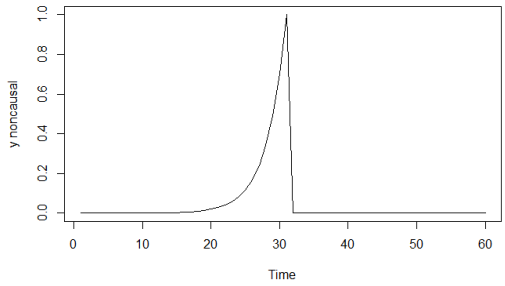
$$\begin{aligned}
y_{32} &= 0.7y_{33} + \epsilon_{32} = 0.7 \cdot 0 + 0 = 0; \\
y_{31} &= 0.7y_{32} + \epsilon_{31} = 0.7 \cdot 0 + 0 = 0; \\
y_{30} &= 0.7y_{31} + \epsilon_{30} = 0.7 \cdot 0 + 1 = 1; \\
y_{29} &= 0.7y_{30} + \epsilon_{29} = 0.7 \cdot 1 + 0 = 0.7; \\
y_{28} &= 0.7y_{29} + \epsilon_{28} = 0.7 \cdot 0.7 + 0 = 0.49, \\
&\vdots
\end{aligned}$$

which is exactly the same as for the causal model but then in reverse time: we see an increase until t = 30 and then an immediate drop to zero. Observe that the value of the autoregressive parameter determines how strong the exponential growth of the bubble is and the need for a fat-tailed error distribution stems from the fact that an extreme value has to be drawn (in our example at t = 30) in order to create the paths discussed

before. Figure 2 shows simulated paths for both causal and noncausal AR(1) with an autoregressive parameter value of 0.7.




(a) Simulated MAR(1,0),i.e., causal AR(1), process with $\phi_1 = 0.7$.




(b) Simulated $MAR(0,1)$,i.e., noncausal $AR(1)$, process with $\varphi_1 = 0.7$.

Figure 2: Simulated process from causal and noncausal components, $T = 60$.

Note that the component process jointly determine the bubble patterns. In particular, the component $u$ with parameter $\varphi$ determines the growth phase of the bubble, while the $v$ component and parameter $\phi$ determine the bubble burst, i.e., in the causal case (a), a jump is followed by an exponential decrease and consequently cannot mimic the typical bubble pattern, while (b) in the noncausal case, one observes a process of exponential growth followed by a drop. In summary, by combining the causal and noncausal dynamics in the mixed causal-noncausal model, one can create even more interesting patterns (like e.g. asymmetric cycles) that previously could only be obtained using highly nonlinear and complex models, as exemplified in Figure 3.

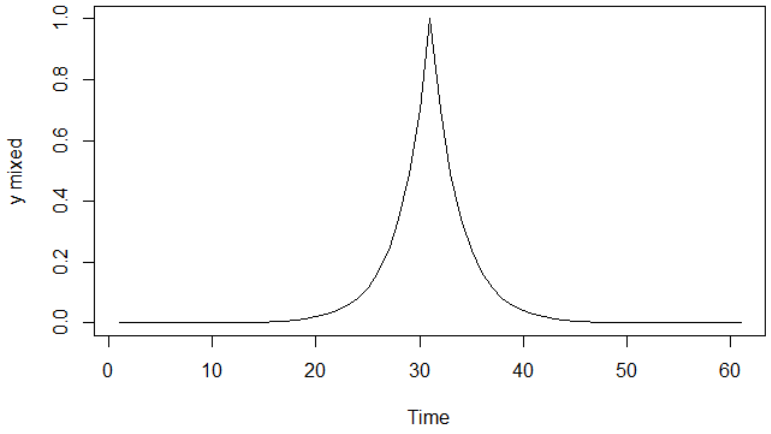



Figure 3: Simulated mixed MAR(1,1) process with $\phi_1 = \varphi_1 = 0.7$.

# 3 Inferential approaches

Throughout this chapter, the different approaches for estimation of the parameters of the mixed model, used in this work, will be presented. Initially, the maximum likelihood method is treated, which uses an approximate likelihood function. Next, Bayesian methods are developed, using the full likelihood through different mechanisms.

## 3.1 Approximated maximum likelihood estimation - AML

Maximum likelihood estimation of the parameters of a noncausal autoregression was first studied by Breidt et al. (1991) and Davis (2006) generalized it allowing the distribution of the error term to depend on the additional parameter vector $\boldsymbol{\lambda}$.

As described previously, in the mixed causal-noncausal models, $y_t$ depends on its past as well as future values. Thus, if the error process $\epsilon_t$ has a density $f_\epsilon(\epsilon_t)$, the joint density of the data $y_t$ can be approximated by

$$\prod_{t=r+1}^{T-s} f_\epsilon(\phi, \varphi, \boldsymbol{\lambda}|y_t),$$

where the vector $\boldsymbol{\lambda}$ collects distributional parameters necessary to characterize the density.

Relatively little is known about the properties of the MLE of mixed causal-noncausal models if $f_\epsilon(\epsilon_t, \boldsymbol{\lambda})$ is not assumed to be known.

Following Breidt et al. (1991), the non-Gaussianity assumption ensures the identifiability of causal and noncausal parts. In particular, if $\epsilon_t$ is a sequence of i.i.d. t-distribution zero mean random variables with $\nu(\nu > 0)$ degrees of freedom and scale $\sigma$ $(\sigma > 0)$, one can find the corresponding approximated log-likelihood function as proposed by Hecq, Lieb e Telg (2015) and the approximated maximum likelihood estimator. Hence,

$$
\begin{aligned}
l_{\mathbf{y}}(\boldsymbol{\theta}|y_1,\ldots,y_t) &= (T-(r+s))[ln(\Gamma((\nu+1)/2)) - ln(\sqrt{\nu\pi}) \\
&- ln(\Gamma(\nu/2)) - ln(\sigma)] - (\nu+1)/2 \sum_{t=r+1}^{T-s} ln(1+((\phi(L)\varphi(L^{-1})y_t)/\sigma)^2/\nu).
\end{aligned} \tag{3.1}
$$

Thus, the MLE corresponds to the solution of the problem:

$$
\hat{\boldsymbol{\theta}}_{ML} = argmax_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l_{\mathbf{y}}(\boldsymbol{\theta}|y_1,\ldots,y_t), \tag{3.2}
$$

where $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\varphi}, \sigma, \nu)$, $\boldsymbol{\Theta}$ is a permissible parameter space containing the true value of $\boldsymbol{\theta}$.

Since an analytical solution of the score function is not directly available, we use The MARX package witch offers functions to simulate, estimate and select mixed causal-noncausal autoregressive models, possibly including exogenous regressors. The R package MARX used in this study is freely available at https://CRAN.R-project.org/package=MARX. See Hecq, Lieb e Telg (2017) for instructions on how to use the package.

For $\nu > 2$ and $E(|\epsilon_t|^2 < \infty)$, Fonseca, Ferreira e Migon (2008) have shown that the MLE is $\sqrt{T}$-consistent and assymptotically normal with

$$
\sqrt{T}(\hat{\phi}_{iML} - \phi) \sim N\left(0, \frac{\nu+3}{\nu+1}\sigma^2 E[v_t^2(\varphi)]^{-1}\right), i = 1,2,\ldots,r, \tag{3.3}
$$

$$
\sqrt{T}(\hat{\varphi}_{jML} - \varphi) \sim N\left(0, \frac{\nu+3}{\nu+1}\sigma^2 E[u_t^2(\phi)]^{-1}\right), j = 1,2,\ldots,s, \tag{3.4}
$$

$$
\sqrt{T}(\hat{\sigma^2}_{ML} - \sigma^2) \sim N\left(0, \frac{\nu+3}{\nu}\sigma^2(2n)^{-1}\right), \tag{3.5}
$$

$$
\sqrt{T}(\hat{\nu}_{ML} - \nu) \sim N\left(0, \frac{4}{n}\left[\Psi'(\frac{\nu}{2}) - \Psi'(\frac{\nu+1}{2}) - \frac{2(\nu+5)}{\nu(\nu+1)(\nu+3)}\right]^{-1}\right), \tag{3.6}
$$

where $v_t(\varphi) = \varphi(L^{-1})y_t$ and $u_t(\phi) = \phi(L)y_t$. However, if $\nu \to \infty$, the model parameters cannot be consistently estimated anymore because the log-likelihood approaches the Gaussian distribution.

As a result of the asymptotic normal approximation, we can construct confidence intervals for parameters. The limiting distribution of these parameters could also be considered in the infinite variance setting.

## 3.2 Econometrics Bayesian Formulation - EBF

Lanne e Saikkonen (2008) studied the maximum likelihood (ML) estimation of the noncausal model 2.1 and proposed a Bayesian estimation and forecasting procedure for noncausal autoregressive models (Lanne, Luoma e Luoto (2012)). They showed that there exists a linear transformation $\mathbf{z}' = \boldsymbol{B}\boldsymbol{A}\mathbf{y}'$, where $\mathbf{z} = (v_1, \ldots, v_r, \epsilon_{r+1}, \ldots, \epsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$, $\mathbf{y} = (y_1, \ldots, y_T)$ and $\boldsymbol{B}$ and $\boldsymbol{A}$ are high dimensional matrix functions of parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_r)'$ and $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_s)'$, in which, $\boldsymbol{B}$ has unity determinant Lanne e Saikkonen (2013). This resut as well as the linear transformation is demonstrated in Appendix C.

In this way, given the independence conditions of the components of $\boldsymbol{z}$ (this can be observed in equation 2.6), the exact likelihood can be expressed as:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta}) &= p(v_1, \ldots, v_r|\boldsymbol{\theta}) \times |A(\boldsymbol{\phi}, \boldsymbol{\varphi})| \prod_{t=r+1}^{T-s} f_\epsilon(\phi, \varphi, \nu|y_t) \times p(u_{T-s+1}, \ldots, u_T|\boldsymbol{\theta}) \quad (3.7) \\ &= p(\varphi(L^{-1})y_1, \ldots, \varphi(L^{-1})y_r|\boldsymbol{\theta}) \\ &\times |A(\boldsymbol{\phi}, \boldsymbol{\varphi})| \prod_{t=r+1}^{T-s} f_\epsilon(\phi, \varphi, \nu|y_t) p(\phi(L)y_{T-s+1}, \ldots, \phi(L)y_T|\boldsymbol{\theta}). \end{aligned}$$

But it is not possible to express the joint distribution of $v_1, \ldots, v_r$ and $u_{T-s+1}, \ldots, u_T$ in a closed form, because the joint distribution of independently t-distributed random variables is not a multivariate t-distribution. For this reason, the authors consider an approach which employs errors $\boldsymbol{\epsilon}^- = (\epsilon_{-M}, \ldots, \epsilon_{-1}, \epsilon_0)$ and $\boldsymbol{\epsilon}^+ = (\epsilon_{T+1}, \epsilon_{T+2}, \ldots, \epsilon_{T+M})$ as an alternative by using the moving average representation of $u_t$ and $v_t$ defined as $v_t \approx \sum_{j=0}^{M} \alpha_j \epsilon_{t-j}$ and $u_t \approx \sum_{j=0}^{M} \beta_j \epsilon_{t+j}$, where the integer $M$ is assumed to be sufficiently large to make the approximation error negligible (in our simulation study, the truncation parameter M was set at 30 as suggested by the authors of this approach).

Given the errors $\boldsymbol{\epsilon}^-$ and $\boldsymbol{\epsilon}^+$ and recalling $v_t = \varphi(L^{-1})y_t$. Rewriting $\epsilon_t = \phi(L)v_t$, the Jacobian method can be applied:

$$\begin{aligned} p(v_1, \ldots, v_r|\boldsymbol{\epsilon}^-, \boldsymbol{\theta}) &= \left|\frac{d\epsilon_t}{dv_t}\right| f_{\epsilon_t}(\epsilon_t|\boldsymbol{\epsilon}^-, \boldsymbol{\theta}) \\ &= \left|\frac{d\phi(L)v_t}{dv_t}\right| f_{\epsilon_t}(\phi(L)v_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^-), \nu) \\ &= \left|\frac{d(v_t - \phi_1 v_{t-1} - \ldots - \phi_r v_{t-r})}{dv_t}\right| f_{\epsilon_t}(\phi(L)v_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^-), \nu) \\ &= f_{\epsilon_t}(\phi(L)v_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^-), \nu). \end{aligned} \tag{3.8}$$

The same reasoning can be develop to write conditional densities of $u$, we just need to use the fact that $u_t = \phi(L)y_t \rightarrow \epsilon_t = \varphi(L^{-1})u_t$.

