Modelling and forecasting economic time series with mixed causal-noncausal models

Doctoral thesis

MODELLING AND FORECASTING ECONOMIC TIME SERIES WITH MIXED CAUSAL-NONCAUSAL MODELS

Elisa Voisin

2022

This book was typeset by the author using LaTeX.

ISBN: 978-94-6469-118-4
Printed in the Netherlands by ProefschriftMaken

MODELLING AND FORECASTING ECONOMIC TIME SERIES WITH MIXED CAUSAL-NONCAUSAL MODELS

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. Habibović,
in accordance with the decision of the Board of Deans,
to be defended in public
on Tuesday, 20 December 2022, at 16.00 hours

by

Elisa Marie Voisin

Supervisor
Prof. dr. Alain Hecq

Co-supervisor
Dr. Ines Wilms

Assessment Committee
Prof. dr. Bertrand Candelon (chair)
Prof. dr. Joann Jasiak, York University, Canada
Dr. Stephan Smeekes
Prof. dr. Jean-Michel Zakoïan, ENSAE-CREST, France

In memory of my aunt and godmother, Gigie

Acknowledgments

My time in Maastricht started in 2013 with my broken English and the discovery of bitterballen. During my studies here I have had the chance to meet so many great people. Some of whom I would like to thank in particular (more or less chronologically).

I would like to start by thanking Louisdo and Simon with all my heart, I could not have dreamt of better friends to begin this adventure with nine years ago. I would also like to thank Frederique, Julie, Mattia and Ger for being such good friends. You guys have always been an essential support for me. In short, I lost a tooth but earned lifetime friends in the process. I cannot mention everyone so let me say *thank you* to all my friends that I have met during the bachelor and the master. You made my student years extraordinary and you made me love Maastricht so much that I decided to stay four more years..! To be fair, it might have required a bit more convincing and as my supervisor Alain said to me when I was in doubt: "four years is nothing". Indeed, time flew and it's already over.

I want to thank you, Alain, for motivating me. I also want to thank you for your guidance, your availability and your patience over the last years (and also for planning most meetings after 10am). Thank you, Ines, for your help, especially during the critical time of the last stages of the thesis. I would also like to thank the members of the assessment committee, Bertrand Candelon, Joann Jasiak, Stephan Smeekes and Jean-Michel Zakoïan, for their careful reading and approval of my thesis.

I am thankful for all the people I got the chance to meet during my PhD, at university, conferences and summer schools. I am also grateful to have had the opportunity to work with João Issler and to go to Rome for three months to work with Gianluca Cubadda. I would like to thank all the PhDs that I have met at KE and also some from other

departments (in spite of the reputation of KE people hiding in their cellar...). Lunch and coffee breaks were always a good way to get my mind off research. It always felt like an adventure, walking towards the coffee machine, wondering whether it would be working or not, or whether the coffee would smell like sewage water.

There are a few more people that I would like to thank personally. First, Moritz and Dewi, who also started this adventure 9 years ago with me! Momo, from teammates in macro to defending our PhDs a few weeks apart, I am grateful for all the moments we shared together! Dewi, you were the perfect office mate. We always found the right balance between gossiping, complaining and working! Rasmus, I wish you had started your PhD at the same time as us. That way, you wouldn't have abandoned us so early... You are such a precious friend (and not just because you do amazing BBQs). From the moment you left, I have missed you and I have missed having a (or some...) Triple Hop at Gouv' with you.

I really enjoyed the time with all of you at uni and also outside, having after-work biertjes... until the plot twist in the story when all bars closed and we were forced to keep 1.5m away from each other (for quite a while). So, although nothing much happened in this 2 to 6 years (difficult to say) time-warp, I was surrounded by my most loyal friends: Ajax and Haga. We would often meet with their two Polish flatmates and an unemployed German. Though, contrary to Ajax and Haga, those three people could actually talk back to me... More sentimentally: you three, Alex, Adam and Monika, have been my greatest support, not only during this period but during the whole PhD. Alex, I am ever so grateful to have you by my side for four years now and I cannot thank you enough for everything you have done for me. I look so much forward to this new adventure with you. Adam (though everyone knows you as Mama J), I could not have hoped for a better friend, co-worker and office-ball partner this whole time. We make such a great team, I mean, two nerds who laugh at very mature jokes and who like absolutely proper techno, it couldn't have gone wrong (or maybe that's where it went wrong). Monika, I am sorry for all

the econometrics talks you had to endure! Thank you so much for cheering me up countless times with the perfect comfort food and for laughing with me about worteltjes. I am so thankful to have shared this PhD Tour (which had little to do with the PhD) with the both of you. I hope we will keep our tradition of family holidays and that our love for techno-karaoke and *too great meatloaf* will never die!

Fortunately, after a while we got the possibility to travel again and I am grateful that we managed to organise my research visit to Tor Vergata. For that I would like to thank once more Gianluca for inviting me and all the PhDs there for being so welcoming. Special thanks to Filippo, Luca and (the non-Italian) Hanno for all the help, either research-wise or to get around and discover Rome! Hanging out with you guys was every time a lot of fun. Francesco, I am glad you were my topic/courses/Rome partner, you have been a great help and I always enjoyed spending time with you! Marie and François, I give you an absolutely not awkward thumb-up-bump for joining the schnitzel+bowling-waterslide=bouldering crew! I hope you will keep the tradition going (based on what I have witnessed I fully trust you François to do so, even with strangers). It was such a nice way to close this chapter with you guys! There are of course so many other people that I would like to thank, some who left before me and others that I am leaving now. Once again, sorry I cannot mention everyone, but I thank you all for the times we shared together.

Lastly, I would like to thank my friends back in France and my family for their love and support. More specifically, the greatest thanks to my parents and brother. Being far away from you (and the *toutous*) is always difficult, but all the calls and the *apéros* we shared on facetime always made me feel like I was with you for a little while. Thank you for everything.

Elisa Voisin, October 2022

Contents

1
Introduction

A time series is a variable that is observed over time at a certain frequency. It can for instance be a stock price that is tracked every minute or gross domestic product measured quarterly. Time series models therefore intend to explain the evolution over time of these variables. Most linear models however require stationary series, namely variables fluctuating around their mean. Nonetheless, stationary time series can sometimes be characterised by some non-linear features. For instance, locally explosive episodes, which are long lasting increases followed by a sudden crash, are difficultly modeled with linear models. Such episodes, defined in this thesis as *bubbles*, can be observed in many commodity prices, inflation rates, stock prices or cryptocurrencies for instance. Most models that intend to capture these processes are thus either highly non-linear or constructed to alternate between distinct specifications, requiring the estimation of a substantial amount of coefficients.

As an example, Figure 1.1 depicts West Texas Intermediate (WTI) and Brent crude oil monthly prices between mid 1987 to late 2020.

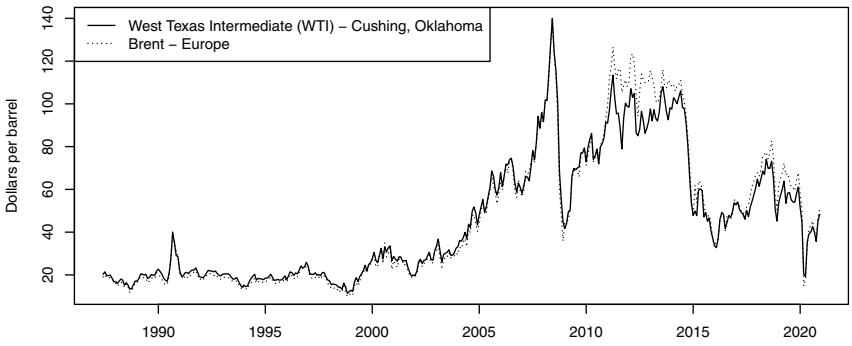


Figure 1.1: WTI and Brent crude oil prices

The episode in 2008, observed in both series, clearly represents the above-mentioned bubble pattern. Commodities are at the bottom of the supply chain leading to global consumption. Hence, persistent increases in commodity prices will induce a surge in marginal costs along the supply chains. This will in turn translate to a persistent increase in inflation, thus also affecting households. Understanding the dynamics and behaviours of such processes is therefore of great importance for investors but also for policy makers. There can be many reasons for these bubbles, it can be the anticipation of a policy implementation after an announcement, speculation, supply shortages, etc. The fact is that we do not always know the reason, and more specifically, rarely do we know when the sudden drop will happen. This emphasises the importance of not only being able to model them but also to forecast them.

One could for instance think of various variables that help explain the price of oil over time. However, their inclusion in a model would imply the necessity to have the predictions of these variables to forecast the price of oil itself. For this reason, autoregressive (AR) models became very popular. They proved to provide great fit to the

data in many applications and often demonstrated better forecast performances.

An AR model explains a variable today with its own past observations. An AR(p) process, say y_t, depends on its last p values and the model looks as follows,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t.$$

The roots of the lag polynomial $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$ all lie outside the unit circle for stationary processes, with L being the lag operator $L^i y_t = y_{t-i}$. The error term ε_t follows an *i.i.d.* distribution, often assumed to be Gaussian. The next value y_{t+1}, for instance, would easily be predicted using the last p observed values of the process. Such models have proven to outperform a significant amount of models in various areas of applications. However, they are unable to capture non-linear dynamics. Hence, the need to construct parsimonious and stationary models for series characterised by non-linear features gave rise, in line with the aforementioned autoregressive models, to mixed causal-noncausal (MAR) models.

MAR models are autoregressive models not only using past values of the process but also future values. The model is constructed in a multiplicative structure of the backward- and forward-looking lag polynomials as follows,

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t, \tag{1.1}$$

where L^{-i} is the lead operator, $L^{-i} y_t = y_{t+i}$. An MAR(r, s) process, wifh r lags and s leads, has therefore a lag polynomial $\Phi(L)$ of order r and a lead polynomial $\Psi(L^{-1})$ of order s. While having both lag and lead polynomials stationary, this model manages to capture non-linear features such as bubbles. A condition for the identification of MAR models is the non-Gaussianity of the error term, which appears to be quite common in practice and thus not restrictive.

An MAR process can be filtered into a purely causal and a purely noncausal component. The following filtration of the process y_t,

$$u_t = \Phi(L)y_t,$$

makes u_t a purely noncausal MAR$(0, s)$ process,

$$\Psi(L^{-1})u_t = \varepsilon_t.$$

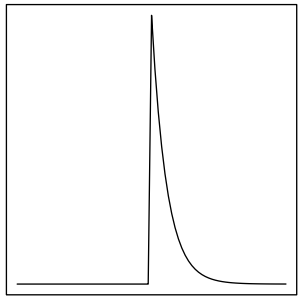
This purely noncausal component u_t will be the process of interest in most of the chapters of this thesis since it is the forward looking component of MAR models that makes derivations or forecasts more intricate. Results are then easily extendable to mixed processes.

The increasing interest in MAR models and their numerous empirical applications in the literature demonstrate their versatility and the benefits of using these models in many areas of applications. MAR models have proven to provide a better fit than purely causal models for inflation rates (Lanne and Saikkonen, 2011), Bitcoin prices (Hencic and Gouriéroux, 2015), crude oil prices (Gourieroux, Jasiak, and Tong, 2021) and many other commodity prices and financial times series (see among others Hecq, Lieb, and Telg, 2016; Fries and Zakoïan, 2019a). However, the literature employing and analysing MAR models for prediction purposes is still very limited. This thesis hence focuses primarily on the forecasting aspect of these models.

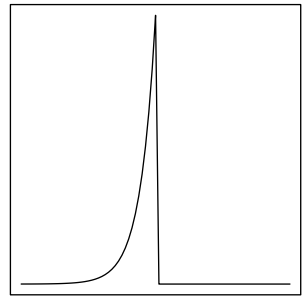
The thesis first aims attention at univariate models with a unique lead, since it is already sufficient to generate the bubble pattern in a series. As an example, Figure 1.2 shows the differences in trajectories of transitory shocks in purely causal, noncausal and mixed processes, namely with a lag and/or a lead. In a purely causal process (graph (a)), shocks have a sudden impact which will wear off over time, while for a purely noncausal process (graph (b)) the

impact is mirrored, there is an anticipating effect until the time of the shock. Logically, for mixed processes (graph (c) and (d)), shocks have an impact both before and after, and the symmetry of the impact around the shock depends on the magnitude of the coefficients.

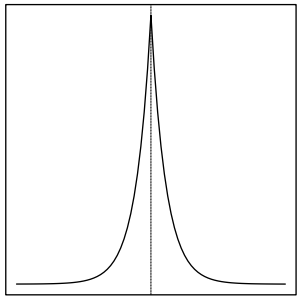
(a) MAR(1,0) with $\phi = 0.8$



(b) MAR(0,1) with $\psi = 0.8$



(c) MAR(1,1) with $(\phi, \psi) = (0.8, 0.8)$



(d) MAR(1,1) with $(\phi, \psi) = (0.3, 0.8)$

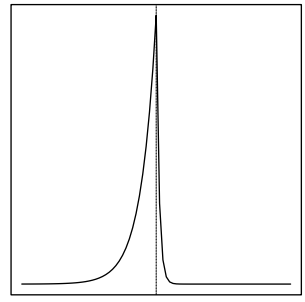


Figure 1.2: Effects of a lag and a lead on transitory shocks of MAR processes

Forecasting a purely causal process is straightforward, however this is not the case in the presence of a lead. Indeed, there is the need to forecast future values which are dependent on even farther future values. However, this can be overcome and it is possible to recover, or approximate, the conditional distribution of the process. That is, the

probability density function of the forecasted values given the past observed points. This allows to recover, especially during explosive episodes, the bimodality of the conditional distribution of MAR processes, which would not be captured by purely backward-looking models. This bimodality assigns probabilities to the two distinct potential outcomes that can happen, namely a further increase or a sudden drop. These bimodal densities carry therefore much more information and are more representative of future outcomes than point forecasts or than unimodal distributions derived from purely causal AR models for instance.

Chapter 2 first analyses MAR(0,1) processes with Cauchy-distributed errors, for which a closed-form expression of the predictive density exists. This allows to have a benchmark of the performance of the approximation methods. The first method was developed by Lanne, Luoto, and Saikkonen (2012) and is based on simulations of future errors to approximate for instance point forecasts or conditional cumulative probabilities, at a given forecast horizon. However, this approach becomes quite computationally demanding when the whole conditional density is needed. The second method was developed by Gouriéroux and Jasiak (2016) and allows to obtain the whole conditional density using sample approximations. While this approach is much more efficient in obtaining predictive densities, it depends on the sample size and also becomes computationally demanding as the forecast horizon increases. To account for the latter, Gouriéroux and Jasiak (2016) developed a sampling importance resampling (SIR) algorithm, which recovers the conditional density from an instrumental model from which it is easier to forecast. We find that for median values, all approaches yield similar results, however we start observing discrepancies as the process starts deviating.

The first issue one may face when analysing time series is the non-stationarity of the data. Taking the first difference of the series

will for instance eliminate the bubble-looking patterns that are of interest. Furthermore, time series are often stationary around a (non)linear trend. It is the latter that we consider in **Chapter 3**, which first analyses the impact of different detrending methods on the identification of MAR models in a simulation study. The chapter then turns to the forecast of the oil prices shown in Figure 1.1 during the COVID-19 pandemic outbreak. We employ the Hodrick-Prescott (HP) filter developed by Hodrick and Prescott (1997) but also two polynomial trends of orders 4 and 6 to illustrate the variations in results based on the detrending method applied. A criticism of the HP-filter or large order polynomial trends is their lack of interpretation. Hence, we also detrend the series using the US crude oil Strategic Petroleum Reserves (SPR), which are last resort reserves. They are much less volatile than crude oil stocks for instance and appear to follow the same trend as crude oil prices.

Chapter 4 proposes a short-term credibility index for Central banks that employ an inflation targeting system. The chapter is more specifically applied to Brazil, where the central bank provides a target inflation and tolerance bounds around it since 1999. The inflation rate is modelled using MAR models and we forecast the inflation rate during a rather stable period, which allows us to use both the method of Lanne, Luoto, and Saikkonen (2012) and the SIR algorithm developed by Gouriéroux and Jasiak (2016) to forecast farther horizons. We thus estimate the probabilities that the inflation rate will remain within the target bounds announced by the Central bank 1, 3 and 6 months ahead and investigate the added value of including three key drivers of the economy in the model. We then use the obtained one-month ahead probabilities as a short-term credibility index of the Central bank targeting system. We suggest employing receiver operating characteristic curves to determine the optimal probability threshold from which the bank is predicted to be credible. The literature on the credibility of Central banks is quite extensive but mostly focuses long term credibility based on people's beliefs.

We argue that our short-term credibility index is complementary to long-term indices, and together provide a more detailed and clearer picture of the current and predicted credibility of the Central Bank inflation targeting system.

Chapter 5 addresses more general forms of MAR models and investigates cases with more than one lead. This Chapter gathers some partial results and some more promising results for future applied research with MAR(r, s) models. Analogously to Chapter 2, the initial motivation of this chapter is to obtain a theoretical benchmark of the predictive density to evaluate the performance of the approximations methods. The first part of the chapter hence attempts to find the closed-form expression of the conditional distribution of an MAR(0,2) process with Cauchy-distributed errors, as an extension of the findings of Gouriéroux and Zakoïan (2013) for processes with one lead. We fist present the marginal distribution of the MAR(0,2) process, which is also Cauchy, but unfortunately, we were not able to derive the theoretical conditional distribution. This however, does not prevent the use of the approximation methods and their results can be analysed based on the knowledge we know from the first chapters. The second part of this chapter proposes a prediction method built on the two existing ones. It allows to estimate the conditional density using simulations in a more efficient way than from the approach of Lanne, Luoto, and Saikkonen (2012). This proposed method is compared to the sample-based approach of Gouriéroux and Jasiak (2016) on simulated series and both approaches are then illustrated on the price index of all metals, strongly forward looking as it is identified as an MAR(0,2).

Chapter 6 turns to the multivariate framework of MAR models, denoted as VMAR. We often notice similarities in the locally explosive episodes in series such as in many commodity prices for instance or cryptocurrencies. This chapter therefore proposes ways to detect such commonalities in the noncausal components. A particularity

of the VMAR model is that the multiplicative structure as defined in (1.1) is not commutative. In the univariate setting, we have that $\Phi(L)\Psi(L^{-1}) = \Psi(L^{-1})\Phi(L)$, however, this is not the case in the multivariate model, given the non commutativity property of the matrix product. It implies that there exist two distinct, yet equivalent, representations of the same process with different coefficient matrices. The chapter therefore first investigates the conditions for the presence of common bubbles (CB), for each of the representations. We find that the representation starting with the lead polynomial makes the detection of CB easier. From that, we propose a likelihood ratio test comparing an unrestricted model which does not have commonalities with a restricted model on which commonalities are imposed. We also consider model selection using information criteria to detect CB. We present a simulations study and the approach is then illustrated on three commodity indices. While they follow similar patterns we were not able to detect commonalities neither in the bivariate combinations nor in the trivariate models.

2

Forecasting bubbles with mixed causal-noncausal autoregressive models

Adapted from: Alain Hecq and Elisa Voisin (2021). "Forecasting bubbles with mixed causal-noncausal autoregressive models". In: *Econometrics and Statistics* 20, pp. 29–45.

Abstract

Density forecasts of locally explosive processes are investigated using mixed causal-noncausal models, namely time series models with both lag and lead components. In the absence of theoretical expressions for the predictive density for a large range of potential error distributions, two approximation methods are analysed and compared using Monte Carlo simulations. The focus is on the prediction of one-step ahead probabilities of turning points during bubble episodes. A thorough analysis provides some guidance in using these approximation methods during extreme events, with the suggestion to consider both approaches together as they jointly carry more information. The analysis is illustrated with an application on Nickel prices, focusing on the financial crisis bubble.

2.1 Introduction

Locally explosive episodes have long been observed in financial and economic time series. Such patterns, often observed in stock prices, can be triggered by anticipation or speculation. Given this forward-looking aspect, expectation models have been prevalent for modelling them. As shown for instance by Gouriéroux, Jasiak, and Monfort (2016), equilibrium rational expectation models admit a multiplicity of solutions, and some of them feature such speculative bubble patterns. In this chapter, speculative bubbles, or simply bubbles refer to as processes characterised by a rapid and persistent increase followed by a sudden crash. Some authors define bubbles as deviations from the fundamental solution of a present value type model (e.g. Diba and Grossman, 1988). Those bubbles might not have the non-linear pattern that we investigate in this chapter. From an empirical point of view, models employed to capture bubble features range from simple approaches, such as single bubble models with a constant probability of crash, to rather complex models depending on numerous parameters (e.g. see the autoregressive specifications with breaks in the autoregressive roots of Phillips, Wu, and Yu, 2011). Although those latter models may a posteriori fit the data well, they are either not designed for prediction purposes or render predictions intricate due to their dependence on extensive parameters estimation. Mixed causal-noncausal autoregressive (hereafter MAR) models are hence an appealing alternative to model such type of stationary non-linear time series.

MAR processes are univariate autoregressive models incorporating not only lags but also leads of the dependent variable, with potentially heavy-tailed errors (e.g. Cauchy and Student's t distributions). While being parsimonious, MAR models generate non-linear dynamics such as locally explosive episodes in a strictly stationary setting (Fries and Zakoïan, 2019b). So far, the focus has mainly been put on the identification and the estimation of such models. Lanne, Luoto, and

Saikkonen (2012), Hencic and Gouriéroux (2015) and Hecq, Lieb, and Telg (2016) show that the inclusion of noncausal components (leads) in autoregressive models is favoured, explaining respectively the observed bubbles in inflation series, in Bitcoin prices and in the demand of solar panels in Belgium. A logically arising interrogation concerns the ability of MAR models to forecast. Gouriéroux and Zakoïan (2017) derive conditional point and density forecasts of purely noncausal MAR(0,1) processes with Cauchy-distributed errors. Fries (2018) derives up to the fourth conditional moments of α-stable MAR(0,1) processes and shows that during explosive episodes the conditional distribution simplifies to a Bernoulli process putting weight on a crash and a further increase. Other distributions however, such as Student's t distributions for instance, leads to an absence of closed-form results for the conditional moments and distribution. Due to the absence of theoretical formulas to predict such processes, Lanne, Luoto, and Saikkonen (2012) and Gouriéroux and Jasiak (2016) developed numerical approximation methods based on simulations and on past realised values respectively.

A key feature of MAR models is their ability to generate locally explosive episodes. Yet, the literature regarding the proficiency of such models to predict turning points during explosive episodes as well as the performance of the aforementioned approximation methods remains scarce (see also Lanne, Nyberg, Saarinen, et al., 2012; Gouriéroux, Hencic, and Jasiak, 2018). The first contribution of this chapter is consequently to focus on extreme episodes and to analyse and compare in details the two numerical approaches developed for forecasting MAR process. The chapter considers MAR(r,1) processes with unconstrained r number of lags, a unique positive lead coefficient and a fat-tail error distribution (possibly with infinite variance). We compare approximation methods to theoretical results and then assess their performance when no-closed form expressions exist.

We show that at the outset of a bubble the predictive density of the process splits into a bi-modal distribution, putting weights on a crash and on a further increase. This corroborates Fries's (2018) results for α-stable MAR(0,1) processes and would not be detected with standard Gaussian ARMA models. We detect that both approximation methods yield similar and accurate probabilities during stable episodes but differ during bubbles. Overall, we find that the simulations-based approach of Lanne, Luoto, and Saikkonen (2012) is a good approximation of theoretical probabilities, given an adequate number of simulations, and that the sample-based method of Gouriéroux and Jasiak (2016) is characterised by a learning mechanism. The latter is composed of the theoretical probabilities that are inflated or deflated by past events in the sample. Our results show that combining results obtained from the two methods can help decompose the sample-based probabilities into (a proxy of) the theoretical probabilities and the learnt probabilities, when no closed-form exist.

The chapter is constructed as follows. Section 2.2 introduces mixed causal-noncausal autoregressive models. Section 2.3 discusses how they can be used for prediction when the conditional moments and densities admit closed-form expressions. In Section 2.4 are presented the simulations-based forecasting approach proposed by Lanne, Luoto, and Saikkonen (2012), followed by the sample-based method proposed by Gouriéroux and Jasiak (2016). Both approaches are first compared to Cauchy-derived closed-form results to illustrate their performance and are then applied to Student's t-distributed processes. In Section 2.5 both approximation methods are illustrated with an application to a detrended Nickel prices series that has been found to follow an MAR(1,1) process. Section 2.6 concludes.

2.2 Mixed causal-noncausal autoregressive models

Consider the univariate MAR(*r*,*s*) process defined as follows,

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,$$

where L and L^{-1} are respectively the lag and lead operators; Φ and Ψ are two invertible polynomials with degrees r and s respectively. That is, $\Phi(L) = (1-\phi_1 L-\cdots-\phi_r L^r)$ and $\Psi(L^{-1}) = (1-\psi_1 L^{-1}-\cdots-\psi_s L^{-s})$ with roots strictly outside the unit circle. The error term ε_t is i.i.d, following a non-Gaussian distribution. This assumption, not empirically restrictive since non-normality is widely observed in financial and economic time series, is necessary to achieve the identification of the model. An MAR(*r*,*s*) model can also be expressed as a causal *AR* model where y_t depends on its own past,

$$\Phi(L)y_t = u_t, \tag{2.1}$$

where u_t denotes the purely noncausal component of the errors ε_t, which itself depends on its own future,

$$\Psi(L^{-1})u_t = \varepsilon_t. \tag{2.2}$$

Alternatively, we can also filter the process as $\Phi(L)v_t = \varepsilon_t$ with $\Psi(L^{-1})y_t = v_t$ to obtain the backward component of the errors, v_t. The process y_t admits a stationary infinite two-sided *MA* representation and depends on past, present and future values of ε_t,

$$y_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}.$$

The case in which all coefficients a_i for $-\infty < i \leq 0$ (resp. $0 \leq i < \infty$) are equal to 0, corresponds to a purely causal (resp. noncausal) model.

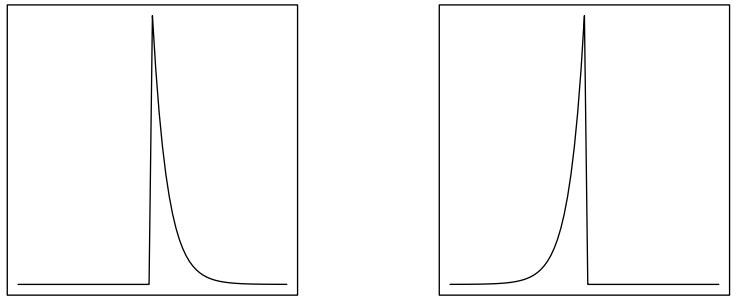
In spite of their apparent simplicity and their parsimonious

representation, MAR models often provide a better fit to economic and financial data than ARMA models as they capture non-linear dynamics such as bubbles or asymmetric cycles. Figure 2.1 displays how the presence of a lag, a lead, or both, affects the shape of transitory shocks in MAR processes. Purely causal (resp. noncausal) processes are only affected by a shock after (resp. before) the impact; this is shown in graph (a) (resp. (b)). Consequently, mixed processes are affected both in anticipation and after the shock; the shape of the locally explosive episode (mostly forward or backward looking) depends on the magnitude of the lag and lead coefficients. When the coefficients are identical (c) the effects of the shock are symmetric around the impact. When the coefficient of the lead is higher (d), the explosive episode is more analogous to what we refer to as a bubble with an asymmetry around the peak, making the increasing phase more persistent than the decrease.

The usual practice for estimating and identifying MAR models is as follows. It is well known that methods based on first and second moments (e.g. OLS) are unable to distinguish between purely causal, noncausal or mixed processes as their autocovariance functions are identical. Fitting an autoregressive model by OLS (the so called pseudo causal model) allows however to estimate the sum of leads and lags components, p. Note that a non-Gaussian MLE can give misleading results (Gouriéroux and Jasiak, 2018). Subsequently, the respective numbers of lags (r) and leads (s), such that $r + s = p$, can be estimated by an approximate maximum likelihood (hereafter AML) approach (Lanne and Saikkonen, 2011). The selected model is the one maximising the AML with respect to the positive integers r and s and the continuous parameters $\Omega = (\Phi, \Psi, \Theta)$, where $\Phi = (\phi_1, \ldots, \phi_r)$, $\Psi = (\psi_1, \ldots, \psi_s)$ and Θ comprises the error distribution parameters, such as the scale or location for instance. The AML estimator is defined as follows,

$$\left(\hat{\Phi}, \hat{\Psi}, \hat{\Theta}\right) = argmax_{\Phi,\Psi,\Theta} \sum_{t=r+1}^{T-s} \ln\left[g\Big(\Phi(L)\Psi(L^{-1})y_t; \Theta\Big)\right],$$

(a) MAR(1,0) with $\phi = 0.8$ (b) MAR(0,1) with $\psi = 0.8$



(c) MAR(1,1) with $(\phi,\ \psi) = (0.8,\ 0.8)$ (d) MAR(1,1) with $(\phi,\ \psi) = (0.3,\ 0.8)$

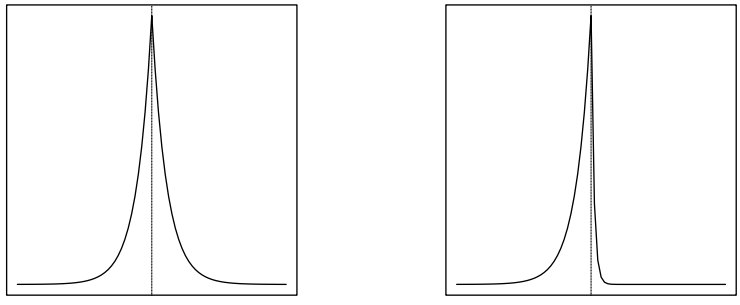


Figure 2.1: Effects of a lag and a lead on transitory shocks of MAR processes

where g denotes the *pdf* of the error term, satisfying the regularity conditions (Andrews, Davis, and Breidt, 2006). Lanne and Saikkonen (2011) show that the resulting (local) maximum likelihood estimator is consistent, asymptotically normal and that $(\hat{\Psi}, \hat{\Phi})$ and $\hat{\Theta}$ are asymptotically independent, for Student's t-distributed errors with finite variance. Since an analytic solution of the maximisation problem at hand is not directly available, numerical gradient-based procedures can be employed. Hecq, Lieb, and Telg (2016) indicate that estimating MAR models is easier for more volatile series since the convergence of the estimator is empirically faster for distributions with fatter tails.

They propose an alternative way to obtain the standard errors, a method implemented in the R package MARX (Hecq, Lieb, and Telg, 2017a).

While with a unique positive lead coefficient, say ψ, explosive episodes increase at a fixed rate ψ^{-1} until a sudden crash, other specifications induce more complex patterns not resembling the bubble pattern that this analysis focuses on. This chapter hence only provides results for stationary MAR(*r*,1) processes,

$$\Phi(L)(1-\psi L^{-1})y_t = \varepsilon_t,$$

with $\psi > 0$ and where $\Phi(L)$ is the lag polynomial of order $r \geq 0$. We will refer to $u_t = \Phi(L)y_t$ as the purely non-causal component of the process. Throughout the analysis we consider three distributions: the standard Cauchy, the Student's *t*(2) and the Student's *t*(3) distributions. Furthermore, the data generating process (hereafter *dgp*) is assumed correctly identified to disregard the estimation uncertainty.

2.3 Predictions using closed-form expressions

Conditional expectations can be used to predict the next points but forecasting densities is more informative to analyse the probabilities of potential future paths. The latter can be employed to evaluate the probabilities of a turning point in an explosive episode. However, the anticipative aspect of MAR models complicates their use for predictions. Results are not as straightforward as they could be with purely backward-looking ARMA models. Although in some cases point and density forecasts can be directly obtained from the assumed errors distribution, they sometimes need to be approximated.

Given a correctly identified *dgp*, the information sets $(y_1, \ldots, y_T, y^*_{T+1}, \ldots, y^*_{T+h})$ and $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-1}, u_T, u^*_{T+1},$

$\ldots, u^*_{T+h})$, where $v_t = \Phi(L)^{-1}\varepsilon_t$ and $u_t = (1 - \psi L^{-1})^{-1}\varepsilon_t$, are equivalent (Gouriéroux and Jasiak, 2016). The asterisk indicates future values of the variables to be forecasted. This equivalence allows to predict future values of the mixed process y from predictions of the forward-looking component of the process, namely u, using the relation $\Phi(L)y_t = u_t$. Most prediction methods hence aim attention at purely noncausal processes.

2.3.1 Point predictions

The linear projection on the past does not correspond, for MAR processes, to the conditional expectation (Gouriéroux and Jasiak, 2018). Gouriéroux and Zakoïan (2017) derive the first two conditional moments of MAR(0,1) processes, here denoted as u_t. They show that for Cauchy processes, with a positive lead coefficient, the conditional point forecast of u_{T+1} is a random walk,

$$\mathbb{E}[u_{T+1}|\mathcal{F}_T] = u_T. \tag{2.3}$$

This is puzzling since despite being stationary, the Cauchy-distributed process u_{T+1} has a unit root in its conditional expectation. Fries (2018) provides up to the fourth conditional moments of α-stable-distributed MAR(0,1) processes, given some admissible parametrisation of the tail and asymmetry parameters. He also derives the limiting distribution of those four moments when the process diverges. He shows that during an explosive episode, the computation of those moments gets considerably simplified and are characteristic of a weighted Bernoulli distribution charging probability $\psi^{\alpha h}$ to the value $\psi^{-h}u_T$ and $(1 - \psi^{\alpha h})$ to value zero, for a tail parameter $0 < \alpha < 2$ and a forecast horizon $h \geq 1$. Those results indicate that along a bubble, the process can only either keep on increasing at fixed rate or drop to zero. For Cauchy-distributed processes ($\alpha = 1$), the mean forecast during an explosive episode remains equal to Equation (2.3), yet for other α-stable distributions the conditional expectation may be drastically simplified.

Overall, during an explosive episode, the conditional expectation of an MAR(0,1) process is a weighted average of the crash and the further increase (hence the random walk for Cauchy-distributed processes). Clearly, point forecasts in this case lose substantial information and density forecasts would consequently be more instructive.

2.3.2 Density predictions

The joint conditional predictive density (as named by Gouriéroux and Jasiak, 2016) or the causal transition distribution (as named by Gouriéroux and Zakoïan, 2017) of the h future values of an MAR(0,1) process, $(u^*_{T+1}, \ldots, u^*_{T+h})$, which is a Markov process of order one, can be obtained given the value of the last observed point u_T. While the interest is on predicting the future given contemporaneous and past information, it is only possible, by the model definition, to derive the density of a point conditional on its future point. Bayes' Theorem is applied repeatedly until all conditional *pdf*'s are conditioned on the value of future points. The resulting predictive density of an h-step ahead forecast is as follows,

$$
\begin{aligned}
&l(u^*_{T+1}, \ldots, u^*_{T+h}|u_T) \\
&= l(u_T, u^*_{T+1}, \ldots, u^*_{T+h-1}|u^*_{T+h}) \times \frac{l(u^*_{T+h})}{l(u_T)} \\
&= \left\{ l(u_T|u^*_{T+1}, \ldots, u^*_{T+h}) l(u^*_{T+1}|u^*_{T+2}, \ldots, u^*_{T+h}) \ldots l(u^*_{T+h-1}|u^*_{T+h}) \right\} \\
&\qquad \times \frac{l(u^*_{T+h})}{l(u_T)},
\end{aligned}
$$

where l denotes densities associated with the noncausal process u_t. The process u_t is a Markov process of order one, hence the conditioning information set of the densities can be reduced to a single future point. Furthermore, Equation (2.2) states that $\varepsilon_t = u_t - \psi u_{t+1}$, thus, given the value of u_{t+1}, the conditional density of u_t is equivalent to the density of ε_t (which is known) evaluated at the point $u_t - \psi u_{t+1}$.