Naturally, the joint distribution of $(\boldsymbol{\epsilon}^-, v_1, \ldots, v_r)$ and $(u_{T-s+1}, \ldots, u_T, \boldsymbol{\epsilon}^+)$ can be obtained by using the basic rules of conditional probability, i.e.,

$$\begin{aligned} p(\boldsymbol{\epsilon}^-, v_1, \ldots, v_r|\boldsymbol{\theta}) &= p(\boldsymbol{\epsilon}^-|\boldsymbol{\theta}) \times p(v_1, \ldots, v_r|\boldsymbol{\epsilon}^-, \boldsymbol{\theta}) \\ &= \prod_{t=-M}^{0} f_\sigma(\epsilon, \nu) \prod_{t=1}^{r} f_{\epsilon_t}(\phi(L)v_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^-), \nu), \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} p(u_{T-s+1}, \ldots, u_T, \boldsymbol{\epsilon}^+|\boldsymbol{\theta}) &= p(\boldsymbol{\epsilon}^+|\boldsymbol{\theta}) \times p(u_{T-s+1}, \ldots, u_T|\boldsymbol{\epsilon}^+, \boldsymbol{\theta}) \\ &= \prod_{t=T-s+1}^{T} f_\sigma(\epsilon_t, \nu) \prod_{t=T+1}^{T+M} f_{\epsilon_t}(\varphi(L^{-1})u_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^+), \nu). \end{aligned} \tag{3.10}$$

By using the elements of $\boldsymbol{\epsilon}^-$ and $\boldsymbol{\epsilon}^+$, an approximation of the likelihood function can be obtained as:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(\boldsymbol{\epsilon}^-, \boldsymbol{y}, \boldsymbol{\epsilon}^+|\theta) d\boldsymbol{\epsilon}^- d\boldsymbol{\epsilon}^+, \tag{3.11}$$

where

$$
\begin{aligned}
p(\boldsymbol{\epsilon}^-, \boldsymbol{y}, \boldsymbol{\epsilon}^+|\boldsymbol{\theta}) &= p(\boldsymbol{\epsilon}^-, v_1, \ldots, v_r|\boldsymbol{\theta}) \\
&\times A(\boldsymbol{\phi}, \boldsymbol{\varphi})| \prod_{t=r+1}^{T-s} f_\epsilon(\phi, \varphi, \nu|y_t) \\
&\times p(u_{T-s+1}, \ldots, u_T, \boldsymbol{\epsilon}^+|\boldsymbol{\theta}) \\
&= \prod_{t=-M}^{0} f_\sigma(\epsilon_t, \nu) \prod_{t=1}^{r} f_{\epsilon_t}(\phi(L)v_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^-), \nu) \\
&\times A(\boldsymbol{\phi}, \boldsymbol{\varphi})| \prod_{t=r+1}^{T-s} f_\epsilon(\phi, \varphi, \nu|y_t) \\
&\times \prod_{t=T-s+1}^{T} f_\sigma(\epsilon_t, \nu) \prod_{t=T+1}^{T+M} f_{\epsilon_t}(\varphi(L^{-1})u_t(\boldsymbol{\theta}, \boldsymbol{\epsilon}^+), \nu).
\end{aligned} \tag{3.12}
$$

The elements of $\boldsymbol{\epsilon}^+$ and $\boldsymbol{\epsilon}^-$ are unobserved variables, whose posterior densities are obtained by simulation methods along with the unknown parameters. With this estimates, they can also calculate:

$u_{T+i} = \epsilon_{T+1} + \sum_{j=1}^{s} \varphi_j u_{T+i+j}$ and $v_{-i} = \epsilon_i + \sum_{j=1}^{r} \phi_j v_{-i-j}, i = 0, 1, \ldots, M.$

In order to build a posterior sampler, the authors admitted prior independence of the parameters and assumed that each disturbance $\epsilon_t$ follows Student's-t distribution with location parameter 0, scale $\omega^{-1/2}$, where $\omega = \sigma^{-2}$ and $\nu$ degrees of freedom. In addition, they employed the following representation for the errors to have a high acceptance rate of candidate draws:

$$\epsilon_t = \tilde{w}^{-1/2}\eta_t, t = 1, \ldots, T,$$

where $\eta_t$ are i.i.d. $N(0, \omega^{-1})$ and $\nu\tilde{\omega}_t \sim \chi^2(\nu)$.

The priors proposed were:

$$p(\phi) \propto N(\phi_0, \Phi_0^{-1})I_S(\phi),\ p(\varphi) \propto N(\varphi_0, \Psi_0^{-1})I_S(\varphi),\ p(\omega) \propto 1/\omega, p(\nu) \propto \frac{1}{\nu_0}e^{\nu_0 - 2}, \nu_0 > 2$$

where $I_S(.)$ is an indicator function in the stationary region and $\phi_0, \varphi_0, \Phi_0$ and $\Psi_0$ are known hyperparameters.

After all, the joint posterior density of $\boldsymbol{\epsilon}^-$, $\boldsymbol{\epsilon}^+$ and $\boldsymbol{\theta}$ can then be expressed as

$$p(\boldsymbol{\theta}, \boldsymbol{\epsilon}^-, \boldsymbol{\epsilon}^+|\boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{\epsilon}^-, \boldsymbol{y}, \boldsymbol{\epsilon}^+|\boldsymbol{\theta}). \tag{3.13}$$

As the integral in 3.11 cannot be computed analytically, in Appendix D it is shown that an applicable numerical solution can be obtained once the distribution of errors $\epsilon_t$ has been chosen.

## 3.3 Bayesian Approach via Data Augmentation - BDA

Our idea is straightforward. Tanner e Wong (1987) and Chib e Greenberg (1995) exemplify the use of Metropolis Hastings by an AR (2) model. Through this approach, they explore the stationary distribution of $y_1, y_2$ and describe the likelihood from 3 onwards. Moreover, we have taken into consideration the paper of Nandram e Petruccelli (1997), in which autoregressive causal models AR(p) are considered by employing and taking prior specifications for the latent variables $y_i^{(0)}$ (it is unobserved initial values of $\boldsymbol{y}$). Thus, we assume prior distributions for unobserved variables, considering, however, a mixed model and not only a purely causal AR model. That is, $\boldsymbol{y}^* = (\boldsymbol{y}^*_{(I)}, \boldsymbol{y}^*_{(F)})$, where $\boldsymbol{y}^*_{(I)} = (y^*_{1-r}, \ldots, y^*_0)$ represents the initial unobserved values and $\mathbf{y}*_{(I)} = (y^*_{T+1}, \ldots, y^*_{T+s})$ the final unobserved values. Hence, Bayesian analyses can be developed.

By employing the latent variables $\mathbf{y}*$, we are able to use all the data because we do not need to condition on the first $r$ and on the last $s$ observations. For this reason, we can obtain the full likelihood. Furthermore, as it has been pointed out above, we also consider independent errors with two different error distributions in hierarchic structure: Student-t and Slash. Since we assume the components $\mathbf{y}*$ and $\boldsymbol{\theta}$ are independent, the exact likelihood can be expressed integrating out the latent variables:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta}) &= \int \ldots \int p(\mathbf{y}, \mathbf{y}^*|\boldsymbol{\theta}) d\mathbf{y}^* \\ &= \int \ldots \int p(\mathbf{y}|\mathbf{y}^*, \boldsymbol{\theta}) p(\mathbf{y}^*, \boldsymbol{\theta})) d\mathbf{y}^* \\ &= \int \ldots \int p(\mathbf{y}|\mathbf{y}^*, \boldsymbol{\theta}) p(\mathbf{y}^*) p(\boldsymbol{\theta}) d\mathbf{y}^*. \end{aligned} \tag{3.14}$$

As a result of the Bayes Theorem, we are able to compute the posterior distributions of $\theta$ and $\mathbf{y}^*$. In doing so,

$$p(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) = \frac{p(\boldsymbol{\theta}) p(\mathbf{y}^*) p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}. \tag{3.15}$$

We can see that we need to calculate the marginal likelihood $p(\mathbf{y})$. A closed-form solution does not exist, though. Aside from that, our models might require a large number of parameters. This means that our prior distributions could potentially be moderate dimensional. This, in turn, means that our posterior distributions will also be high dimensional. So, we must turn to a numerical approximation method instead. Specifically, we will use Monte Carlo methods.

In order to make easier inferences about the parameters of the model, we assume the general hierarchical model with two levels and full conditional distributions can be obtained in closed form, except for a single parameter, for which we use a Metropolis step in Gibbs sampler. Definitely, as we will use Gibbs sampling algorithm for almost all components, it will impact the speed of convergence positively.

Let us take

$$\begin{aligned} \epsilon_t|\sigma^2,\delta_t &\overset{ind}{\sim} N(0,\sigma^2\delta_t^{-1}), t=1,\ldots,T, && (3.16)\\ \delta_t|\alpha &\overset{iid}{\sim} f(\delta_t|\alpha), && (3.17) \end{aligned}$$

where $\delta_t$ represents the unobserved latent variable introduced in the model to allow for feasible inferences regarding the parameters of the model and $\alpha$ is the component associated with the marginal distribution of the error terms, i.e., $\alpha$ represents the degrees of freedom in a Student-t model ($\alpha=\nu$) and a shape parameter in a Slash model ($\alpha=\lambda$).

The joint probability distribution of the observations conditional on the parameters can, therefore, be expressed as:

$$L(\boldsymbol{\theta},\mathbf{y}^*|\mathbf{y}) = \prod_{t=1}^{r} f_{\epsilon_t}(y_t|\boldsymbol{\theta},\mathbf{y}^*_{(\mathbf{I})}) \times \prod_{t=r+1}^{T-s} f_{\epsilon_t}(y_t|\boldsymbol{\theta}) \times \prod_{t=T-s+1}^{T} f_{\epsilon_t}(y_t|\boldsymbol{\theta},\mathbf{y}^*_{(\mathbf{F})}) \quad (3.18)$$

where $\boldsymbol{\theta}=(\sigma^2,\boldsymbol{\delta},\alpha,\boldsymbol{\phi},\boldsymbol{\varphi})$, $\boldsymbol{\delta}=(\delta_1,\delta_2,\ldots,\delta_T)$.

To explain this likelihood better, consider a $MAR(1,1)$. Then we have:

$$\begin{aligned} (1-\phi_1(L))(1-\varphi_1(L^{-1}))y_t &= \epsilon_t && (3.19)\\ y_t-\phi_1 y_{t-1}-\varphi_1 y_{t+1}+\phi_1\varphi_1 y_t &= \epsilon_t \\ (1+\phi_1\varphi_1)y_t-\phi_1 y_{t-1}-\varphi_1 y_{t+1} &= \epsilon_t, \end{aligned}$$

and the dependence of $y_t$ on future value $y_{t+1}$ and past value $y_{t-1}$ is evident. As $r = s = 1$, the first observation depends on $\mathbf{y}^*_{(\mathbf{I})} = y^*_0$ and the last depends on $\mathbf{y}^*_{(\mathbf{F})} = y^*_{T+1}$. In this case, we can construct the full likelihood as:

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) &= f_{\epsilon_t}(y_1|\boldsymbol{\theta}, y^*_0) \times \prod_{t=2}^{T-1} f_{\epsilon_t}(y_t|\boldsymbol{\theta}) \times f_{\epsilon_t}(y_T|\boldsymbol{\theta}, y^*_{T+1}) \\ &= \frac{\delta_1^{1/2}}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{\delta_1}{2\sigma^2}\left[(1+\phi_1\varphi_1)y_1 - \phi_1 y^*_0 - \varphi_1 y_2\right]^2\right\} \\ &\times \prod_{t=2}^{T-1} \frac{\delta_t^{1/2}}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{\delta_t}{2\sigma^2}\left[(1+\phi_1\varphi_1)y_t - \phi_1 y_{t-1} - \varphi_1 y_{t+1}\right]^2\right\} \\ &\times \frac{\delta_T^{1/2}}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{\delta_T}{2\sigma^2}\left[(1+\phi_1\varphi_1)y_T - \phi_1 y_{T-1} - \varphi_1 y^*_{T+1}\right]^2\right\}. \end{aligned} \tag{3.20}$$

In addition to the likelihood function, Bayesian analysis requires the specification of prior distributions of the parameters of interest. We assume prior independence of the parameters $\boldsymbol{\theta}$ and $\mathbf{y}^*$. The autoregressive parameters are modelled as:

$$\phi_i \overset{iid}{\sim} N(0, \sigma^2_\phi), i = 1, 2, \ldots, r \text{ and } \varphi_j \overset{iid}{\sim} N(0, \sigma^2_\varphi), j = 1, 2, \ldots, s.$$

We also assume the unobservable variables are normally distributed:

$$y^*_{(k)} \sim N(0, \sigma^2_{y^*}), k = 1, \ldots, r+s, . \tag{3.21}$$

Besides, we take the prior distribution for $\sigma^2$ to be inverse gamma. More specifically, we consider that:

$$\sigma^2 \overset{iid}{\sim} IG(a_{\sigma^2}, b_{\sigma^2}). \tag{3.22}$$

At last, the priors of $\alpha$ and $\boldsymbol{\delta}$ take into account the specification of the error term. Hence, this particular specification will be detailed on subsections 3.3.1 and 3.3.2.

Using the above priors, the full likelihood in equation 3.20 and Bayes Theorem as expressed in 3.15, the posterior distribution of $\boldsymbol{\theta}$ is defined as:

$$
\begin{aligned}
p(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) &\propto L(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) \times p(\boldsymbol{\theta}) p(\mathbf{y}^*) && (3.23)\\
&\propto f_{\epsilon_t}(y_1|\boldsymbol{\theta}, y_0^*) \times \prod_{t=2}^{T-1} f_{\epsilon_t}(y_t|\boldsymbol{\theta}) \times f_{\epsilon_t}(y_T|\boldsymbol{\theta}, y_{T+1}^*) && (3.24)\\
&\times \prod_{t=1}^{T} p(\delta_t) \times \prod_{i=1}^{r} p(\phi_i) \times \prod_{j=1}^{s} p(\varphi_j) \\
&\times p(\sigma^2) \times p(\alpha) \times \prod_{k=1}^{r+s} p(y_k^*).
\end{aligned}
$$

Given the posterior distribution 3.23, the conditional posterior distributions can be easily calculated:

1.
$$
\phi_i|\boldsymbol{\theta}_{-\phi_i}, \mathbf{y}^*, \mathbf{y} \sim N\left(\frac{S_1}{S_2}, \frac{\sigma^2\sigma_\phi^2}{S_2}\right), i = 1, \ldots, r \quad (3.25)
$$

2.
$$
\varphi_j|\boldsymbol{\theta}_{-\varphi_j}, \mathbf{y}^*, \mathbf{y} \sim N\left(\frac{S_3}{S_4}, \frac{\sigma^3\sigma_\phi^4}{S_4}\right), j = 1, \ldots, s \quad (3.26)
$$

3.
$$
\sigma^2|\boldsymbol{\theta}_{-\sigma^2}, \mathbf{y}^*, \mathbf{y} \sim GI\left(\frac{T}{2} + a_{\sigma^2}, \sum_{t=1}^{T} \delta_t \frac{\left[(1-\phi(L))(1-\varphi(L^{-1}))y_t\right]^2}{2} + b_{\sigma^2}\right). \quad (3.27)
$$

4.
$$
p(y_{(k)}^*|\boldsymbol{\theta}, \mathbf{y}_{-j}^*, \mathbf{y}) \propto L(\theta, \mathbf{y}^*|\mathbf{y}, \delta) \times p(y_{(k)}^*), k = 1, \ldots, r+s. \quad (3.28)
$$