Since $u_\tau = \psi u_{\tau+1} + \varepsilon_\tau$ and because $u_{\tau+1}$ and ε_τ are independent for all $1 \leq \tau \leq T$, we have $f_{u_\tau|u_{\tau+1}}(x) = f_{\varepsilon_\tau+\psi u_{\tau+1}|u_{\tau+1}}(x) = f_{\varepsilon_\tau|u_{\tau+1}}(x - \psi u_{\tau+1}) = f_\varepsilon(x - \psi u_{\tau+1})$. For simplicity, the density distributions related to u_t (resp. ε_t) are just denoted by l (resp. g). That is, for any assumed error distribution g we have,

$$\begin{aligned} & l(u^*_{T+1}, \ldots, u^*_{T+h}|u_T) \\ & = \left\{ l(u_T|u^*_{T+1}) l(u^*_{T+1}|u^*_{T+2}) \ldots l(u^*_{T+h-1}|u^*_{T+h}) \right\} \times \frac{l(u^*_{T+h})}{l(u_T)} \\ & = g(u_T - \psi u^*_{T+1}) \ldots g(u^*_{T+h-1} - \psi u^*_{T+h}) \times \frac{l(u^*_{T+h})}{l(u_T)}. \end{aligned} \tag{2.4}$$

Whereas with Student's *t*-distributed errors the stationary *pdf* of u_t does not admit a closed-form expression, Gouriéroux and Zakoïan (2013) present closed-form solutions for the predictive density of purely noncausal MAR(0,1) processes when the errors are Cauchy-distributed. We know that

$$u_t = \psi u_{t+1} + \varepsilon_t = \sum_{i=0}^{\infty} \psi^i \varepsilon_{t+i},$$

and they show that the characteristic function of the infinite sum corresponds to that of a Cauchy with scale parameter $\frac{\gamma}{(1-\psi)}$, where γ is the scale of the Cauchy distribution of the errors ε_t. Hence, the purely noncausal component u_t of an MAR(*r*,1) Cauchy process follows a $Cauchy\left(0, \frac{\gamma}{(1-\psi)}\right)$ distribution and its predictive density can be com-

puted as follows,

$$
\begin{aligned}
&l(u^*_{T+1},\dots,u^*_{T+h}|u_T)\\
&=\frac{1}{(\pi\gamma)^h}\left(\frac{1}{1+\frac{(u_T-\psi u^*_{T+1})^2}{\gamma^2}}\cdots\frac{1}{1+\frac{(u^*_{T+h-1}-\psi u^*_{T+h})^2}{\gamma^2}}\right)\\
&\times\frac{\gamma^2+(1-\psi)^2u_T^2}{\gamma^2+(1-\psi)^2(u^*_{T+h})^2}.
\end{aligned}
\tag{2.5}
$$

To illustrate the evolution of the predictive density when the series departs from central values, Figure 2.2 shows one-step ahead forecasts for different levels (0.79, 9.81 and 63.53) corresponding to quantiles 0.55, 0.85 and 0.975 of a purely noncausal process with a lead coefficient of 0.8 and standard Cauchy-distributed errors. By using quantiles, explosive episodes can be compared between different distributions and parameters. While the predictive distribution is uni-modal for median-level values, it splits and becomes bi-modal when the series departs from such values. As the series diverges, the bi-modality of the conditional distribution becomes more evident. The two modes correspond to a drop to 0 and a continuous increase at rate 0.8^{-1} and each event has probability 0.2 and 0.8 respectively. Those results corroborate Fries's (2018) findings in the case of diverging Cauchy-distributed MAR(0,1) series. Note that results are analogous for any lead coefficient ψ, with corresponding probabilities of a crash equal to $1-\psi$. Bi-modality in this chapter will therefore designate the split of the conditional density and not the bi-modality sometimes observed in the estimation of the coefficients of MAR models that are subject to starting values (Hecq, Lieb, and Telg, 2016; Bec, Nielsen, and Saïdi, 2020b))

Figure 2.3 shows the evolution, as the variable increases, of two-step ahead forecasts of the same process, namely a purely noncausal process with lead coefficient 0.8 and Cauchy-distributed errors,

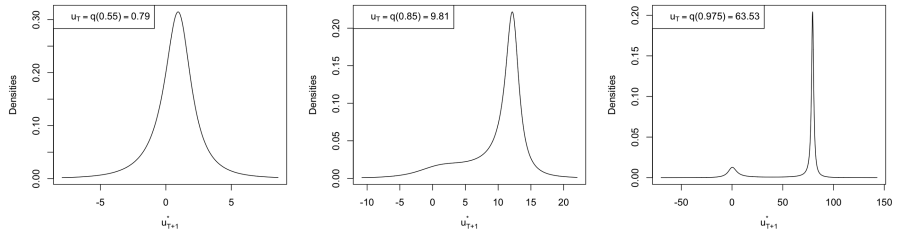


Figure 2.2: Evolution of the 1-step ahead predictive density as the level of the series increases for a Cauchy MAR(0,1) with $\psi = 0.8$.

evaluated at the same points as in Figure 2.2. The graphs depict the joint density of u^*_{T+1} (x-axis) and u^*_{T+2} (y-axis). The bi-modality of the predictive density can also be observed here. For high levels of the series, the split of the distribution is evident; at each step the series can either keep on increasing or drop to zero, and the latter corresponds to an absorbing state. An interpretation of the regions of the graphs with respect to potential future shapes of the series was given by Gouriéroux and Jasiak (2016).

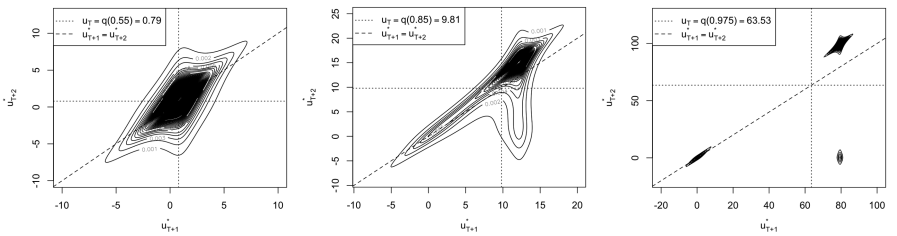


Figure 2.3: Evolution of the 2-step ahead joint predictive density as the level of the series increases for an MAR(0,1) with $\psi = 0.8$ and Cauchy-distributed errors

For MAR(r,1) processes, one-step ahead density forecasts consist in shifting the predictive density of the purely non-causal component by the causal part of the process, namely by $\phi_1 y_T + ... + \phi_r y_{T-r+1}$. For

an h-step ahead forecast, with $h \geq 1$, the predictive density of y^*_{T+h} depends on the joint conditional density of $(u^*_{T+1}, \ldots, u^*_{T+h})$. One way of approaching this is to express the predictive density (2.5) in terms of y^*_{T+k}, with $1 \leq k \leq h$. For instance, for the following MAR(1,1) process,

$$(1 - \phi L)(1 - \psi L^{-1}) y_t = \varepsilon_t,$$

where $\varepsilon_t \sim$ Cauchy$(0, \gamma)$, the joint predictive density of $(y^*_{T+1}, \ldots, y^*_{T+h})$ can be obtained by substituting the noncausal process u_t in (2.5) as follows,

$$\begin{aligned} l(y^*_{T+1}, \ldots, y^*_{T+h} | y_T, y_{T-1}) = \frac{1}{(\pi\gamma)^h} \times \frac{1}{1 + \frac{\left((y_T - \phi y_{T-1}) - \psi(y^*_{T+1} - \phi y_T)\right)^2}{\gamma^2}} \\ \times \ldots \frac{1}{1 + \frac{\left((y^*_{T+h-1} - \phi y^*_{T+h-2}) - \psi(y^*_{T+h} - \phi y^*_{T+h-1})\right)^2}{\gamma^2}} \\ \times \frac{\gamma^2 + (1 - \psi)^2 (y_T - \phi y_{T-1})^2}{\gamma^2 + (1 - \psi)^2 (y^*_{T+h} - \phi y^*_{T+h-1})^2}. \end{aligned}$$

Overall, density predictions yield a more complete forecast as they carry more information regarding potential future patterns of the series than conditional expectations. They cannot be easily graphically displayed for forecast horizons larger than 2 as we investigate joint predictions. Yet, results can be used to compute the probabilities regarding future patterns. Nonetheless, as indicated by Fries (2018) for α-stable distributions, explosive episodes seem to be memoryless and as the series diverges, the probabilities of a crash tend to the constant $1 - \psi^{\alpha h}$ for a given horizon $h \geq 1$. Even though as $h \to \infty$ this probability tends to 1, it may not be very realistic when it comes to real data. We may expect the probabilities of a crash in stock prices for instance to increase with the level of prices. Furthermore, the assumption of other fat-tail distributions (e.g. Student's t) generally leads to the absence of closed-form expressions for the conditional moments and densities. The next section presents two approaches to approximate conditional

densities in such circumstances; the first one is based on simulations (Lanne, Luoto, and Saikkonen, 2012) and the second one uses sample counterparts (Gouriéroux and Jasiak, 2016).

2.4 Predictions using approximation methods

The upcoming subsections first provide the general formulas for *h*-step ahead approximated predictive densities of MAR(*r*,1) processes before focusing on one-step ahead predictions. The two approximation methods are analysed in details for such horizon. We first investigate the case for which theoretical closed-form expressions exist, that is when the errors are Cauchy-distributed. This allows to gauge the approximation of both approaches prior to analysing and comparing them with error distributions for which no closed-form expressions exist. For this, we employ Student's t distributions with 2 and 3 degrees of freedom (respectively with infinite and finite variance). Note that results for processes with higher lead orders can be found in the respective articles of Lanne, Luoto, and Saikkonen (2012) and Gouriéroux and Jasiak (2016).

2.4.1 Predictions using simulations-based approximations

Lanne, Luoto, and Saikkonen (2012) base their methodology on the fact that the noncausal component of the errors, u_t, can be expressed as an infinite sum of future errors, which in the MAR(*r*,1) case is as follows,

$$u_t = \Psi(L^{-1})^{-1}\varepsilon_t = \sum_{i=0}^{\infty} \psi^i \varepsilon_{t+i}.$$

Since stationarity is assumed, and because in applications ψ rarely exceeds 0.9, the sequence (ψ^i) converges rapidly to zero. Hence, they assumed that there exists an integer M large enough so that any future

point of the noncausal component of the errors can be approximated by the following finite sum,

$$u^*_{T+h} \approx \sum_{i=0}^{M-h} \psi^i \varepsilon^*_{T+h+i}, \tag{2.6}$$

for any forecast horizon $h \geq 1$.

As explained before, any point forecast y^*_{T+h}, with $h \geq 1$, of an MAR(r,1) process depends on the sequence forecast $(u^*_{T+1}, \ldots, u^*_{T+h})$. Employing the approximations of (2.6), it is therefore sufficient to forecast a sequence of M future errors $(\varepsilon^*_{T+1}, \ldots, \varepsilon^*_{T+M})$ – which we will denote ε^*_+ – to obtain approximated predictions for the variable of interest. Instead of deriving an M-dimensional conditional joint density function (Lanne, Luoto, and Saikkonen (2012) use $M = 50$), they introduce a way to obtain conditional point and cumulative density forecasts. While the approach they propose to estimate the standard errors of the coefficients requires finite moments for the errors distribution, this restriction is not necessary for their forecasting method (Lanne and Saikkonen, 2011).

Using the companion form of an MAR(r,1) model, y^*_{T+h} can be expressed by recursion as the sum of a known component (at time T) and a combination of h future values of u_t, where the latter, based on (2.6), can be approximated from M future errors,

$$\begin{aligned} y^*_{T+h} &= \iota' \mathbf{\Phi}^h \mathbf{y}_T + \sum_{k=0}^{h-1} \iota' \mathbf{\Phi}^k \iota u^*_{T+h-k} \\ &\approx \iota' \mathbf{\Phi}^h \mathbf{y}_T + \sum_{k=0}^{h-1} \iota' \mathbf{\Phi}^k \iota \sum_{i=0}^{M-h+k} \psi^i \varepsilon^*_{T+h-k+i}, \end{aligned} \tag{2.7}$$

$$\mathbf{y}_T = \begin{bmatrix} y_T \\ y_{T-1} \\ \vdots \\ y_{T-r+1} \end{bmatrix}, \quad \mathbf{\Phi} = \underset{(r \times r)}{\begin{bmatrix} \phi_1 & \phi_2 & \dots & \dots & \phi_r \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}} \quad \text{and} \quad \iota = \underset{(r \times 1)}{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}.$$

Let $g(\varepsilon_+^*|u_T)$ be the conditional joint distribution of the M future errors, which, using Bayes' Theorem can be expressed as follows,

$$g(\varepsilon_+^*|u_T) = \frac{l(u_T|\varepsilon_+^*)}{l(u_T)} g(\varepsilon_+^*).$$

Thus, for any function q,

$$\begin{aligned} \mathbb{E}\Big[q(\varepsilon_+^*)|u_T\Big] &= \int q(\varepsilon_+^*) g(\varepsilon_+^*|u_T) d\varepsilon_+^* \\ &= \frac{1}{l(u_T)} \int q(\varepsilon_+^*) l(u_T|\varepsilon_+^*) g(\varepsilon_+^*) d\varepsilon_+^* \\ &= \frac{\mathbb{E}_{\varepsilon_+^*}\Big[q(\varepsilon_+^*) l(u_T|\varepsilon_+^*)\Big]}{l(u_T)}. \end{aligned} \tag{2.8}$$

Similarly as before, $l(u_T|\varepsilon_+^*)$ can be obtained from the error distribution g. Yet, since it is conditional on ε_+^* instead of u_{T+1}^*, we can only obtain an approximation. Using this approximation and the Iterated Expectation theorem, the marginal distribution of u_T can be approximated as follows,

$$l(u_T) = \mathbb{E}_{\varepsilon_+^*}\left[l(u_T|\varepsilon_+^*)\right] \approx \mathbb{E}_{\varepsilon_+^*}\left[g\left(u_T - \sum_{i=1}^{M} \psi^i \varepsilon_{T+i}^*\right)\right].$$

Overall, by plugging the aforementioned approximation in (2.8), we

obtain

$$\mathbb{E}\Big[q(\varepsilon_+^*)|u_T\Big] \approx \frac{\mathbb{E}_{\varepsilon_+^*}\left[q(\varepsilon_+^*)g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\Big)\right]}{\mathbb{E}_{\varepsilon_+^*}\left[g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^*\Big)\right]}.$$

Let $\varepsilon_+^{*(j)} = \Big(\varepsilon_{T+1}^{*(j)}, \ldots, \varepsilon_{T+M}^{*(j)}\Big)$, with $1 \leq j \leq N$, be the *j*-th simulated series of *M* independent errors, randomly drawn from the assumed distribution of the process. Assuming that the number of simulations N is large enough, the conditional expectation of interest can be approximated as follows,

$$\mathbb{E}\Big[q(\varepsilon_+^*)|u_T\Big] \approx \frac{N^{-1}\sum_{j=1}^{N} q\Big(\varepsilon_+^{*(j)}\Big)g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^{*(j)}\Big)}{N^{-1}\sum_{j=1}^{N} g\Big(u_T - \sum_{i=1}^{M}\psi^i\varepsilon_{T+i}^{*(j)}\Big)}. \tag{2.9}$$

Based on Equation (2.7), for any MAR(*r*,1) process and for any forecast horizon $h \geq 1$, choosing $q(\varepsilon_+^*) = \mathbf{1}(\iota'\mathbf{\Phi}^h\mathbf{y}_T + \sum_{k=0}^{h-1}\iota'\mathbf{\Phi}^k\iota\sum_{i=0}^{M-h+k}\psi^i\varepsilon_{T+h-k+i}^* \leq x)$ in (2.9) provides an approximation of $\mathbb{P}(y_{T+h}^* \leq x|\mathcal{F}_T)$. With this approach, we can therefore obtain directly the cumulative probabilities of y_{T+h}^* and by computing its value for all possible x we can obtain the whole conditional *cdf* of y_{T+h}^*, independent of previous forecasted points,

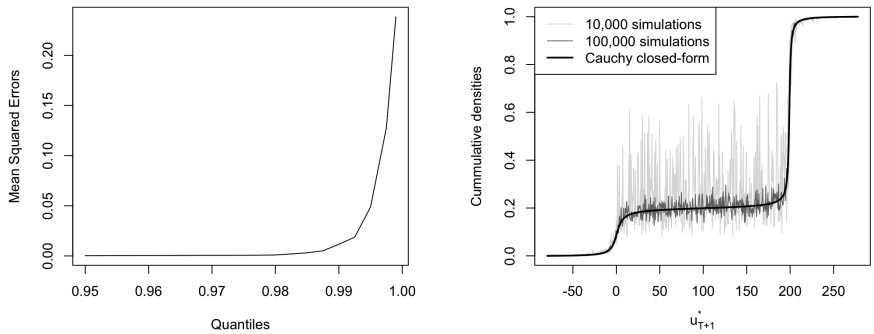
$$\begin{aligned}\mathbb{P}\Big(y_{T+h}^* \leq x|\mathcal{F}_T\Big) &= \mathbb{E}\Big[\mathbf{1}(y_{T+h}^* \leq x)|\mathcal{F}_T\Big] \\ &\approx \mathbb{E}\left[\mathbf{1}\left(\iota'\mathbf{\Phi}^h\mathbf{y}_T + \sum_{k=0}^{h-1}\iota'\mathbf{\Phi}^k\iota\sum_{i=0}^{M-h+k}\psi^i\varepsilon_{T+h-k+i}^* \leq x\right)\middle| u_T, y_T\right].\end{aligned}$$

As an illustration, let us consider again an MAR(0,1) process with a lead coefficient of 0.8 and Cauchy-distributed errors. The complete one-step ahead predictive *cdf* is computed for all potential x and is approximated using the $N = 10\,000$ simulations suggested by Lanne, Luoto, and Saikkonen (2012) at each iteration and we increase the truncation parameter M to 100. It simplifies as follows,

$$\begin{aligned}\mathbb{P}\Big(u^*_{T+h} \leq x|\mathcal{F}_T\Big) &\approx \mathbb{E}\left[\mathbf{1}\left(\sum_{j=0}^{100-1} 0.8^j \varepsilon^*_{T+1+j} \leq x\right)\middle| u_T\right] \\ &\approx \frac{\sum_{j=1}^{10\,000} \mathbf{1}\Big(\sum_{j=0}^{100-1} 0.8^j \varepsilon^{*(j)}_{T+1+j} \leq x\Big) g\Big(u_T - \sum_{i=1}^{100} 0.8^i \varepsilon^{*(j)}_{T+i}\Big)}{\sum_{j=1}^{10\,000} g\Big(u_T - \sum_{i=1}^{100} 0.8^i \varepsilon^{*(j)}_{T+i}\Big)}.\end{aligned} \tag{2.10}$$

The Mean Squared Errors (henceforth MSE) between the approximated and the theoretical cumulative probabilities are presented on graph (a) in Figure 2.4. They are computed for different values of u_T, corresponding to quantiles ranging from $Q(0.95)$ to $Q(0.999)$ of the MAR process. The theoretical *cdf* is computed using closed-form expressions and the approximated cumulative probabilities evaluated using Equation (2.10). They are computed for a large number of points x (as shown in graph (b) of Figure 2.4) and we report the MSE over all points. Namely we report the average squared distance from the theoretical cumulative probabilities to the ones obtained from simulations. The MSEs increase with the level of the series (from 0.0002 to 0.2384 for quantiles 0.95 to 0.999). For illustration, graph (b) in Figure 2.4 compares the probabilities obtained with $10\,000$ and $100\,000$ simulations with the theoretical *cdf* for quantile 0.99. The average discrepancies between the estimated and theoretical cumulative probabilities significantly decrease for larger number of simulations (light grey against dark grey line) and

results converge towards the theoretical distribution (black line). For such process, the quantile 0.99 correspond to a value of u_T of around 160 and we can see that with 10 000 simulations (light grey line), the probability that the process at u_{T+1} will be at most 50 is around 0.6 while it theoretically should be roughly 0.2. In this example, the method overestimates the theoretical probabilities by up to 0.4 with 10 000 simulations, and only by up to 0.05 with 100 000 simulations. Furthermore, it is important to note that the bi-modality of the conditional distribution of an explosive episode is captured by this approach.



(a) Evolution of the MSE of the cumulative probabilities estimations with 10 000 simulations for increasing quantiles.

(b) Estimated cumulative probabilities for Q(0.99) using 10 000 and 100 000 simulations compared to the theoretical *cdf*.

Figure 2.4: Sensitivity of estimations to the number of simulations for an MAR(0,1) with $\psi = 0.8$ and Cauchy distributed errors.

Figure 2.5 depicts the empirical distribution of 1 000 iterations of the same experiment with different number of simulations. Since this chapter focuses on the investigation of turning point of explosive episodes, each iteration consists in computing the probabilities of a decrease of at least 25% when the last observed value is equal to 318.3, corresponding to the quantile 0.995 of the process. The theoretical probability of a crash is equal to 0.2 and as the number

of simulations increases, results converge to this value. For lower number of simulations the same experiment repeated twice may lead to contradicting results. This stems from the fact that we investigate explosive episodes. Recall that bubbles are triggered by a future extreme value in the error terms and if no simulated paths among all simulations can trigger such increase, then the probabilities may be significantly misleading. Conclusions are similar for different lead coefficients (results available upon request), the simulations-based probabilities are a good approximation of theoretical Cauchy-derived probabilities ($(1-\psi)$ during explosive episodes), when the number of simulations is coherently chosen w.r.t. the level of the series.

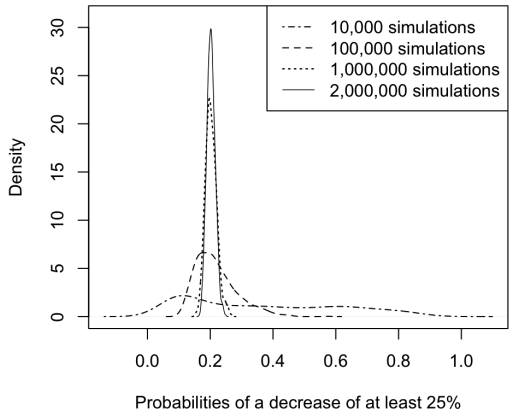


Figure 2.5: Empirical distributions of 1 000 repeated forecasts using different numbers of simulations for an MAR(0,1) process with $\psi = 0.8$ and Cauchy distributed errors evaluated at quantile 0.995.

Previous results help evaluating the ability of the described method to approximate theoretical results, however, analysing its performance for distributions lacking closed-form predictions needs to be performed. We consider Student's $t(2)$ and $t(3)$ distributions and reiterate the same analysis as above. Values corresponding to identical quantiles vary between distributions. While for a

Cauchy ($t(1)$) distributed MAR(0,1) process with $\psi = 0.8$, Q(0.995) corresponds to a value of 318.3, it corresponds to 17.35 and 8.75 for $t(2)$ and $t(3)$ respectively. Note that quantiles of processes with Student's t-distributed errors were empirically estimated. The rate of increase remains identical (here, 0.8^{-1}) across distributions, hence the bi-modality (crash and increase) given a last observed quantile is less evident for larger degrees of freedom, due to the lower corresponding values. That is, probabilities of a crash in such cases will be sensitive to the defined magnitude of the crash.

Figure 2.6 shows the approximated cumulative probabilities computed at last values corresponding to quantile 0.995 for both Student's $t(2)$ and $t(3)$ distributed MAR(0,1) processes, with $\psi = 0.8$ (in analogy to graph (b) of Figure 2.4). We can see that indeed, given a quantile, the bi-modality of the conditional distribution is less apparent for larger degrees of freedom. We also find that approximations are less sensitive to the number of simulations for larger degrees of freedom. This stems from the fact that while the last value of each series correspond to identical quantiles, the values induced by the natural rate of increase of the bubble do not. For instance, 318.3 and 8.75 both correspond to the quantile 0.995 for the Cauchy- and $t(3)$-distributed processes, but $0.8^{-1} \times 318.3$ and $0.8^{-1} \times 8.75$ in turn correspond respectively to quantiles 0.999 and 0.986. The latter is by definition most likely to be drawn in the simulations, implying that for such level of the series the method is less sensitive to the number of simulated paths for larger degrees of freedom. We also observe that as the number of simulations increases, the probabilities converge towards a unique function.

To generalise our first findings, Table 2.1 displays the probabilities of a decrease of at least 25% once a value corresponding to quantile 0.995 is reached for the three distributions and for three distinct lead coefficients. Theoretical probabilities are also reported for Cauchy ($t(1)$) distributed errors obtained with the closed-form expression

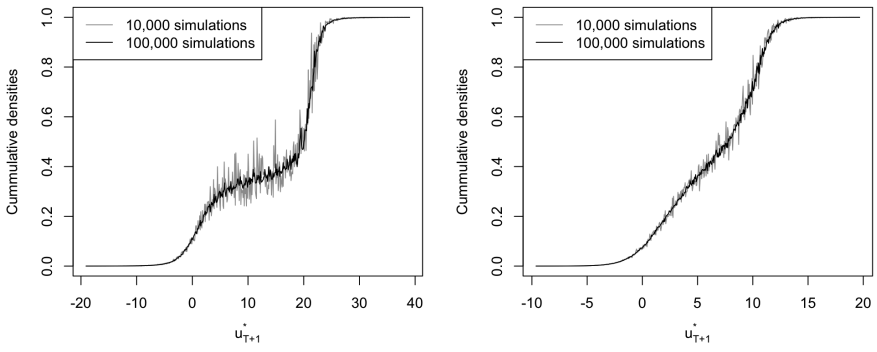


Figure 2.6: Estimated cumulative probabilities evaluated at Q(0.995) using 10 000 and 100 000 simulations for MAR(0,1) processes with $\psi = 0.8$ and with $t(2)$ (left) or $t(3)$ (right) distributed errors.

(2.5). Simulations-based probabilities are obtained with 1 000 000 simulations and we report the average of 1 000 iterations.

Table 2.1: Probabilities of a crash of at least 25% when quantile 0.995 is attained

Lead coefficient	Theoretical	Simulations-based		
ψ	Cauchy/$t(1)$	$t(1)$	$t(2)$	$t(3)$
0.2	.794	.793	.937	.960
0.5	.497	.499	.718	.792
0.8	.201	.203	.358	.435

Reported probabilities for the simulations-based approach are the average over 1 000 forecasts using 1 000 000 simulations.

We find that the theoretical and simulations-based probabilities for $t(1)$-distributed errors do not differ by more than 0.002 on average. Since approximations are much more sensitive to the number of simulations for lower degrees of freedom, as was indicated by Figure 2.6, we can expect the probabilities obtained for the $t(2)$ and $t(3)$ distributions to differ from theoretical probabilities by not more than

0.002 on average, given the attained quantile. The probabilities of a turning point significantly increase with the degrees of freedom of the distribution and with lower lead coefficients. For the investigated quantile, a process with $t(3)$-distributed errors and a lead coefficient of 0.2 has a probability of only 0.04 to keep on increasing as opposed to 0.206 for the Cauchy-distributed process. Furthermore, simulations (results available upon request) indicate that as the series diverges, the probabilities of a crash tend to a constant for all processes. For MAR(0,1) processes with a lead coefficient of 0.8 for instance, the probabilities of a downturn tends to 0.2, 0.36 and 0.48 as the bubble increases for $t(1)$, $t(2)$ and $t(3)$ respectively. Intriguingly, these results are conforming with the crash probabilities (at horizon 1) of explosive MAR(0,1) α-stable processes, $1 - \psi^{\alpha}$, with $\psi = 0.8$ and α equal to the degrees of freedom of the Student's t distribution. This might suggest that Fries's (2018) findings hold for more general error distributions.

Overall, Lanne, Luoto, and Saikkonen's (2012) approach behaves coherently with what was expected from the degrees of freedom, the level of the series and the lead coefficients. This method seems to provide good approximations of theoretical probabilities. With a sufficiently large number of simulations, the probabilities obtained with Cauchy errors converge to the theoretical distribution. While for $t(2)$ and $t(3)$ results cannot be compared to a theoretical benchmark, estimated probabilities also converge to a unique distribution which we can expect to be a proxy for the theoretical conditional cumulative density. We will therefore employ the simulations-based probabilities (given a sufficiently large number of simulations) as proxies for theoretical probabilities when no closed-forms exist. Analogously to theoretical probabilities, the simulated ones also tend to a constant during explosive episodes. That is, after some threshold probabilities of crash remain constant, which as mentioned before does not seem very realistic when we think of prices for instance.

2.4.2 Predictions using sample-based approximations

This section studies the approach proposed by Gouriéroux and Jasiak (2016). Based on past values of the series, this method does not require closed-form expressions for the marginal distribution of the noncausal process u_t. They derive a sample-based estimator of the predictive density (2.4). The marginal distributions of u_t can be expressed as follows,

$$l(u_t) = \mathbb{E}_{u_{t+1}}[l(u_t|u_{t+1})].$$

Again, the noncausal relationship described in Equation (2.2) is used to evaluate the conditional distribution of $l(u_t|u_{t+1})$ with the distribution of the errors, $g(u_t - \psi u_{t+1})$. While Lanne, Luoto, and Saikkonen (2012) employed simulations to approximate expected values, Gouriéroux and Jasiak (2016) use sample-based counterparts. The expected value here is approximated by the average obtained using all points from the sample for the conditioning variable,

$$l(u_\tau) = \mathbb{E}_{u_{\tau+1}}[g(u_\tau - \psi u_{\tau+1})] \approx \frac{1}{T}\sum_{i=1}^{T}\Big\{g(u_\tau - \psi u_i)\Big\}, \tag{2.11}$$

with $\tau = \{T, T+h\}$, necessary to evaluate (2.4). Hence, given the entire sample, the predictive density for the MAR(0,1) process u_t can be approximated by substituting the marginal densities with their sample counterparts (2.11) in (2.4),

$$\begin{aligned} &l(u^*_{T+1}, \ldots, u^*_{T+h}|\mathcal{F}_T) \\ &\approx g(u_T - \psi u^*_{T+1}) \ldots g(u^*_{T+h-1} - \psi u^*_{T+h}) \frac{\sum_{i=2}^{T} g(u^*_{T+h} - \psi u_i)}{\sum_{i=2}^{T} g(u_T - \psi u_i)}. \end{aligned} \tag{2.12}$$

For centred and symmetrical uni-modal distributions, such as the Cauchy and the Student's t that are employed in this analysis, the probability density function is by definition maximised at zero. That is, the densities g, as they are evaluated in Equation (2.12), are maximised at the points where $u_\tau - \psi u_i = 0$ and tend to zero as

the difference widens. Since at time T all observations up to u_T are used in the estimation, the ratio in Equation (2.12) only varies as a function of u^*_{T+h} and will be maximised for paths that were already undertaken. Furthermore, the denominator $\sum_{i=2}^{T} g(u_T - u_i)$ decreases towards zero as u_T diverges from all past values of the sample (i.e. during an explosive episode). This implies that we can expect this approximation method to put more weight on forecast points corresponding to already observed paths and that this tendency will be exacerbated during bubbles.

Analogously to Section 2.3, given a correctly identified model, the predictive density of an MAR(*r*,1) process can be obtained by substituting the filtered noncausal process u_t by the mixed process y_t in (2.12). As an illustration, the *h*-step ahead predictive density of an MAR(1,1) process with undefined non-Gaussian distribution g is obtained using the whole sample up to time *T* as follows,

$$\begin{aligned} l(y^*_{T+1},\dots,y^*_{T+h}|\mathcal{F}_T) \approx\ & g((y_T - \phi y_{T-1}) - \psi(y^*_{T+1} - \phi y_T)) \times \dots \\ & \dots \times g((y^*_{T+h-1} - \phi y^*_{T+h-2}) - \psi(y^*_{T+h} - \phi y^*_{T+h-1})) \\ & \times \frac{\sum_{i=2}^{T} g(y^*_{T+h} - \phi y^*_{T+h-1} - \psi(y_i - \phi y_{i-1}))}{\sum_{i=2}^{T} g(y_T - \phi y_{T-1} - \psi(y_i - \phi y_{i-1}))}. \end{aligned}$$

For comparison purposes with the simulations-based approach, let us again consider an MAR(0,1) process with a lead coefficient of 0.8 and standard Cauchy-distributed errors. For median levels of the series, results are similar between closed-form and sample-based predictions regardless of past behaviours. However, as the series departs from central values, discrepancies emerge and are path-dependent. To illustrate this feature, Figure 2.7 shows one-step ahead density forecasts performed at time $T = 200$ of two different MAR(0,1) trajectories, both ending at 63.53 corresponding to quantile 0.975. Series 1 (left) only has smaller explosive episodes before the one

at which predictions are performed while series 2 already lived a more considerable bubble before. The one-step ahead density predictions are estimated using formula (2.12) and compared to closed-form Cauchy results derived from (2.5) in the bottom graphs. The estimator captures the split of the density of the explosive episode but we can observe discrepancies between the estimated and theoretical densities. For series 1 (left), the estimator overestimates the probabilities corresponding to a crash as all past values are lower than the last observed point. For series 2 (right), it overestimates potential increases, which is explained by the fact that the series already attained such point before and kept on increasing. Predictions with this approach are therefore case-specific and can be characterised by a learning mechanism based on past behaviours. Probabilities can be empirically derived from the obtained predictive densities. For instance, the probabilities of a decrease of at least 25% are theoretically 0.205 for such process but are respectively equal to 0.557 and 0.263 for series 1 and 2. The difference stems from the aforementioned learning mechanism. The choice of the event (e.g. magnitude of the crash) and thus the threshold used to calculate the probabilities may have a considerable impact on the results. Theoretically, for such processes and quantiles, the probabilities of a drop of 75% are only 0.032 lower than for a drop of 25%. This indicates that the arbitrary definition of a crash (e.g. a drop of 25% or of 75%) does not significantly affect the resulting probabilities. However, for series 1 (resp. 2), the probabilities of a drop of at least 75% are 0.11 (resp. 0.05) lower than for a drop of 25%. The learning mechanism can hence induce substantial probabilities for scenarios in between the crash and the further increase.

While the simulations-based approach previously described only depends on parameters estimated and the last observed value, the sample-based method depends also on the entire trajectory and the sample size. Hence, in the intent to generalise results in a similar way as in the previous sections, the parameters considered

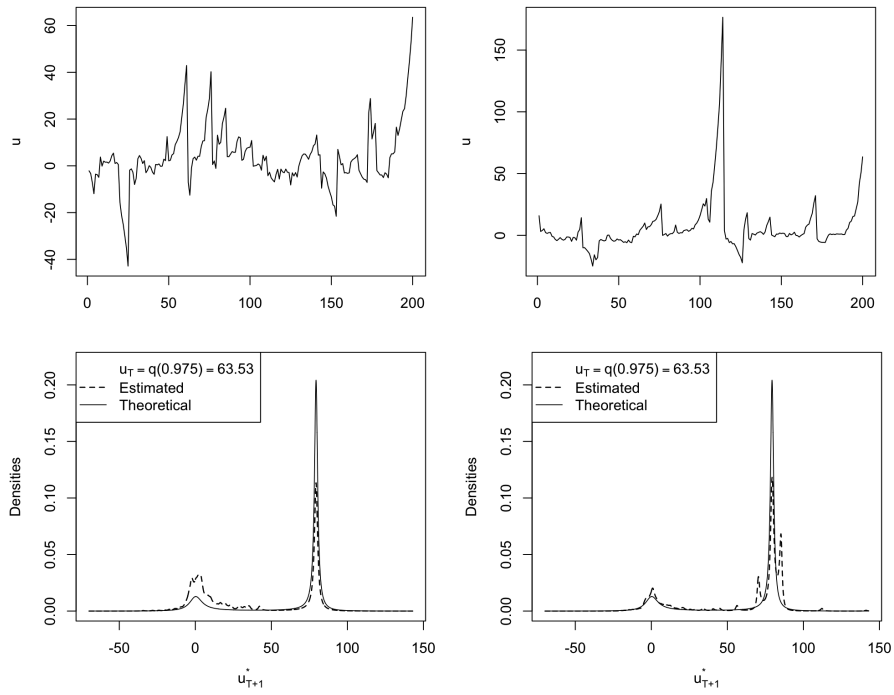


Figure 2.7: Comparison between estimated and theoretical 1-step ahead predictive densities for Cauchy MAR(0,1) with $\psi = 0.8$ evaluated at quantile 0.975 for 2 distinct trajectories.

in Table 2.1 also need to interact with various sample sizes and over multiple trajectories. We first simulate sets of 1 000 different MAR(0,1) trajectories (with Cauchy-distributed errors for the purpose of comparing with theoretical results) ending at quantiles 0.99 or 0.995, with lead coefficients 0.2, 0.5 or 0.8 and sample size 100, 200, 500 or 1000. This results in 24 sets of 1 000 replications and for all of them we compute probabilities of a crash of at least 25%, performed at the last observed point. As an illustration, Figure 2.8 reports the distribution of the 1 000 probabilities obtained from given groups of trajectories, for the two quantiles, the three lead coefficients

(varying from top to bottom) but only two sample sizes (500 in the left column and 1 000 in the right column). Each graph depicts the distributions of probabilities derived from the same processes with identical lead coefficient and sample size, and ending at one of the two considered quantiles (Q(0.99) represented by the black line and Q(0.995) by the filled grey density function). The vertical dashed line represents the theoretical probabilities (obtained from closed-form expressions). Overall, the probabilities diverge more from one trajectory to another for: *(i)* larger lead coefficients (lower graphs compared to top ones), *(ii)* larger sample sizes (right column compared to left) and *(iii)* low quantiles (black line compared to grey area). Larger lead coefficients imply longer-lasting bubbles and therefore more extreme points over the whole sample and enlarging the sample size increases the overall number of extreme episodes. A lower quantile implies larger occurrence of the given value in the series. Hence, as explained above, previous bubbles and thus the frequency of observing similar values in the sample significantly affect probabilities obtained with this approach. Based on its learning mechanism, if a similar level has already been attained in the past, this method will indicate large probabilities that the series will keep on increasing, as it already happened before. That is, for two distinct replications of the same process, ending at the exact same point, one trajectory may have probability zero of crashing while the other may have a probability of 0.6 for instance. However, probabilities are upper bounded, where the upper bound is also the most recurrent probability (main mode of the distribution). We can see that for lower lead coefficient, this mode is closer to the theoretical probability. Hence, the larger the lead coefficient, the more the sample-based method tends to overestimate probabilities of a crash compared to theoretical probabilities. This is also for larger lead coefficients that the probabilities between trajectories vary the most (fatter left tail).

Table 2.2 summarises the same results as Table 2.1 does for the simulations-based approach. However, due to the case-specificity

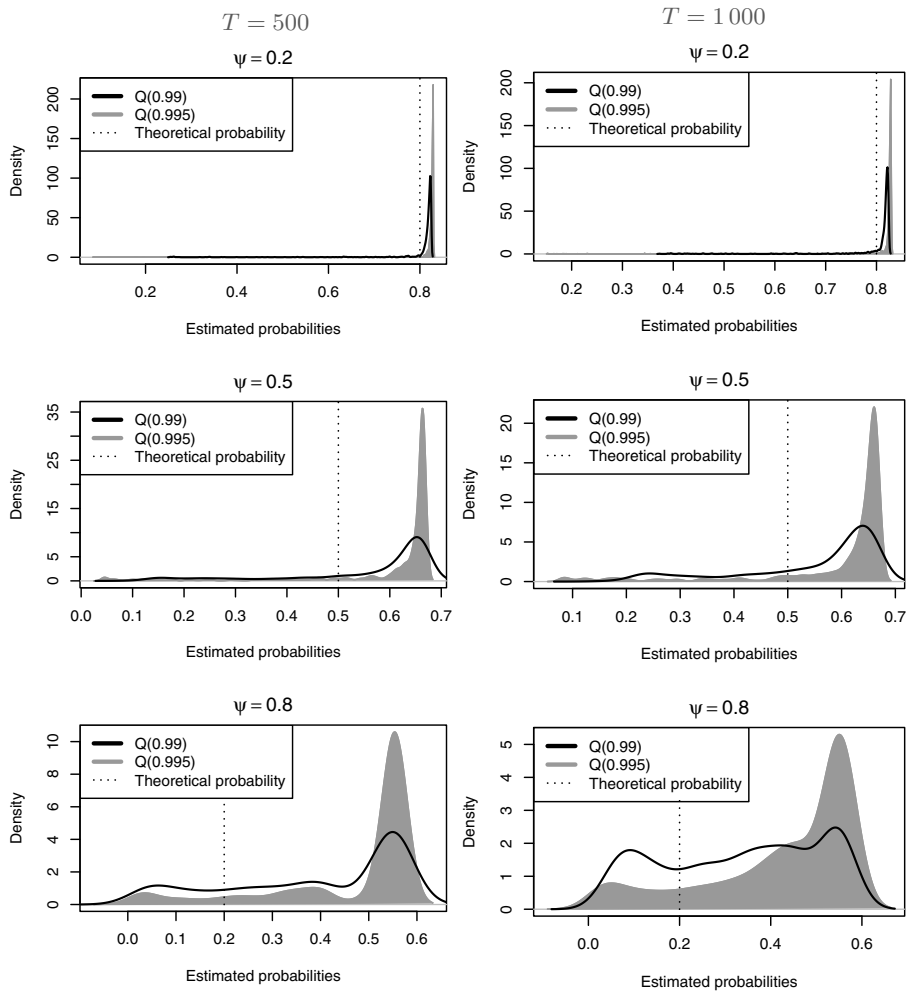


Figure 2.8: Distributions of estimated probabilities of a crash of at least 25% from 1 000 different trajectories evaluated at two different quantiles with distinct lead coefficients and sample sizes.

of the sample-based method, results must be presented for various sample sizes (100, 200, 500 and 1 000) and be summarised over multiple trajectories of the same process. We report the main mode and the first quartile of the distribution of the 1 000 probabilities obtained from different replications of the same process, ending at the same point. The first quartile indicates how much probabilities vary from one trajectory to another and the mode indicates the upper-bound and by definition the most recurrent probability over the 1 000 replications. Analogously to Table 2.1 we consider Student's $t(1)$-, $t(2)$- and $t(3)$-distributed processes. Probabilities are compared to theoretical probabilities (for $t(2)$ and $t(3)$, the results obtained in Table 2.1 are taken as proxies for the theoretical probabilities). By comparing the theoretical probabilities with the column 'Mode' (the most frequently observed probabilities over the 1 000 replications) we notice that the tendency of the sample-based method to overestimate the probabilities of a crash is lower for larger degrees of freedom in the errors distribution given a lead coefficient. For a lead coefficient of 0.2 and a Cauchy-distributed process for instance, the mode is 0.034 above theoretical probabilities while it is at most 0.002 above for a $t(3)$-distributed process. Conclusions drawn from the illustrative examples of Cauchy-distributed processes in Figure 2.8 can be generalised for the three distributions, namely that an increase in the sample size and in the lead coefficient lead to significantly more varying results, even though those discrepancies overall decrease with larger degrees of freedom. Nevertheless, due to the dependence on past points and the case-specificity of this approach, it is rather challenging to demonstrate theoretical guarantees or convergence of this approximation method during explosive episodes.

The focus of this chapter is on one-step ahead forecasts but farther predictions are possible, however increasingly computationally demanding. Gouriéroux and Jasiak (2016) propose a method to tackle this issue by elaborating a Sampling Importance Resampling (SIR) algorithm. The algorithm aims at recovering a predictive

Table 2.2: Sample-based probabilities of a crash of at least 25% evaluated at Q(0.995) for 1 000 replications of each model

ψ	Sample size	$t(1)$			$t(2)$			$t(3)$		
		Theor.	1st Q.	Mode	Simul.	1st Q.	Mode	Simul.	1st Q.	Mode
0.2	100	.794	.828	.828	.937	.941	.941	.960	.961	.962
	200		.828	.828		.941	.941		.961	.961
	500		.825	.828		.941	.941		.961	.961
	1000		.824	.828		.940	.941		.961	.961
0.5	100	.497	.664	.665	.718	.772	.776	.792	.808	.819
	200		.657	.665		.756	.776		.802	.818
	500		.629	.665		.738	.775		.795	.819
	1000		.611	.665		.714	.775		.789	.815
0.8	100	.201	.555	.556	.358	.544	.597	.435	.428	.606
	200		.553	.556		.458	.603		.391	.590
	500		.404	.556		.259	.607		.369	.491
	1000		.343	.556		.290	.605		.379	.416

Theor. corresponds to theoretical probabilities (Simul. correspond to the simulations-based probabilities reported in Table 2.1 and used as proxies for theoretical probabilities). Are also reported the 1st quantile and the mode of the distribution of the 1 000 probabilities of each scenario.

density based on simulations from a misspecified instrumental model from which it is easier to simulate. They suggest using a Gaussian AR model of order s (here an $AR(1)$) to simulate the process u_t. Results are available upon request but overall this approach recovers the intended densities for median levels of the series but fails to recover both the parts corresponding to the crash and to the increase during explosive episodes. The failure of the algorithm for high levels of the series stems from the intention to recover a bi-modal distribution from a uni-modal distribution. If the variance of the uni-modal instrumental distribution is not large enough to cover both modes of the sample-based density, the algorithm will not be able to recover the whole conditional

distribution. The shape of the Normal distribution significantly depends on past behaviours of the series since the variance is estimated as the variance of the residuals of the MAR model. Hence, for more volatile series, the variance of the instrumental Normal distribution will be larger, yet, as the variable increases and the two modes diverge, there will always be a point from which the SIR algorithm does not succeed in recovering the density anymore. While the algorithm is not needed for one-step ahead predictions when computations are not too demanding, the conclusions drawn are the same for larger forecasts horizons of explosive episodes, for which the algorithm may miss parts of the joint conditional density.

Gouriéroux, Hencic, and Jasiak (2018) find that the quality of forecasts diminishes when the series follows an explosive episode. Indeed, approximations errors amplify with the level of the series, and there is a point from which the SIR algorithm does not recover the whole density anymore. Yet, we find that the sample-based estimator captures the split of the conditional density as the series departs from central values and comprises both the crash and increase parts of the predictive density. Furthermore, it yields time varying probabilities based on its learning mechanism. While sample-based predictive densities based on Student's t-distributions cannot be compared to closed-form predictions, results corroborate the conclusions drawn with Cauchy. Thinner tails in the errors distribution lead to higher probabilities of crash for given quantiles of an MAR process. A limitation is that when closed-form results are not available, we cannot disentangle how much of the derived probabilities are induced by the underlying distribution and how much by past behaviours. To tackle this, the probabilities estimated with the simulations-based approach of Lanne, Luoto, and Saikkonen (2012) can be used as benchmark as they seem to be good approximations of theoretical probabilities. Such data-driven approach alleviates the issue of constant probabilities that theory or the simulations-based method suggests during explosive episodes. Yet, this is at the costs of heavy

computations that increase with the forecast horizon.

2.5 Empirical analysis

We now empirically analyse the two approaches presented in Section 2.4. Karapanagiotidis (2014) and Lof and Nyberg (2017) find evidence that non-causal models generally provide better fits for commodity price series. We hence forecast the bubble pattern in commodity prices and in particular in the monthly Global price of Nickel. The series is obtained from the International Monetary Fund and spans the period from January 1980 to September 2019. There seems to be a positive trend in the data but making the series stationary is far from obvious. Indeed, usual unit root tests do not perform well for this type of variable with very large spikes. For instance ADF tests would reject the null of a unit root against both a mean and a trend reverting alternative. A conclusion that does not seem satisfactory from the graphs of the data. However, transformations of the data may affect the dynamics of the process, like a X-11 seasonal filter does for instance (see Hecq, Telg, and Lieb, 2017). It might also well be that the series is stationary around a shift in mean. Hencic and Gouriéroux (2015) use a cubic deterministic trend to isolate the bubble in the Bitcoin. Hecq and Voisin (2022) find that underestimating the order of a polynomial trend can have substantial impacts on the dynamics of the series and show that the Hodrick-Prescott (hereafter HP) filtering can be used as an alternative that does not require any assumption regarding the order of the trend. They find that with the penalising parameter suggested by Ravn and Uhlig (2002), i.e. $\lambda = 129\,600$ for monthly time series, the filtering preserves the bubble features of the data and thus provides a stationary time series with locally explosive episodes (which would disappear when taking the returns).

The HP-detrended series is reported in Figure 2.9. We first estimate

an autoregressive model by OLS on the whole HP-detrended Nickel price series. Information criteria (AIC, BIC and HQ) all advocate a pseudo lag length of $p = 2$. The three possible MAR(*r*,*s*) specifications are consequently an MAR(2,0), an MAR(1,1) or an MAR(0,2). Using the MARX package of Hecq, Lieb, and Telg (2017a) an MAR(1,1) with a *t*-distribution with a degree of freedom of 1.45 and a scale parameter of 390 is favoured. The value of the causal and the noncausal parameters are respectively 0.66 and 0.73. We are consequently in the situation in which the predictive density does not admit closed-form expressions (although not very far from the Cauchy) but the sample- and simulations-based approaches can be used.

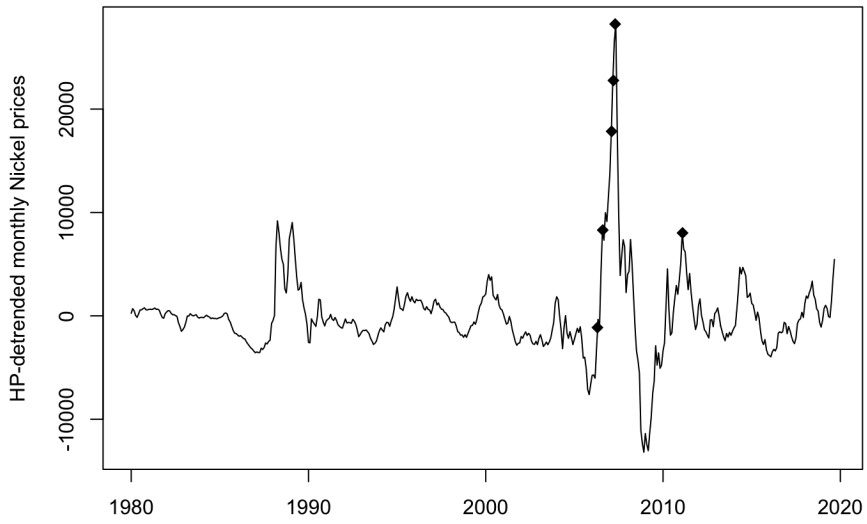


Figure 2.9: HP-detrended monthly Nickel prices series and the points at which forecasts are performed.

We aim attention at the main explosive episode, which crashed in June 2007. To investigate the evolution of the predicted probabilities along the bubble with settings as close as possible to the assumptions

made throughout this chapter, we assume that the model is correctly identified (parameters estimated over the whole sample) at each point of interest and hence perform in-sample forecasts. The points at which we perform predictions are represented by the diamonds on the trajectory in Figure 2.9. We investigate five points along the main explosive episode and one after to capture the effects of the inclusion of the crash in the predictions. Each point is assigned an index between 1 and 6 indicating their order of arrival. At each point, we compute the one-step ahead sample-based predictive density and compute probabilities for three different magnitudes of crash derived from both the sample- and simulations-based approaches. Since simulations-based estimations are good approximations (with a large enough number of simulations) of theoretical results, we consider them as a proxy for theoretical probabilities to which sample-based probabilities are compared.

Results are reported in Table 2.3. The quantiles corresponding to each of the six points were empirically evaluated, based on the estimated model, and are presented in the second column. The whole sample up to the points of interest was used to approximate the conditional density function in the sample-based approach. For the simulations-based method, given the degrees of freedom estimated for the errors distribution and the quantiles to be investigated, 5 000 000 simulations were employed at each iteration, to ensure the accuracy of the results. We report the probabilities obtained with the sample-based approach (columns 'samp.'), the simulations-based approach ('sims.') as well as the difference between the two resulting probabilities ('diff.'). We investigate the probabilities of three magnitudes of decreases, up to 40% ('$< 60\%y_T$'). We do not consider larger drops since with a lag coefficient of 0.66, the left mode of the conditional distribution is located at 66% of the last observed value and the probabilities of larger drops quickly decay to zero. This can be seen in the column $< 60\%y_T$, the probabilities are already almost all below 0.10 for points along the bubble. Hence, let us focus on

Table 2.3: One-step ahead probabilities of events for detrended monthly Nickel prices at six different point in time

Point	Quantile	$< y_T$			$< 80\%y_T$			$< 60\%y_T$		
		samp.	sims.	diff.	samp.	sims.	diff.	samp.	sims.	diff.
1	.376	.744	.771	-.028	.842	.844	-.003	.904	.886	.018
2	.951	.536	.328	.207	.459	.247	.212	.128	.075	.053
3	.985	.600	.360	.239	.533	.314	.219	.059	.048	.011
4	.989	.603	.365	.238	.548	.331	.217	.040	.039	.001
5	.993	.616	.380	.236	.541	.343	.198	.024	.028	-.004
6	.949	.331	.323	.080	.263	.240	.023	.078	.070	.008

The quantiles were evaluated with simulations based on the estimated model. Samp. (resp. sims.) represents sample-based (resp. simulations-based) probabilities. For the simulations-based approach the following settings were employed: $M = 100$ and $N = 5\,000\,000$. Results with this approach are assumed to be proxies for the theoretical ones.

the first two blocks of results, namely probabilities of a decrease and probabilities of a crash of at least 20%.

Points 1 to 5 capture the evolution from the outset to the peak of the bubble. During this episode the series departs from below median values (Q(0.376)) to reach a value corresponding to a quantile of 0.993 at the top of the bubble. Columns 'diff.' indicate that, except for point 1 (for which both methods yield roughly the same results), the sample-based approach always overestimates probabilities compared to the ones obtained with the simulations. Once the series enters the explosive episode, the discrepancies between the two approaches widen. For probabilities of a decrease for instance ('$< y_t$'), the differences of the probabilities between the two methods increase until the series has exceeded all past points (remains almost constant from point 3 to 5). The discrepancies for the probabilities of larger crashes vary more due to the significant probabilities assigned by the sample-based approach to points between the crash and the natural rate of increase. Overall, the results are in line with Table 2.2 from

Section 2.4 for such lead coefficient, sample size and t distribution, for which the sample-based methods usually overestimates probabilities of a crash. The difference between the probabilities of the two methods represents how much of the sample-based probabilities are induced by its learning mechanism. This indicates that once all past values are exceeded, the uncertainty added to the theoretical probabilities remains more or less constant. Not only does the difference between the probabilities of the methods remains constant after the series has exceeded all past values, but also the probabilities of each method themselves. We can see that after point 3, probabilities do not increase by more than 0.02. This suggests that both theoretical and learnt probabilities would probably remain constant along the bubble if it were to increase more after point 5. Additionally, note that the simulated probabilities of a downturn (column '$< y_T$') from points 3 to 5 are very close to the probabilities of a crash of an explosive α-stable MAR(0,1) process, $1 - \psi^\alpha$, with $\psi = 0.73$ and $\alpha = 1.45$, suggesting again the applicability of Fries's (2018) results to more general errors distributions. Furthermore, while probabilities indicate almost the same probabilities for the series to drop by 5% than to drop by 20%, they suggest that the crash would most likely not be larger than 40%; and indeed, after point 5, when the bubble burst, the series dropped by 37%.

Point 6 was chosen to illustrate the learning mechanism of the sample-based method and the value at point 6 corresponds to a quantile slightly lower than point 2. Hence, as expected, since simulations-based probabilities are not affected by past values, probabilities of a decrease are slightly lower as well. However, as the main explosive episode is now included in the sample-based predictions, the learning mechanism suggests less risk of a crash as the series already entered a bubble after reaching this point before. This leads to probabilities of a decrease 0.205 lower than at point 2 and 0.196 lower for a crash of at least 25%. The discrepancies between the sample- and simulations-based probabilities are now less than

0.08 for the two magnitudes.

To wrap up, the use of both approaches when the distribution of the errors does not allow for closed-form expressions can help disentangle how much probabilities in the sample-based approach are induced by past behaviours. Indeed, even if probabilities of a turning point were continuously increasing with the level of the variable, the discrepancy between the probabilities of the two approaches, capturing the variation in uncertainty regarding the downturn may remain constant after some point as we have seen in this empirical example. The two approaches carry different information; on the one hand, the sample-based approach relies on past behaviours and is usually yielding higher probabilities of turning points. On the other hand the simulations-based approach yields probabilities solely induced by the underlying model. Yet, both methods capture the bi-modality of the conditional density during explosive episodes, which indicates the outset of a bubble. The two approaches can be used individually or combined, based on preferences and on beliefs regarding the process.

2.6 Conclusion

This chapter analyses and compares in details two approximation methods developed to forecast mixed causal-noncausal autoregressive (MAR) processes. MAR models can be employed in a parsimonious setting to model nonlinear dynamics that would otherwise require much more complex modelling. This chapter aims attention at predictive densities rather than point forecasts as they are more informative, especially in the case of explosive episodes, and focuses on one-step ahead probabilities of turning point of MAR(r,1) processes.

The simulations-based (Lanne, Luoto, and Saikkonen, 2012) and sample-based (Gouriéroux and Jasiak, 2016) methods are first compared to theoretical results using various MAR processes with Cauchy-distributed errors to have a benchmark of their respective performance. The two approximation approaches are then applied to Student's t-distributed processes, with 2 and 3 degrees of freedom (cases for which no closed-form expression exist). We find that simulations-based predictive probabilities are a good approximation of theoretical results obtained with Cauchy-distributed errors, given a large enough number of simulations in the approximations. We show that with $t(2)$- and $t(3)$-distributed errors, the predictive cumulative probabilities also converge (as the number of simulations increases) to a unique distribution. Probabilities of a downturn during explosive episodes are conforming with Fries's (2018) findings for α-stable distribution which might indicate their applicability to more general error distributions. We therefore suggest employing the simulations-based probabilities as proxy for theoretical probabilities, given a sufficient number of simulations. We further find that sample-based predictive densities start to deviate from both the closed-form (when available) and the simulations-based densities when the series depart from its central values. The discrepancies tend to increase with the lead coefficient, the sample size and for low degrees of freedom of the Student's t distribution. The sample-based approach gives time-varying probabilities and depends on how similar the event under investigation is to past events; it is therefore case specific and is characterised by a learning mechanism. This approach yields results that are a mixture of probabilities ensuing from the underlying distribution and probabilities learnt from past behaviours of the series. Note that both methods capture the bi-modality of the conditional distribution as the series diverges from central values, which is an indicator of a potential bubble outset and which was expected from Fries's (2018) results.

We illustrate the two methods with detrended monthly Nickel price

series. When the underlying distribution does not admit closed-form expressions for the predictive densities, the only way to disentangle the theoretical probabilities from the learnt probabilities is to employ the simulations-based probabilities as proxy for the theoretical ones. The approximation prediction methods may lack theoretical grounds but provide valuable information based on the estimated model and on past behaviours of the series in a parsimonious way. The information provided by both approaches could for instance be employed to construct investment strategies but further research should be carried with respect to the design and the performance of such practice with a comparison to existing methods.

3

Predicting crashes in oil prices during the COVID-19 pandemic with mixed causal-noncausal models

Adapted from: Alain Hecq and Elisa Voisin (2022). "Predicting bubble bursts in oil prices during the COVID-19 pandemic with mixed causal-noncausal models". In: *Advances in Econometrics in honor of Joon Y. Park*. Forthcoming.

Abstract

This chapter aims at shedding light upon how transforming or detrending a series can substantially impact predictions of mixed causal-noncausal (MAR) models, namely dynamic processes that depend not only on their lags but also on their leads. MAR models have been successfully implemented on commodity prices as they allow to generate nonlinear features such as locally explosive episodes (denoted here as bubbles) in a strictly stationary setting. We consider multiple detrending methods and investigate, using Monte Carlo simulations, to what extent they preserve the bubble patterns observed in the raw data. MAR models relies on the dynamics observed in the series alone and does not require economical background to construct a structural model, which can sometimes be intricate to specify or which may lack parsimony. We investigate oil prices and estimate probabilities of crashes before and during the first 2020 wave of the COVID-19 pandemic. We consider three different mechanical detrending methods and compare them to a detrending performed using the level of strategic petroleum reserves.

3.1 Introduction

This chapter aims at forecasting Brent and WTI oil price series during the first wave of the COVID-19 pandemic outbreak in 2020 using the recent literature on mixed causal-noncausal autoregressive models (hereafter MAR). Namely, time series processes with lags but also leads components and non-Gaussian errors. This new specification can, in a parsimonious way, model locally explosive episodes in a strictly stationary setting. It can therefore capture nonlinear features such as bubbles (which is defined here as a persistent increase followed by a sudden crash), often observed in commodities prices, while standard linear autoregressive models (e.g. ARMA models) cannot do so. MAR models have successfully been implemented on several commodity price series (see inter alia Hecq and Voisin, 2021; Hecq, Issler, and Telg, 2020; Fries and Zakoïan, 2019a; Gouriéroux and Zakoïan, 2017; Cubadda, Hecq, and Telg, 2019; Lof and Nyberg, 2017; Karapanagiotidis, 2014).[1] Similarly to Gouriéroux and Zakoïan (2013), our goal when introducing a lead component in oil prices is not to provide an economic justification for the existence of a rational bubble. However, the link with a present value model between prices and dividends (Campbell and Shiller, 1987) can enrich the discussion and it also explains the difficulties to find economic fundamentals for oil prices. This motivates our choice to use proxies such as technical methods to extract the bubble component. Let us indeed consider a general model (see Diba and Grossman, 1988) in which the real current stock price P_t is linked to the present value of next period's expected stock price

[1] An alternative strategy to ours is to consider autoregressive processes with breaks in coefficients. Indeed, autoregressive processes with successively unit roots, explosive and stable stationary episodes are also able to capture locally explosive episodes. See among many others Phillips, Wu, and Yu (2011) and the survey papers by Homm and Breitung (2012) or Bertelsen (2019). Yet, for the purpose of forecasting, we argue for the choice of a model with constant coefficients as more adequate.

P_{t+1}, dividend payments D_{t+1} and an unobserved variable u_{t+1},

$$P_t = \frac{1}{1+r}\mathbb{E}_t[P_{t+1} + \alpha D_{t+1} + u_{t+1}], \tag{3.1}$$

with $\mathbb{E}_t$ the conditional expectation given the information set known at time *t*. The discount factor is $\frac{1}{1+r}$ with r being a time-invariant interest rate. The general solution of (3.1) is (e.g. Diba and Grossman, 1988)

$$P_t \;=\; \sum_{i=1}^{\infty}\left(\frac{1}{1+r}\right)^i \mathbb{E}_t[\alpha D_{t+i} + u_{t+i}] + B_t \;=\; P_t^F + B_t, \tag{3.2}$$

where the actual price deviates from its fundamental value P_t^F by the amount of the rational bubble B_t. As shown by Gourieroux, Jasiak, and Monfort (2020), MAR processes provide stationary solutions for the modeling of the bubbles component in (3.2) (see also Fries, 2021).

However, oil prices are challenging time series to forecast and model (see Baumeister and Kilian,2016 and for a survey on oil prices forecasting see Alquist, Kilian, and Vigfusson, 2013). Unlike for equity prices, measuring commodities fundamentals might not be as straightforward (Brooks, Prokopczuk, and Wu, 2015). Pindyck (1993) and Alquist and Kilian (2010) consider the convenience yield, that is, a premium associated with holding an underlying product instead of derivative securities or contracts. It typically increases when costs associated with physical storage are low. Yet, not only is the convenience yield not easy to measure but there also are other factors driving each of the demand and supply side of crude oil: the level of stocks, economic activity, geopolitical considerations, shifts in expectations regarding the oil market, etc. While there is a large literature on modeling and forecasting the price of oil using structural models that incorporate economic fundamentals (see Kilian and Zhou, 2020b), our model is parsimonious and exploits the statistical properties of oil prices only.

As can be seen in Figure 3.3 in Section 3.4, oil prices series do not appear to be stationary over time. Consequently, before estimating MAR models we intend to extract a smooth time-varying trend to render the series stationary without affecting the dynamics. By extracting a trend from the series we do not claim to identify the fundamental values of oil prices but instead detrend the series while preserving the dynamics of the prices in the remaining cycle and more specifically the noncausal component. As such, we obtain stationary series that retain their forward-looking aspect and which can be modeled as MAR processes. Obviously, a wrong detrending can give misleading results if it alters the dynamics of the cycle. Consequently, investigating the impact of different technical detrending filters on the identification of MAR models is the first contribution of this paper. Similarly to what Canova (1998) does for business cycles, we investigate the extent to which the identification of causal and noncausal dynamics are sensitive to different filters. We then study the consequences on the predictive densities of oil prices after applying different detrending methods. Inspired by the work of Kilian and Murphy (2014), who constructed a structural VAR model of the global market for crude oil, we make use of US crude oil Strategic Petroleum Reserve (SPR), a sub-part of total petroleum stocks, for a potential trend in oil prices in Section 3.4. Hence, the second contribution of this chapter is to compare the MAR estimations and predictions of oil price series after using technical detrending with the results obtained after detrending with the SPR levels.

The rest of this chapter is as follows. Section 3.2 describes mixed causal-noncausal models and explains the different technical detrending methods employed in this analysis, leaving the locally explosive components in the cycle. In Section 3.3, the impact of the different detrending filters on model identifications is investigated using a Monte Carlo study, based on trends estimated in oil prices series. We investigate the identification of the models but also the

magnitude of the coefficients estimated as they are the main drivers of the predictions. Section 3.4 analyses the impact of these filters on the WTI and the Brent crude oil price series for ex-post and real-time analyses. We compare the results with those obtained after detrending with US SPR levels. We show how each detrending approach affects probabilities that oil price crashes in the period capturing the first 2020 wave of the COVID-19 pandemic. Section 3.5 concludes.

3.2 Mixed causal-noncausal models and filtering

3.2.1 The model

MAR(r, s) denotes dynamic processes that depend on their r lags as for usual autoregressive processes but also on their s leads in the following multiplicative form

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t, \tag{3.3}$$

with L the backward operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. When $\Psi(L^{-1}) = (1 - \psi_1 L^{-1} - ... - \psi s L^{-s}) = 1$, namely when $\psi_1 = ... = \psi_s = 0$, the univariate process y_t is a purely causal autoregressive process, denoted MAR(r,0) or simply $AR(r)$ model, $\Phi(L)y_t = \varepsilon_t$. Reciprocally, the process is a purely noncausal MAR$(0, s)$ model $\Psi(L^{-1})y_t = \varepsilon_t$, when $\phi_1 = ... = \phi_r = 0$ in $\Phi(L) = (1 - \phi_1 L - ... - \phi_r L^r)$. The roots of both the causal and noncausal polynomials are assumed to lie outside the unit circle, that is $\Phi(z) = 0$ and $\Psi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series y_t admits a two-sided moving average representation $y_t = \sum_{j=-\infty}^{\infty} \gamma_j \varepsilon_{t-j}$, such that $\gamma_j = 0$ for all $j < 0$ implies a purely causal process y_t (with respect to ε_t) and a purely noncausal model when $\gamma_j = 0$ for all $j > 0$ (Lanne and Saikkonen, 2011). Error terms ε_t are assumed iid (and not only weak white noise) non-Gaussian (with potentially infinite variance) to ensure the identifiability of the causal and the noncausal parts (Breidt et al., 1991; Gouriéroux and Zakoïan, 2015). While noncausal models are strictly

stationary, their conditional moments are time-varying. A purely stationary noncausal MAR(0,1) Cauchy-distributed process, has a unit root in its conditional mean and exhibit ARCH-type effects (see Gouriéroux and Zakoïan, 2017; Cavaliere, Nielsen, and Rahbek, 2018).

Figure 3.1 shows a purely causal (a) and a purely noncausal (b) trajectories induced by the same Student's $t(2)$-distributed errors, both with coefficient 0.8 and 200 observations. For the purely causal process, a shock is unforeseeable and affects the series only once it happened, inducing a large jump in the series. On the other hand, for purely noncausal processes, a shock impacts the process ahead of time, mirroring the purely causal trajectory. Indeed, we see that the series already reacts to a positive shock by increasing until a sudden crash, creating bubble patterns. This anticipative aspect is widely observed in financial and economics time series. The detrended Brent crude oil prices as shown in Figure 3.5 noticeably exhibit such features, the most apparent episode being the 2008 financial crisis. A combination of causal and noncausal dynamics consequently creates some asymmetry around a shock, varying with the magnitude of the respective coefficients.

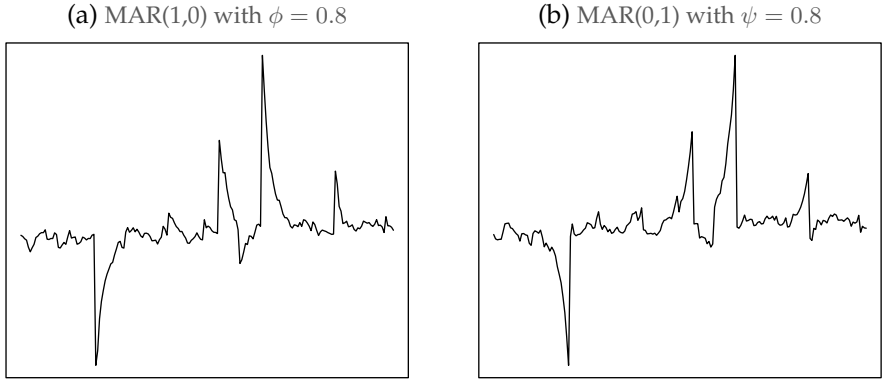


Figure 3.1: Purely causal (a) and noncausal (b) trajectories

The advantage with oil prices is that they already underwent bubbles in the past, and those previous locally explosive episodes will help identifying MAR models. In the case where series are for the first time following a long and abnormal increase, an explosive process is difficult to distinguish from a stationary locally explosive one.

The focus of this chapter is on the probabilities of crashes. Predictions are performed using the approximation methods of Gouriéroux and Jasiak (2016) and Lanne, Luoto, and Saikkonen (2012) since no closed-form of the predictive density exists when the errors of the process follow a Student't distribution. For a detailed analysis of the two approximation methods see Hecq and Voisin (2021).[2]

3.2.2 Filtering the data

The requirement of y_t being stationarity for both lag and lead polynomials gave rise to different strategies to transform nonstationary series to stationary ones. Hecq, Issler, and Telg (2020) and Cubadda, Hecq, and Telg (2019) assume[3] that their commodity price series are $I(1)$ and work with the returns Δy_t. However, this operation eliminates most of the locally explosive behaviors and the transformed series consist of many spikes instead.

In this paper, we capture the trending behavior of the observed series denoted $\tilde{y}_t$ in different ways using the general form

$$\tilde{y}_t = f_t + y_t,$$

where

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t.$$

[2] A description of the methods used in this analysis can be found in Section 2.4.
[3] The locally explosive features of the data make unit root tests doubtful.

In this framework, $\tilde{y}_t$ is the (potentially nonstationary) observed series and f_t a generic trend function. The deviation of $\tilde{y}_t$ from its trend is an MAR(r, s) process. Several authors, although sometimes not explicitly, use this decomposition. Cavaliere, Nielsen, and Rahbek (2018) opt for the choice of a particular time period with no trend and hence use only an intercept $f_t = \mu$. Hencic and Gouriéroux (2015) detrend $\tilde{y}_t$ using a polynomial trend function of order three. In summary, we could consider several choices among the following deterministic trends,

$$
\begin{aligned}
f_t^{(1)} &= \mu, \\
f_t^{(2)} &= \mu + \beta D_t, \quad \text{with } D_t = 1 \text{ when } t \geq t_{break} \text{ and } 0 \text{ otherwise}, \\
f_t^{(3)} &= \alpha_0 + \alpha_1 t + ... + \alpha_k t^k, \quad \text{with } k \in \mathbb{Z}^+ \text{ and } t = 1, 2, \ldots, T.
\end{aligned}
$$

Note (see Section 3.4) that since a larger order of polynomial allows for more flexibility, we consider polynomial trends of order four and six for the trending pattern of the monthly oil prices series considered in this analysis. More complex trends, constructed as a combination of the aforementioned examples could also be considered, such as (multiple) breaks in trends for instance.

Hecq and Voisin (2021) use the Hodrick-Prescott filter (HP) before detecting bubbles in Nickel monthly prices. The HP filter, as opposed to the aforementioned deterministic trends, extracts the trend process $f_t^{(4)}$ via a minimisation that relies on a penalising parameter denoted λ.

$$
\min_{\{f_t^{(4)}\}_{t=1}^T} \left\{ \sum_{t=1}^{T} (\tilde{y}_t - f_t^{(4)})^2 + \lambda \sum_{t=3}^{T} \left[(f_t^{(4)} - f_{t-1}^{(4)}) - (f_{t-1}^{(4)} - f_{t-2}^{(4)}) \right]^2 \right\}.
$$

The larger this parameter, the smoother the trend component is (that is, with λ approaching infinity, the extracting trend becomes linear). For details about the HP filter see Hodrick and Prescott (1997). It is now commonly accepted to use $\lambda = 1\,600$ for quarterly data. For other

frequencies, the rule of thumb consists in adjusting the parameter to the frequency relative to quarterly data,

$$\lambda = \left(\frac{\text{number of observations per year}}{4} \right)^{i} \times 1\,600,$$

with either $i = 2$ (Backus and Kehoe, 1992) or $i = 4$ (Ravn and Uhlig, 2002), yielding respectively a penalising parameter of 14 400 and 129 600 for monthly series. Most criticisms of the HP filter concern its application on series with complicated stochastic and deterministic trends. Phillips and Shi (2019) propose an adaptation of the filter improving its accuracy for such series.[4] We investigate in Section 3.3 the potential dynamic distortions that can be induced by HP filtering (see among others Hamilton, 2018) but find no significant distortions of the mixed causal-noncausal dynamics.

Note that we are not interested in the exact value of a forecast but rather in its direction and potential magnitude. This is why we extract smooth trends to preserve the dynamics in the series. This allows to estimate predictive densities of oil prices based on the statistical properties of the data alone in a parsimonious way, and not from the construction of complicated structural models. However, wrongly detrending the series could have a significant impact on the estimation of the noncausal dynamics of the process, which could in turn strongly under- or over-estimate the longevity of explosive episodes and therefore of the probabilities of crashes and of turning points.

3.3 Monte Carlo analysis - Effects of detrending

The aim of this section is to analyse the effect of wrongly detrending a series, both on the identification of the MAR model and on the

[4] In our case, the proposed boosting algorithm absorbs too much dynamics and captures the bubble in the trend component.

subsequent predictions performed with the resulting model. We base this analysis on stylised facts observed in oil prices series.

3.3.1 Accuracy of detrending

We simulate 5 000 trajectories for 12 distinct data generating processes (hereafter *dgp*), composed of a trend and a stationary dynamic process denoted as cycle. All *dgp*s are generated by Student's *t*-distributed errors with 2 degrees of freedom, a value frequently observed in financial time series, and with 400 observations. For the cycles, we consider purely noncausal processes with a lead coefficient of 0.8, purely causal processes with a lag coefficient of 0.6 and mixed causal-noncausal processes with a lag coefficient of 0.6 and a lead coefficient of 0.8. The heavy-tailed distribution generates extreme values, inducing bubble-like phenomena in processes with noncausal components. We are interested in mostly forward looking processes characterised by long lasting bubbles hence the choice of coefficients. We consider three different deterministic trends: a linear trend with breaks (denoted *breaks*) and two polynomial trends up to orders 4 and 6 (denoted respectively τ^4 and τ^6 for simplicity). The coefficients of the trends were estimated on the monthly WTI crude oil prices series between 1986 and 2019. Figure 3.2 depicts the three mentioned trends to which purely causal, noncausal and mixed causal-noncausal trajectories are added. Additionally, we consider processes with an intercept only. This results overall in 12 sets of 5 000 trajectories of the form $\tilde{y}_t = f_t + y_t$.