In particular, considering MAR(1,1), i.e., $\boldsymbol{y}^* = (y_0, y_{T+1})$ :

$$
y_{(1)}^* = y_0^*|\theta, y \sim N\left(\frac{\delta_1\sigma_{y^*}^2\phi_1\left((1+\phi_1\varphi_1)y_1 - \varphi_1 y_2\right)}{\delta_1\sigma_{y^*}^2\phi_1^2 + \sigma^2}, \frac{\sigma^2\sigma_{y^*}^2}{\delta_1\sigma_{y^*}^2\phi_1^2 + \sigma^2}\right)
$$

$$
y_{(2)}^* = y_{T+1}^*|\theta, y \sim N\left(\frac{\delta_t\sigma_{y^*}^2\varphi_1\left((1+\phi_1\varphi_1)y_T - \phi_1 y_{T-1}\right)}{\delta_t\sigma_{y^*}^2\varphi_1^2 + \sigma^2}, \frac{\sigma^2\sigma_{y^*}^2}{\delta_t\sigma_{y^*}^2\varphi_1^2 + \sigma^2}\right)
$$

where $S_1 = \sum_{t=1}^{T} \delta_t \left(\sum_{i=1}^{s} \varphi_i y_{t+i-1} - y_{t-j}\right)\left(\sum_{i=1}^{s} \varphi_i y_{t+i} - y_t + \sum_{k\neq j}^{r} \phi_k (y_{t-k} - \sum_{i=1}^{s} \varphi_i y_{t+i-2})\right)\sigma_\phi^2$

$S_2 = \sum_{t=1}^{T} \sigma_\phi^2 \delta_t \left(\sum_{i=1}^{s} \varphi_i y_{t+i-1} - y_{t-j}\right)^2 + \sigma^2,$

$S_3 = \sum_{t=1}^{T} \delta_t \left( \sum_{i=1}^{r} \phi_i y_{t-i+1} - y_{t+j} \right) \left( \sum_{i=1}^{r} \phi_i y_{t-i} - y_t + \sum_{k \neq j}^{s} \varphi_k (y_{t+k} - \sum_{i=1}^{r} \phi_i y_{t-i+2}) \right) \sigma_\phi^2$ and
$S_4 = \sum_{t=1}^{T} \sigma_\phi^2 \delta_t \left( \sum_{i=1}^{r} \phi_i y_{t-i+1} - y_{t+j} \right)^2 + \sigma^2$.

In the same way as mentioned above, the conditional posterior distribution of $\delta_t$ and $\alpha$ will depend on the specification of error terms. Therefore, it is detailed on the next subsections. Observe that all full conditional distributions presented above have closed form, consequently, the Gibbs algorithm was adopted.

### 3.3.1 Student-t Error Terms

Let $\epsilon_t|\sigma^2, \nu$ be a Student-t with $\nu$ degrees of freedom, zero mean and scale $\sigma$. This model can be rewritten in an hierarchical setting:

$$\epsilon_t|\sigma^2, \delta_t \overset{ind}{\sim} N(0, \sigma^2 \delta_t^{-1}), t = 1, \ldots, T, \tag{3.29}$$

$$\delta_t|\nu \overset{iid}{\sim} Ga(\nu/2, \nu/2), \tag{3.30}$$

Observe that this model has two levels of hierarchy with $\nu$ appearing on the second level only. In this way, to make complete posterior analyses, we just need incorporate the priors for $\nu$ and $\delta_t$.

As suggested $\nu$ by Fonseca, Ferreira e Migon (2008), we consider the Jeffreys-rule prior to $\nu$ , to be denoted by $\pi^J(\nu)$:

$$\pi^J(\nu) \propto \left( \frac{\nu}{\nu+3} \right)^{1/2} \left\{ \psi' \left( \frac{\nu}{2} \right) - \psi' \left( \frac{\nu+1}{2} \right) - \frac{2(\nu+3)}{\nu(\nu+1)^2} \right\} \left( \frac{\nu+1}{\nu+3} \right)^{1/2},$$

where $\psi(a) = d\{log\Gamma(a)\}/da$ and $\psi'(a) = d\{log\psi(a)\}/da$.

Thus, the following conditional posterior densities result:

$$\delta_t|\boldsymbol{\theta}_{\delta_t}, \mathbf{y}^*, \mathbf{y} \sim G\left( \frac{\nu+1}{2}, \frac{nu}{2} + \frac{[(1-\phi(L))(1-\varphi(L^{-1}))y_t]^2}{2\sigma^2} \right), t = 1, \ldots, T, \tag{3.31}$$

and

$$\nu|\boldsymbol{\theta}_{-\nu}, \mathbf{y}^*, \mathbf{y} \propto \left[\prod_{t=1}^{T} \frac{(\nu/2)^{(\nu/2)}}{\Gamma(\nu/2)} \delta_t^{\nu/2-1}\right] exp\left\{-\nu\left(\sum_{t=1}^{T} \delta_t/2\right)\right\} \times \pi^J(\nu). \tag{3.32}$$

Note that only the full posterior conditional for $\nu$ has no closed form, so a Metropolis-Hastings step was adopted and the log-normal distribution centered on the current value was used as the proposal distribution.

### 3.3.2 Slash Error Terms

In the econometric literature the Student's-t distribution (and the Cauchy as particular case) has almost exclusively been used in applications, since it is a choice which allows to incorporate heavy tail behaviour, but it should be straightforward to extend the method to other distributions, such as Slash distribution, which has a fatter tail than Student-t.

Analogously, let $\epsilon_t|\theta \sim S(0, \sigma^2, \lambda)$ be a Slash model with shape parameter $\lambda, \lambda > 0$. This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\lambda \uparrow \infty$. This is clearer when the model is rewritten in an hierarchical setting (Abanto-Valle et al. (2012)), i.e.,

$$\epsilon_t|\sigma^2, \delta_t \overset{ind}{\sim} N(0, \sigma^2\delta_t^{-1}), t = 1, \ldots, T, \tag{3.33}$$

$$\delta_t|\lambda \overset{iid}{\sim} Be(\lambda, 1), \tag{3.34}$$

In this case, we have

$$\delta_t|\boldsymbol{\theta}_{\delta_t}, \mathbf{y}^*, \mathbf{y} \sim G\left(\lambda + 1/2, \frac{\left[(1-\phi(L))(1-\varphi(L^{-1}))y_t\right]^2}{2\sigma^2}\right), t = 1, \ldots, T. \tag{3.35}$$

Assuming that $\lambda \sim G(a_\lambda, b_\lambda)$,in what follows, we obtain the full posterior conditional distribution of $\lambda$ as:

$$\lambda|\boldsymbol{\theta}_{-\lambda}, \mathbf{y}^*, \mathbf{y} \propto \left[\prod_{t=1}^{T} \delta_t^{\lambda-1}\right] exp\left\{-b_\lambda\lambda\right\} \lambda^{T+a_\lambda-1}. \tag{3.36}$$

Again, steps of the Metropolis algorithm are necessary.

### 3.3.3 Model Selection

Model selection becomes a more complicated empirical issue than in conventional causal autoregressive models, once allowance for noncausality is made. In particular, in addition to the order of the autoregression, the number of leads and lags to include, must be decided upon.

Breidt et al. (1991) suggested a model selection procedure based on maximizing the likelihood function. In other words, all purely causal, noncausal and mixed models of a given order ($p$) are estimated, and the model yielding the greatest value of the likelihood function is selected. This propose, however, is computationally costly.

In this section, we consider the penalized regression methods that allow us to select, contract and estimate the coefficients of the model. This method fits the model based on some loss function, subject to a constraint on coefficient values. Specially, we use the Lasso of Tibshirani (1996), which estimates regression coefficients through L1-constrained least squares.

Tibshirani (1996) argues that the lasso technique can be seen from a Bayesian perspective by assigning a prior exponential distribution to the regression coefficients of the model. Motivated by this, Park e Casella (2008) propose the Gibbs sampler for Bayesian lasso, writing the double exponential distribution as a mixture on the scale of normal distributions.

Consider a regression problem with high dimensionality, where $\boldsymbol{y} \sim N(X\beta, \sigma^2 I_n)$, with vector of dimension $p \times 1$ of coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \beta_p)'$, matrix of regressors, $X$, of dimension $n \times p$ and response $\boldsymbol{y} = (y_1, \ldots, y_n)'$ of dimension $n \times 1$.

For convenience, thorough the addition of a penalty of the form $\eta \sum_{j=1}^{p} |\beta_j|$ to the residual sum of squares, Lasso estimates often are viewed as L1-penalized least squares estimates, that is, $\boldsymbol{\beta}$ are obtained by minimizing

$$Q_y(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\beta) + \eta \sum_{j=1}^{p} |\beta_j|, \tag{3.37}$$

for some $\eta \geq 0$.

Note that the sum in the equation in 3.37 can be interpreted as the gaussian logarithm of beta posteriori density, where the first portion of the sum represents the log likelihood and the second part, the priori distribution of the $\beta's$, that is, $p(\beta|y) \propto exp\{-Q_y(\beta)\}$. Therefore, the estimation of the penalized regression coefficients can be interpreted as the mode of the posterior distribution when the regression parameters are independent and identically distributed, with prior distribution given by

$$p(\beta_j|\eta) = \frac{\eta}{2\sqrt{\sigma^2}} e^{-\eta|\beta_j|/\sqrt{\sigma^2}}, \;\; j = 1, 2, \ldots, p, \tag{3.38}$$

which is as the double-exponential distribution.

Under the approach of mixed models, a zero centred prior distribution is specified for both causal and non-causal autoregressive parameters,

$$p(\phi_1, \ldots, \phi_r) \propto \prod_{i=1}^{r} e^{-\eta|\phi_i|/\sqrt{\sigma^2}}, \tag{3.39}$$

and

$$p(\varphi_1, \ldots, \varphi_s) \propto \prod_{j=1}^{s} e^{-\eta|\varphi_j|/\sqrt{\sigma^2}}. \tag{3.40}$$

Next, as incorporated by Schmidt e Makalic (2013), we take a gamma prior for $\eta$ with shape $a_\eta$ and scale $b_\eta$, because the resulting conjugacy allows easy extension of the Gibbs sampler. That is,

$$p(\eta) \propto \eta^{a_\eta - 1} e^{-b_\eta \eta}, \;\; \eta > 0, a_\eta > 0 \text{ and } b_\eta > 0. \tag{3.41}$$

As we will apply, in our simulation studies, this model selection only for Student-t errors, for the other parameters, we use the same prior distributions defined in subsection 3.3.1.

The joint posterior distribution under this prior specifications is defined as:

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{y}^*, \eta|y_1, y_2, \ldots, y_T) \;&\propto\; L(\boldsymbol{\theta}, \mathbf{y}^*, \mathbf{y}) \\ &\times \prod_{i=1}^{r} \frac{\eta}{2\sqrt{\sigma^2}} e^{-\eta|\phi_i|/\sqrt{\sigma^2}} \times \prod_{j=1}^{s} \frac{\eta}{2\sqrt{\sigma^2}} e^{-\eta|\varphi_j|/\sqrt{\sigma^2}} \\ &\times \eta^{a_\eta - 1} e^{-b_\eta \eta} \times p(\sigma^2, \boldsymbol{\delta}, \nu, \mathbf{y}^*). \end{aligned} \tag{3.42}$$

Observe that the posterior distribution does not have a closed form and therefore MCMC methods can be used to approximate it. The complete conditional distributions for the parameters are in Appendix E.

### 3.3.4 Forecasting

As pointed out above, prior incorporation of the past and future unobserved components into the analysis allows for computing optimal forecasts from the mixed autoregressive model (1) in a straightforward manner, even if it seems counter-intuitive to forecast a process that depends on future values.

The posterior predictive distribution of $y_{t+h}$ given $y$ is calculated by marginalizing the distribution of $p(y_{t+h})$ given $\boldsymbol{\theta}$ and $\boldsymbol{y}^* = (y^*_{1-r}, \ldots, y^*_0, y^*_{T+1}, \ldots, y^*_{T+s}, \ldots, y^*_{T+s+h})$ over the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{y}^*)|\boldsymbol{y}$ :

$$p(y_{t+h}|\boldsymbol{y}) = \int \ldots \int p(y_{t+h}|\boldsymbol{\theta}, \boldsymbol{y}^*) p(y_1, \ldots, y_t|\boldsymbol{y}^*, \boldsymbol{\theta}) p(\boldsymbol{y}^*) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{y}^* \tag{3.43}$$

This integral cannot be computed analytically. However, it is possible to obtain a sample of future observations from the posterior predictive distribution via the MCMC algorithm with the distribution $\epsilon^{(i)}_{T+h} \sim t_{\nu^{(i)}}(0, \sigma^{2^{(i)}}), i = 1, 2, \ldots, N$, where $N$ represents the number of sample from the posterior density of $\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}$.

In particular, for mixed models with lag and lead one, we can conclude from equation 3.19 that:

$$y_t = \frac{1}{1 + \phi_1 \varphi_1} \left( \phi_1 y_{t-1} + \varphi_1 y_{t+1} + \epsilon_t \right), t = 1, 2, \ldots, T, \tag{3.44}$$

Thus,

$$y_{T+h} = \frac{1}{1 + \phi_1 \varphi_1} \left( \phi_1 y_{T+h-1} + \varphi_1 y_{T+h+1} + \epsilon_{T+h} \right). \tag{3.45}$$

Considering $r = s = 1$ and $h = 1$, i.e., one step-ahead, we have $\boldsymbol{y}^* = (y^*_0, y^*_{t+1}, y^*_{t+2})$. So, we can simulate values from the predictive density $p(y_{T+1}|\mathbf{y})$. To do this, we follow the formula in 3.45, and proceed in three steps:

- First, draw N sample from the posterior density $p(\boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y})$ i.e., $\theta^1, \theta^2, \ldots, \theta^N$ and $\mathbf{y}^{*1}, \mathbf{y}^{*2}, \ldots, \mathbf{y}^{*N}$.
- Generate $\epsilon^{(i)}_{T+1} \sim t_{\nu^{(i)}}(0, \sigma^{2^{(i)}}), i = 1, 2, \ldots, N$.