Four detrending methods are employed for each trajectories, with the general form $\tilde{y}_t = \hat{f}_t + \hat{y}_t$. Estimated polynomial trends of orders 4 and 6 and HP filters with $\lambda = 14\,000$ and $\lambda = 129\,600$ are applied (respectively denoted t^4, t^6, HP_1 and HP_2).[5] To gauge and compare

[5]The estimated polynomial trends are denoted t^4 and t^6 to distinguish them from

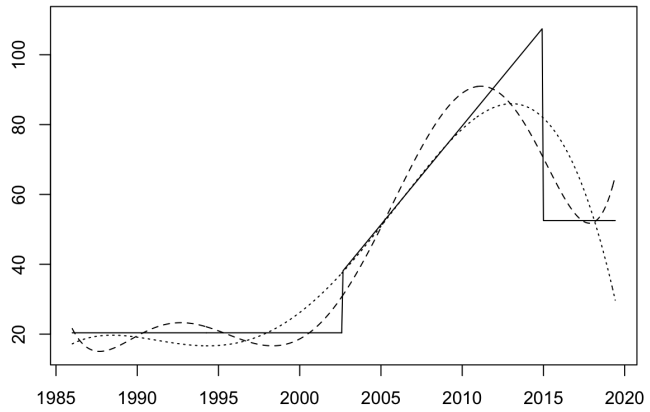


Figure 3.2: Trends estimated on WTI oil prices series

the accuracy of the detrending methods, Table 3.1 shows the average mean square errors (MSE) between the true cycle of $\tilde{y}_t$ (y_t) and the one obtained after detrending ($\hat{y}_t$). The average MSEs are computed over the 5 000 replications of each *dgp* and for the four detrending approaches,

$$MSE_{k,d} = \frac{1}{5\,000}\sum_{i=1}^{5\,000}\frac{1}{400}\sum_{t=1}^{400}(y_t^{(k,i)} - \hat{y}_t^{(k,i,d)})^2,$$

where *k* indicates the *dgp*, *d* the detrending method used, and *i* the *i*-th replication with $1 \leq i \leq 5\,000$.

The MSEs are minimised when the correct polynomial trend is employed or when the lower order is employed (4 in this case) in the absence of trend in the *dgp*. However, underestimating the order of the polynomial trend leads to significantly larger discrepancies. Distortions between the true cycle and the detrended series are

the polynomial trends part of the *dgp*s τ^4 and τ^6.

larger for mixed causal-noncausal processes than for purely causal or noncausal processes. Furthermore, in the presence of noncausal dynamics the HP filter with $\lambda = 14\,400$ (HP_1) distorts more the series than HP_2. Hence, we can expect that a low penalising parameter in the HP filter mostly captures some of the noncausal dynamics. However, HP_1 distorts the least the cycles to which the linear trend with breaks was added. It is the method that best manages to mimic this non-smooth trend due to this flexibility induced by its low penalising parameter.

Table 3.1: Average Mean Squared Errors between true cycles and detrended series

DGP	Detrended with			
	t^4	t^6	HP_1	HP_2
MAR(0,1) + no trend	5.23	7.61	11.44	7.15
MAR(0,1) + τ^4	4.55	6.03	9.62	7.50
MAR(0,1) + τ^6	62.42	6.38	11.35	11.26
MAR(0,1) + breaks	79.02	55.78	31.84	47.65
MAR(1,1) + no trend	22.69	31.05	48.58	30.81
MAR(1,1) + τ^4	42.74	65.18	91.42	57.60
MAR(1,1) + τ^6	85.91	39.57	61.21	43.02
MAR(1,1) + breaks	101.48	86.93	78.36	77.18
MAR(1,0) + no trend	1.20	1.64	2.55	1.58
MAR(1,0) + τ^4	0.96	1.34	2.14	2.70
MAR(1,0) + τ^6	59.24	2.45	4.21	6.73
MAR(1,0) + breaks	76.42	52.19	26.30	44.10

Are reported the average MSEs over 5 000 trajectories with sample size $T = 400$. HP_1 corresponds to the HP filter with $\lambda = 14\,400$ and HP_2 to the HP filter with $\lambda = 129\,600$.

3.3.2 Effects of detrending on model identification

To investigate the impact of detrending on dynamic processes, we perform MAR estimations on the raw and detrended series from each *dgp*. The estimation of MAR models first consists in estimating the pseudo causal lag order. Since the autocorrelation structure of mixed or purely causal and noncausal processes are identical, we can estimate the order of autocorrelation (p) with information criteria by OLS. Once this order p is estimated, the identification of the lag and lead orders (r and s respectively) is performed by maximum likelihood among all MAR(r,s) models such that $r + s = p$ (Lanne and Saikkonen, 2011). We do so using the MARX package in R (Hecq, Lieb, and Telg, 2017b).

Table 3.2 presents the frequencies of identifying wrong models in each of the 12 *dgp*, based on the detrending methods, with a maximum pseudo causal lag order of 4.[6] Proportions of a wrongly identified the pseudo lag order in the first step of the estimation using BIC are reported ($p \neq 1$ and $p \neq 2$), as well as the proportions of wrongly identified MAR models, namely when at least one of the lag or lead order mis-identified. We also report the frequency with which no noncausal dynamics is identified ($s = 0$). For the purely causal processes we only report in the last column ($s > 0$), i.e. the frequency with which spurious noncausal dynamics is detected.

Let us first focus on the models with noncausal dynamics (the MAR(0,1) and MAR(1,1) *dgp*s) for which we report the frequencies with which we over- or underestimate the pseudo causal lag order in the first step of the estimation. We can see that HP_1 under-performs relative to the other approaches. Indeed, around twice as many lag orders are wrongly estimated in the first step on average, with a maximum of 22.84% for the MAR(1,1) processes with breaks in the

[6] Results when the pseudo lag order is fixed to the correct one ($p = 1$ or $p = 2$ for mixed models) are available upon request.

linear trend. However, this non-smooth trend seems to be difficult to capture by the filters considered in this analysis. We can see from the last five rows of Table 3.2 that detrending this type of processes with breaks – with the four methods employed here – does not improve the correct identification of the orders of the model, and can even make it worse for MAR(1,1). This can be explained by the construction of the trend, mimicking somehow a bubble pattern, with a long and persistent expansion when the linear trend is present and followed by a sudden crash when the series returns to a stationary process. This might be mistaken for noncausal dynamics, ensuring a non zero lead order identification when the series is not detrended. This claim is supported by the results in the last column, indicating large proportions of wrongly detected noncausal dynamics for each detrending approaches, with 7.54% for HP_1 and more than 28% for the others. For the *dgp*s with other trends (or only intercept) HP_1 wrongly estimates the pseudo causal lag order at most 10.78% of the time. For the three other detrending methods the pseudo lag order is wrongly identified in less than 7.3% of the cases. Note that when the lag order is wrongly identified, it is almost always due to over-identification. The discrepancy between the two HP filters is explained by the low penalising parameter in HP_1 allowing the trend to mimic the series too much. By that, some of the dynamics of the MAR process are absorbed by the trend.

It is notably more harmful not to detrend when necessary than the contrary. As can be seen on the upper rows of Table 3.2, applying polynomial trends or HP_2 do not increase the proportions of wrongly identified models by more than 1.6% compared to estimations on the raw series. However, when the existing trend is ignored, the pseudo lag order is wrongly estimated twice as much on the raw series than for the detrended series, and the MAR models are wrongly identified up to 6 times more than the best performing detrending method. Furthermore, the incorrect identification of the pseudo lag order p accounts for most of the proportion of wrongly identified

MAR models. If p is correctly estimated, the model is also correctly identified in more than 99% of the cases. Note that the pseudo causal lag order identified is never zero, meaning that no detrending completely absorbs all dynamics. Besides, in no more than 0.62% the detrending methods killed the noncausal dynamics, as is indicated by the columns $s = 0$.

Let us now consider the last column, displaying the results for purely causal processes. We here investigate whether detrending can create spurious noncausal dynamics ($s > 0$). We find that (ignoring the *dgp* composed of the trend with breaks) as long as the polynomial trend order is not underestimated, in less than 3.46% of the cases noncausal dynamics was wrongly detected. For the processes with a polynomial trend of order 6, detrending with a polynomial trend of order 4 creates spurious noncausal dynamics in 60.02% of the cases.

Overall, for a *dgp* with noncausal dynamics, the impact of ignoring a trend is quite significant while detrending when not necessary has negligible effects on model identification. Both the polynomial trends and the HP filter with $\lambda = 129\,600$ (HP_2) perform equally well with respect to identifying the correct orders of the model. Choosing a penalising parameter λ too low alters the dynamics of the process as shown by the results from HP_1. All of the approaches almost always retain the noncausal dynamics, but rarely create spurious noncausal dynamics when nonexistent in the *dgp* (except when the polynomial trend order is underestimated). The lead order is not always the correct one but in less than 0.62% for all cases no noncausal dynamics is found. The presented results only report identification of the model lag and lead orders. To have a better understanding of the impact of the detrending methods on the dynamics, focus needs to be put on the impact on the estimated coefficients and parameters of the models identified.

Table 3.2: Percentages of mis-identified MAR models

Detrending method	$p \neq 1$	wrong MAR	$s = 0$	$p \neq 2$	wrong MAR	$s = 0$	$s > 0$
	MAR(0,1) + no trend			MAR(1,1) + no trend			MAR(1,0) + no trend
raw	5.50	5.50	0.00	4.74	4.74	0.00	0.52
t^4	5.46	5.58	0.14	4.64	4.68	0.04	0.72
t^6	5.70	5.86	0.22	4.94	5.04	0.06	0.88
HP_1	10.78	11.24	0.52	8.18	8.44	0.22	1.86
HP_2	6.84	7.10	0.28	5.56	5.66	0.06	1.02
	MAR(0,1) + τ^4			MAR(1,1) + τ^4			MAR(1,0) + τ^4
raw	12.10	43.84	35.04	9.70	16.38	7.22	32.76
t^4	6.44	6.72	0.28	4.70	4.74	0.04	0.78
t^6	6.76	6.96	0.20	4.86	4.94	0.08	0.76
HP_1	10.28	10.88	0.60	7.64	7.90	0.22	2.52
HP_2	6.24	6.56	0.32	5.36	5.56	0.14	1.92
	MAR(0,1) + τ^6			MAR(1,1) + τ^6			MAR(1,0) + τ^6
raw	13.18	36.14	26.04	9.04	15.56	7.12	35.44
t^4	7.30	7.36	0.08	4.00	4.14	0.04	60.02
t^6	6.54	6.68	0.14	4.48	4.68	0.04	0.92
HP_1	9.40	9.84	0.44	7.90	8.24	0.22	2.86
HP_2	5.94	6.12	0.18	4.86	5.12	0.06	3.46
	MAR(0,1) + breaks			MAR(1,1) + breaks			MAR(1,0) + breaks
raw	4.54	4.92	0.68	6.34	7.68	1.38	94.60
t^4	3.44	4.00	0.60	8.68	8.74	0.22	38.24
t^6	3.40	3.86	0.58	10.70	10.86	0.24	28.24
HP_1	4.00	4.58	0.62	22.84	23.24	0.26	7.54
HP_2	3.38	3.70	0.40	12.70	12.82	0.18	28.92

During the first stage of the model identification, the maximum number of lags in the pseudo lag model is set to 4. Results are in percentages of the 5 000 trajectories. $T = 400$. HP_1 corresponds to the HP filter with $\lambda = 14\,400$ and HP_2 to the HP filter with $\lambda = 129\,600$.

Detailed results on the impact on estimated coefficients are available in Appendix A. Overall, we find that due to low penalisation, HP_1 absorbs too much of the dynamics (mostly the noncausal ones) in the resulting trend. Hence, for monthly data, we advise to use the HP

filter with penalisation parameter 129 600. It is also rather harmful to underestimate the order of the polynomial trend, which results in a significantly larger lead coefficient. When the fundamental trend consists of breaks (mimicking bubbles), the smooth detrending methods do not succeed in capturing the trend and this translates in much more persistent noncausal dynamics. We also investigate the effect of detrending white noise series; while for the raw series, 6.82% of the models were identified with dynamics, 7.34% were identified with dynamics for the HP filtered series with penalising parameters 129 600. Hence we find no significant creation of dynamics when applying the HP filter to a white noise.

3.4 Predicting crashes in oil prices

This section investigates the impact of detrending both for in-sample and real-time analyses. WTI and Brent crude oil monthly prices series are employed, ranging from June 1987 to December 2020. The series consist of end-of-period prices, which enables us to adequately time our analysis based on the outbreak of the COVID-19 pandemic and the appearances of worldwide regulations and lock-downs to counter its spread. Figure 3.3 shows that both series are characterised by bubble episodes, which we define in this chapter as rapidly increasing episodes followed by a sharp decline, the main one being during the financial crises in 2008. The series are also characterised by various sudden crashes. The highlighted gray bar represents the period of interest in this analysis. The earliest point of the period is December 2019; at this point almost no information was available on the coronavirus and no worldwide outbreak had already taken place. Then, we can see that as the outbreak started and regulations were increasingly being imposed worldwide, the price of crude oil significantly dropped. Brent crude oil prices fell from around $68 at the end of December 2019 to around $15 by the end of March 2020, point at which most European countries imposed national lock-downs. The restrictions of movement within and between countries

thus induced a sharp and sudden decrease in the demand for crude oil.

As shown in Figure 3.3, the series are probably nonstationary but considering their growth rate would eliminate the locally explosive episodes that are interesting to exploit. The two series appear almost identical until the 2008 financial crisis, period from which we can observe more apparent discrepancies. The last part of the samples is rather noisy and volatile, and estimating a trend on such a part is not straightforward.[7] We seek to extract a smooth trend without affecting the dynamics of the series. Based on the findings of Section 3.3, we consider the deterministic polynomial trends of orders 4 and 6 as well as the HP filter with $\lambda = 129\,600$ (denoted t^4, t^6 and HP respectively). We furthermore employ an economic variable – described in the following section – as another trend to compare economically motivated detrending with mechanical detrendings. The analysis focuses on the probabilities for oil prices to drop and investigates the potential magnitude of such decrease. We first consider an in-sample analysis, that is, the trends and the MAR models are estimated over the whole sample, from June 1987 to December 2020. Then, we fix the estimated parameters and use this information to perform one-month ahead density forecasts for the months of January, February, March and April 2020. The in-sample analysis includes as much information as possible and therefore reduces estimation uncertainty. We then compare the in-sample analysis to a real-time forecast exercise. In the real time analysis, we re-estimate the trends and the MAR models at each point of the period of interest. That is, we consider an expanding sample and perform one-month ahead density forecasts for points that are out-of-sample.

[7] Figure 3.11 in Appendix B shows the oil prices deflated with the consumer price index.

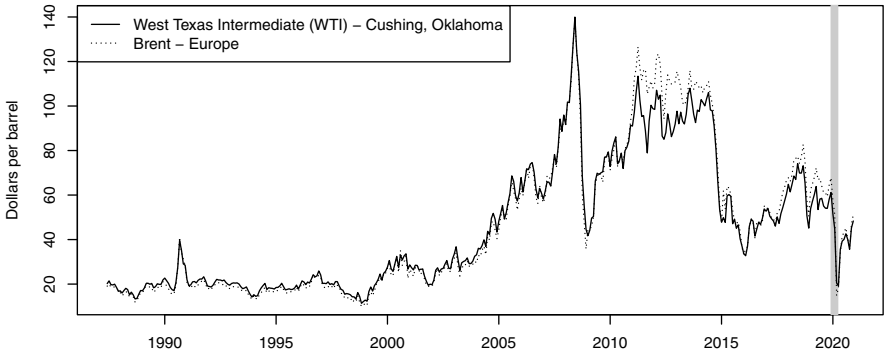


Figure 3.3: Monthly crude oil prices

3.4.1 Economic variables to detrend series

There is an extensive literature on modeling oil prices using economic variables. As an example, Kilian and Murphy (2014) construct a structural VAR model for the real price of oil, making use of stationary transformations of economic variables, namely the real economic activity index constructed in Kilian (2009) as well as inventories and production of crude oil. In this analysis we however do not construct a structural model for the price of oil, but instead we investigate ways of detrending prices without altering the inherent dynamics of the process. As such, we suggest employing the US crude oil Strategic Petroleum Reserve (hereafter SPR) levels. These reserves were established primarily to reduce the impact of disruptions in supplies of petroleum stocks (Kilian and Zhou, 2020a). This variable therefore incorporates not only expectations regarding the economic activity but also regarding the production of crude oil. US SPR stock is depicted against WTI crude oil prices in Figure 3.4.

SPR is significantly less volatile than total crude oil stocks as it is a last resort reserve and is not often made use of as it requires approval of the

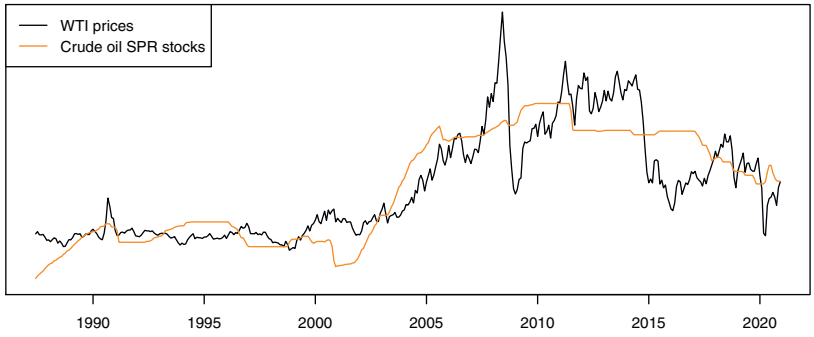


Figure 3.4: Raw WTI prices and US crude oil SPR stocks

US President.[8] This characteristic of the series makes it a good candidate for the smooth trend we intend to extract from oil price series. We hence detrend prices (both nominal and real) by taking the residuals from a standard OLS regression of prices on crude oil SPR levels.[9]

3.4.2 In-sample analysis

To save space, Figure 3.5 only depicts the detrended Brent series,[10] after the polynomial trends, the HP filter and SPR levels were used to detrend the whole sample. The *SPR*-detrended series consists of the residuals obtained from a standard OLS regression of the prices on the SPR levels. We can see that the *HP*-detrended series (black solid line) and the t^6-detrended series (dashed line) are very much alike over the majority of the sample. The polynomial trend of order 4 (dotted line)

[8] Limited release can be allowed by the Secretary of Energy for crude oil loans to non-governmental entities, as is described by the Energy department of the US.

[9] As shown in Figure 3.3, WTI and Brent price series seem to follow a similar trend; we therefore also employ US SPR stocks to detrend Brent prices. The remaining cycles look similar to the *HP*-detrended series and are therefore stationary.

[10] Data and results for the prices-adjusted series that are not presented here are available upon request.

however seems to induce some more variations than the other two mechanical detrending. This is especially visible at the beginning and at the end of the sample, stemming from the lack of flexibility of such trend due to its lower order. This latter detrending method suggests that the end of 2020 is as extreme as the period between 2010 and 2014, during which prices were in fact twice as large. We can see that overall *SPR*-detrended series follows a similar pattern than the others but shows slightly more persistent dynamics. This could stem from the fact that the SPR series, while being rather smooth, still displays more dynamics than the 3 other trends considered here. Hence, it could slightly alter the dynamics in the remaining cycle. *SPR*-detrended series has a correlation of 0.84 with both *HP*- and t^6-detrended series. Note that until the end of the 1980s, there was a persistent increase in SPR due to the creation and initial filling of the reserves which started in 1977, explaining the induced downward trend at the beginning of the detrended sample.

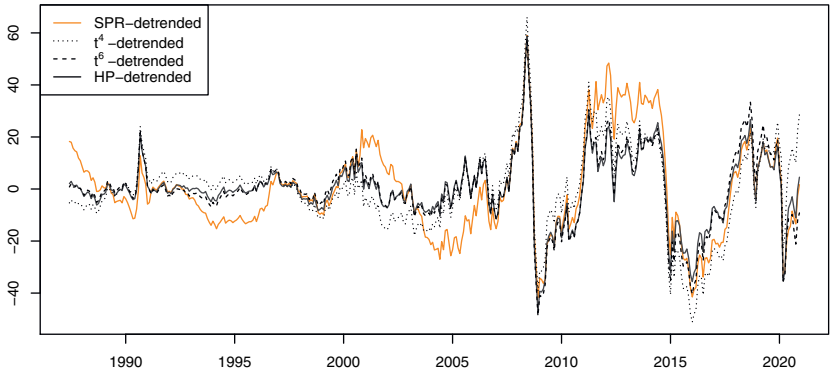


Figure 3.5: Detrended Brent prices

We estimate MAR models with Student's t-distributed errors and set the maximum pseudo lag length in the first stage on the estimation

to 4. All resulting models are MAR(1,1) and are reported in Table 3.3. We report the lag and lead coefficients as well as the degrees of freedom of the distribution and their respective standard errors in parentheses. Models estimated on series that were detrended with a polynomial trend of order 6 and with the HP filter are the most similar, as suggested by Figure 3.5. Models estimated after detrending with the polynomial trend of order 4 slightly deviate from the two others and always have a larger lead coefficient, hence indicating more persistence in the explosive episodes. Recall that Section 3.3 suggests that underestimating the trend order in mixed causal-noncausal models induces on average an overestimation of the noncausal coefficient. All series are mostly forward looking, as the lead coefficients are at least 0.8 while the lag coefficients are at most 0.31.[11] We can see that, as expected, *SPR*-detrended series are slightly more persistent in their noncausal dynamics with a lead coefficient up to 0.1 larger than other detrending methods and also slightly larger degrees of freedom induced by more persistent extreme events. The identification of the dynamics is overall consistent across series and their transformation. Note that adjusting the series for inflation leads to larger estimated degrees of freedom for the Student's t distribution but overall to similar dynamics.

Lacking closed-form expressions for the predictive densities, we use the two data-driven approaches mentioned in Section 3.2. We employ the simulations-based approach of Lanne, Luoto, and Saikkonen (2012), which only depends on the model estimated and the last observed point and compare, it approximate the density by use of simulations. We compare this method with the sample-based approach of Gouriéroux and Jasiak (2016), which uses past values in

[11] The bi-modality of the coefficient distribution in the estimation can lead, in the optimisation of the likelihood function, to a local maximum (Bec, Nielsen, and Saïdi, 2020b). This phenomenon is subject to initial values and can induce a switch between the lag and lead coefficients. This was however thoroughly checked in the analysis.

Table 3.3: Estimated MAR models

Series	MAR(1,1) estimations per detrending method											
	t^4			t^6			HP			SPR		
	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$
WTI	0.25	0.88	2.25	0.29	0.82	1.93	0.29	0.80	1.85	0.24	0.90	2.60
	(0.03)	*(0.01)*	*(0.36)*	*(0.03)*	*(0.02)*	*(0.29)*	*(0.03)*	*(0.02)*	*(0.28)*	*(0.03)*	*(0.01)*	*(0.37)*
WTI_{real}	0.22	0.88	3.05	0.26	0.83	2.75	0.26	0.81	2.63	0.22	0.91	3.44
	(0.04)	*(0.02)*	*(0.49)*	*(0.04)*	*(0.02)*	*(0.50)*	*(0.04)*	*(0.02)*	*(0.49)*	*(0.04)*	*(0.02)*	*(0.69)*
Brent	0.31	0.89	1.93	0.31	0.86	1.82	0.31	0.83	1.83	0.31	0.92	2.18
	(0.03)	*(0.01)*	*(0.27)*	*(0.03)*	*(0.02)*	*(0.33)*	*(0.03)*	*(0.02)*	*(0.31)*	*(0.03)*	*(0.01)*	*(0.34)*
$Brent_{real}$	0.26	0.90	2.59	0.27	0.86	2.48	0.27	0.84	2.45	0.25	0.92	2.90
	(0.04)	*(0.02)*	*(0.55)*	*(0.04)*	*(0.02)*	*(0.56)*	*(0.04)*	*(0.02)*	*(0.57)*	*(0.04)*	*(0.02)*	*(0.53)*

The models are obtained with a maximum pseudo lag order of 4 and for each series the model identified was an MAR(1,1). ϕ is the lag coefficient, ψ is the lead coefficient and γ the degrees of freedom of the Student's t distribution. The polynomial trend are trends up to the order indicated and the HP filtering is performed with a penalisation parameter $\lambda = 129\,600$. In parentheses are reported the standard error of the coefficients estimated obtained with the MARX package (Hecq, Lieb, and Telg, 2017b).

the forecasting step to approximate the conditional density. Table 3.4 shows the one-month ahead probabilities that the series will decrease (hence be lower than its last observed value) and the probabilities that the series will drop by more than 1 standard deviation (the standard deviations are calculated empirically over the whole sample). Forecasts are performed for January, February, March and April 2020 and results from the two prediction methods are reported for each of the detrended nominal series. We focus on the nominal series as they are the prices people observe and because the estimated models for real series are noticeably similar.[12] While we advocate the use of predictive densities to get the best picture of potential future prices, we choose 2 arbitrary probabilities to present for a matter of comparison and to save space. Nonetheless, the probabilities for any

[12]Results for price-adjusted series can be found in Appendix B, probabilities slightly vary however the patterns described in the results for nominal series are identical.

event can be computed from the methods used here, and they could for instance be employed in the construction of risk measures.

Table 3.4: One-step ahead probabilities

Series	Detrended with	Jan.		Feb.		Mar.		Apr.	
		samp.	sims.	samp.	sims.	samp.	sims.	samp.	sims.
		Probability of a decrease							
	t^4	.444	.423	.784	.762	.722	.681	.828	.825
	t^6	.414	.437	.873	.851	.726	.748	.583	.705
	HP	.411	.422	.869	.836	.701	.730	.544	.675
	SPR	.432	.440	.808	.768	.691	.687	.738	.781
WTI									
		Probability of a decrease $>$ 1 s.d.							
	t^4	.052	.044	.016	.017	.006	.008	.018	.015
	t^6	.041	.042	.007	.011	.004	.005	.177	.322
	HP	.047	.045	.007	.012	.005	.006	.227	.399
	SPR	.012	.010	.005	.005	.002	.003	.005	.014
		Probability of a decrease							
	t^4	.379	.346	.806	.800	.718	.696	.879	.852
	t^6	.398	.400	.886	.864	.792	.768	.569	.770
	HP	.397	.396	.880	.853	.757	.745	.500	.720
	SPR	.386	.390	.861	.824	.786	.731	.678	.860
$Brent$									
		Probability of a decrease $>$ 1 s.d.							
	t^4	.044	.035	.016	.016	.007	.008	.071	.106
	t^6	.034	.034	.004	.011	.001	.005	.308	.552
	HP	.037	.038	.008	.012	.003	.006	.350	.591
	SPR	.012	.010	.003	.006	.000	.003	.025	.065

For the simulations-based approach (sims.) the truncation parameter $M = 100$ and 1 000 000 simulations were used. Standard deviations (s.d.) are calculated over the detrended samples and are around 15 for all nominal series.

At the end of December 2019 oil prices were around \$60 per barrel, they had been fluctuating around this price over the last three years. All detrending methods yield values for December that are above the 90^{th} percentile of the samples, suggesting high but not extreme levels. At that point in time, no international alerts regarding the risk of a pandemic had been made yet. Probabilities that prices will drop in

January are roughly 0.4 for all series and for both forecasting methods. However, probabilities that prices will drop by more than 1 standard deviation are at most 0.052. This confirms that crude oil prices are in a period of volatile and rather high prices, but it does not suggest a bubble behavior with a potential large drop. This can also be seen by the difference between the sample-based and simulations-based predictions. Hecq and Voisin (2021) show that discrepancies between the two approaches mostly arise during extreme episodes. Here, they do not differ by more than 3.3% for the probabilities of a decrease, and by no more than 0.9% for the probabilities of a sharper decrease.

At the end of January 2020, international alerts regarding the spread of the novel coronavirus had been made, which induced an unforeseeable drop in prices. Yet, the t^4-detrended series only fell by half a standard deviation and the other two by 75% (resp. 80%) of a standard deviation for the Brent (resp. WTI) series. Values remained however above median values. Forecasts based on both methods suggest a continuity in the decrease for February with probabilities ranging from 0.76 to 0.88, yet, they indicate almost zero probability that the drop will be substantial (more than a standard deviation). They hence suggest a return to median values, meaning a return to fundamental prices. Both prediction methods again provide results diverging by no more than 3.3%. By the end of February 2020, mass gatherings started to be forbidden and the first advice for the quarantine of individuals to contain the spread of the virus had be made. The increasing worldwide pressure hence kept pushing prices down. Yet, no decrease in the detrended series was larger than 60% of a standard deviation, which was once again in line with the predictions. The series reached their median levels, forecasts for March suggested that series would remain stable around those values, yet favoring a further slight decrease as prices had been declining for the last three consecutive periods. Probabilities of a sharp drop decreased even more towards zero and both prediction methods yielded again similar probabilities.

In March 2020 the worldwide situation worsened significantly and the World Health Organisation declared COVID-19 a global pandemic. Many countries imposed strict movement restrictions within and across borders, and curfews and lock-downs were implemented. This sudden drop in crude oil demand led to a considerable fall in prices, WTI prices fell by 55% and Brent prices by 71%. Values of the detrended series fell by more than 2 standard deviations and reached the 2^{nd} and 3^{rd} percentile for HP-, SPR- and t^6-detrending. This indicates a negatively explosive episode, and therefore a negative bubble below fundamental prices. The t^4-detrending values correspond to at least the 10^{th} percentile, suggesting a less extreme episode, compared to the previous behavior of the series. Until this point both predicting methods yielded similar probabilities. However, the discrepancy between the probabilities now attain 0.24 difference, where the simulations-based probabilities of a decrease are always larger than the sample-based probabilities. Hecq and Voisin (2021) show that the discrepancies between the sample- and simulations-based approaches widen during explosive episodes. This is why probabilities for t^4-detrending series are still very similar across the forecasting methods as opposed to the other detrending methods. They also show that the larger the lead coefficient, the more the sample-results tend to yield larger probabilities of a turning point than the ones computed with simulations. This stems from the fact that the series had attained a few times this point before (in 2008 and in 2015) and turned back towards median value. It is therefore, based on the learning mechanism of the sample-based approach, less likely that the series will keep on decreasing. It is important to notice that even though prices dropped significantly, probabilities that they will keep on decreasing are lower than before for HP- and t^6-detrended series as well as for SPR-detrended Brent. However, compared to previous forecasts, probabilities now suggest that if the series actually kept on decreasing, it could likely be by more than 1 standard deviation as it has now entered an explosive episode. SPR-detrended

WTI series has a larger probabilities of decrease than for the previous month, however, as can be notice, the probabilities of the sharper decrease for both SPR-detrended series are much closer to 0 than with other detrending. This stems from the larger degrees of freedom as well as larger lead coefficient and slightly lower lag coefficient.

Figure 3.6 illustrates the evolution of the predictive densities of the *HP*-detrended Brent series over the time span. On the $x - axis$ are the predictions and on the $y - axis$ their corresponding probability density. The vertical dashed line corresponds to the last value, that is, in graph (a), the vertical line is the detrended value of Brent prices for December 2019. We can clearly observe the bi-modality of the distribution when the series deviates from median values, as shown for the forecasts of January and April, which exacerbate during the explosive negative episode. The range and shape of the density also explains the discrepancies between probabilities of a decrease and probabilities of a decrease of more than 1 standard deviation.[13]

To illustrate the valuable information provided by the predictive densities of MAR models, graph (a) of Figure 3.7 depicts the predictive density for April 2020 using a Gaussian $AR(2)$ model instead of an MAR(1,1) on HP-detrended Brent prices. The predictive density is obtained using the closed-form of the conditional normal distribution. We can see that the mode of the density corresponds to a further decrease, but it now lacks the bi-modality and therefore does not suggest a return to central values as does the MAR predictive density shown on graph (d) of Figure 3.6. As such, once the series enters a locally – here negative – explosive episode, the $AR(2)$ only predicts a continuing decrease of the prices. Graph (b) of Figure 3.7 displays the sample-based predictive density of the SPR-detrended Brent series. We can see that the larger lead coefficient implies a lower rate of decrease (this can be seen as the distance between the two

[13] Results for all other series, available upon request, follow a similar pattern.

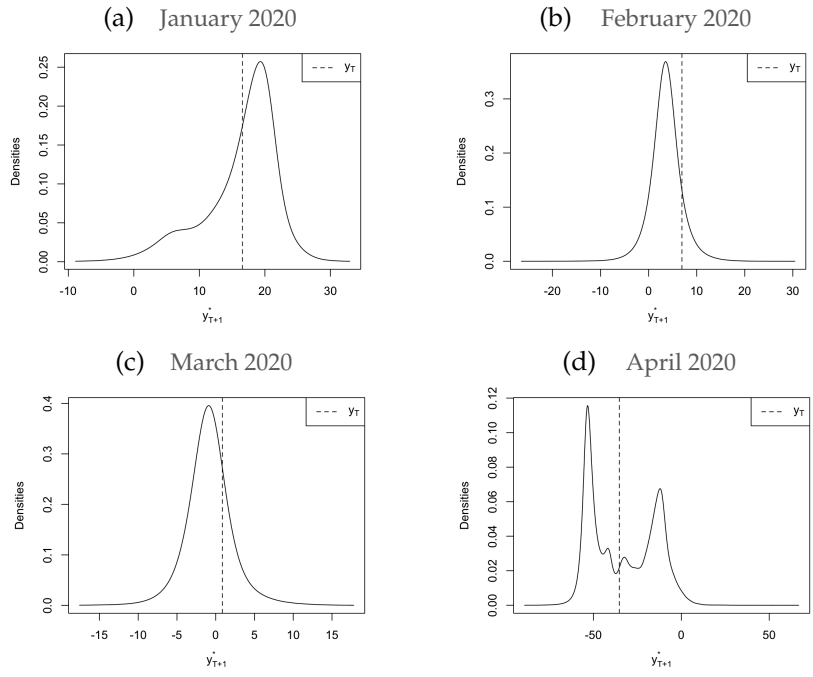


Figure 3.6: One-step ahead predictive densities of *HP*-detrended Brent prices obtained with the sample-based prediction method.

modes), but it indicates larger probabilities of a further decrease as large lead coefficients imply longer lasting explosive episodes. This is why the right mode, which corresponds to a return to central values has a much lower weight on the density.

The MAR models employed here are univariate, hence no exogenous information is incorporated, as opposed to MARX models (see Hecq, Issler, and Telg (2020) and Hecq, Issler, and Voisin (2022)). Disregarding exogenous variables facilitates forecasting but can sometimes lead to consequential lack of information. For instance, it is expected that crude oil prices should be lower-bounded as they

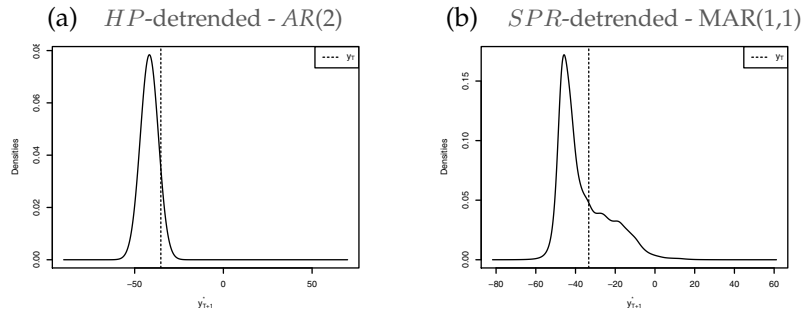


Figure 3.7: One-step ahead predictive densities of *HP*-detrended Brent prices obtained with the sample-based prediction method.

cannot decrease indefinitely and become increasingly negative. Simulations-based probabilities cannot not take that into account as they are only based on the model estimated. Sample-based probabilities however, since prices have never become negative (or at least not long enough to be visible on monthly series), will tend to limit the probabilities that it will happen in the future, even without incorporating additional information within the model, based on its learning mechanism.

Overall, *HP*- and t^6-detrending provide similar results both for estimation and predictions. SPR-detrending, as mentioned earlier yields slightly different dynamics which might stem from the dynamics that are inherent to the stock variable itself. Detrending with t^4 yields slightly different results for the estimation but which in turn yields quite different results for predictions. We saw in Figure 3.5 that detrending with a polynomial trend of order 4 induced different dynamics in the remaining cycle than the other mechanical detrending. This also corroborates the results found in Section 3.3 about the risks of underestimating the order of a polynomial trend on the dynamics of the series. We can also see in Figure 3.5 that the main differences between all detrending methods appear at the end

of the sample. HP and SPR detrending are almost identical while t^6 provides slightly lower values. On the other hand, t^4-detrending yields significantly larger value than the others for the end of the sample.

3.4.3 Real-time analysis

To illustrate the difficulties and the limitations of detrending and forecasting in real time, we compare the results obtained in real time to the ones obtained in-sample for Brent prices with t^4, t^6 and HP detrending. We did not include SPR detrending in this Section as we are interested in detrending methods that are affected by sample expansion and while with SPR detrending we still need to re-estimate the model at each point, the trend itself does not change. Table 3.5 shows the estimated MAR models for the expanding samples after each detrending. We can see that the expansion of the sample, even with the inclusion of the large drop of March 2020 did not affect the identification of the model nor the dynamics. Lead and lag coefficients vary by no more than 0.03. The estimated degrees of freedom of the Student's t distribution are rather stable until the data point of March is included, which induced decrease between 0.07 and 0.1 for all series, getting therefore closer to the parameter estimated ex-post. This stability in the estimation of the models suggest that probabilities should not significantly differ either.