- Calculate $y_{T+1}$ as expressed in equation 3.45 by using values of $\boldsymbol{\theta}, \mathbf{y}^*$ and $\boldsymbol{\epsilon}$ generated in steps 1 and 2.

### 3.3.5 Mixed model X purely causal and purely non-causal models

So far, we have assumed that the order of the non-causal AR model specification is supposed to be widely known, although it is obviously never the case. Instead, the orders $r$ and $s$ of the polynomials $\phi(L)$ and $\varphi(L^{-1})$, respectively, must in practice be determined based on the data.

The "Bayesian way"to compare models is to compute the marginal likelihood of each model $p(\boldsymbol{y}|M_{ij})$, i.e the probability of the observed data $y$ given the $M_{ij}$ model, $i = 0, \ldots, rmax$ and $j = 0, \ldots, smax$. This quantity, the marginal likelihood, is just the normalizing constant of Bayes' theorem. We can see this if we write Bayes' theorem and explicit the fact that all inferences are model-dependent.

$$p(\boldsymbol{\theta}|\boldsymbol{y}, M_{ij}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, M_{ij})p(\boldsymbol{\theta}|M_{ij})}{p(\boldsymbol{y}|M_{ij})}.$$

Usually, when doing inference, we do not need to compute this normalizing constant. However, for model comparison and model averaging the marginal likelihood is an important quantity. It is given by:

$$p(\boldsymbol{y}|M_{ij}) = \int \ldots \int p(\boldsymbol{\theta}|M_{ij})p(\boldsymbol{y}^*, \boldsymbol{y}|\boldsymbol{\theta}, M_{ij})d\boldsymbol{y}^* d\boldsymbol{\theta}.$$

Computing the marginal likelihood described above is, generally, a hard task because it is an integral of a highly variable function over a high dimensional parameter space. In general this integral needs to be solved numerically using reasonably sophisticated methods. So, we decided to use the harmonic mean identity proposed by Raftery et al. (2006). This suggests that the integrated likelihood $p(y)$ can be approximated by the sample harmonic mean of the likelihoods,

$$\hat{p}_{HM}(\boldsymbol{y}) = \left[\frac{1}{N}\sum_{t=1}^{N}\frac{1}{p(\boldsymbol{y}|\boldsymbol{\theta}^t)}\right]^{-1},$$

based on $N$ draws $\theta^1, \theta^2, \ldots, \theta^B$ from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$.

Given the marginal likelihoods $p(\boldsymbol{y}|M_{ij})$ of each of the models $M_{ij}$, model selection can be based on the posterior model probabilities. By assuming that our set of models is exhaustive, we obtain from Bayes' theorem that

$$p(M_{ij}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|M_{ij})p(M_{ij})}{\sum_{i=0}^{r}\sum_{j=0}^{s} p(\boldsymbol{y}|M_{ij})p(M_i j)}.$$

If we assume that all the models are equally likely a priori, the expression above can be written as:

$$p(M_{ij}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|M_{ij})}{\sum_{i=0}^{r}\sum_{j=0}^{s} p(\boldsymbol{y}|M_{ij})}.$$

# 4 Analysis of results

The objective of this chapter is to present and discuss the results obtained from the estimation methods presented in Chapter 3. The results will be obtained in two different situations: two simulated studies and one application to real data. Thus, subsection 4.1.1 deals with a simulated study to compare the three methods in relation to sample size and degrees of freedom. The second simulated study (subsection 4.1.2) is based entirely on our model proposal. First, it compares the effect of the mixed model in distinguishing causal and noncausal componentes. Then, we evaluete the effect of different values of the polynomial coefficientes and, finally, we compare two different error distributions. In Section 4.2, an application will be performed on real data, in which the same behaviour presented by our method in the simulated study is expected.

## 4.1 Simulations

For Bayesian analysis, the priors specification were $\phi_i \sim N(0,1), i = 1,\ldots,r$, $\varphi_j \sim N(0,1), j = 1,\ldots,s$, $\sigma^2 \sim IG(3,2)$, $y^*_{(k)} \sim N(0,1000), k = 1,\ldots,r+s$ that is, non-informative priors. Chains of length 10,000 were simulated, and the first 1,000 simulations were excluded.

From the frequentist point of view, we use the maximum likelihood estimator as a point estimator and in the Bayesian approaches, we use the posterior mean, i.e.,

$$\hat{\theta}_{iAML} = argmax_{\theta_i \in \Theta} l_{\mathbf{y}}(\theta | y_1, \ldots, y_t), \tag{4.1}$$

$$\hat{\theta}_{iEBF} = E[\theta_i|\mathbf{y}], \text{ and } \hat{\theta}_{iBDA} = E[\theta_i|\mathbf{y}] \tag{4.2}$$

Considering $R$ replications, we will take the mean as the final measure, i.e., $\hat{\theta}_t = \sum_{i=1}^{R} \frac{\hat{\theta}_i}{R}$, $t = 50, 100, 200$ *and* $500$, where $\hat{\boldsymbol{\theta}} = (\hat{\phi}, \hat{\varphi}, \hat{\sigma}, \hat{\nu}, \hat{\lambda})$.

As a comparative model criterion, we will use the mean absolute percentage error (MAPE) and the score interval (SI) that summarizes the size of the interval estimates penalizing when the interval does not contemplate the true value of the quantity to be estimated. Although these measures are used for predictions, we will adapt to the coefficients of the autoregressive mixed model.

Let $(q_{1i}; q_{2i})$ be a $(1-\epsilon)100\%$ credibility (or confidence) interval for $\theta_i$ and $\theta$ the true value of the parameter. Hence, the score interval is given by:

$$SI = \left[(q_{1i}; q_{2i}) + \frac{2}{\epsilon}(q_{1i} - \theta_i)I_{(\theta_i < q_{1i})} + \frac{2}{\epsilon}(\theta_i - q_{2i})I_{(\theta_i > q_{2i})}\right]. \tag{4.3}$$

We used $\epsilon = 0.05$ resulting in an interval with 95% level of credibility. The SI measure is calculated for each of the replicates and the final measure is the mean IS's. This criterion is oriented negatively, that is, the lower its value, the better.

Finally, the MAPE is defined by the formula:

$$MAPE = \frac{1}{R}\sum_{i=1}^{R}\frac{|\hat{\theta} - \theta|}{\theta}. \tag{4.4}$$

### 4.1.1 Comparison between inferential approaches

In this section, we aim to compare the three methods described above, assuming only Student-t distribution for error terms, under varying sample sizes and degrees of freedom. The data was generated by a mixed causal-noncausal model of lag and lead order one:

$$(1 - 0.3L)(1 - 0.7L^{-1})y_t = \epsilon_t,$$

where $\epsilon_t \sim t_\nu(0, 1), t = 1, \ldots, T$ and $\nu = 2.5$ and $4.0$. The number of replications was $R = 100$ in all experiments (under different sample sizes: $T = 50, 100, 200, 500$).

Table 1 summarizes the results of the simulation study in terms of means and score intervals (SI) of the estimated parameters over all replications. The entry AML corresponds to the approximated maximum likelihood computed as the average over the 100 replications, while EBF and BDA correspond to Bayesian approaches defined in Sections 3.2 and 3.3, respectively. The column Time indicates computational time for a single replicate.

Table 1: Point estimation and score interval of the three estimators described in this paper of simulated MAR(1,1) process.

| Methods | $\phi = 0.3$, $\varphi = 0.7$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $T$ \ $\nu$ | 2.5 | | | | 4.0 | | | | Time |
| | | $\hat{\phi}$ | $SI_{(\phi)}$ | $\hat{\varphi}$ | $SI_{(\varphi)}$ | $\hat{\phi}$ | $SI_{(\phi)}$ | $\hat{\varphi}$ | $SI_{(\varphi)}$ | |
| AML | 50 | 0.37 | 2.18 | 0.60 | 1.91 | 0.35 | 2.54 | 0.58 | 2.44 | 0.001 min |
| EBF | | 0.29 | 0.75 | 0.63 | 0.65 | 0.27 | 1.10 | 0.64 | 0.92 | 2.091 min |
| BDA | | 0.26 | 0.73 | 0.63 | 0.63 | 0.26 | 1.04 | 0.61 | 0.93 | 0.034 min |
| AML | 100 | 0.32 | 0.83 | 0.67 | 0.67 | 0.32 | 1.58 | 0.65 | 1.48 | 0.019 min |
| EBF | | 0.28 | 0.53 | 0.68 | 0.39 | 0.25 | 0.73 | 0.68 | 0.91 | 4.110 min |
| BDA | | 0.27 | 0.51 | 0.68 | 0.37 | 0.24 | 0.75 | 0.67 | 0.87 | 0.067 min |
| AML | 200 | 0.30 | 0.35 | 0.69 | 0.18 | 0.30 | 0.64 | 0.68 | 0.59 | 0.030 min |
| EBF | | 0.28 | 0.34 | 0.68 | 0.22 | 0.25 | 0.51 | 0.69 | 0.34 | 10.120 min |
| BDA | | 0.27 | 0.33 | 0.69 | 0.21 | 0.24 | 0.50 | 0.69 | 0.32 | 0.13 min |
| AML | 500 | 0.30 | 0.15 | 0.70 | 0.11 | 0.30 | 0.36 | 0.69 | 0.28 | 0.070 min |
| EBF | | 0.30 | 0.19 | 0.70 | 0.13 | 0.25 | 0.39 | 0.71 | 0.20 | 26.210 min |
| BDA | | 0.29 | 0.21 | 0.70 | 0.12 | 0.26 | 0.37 | 0.70 | 0.18 | 0.367 min |

First, we have seen in Table 1 that results are obviously closer to accuracy for larger samples and for smaller degrees of freedom for all methods. Therefore, apparently, we can observe that the heavier the tails are, the more accurate the estimation of the parameters is. In addition, the results of the Bayesian approaches are found to be very similar and they have smaller IS when compared to the AML method, and this advantage is greater when $T$ is small. In terms of computational time, our proposed model (BDA) is faster than the Econometrics Bayesian Formulation.

We also consider in this simulation study, mean absolute percentage error and it can be seen in Figures 4 and 5.

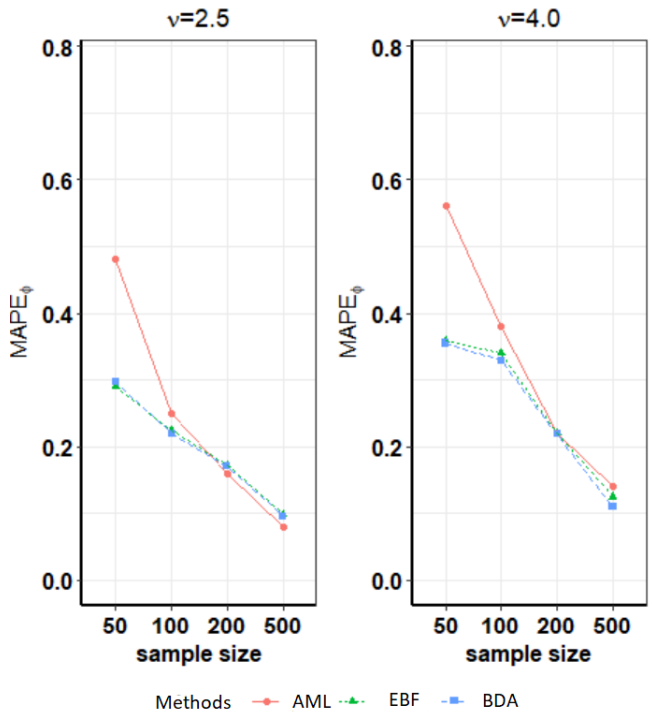



Figure 4: MAPE of $\phi$ based on $T = 50, 100, 200, 500$.

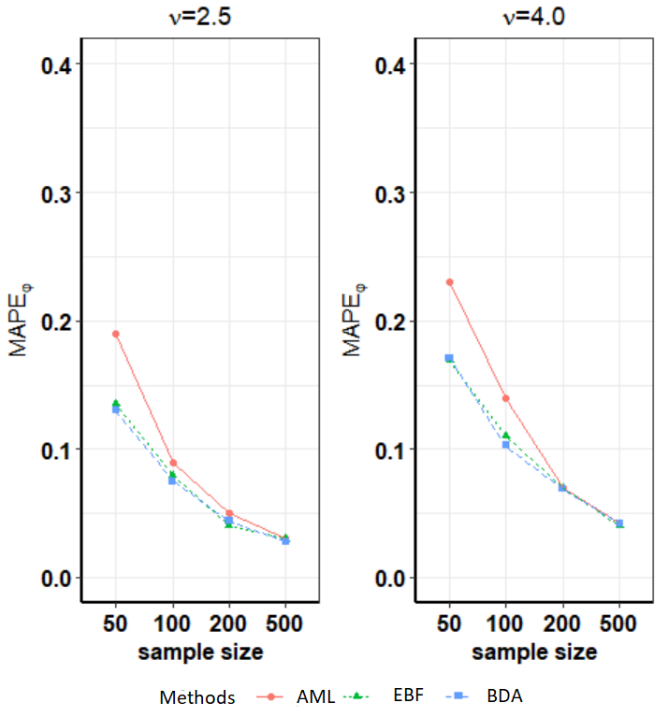



Figure 5: MAPE of $\varphi$ based on $T = 50, 100, 200, 500$.

Once again, we can note that for a sample size up to 100, the approximate method produces errors far superior to Bayesian methods.

### 4.1.2 Simulations based on our proposed approach

In the previous section, we did not observe differences in the results of Bayesian methods, except computational time. Therefore, from now on, we will develop some simulated studies in which we will only use our proposal method, since it is simpler and computationally faster. Moreover, we want to focus our attention on the results of our model.

#### 4.1.2.1 Results - Model comparison

In order to study the ability of Bayesian model selection in discriminating between causal and noncausal specifications, we conducted a simulation experiment. Throughout, the results are based on 100 simulations of a series of $T = 500$ observations. In order to keep the number of simulations reasonable, we continue to restrict our attention to a simple case with $rmax = smax = 1$. In this way,

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

where $\epsilon_t \sim t_{2.5}(0, 1), t = 1, \ldots, T$. For each simulation, we estimate four different models: a white noise model $M_{00}(r = 0, s = 0)$, a purely causal model $M_{10}(r = 1, s = 0)$, a purely noncausal model $M_{01}(r = 0, s = 1)$ and a mixed model $M_{11}(r = 1, s = 1)$. Then, we calculated $p(y|M_{11})$, $p(y|M_{10})$ , $p(y|M_{01})$ and $p(y|M_{00})$. After, as explained in subsection 3.3.5, we computed the proportion of replicates that $p(M_{11}|y) > 0.5$, by indicating that a model is selected if its posterior model probability exceeds 50%.

Table 2: Proportion of times when $p(M_{11}|y) > 0.5$

| $\phi_1$ \ $\varphi_1$ | 0.0 | 0.3 | 0.7 | 0.9 |
|---|---|---|---|---|
| 0.0 | 0.58 | 0.69 | 0.71 | 0.75 |
| 0.3 | 0.61 | 0.90 | 0.94 | 1.00 |
| 0.7 | 0.66 | 0.95 | 1.00 | 1.00 |
| 0.9 | 0.62 | 0.97 | 1.00 | 1.00 |

We can observe that the greater the parameters $\phi_1$ and $\varphi_1$, the greater the probability of the true noncausal processes is. Moreover, the procedure seems to perform fairly well in discriminating between causality and non-causality, selecting a mixed process in over 69% of the replicates whenever the true value of $\varphi_1$ is greater than or equal to 0.3 according to the 50% rule above.

We, thus, conclude that mixed models are robust enough to accommodate purely causal and purely noncausal models. This statement becomes clearer if we look at the values of the polynomials estimated in the following Table. Note that, even though, at least one of the components is zero (1st line or 1st column), the inference procedure is able to estimate its value well.

Table 3: Point estimates for each model with different pairs $\phi$ and $\varphi$ in the form $(\hat{\phi}, \hat{\varphi})$.

| $\phi_1$ \ $\varphi_1$ | 0.0 | 0.3 | 0.5 | 0.8 |
|---|---|---|---|---|
| 0.0 | (0.00,-0.01) | (0.00,0.29) | (0.00,0.50) | (-0.01,0.80) |
| 0.3 | (0.29,0.00) | (0.29,0.30) | (0.30,0.49) | (0.29,0.80) |
| 0.5 | (0.49,0.01) | (0.49,0.30) | (0.50,0.50) | (0.50,0.79) |
| 0.8 | (0.80,0.00) | (0.80,0.29) | (0.79,0.50) | (0.79,0.80) |

Besides observing the good performance of the mixed model in recovering its causal and noncausal components, we would like to compare the performance of these models, by using mean absolute percentage error (MAPE) as expressed in equation 4.4. However, note that MAPE has a significant disadvantage; it produces infinite or undefined values when the true values are zero or close to zero. Several alternative measures have been proposed to address this issue. So, we will use the relative difference ($d_r$), proposed by Bennet, Briggs e Badalamenti (2008) , in which, we can say that the absolute difference is being scaled by some function of the values $\theta$ and $\hat{\theta}$, say $f(\theta, \hat{\theta})$,

$$d_r = \frac{1}{R} \sum_{i=1}^{R} \left| \frac{\theta - \hat{\theta}_i}{f(\theta, \hat{\theta}_i)} \right|. \tag{4.5}$$

We consider $f(\theta, \hat{\theta}) = max(|\theta|, |\hat{\theta}|)$ and table 4 provides the $d_r$ for each model in the form $(d_{r_\phi}, d_{r_\varphi})$.

Table 4: Relative difference for each model with different pairs $\phi$ and $\varphi$ in the form $(d_{r_\phi}, d_{r_\varphi})$.

| $\phi_1$ \ $\varphi_1$ | 0.0 | 0.3 | 0.5 | 0.8 |
|---|---|---|---|---|
| 0.0 | (1.00;1.00) | (1.00;0.07) | (1.00;0.04) | (1.00;0.01) |
| 0.3 | (0.06;1.00) | (0.06;0.06) | (0.07;0.04) | (0.06;0.02) |
| 0.5 | (0.05;1.00) | (0.04,0.06) | (0.05;0.05) | (0.04;0.02) |
| 0.8 | (0.02;1.00) | (0.02;0.07) | (0.02;0.04) | (0.02;0;02) |

We see that in purely causal (noncausal) models presented in the first column (first line), except for the first element, which represents a white noise, the greater the value of the noncausal (causal) component, the smaller the relative error is.

Whenever $\phi_1$ is greater than $\varphi_1$ (lower triangle), the smaller the error associated with $\phi_1$. If $\phi_1$ is smaller than $\varphi_1$, the smaller the error associated with $\varphi_1$ (upper triangle). Finally, If $\phi_1$ is equal to $\varphi_1$, the errors for each component are the same (main diagonal). In summary, the greater are the parameters $\phi_1$ and $\varphi_1$, the smaller is the relative error of each component of the processes. Furthermore, values of at least 0.3 for both components generate errors of up to 7%.

#### 4.1.2.2 Results - Student-t versus Slash error terms

So far, it has been widely assumed in this chapter that the distribution of errors is t-student. On the other hand, we can use another heavy tailed distribution with hierarchical representation as it is described in Section 3. So, let us consider the error terms $\epsilon_t$ are assumed to follow a Slash distribution with scale parameter unity and shape equal to 2.5. That is:

$$\epsilon_t \sim S(0, \sigma^2, \lambda), t = 1, \ldots, T, \tag{4.6}$$

for $T = 50$ and $T = 500$, $\sigma^2 = 1$ and $\lambda = 2.5$.

We will estimate the parameters of this model (4.6) considering both the Slash error (that is, coinciding with the true distribution error) and the Student-t. The true parameters polynomial values are set to $\phi_1 = \varphi_1 = 0.3$.

Figures 6 and 7 show the histogram of the posterior means based on 100 simulations, assuming that the models were generated by Slash (1st line) and Student-t (2st line) errors considering $T = 50$ and $T = 500$, respectively. The dotted vertical blue line represents the true polynomial values, while the red lines represent the average posterior mean over all replications (solid line) and the average of 2.5% and $97,5\%$ quantiles (dashed line).

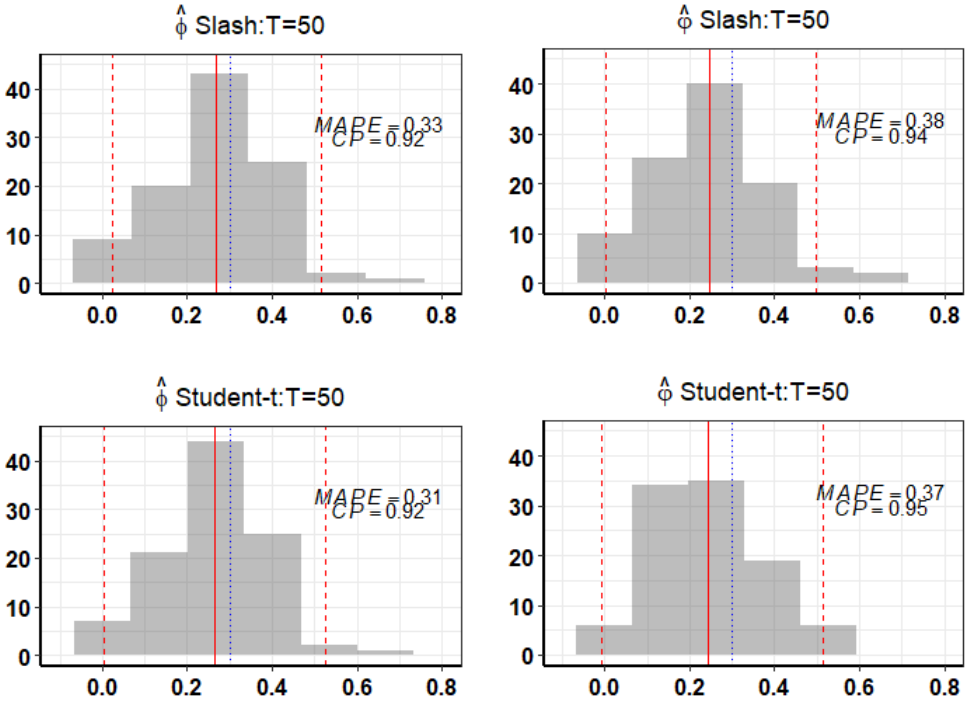



Figure 6: Histograms of estimators based on Slash and Student-t error distributions of a model generated by a Slash error distribution and 50 observations.

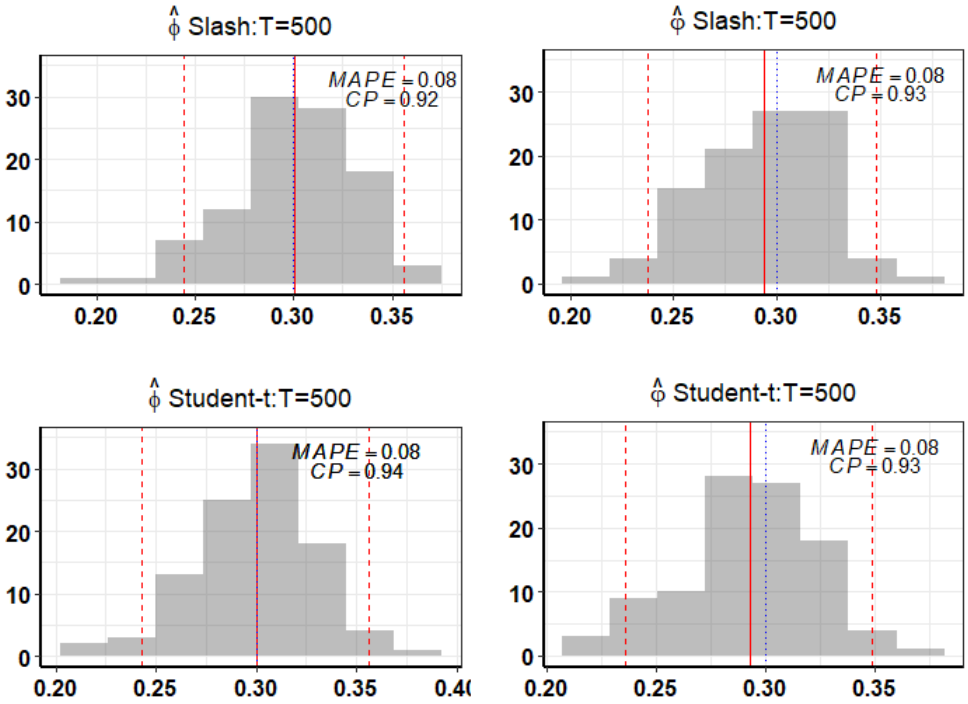



Figure 7: Histograms of estimators based on Slash and Student-t error distributions of a model generated by a Slash error distribution and 500 observations.

We can conclude that in both cases ($T = 50$ and $T = 500$), the Student-t distribution presented point estimates very close to the true value of $\phi$ and $\varphi$, even if the data were

generated by a Slash distribution. Moreover, the results of estimators based on Slash and Student-t errors are found to be very similar. This can be confirmed by looking at the MAPE and coverage probability (CP) shown in the Figures above. Finally, we can also conclude that the larger the sample size, the more precise the resulting estimates.

#### 4.1.2.3 Results - Model selection

In this section, we concentrate on Lasso formulation to model selection as described in Section 3. Therefore, a simulated study is proposed with the objective of selecting the order of the model. For this purpose, an initial order $p = r+s$ is set and the Lasso method is allowed to estimate both the number of lags and leads as well as the parameters.

We consider autoregressive time series of length $T = 500$. The following cases will be analysed separately:

- time series of the first 50 observations;
- time series of the first 100 observations;
- time series of the first 200 observations;
- time series of the 500 observations.

The data are generated by a mixed causal-noncausal models, where $\epsilon_t$ are i.i.d $t_{2.5}(0, 1)$, following the structures:

$$M1 - MAR(1,1) : (1 - 0.3L)(1 - 0.7L^{-1})y_t = \epsilon_t; \tag{4.7}$$

$$M2 - MAR(2,1) : (1 - 0.4L - 0.3L^2)(1 - 0.5L^{-1})y_t = \epsilon_t; \tag{4.8}$$

$$M3 - MAR(2,3) : (1 - 0.5L - 0.3L^2)(1 - 0.2L^{-1} - 0.3L^{-2} - 0.4L^{-3})y_t = \epsilon_t. \tag{4.9}$$

Note that the proposed models have 2, 3 and 5 significant orders, respectively. It is considered a total order of p = 6 and verified if it is possible to recover the significant orders $r$ and $s$ of the models above from the Lasso method. We applied the inference

procedure set forth in Subsection 3.3.3 over all 100 replications according to the models in (4.7), (4.8) and (4.9).

In general, it is said that a lag or lead is not significant if the interval of 100% credibility of the coefficient associated with it, contemplate the value zero. If the interval does not contemplate the zero, the lag or the lead is said to be significant. We will use intervals of 95% credibility in the analyses performed in this simulated study.

Figure 8 shows the proportion of times each order $p$ is significant for the model M1 out of 6 possible orders based on 100 simulations and considering the four series sizes.