To investigate the sensitivity of the detrending methods to the addition of new data points, Figure 3.8 shows how the detrended series vary based on the stopping point of the sample. The dashed line corresponds to the ex-post detrended series, hence when all data points until December 2020 are included. Then, the expanding samples are depicted from the light blue curve (sample stopping in December 2019) to the black curve (sample until March 2020). While detrending with t^4 induced the most spurious dynamics over the

Table 3.5: Estimated MAR models on different Brent prices samples

Sample	MAR(1,1) estimations per detrending method								
	t^4			t^6			HP		
	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$
In-sample	0.31	0.89	1.93	0.31	0.86	1.82	0.31	0.83	1.83
	(0.03)	*(0.01)*	*(0.27)*	*(0.03)*	*(0.02)*	*(0.33)*	*(0.03)*	*(0.02)*	*(0.31)*
$\rightarrow$ Dec	0.30	0.89	2.06	0.29	0.86	1.98	0.29	0.84	1.97
	(0.03)	*(0.01)*	*(0.27)*	*(0.03)*	*(0.02)*	*(0.33)*	*(0.03)*	*(0.02)*	*(0.31)*
$\rightarrow$ Jan	0.30	0.89	2.05	0.29	0.86	1.97	0.28	0.84	1.97
	(0.03)	*(0.01)*	*(0.31)*	*(0.03)*	*(0.02)*	*(0.33)*	*(0.03)*	*(0.02)*	*(0.31)*
$\rightarrow$ Feb	0.30	0.89	2.07	0.31	0.85	1.97	0.30	0.83	1.97
	(0.03)	*(0.01)*	*(0.31)*	*(0.03)*	*(0.02)*	*(0.32)*	*(0.03)*	*(0.02)*	*(0.31)*
$\rightarrow$ Mar	0.30	0.89	1.97	0.31	0.85	1.89	0.30	0.83	1.90
	(0.03)	*(0.02)*	*(0.36)*	*(0.03)*	*(0.02)*	*(0.36)*	*(0.03)*	*(0.02)*	*(0.33)*

See Table 3.3.

sample, it seems, as well as the *HP* filter, to be less affected by the addition of the new points than the t^6-detrending. In graph (a), we can see that the 4 detrended series are almost identical, even once the point for March is added. In graph (c), corresponding to the *HP*-detrended series, we can see that the 3 first detrended series are almost identical but that the inclusion of March creates a slight shift in the detrended series. In this case also, the inclusion of even later points will induce further shifts of the estimated trend. However, for the polynomial trend of order 6, as depicted in graph (b), we can see that the inclusion of each point creates a noticeable shift in the estimated trend. From this, we expect the t^6-detrended series to be the ones for which the probabilities differ the most from the in-sample probabilities. Indeed, even if the estimated model is almost identical, the substantial discrepancies between the real-time and ex-post detrended series may impact probabilities, especially during (mildly) explosive episodes.

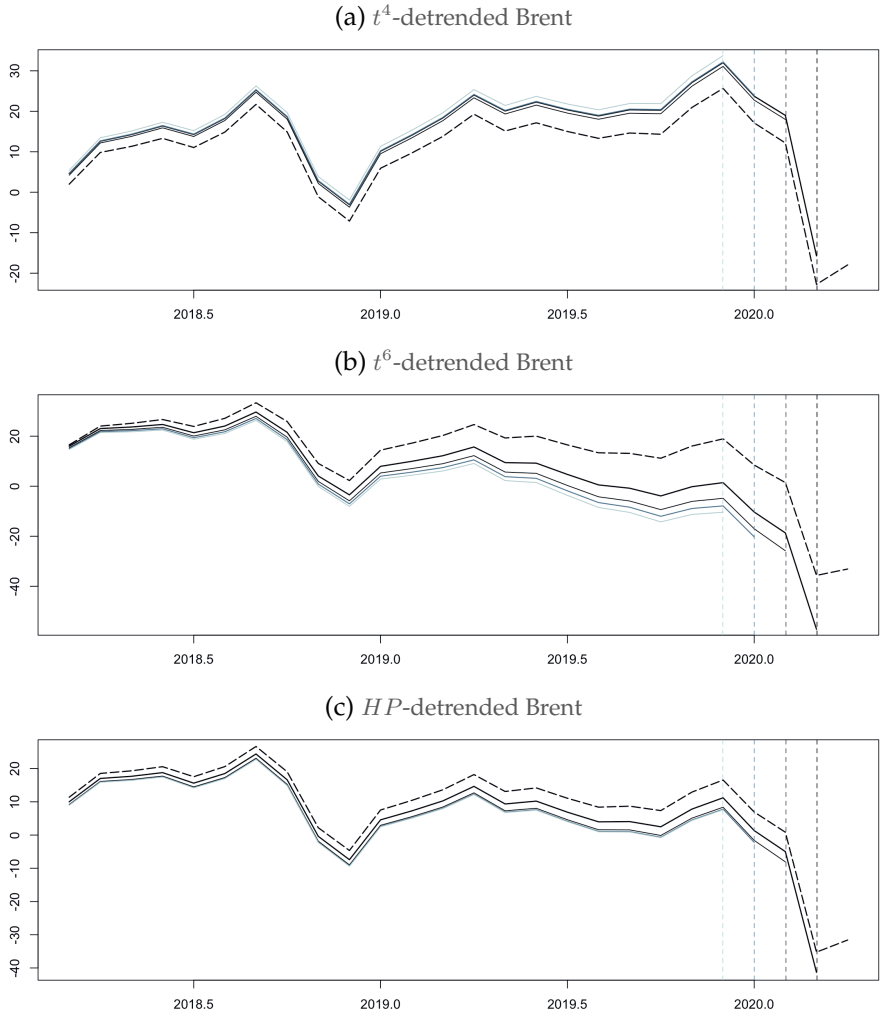
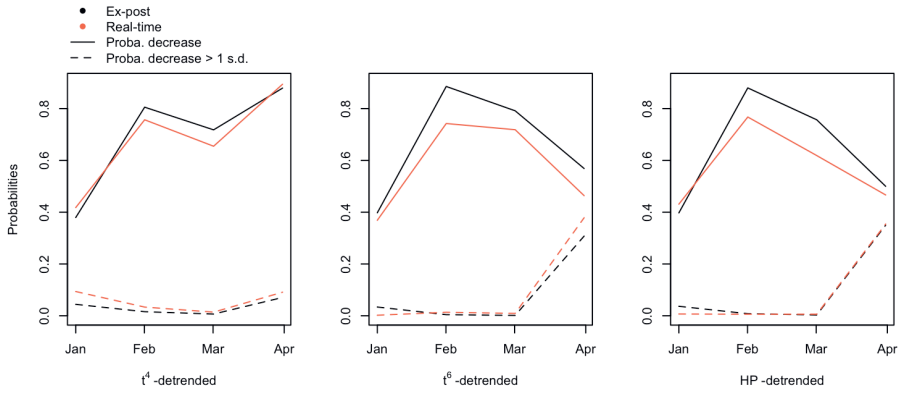


Figure 3.8: In-sample (dashed curve) vs real-time detrending of Brent prices

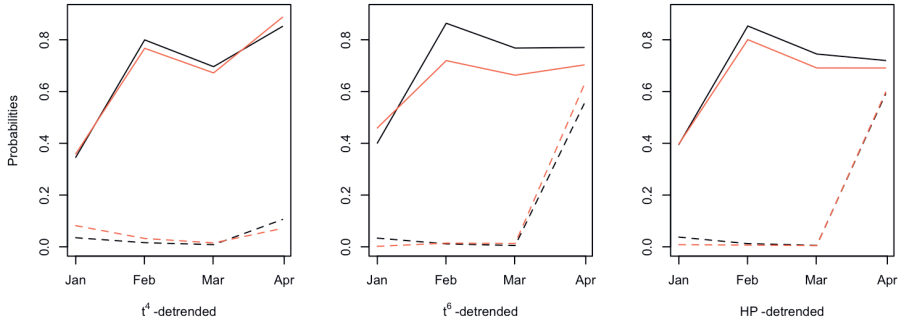
Figure 3.9 depicts the evolution of the one-month ahead probabilities with expanding window. In black are the in-sample probabilities and in orange the real-time probabilities. For the real-time analysis,

the trend and the model is re-estimated at each point. The full lines are the probabilities of a decrease and the dashed lines are the probabilities of a decrease of more than 1 standard deviation. Graph (a) (resp. (b)) represents the sample-based (resp. simulations-based) probabilities. As expected, the simulations-based probabilities are the least affected by the re-estimation of the model at each point, since we did not observe significant alteration in the estimations. However, as shown in Figure 3.8, it is indeed the t^6-detrending that is the most sensitive to the expansion of the sample. Furthermore, we can see that mostly the probabilities of a decrease are affected, as the probabilities of a drop of more than 1 standard deviation are not significantly deviating from the in-sample probabilities. Overall, this indicates that real-time forecasting would have indicated on average lower probabilities of a decrease, at each point and for both approaches. Yet, it would have indicated equal, if not slightly higher, probabilities for the larger drop. Hence, probabilities of more extreme events, namely the tails of the predictive densities, seem to be the least affected by alteration of the trend.

Overall, it seems that the *HP*-filter is the least sensitive to the change of sample size within this analysis. Results with t^4-detrending also emphasises the risks of underestimating the order of the trend. Moreover, while the *HP*-filter and the polynomial trend of order 6 perform similarly in this analysis, assuming the order of a polynomial trend requires additional understanding regarding the deviations of the series from its fundamental trend. Deterministic trends appear also to be more sensitive to the addition of points in a real-time exercise than the HP filter. Furthermore, while simulations-based probabilities are not characterised by the learning mechanism of the sample-based approach, they are less affected by expanding samples, as long as the model estimated remains consistent. However, as mentioned earlier, with a model that lacks exogenous information, the sample-based approach relying more on past behavior can potentially offset the shortcomings.



(a) Sample-based probabilities



(b) Simulations-based probabilities

Figure 3.9: Evolution of in-sample (black solid and dashed lines) and real-time (orange solid and dashed lines) probabilities over time

3.5 Conclusion

This chapter aims at shedding light upon how transforming or detrending a series can substantially impact predictions of mixed causal-noncausal models. Assuming a polynomial trend of order 4 for WTI and Brent series probably alters the dynamics in the remaining cycle. The HP filter (with penalising parameter $\lambda = 129\,600$) does not

require any further assumptions with respect to the trend and can therefore be an adequate filter in cases where the trend is unknown. Knowing the actual trend or using exogenous variables for it is also not straightforward. We use US crude oil strategic petroleum reserves (SPR) to detrend oil price series to illustrate this option. We show that by detrending with SPR we obtain similar results to the HP and polynomial trend of order 6 detrending. However, detrending with a variable that has seasonality or dynamics will alter the dynamics left in the cycle. Overall, caution is needed when detrending a series, and some filtering such as polynomial trends may require additional understanding regarding the deviations of the series from its fundamental trend. Nonetheless, once the series is detrended, resulting in a stationary series, using MAR models is a straightforward approach to model nonlinear time series. They capture the locally explosive episodes observed in oil prices in a strictly stationary setting. While the bi-modality of the predictive density would not be detected with standard Gaussian ARMA models, it could be detected with complex nonlinear models, but such model lacks the parsimonious characteristic of MAR models. The data-driven prediction methods may lack theoretical grounds but provide valuable information based on the estimated model and on past behaviors of the series in a parsimonious way. This chapter focuses on one-step ahead predictions of decrease in crude oil prices during the first wave of the COVID-19 pandemic.

Appendix A Impact of detrending on estimated coefficients

We now investigate the persistence of the dynamics from the magnitude of the estimated coefficients. For instance, a lower lead coefficient will indicate shorter lived bubbles compared to the true generated process and thus increases the probabilities of a crash during an explosive episode. The same goes for larger degrees of freedom when the errors follow a Student's t distribution: larger degrees of freedom correspond to thinner tails, and thus rarer extreme values and thus makes less probable long lasting explosive episodes.

We investigate the distribution of the estimated coefficients given a correctly identified model. Frequencies of wrongly identified models per *dgp* and detrending method are shown in the columns 'wrong *MAR*' of Table 3.2. Hence, proportions of correctly identified models range between 76.76% and 96.3% of the 5 000 replications, but are almost always above 90%. Figure 3.10 reports the box plots of estimated coefficients for the purely noncausal (left column) and mixed causal-noncausal (center and right columns, for the lag and lead coefficients respectively) processes after each of the four detrending approaches is applied. We indicate the true coefficients, 0.6 and 0.8 for the lag and lead respectively, by the vertical dotted line. The box plots indicate the minimum, maximum, the interquartile range and the median. The HP_1-filtered series (with $\lambda = 14\,000$) are on average characterised by lower estimated lead and lag coefficients than the other detrended series. This is due to the low penalisation of the filter, capturing too much of the dynamics, reducing the persistence of the true noncausal process. Furthermore, we can see that using polynomial trends does not affect estimations of the coefficients, on average, as long as the order of the trend estimated is at least that of the true trend. That is, underestimating the order of the trend leads to an alteration of the dynamics and in our case, to more persistent noncausal dynamics. The HP_2 filter performs similarly to

t^6, but we can expect that if the true trend was a higher order, HP_2 would perform better. The constructed linear trend with breaks leads to much larger noncausal coefficients for all detrending methods. The second break in the trend mimics the crash of a bubble and the long expansion preceding it leads to the identification of the model with a larger lead coefficient, which corroborates the earlier findings. Importantly, lag coefficients are on average correctly identified (the distributions of the estimated degrees of freedom, available upon request, show that they are not significantly affected by the detrending either). A wrong detrending therefore mostly affects the noncausal dynamics of the processes.

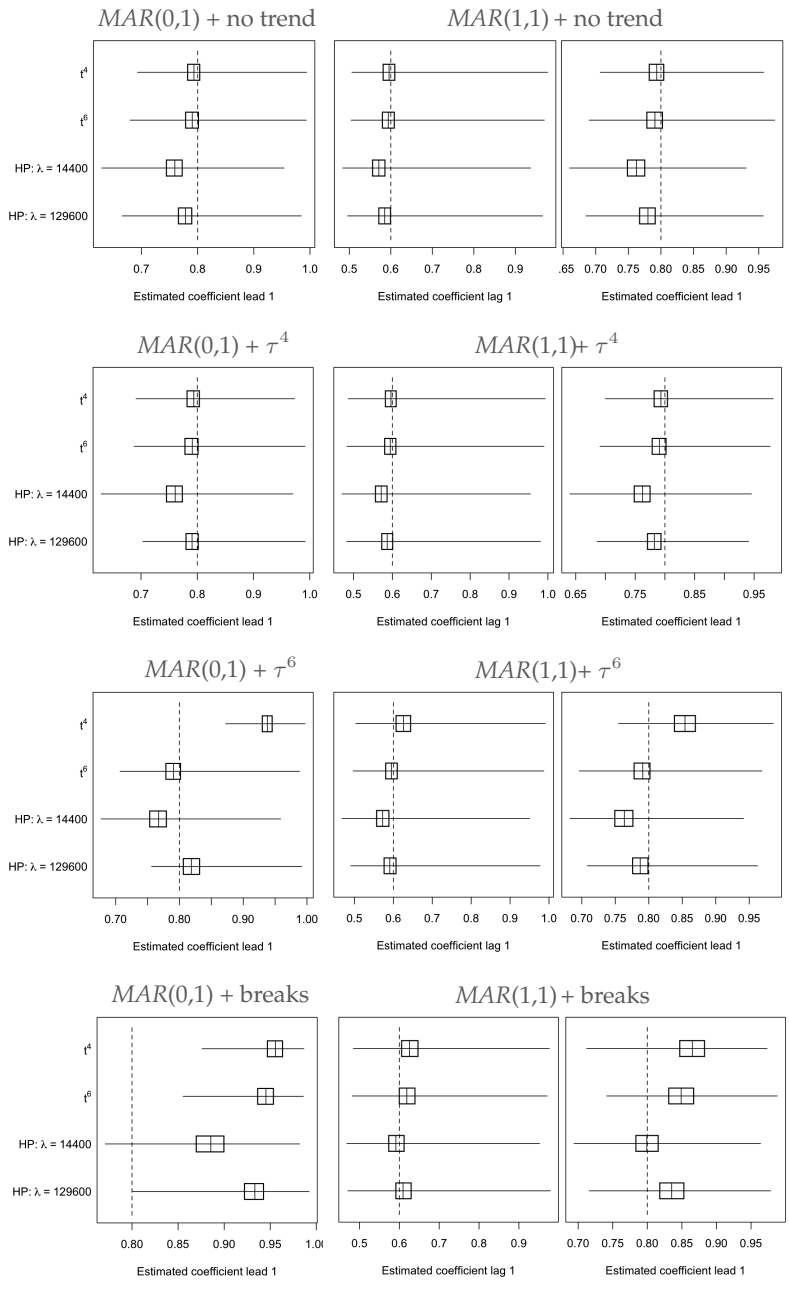


Figure 3.10: Distribution of estimated *MAR* coefficients

Appendix B Results for price adjusted series

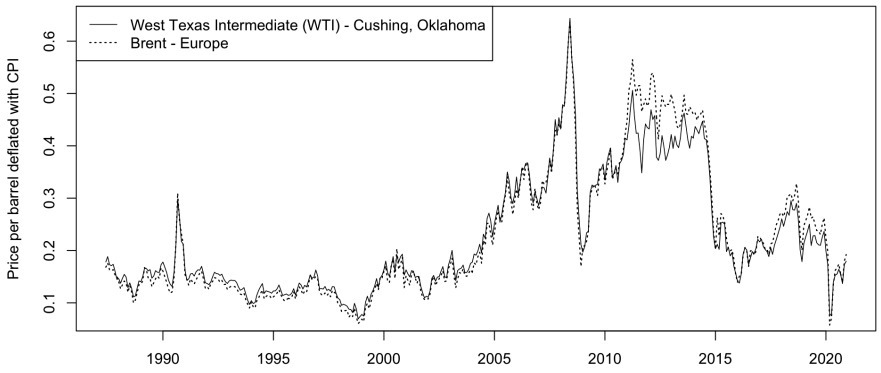


Figure 3.11: Monthly crude oil prices deflated with Consumer Price Index

Table 3.6: One-step ahead probabilities

Series	Detrended with	Jan.		Feb.		Mar.		Apr.	
		samp.	sims.	samp.	sims.	samp.	sims.	samp.	sims.
		Probability of a decrease							
	t^4	.533	.476	.735	.704	.668	.637	.794	.789
	t^6	.511	.503	.775	.753	.646	.662	.696	.656
	HP	.501	.492	.765	.738	.645	.642	.660	.626
	SPR	.470	.446	.637	.656	.529	.586	.639	.693
WTI_{real}									
		Probability of a decrease $>$ 1 s.d.							
	t^4	.034	.030	.014	.016	.008	.008	.010	.007
	t^6	.030	.029	.010	.013	.005	.006	.064	.090
	HP	.034	.032	.010	.015	.006	.008	.104	.125
	SPR	.002	.002	.002	.002	.001	.001	.002	.007
		Probability of a decrease							
	t^4	.461	.428	.739	.714	.666	.639	.856	.850
	t^6	.485	.468	.790	.765	.693	.677	.721	.736
	HP	.476	.462	.780	.750	.675	.656	.651	.675
	SPR	.526	.440	.715	.705	.606	.627	.643	.797
$Brent_{real}$									
		Probability of a decrease $>$ 1 s.d.							
	t^4	.033	.026	.015	.015	.008	.009	.027	.023
	t^6	.027	.026	.009	.013	.000	.007	.134	.215
	HP	.030	.029	.011	.015	.000	.008	.210	.294
	SPR	.002	.003	.002	.003	.001	.002	.007	.020

For the simulations-based approach (sims.) the truncation parameter $M = 100$ and 1 000 000 simulations were used. Standard deviations (s.d.) are calculated over the detrended samples.

4

A short term credibility index for central banks under inflation targeting: an application to Brazil

Adapted from: Alain Hecq, João Victor Issler, and Elisa Voisin (2022). *A short term credibility index for central banks under inflation targeting: an application to Brazil*. DOI: 10.48550/ARXIV.2205.00924.

Abstract

This chapter uses predictive densities obtained via mixed causal-noncausal autoregressive models to evaluate the statistical sustainability of Brazilian inflation targeting system with the tolerance bounds. The probabilities give an indication of the short-term credibility of the targeting system without requiring modelling people's beliefs. We employ receiver operating characteristic curves to determine the optimal probability threshold from which the bank is predicted to be credible. We also investigate the added value of including experts predictions of key macroeconomic variables.

4.1 Introduction

With the introduction of the Real Plan in July 1994, and later in 1999 with the *Inflation Targeting Regime*, actual Brazilian inflation has been substantially smaller compared to the previous hyperinflation period of the 1970s and 80s. Annual inflation computed with the Extended National Consumer Price Index (IPCA[1]), the reference variable for the Brazilian *Inflation-Targeting Regime*, is 6.07% on average between January 1997 and October 2020. Figure 4.1 displays the IPCA annual inflation rate for that period, together with the upper and lower tolerance bounds set within the regime. These tolerance bounds were 2 to 2.5 percentage points from the target inflation until 2017 and are now 1.5 percentage points above and below the target. With the noticeable exception of years 2001-2003 and 2015-2016, the Central Bank of Brazil (BCB) succeeded to ensure that the IPCA's annual inflation remains within or close to the tolerance interval.

The question that we want to answer in this chapter is whether, in any given month, it is sustainable for inflation to stay within the actual tolerance bounds set in advance by the *Inflation-Targeting Regime* of the Brazilian Central Bank (BCB). We do this by computing the conditional probability that actual inflation stays within the bounds in the near future (1- to 6-month ahead) using a forward-looking approach that allows current inflation to depend on future inflation as well on lagged inflation. This is the first original contribution of this paper.

[1]The IPCA targets population families with household income ranging from 1 to 40 minimum wages. This income range guarantees a 90% coverage of families living in 13 geographic zones: metropolitan areas of Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Vitória, Rio de Janeiro, São Paulo, Curitiba, Porto Alegre, as well as the Federal District and the cities of Goiânia and Campo Grande. Basket items include Food and Beverages, Housing, Household Articles, Wearing Apparel, Transportation, Health and Personal Care, Personal Expenses, Education and Communication.

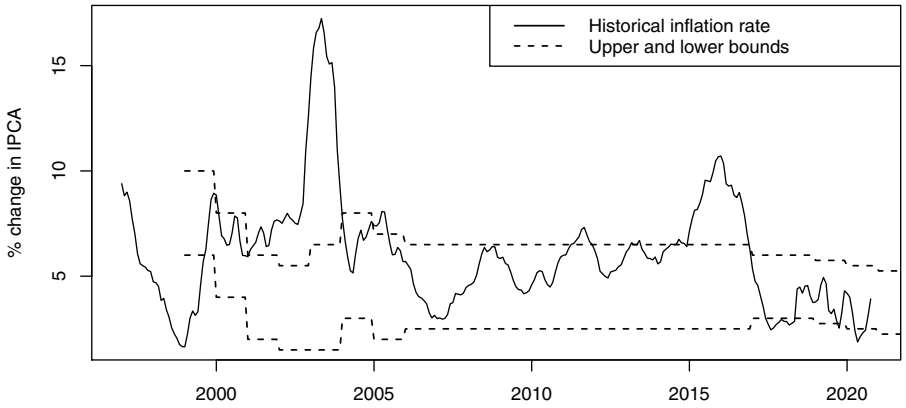


Figure 4.1: Annual inflation rate in Brazil and the target bounds

Applying the techniques discussed in this chapter to other countries is straightforward, as long as we are able to obtain their respective targets and bounds. So, the second original contribution of this chapter is to propose a new methodological approach to measure the credibility of central banks under an *Inflation-Targeting Regime*. Due to the nature of the problem we were set to answer, this methodology is related to short-term credibility as opposed to longer term credibility indices that have been proposed in the literature. Hence, we provide a complement to these techniques.

Although we employ a forward-looking approach, the information used to compute these probabilities across time only condition on current information. Indeed, we use a mixed causal-noncausal model – MAR(r, s)) – a process with a *lag* polynomial of order r (as in the usual autoregressive AR(r) model), but with a *lead* polynomial of order s as well. The introduction of a forward component makes sense for modelling inflation, since many economic models link current inflation to their future expected value which is captured by

the forward component introduced in the MAR(r, s) class. Moreover, this class also allows modelling parsimoniously nonlinear features in inflation, which the simpler linear setup misses. MAR models were introduced by Lanne, Luoto, and Saikkonen (2012) and refined by Gouriéroux and Jasiak (2016). Here we use the framework proposed in Hecq and Voisin (2021) and Hecq and Voisin (2022).

We do showcase the Brazilian experience, despite the fact that the techniques employed in this chapter have a broader application. Inflation has been a major problem for this country over the years and inflation has been in and out the bounds according to Figure 4.1. So, we would have two polar cases at hand: when the central bank policy had the desired effect on inflation and when it had not. Another interesting feature about Brazil is that, as far as we know, the BCB has the most complete data base on inflation expectations on earth – The *Focus* data base. It contains high frequency (daily) data on inflation expectations (also on other key macroeconomic variables) from current month all the way to 12 months ahead, and then on yearly intervals. This allows expectations data to be used to help forecast future inflation as a proxy for it, employing the MARX model proposed by Hecq, Issler, and Telg (2020), where strictly-exogeneous regressors are used in the model.

Most of the literature on central-bank credibility measures whether people expect or not that the central bank will meet their target; see Blinder (2000). This implies the need for modelling people's expectations regarding future inflation. As shown by Issler and Soares (2022), and the references therein, it is not straightforward to compute people's beliefs, and results can change substantially based on the method being employed. This chapter instead focuses on the actual inflation process and asks whether it will remain within the announced bounds. It does not require modelling people's beliefs. It is based on the econometric properties and the dynamics of inflation using a forward-looking approach.

Empirically, the chapter first identifies the MAR(r, s) model on IPCA annual inflation. We found that the 12-month inflation rate follows a MAR(1,1) process with the error term having a Student's t distribution with 3.25 degrees of freedom. This is our benchmark case, but we also experiment with alternative specifications. Using the approach in Hecq and Voisin (2021) and Hecq and Voisin (2022), and our benchmark specification, we evaluate the probabilities that the inflation rate stays on track in the future within the announced target bounds.

Finally, in our last original contribution, we compare the empirical results of our proposed short-term index to those of the literature by employing a receiver operating characteristic curve (ROC curve), something that is not common in the literature. For every measure of credibility, the ROC curve plots the true-positive and the false-positive rates for different probability thresholds used in classifying states (credible vs. non-credible). Our results show promise for the method propose in this chapter as a complement to the existing ones, which are focused on credibility at longer horizons.

The rest of the chapter is as follows. Section 4.2 provides a summary of the model and estimation methods used here. Section 4.3 summarizes the main methods that have been developed to forecast with MAR models. Section 4.4 provides the estimated probabilities for the inflation to stay within announced bounds at various horizons. In Section 4.5 we ask whether adding the information of experts forecasters from the *Focus* database helps forecasting the conditional probabilities computed here. Section 4.6 compares our short-term credibility measure with existing credibility indices presented in the literature. Section 4.7 concludes.

4.2 Mixed causal-noncausal models

4.2.1 Notation

An MAR(r, s) process y_t depends on its r lags as for usual autoregressive processes but also on its s leads in the following multiplicative form

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,$$

with L is the backshift operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. When $\Psi(L^{-1}) = (1 - \varphi_1 L^{-1} - ... - \varphi_s L^{-s}) = 1$, namely when $\varphi_1 = ... = \varphi_s = 0$, the process y_t is a purely causal autoregressive process, denoted AR(r,0) or simply AR(r) model $\Phi(L)y_t = \varepsilon_t$. The process is a purely noncausal AR($0, s$) model $\Psi(L^{-1})y_t = \varepsilon_t$, when $\phi_1 = ... = \phi_r = 0$ in $\Phi(L) = (1 - \phi_1 L - ... - \phi_r L^r)$. The roots of both the causal and noncausal polynomials are assumed to lie outside the unit circle, that is $\phi(z) = 0$ and $\varphi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series y_t admits a two-sided moving average (MA) representation $y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$, such that $\psi_j = 0$ for all $j < 0$ implies a purely causal process y_t (with respect to ε_t) and a purely noncausal model when $\psi_j = 0$ for all $j > 0$ (Lanne and Saikkonen, 2011). Error terms ε_t are assumed *i.i.d.* (and not only weak white noise) non-Gaussian to ensure the identifiability of the causal and the noncausal part (Breidt et al., 1991).

There is a increasing literature making use of MAR models; see among others Karapanagiotidis (2014), Hencic and Gouriéroux (2015), Gouriéroux and Jasiak (2016), Lof and Nyberg (2017), Hecq and Sun (2021), Bec, Nielsen, and Saïdi (2020a), Gourieroux, Jasiak, and Tong (2021), Gourieroux, Hencic, and Jasiak (2021).

4.2.2 Estimation results on IPCA

Let π_t denote the year-on-year inflation rate in Brazil at time t. The hybrid New Keynesian Phillips Curve (NKPC) regression is such as

$$\pi_t = \gamma_f \mathbb{E}_t[\pi_{t+1}] + \gamma_b \pi_{t-1} + \beta x_t + \epsilon_t,$$

where $\mathbb{E}_t[\cdot]$ the conditional expectation at time t, x_t is a measure for marginal costs, which is not directly observable (a potential proxy can be the output gap) and ϵ_t an $i.i.d.$ error term. Adding and subtracting $\gamma_f \pi_{t+1}$ and rearranging terms, gives

$$\pi_t = \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + \underbrace{\beta x_t + \gamma_f \left(\mathbb{E}_t[\pi_{t+1}] - \pi_{t+1}\right) + \epsilon_t}_{\equiv \eta_{t+1}},$$

where the newly defined disturbance term η_{t+1} consists of three different parts: (i) the expectation error $(\mathbb{E}_t[\pi_{t+1}] - \pi_{t+1})$ which is assumed $i.i.d.$ following the literature on rational expectations models, (ii) the marginal costs variable x_t and (iii) an $i.i.d.$ error ϵ_t. Hence, the time series properties of η_{t+1} will depend on those of x_t, but x_t is assumed to be adequately approximated by a finite-order autoregression (Lanne and Luoto, 2013). Subsequently, the newly obtained equation is divided by γ_f and lagged by one period to obtain $(1 - \gamma_f^{-1} L + \gamma_f^{-1}\gamma_b L^2)\pi_t = -\gamma_f^{-1}\eta_t$. Next, $a(z) \equiv (1 - \gamma_f^{-1} z + \gamma_f^{-1}\gamma_b z^2)$ can be written as the product of two polynomials, i.e., $a(z) = (1 - \phi z)(1 - \varphi^* z)$ with $|\phi| < 1$ and $|\varphi^*| > 1$ for plausible values of γ_f and γ_b leading to a stable mixed causal noncausal formulation. See Lanne and Luoto (2013) for details. The MAR models we consider in this chapter are not a direct mapping of the NKPC but rather an estimation of the dynamics emanating from the transformations of the Philips curve mentioned above.

We have used the MARX package developed by Hecq, Lieb, and Telg (2017b)[2] and found on the whole sample an MAR(1,1) with a Student's

[2]We look at different starting value to avoid the bimodality trap of the MAR(r,s)

t with 3.25 degrees of freedom. Standard errors computed as in Hecq, Lieb, and Telg (2016) are in brackets.

$$(1 - \underset{(0.035)}{0.58}\, L)(1 - \underset{(0.016)}{0.94}\, L^{-1})\pi_t = \varepsilon_t, \qquad \varepsilon_t \sim t(3.25) \tag{4.1}$$

To analyse the stability of the estimation we recursively estimate the orders of the MAR(*r*,*s*) model and the corresponding coefficients with an expanding window. The initial sample goes from January 1997 to April 2005 (100 data points) and the last one goes to January 2020 (277 data points). Note that at each point the model identified was an MAR(1,1), we hence only provide the graphs of the recursive estimates of the lag coefficient, the lead coefficient and the degrees of freedom of the Student's t distribution in Figure 4.2 with the corresponding 95% confidence interval.

We can see that adding data points does not affect the value of the estimated coefficients. Over the 177 points added, the value of the lag coefficients slightly decreases from 0.62 to 0.58 whereas the lead coefficient varies between 0.936 and 0.948 and stabilises around 0.944 towards the end of the sample. The degrees of freedom vary between 3 and 3.7 and converges towards 3.25 towards the end of the sample. Hence, the slightly different models estimated in real time for recursive out of sample forecasts will not significantly affect the calculated probabilities and they will therefore be comparable. Note that from the correlogram of the residuals we might detect a seasonal MA(12) component. This feature is probably due to the construction of the year on year monthly inflation. Over differencing at some seasonal frequencies can introduce that moving average pattern. We keep working with this annual series as this is the target variable considered by the central bank. For our investigation that component

estimated coefficients. This is not implemented in MARX so far. Note that Hecq, Issler, and Telg (2020) show how to estimate such a model directly without solving for the exogenous variable x_t. This will be estimated in Section 4.5.

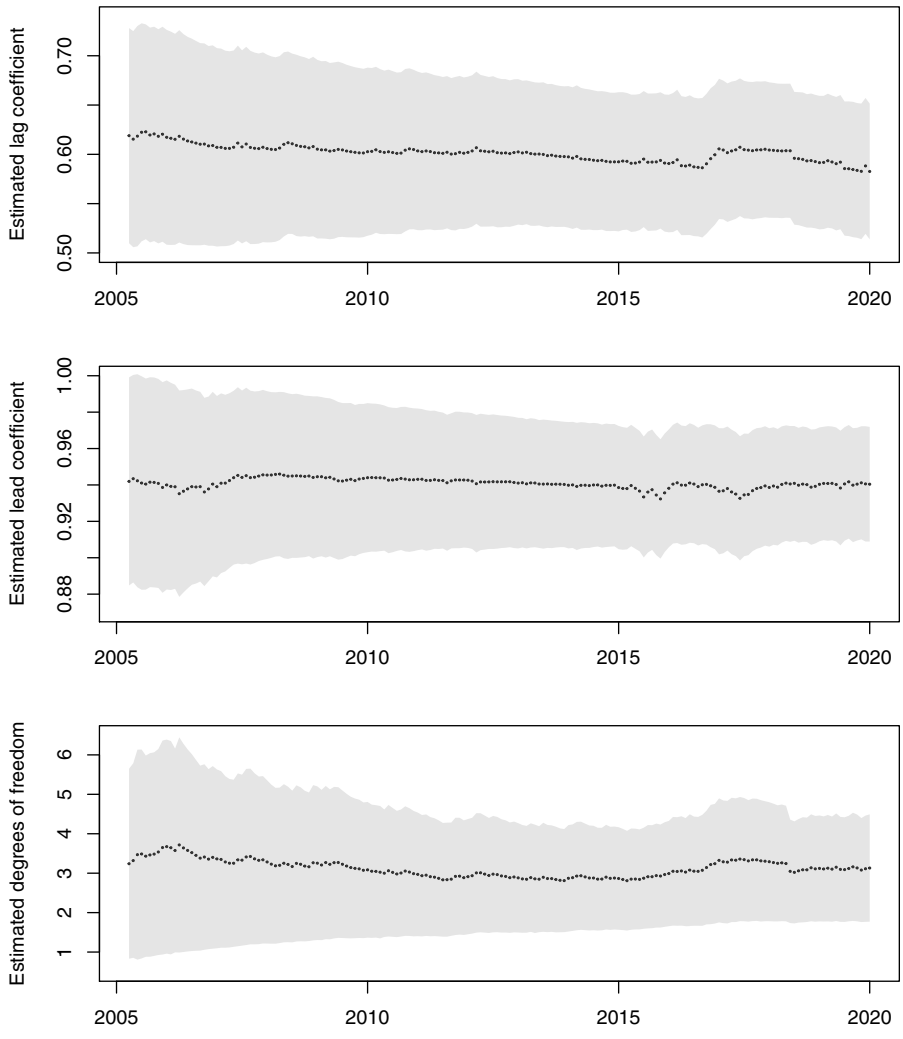


Figure 4.2: Recursive estimates of the coefficients based on end point of sample and their 95% C.I.

mainly impacts standard errors. Mixed models with MA components (genuine or spurious) are also out of the scope of this paper.

Expanding the multiplicative model (4.1) we obtain,

$$\pi_t = \frac{0.58}{1 + 0.58 \times 0.94}\pi_{t-1} + \frac{0.94}{1 + 0.58 \times 0.94}\pi_{t+1} + \frac{1}{1 + 0.58 \times 0.94}\varepsilon_t$$

$$= 0.38\pi_{t-1} + 0.61\pi_{t+1} + \varepsilon_t^*.$$

The obtained coefficients for the lag and lead are of the same magnitude as in the literature on the NKPC, and we can notice that they add up to close to unity, which is often a restriction imposed for model identification (see among others Gali, Gertler, and Lopez-Salido, 2005; Nason and Smith, 2008).

4.3 Predicting the probabilities to stay in the bounds

With Cauchy- and Levy-distributed errors, the conditional density of an MAR(*r*,*s*) process admits closed-form expressions, this is however not the case for Student's *t* distributed processes (Gouriéroux and Zakoïan, 2017). Assuming Student's *t* errors offers more flexibility than the Cauchy distribution, which might be too extreme, especially for inflation series that are not particularly volatile. Hence, in the absence of closed-form expressions for the predictive density, two approximations methods have been developed. The first one is based on simulations and was proposed by Lanne, Luoto, and Saikkonen (2012). The second one employs past realised values instead of simulations and was proposed by Gouriéroux and Jasiak (2016). However, as the latter becomes too computationally demanding when the forecast horizon increases, Gouriéroux and Jasiak (2016) proposed a Sampling Importance Resampling (henceforth SIR) algorithm facilitating longer horizon forecasts with this method. While the algorithm does not always work for extreme events, it provides

accurate results when the variables are rather stable, which is the case with the inflation series that we investigate here. Overall, both approaches use the decomposition of the mixed process into a causal and a noncausal component as such

$$\begin{aligned} u_t &\equiv \Phi(L)\pi_t \\ v_t &\equiv \Psi(L^{-1})\pi_t. \end{aligned} \tag{4.2}$$

The process u_t is the purely noncausal component of the errors, on which we will focus. In this analysis, since the inflation series is an MAR(1,1), the process u_t is a purely noncausal process of order 1,

$$\begin{aligned} \Psi(L^{-1})u_t &= \varepsilon_t \\ u_t &= \psi u_{t+1} + \varepsilon_t. \end{aligned} \tag{4.3}$$

4.3.1 Simulations-based approach

The purely noncausal component of the errors, u_t, can be expressed as an infinite sum of future error terms in its MA representation. The stationarity of the process ensures the existence of an integer M large enough to approximate this infinite representation as such (Lanne, Luoto, and Saikkonen, 2012),

$$u_t \approx \sum_{i=0}^{M} \psi^i \varepsilon_{t+i}, \tag{4.4}$$

Let lb_t and ub_t be the lower and upper bound for inflation in Brazil assessed for time t. We are interested in the conditional probabilities that inflation will be within the bounds at a given horizon h, where T

is the last observed point in the sample,

$$\begin{aligned}\mathbb{P}\Big(lb_{T+h} \leq \pi^*_{T+h} \leq ub_{T+h}|\mathcal{F}_T\Big) &= \mathbb{P}\Big(\pi^*_{T+h} \leq ub_{T+h}|\mathcal{F}_T\Big) - \mathbb{P}\Big(\pi^*_{T+h} \leq lb_{T+h}|\mathcal{F}_T\Big) \\ &= \mathbb{E}_T\Big[\mathbf{1}(\pi^*_{T+h} \leq ub_{T+h}) - \mathbf{1}(\pi^*_{T+h} \leq lb_{T+h})\Big]\end{aligned} \tag{4.5}$$

The indicator function $\mathbf{1}()$ is equal to 1 when the condition is met and 0 otherwise.

Since $\pi_t = \phi\pi_{t-1} + u_t$, by recursive substitution and using the approximation equation (4.4), we obtain,

$$\begin{aligned}\pi_{T+h} &= \phi^h \pi_T + \sum_{i=0}^{h} \phi^i u_{T+h-i} \\ &\approx \phi^h \pi_T + \sum_{i=0}^{h} \sum_{j=0}^{M-h-i} \phi^i \psi^j \varepsilon_{T+h-i+j},\end{aligned} \tag{4.6}$$

where M is the truncation parameter introduced in Equation (4.4). Substituting this approximation in (4.5), an approximation of the conditional probabilities is the following,

$$\begin{aligned}&\mathbb{P}\Big(lb_{T+h} \leq \pi^*_{T+h} \leq ub_{T+h}|\mathcal{F}_T\Big) \approx \\ &\qquad \mathbb{E}_T\Bigg[\mathbf{1}\Bigg(\pi_T + \sum_{i=0}^{h} \sum_{j=0}^{M-h-i} \phi^i \psi^j \varepsilon_{T+h-i+j} \leq ub_{T+h}\Bigg) \\ &\qquad\qquad - \mathbf{1}\Bigg(\pi_T + \sum_{i=0}^{h} \sum_{j=0}^{M-h-i} \phi^i \psi^j \varepsilon_{T+h-i+j} \leq lb_{T+h}\Bigg)\Bigg]\end{aligned} \tag{4.7}$$

Given the information set known at time T, the indicator functions in (4.7) are only functions of the M future errors, $\varepsilon^*_+ = \Big(\varepsilon^*_{T+1}, \ldots, \varepsilon^*_{T+M}\Big)$. Let $q(\varepsilon^*_+)$ be the function providing the value

of the difference between the two indicator functions. Furthermore, let $\varepsilon_{+}^{*(j)} = \left(\varepsilon_{T+1}^{*(j)}, \ldots, \varepsilon_{T+M}^{*(j)}\right)$, with $1 \leq j \leq N$, be the *j*-th simulated series of M independent errors, randomly drawn from the errors distribution, here a Student's *t*(3.25). Assuming that the number of simulations N and the truncation parameter M are large enough, the probability that the inflation rate (which follows an *MAR*(1,1) process) will remain within the bounds in h months can be approximated as such (Lanne, Luoto, and Saikkonen, 2012), [3]

$$\begin{aligned}\mathbb{P}\Big(lb_{T+h} \leq \pi_{T+h}^{*} \leq ub_{T+h} | \mathcal{F}_T\Big) &\approx \mathbb{E}_T\left[q(\varepsilon_{+}^{*})\right] \\ &\approx \frac{\sum_{j=1}^{N} q\Big(\varepsilon_{+}^{*(j)}\Big) g\Big(u_T - \sum_{i=1}^{M} \psi^i \varepsilon_{T+i}^{*(j)}\Big)}{\sum_{j=1}^{N} g\Big(u_T - \sum_{i=1}^{M} \psi^i \varepsilon_{T+i}^{*(j)}\Big)},\end{aligned} \tag{4.8}$$

where g is the *pdf* of the Student's *t*(3.25) distribution.

Hecq and Voisin (2021) results show that with Cauchy-distributed errors, this approach is a good estimator of theoretical probabilities but are significantly sensitive to the number of simulations N during locally explosive episodes. For Student's *t* distributions however, results cannot be compared to theoretical ones, but as the number of simulations gets larger, the derived densities converge to a unique function. Overall Hecq and Voisin (2021) show that: *(i)* for degrees of freedom close to 3 and *(ii)* during stable episodes, this approach yields consistent results that are not significantly sensitive to the number of simulations, as long as it is reasonably large. Hence, choosing the right number of simulations per iteration should not be a worry in this analysis.

[3] See Section 2.4.1 for more detailed derivations of the estimator.

4.3.2 Sample-based approach

As an alternative to using simulations, Gouriéroux and Jasiak (2016) employ all past observed values of the process to approximate the marginal distributions of noncausal processes. They propose the following sample-based approximation of the predictive density of an MAR(0,1),[4]

$$\begin{aligned} &l(u^*_{T+1},\dots,u^*_{T+h}|\mathcal{F}_T) \\ &\approx g(u_T-\psi u^*_{T+1})\dots g(u^*_{T+h-1}-\psi u^*_{T+h})\frac{\sum_{i=2}^{T} g(u^*_{T+h}-\psi u_i)}{\sum_{i=2}^{T} g(u_T-\psi u_i)}, \end{aligned} \tag{4.9}$$

where g is the *pdf* of a Student's t distribution with 3.25 degrees of freedom.

Given a correctly identified model and based on the equivalence of the information sets $(\pi_1,\dots,\pi_T,\pi^*_{T+1},\dots,\pi^*_{T+h})$ and $(v_1,\varepsilon_2,\dots,\varepsilon_{T-1},u_T,u^*_{T+1},\dots,u^*_{T+h})$ (Gouriéroux and Jasiak, 2016), where $v_t=\pi_t-\psi\pi_{t+1}$, the predictive density of the *MAR*(1,1) process π_t can be obtained by substituting the filtered noncausal process u_t by the mixed process π_t in (4.9),

$$\begin{aligned} l(\pi^*_{T+1},\dots,\pi^*_{T+h}|\mathcal{F}_T) \approx\ & g((\pi_T-\phi\pi_{T-1})-\psi(\pi^*_{T+1}-\phi\pi_T))\times\dots \\ &\dots\times g((\pi^*_{T+h-1}-\phi\pi^*_{T+h-2})-\psi(\pi^*_{T+h}-\phi\pi^*_{T+h-1})) \\ &\times\frac{\sum_{i=2}^{T} g(\pi^*_{T+h}-\phi\pi^*_{T+h-1}-\psi(\pi_i-\phi\pi_{i-1}))}{\sum_{i=2}^{T} g(\pi_T-\phi\pi_{T-1}-\psi(\pi_i-\phi\pi_{i-1}))}. \end{aligned} \tag{4.10}$$

Evidently, evaluating the conditional joint density over all possible outcomes becomes considerably computationally demanding as the forecast horizon increases. This is why Gouriéroux and Jasiak (2016)

[4] See Section 2.4.2 for more detailed derivations of the estimator.

developed a Sampling Importance Resampling (henceforth SIR) algorithm to counter this computational limitation. We provide details and describe the algorithm in the subsequent Section. We will therefore employ estimator (4.10) for a forecast horizon of 1 (see (4.11)) and will use the SIR algorithm for horizons of 3 and 6 months.

$$l(\pi^*_{T+1}|\mathcal{F}_T) \approx g\Big(\pi_T - \phi\pi_{T-1} - \psi(\pi^*_{T+1} - \phi\pi_T)\Big) \times \frac{\sum_{i=2}^{T} g\Big(\pi^*_{T+1} - \phi\pi^*_T - \psi(\pi_i - \phi\pi_{i-1})\Big)}{\sum_{i=2}^{T} g\Big(\pi_T - \phi\pi_{T-1} - \psi(\pi_i - \phi\pi_{i-1})\Big)}. \tag{4.11}$$

This estimator provides predicted probabilities that are a combination of theoretical probabilities and probabilities induced by past events; results are therefore case-specific and are based on a learning mechanism (Hecq and Voisin, 2021). For values close to the median this approach provides accurate and similar results to theoretical probabilities (when available) and to the simulations-based method of Lanne, Luoto, and Saikkonen (2012) for one-step ahead forecasts. Discrepancies widen as the level of the series increases.

4.3.3 Sampling Importance Resampling algorithm

As previously mentioned, estimator (4.10) developed by Gouriéroux and Jasiak (2016) is substantially computationally demanding to employ for long forecast horizons and simulating from this distribution is rather intricate. The authors then proposed a SIR algorithm to counter this. The algorithm consists in simulating potential paths of future noncausal components u_t's from an instrumental misspecified model from which it is easier to simulate. The distribution (4.10) of interest is then recovered using a weighted resampling of the simulations and the relation (4.3) between the inflation rate π_t and its noncausal component u_t. Gouriéroux and Zakoïan (2013) note that a Markov process

in reverse time is also a Markov process – of the same order – in calendar time with non-linear dynamics. Hence, since in this analysis u_t is a non-causal MAR(0,1) process, it could be expressed as a causal AR(1) process, with non-linear dynamics. For a study on misspecified causal analysis of noncausal processes see Gouriéroux and Jasiak (2018). Following Gouriéroux and Jasiak (2016), we employ a Gaussian AR(1) model as instrumental model for the algorithm,

$$u_t = \tilde{\rho} u_{t-1} + \tilde{\varepsilon}_t. \tag{4.12}$$

The parameter $\tilde{\rho}$ is estimated using standard OLS on the observed values u_t filtered from the initial $MAR(1,1)$ process π_t. The errors $\tilde{\varepsilon}_t \sim IIN(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the MAR residuals variance, $\tilde{\cdot}$ indicates estimation from the instrumental model, and $\hat{\cdot}$ from the initial $MAR(1,1)$ model. The conditional predictive density for the instrumental process is as follows,

$$\begin{aligned}
&\tilde{F}(u^*_{T+1}, \dots, u^*_{T+H} | u_T) \\
&= \tilde{l}(u^*_{T+H} | u^*_{T+H-1}) \tilde{l}(u^*_{T+H-1} | u^*_{T+H-2}) \dots \tilde{l}(u^*_{T+1} | u_T) \\
&= f(u^*_{T+H} - \tilde{\rho} u^*_{T+H-1}) f(u^*_{T+H-1} - \tilde{\rho} u^*_{T+H-2}) \dots f(u^*_{T+1} - \tilde{\rho} u_T),
\end{aligned} \tag{4.13}$$

where $\tilde{F}$ is the predictive conditional distribution of h future u_t's from the instrumental model and f the *pdf* of a normal distribution with mean zero and variance $\hat{\sigma}^2$. Even though this model is clearly misspecified, the resampling step should automatically correct for the induced misspecifications (Gouriéroux and Jasiak, 2016). The algorithm for h-step ahead predictions is as follows (see also Gourieroux, Hencic, and Jasiak, 2021),

1. **Sampling:** Draw K series of h independent values $\tilde{\varepsilon}$ from a normal distribution $N(0, \hat{\sigma}^2)$. Using the recursive equation (4.12) and the last observed value u_T, compute the K simulated paths $(\tilde{u}^i_{T+1}, \dots, \tilde{u}^i_{T+H})$, respectively stacked in $\tilde{U}^i$'s, $i = 1, \dots, K$.

2. **Importance:** Denote $\Pi(\tilde{U}^i)$ the conditional density (4.10) evaluated at the path $(\tilde{u}^i_{T+1}, \ldots, \tilde{u}^i_{T+H})$. Compute the weights $w_i = \hat{\Pi}(\tilde{U}^i)/\tilde{F}(\tilde{U}^i)$ with $i = 1, \ldots, K$. Namely, the ratio between the value of the density the algorithm intends to recover and the value of the instrumental density.

3. **Resampling:** Draw with probability weighting based on the previously computed weights and replacement S paths $(\tilde{u}^s_{T+1}, \ldots, \tilde{u}^s_{T+H})$, $s = 1, \ldots, S$, from the K simulated series in the sampling step, with respective weights w_i.

Once the set of S re-sampled $\tilde{U}^{i\prime}$s is obtained, each simulated path $(\tilde{u}^s_{T+1}, \ldots, \tilde{u}^s_{T+H})$, $s = 1, \ldots, S$ can be transformed into the corresponding future path for the variable of interest $(\pi^*_{T+1}, \ldots, \pi^*_{T+H})$ using on the causal relation in Equation (4.3).

As stated by Gouriéroux and Jasiak (2016), if the number of initial simulations K is large enough, simulating from the instrumental Gaussian model and applying the SIR algorithm is equivalent to simulating directly from the distribution of interest. It avoids simulating from a too complicated distribution, or avoids the simulation of too many extreme values. Hecq and Voisin (2021) find that this approach may not function well during extreme events as the instrumental density becomes too different from the one to recover. The instrumental density (derived from normally distributed errors) is unimodal while the predictive density of a purely noncausal process during an explosive episode is bi-modal with a significantly larger range. This implies that during extreme events the instrumental conditional density is sometimes zero where the target distribution is not. This implies that during extreme episode the distribution derived from the algorithm may not converge to the target one since parts of the distribution are not simulated in the first step of the algorithm, regardless of the number of simulations. Further research should be done to improve the algorithm in such cases. However, we focus here on stable periods and this

limit of the algorithm does therefore not affect our analysis.

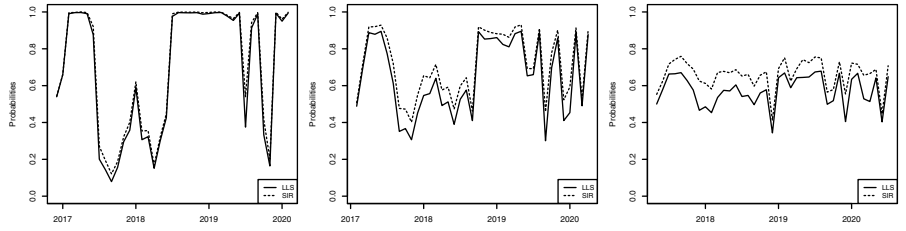
4.4 Forecasting with MAR model

Using the model estimated in Section 4.2, we perform pseudo-real-time 1-, 3- and 6-months forecasts of the probabilities to remain within the bounds with expanding window from November 2016 to January 2020. We hence put emphasis on short-term predictions which could provide warnings that the year-on-year inflation will be outside the target bounds in the near future. We first analyse the impact of the choice of forecasting method as well as the impact of the forecast horizon on the probabilities. As shown previously in Figure 4.2, the estimation of the model is stable and particularly so after 2016. Therefore, re-estimating the model with expanding window at each point of forecasts does not have a significant impact on the results and represents a good real-time forecasting analysis. We can hence obtain 39 pseudo real-time probability forecasts for each of the three horizons and each of the two methods.

Figure 4.3 depicts the differences between the predictions made with the simulations based method (solid line) of Lanne, Luoto, and Saikkonen (2012)[5], and the SIR algorithm (dashed line) employing the sample-based method of Gouriéroux and Jasiak (2016)[6]. Graph (a) shows the 1-month ahead forecasts performed with increasing sample at each point. Graph (b) and (c) correspond to 3- and 6-months ahead forecasts respectively. All forecasts are performed at the same 39 points, starting in November 2016. Note that as the analysis does not regard extreme episodes, the 1-step ahead density forecasts obtained with the SIR algorithm and the sample-based method of Gouriéroux and Jasiak (2016) were identical[7]. For a thorough

[5] Employing 1 000 000 simulations at each iteration.
[6] Employing 100 000 simulations in the first step and 10 000 resampling forecasts.
[7] Results available upon request.

analysis of their respective performance see Hecq and Voisin, 2021. Since they perform similarly and because the SIR algorithm is less computationally demanding we only present the SIR results alongside the simulations-based results (LLS).



(a) 1-month ahead (b) 3-months ahead (c) 6-months ahead

Figure 4.3: Probability forecasts for inflation to remain within the bounds

We can see that the differences between the two methods increase with the horizon. For one-step ahead forecasts, both approaches yield almost the same results at each point in time, while for the 6-months ahead forecasts they differ on average by 0.09, with the SIR probabilities always larger than LLS. This indicates that LLS predicts more volatile fluctuations in the inflation rate than the SIR for longer horizons, hence suggesting lower probabilities to remain within the bounds. Furthermore, we can see that while probabilities vary significantly at a 1-month horizon (from 0.08 to 1), they converge around 0.6 for both methods as the horizon increases to 6 months. That is, as the forecast horizon increases, the uncertainty prevails, regardless of being within or outside the bounds at the moment of the forecast, which yields almost 50/50 chance to meet the target at at 6-months horizon.[8] This is because we are investigating a period during which there is no explosive episodes. Inflation is relatively stable though close to the lower bound of the target.

[8] Table 4.1 in Appendix A summarizes the results.

4.5 MARX model

4.5.1 Model representation

MAR models with additional strictly exogenous variables have been introduced by Hecq, Issler, and Telg (2020). The so called MARX model allows to estimate the effects of covariates without aggregating them in the error term as in the MAR(r, s) model of Section 4.2. The MARX(r, s, q) for a stationary time series π_t (here, the annual IPCA inflation rate) reads as follows

$$\phi(L)\varphi(L^{-1})\pi_t - \beta' X_t = \varepsilon_t, \tag{4.14}$$

where $\phi(L)$ and $\varphi(L^{-1})$ are the lag and lead polynomials of order r and s with $r + s = p$ and q is the number of strictly exogenous variables. We still assume that the roots of both polynomials lie outside the unit circle. When $q = 0$, the process reduces to a standard MAR(r, s). In case $q > 0$, the process no longer has a strictly stationary solution solely in terms of ε_t, but involves additionally X_t (consisting of q exogenous variables). That is,

$$\pi_t = \gamma(L, L^{-1})\varepsilon_t + \gamma(L, L^{-1})\beta' X_t = \sum_{j=-\infty}^{\infty} \gamma_j z_{t-j}, \tag{4.15}$$

where $z_{t-j} = \varepsilon_{t-j} + \sum_{i=1}^{q} \beta_i x_{i,t-j}$ and $\gamma(L, L^{-1})$ is an operator satisfying $\gamma(L, L^{-1})\phi(L)\varphi(L^{-1}) = 1$ such that π_t has a two-sided MA-representation augmented with past, current and future values of X_t. Consequently, denoting $\tilde{x}_{i,t} = \beta_i x_{i,t}$, π_t consists of a two-sided MA representation and the sum of q processes $\tilde{x}_{i,t}$ that are passed through a two-sided linear filter with coefficients resulting from inverting the product $[\phi(L)\varphi(L^{-1})]$. Similarly to the MAR, the MARX can be decomposed in u and v components, namely the noncausal and causal components respectively,

$$u_t \equiv \phi(L)\pi_t \leftrightarrow \varphi(L^{-1})u_t - \beta' X_t = \varepsilon_t, \tag{4.16}$$

$$v_t \equiv \varphi(L^{-1})\pi_t \leftrightarrow \phi(L)v_t - \beta' X_t = \varepsilon_t. \tag{4.17}$$

This (u, v) representation is useful to simulate variable and to forecast MAR processes. See Hecq, Issler, and Telg (2020) for details on the model.

4.5.2 Data

We evaluate to what extent key drivers of inflation identified by the BCB, have an influence on the probability that inflation stays within the target bounds. The variables we consider are the year-on-year percentage change in industrial production[9] (ip_t), the year-on-year percentage change in the Real/US$ exchange rate (ex_t) where the rate refers to the last working day of the period and the year-on-year percentage change in the Selic target interest rate (ir_t). We denote $X_t = (ip_t, ex_t, ir_t)$ the stacked stationary exogenous variables at time t. We use published indicators of the three variables (retrieved from FRED and from the Central Bank of Brazil databases). For coherence and comparison purpose we use the same sample as for Section 4.2, namely from January 1997 to January 2020.

4.5.3 Estimation results

The aim of this section is to analyse the added value of augmenting the MAR(1,1) model identified in Section 4.4 with exogenous variables. Hence, we identify the MARX(1,1,q) model best fitting the data, using the strategy proposed in Hecq, Issler, and Telg (2020). We first estimate an ARDL model using a maximum likelihood approach and find that BIC favors an ARDL(2,0,0,0), namely an AR(2) with only contemporaneous values of the three regressors. The increase of the goodness of

[9] We prefer to use the industrial production index as a measure of economic activity over the GDP that is available quarterly.

fit is however not substantial, the $\bar{R}^2$ increases from 0.977 to 0.980. The estimated model is the following (HCSE standard errors in brackets),

$$\pi_t = \underset{(.05)}{1.46}\pi_{t-1} - \underset{(.05)}{0.50}\pi_{t-2} + \underset{(.56)}{2.21}ip_t + \underset{(.17)}{0.87}ex_t + \underset{(.09)}{0.23}ir_t + \eta_t.$$

All three regressors have a significant impact on annual inflation and only their contemporaneous values are selected. The value of the Jarque and Bera normality test is 185.3 with a $p - value < 0.0001$.

Following the methodology described in Hecq, Issler, and Telg (2020) we compare all MARX(1,1,3) considering all possible time indices of the exogenous variables (namely their lag, lead or contemporaneous value). The model maximizing the likelihood function is the following MARX(1,1,3) with a $t(3.39)$ (standard errors in brackets),

$$(1 - \underset{(.04)}{0.50}L)(1 - \underset{(.01)}{0.97}L^{-1})\pi_t = -\underset{(.38)}{1.64}ip_{t+1} - \underset{(.09)}{0.53}ex_{t+1} - \underset{(.08)}{0.04}ir_{t+1} + \varepsilon_t. \tag{4.18}$$

That is, the model maximizing the likelihood function includes the lead of the three exogenous variables. The year-on-year inflation rate is therefore influenced by anticipations in the key economic variables considered here. The coefficients might not represent actual effects but instead corrections of the effect already captured by the coefficient on the lead of the inflation rate. We do not focus on that.[10] We can see that the lag coefficient has slightly decreased (by 0.04) while the lead coefficient is now slightly higher (by 0.03).

[10] We acknowledge the fact that these regressors might not be fully strictly exogenous. Yet, we provide this analysis to illustrate the use of MARX models and take advantage of the availability of frequent forecasts for these variables which enables using MARX models to forecast without needing to predict the exogenous variables separately.

4.5.4 Predictions with Focus data

Hecq, Issler, and Telg (2020) find that the forecasts obtained in a MARX are superior to the ones of a MAR model when the future values of the regressors are known. The gain diminishes when the regressors must be forecasted as well, using an ARMA model for instance. The choice of exogenous variables in this analysis was motivated by the fact that those variables are forecasted and updated on a daily basis by experts in the Focus database maintained by the Brazilian Central Bank.[11] Previous periods are re-evaluated until the official numbers are published and forecasts for future periods are available at various horizons. This allows to perform forecasts using MARX models without modelling and forecasting separately the exogenous variables.

In this study, when performing predictions, we take the overall median of experts' forecasts for the future values of X in real time. This means for instance that in May 2019 we take the forecasts of the explanatory variables for the next months made at the end of May 2019. Furthermore, to ensure coherence in the data, we also replace the last three data points (in this example we replace March, April and May 2019) by the last vintages available at the point at which the forecast is preformed.

Alike the MARX process π_t, the MA representation of the noncausal component u_t is also augmented by contemporaneous and future val-

[11] See https://www.bcb.gov.br/en/monetarypolicy/marketexpectations. The Brazilian survey on economic forecasters is unique. Everyday, a set of experts (from banks, fund managers, brokers, consulting companies, etc) give their evaluation on the future of inflation rate as well as for several key macroeconomic variables regarding the Brazilian economy. Most variables are published before the release of the Brazilian consumer price index, which is made public around the 10th of the subsequent month. Some variables have some delays and they also often continue to be forecasted for several vintages after the end of the corresponding month.

ues of X_t,

$$\begin{aligned} u_t &= \varphi(L^{-1})^{-1}[\beta' X_t + \varepsilon_t] \\ &= \sum_{i=0}^{\infty} \psi^i [\beta' X_{t+i} + \varepsilon_{t+i}]. \end{aligned} \tag{4.19}$$

In analogy to Equation (4.4), the method of Lanne, Luoto, and Saikkonen (2012) – if possibly extended to MARX models – would require a truncation parameter M large enough to approximate (4.19). This implies that M future values of the exogenous variables would be needed to perform forecasts with such approach. Lanne, Luoto, and Saikkonen (2012) suggest using M=50 which represents more than four years of monthly predictions.

On the other hand, the method of Gouriéroux and Jasiak (2016) only requires as many predictions of the exogenous variables as the forecast horizon. This makes this approach the most suitable for this analysis. As explained in Section 4.4, for one-step ahead forecasts, the estimator (4.11) can be employed, but as the forecast horizon increases the SIR algorithm alleviates the computational limitations of the estimator. The SIR approach simulates the noncausal component of the process using a purely causal instrumental model and then transforms those simulations into simulations of the variable of interest using the causal relation between π_t and u_t described in (4.2).

We use the following instrumental model for the SIR algorithm,

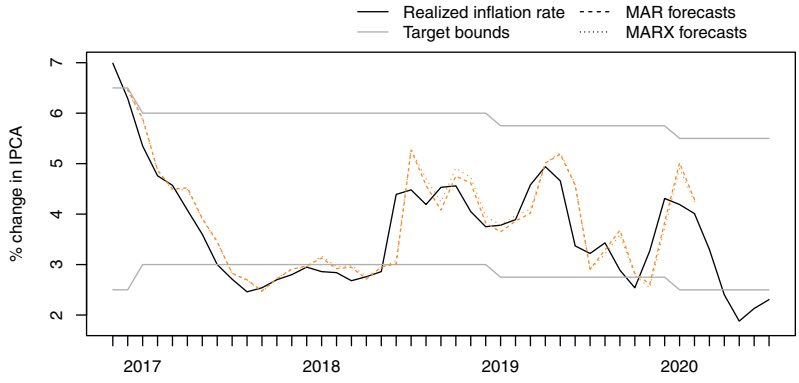
$$u_t = \rho u_{t-1} + \eta_1 ip_t + \eta_2 ex_t + \eta_3 ir_t + \epsilon_t, \tag{4.20}$$

where $\epsilon_t \sim N(0, \sigma^2)$. This instrumental model is the pseudo causal model obtained by inverting the time indices in the noncausal representation of u_t. The parameters are obtained using standard OLS on the whole sample.

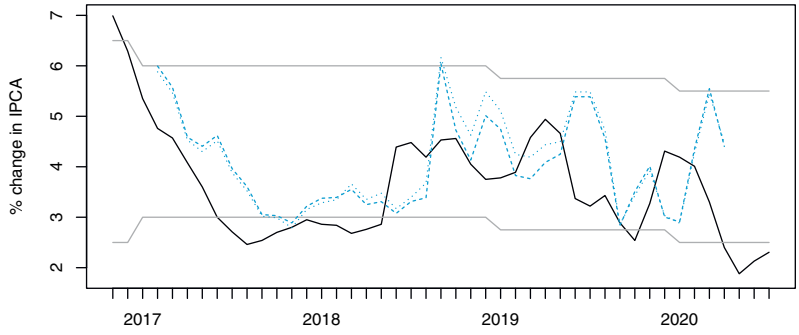
As mentioned, to resemble a real time forecast, we replace the values for the last three observations (from T-2 to T) by the evaluations made by experts for the exogenous variables at time T, and we take the six consecutive months forecasts made at the same point in time. Alike Section 4.4, we perform 1-, 3- and 6-months ahead forecasts at the end of each month from November 2016 to January 2020. The SIR algorithm is employed the same way as for the MAR. We re-estimate the model every time, yet we fix the time indices of the variables to be the ones in model (4.18).[12]

Figure 4.4 shows point forecasts obtained from the MAR model of Section 4.4 (dashed lines) compared to the point forecasts obtained with the MARX model (dotted lines). Forecasts at the different horizons are compared to realized inflation rate (black solid line) and the announced target bounds (grey solid lines). We can see that the inclusion of exogenous variables does not significantly alter the predictions, especially for short-term horizons. An explanation to this is that forecasts of yearly changes in the exogenous variables around that time were very close to 0 as we are investigating a stable period. We can however notice that when the inclusion of exogenous variables has an impact, it becomes increasingly noticeable at larger horizons. We can therefore expect that during unstable and more volatile periods the inclusion of exogenous variables should influence more significantly forecasts and predictive densities. Another explanation is that inflation might already be influenced by expectations of the exogenous variables. Hence, their inclusion might most of the time not be enough to significantly alter predictions.

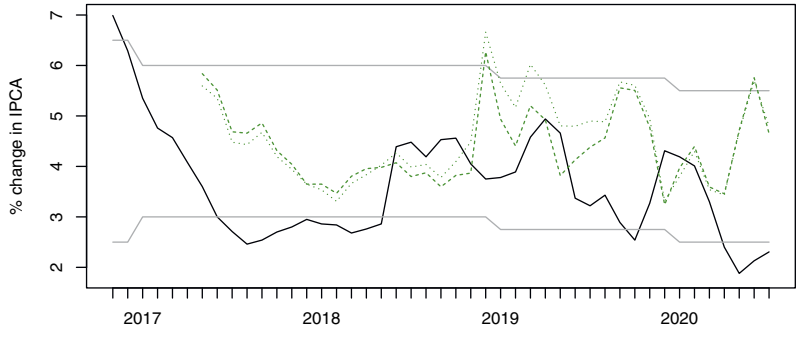
[12] Note that estimation is stable for all coefficients and predictions are thus easily comparable from one point in time to another.



(a) 1-month ahead



(b) 3-month ahead



(c) 6-month ahead

Figure 4.4: MAR vs MARX

4.6 Short-term credibility

There is a long tradition of evaluating the credibility of central banks. Svensson (2000) for instance measures credibility as the distance between expected and targeted inflation. A commonly used definition of central bank credibility is the following: *a central bank is credible if people believe it will do what it says* (Blinder, 2000). This definition however implies the need for modelling people's beliefs, which is not straightforward and can induce significantly different results based on the assumptions and model construction (see the reference list in Issler and Soares, 2022). Some construct the credibility as an inverse function of the gap between expectation and target, using different ad-hoc thresholds for what is considered credible or not (see among others Cecchetti and Krause, 2002 and Mendonça and Souza, 2007). Bomfim and Rudebusch (2000) construct expected inflation as a weighted average of the target and past inflation rate and interpret credibility as the weight of the latter. Dovern, Fritsche, and Slacalek (2012) use professional forecasters' predictions and interpret the discrepancies between them as an indication of lack of credibility. Issler and Soares (2022) propose a bias-corrected measure of inflation expectation using survey data and construct a credibility index based on whether the target falls within the confidence interval of their expectation measure. Most credibility measures are constructed for long-term horizons, such as 12 months ahead for instance, so that short-term shocks to inflation vanish.

We take a different stand-point and measure credibility not from the perspective of people' beliefs but from the dynamics in past inflation rates, building on the forward-looking characteristics of MAR models. That is, we use the probabilities that yearly inflation remains within the target bounds as an indication of whether the Central bank's target is currently credible or not. This measure therefore only relies on the statistical and dynamic properties of inflation and not on people's beliefs. To retain and employ all the dynamics

in realised inflation rates, we use monthly year-on-year inflation rate, which implies that our forecasts are based on shorter time horizons than the one-year-ahead expectations used in most of the aforementioned methods to build the credibility indices. We therefore measure short-term credibility and not long-term, which explains the discrepancies obtained with other approaches. Indeed, short-term predictions carry more shocks than one-year ahead forecasts. Our approach can be used in real time, here on a monthly basis, as an early warning that yearly inflation will exit the bounds in the near future.

For comparison purposes we use the same time span as the analysis of Issler and Soares (2022), namely from January 2007 to April 2017. Figure 4.5 compares our short-term credibility measurement with the longer-term credibility index proposed in Issler and Soares (2022) – denoted as *IS*. The short-term credibility index is the 1-month ahead simulations-based probabilities (LLS) that yearly inflation will remain within the target bounds (the red solid line), derived from the forward-looking MAR model. The discrepancies between the two methods stem from the horizon and perspective of each index. Recall that the long-term index is constructed a year ahead, that is, in 2013 for instance, people did not believe the Central bank would be credible in 2014. However, on a shorter-term basis, it seemed much more likely that the target would be met, hence a much larger short-term index. Our short-term approach does not investigate whether people trust the central bank to reach its goals in a year but instead whether it seems likely to happen at short horizons based on realized inflation rates. This makes the two approaches complementary.

To provide a better understanding of the indices we construct the receiver operating characteristic (ROC) curves. It represents the accuracy of the credibility measures as the threshold from which we consider the Central bank to be credible varies. That is, for a given threshold x and a given forecasted credibility index, the Central bank

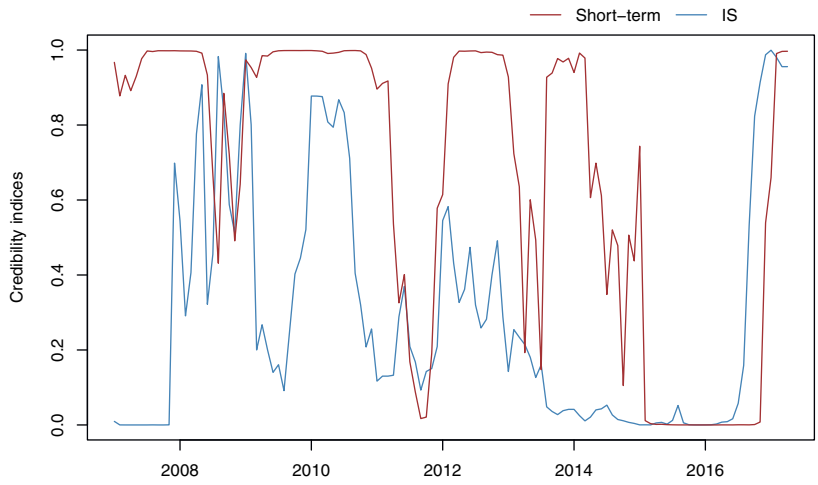


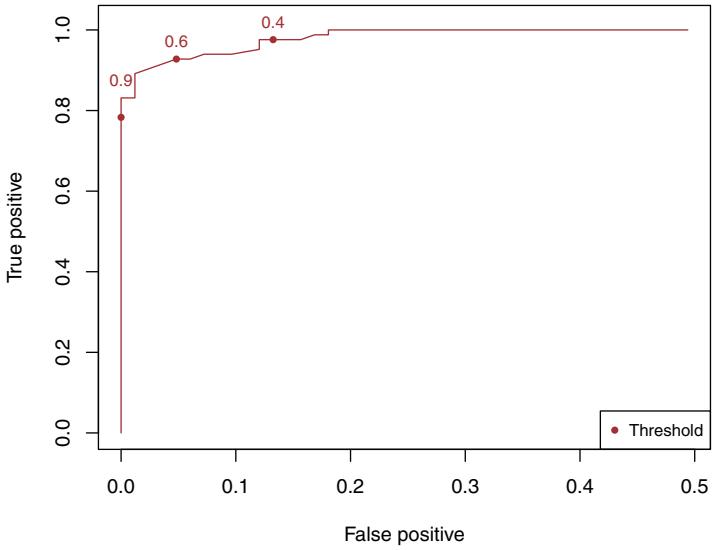
Figure 4.5: Comparison credibility indices

is predicted to be credible if the index is above x and we compare it to the actual inflation rate in the same month. We can then compute the rates of true and false positives for each given threshold and index. A true positive is when the bank is predicted to be credible and the actual inflation rates turns out to be within the tolerance bounds and a false positive is when the bank was predicted to be credible but the inflation rate ended up outside the bounds. We consider the long term indices that part of the analysis by Issler and Soares (2022), namely the index they construct (IS), *DM* is the index developed by Mendonça and Souza (2007), *CK* the index by Cecchetti and Krause (2002), *DGMS* by Mendonça and Souza (2009) and *LL* by Levieuge, Lucotte, and Ringuedé (2018). Figure 4.6 depicts the ROC curves for each of the indices. Graph (a) is our short-term index and graph (b) is for the long-term indices. It was expected to find that the short-term approach would outperform long-term forecasts in detecting when the inflation rate will exit the target bounds since predictions are made for a much closer horizon. For all indices, the ROC curves go

from the top right corner when the threshold is set to 0 and to the bottom left corner when the threshold is set to 1. The sensitivity of each index to changes in the thresholds in between those extremes vary significantly from one to another. We see on graph (a) for the short-term index that thresholds between 0.4 and 0.9 provide a true positive rate above 0.8 and a false positive rate of less than 0.15. Hence this approach is not particularly sensitive to the choice of threshold for accurate results. The choice of threshold then only depends on the preference of the practitioner regarding the wanted true and false positive rates. We can however notice that long-term indices are much more sensitive to the choice of thresholds. We only depict the points for thresholds 0.4 and 0.9, but we can first notice how, for most methods, changes in thresholds in between those values will yield a very different results accuracy. Naturally, the longer horizon of these predictions imply a loss in precision, however, the results across long term indices are noticeably different. Some methods tend to overestimate the credibility, while others tend to underestimate it. If we compare our short-term index with the IS long-term index, which is not dependent on ad-hoc benchmarks or thresholds, we notice that people believe the Central bank to be much less credible that it actually is a year ahead. Indeed, any probability threshold in the ROC curve above 0.4 yield rather low true positive rates.