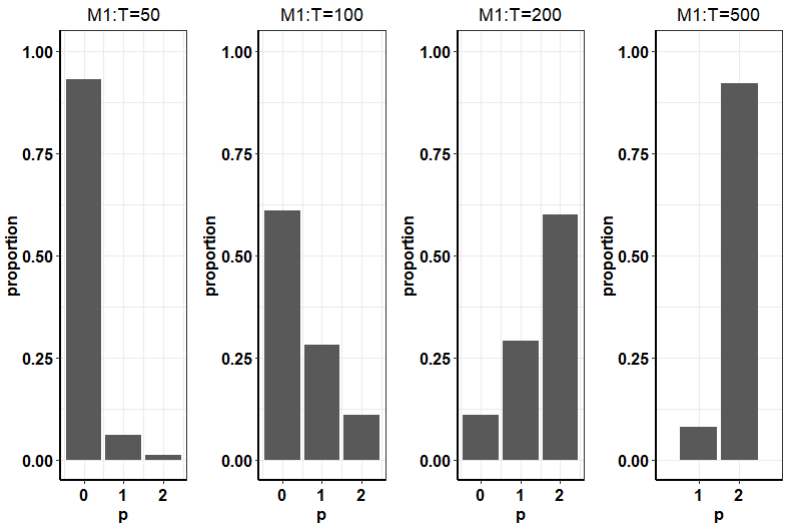



Figure 8: Barplot of the number of nonzero coefficients of model M1 over the 100 replications for $p = 6$ and different values of $T$.

We can observe that considering $T = 200$ and $T = 500$, the method retrieves two nonzero coefficients, and it is most evident in the case of $T = 500$, in which a model of order two has more than 80% chance of being selected.

The results for the M2 model are similar in the sense of choosing the correct order for a sample size of at least 200 observations and it can be observed in Figure 9. However, for model M3, and as we can found in Figure 10, only for a sample size of $T = 500$, the correct total order can be selected.

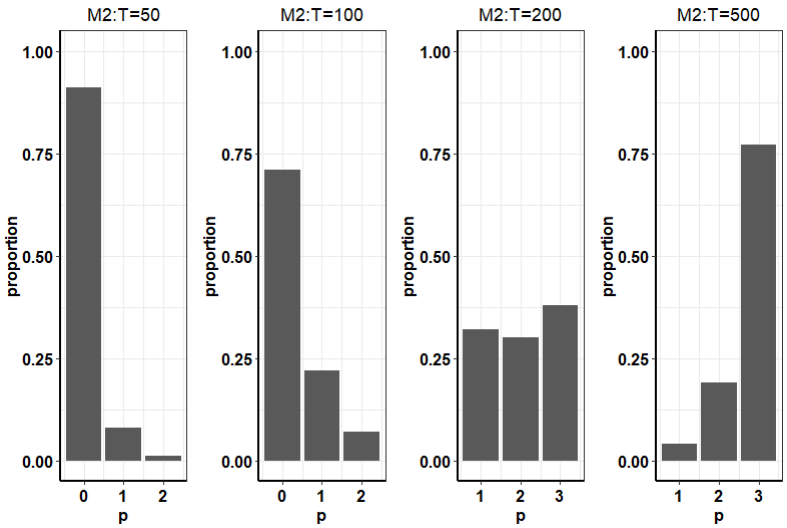



Figure 9: Barplot of the number of nonzero coefficients of model M2 over the 100 replications for $p = 6$ and different values of $T$.

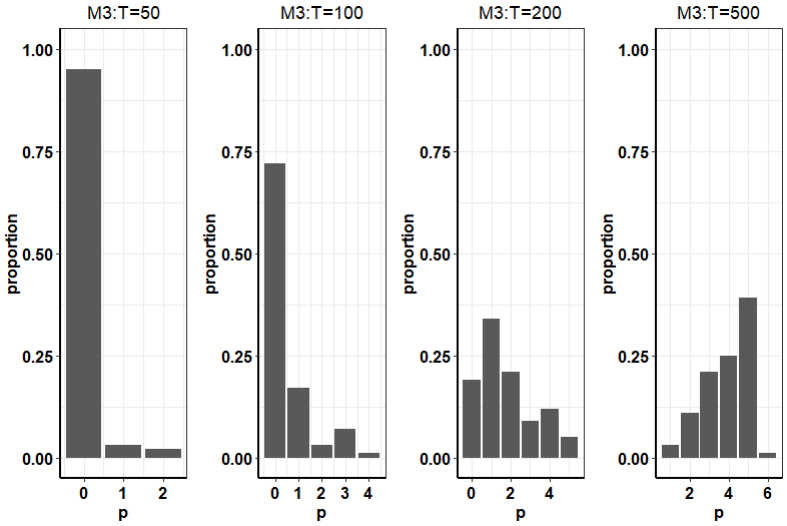



Figure 10: Barplot of the number of nonzero coefficients of model M3 over the 100 replications for $p = 6$ and different values of $T$.

These graphs show that the larger the size of the series, the more information is obtained and the greater the chance of recovering the lags and the leads of the original model. This conclusion can be complemented and retained in the tables below.

Table 5: Relative number of times, in the 100 simulations, where each one of the nonzero coefficients of each model were correctly recovered, for the different values of $T$.

(a) Model M1

| T | $\phi_1$ | $\varphi_1$ |
|---|---|---|
| 50 | 0.03 | 0.05 |
| 100 | 0.24 | 0.26 |
| 200 | 0.86 | 0.63 |
| 500 | 1 | 0.92 |

(b) Model M2

| T | $\phi_1$ | $\phi_2$ | $\varphi_1$ |
|---|---|---|---|
| 50 | 0.02 | 0.03 | 0.08 |
| 100 | 0.19 | 0.18 | 0.17 |
| 200 | 0.61 | 0.59 | 0.45 |
| 500 | 0.93 | 0.90 | 0.80 |

(c) Model M3

| T | $\phi_1$ | $\phi_2$ | $\varphi_1$ | $\varphi_2$ | $\varphi_3$ |
|---|---|---|---|---|---|
| 50 | 0.02 | 0.00 | 0.03 | 0.02 | 0.00 |
| 100 | 0.25 | 0.01 | 0.14 | 0.07 | 0.01 |
| 200 | 0.66 | 0.14 | 0.47 | 0.35 | 0.14 |
| 500 | 0.93 | 0.64 | 0.89 | 0.87 | 0.55 |

Note, also, that the higher the total order of the model, the harder it is to retrieve it correctly, and more observations are required. As it can be seen, in all cases where T = 500, it was possible to recover, with high proportion, the nonzero coefficients of the model.

Finally, we are interested in evaluating a summary of the posterior of the autoregressive coefficients. Instead of relying on only one simulation, we consider all 100 simulations. Hence, the posterior mean, as well as the 2,5% and 97% quantiles, are represented by the mean of these statistics in all replications, i.e.

$$\hat{\theta} = \frac{1}{100}\sum_{i=1}^{100}\hat{\theta}_i \text{ and } IC(\hat{\theta}) = \left(\frac{1}{100}\sum_{i=1}^{100} q_{\theta_i}(0.025); \frac{1}{100}\sum_{i=1}^{100} q_{\theta_i}(0.975)\right),$$

where $\theta$ represents the autoregressive components, $q(k)$ the k-th q-quantile and $\hat{\theta}_i$ the posterior mean.

These summaries are illustrated in Figure 11 with posterior means represented by black point, while the true value by blue point. Besides, average 2,5% and 97% quantiles are represented by the extremes of the vertical segments. Observe that the dashed red line indicates zero, and it is important to note that the coefficients whose intervals contain zero are considered to be non-significant.

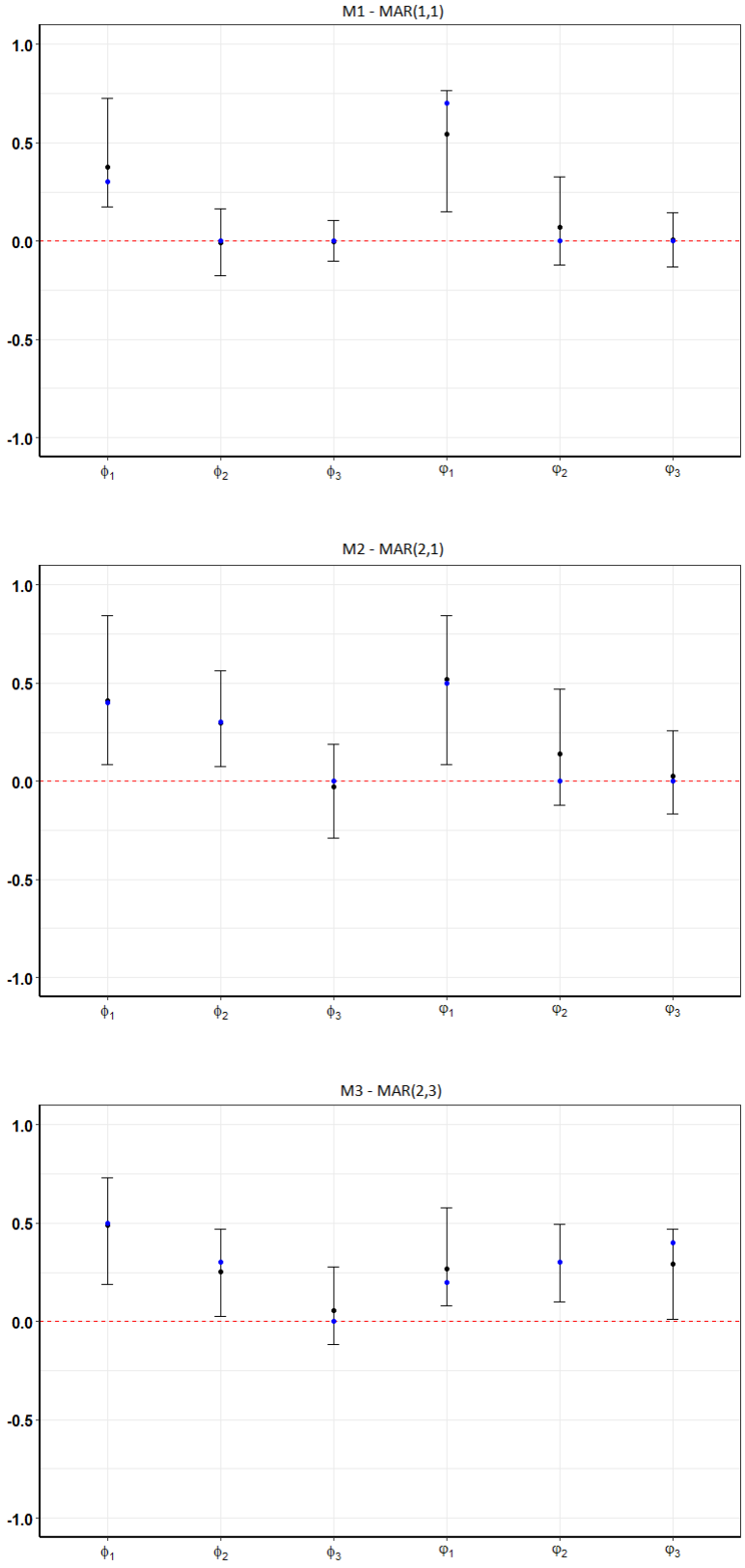



Figure 11: Summary of the posterior autoregressive coefficients for each of the models, considering $T = 500$, with the average posterior mean represented by the black point, the true value of the coefficients by the blue point, and the average of 2.5% and 97.5% quantiles by the extreme of the vertical segments.

## 4.2 Application - United States housing bubble

### 4.2.1 The Data

In our empirical study, we consider the United States housing bubble from the first part of year 1980 to first part of 2016, that includes a bubble, which peaked in early 2006, started to decline in 2006 and 2007, and reached new lows in 2012. More specifically, the sample consists of 145 observations on the quarterly prices of the prices in real terms. The dynamics of the data is displayed in Figure 12. We observe a nonlinear trend as well as the bubble that peaked at last quarter of 2006 (vertical dashed line).

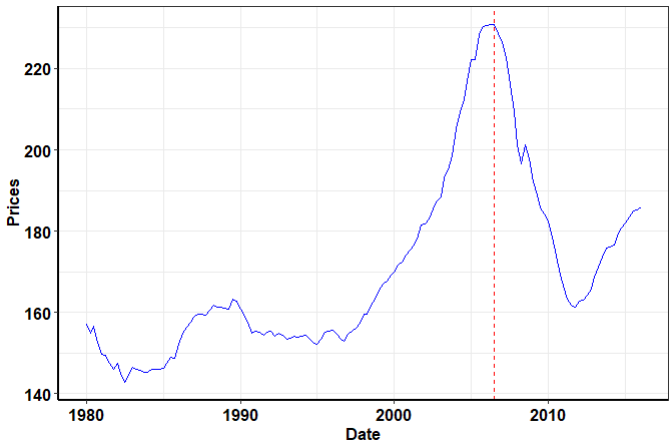



Figure 12: USA house - prices in real terms, 1980-2016.

The familiar bubble and burst patterns are obvious in the serie. To some extent, the bubble could have been foreseen for several reasons such as low mortgage interest rates, low short-term interest rates, relaxed standards for mortgage loans, and irrational exuberance on the part of loan originators coupled with mania for home ownership. The combination of these factors, coupled with failure to heed warnings, led the housing bubble to be more extreme and the resulting credit crisis to be more severe, resulting in the recession.

Consequently, such problems have created a boom and then the burst phenomenon characteristic of such a bubble.

### 4.2.2 Implementation of MAR(r,s) model

In this section, we apply our proposed model discussed above to modelling United States housing prices over the period January 1980 - May 2016 that includes a bubble, which bursted on First quarter of 2006 and compare the performance of the approximated

maximum likelihood (by using the package MARX), assuming Student-t distribution errors. We implemented the lasso method and arrived at the conclusion that only $\phi_1$ and $\varphi_1$ were significant, that is, with intervals of credibility without zero.

The parameter estimates as well as confidence interval (AML approach) and credible interval (BDA approach) are provided in Table 6. Both roots of polynomials lie outside the unit circle.