Overall, while our short-term credibility index might be affected by short-term shocks in the inflation rate, it does not require any modelling of people's beliefs, or does not require any ad-hoc decisions of benchmarks. It can serve as an early warning of exiting the target bounds. The different perspective that our measure takes makes it complementary to longer-horizons credibility indices. Indeed, the long term credibility index built from people's beliefs combined with short-term probabilities to meet the target provide a clearer and more complete picture of the reliability of inflation targeting system.

(a) Short-term index



(b) Long term indices

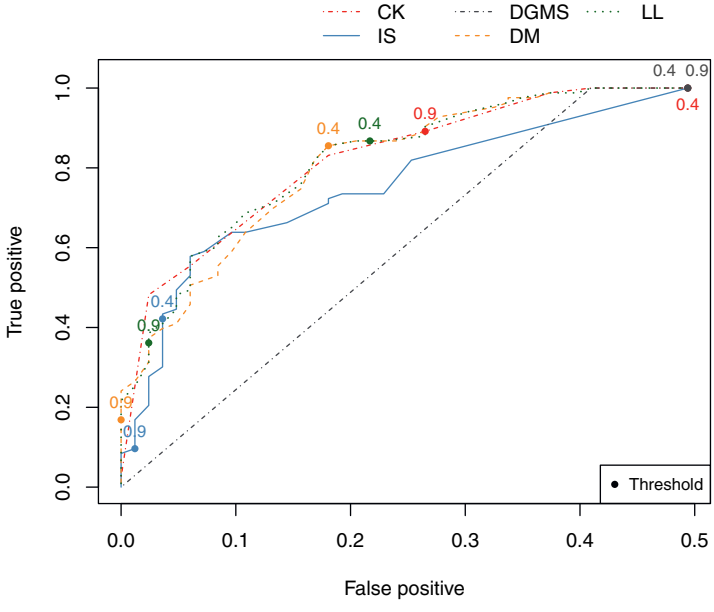


Figure 4.6: ROC curves

4.7 Conclusion

This chapter investigates the probabilities that the Brazilian inflation rate remains within the target bounds determined by the central bank. We estimate mixed causal-noncausal models with leads and lags of the inflation rate and the first noticeable results are the stability of the parameters estimation over time. Contrary to usual practice when employing mixed causal noncausal models, which consists in forecasting bubble bursts, we choose the stable period between 2017 and 2020 for the forecasting exercise. This allows us to forecast at farther horizons using a sampling importance resampling algorithm. We perform 1-, 3- and 6-months ahead probability forecasts. For short-horizons forecasts we correctly track probabilities to stay in the target bounds and interpret those results as the short-term credibility of the inflation targeting system based on historical inflation without the need to model people's beliefs. In longer horizon, probabilities seem to converge to a 50/50 result. We augment the univariate model with key economic variables and results show that during stable times, the addition of those variables do not impact significantly the predictions obtained from the univariate model. We then compare our measure of short-term credibility of the central bank with existing indices in the literature. We find that the distinct horizon and perspective used to build the indices makes our short-term credibility index complementary to the existing long-term indices. We employ receiver operating characteristic curves to determine the adequate probability threshold from which the Central bank is predicted to be credible or not.

Appendix A Summary of forecasted probabilities

Table 4.1: Probabilities for inflation to remain within the bounds at the indicated months based on different forecast horizons and methods

		Dates										
		Dec-16	Jan-17	Fev-17	Mar-17	Avr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17
1-M	LLS	0.540	0.659	0.991	0.996	0.997	0.991	0.877	0.201	0.143	0.079	0.156
	SIR	0.542	0.660	0.996	0.997	1.000	0.996	0.921	0.271	0.194	0.121	0.188
3-M	LLS			0.491	0.694	0.887	0.879	0.894	0.771	0.612	0.351	0.369
	SIR			0.498	0.720	0.916	0.914	0.930	0.862	0.721	0.471	0.482
6-M	LLS						0.500	0.580	0.663	0.664	0.671	0.628
	SIR						0.553	0.626	0.717	0.742	0.759	0.720
		Nov-17	Dec-17	Jan-18	Fev-18	Mar-18	Avr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18
1-M	LLS	0.295	0.358	0.590	0.307	0.323	0.151	0.301	0.426	0.976	0.996	0.996
	SIR	0.322	0.401	0.620	0.354	0.356	0.168	0.321	0.447	0.990	0.999	0.999
3-M	LLS	0.307	0.448	0.547	0.560	0.640	0.488	0.510	0.393	0.523	0.574	0.411
	SIR	0.404	0.534	0.638	0.647	0.709	0.593	0.589	0.458	0.593	0.644	0.454
6-M	LLS	0.576	0.466	0.485	0.453	0.536	0.575	0.571	0.604	0.541	0.547	0.497
	SIR	0.690	0.625	0.612	0.582	0.671	0.678	0.670	0.687	0.652	0.661	0.597
		Oct-18	Nov-18	Dec-18	Jan-19	Fev-19	Mar-19	Avr-19	May-19	Jun-19	Jul-19	Aug-19
1-M	LLS	0.995	0.996	0.988	0.991	0.996	0.997	0.977	0.954	0.993	0.375	0.913
	SIR	0.999	1.000	0.996	0.997	1.000	0.999	0.982	0.964	0.998	0.540	0.943
3-M	LLS	0.892	0.852	0.856	0.862	0.823	0.810	0.882	0.896	0.653	0.658	0.886
	SIR	0.931	0.897	0.880	0.881	0.891	0.876	0.917	0.935	0.693	0.696	0.915
6-M	LLS	0.560	0.578	0.344	0.643	0.669	0.589	0.643	0.645	0.647	0.674	0.679
	SIR	0.657	0.675	0.412	0.693	0.751	0.625	0.692	0.741	0.725	0.756	0.750
		Sep-19	Oct-19	Nov-19	Dec-19	Jan-20	Fev-20	Mar-20	Avr-20	May-20	Jun-20	Jul-20
1-M	LLS	0.989	0.333	0.164	0.993	0.95	0.994					
	SIR	0.996	0.415	0.201	0.998	0.961	0.998					
3-M	LLS	0.306	0.695	0.857	0.410	0.453	0.882	0.492	0.877			
	SIR	0.454	0.778	0.904	0.510	0.579	0.910	0.527	0.895			
6-M	LLS	0.499	0.518	0.667	0.405	0.635	0.667	0.529	0.514	0.642	0.403	0.649
	SIR	0.565	0.579	0.729	0.555	0.723	0.716	0.656	0.665	0.688	0.446	0.708

Probabilities for inflation to remain between the bounds. LLS correspond to the simulations-based approach proposed by Lanne, Luoto, and Saikkonen (2012), 1 000 000 were employed. SIR correspond to the sampling importance resampling algorithm based on the sample-based approach proposed by Gouriéroux and Jasiak, 100 000 simulations were used in the first step, and 10 000 in the resampling step. Shaded columns represent months during which inflation was below the bounds.

5

Mixed causal-noncausal processes with multiple leads

Abstract

We investigate, as an extension to the case with one lead, the possibility to derive the closed-form of the conditional distribution of Cauchy-distributed mixed causal-noncausal processes with two leads. As a first step we derive the marginal distribution of the process, which is also a Cauchy distribution. The second step is the derivation of a bivariate distribution necessary to obtain a closed-form of the conditional distribution. This step has not been successful yet and is therefore left for further research. Then, we propose a prediction approach which builds on the two existing methods developed to approximate the conditional distribution of MAR(r, s) processes in the absence of closed-form. We compare the results obtained from the proposed approach with the densities obtained from its sample-based counterpart, developed by Gouriéroux and Jasiak (2016), in a simulation study. We then forecast the price index of all metals as a real life illustration of the two methods.

5.1 Introduction

Mixed causal-noncausal autoregressive (hereafter MAR) models have proven to be suitable in modelling stationary processes characterised by non-linear features. MAR models can capture these non-linear characteristics while being parsimonious and within a strictly stationary setting. An MAR(r, s) is defined as follows,

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t,$$

where L is the lag operator, $L^i y_t = y_{t-i}$ and L^{-1} the lead operator, $L^{-i}y_t = y_{t+i}$. Both lag polynomials $\Phi(L) = (1 - \phi_1 L - \ldots - \phi_r L^r)$ and $\Psi(L^{-1}) = (1 - \psi_1 L^{-1} - \ldots - \psi_s L^{-s})$ have roots strictly outside the unit circle to fulfill the stationarity condition. Note that the error term ε_t must be non-Gaussian for identification purposes. The mixed process can then be decomposed into a purely causal and a purely noncausal components. The following filtration,

$$u_t = \Phi(L)y_t,$$

makes u_t a purely noncausal process,

$$\Psi(L^{-1})u_t = \varepsilon_t. \tag{5.1}$$

This purely noncausal component u_t is the process of interest in the chapter. Indeed, it is the forward-looking part of the process that makes derivations and forecasts more intricate. Results are then easily extendable to mixed processes.

The literature regarding the estimation of MAR(r, s) models is now extensive, however, the literature regarding the forecasts of MAR processes, and especially with multiple leads, is still scarce. This chapter thus focuses on the predictions of MAR processes with more than one lead. It first aims attention at the theoretical aspect and then turns to more empirical analyses.

Assuming a Cauchy distribution for the error term ε_t is rather extreme and restrictive compared to a Student's t-distribution, which provides more flexibility as the degrees of freedom adjust to the recurrence of extreme values. However, out of the two distributions only the Cauchy entails a closed-form expression of the conditional distribution of MAR(0,1) processes (Gouriéroux and Zakoïan, 2013). Facing the lack of closed-form for suitable model specifications, Lanne, Luoto, and Saikkonen (2012) and Gouriéroux and Jasiak (2016) proposed approximation methods, using respectively simulations and sample-based approximations. Hecq and Voisin (2021) evaluates the performance of these approaches for MAR$(0,1)$ processes by first comparing them to the closed-form results with Cauchy-distributed errors. Since a $t(1)$-distribution coincides with a Cauchy(0,1), this allows to obtain a benchmark of their performance for the most extreme t-distribution to better understand how they behave even in the absence of closed-form expressions.

One lead is enough to capture locally explosive episodes, yet processes are often identified, in applied research, as MARs with more than one lead. For instance, Fries and Zakoïan (2019a) detect 2 to 3 leads in Soybean prices, Sugar prices and Shiller price/earning ratio and Giancaterini and Hecq (2022) identify an MAR(2,2) for the variation of daily COVID-19 deaths in Belgium. This shows the importance of also having a theoretical understanding of the conditional distribution of more general MAR(r,s). The first motivation of this chapter is thus to extend the findings of Gouriéroux and Zakoïan (2013) to Cauchy-distributed MAR$(0,2)$ processes.

The first step we take is the investigation of the marginal distribution of the process u_t. We find that it is Cauchy-distributed and obtain a closed-form expression of the scale when the roots of the lead polynomial are real. When the roots are complex however, we do not find a closed-form and thus approximate it. The second step is the derivation of the joint distribution of the bivariate process

(u_t, u_{t+1}), which is necessary to obtain the conditional distribution of interest $f(u_t|u_{t+1}, u_{t+2})$. We determine that it is a bivariate Cauchy distribution but do not find a general form of the parameters of the distribution. We therefore leave this inquiry for future research.

While we could not compare the approximation methods with the Cauchy-derived closed-form results as was done in Hecq and Voisin (2021) for models with one lead, the approximation methods can still be used and compared between one another. The simulations-based method of Lanne, Luoto, and Saikkonen (2012) is derived to obtain, at a given horizon, a point forecast or a conditional cumulative density value. Throughout this thesis, we argue for the use of density forecasts instead of point forecasts, as they carry much more information. Hence, estimating the entire conditional distribution with this approach is quite computationally demanding. The approach of Gouriéroux and Jasiak (2016) allows to approximate the entire conditional distribution using sample averages. While this approach is sensitive to the sample size, it is more efficient in obtaining the complete predictive density. This approach however becomes increasingly computationally demanding as the forecast horizon *H* increases. Indeed, it predicts the density of the whole path up to *T*+*H* whereas the method of Lanne, Luoto, and Saikkonen (2012) only predicts the density of the point *T*+*H*.

The second motivation and contribution of this chapter is the proposition of a forecasting method combining the two existing ones. We suggest employing simulations in the spirit of Lanne, Luoto, and Saikkonen (2012), to approximate the conditional distribution like Gouriéroux and Jasiak (2016) do, using simulations-based averages instead. We compare the results of this proposed approach with the sample-based methods on simulated MAR$(0, 2)$ series with different sets of parameters. We investigate their respective sensitivity to the number of simulations and to the sample size. Then, to illustrate the methods we predict the last 12 months of the metals price index,

which is a period of volatile prices due to the impact of the COVID-19 pandemic, in a pseudo-real life forecasting exercise.

The chapter is constructed as follows. Section 5.2 presents the derivations of the marginal densities of MAR$(0, 2)$ Cauchy-distributed processes. Section 5.3 describes the existing approximation methods to forecast MAR$(0, s)$ processes and introduces the new simulations-based approach for approximating the conditional density. Section 5.4 compares the sample-based approach of Gouriéroux and Jasiak (2016) and the proposed simulations-based counterpart in a simulation study. Section 5.5 then illustrates the methods on detrended metals price index, which are identified as an MAR$(0, 2)$. Section 5.6 then concludes.

5.2 MAR(0,2) with Cauchy distributed errors

This section intends to derive the conditional density of a Cauchy-distributed MAR$(0, 2)$ process, as an extension of the findings of Gouriéroux and Zakoïan (2013) for the MAR$(0, 1)$ case. Recall that an MAR$(0, 1)$ process is defined as $u_t = \psi u_{t+1} + \varepsilon_t$. The *H*-step ahead predictive density of an MAR$(0, 1)$ process is the following,

$$\begin{aligned} f(u_{T+1}, \ldots, u_{T+H}|u_T) &= \frac{f(u_T, u_{T+1}, \ldots, u_{T+H})}{f(u_T)} \\ &= \frac{f(u_T|u_{T+1}) \times f(u_{T+1}|u_{T+2}) \ldots f(u_{T+H-1}|u_{T+H}) \times f(u_{T+H})}{f(u_T)}. \end{aligned} \tag{5.2}$$

The conditional densities in the form of $f(u_t|u_{t+1})$ are the *pdf* of the errors' distribution, evaluated at the point $u_t - \psi u_{t+1}$. The part of the density that is often unknown or does not have a closed-form expression is the ratio of the marginal densities. Gouriéroux and Zakoïan (2013) show that if the error term ε_t is Cauchy-distributed, then the predictive density (5.2) admits a closed-form. Indeed, if $\varepsilon_t \sim Cauchy(0, \gamma)$,

the marginal distribution of u_t is a $Cauchy\left(0, \frac{\gamma}{1-|\psi|}\right)$. Other errors' distributions, such as Student's *t*, while being more flexible and providing a better fit than Cauchy, lead to the absence of closed-form for the predictive density. The one-step ahead predictive density of a Cauchy-distributed MAR$(0,1)$ process therefore corresponds to the following,

$$\begin{aligned} f(u_{T+1}|u_T) &= \frac{f(u_T|u_{T+1}) \times f(u_{T+1})}{f(u_T)} \\ &= \frac{1}{\pi\gamma} \times \frac{\gamma^2}{(u_T - \psi u_{T+1})^2 + \gamma^2} \times \frac{\gamma^2 + (1-|\psi|)^2 u_T^2}{\gamma^2 + (1-|\psi|)^2 u_{T+1}^2}. \end{aligned}$$

Let us now consider a purely noncausal MAR$(0,2)$ process,

$$u_t = \psi_1 u_{t+1} + \psi_2 u_{t+2} + \varepsilon_t, \qquad \text{with} \quad \varepsilon_t \sim Cauchy(0,\gamma).$$

As a natural extension of the MAR(0,1) predictive density (5.2), the *H*-step ahead conditional distribution of an MAR$(0,2)$ process is the following,

$$\begin{aligned} &f(u_{T+1}, \ldots, u_{T+H}|u_{T-1}, u_T) \\ &= \frac{f(u_{T-1}, u_T, u_{T+1}, \ldots, u_{T+H})}{f(u_{T-1}, u_T)} \\ &= f(u_{T-1}|u_T, u_{T+1})\, f(u_T|u_{T+1}, u_{T+2}) \ldots f(u_{T+H-2}|u_{T+H-1}, u_{T+H}) \\ &\qquad \times \frac{f(u_{T+H-1}, u_{T+H})}{f(u_{T-1}, u_T)}. \end{aligned} \tag{5.3}$$

Analogously to the MAR$(0,1)$ case, the conditional distributions in the form of $f(u_t|u_{t+1}, u_{t+2})$ are the *pdf* of a $Cauchy(0,\gamma)$ evaluated at the point $u_t - \psi_1 u_{t+1} - \psi_2 u_{t+2}$. However, to obtain a closed-form of the predictive density 5.3, we now need to determine the joint distribution of the bivariate process $f(u_t, u_{t+1})$. To do so, we first investigate the marginal distribution of the process u_t.

Consider the MA representation of the MAR$(0, 2)$ process,

$$u_t = (1 - \psi_1 L^{-1} - \psi_2 L^{-2})^{-1} \varepsilon_t.$$

Let λ_1 and λ_2 be the inverted roots of the lag polynomial $(1 - \psi_1 L^{-1} - \psi_2 L^{-2})$. The coefficients of the MA representation can thus be obtained as follows,

$$\begin{aligned}
(1 - \psi_1 L^{-1} - \psi_2 L^{-2})^{-1} \varepsilon_t &= \left(\left(1 - \lambda_1 L^{-1}\right)\left(1 - \lambda_2 L^{-1}\right)\right)^{-1} \varepsilon_t \\
&= \frac{\lambda_1}{\lambda_1 - \lambda_2} \frac{1}{1 - \lambda_1 L^{-1}} \varepsilon_t - \frac{\lambda_2}{\lambda_1 - \lambda_2} \frac{1}{1 - \lambda_2 L^{-1}} \varepsilon_t \\
&= \sum_{j=0}^{\infty} \left(\frac{\lambda_1}{\lambda_1 - \lambda_2} \lambda_1^j - \frac{\lambda_2}{\lambda_1 - \lambda_2} \lambda_2^j \right) \varepsilon_{t+j} \\
&= \sum_{j=0}^{\infty} \left(\frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right) \varepsilon_{t+j}.
\end{aligned}$$

To fulfill the stationarity condition, both inverted roots must be inside the unit circle.

Substituting the process u_t with its MA representation in the general form of its characteristic function, we obtain,

$$\mathbb{E}\left[e^{i\tau u_t}\right] = \mathbb{E}\left[e^{i\tau \sum_{j=0}^{\infty} \left(\frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right) \varepsilon_{t+j}} \right] = \prod_{j=0}^{\infty} \mathbb{E}\left[e^{i\tau \left(\frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right) \varepsilon_{t+j}} \right].$$

Recall that the error term ε_t is *i.i.d* $Cauchy(0,\ \gamma)$, hence the above entity is the infinite product of the characteristic functions of processes with distribution $Cauchy\left(0,\ \left|\frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2}\right| \gamma\right)$. Thus, writing out the functional form of the characteristic functions, we can recover the charac-

teristic function of the process u_t,

$$\prod_{j=0}^{\infty} \mathbb{E}\left[e^{i\tau \left(\frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right) \varepsilon_{t+j}} \right] = \prod_{j=0}^{\infty} e^{-|\tau| \left| \frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right| \gamma}$$
$$= e^{-|\tau| \sum_{j=0}^{\infty} \left| \frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right| \gamma}.$$

It entails that the characteristic function of the process u_t is that of a Cauchy-distributed process with scale $\gamma \sum_{j=0}^{\infty} \left| \frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right|$. Hence, the marginal distribution of the MAR$(0, 2)$ Cauchy-distributed process u_t is the following,

$$u_t \sim Cauchy \left(0,\ \gamma \sum_{j=0}^{\infty} \left| \frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right| \right). \tag{5.4}$$

The upcoming sections investigate the convergence of the infinite sum in the scale parameter. We first analyse the scale when the roots of the lead polynomial are real and then investigate the case when the roots are complex.

5.2.1 Lead polynomial with real roots

When the roots of the lag polynomials $(1 - \psi_1 L^{-1} - \psi_2 L^{-2})$ are real, that is when the determinant $\Delta \equiv \psi_1^2 + 4\psi_2 \geq 0$, the inverted roots λ_1 and λ_2 are defined as such,

$$\lambda_1 = \frac{\psi_1 - \sqrt{\psi_1^2 + 4\psi_2}}{2} \qquad \lambda_2 = \frac{\psi_1 + \sqrt{\psi_1^2 + 4\psi_2}}{2}.$$

For any admissible combination of the lead coefficients yielding real

roots, we therefore have that,

$$\begin{aligned}\lambda_1 - \lambda_2 &= \frac{\psi_1 - \sqrt{\psi_1^2 + 4\psi_2}}{2} - \frac{\psi_1 + \sqrt{\psi_1^2 + 4\psi_2}}{2} \\ &= -\sqrt{\Delta} < 0.\end{aligned}$$

This means that for all potential real inverted roots of the lag polynomial as they are defined above we have $\lambda_1 < \lambda_2$.

It implies that there are two possible scenarios. The first one is $|\lambda_1| > |\lambda_2|$, which is always satisfied when $\lambda_1 < \lambda_2 < 0$ and for some combinations of coefficients satisfying $\lambda_1 < 0 < \lambda_2$. This scenario requires $\psi_1 < 0$ for the conditions to met. The second scenario, namely $|\lambda_1| < |\lambda_2|$, is always satisfied when $0 < \lambda_1 < \lambda_2$ and for some combinations of coefficients satisfying $\lambda_1 < 0 < \lambda_2$. As opposed to the first one, this scenario requires $\psi_1 > 0$.

Let us first focus on the first scenario, namely $|\lambda_1| > |\lambda_2|$. Since we know that $\lambda_1 < \lambda_2$, it implies that, for an arbitrary number i, $\lambda_1^i > \lambda_2^i$ when i is even and $\lambda_1^i < \lambda_2^i$ when i is odd. Therefore, the infinite sum from the scale parameter in (5.4) can be expressed as follows,

$$\begin{aligned}\sum_{j=0}^{\infty} \left| \frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right| &= \frac{1}{|\lambda_1 - \lambda_2|} \sum_{j=0}^{\infty} \left| \lambda_1^{j+1} - \lambda_2^{j+1} \right| \\ &= \frac{1}{\lambda_2 - \lambda_1} \sum_{j=0}^{\infty} (-1)^j (\lambda_2^{j+1} - \lambda_1^{j+1}) \\ &= \frac{\lambda_2}{\lambda_2 - \lambda_1} \sum_{j=0}^{\infty} (-\lambda_2)^j - \frac{\lambda_1}{\lambda_2 - \lambda_1} \sum_{j=0}^{\infty} (-\lambda_1)^j \\ &= \frac{\lambda_2}{\lambda_2 - \lambda_1} \frac{1}{1 + \lambda_2} - \frac{\lambda_1}{\lambda_2 - \lambda_1} \frac{1}{1 + \lambda_1} \\ &= \frac{1}{(1 + \lambda_1)(1 + \lambda_2)} = \frac{1}{1 + \lambda_1 + \lambda_2 + \lambda_1 \lambda_2}.\end{aligned}$$

Note that $\lambda_1\lambda_2 = -\psi_2$. Hence, by substituting the inverted roots with their values, we find that the infinite sum converges to the following,

$$\sum_{j=0}^{\infty}\left|\frac{\lambda_1^{j+1}-\lambda_2^{j+1}}{\lambda_1-\lambda_2}\right| = \frac{1}{1+\frac{\psi_1-\sqrt{\Delta}}{2}+\frac{\psi_1+\sqrt{\Delta}}{2}-\psi_2} = \frac{1}{1+\psi_1-\psi_2}.$$

Let us now consider the second scenario, $|\lambda_1| < |\lambda_2|$. This scenario implies that $\lambda_1^i < \lambda_2^i$ for any arbitrary number i. Hence, the infinite sum from the scale parameter in (5.4) can be expressed as follows,

$$\begin{aligned}\sum_{j=0}^{\infty}\left|\frac{\lambda_1^{j+1}-\lambda_2^{j+1}}{\lambda_1-\lambda_2}\right| &= \frac{\lambda_2}{\lambda_2-\lambda_1}\sum_{j=0}^{\infty}\lambda_2^j - \frac{\lambda_1}{\lambda_2-\lambda_1}\sum_{j=0}^{\infty}\lambda_1^j \\ &= \frac{\lambda_2}{\lambda_2-\lambda_1}\frac{1}{1-\lambda_2} - \frac{\lambda_1}{\lambda_2-\lambda_1}\frac{1}{1-\lambda_1} \\ &= \frac{1}{(1-\lambda_1)(1-\lambda_2)} = \frac{1}{1-\lambda_1-\lambda_2+\lambda_1\lambda_2}.\end{aligned}$$

Again, by substituting the inverted roots with their values we find that the infinite sum converges to the following,

$$\sum_{j=0}^{\infty}\left|\frac{\lambda_1^{j+1}-\lambda_2^{j+1}}{\lambda_1-\lambda_2}\right| = \frac{1}{1-\frac{\psi_1-\sqrt{\Delta}}{2}-\frac{\psi_1+\sqrt{\Delta}}{2}-\psi_2} = \frac{1}{1-\psi_1-\psi_2}.$$

Since the first scenario requires $\psi_1 < 0$ and the second scenario requires $\psi_1 > 0$, the marginal distribution of the process u_t, as defined in (5.4), can be generalized, when the roots of the lag polynomial are real, as follows,

$$Cauchy\left(0,\ \frac{\gamma}{1-|\psi_1|-\psi_2}\right). \tag{5.5}$$

5.2.2 Lead polynomial with complex roots

Let us now turn to the case when the roots of the lag polynomial $(1 - \psi_1 L^{-1} - \psi_2 L^{-2})$ are complex. It thus implies a negative determinant $\Delta = \psi_1^2 + 4\psi_2 < 0$ and the two inverted roots λ_1 and λ_2 are now as follows,

$$\lambda_1 = \frac{\psi_1}{2} - i\frac{\sqrt{-\Delta}}{2} \qquad \lambda_2 = \frac{\psi_1}{2} + i\frac{\sqrt{-\Delta}}{2}.$$

Complex roots can only be obtained with a negative second lead coefficient, and more specifically when $\psi_2 < -\frac{\psi_1^2}{4}$.

Let r be the modulus of the two conjugate complex roots, which must be smaller than 1 to satisfy stationarity conditions. Hence,

$$r \equiv |\lambda_1| = |\lambda_2| = \sqrt{\left(\frac{\psi_1}{2}\right)^2 + \left(\frac{\sqrt{-\Delta}}{2}\right)^2} = \sqrt{-\psi_2} < 1,$$

which implies that $-1 < \psi_2 < -\frac{\psi_1^2}{4}$. Furthermore, let us denote $\theta \equiv -arctan\left(\frac{\sqrt{-\Delta}}{\psi_1}\right)$ the argument of λ_1, and thus $-\theta$ the argument of λ_2. The infinite sum of the scale parameter in (5.4) can thus be expressed in terms of the polar forms of the inverted roots as follows,

$$\begin{aligned}\sum_{j=0}^{\infty}\left|\frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2}\right| &= \sum_{n=1}^{\infty}\left|\frac{\lambda_1^{n} - \lambda_2^{n}}{\lambda_1 - \lambda_2}\right| \\ &= \sum_{n=1}^{\infty}\frac{\left|r^n(cos(n\theta) - i\,sin(n\theta)) - r^n(cos(n\theta) + i\,sin(n\theta))\right|}{\sqrt{-\Delta}} \\ &= \sum_{n=1}^{\infty}\frac{\left|2r^n i\,sin(n\theta)\right|}{\sqrt{-\Delta}} = \frac{2}{\sqrt{-\Delta}}\sum_{n=1}^{\infty} r^n\left|sin(n\theta)\right|.\end{aligned}$$

We were not able to derive a closed-form expression of the latter and could therefore not provide a closed-form of the scale of the marginal distribution of the process u_t when its roots are complex. However, since $0 < r = \sqrt{-\psi_2} < 1$ and $0 \leq |sin(n\theta)| \leq 1$, the infinite sum converges at a faster rate than $\sum_{n=1}^{\infty} r^n$ and can therefore be approximated as such,

$$\begin{aligned}\sum_{j=0}^{\infty} \left| \frac{\lambda_1^{j+1} - \lambda_2^{j+1}}{\lambda_1 - \lambda_2} \right| &= \frac{2}{\sqrt{-\Delta}} \sum_{n=1}^{\infty} r^n |sin(n\theta)| \\ &\approx \frac{2}{\sqrt{-\Delta}} \sum_{n=1}^{K} \sqrt{-\psi_2}^n \left| sin\left(n \; arctan\left(\frac{\sqrt{-\Delta}}{\psi_1} \right) \right) \right|, \end{aligned} \tag{5.6}$$

where K is large enough for approximation errors to be negligible. Note that since $arctan(x) = -arctan(-x)$ and $sin(x) = -sin(-x)$, the scale does not depend on the sign of ψ_1 but only on $|\psi_1|$, analogously to the case with real roots.

5.2.3 Summarising results

Let us define the scale of the marginal distribution of the MAR$(0, 2)$ process u_t with $Cauchy(0, 1)$-distributed errors as δ. That is,

$$u_t \sim Cauchy(0, \delta).$$

Table 5.1 displays the scale δ for all feasible combinations of lead coefficients fulfilling the stationarity conditions, where we only consider increments of 0.1. The scale, for both real and complex roots, is dependent on the sign and value of ψ_2 but only on the magnitude of $|\psi_1|$. We thus present the marginal scales for feasible pairs of $(|\psi_1|, \psi_2)$.

The gray cells are the combinations of coefficients corresponding to real roots in the lag polynomial, hence it is the exact values of the scales evaluated using Equation (5.5). The other combinations of coefficients coincide with complex roots, for which the scales are approximated using Equation (5.6) with $K = 200$.

We can notice that when the roots are real, the scale of the marginal distribution increases with the absolute value of both coefficients. When the roots are complex, the scale increases with the absolute value of ψ_2 but decreases when $|\psi_1|$ increases. Note that the scales would increase to even higher values if smaller increments were considered for the most extreme cases. Furthermore, if instead the errors of the MAR(0,2) process are $Cauchy(0, \gamma)$-distributed, then all marginal scales should be multiplied by γ.

Overall, Table 5.1 helps understanding the role of each of the lead coefficients on the scale of the process.

While we found that the marginal distribution of the process is Cauchy, recall that we are interested in the joint distribution of the bivariate process (u_t, u_{t+1}). Note that a random vector is said to have a multivariate Cauchy distribution if any linear combinations of its components follows a Cauchy distribution.

Hence, consider now a linear combination of u_t and u_{t+1}, for arbitrary scalars a and b,

$$\begin{aligned} a\,u_t + b\,u_{t+1} &= a\sum_{i=0}^{\infty}\left(\frac{\lambda_1^{i+1} - \lambda_2^{i+1}}{\lambda_1 - \lambda_2}\right)\varepsilon_{t+i} + b\sum_{i=0}^{\infty}\left(\frac{\lambda_1^{i+1} - \lambda_2^{i+1}}{\lambda_1 - \lambda_2}\right)\varepsilon_{t+1+i} \\ &= a\,\varepsilon_t + \sum_{i=1}^{\infty}\left(\frac{a(\lambda_1^{i+1} - \lambda_2^{i+1}) + b(\lambda_1^{i} - \lambda_2^{i})}{\lambda_1 - \lambda_2}\right)\varepsilon_{t+i}. \end{aligned}$$

Table 5.1: Scale of marginal density of u_t denoted as δ based on coefficients of leads

$\lvert\psi_1\rvert$	ψ_2																		
	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	10.00	5.00	3.33	2.50	2.00	1.67	1.43	1.25	1.11	1.00	1.11	1.25	1.43	1.67	2.00	2.50	3.33	5.00	10.00
0.1	577.10	136.13	56.74	29.71	17.53	11.08	7.26	4.80	3.06	1.11	1.25	1.43	1.67	2.00	2.50	3.33	5.00	10.00	
0.2	559.34	131.71	54.78	28.61	16.83	10.59	6.90	4.52	2.82	1.25	1.43	1.67	2.00	2.50	3.33	5.00	10.00		
0.3	542.86	127.62	52.98	27.60	16.19	10.14	6.57	4.27	2.62	1.43	1.67	2.00	2.50	3.33	5.00	10.00			
0.4	527.53	123.84	51.32	26.68	15.61	9.75	6.29	4.05	2.45	1.67	2.00	2.50	3.33	5.00	10.00				
0.5	513.22	120.31	49.78	25.83	15.07	9.38	6.03	3.86	2.30	2.00	2.50	3.33	5.00	10.00					
0.6	499.82	117.02	48.34	25.04	14.58	9.05	5.79	3.69	2.18	2.50	3.33	5.00	10.00						
0.7	487.24	113.95	47.01	24.31	14.13	8.75	5.58	3.53	2.50	3.33	5.00	10.00							
0.8	475.40	111.06	45.76	23.63	13.71	8.47	5.38	3.39	3.33	5.00	10.00								
0.9	464.25	108.34	44.59	22.99	13.31	8.21	5.20	3.33	5.00	10.00									
1	453.70	105.78	43.49	22.40	12.95	7.97	5.04	5.00	10.00										
1.1	443.72	103.37	42.45	21.84	12.61	7.74	5.00	10.00											
1.2	434.26	101.08	41.47	21.31	12.28	7.53	10.00												
1.3	425.27	98.91	40.54	20.81	11.98	10.00													
1.4	416.71	96.85	39.67	20.34	11.70														
1.5	408.56	94.89	38.83	19.89															
1.6	400.79	93.02	38.04																
1.7	393.36	91.24																	
1.8	386.25																		

In gray are the combinations of coefficients corresponding to real roots in the lag polynomial hence the scales are the exact values evaluated using Equation (5.5). Other entries correspond to complex roots, the scale has been approximated using Equation (5.6) with $K = 200$.

It corresponds to a linear combination of independent Cauchy-distributed errors. This implies that any linear combination of u_t and u_{t+1} follows a Cauchy distribution, which entails that the joint process $(u_t,\, u_{t+1})$ follows a bivariate Cauchy distribution.

We did not find a general representation of the parameters of the bivariate distribution and thus cannot present a closed-form for the conditional density (5.3). We hence leave this inquiry to further research.

As mentioned before, assuming a Cauchy distribution for the error term is rather restrictive. Indeed, in the empirical literature most models are estimated with Student's *t*-distributed error terms with degrees of freedom larger than 1. We however wanted to evaluate the benchmark performance of the existing approximation methods on simulated Cauchy-distributed series, analogously to what was done in Hecq and Voisin (2021) with MAR$(0, 1)$ models. Hence, we for now also leave this for future analyses.