Table 6: Parameter estimates of United States housing prices, 1980-2016.

| $\theta$ | $\hat{\theta}_{AML}$ | $CI_{AML}$ | $\hat{\theta}_{BDA}$ | $CI_{BDA}$ |
|---|---|---|---|---|
| $\phi_1$ | 0.69 | (0.62,0.76) | 0.71 | (0.65,0.77) |
| $\varphi_1$ | 0.97 | (0.96,0.98) | 0.92 | (0.89,0.94) |

From the results abose, we can see that the two methods presented similar estimates. Hence, let us now consider the prediction performance of the MAR(1, 1) for this data. For this purpose we need to predict the future path of untrended process $y$ at some horizon h, (in this case, we selected h=3) that is $y_{T+1}, \ldots, y_{T+3}$, given the avaliable information $y_1, \ldots, y_{142}$. The first column of Table 7 represents the true value of the final 3 observations of the data and the respective forecasts of the approximate methods and our Bayesian approach.

Table 7: Point and interval predictive distribution.

| h | $y_{T+h}$ | $\hat{y}_{T+h_{AML}}$ | $\hat{y}_{T+h_{BDA}}$ | $CI(y_{T+h_{BDA}})$ |
|---|---|---|---|---|
| 1 | 185.0 | 179.6 | 186.9 | [183.24;190.10] |
| 2 | 185.2 | 176.8 | 190.9 | [184.64;194.64] |
| 3 | 186.0 | 174.9 | 191.3 | [182.12;198.91] |

The prediction of the approximate method was based on the MARX package and therefore it has no confidence interval, whereas by our approach, we derive the credibility interval easily (last column in the table above). In all cases, our credibility interval contains the true value of $y_{T+h}$ and point estimates are close to them as well as the maximum likelihood method.

# 5 Conclusions

The main objective of this work was to develop a Bayesian method more intuitive and computationally faster, using a full likelihood of the autoregressive mixed model and compare its estimates with the proposal by the econometrists, as well as the one approximated under a frequentist approach. The data augmentation methodology (our proposal) presented excellent estimates when compared to the approximate method, presenting greater differences mainly in the cases with less observations. In addition, it presented similarity in relation to the econometrics method in terms of the results, but we obtained less computational time.

Although the maximum likelihood method is faster, the Bayesian methods presented are preferable since they do not eliminate any observation from the analysis. Still in this context, it is worth mentioning that Bayesian methods were more accurate in the analyzes.

This study allowed to evaluate the estimates produced by the methods when doing a fully controlled simulated study, in which the true value of the polynomials and the errors assumed to the model were known, and later, through an application in real data. It was verified that the increase in the degree of freedom, when assumed t-student errors, implies in a smaller accuracy of estimation which concomitates with the question of heavy tails. In addition, the higher the value of the polynomials, the better the estimates in terms of relative errors and interval score and therefore, facilitates the identification of the mixed model when presenting greater predictive capacity. Furthermore, we have identified good estimates of the mixed model if we assume Student-t error even though the process generating the data has been Slash.

As a highlight of our proposal, we assign priori to the unobservable variables, write the distribution of errors in a hierarchical structure allowing complete conditional distributions in a closed form. Besides, we suggest the distribution of slash as specification for the errors and we develop the lasso method as criterion of order selection.

The mixed causal-noncausal model has practical relevance in econometrics because of various reasons. They are relatively simple models that are able to generate features that previously could only be obtained using highly nonlinear and complex models. This allows for accurate modelling of processes, especially those that contain asymmetric cycles and speculative bubbles. As these models are only recently introduced in the literature, there are still various settings and contexts in which their behaviour and implications can be studied.

In this work, we have only considered the univariate mixed autoregressive model, but it should be straightforward to extend the method to multivariate approach. Besides, our main goal, for future developments, is to incorporate dynamic structures for the polynomial values because it is interesting to see whether anything can be gained by allowing for time-varying parameters.

# References


ABANTO-VALLE, C. A.; MIGON, H. S.; LACHOS, V. H. et al. Stochastic volatility in mean models with heavy-tailed distributions. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 26, n. 4, p. 402–422, 2012.

BENNET, J. .; BRIGGS, W. L.; BADALAMENTI, A. *Using and Understanding mathematics: A quantitative reasoning approach*. [S.l.]: Pearson Addison Wesley Reading, MA, 2008.

BREIDT, F. J.; DAVIS, R.; LII, K.-S.; ROSENBLATT, M. Maximum likelihood estimation for noncausal autoregressive processes. *J. Multivariate Anal.*, v. 36, n. 2, p. 175–198, 1991.

CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. *The american statistician*, Taylor & Francis Group, v. 49, n. 4, p. 327–335, 1995.

DAVIS, R. A. Maximum likelihood estimation for allpass time series models. 2006.

FONSECA, T. C.; FERREIRA, M. A.; MIGON, H. S. Objective bayesian analysis for the student-t regression model. *Biometrika*, Oxford University Press, v. 95, n. 2, p. 325–333, 2008.

GOURIEROUX, C.; JASIAK, J. Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, Wiley Online Library, v. 37, n. 3, p. 405–430, 2016.

GOURIÉROUX, C.; ZAKOÏAN, J.-M. On uniqueness of moving average representations of heavy-tailed stationary processes. *Journal of Time Series Analysis*, Wiley Online Library, v. 36, n. 6, p. 876–887, 2015.

GOURIÉROUX, C.; ZAKOÏAN, J.-M. Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 79, n. 3, p. 737–756, 2017.

GOURIÉROUX, C.; ZAKOÏAN, J.-M. et al. *Explosive bubble modelling by noncausal process*. [S.l.]: CREST, 2013.

HECQ, A.; LIEB, L.; TELG, S. Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package. *Available at SSRN 3015797*, 2017.

HECQ, A. W.; LIEB, L.; TELG, J. Identification of mixed causal-noncausal models: How fat should we go? 2015.

HENCIC, A.; GOURIÉROUX, C. Noncausal autoregressive model in application to bitcoin/usd exchange rates. In: *Econometrics of Risk*. [S.l.]: Springer, 2015. p. 17–40.

HUANG, J.; PAWITAN, Y. Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 27, n. 4, p. 689–702, 2000.

LANNE, M.; LUOMA, A.; LUOTO, J. Bayesian model selection and forecasting in noncausal autoregressive models. *Journal of Applied Econometrics*, Wiley Online Library, v. 27, n. 5, p. 812–830, 2012.

LANNE, M.; SAIKKONEN, P. Modeling expectations with noncausal autoregressions. 2008.

LANNE, M.; SAIKKONEN, P. Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, De Gruyter, v. 3, n. 3, 2011.

LANNE, M.; SAIKKONEN, P. Noncausal vector autoregression. *Econometric Theory*, Cambridge University Press, v. 29, n. 3, p. 447–481, 2013.

LII, K.-S.; ROSENBLATT, M. Maximum likelihood estimation for nongaussian nonminimum phase arma sequences. *Statistica Sinica*, JSTOR, p. 1–22, 1996.

NANDRAM, B.; PETRUCCELLI, J. D. A bayesian analysis of autoregressive time series panel data. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 15, n. 3, p. 328–334, 1997.

PARK, T.; CASELLA, G. The bayesian lasso. *Journal of the American Statistical Association*, Taylor & Francis, v. 103, n. 482, p. 681–686, 2008.

RAFTERY, A. E.; NEWTON, M. A.; SATAGOPAN, J. M.; KRIVITSKY, P. N. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. bepress, 2006.

ROSENBLATT, M. *Gaussian and non-Gaussian linear time series and random fields*. [S.l.]: Springer Science & Business Media, 2000.

SCHMIDT, D. F.; MAKALIC, E. Estimation of stationary autoregressive models with the bayesian lasso. *Journal of Time Series Analysis*, Wiley Online Library, v. 34, n. 5, p. 517–531, 2013.

TANNER, M. A.; WONG, W. H. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, Taylor & Francis, v. 82, n. 398, p. 528–540, 1987.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996.

WEISS, G. Time-reversibility of linear stochastic processes. *Journal of Applied Probability*, Cambridge University Press, v. 12, n. 4, p. 831–836, 1975.

# ANEXO A – Non-Gaussianity

As introduced by Rosenblatt (2000), the following result characterizes reversibility for linear stationary process sequences under appropriate conditions.

**Theorem 1.** *Consider a linear stationary sequence*

$$y_t = \sum_{i=-\infty}^{\infty} \alpha_i \epsilon_{t-i} = \alpha(L)\epsilon_t$$

*with the $\epsilon_t's$ independent, identically distributed nonconstant random variables. Assume that $\alpha(z) \neq \pm z^r \alpha(z^{-1})$ for any integer $r$. Consider the condition (i) $\epsilon_0$ has infinite second moment and $\{y_t\}$ has its spectral density positive almost everywhere or (ii) $\alpha(z)^{-1} = b(z) = \sum b_i z^i$ with the series converging absolutely in an annulus $\{z : d < |z| < d^{-1}\}$ with $d < 1$ and $b(L)y_t = \sum b_i y_{t-i} = b(L)\alpha(L)\epsilon_t = \epsilon_t$. Under either of these conditions $\{y_t\}$ is reversible if and only if $\epsilon_0$ is Gaussian.*

They also derive the following proposition.

**Proposition 2.** *If $\{y_t\}$ is reversible, the random variable $\epsilon_t = \alpha(L)^{-1}y_t$ has the same probability distribution as $\epsilon_t = \alpha(L^{-1})^{-1}y_t$.*

Consider a purely causal autoregressive model and a purely noncausal autoregressive model with lag and lead $\phi$, respectively. That is:

- $\phi(L)y_t = \epsilon_t$
- $\phi(L^{-1})y_t = \epsilon_t$

We have already remarked that a process is said to be causal if it has the representation

$$y_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i} = \alpha(L)\epsilon_t, \tag{A.1}$$

where $\alpha(L) = \sum_{i=0}^{\infty} \epsilon_i L^i = [\phi(L)]^{-1}$ and $\epsilon_i$ is the coefficient of $L^i$ in the Laurent series expansion of $\phi(L)$. By A.1, we can observe that:

$$\epsilon_t = \alpha(L)^{-1} y_t. \tag{A.2}$$

Similary, one can write a purely noncausal model in its linear representation given by:

$$y_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t+i} = \alpha(L^{-1})\epsilon_t, \tag{A.3}$$

where $\alpha(L^{-1}) = \sum_{i=0}^{\infty} \epsilon_i L^{-i} = [\phi(L^{-1})]^{-1}$ and $\epsilon_i$ is the coefficient of $L^{-i}$ in the Laurent series expansion of $\phi(L^{-1})$. By A.3, we can note that:

$$\epsilon_t = \alpha(L^{-1})^{-1} y_t. \tag{A.4}$$

Hence, if we assume $\{y_t\}$ is a Gaussian process, applying the Theorem 1, $\{y_t\}$ is reversible. Moreover, from Proposition 2, we know that A.2 and A.4 have the same probability distribution.

# ANEXO B – Linear process and heavy tailed distributions

Motivated by Gouriéroux e Zakoïan (2017) , in this Appendix we review the main properties of strong linear processes in the presence of heavy-tailed errors. In particular, they explain how local explosion can arise when the linear process has a noncausal component and heavy-tailed errors.

A process is said to be a strong linear process if it has the representation:

$$Y_t = \sum_{i=-\infty}^{\infty} \alpha_i \epsilon_{t-i} \tag{B.1}$$

where $(\epsilon_t)$ are i.i.d. random variables and $\alpha_i$ is a sequence of real coefficients, satisfying for some $s \in (0, 1)$,

$$E|\epsilon_t|^s < \infty \text{ and } \sum_{i=-\infty}^{\infty} |\alpha_i|^s < \infty.$$

The uniqueness of the MA representation in B.1 with heavy-tailed errors was recently studied by Gouriéroux e Zakoïan (2015).

The trajectory of a strong linear process can be considered as a stochastic combination of baseline deterministic functions.

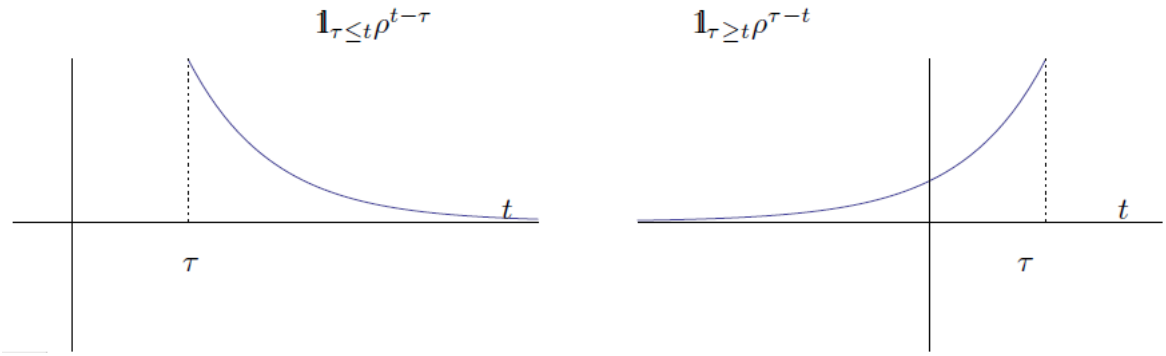
1. A strong purely causal process can be written as: $Y_t = \sum_{\tau=-\infty}^{\infty} \epsilon_\tau I_{\tau \leq t} \alpha_{t-\tau}$.
2. On the other hand, if $Y_t$ is a strong purely noncausal process, we have:

$$Y_t = \sum_{\tau=-\infty}^{\infty} \epsilon_\tau I_{\tau \geq t} \alpha_{\tau-t} \tag{B.2}$$

In the first (second) case, the path of the process $Y_t$ is a combination of baseline paths $Z_\tau(t) = I_{\tau \leq t} \alpha_{t-\tau}$ $(Z_\tau(t) = I_{\tau \geq t} \alpha_{\tau-t})$ with stochastic i.i.d. coefficients $\epsilon_\tau$.

For instance, consider a purely causal and a purely noncausal AR(1),i.e.,

- $Y_t = \phi_1 Y_{t-1} + \epsilon_t$ with $|\phi_1| < 1$ (thus $\alpha_i = \phi_1^i$ for $i \geq 0$);
- $Y_t = \varphi_1 Y_{t+1} + \epsilon_t$ with $|\varphi_1| < 1$ (thus $\alpha_{-i} = \varphi_1^i$ for $i \geq 0$).




Font: Gouriéroux e Zakoïan (2017)

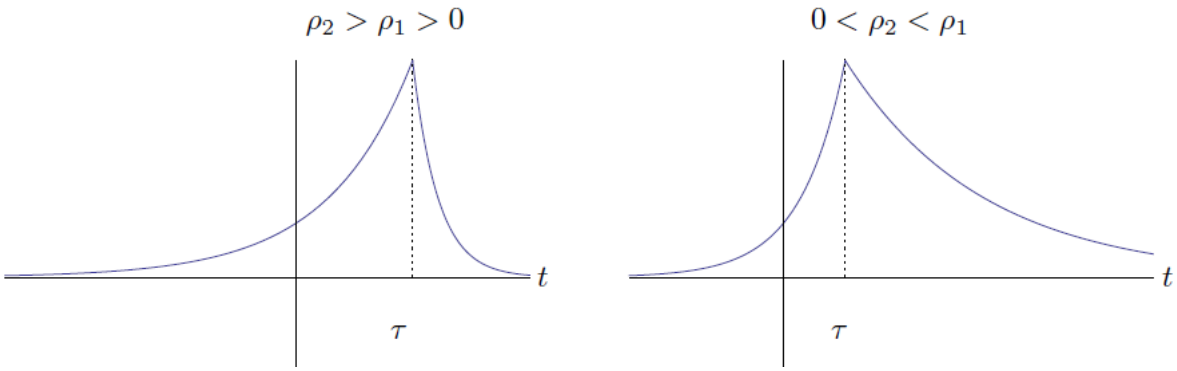
Figure 13: The baseline paths for causal and noncausal AR processes

For causal process (left panel), the baseline path shows an upward jump followed by an exponential decrease, while for noncausal models (right panel), the baseline path features an explosive growth followed by a vertical downturn at $t = \tau$.

Note that, the noncausal $MA(\infty)$ representation B.2 helps understanding the formation of bubbles in the dynamics of noncausal processes. First note that the presence of potentially fat tailed error distributions is likely to produce extreme values of any sign over a finite time period.Now suppose that a very large, say positive, value $\epsilon_\tau$ occurs at time $\tau$. According to B.2 if, for simplicity, the sequence $(\alpha_i)$ is strictly decreasing, for $t \leq \tau$ the weight of that extreme value increases as $t$ approaches $\tau$. This explains the growth phase of the bubble. At $t = \tau + 1$, the extreme value cancels out from the sum and the bubble bursts.

In the mixed process, the path $Y_t$ is a combination of the baseline paths $Z_\tau(t) = I_{\tau<t}\alpha_{t-\tau} + I_{\tau \geq t}\alpha_{\tau-t}$, with stochastic i.i.d. coefficients $\epsilon_\tau$. If the model is a MAR(1,1): $(1 - \phi_1 L)(1 - \varphi_1 L^{-1}) = \epsilon_t$, $|\phi_1| < 1$, $|\varphi_1| < 1$, we get $Z_\tau(t) = (1 - \phi_1\varphi_1)^{-1}(I_{\tau<t}\phi_1^{t-\tau} + I_{\tau \geq t}\varphi_1^{\tau-t})$

Figure 14 shows that the baseline path features an explosive growth followed by an exponential decrease.




Font: Gouriéroux e Zakoïan (2017)

Figure 14: The baseline paths for MAR(1,1) processes

There are two types of asymmetries in the shape of a bubble. Longitudinal asymmetries arise in calendar time when the growth and downturn periods have different lengths, as illustrated in Figure 14. Transversal asymmetries arise when the curvature (resp. the magnitude) at a peak and at a trough are different due to the coefficients of the $MA(\infty)$ representation (resp. to the asymmetric tails of the error distribution).

# ANEXO C – Proof of linear transformation on Econometrics Bayesian model

Define $\boldsymbol{z} = (v_1, \ldots, v_r, \epsilon_{r+1}, \ldots, \epsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$ and $\boldsymbol{y} = (y_1, \ldots, y_T)$ such that $\boldsymbol{z}' = \boldsymbol{B}\boldsymbol{A}\boldsymbol{y}'$. We are interested in the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Note that, as $e_t = \phi(L)v_t$, hence:

$$\boldsymbol{z} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \\ v_{r+1} - \phi_1 v_r - \ldots - \phi_r v_1 \\ \vdots \\ v_{T-s} - \phi_1 v_{T-s+1} - \ldots - \phi_r v_{T-s_r} \\ u_{T-s+1} \\ \vdots \\ u_T \end{pmatrix} = \boldsymbol{B} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \\ v_{r+1} \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{pmatrix}$$

or briefly, $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}$, where

$$\boldsymbol{B} = \left(\begin{array}{ccccc|ccccc|c} \multicolumn{5}{c|}{I_{r\times r}} & \multicolumn{5}{c|}{0_{r\times T-(s+r)}} & 0_{r\times s} \\ \hline -\phi_r & -\phi_{r-1} & -\phi_{r-2} & \dots & -\phi_1 & 1 & 0 & 0 & \dots & 0 & \\ 0 & -\phi_r & -\phi_{r-1} & \dots & -\phi_2 & -\phi_1 & 1 & 0 & \dots & 0 & \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & 0_{T-(s+r)\times s} \\ 0 & 0 & \dots & -\phi_r & -\phi_{r-1} & -\phi_{r-2} & -\phi_{r-3} & \dots & 1 & 0 & \\ 0 & 0 & \dots & 0 & -\phi_r & -\phi_{r-1} & -\phi_{r-2} & \dots & -\phi_1 & 1 & \\ \hline \multicolumn{5}{c|}{0_{s\times r}} & \multicolumn{5}{c|}{0_{s\times T-(s+r)}} & I_{s\times s} \end{array}\right)$$

It is easy to check that the determinant of $\boldsymbol{B}$ is unity because $\boldsymbol{B}$ is an lower triangular matrix with all elements in principle diagonal equal to one. Furthermore, considering $u_t = \phi(L)y_t$ and $v_t = \varphi(L^{-1})y_t$, we can rewrite $\boldsymbol{x}$ as:

$$\boldsymbol{x} = \begin{pmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{pmatrix} = \begin{pmatrix} y_1 - \varphi_1 y_2 \dots \varphi_s y_{s+1} \\ \vdots \\ y_{T-s} - \varphi_1 y_{T-s+1} \dots \varphi_s y_T \\ \\ y_{T-s+1} - \phi_1 y_{T-s} - \dots - \phi_r y_{T-s-r+1} \\ \vdots \\ y_T - \phi_1 y_{T-1} - \dots - \phi_r y_{T-r} \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} y_1 \\ \vdots \\ y_{T-s} \\ \\ y_{T-s+1} \\ \vdots \\ y_T \end{pmatrix} = \boldsymbol{A}\boldsymbol{y}$$

We have not found a simple way to show the structure of matrix $\boldsymbol{A}$, but if we look for simple cases, i.e., a model with lag and lead one, it is simple to conclude that this matrix does not have determinant equal to one.

# ANEXO D – The full conditional distributions for the parameters in the Econometrics Bayesian model

Finally, given the marginal prior densities defined, they showed that the full conditional distributions for $\boldsymbol{\epsilon}^+$, $\boldsymbol{\epsilon}^-$, $\boldsymbol{\phi}$, $\boldsymbol{\varphi}$, $\boldsymbol{\omega}$ and $\tilde{\omega}$ are given by:

$$p(\boldsymbol{\epsilon}^+|\boldsymbol{\phi},\boldsymbol{\varphi},\tilde{\omega},\nu,\boldsymbol{\omega},\boldsymbol{y}) \propto q_1(\boldsymbol{\epsilon}^+)\prod_{t=T+1}^{T+M} t(\epsilon_t|0,\omega^{-1};\nu)$$

$$p(\boldsymbol{\epsilon}^-|\boldsymbol{\phi},\boldsymbol{\varphi},\tilde{\omega},\nu,\boldsymbol{\omega},\boldsymbol{y}) \propto q_2(\boldsymbol{\epsilon}^-)\prod_{t=-M}^{0} t(\epsilon_t|0,\omega^{-1};\nu)$$

$$p(\boldsymbol{\phi}|\boldsymbol{\epsilon}^+,\boldsymbol{\epsilon}^-,\boldsymbol{\varphi},\nu,\boldsymbol{\omega},\boldsymbol{y}) \propto q_2(\boldsymbol{\phi})|A(\boldsymbol{\phi})|N(\bar{\boldsymbol{\phi}},[\boldsymbol{\Phi}_0+wV'\boldsymbol{\Omega}_vV]^{-1})\times I_S(\boldsymbol{\phi})$$

$$p(\boldsymbol{\varphi}|\boldsymbol{\epsilon}^+,\boldsymbol{\epsilon}^-,\boldsymbol{\phi},\tilde{\omega},\nu,\boldsymbol{\omega},\boldsymbol{y}) \propto q_1(\boldsymbol{\varphi})|A(\boldsymbol{\varphi})|N(\bar{\boldsymbol{\varphi}},[\boldsymbol{\Psi}_0+wU'\boldsymbol{\Omega}_uU]^{-1})\times I_S(\boldsymbol{\varphi})$$

$$p(\boldsymbol{\omega}|\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^+,\boldsymbol{\epsilon}^-,\tilde{\omega},\nu,\boldsymbol{y}) \propto \prod_{t=-M}^{0} t(\epsilon_t|0,\omega^{-1};\nu)\prod_{t=T+1}^{T+M} t(\epsilon_t|0,\boldsymbol{\omega}^{-1};\nu)G(T/2,S/2).$$

$$(\nu+\omega[\phi(L)\varphi(L^{-1})y_t]^2)\tilde{\omega}_t|\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^+,\boldsymbol{\epsilon}^-,\boldsymbol{\omega},\nu,\boldsymbol{y} \sim \chi^2_{\nu+1}$$

and

$$p(\nu|\boldsymbol{\epsilon}^{+},\boldsymbol{\epsilon}^{-},\boldsymbol{\omega},\tilde{\omega},\boldsymbol{y}) \propto \sum_{t=-M}^{0} t(e_t|0,\omega^{-1};\nu) \prod_{t=T+1}^{T+M} t(e_t|0,\omega^{-1};\nu)$$

$$\times[2^{\nu/2}\Gamma(\nu/2)]^{-T}\nu^{T\nu/2}\left(\prod_{t=1}^{T}\tilde{\omega}^{(\nu-2)/2}\right) exp\left\{-\left(\frac{1}{\nu_0}+\frac{1}{2}\sum_{t=1}^{T}\tilde{\omega}\right)\nu\right\}$$

See Lanne, Luoma e Luoto (2012) for the expressions of functions $q_1(\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^{+},\boldsymbol{y})$, $q_2(\boldsymbol{\phi},\boldsymbol{\varphi},\boldsymbol{\epsilon}^{-},\boldsymbol{y})$ and matrices $\boldsymbol{\Omega}_u$, $\boldsymbol{\Omega}_v$, $U$ and $V$.

# ANEXO E – Lasso model regression

In order to obtain a posteriori samples of the parameters of interest presented in equation 3.42, techniques such as the Gibbs sampler and the Metropolis-Hastings algorithm are combined. Therefore, we need, again, find the complete conditional distributions of the parameters. With regard to conditional distributions defined in Section 3.15 and in addition to the penalty parameter $\eta$ that was incorporated into the model, the only changes were related to the specification of $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ and these priors depend on $\sigma^2$. For this reason, only the complete conditional distributions of these parameters will need to be recalculated in relation to what was defined in the equations 3.25, 3.26 and 3.27.

Given the marginal prior densities defined in 3.39, 3.40 and 3.22, it is therefore straightforward to verify that the full conditional distributions for $\boldsymbol{\phi}$, $\boldsymbol{\varphi}$, $\sigma^2$ and $\eta$ are given by

$$p(\boldsymbol{\phi}|\boldsymbol{\theta}_{-\phi}, \mathbf{y}^*, \eta, \boldsymbol{y}) \propto \prod_{t=1}^{T} exp\left\{-\frac{\delta_t}{2\sigma^2}\left[\phi(L)\varphi(L)\right] y_t\right\} \times \prod_{i=1}^{r} e^{-\eta|\phi_i|/\sqrt{\sigma^2}}, \tag{E.1}$$

$$p(\boldsymbol{\varphi}|\boldsymbol{\theta}_{-\varphi}, \mathbf{y}^*, \eta, \boldsymbol{y}) \propto \prod_{t=1}^{T} exp\left\{-\frac{\delta_t}{2\sigma^2}\left[\phi(L)\varphi(L)\right] y_t\right\} \times \prod_{j=1}^{s} e^{-\eta|\varphi_j|/\sqrt{\sigma^2}}, \tag{E.2}$$

$$p(\sigma^2|\boldsymbol{\theta}_{-\sigma^2}, \mathbf{y}^*, \eta, \boldsymbol{y}) \propto \prod_{t=1}^{T} \frac{\delta_t^{1/2}}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{\delta_t}{2\sigma^2}\left[\phi(L)\varphi(L)\right] y_t\right\} \tag{E.3}$$

$$\times \prod_{i=1}^{r} \frac{1}{2\sqrt{\sigma^2}} e^{-\eta|\phi_i|/\sqrt{\sigma^2}} \times \prod_{j=1}^{s} \frac{1}{2\sqrt{\sigma^2}} e^{-\eta|\varphi_j|/\sqrt{\sigma^2}}, \tag{E.4}$$

$$p(\eta|\theta, \mathbf{y}^*, \boldsymbol{y}) \sim Ga\left(T + a_\eta, \frac{1}{\sqrt{\sigma^2}}\left[\sum_{i=1}^{r}|\phi_i| + \sum_{j=1}^{s}|\varphi_j| + b_\eta\right]\right). \tag{E.5}$$

Only the density in E.5 is standard and can be readily used to simulate random numbers. Therefore, we adopt a Metropolis–Hastings algorithm described. For $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$

were used normal distributions, evaluated at the current value, as a candidate distribution and for $\sigma^2$, a log normal distribution also evaluated at the current value was assumed.