5.3 Forecasting an MAR(0,*s*) process

Let us now consider the more general MAR$(0, s)$ process,

$$u_t = \psi_1 u_{t+1} + \psi_2 u_{t+2} + \cdots + \psi_s u_{t+s} + \varepsilon_t,$$

where ε_t is for instance Student's *t*-distributed, implying that the conditional density of the MAR$(0, s)$ process does not admit a closed-form expression. The conditional density, extending (5.3) to the case with s

leads, is the following,

$$
\begin{aligned}
&f(u_{T+1},\dots,u_{T+H}|u_{T-s+1},\dots,u_T)\\
&\quad=\frac{f(u_{T-s+1}|u_{T-s+2},\dots,u_{T+1})\dots f(u_{T+H-s}|u_{T+H-s+1},\dots,u_{T+H})}{f(u_{T-s+1},\dots,u_T)}\\
&\qquad\qquad\times f(u_{T+H-s+1},\dots,u_{T+H}).
\end{aligned}
$$

While we do not know the joint distributions $f(u_{T-s+1},\dots,u_T)$ and $f(u_{T+H-s+1},\dots,u_{T+H})$, they can be expressed in terms of the conditional expectations,

$$
\begin{aligned}
&f(u_{T+1},\dots,u_{T+H}|u_{T-s+1},\dots,u_T)\\
&\quad=\frac{f(u_{T-s+1}|u_{T-s+2},\dots,u_{T+1})\dots f(u_{T+H-s}|u_{T+H-s+1},\dots,u_{T+H})}{\mathbb{E}\left[f(u_{T-s+1}|u_{T-s+2},\dots,u_T,u^*_{T+1})\dots f(u_T|u^*_{T+1},\dots,u^*_{T+s})\right]_*}\\
&\qquad\qquad\times\mathbb{E}\Big[f(u_{T+H-s+1}|u_{T+H-s+2},\dots,u^*_{T+H+1})\dots\\
&\qquad\qquad\qquad\dots f(u_{T+H}|u^*_{T+H+1},\dots,u^*_{T+H+s})\Big]_*,
\end{aligned}
\tag{5.7}
$$

where $\mathbb{E}[\cdot]_*$ is the expectation conditioned on the values indicated by a $*$. The product of conditional distributions within the expectations are conditioned on the paths of values $(u^*_{T+1},\dots,u^*_{T+s})$ and $(u^*_{T+H+1},\dots,u^*_{T+H+s})$ respectively.

As described in the previous chapters, when no closed-form exists for the conditional distribution (e.g. when $s>1$ or with Student's t-distributed errors), the approximation methods of Lanne, Luoto, and Saikkonen (2012) and Gouriéroux and Jasiak (2016) can be employed. See Hecq and Voisin (2021) for an analysis of their performance for MAR$(0,1)$ processes and see the respective articles for more details on the methods.

5.3.1 Existing approaches

Gouriéroux and Jasiak (2016) provide a general formula (for any number of lags and leads, forecast horizon and distribution) for estimating the conditional density in the absence of closed-form expression. It is based on (5.7). They propose to approximate the conditional expectations using the sample average, employing all past values,

$$\begin{aligned} f(u_{T+1},\dots,u_{T+H}|u_{T-s+1},\dots,u_T) \approx \\ \frac{g_s(\hat{u}_{T-s+1},\hat{u}_{T-s+2},\dots,u_{T+1})\dots g_s(u_{T+H-s},u_{T+H-s+1},\dots,u_{T+H})}{\sum_{t=1}^{T-s+1}\Big\{g_s(\hat{u}_{T-s+1},\dots,\hat{u}_T,\hat{u}_t)\dots g_s(\hat{u}_T,\hat{u}_t,\dots,\hat{u}_{t+s-1})\Big\}} \\ \times \sum_{t=1}^{T-s+1}\Big\{g_s(u_{T+H-s+1},\dots,u_{T+H},\hat{u}_t)\dots g_s(u_{T+H},\hat{u}_t,\dots,\hat{u}_{t+s-1})\Big\}, \end{aligned} \tag{5.8}$$

where $g_s(u_t,\dots,u_{t+s})$ is the assumed distribution of the error term evaluated at the point $(u_t - \psi_1 u_{t+1} - \cdots - \psi_s u_{t+s})$. The conditional expectations in (5.7) are thus approximated as the average based on the whole observed sample, denoted as $(\hat{u}_1,\dots,\hat{u}_T)$.

Lanne, Luoto, and Saikkonen (2012) on the other hand employ the MA representation of the noncausal component u_t and base their methodology on the assumption that there exists an integer M large enough so that any future point of the noncausal component can be approximated, with negligible error. They therefore base their methodology on the fact that the MAR$(0,s)$ process can be approximated by the following finite sum,

$$u_{T+h} \approx \sum_{i=0}^{M-h} \iota' \begin{bmatrix} \psi_1 & \dots & \psi_{s-1} & \psi_s \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix}^i \iota\, \varepsilon_{T+h+i} = \sum_{i=0}^{M-h} \beta_i\, \varepsilon_{T+h+i}, \tag{5.9}$$

for any $h \geq 1$, where $\iota' = [1 \quad 0 \cdots 0]$ of size $1 \times s$.

Lanne, Luoto, and Saikkonen (2012) propose a general formula to compute the conditional expected value of any function of the *M* future errors, approximated using simulations. Their method allows to obtain point forecasts, or, by choosing the function to be the following indicator function, namely being 1 when the condition is fulfilled and zero otherwise, to obtain an approximation of the conditional cumulative density of the MAR$(0, s)$ process at a given point *x*,

$$\begin{aligned}\mathbb{E}_T\Big[\mathbf{1}(u_{T+h} \leq x)\Big] &= \mathbb{P}\Big(u_{T+h} \leq x | \mathcal{F}_T\Big) \\ &\approx \mathbb{E}_T\left[\mathbf{1}\left(\sum_{j=0}^{M-h+i} \beta_j \varepsilon_{T+h-i+j} \leq x\right)\right].\end{aligned}$$

Hence, by computing its value for all possible x covering the range of potential values for u_{T+h}, we can obtain the whole conditional *cdf* of u_{T+h}. In turn, we can recover the conditional *pdf*. However, this approach becomes quite computationally demanding as each estimated density point requires a large set of simulations and needs to be iterated a significant amount of time over all possible x to obtain the complete conditional density.

Overall, the method of Gouriéroux and Jasiak (2016) is more efficient in obtaining entire predictive densities for short horizons but becomes computationally demanding for farther horizons as it derives the predictive density of the path until $T + H$. The method of Lanne, Luoto, and Saikkonen (2012) on the other hand, while less efficient in obtaining complete predictive densities, can be used to obtain the predictive density at farther horizons as it only predicts the point $T + H$ and not the path leading to it.

5.3.2 Combining approaches

Combining the approaches presented above, we suggest a new way of approximating the conditional density in the absence of closed-form. While Gouriéroux and Jasiak (2016) employ all past values to approximate the expected values in Equation (5.7) by using their sample average, we instead suggest the use of simulations, as do Lanne, Luoto, and Saikkonen (2012). Indeed, especially for small samples, using past values can yield distorted estimated densities as was shown in Hecq and Voisin (2021).

Hence, we suggest approximating (5.7) using simulated paths of $(u^*_{T+1}, \ldots, u^*_{T+H+s})$ instead of observed paths. The distribution of the errors is well known and sets of M *i.i.d* errors can easily be simulated from it. Hence, the nth simulated path, which is denoted as $(u^{(n)}_{T+1}, \ldots, u^{(n)}_{T+H+s})$, is derived from the nth set of M simulated future errors,

$$u^{(n)}_{T+h} \approx \sum_{i=0}^{M-h} \beta_i \varepsilon^{(n)}_{T+h+i}, \qquad \text{for} \quad n = 1, \ldots, N,$$

where β_i are the coefficients from Equation (5.9) and N the overall number of simulated paths.

Equation (5.7) can then be approximated by taking the averages over all N simulated paths. The H-step ahead predictive density of an

MAR$(0, s)$ process is hence approximated as follows,

$$\begin{aligned}
&\hat{f}\left(u_{T+1}, \ldots, u_{T+H} | u_{T-s+1}, \ldots, u_T\right) \\
&= \left(\frac{g_s(u_{T-s+1}, \ldots, u_{T+1}) \ldots g_s(u_{T+H-s}, \ldots, u_{T+H})}{\sum_{n=1}^{N} \left\{ g_s(u_{T-s+1}, \ldots, u_T, u_{T+1}^{(n)}) \ldots g_s(u_T, u_{T+1}^{(n)}, \ldots, u_{T+s}^{(n)}) \right\}} \right) \\
&\qquad \times \sum_{n=1}^{N} \Big\{ g_s(u_{T+H-s+1}, \ldots, u_{T+H}, u_{T+H+1}^{(n)}) \ldots \\
&\qquad\qquad \ldots g_s(u_{T+H}, u_{T+H+1}^{(n)}, \ldots, u_{T+H+s}^{(n)}) \Big\}.
\end{aligned} \tag{5.10}$$

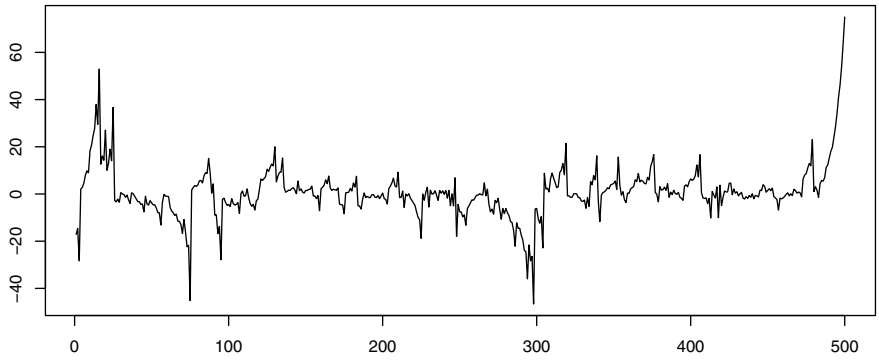
This allows to obtain the the complete predictive density in a more efficient way than with the method of Lanne, Luoto, and Saikkonen (2012). However, the same limitations as the sample-based approach apply when the horizon increases. We will therefore now focus on 1-step ahead forecasts.

5.4 Simulations

To illustrate the application of the aforementioned forecasting methods on MAR$(0, 2)$, we simulate three samples of 10 000 realisations of Student's *t*-distributed error terms, with 1, 2 and 2.5 degrees of freedom. We then transform the three samples into MAR$(0, 2)$ processes by employing two combinations of coefficients, namely $(\psi_1, \psi_2) = (0.5, 0.3)$ and $(1.1, -0.3)$, which correspond to real and complex roots in the lead polynomial respectively. That is, for each of the three samples of 10 000 errors, we obtain two MAR$(0, 2)$ processes.

Figure 5.1 depicts the last 500 data points of the samples derived from the $t(1)$ errors. We only show the last 500 points (out of 10 000) to see

(a) $\psi_1 = 0.5, \psi_2 = 0.3$



(b) $\psi_1 = 1.1, \psi_2 = -0.3$

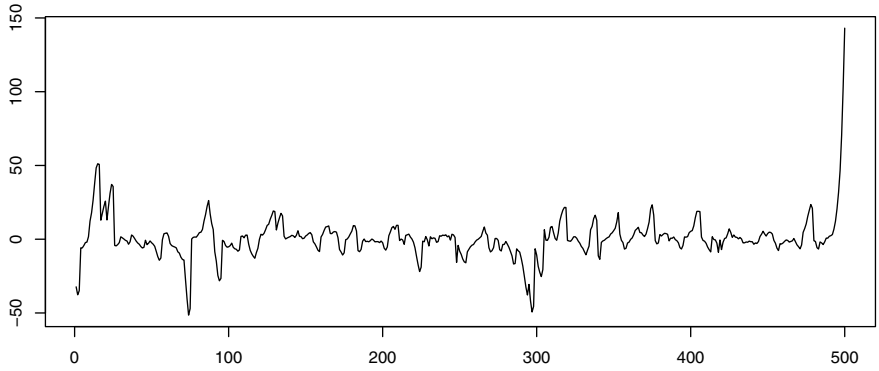
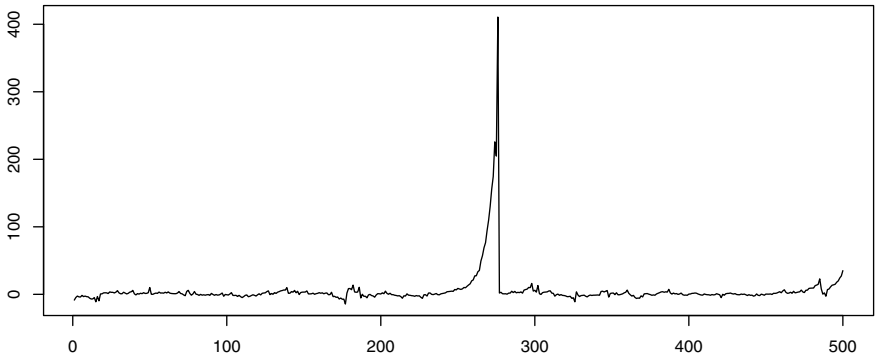


Figure 5.1: Last 500 points of simulated series with $t(1)$ errors

(a) $\psi_1 = 0.5, \psi_2 = 0.3$



(b) $\psi_1 = 1.1, \psi_2 = -0.3$

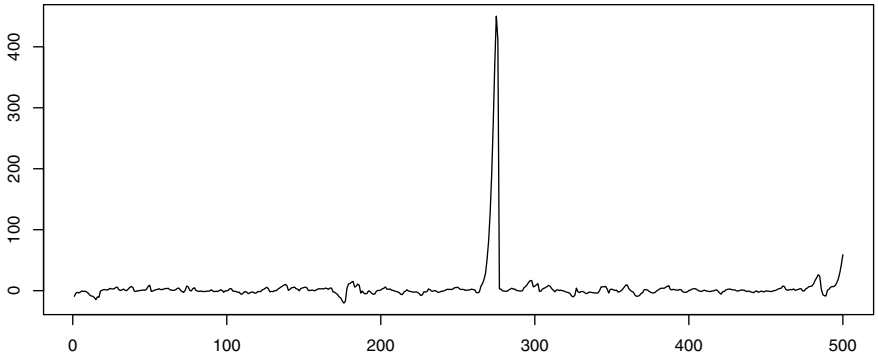
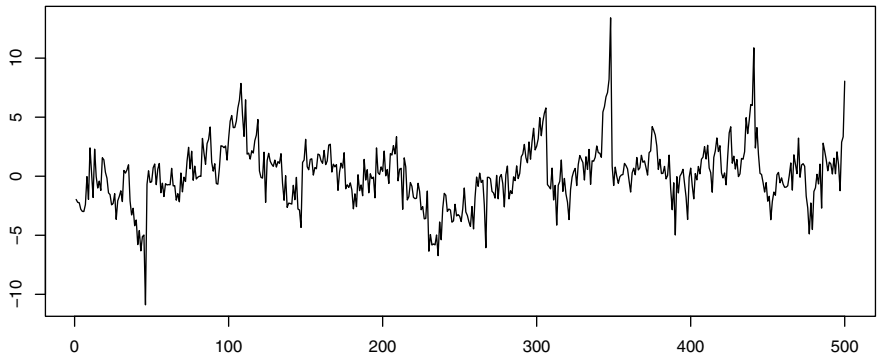


Figure 5.2: Last 500 points of simulated series with $t(2)$ errors

(a) $\psi_1 = 0.5, \psi_2 = 0.3$



(b) $\psi_1 = 1.1, \psi_2 = -0.3$

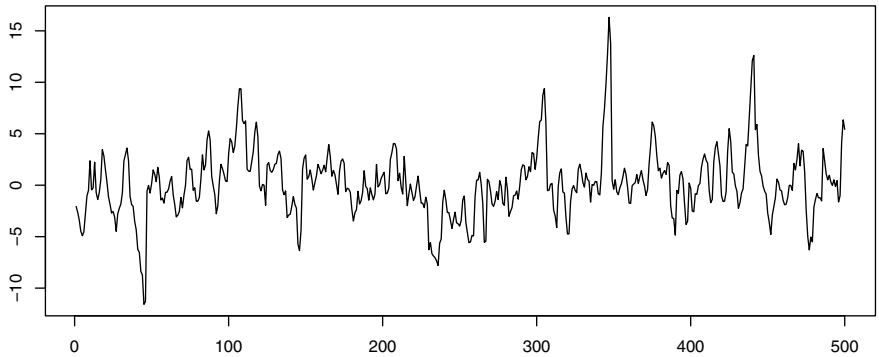


Figure 5.3: Last 500 points of simulated series with $t(2.5)$ errors

more clearly that all series are deviating form central values at the last point of the sample, point at which the forecasts are performed. We can see that the series with real roots, while looking similar to the series with complex roots, seems more noisy with a slightly smaller range, given the same errors trajectory. The same applies for the $t(2)$- and $t(2.5)$-distributed series, as depicted in Figures 5.2 and Figure 5.3 respectively. We can notice that the volatility of the sample significantly decreases as soon as the variance becomes finite, even though the error distribution still has fat tails.

To investigate the impact of the sample size on the forecasting method of Gouriéroux and Jasiak (2016), we take the last 200, 500, 1 000 or the whole sample of length 10 000, so that the last points are always the same. We employ Equation (5.8) with $s = 2$ and $H = 1$ to approximate the 1-step ahead predictive density and denote the method as GJ. For the approach that we propose in this chapter – namely using approximation (5.10) – we employ 10 000, 50 000, 100 000 or 200 000 simulations. For both methods we use the true parameters as we intend to disentangle the sensitivity of the prediction methods from the impact of the sample size on the estimation of the model.

Figures 5.4, 5.5 and 5.6 show the evolution of the one-step ahead predictive densities when the sample size (for the method GJ) and when the number of simulations (for the simulations-based approximations) increase. The figures are for $t(1)$-, $t(2)$- and $t(2.5)$-distributed processes respectively. The top 8 graphs are for coefficients $(\psi_1, \psi_2) = (0.5, 0.3)$ and the bottom graphs for $(\psi_1, \psi_2) = (1.1, -0.3)$. The first row displays the results of the sample-based method with different sample sizes and the bottom row for the approach proposed in this chapter for different number of simulations.

The first noticeable results is that the impact of the sample size and of the number of simulations decreases significantly for processes with

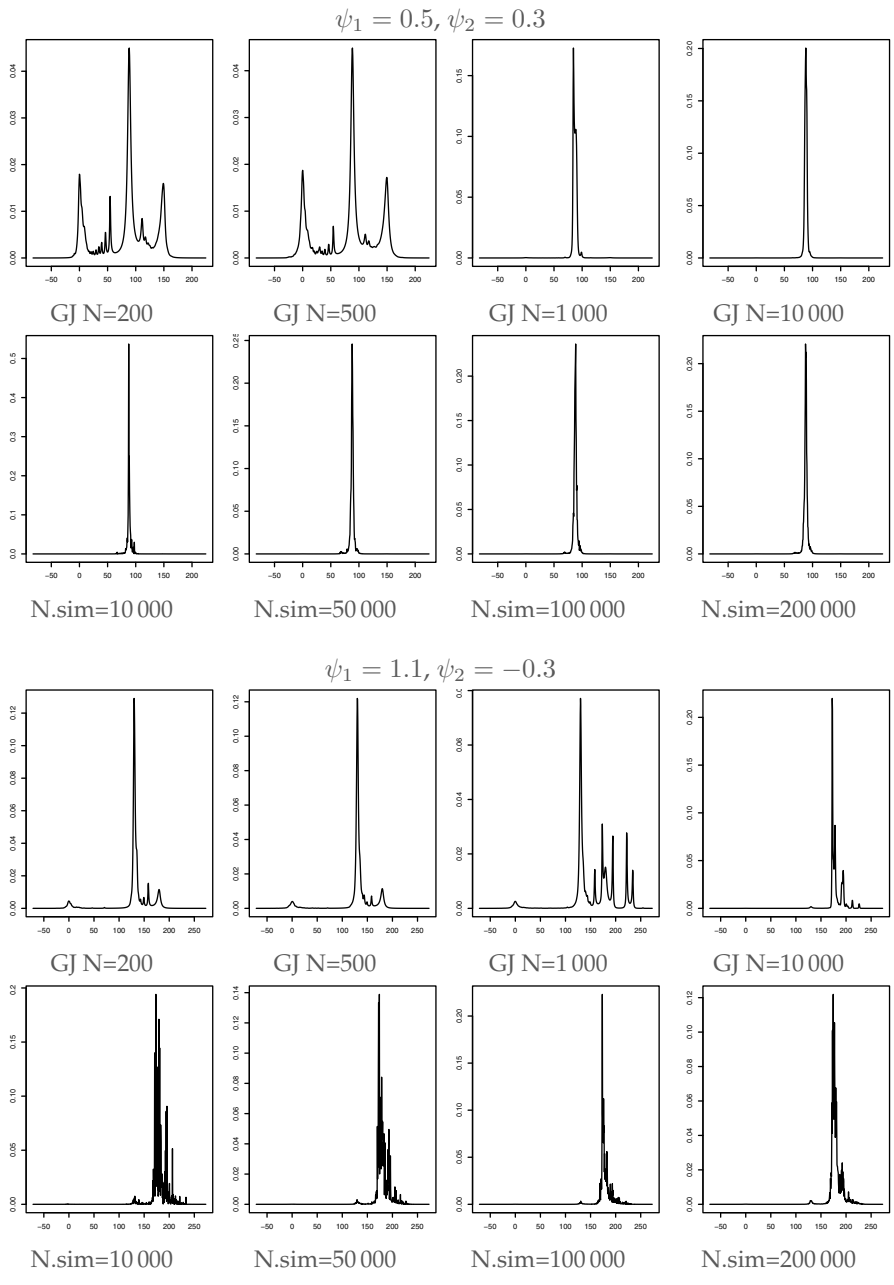


Figure 5.4: Density forecasts of series with $t(1)$ errors

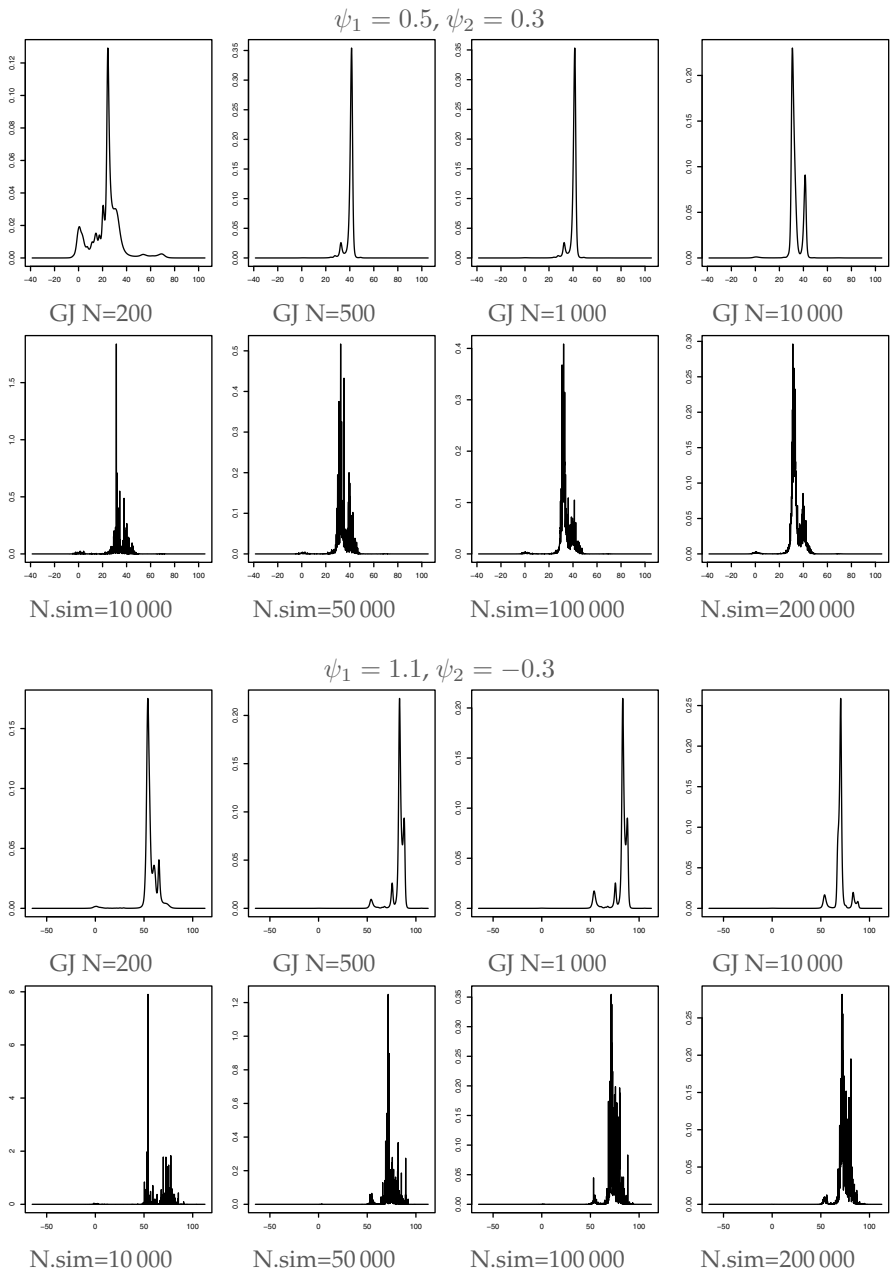


Figure 5.5: Density forecasts of series with $t(2)$ errors

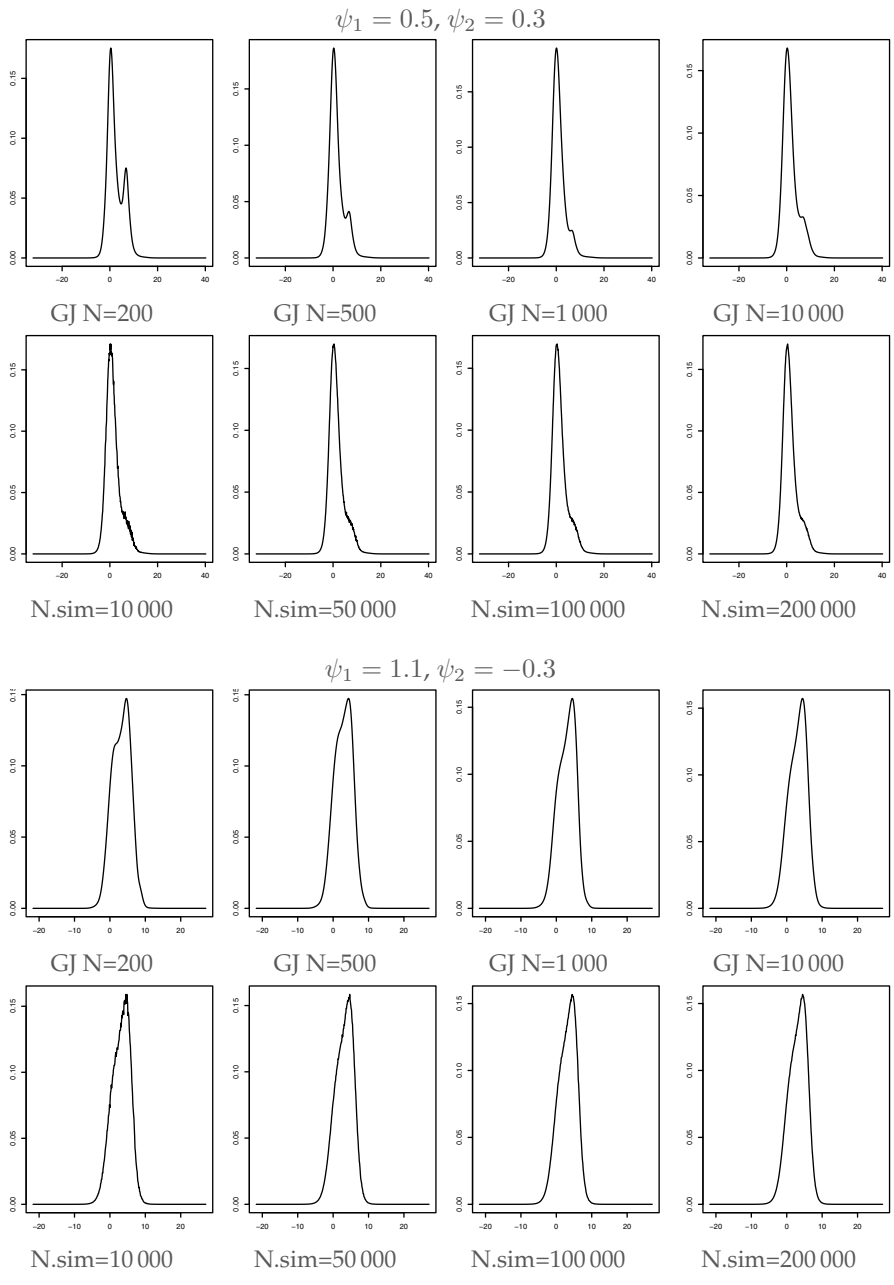


Figure 5.6: Density forecasts of series with $t(2.5)$ errors

finite variance (Figure 5.6). Indeed, due to the significantly lower volatility of the process with $t(2.5)$-distributed errors, a small amount of simulations and a small sample size is sufficient to recover the same predictive density with both methods.

The more noticeable differences between the GJ graphs of Figure 5.4 and 5.5 with different sample sizes is due to extreme values that are much more likely to happen with infinite variance. Based on the learning mechanism that this method has, which relies on past behaviours, the presence, or not, of extreme values in the studied sample can have a notable impact on the obtained conditional density. Figure 5.7 shows the complete simulated samples of the series with coefficients $(\psi_1, \psi_2) = (0.5, 0.3)$ to gauge the range and the recurrence of extreme values based on the distribution of the errors. As expected, both of them decrease as the degrees of freedom increase and it also illustrates how more extreme are the Cauchy-distributed errors.

The simulations-based method that we propose in this chapter yields much noisier results. We notice that the noise decreases as the number of simulations increases. Yet, 200 000 simulations for explosive episodes with errors with infinite variance sometimes does not appear to be enough. For the finite variance case however, as shown in Figure 5.6, 10 000 simulations provides the same results as 200 000.

Overall, results are in line with what Hecq and Voisin (2021) show. The discrepancies between simulations- and sample-based approaches are prevalent in more volatile processes. While the GJ method is affected by the sample size, the approach is much more efficient in obtaining a neat predictive density for large samples. Indeed, the simulations-based approach requires a significant amount of simulations to correct the noisiness of the results. However, as shown by Figure 5.4 with $(\psi_1, \psi_2) = (1.1, -0.3)$, even with a sample of 1 000 points, the density is still noticeably distorted from the learning

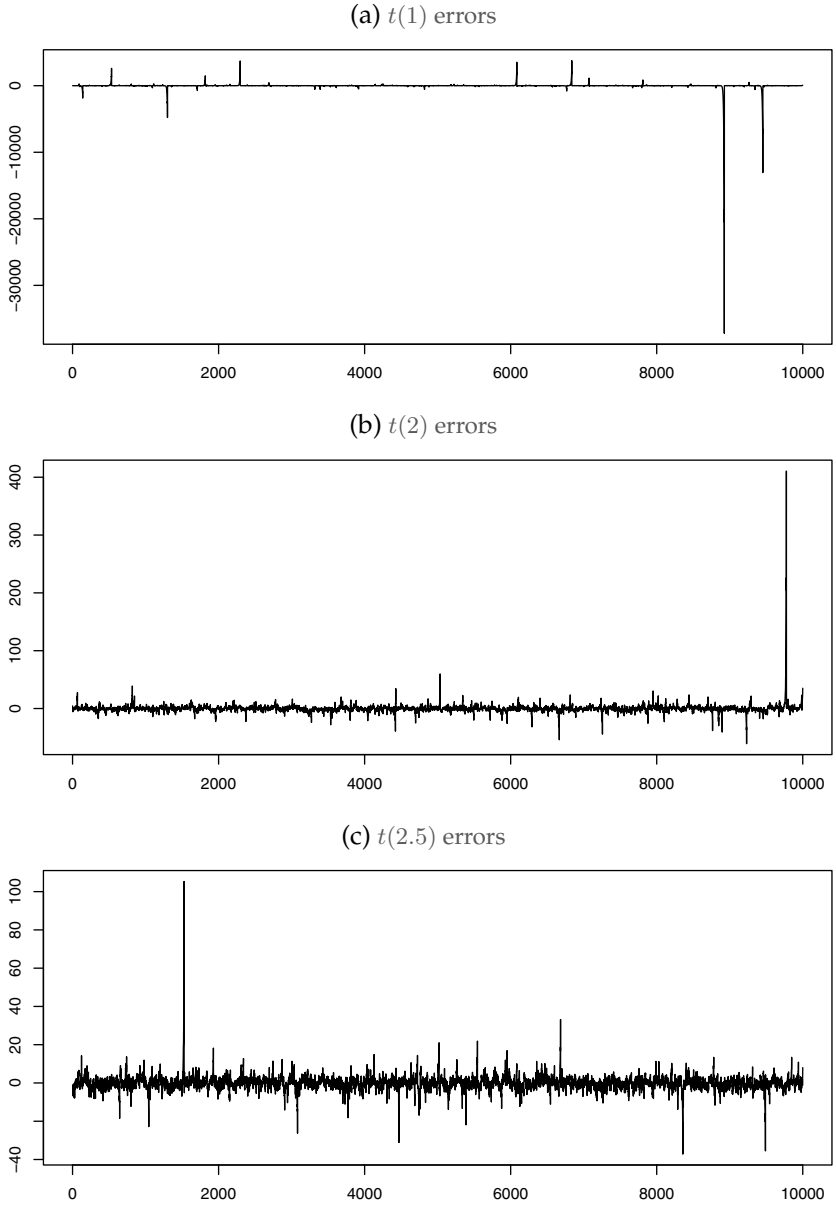


Figure 5.7: Entire sample of series with coefficients $(\psi_1, \psi_2) = (0.5, 0.3)$ for each errors distribution

mechanism of the method. The approaches are thus complementary, with a large enough sample size the GJ method will be more efficient, however when the sample size is too small, the simulations-based approach can correct for it. It also depends on the weight an applied researcher wants to put on past behaviours. If one believes that there is a strong learning process for instance, then the GJ method is more suitable. For more theoretical results, then simulations-based densities will sometimes be more accurate and the number of simulations can always be increased.

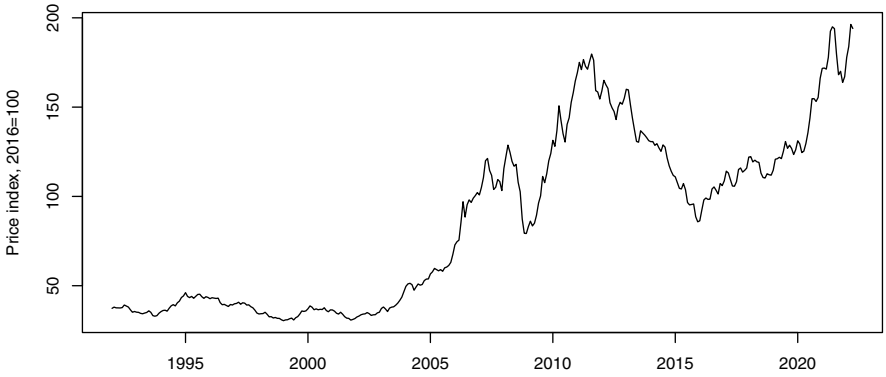
Future research could consist of a more theoretical analysis of the method proposed in this chapter. It could also consists in an analysis including less extreme episodes for instance or investigate 2-step ahead forecasts, which can still be easily visualised graphically. The implications of the 2-dimensional predictive densities obtained from the methods employed in this analysis for the path until *T+2* could be compared to the density forecast of the point at *T+2*, obtained from the method of Lanne, Luoto, and Saikkonen (2012).

5.5 An empirical example

As an empirical example we employ the monthly price index of all metals, retrieved from the IMF. The base year is 2016 and the index includes base and precious metals. The sample ranges from January 1992 until April 2022 yielding a sample of 364 data points. We detrend the series using the Hodrick-Prescott (HP) filter developed by Hodrick and Prescott (1997) and estimate the MAR(r, s) model on the remaining cycle.

The raw and detrended series are depicted in Figure 5.8. The most notable volatile periods are around the financial crisis of 2008 and the period during the COVID-19 pandemic. We will focus on the latter. The volatility in the last period of the sample was first induced by the

(a) All metals price index



(b) *HP*-detrended price index

Figure 5.8: All metals price index

various restrictions imposed to limit the spread of COVID-19 which was then followed by unprecedented levels of government stimuli. All these affected both the supply and the demand sides, which triggered the noticeable volatility in prices.

To estimate the MAR(r, s) model, we first estimate the pseudo lag order p using Schwarz information criterion on a Gaussian causal AR model. We identify the order of autocorrelation p to be equal to 2. Hence, we then compare the likelihood values of all MAR(r, s) models such that $r + s = 2$ with Student's t-distributed errors. The model with the highest likelihood value is an MAR$(0, 2)$.

The detrended metals price index (denoted as y_t) is therefore identified as an MAR$(0, 2)$ and the estimated coefficients are the following,

$$y_t = 1.27y_{t+1} - 0.34y_{t+2} + \varepsilon_t, \qquad \text{with} \quad \varepsilon_t \sim t(1.42).$$

Most commodity prices are indeed found to be forward looking (see for instance Fries and Zakoïan, 2019a; Hecq and Voisin, 2022).

We perform density forecasts using the sample-based approach of Gouriéroux and Jasiak (2016) and the simulations-based counterpart proposed in this chapter on the last 12 months of the sample. We perform pseudo-real life forecasts and thus detrend and re-estimate the model at each point of the predictions with expanding window. The first predictions performed are in May 2021, right before the noticeable peak of June 2021.

Figure 5.9 shows all 12 density forecasts made at the indicated months. The sample-based method is represented by the solid lines and the simulations-based method by the dashed lines, obtained with $200\,000$ simulations.

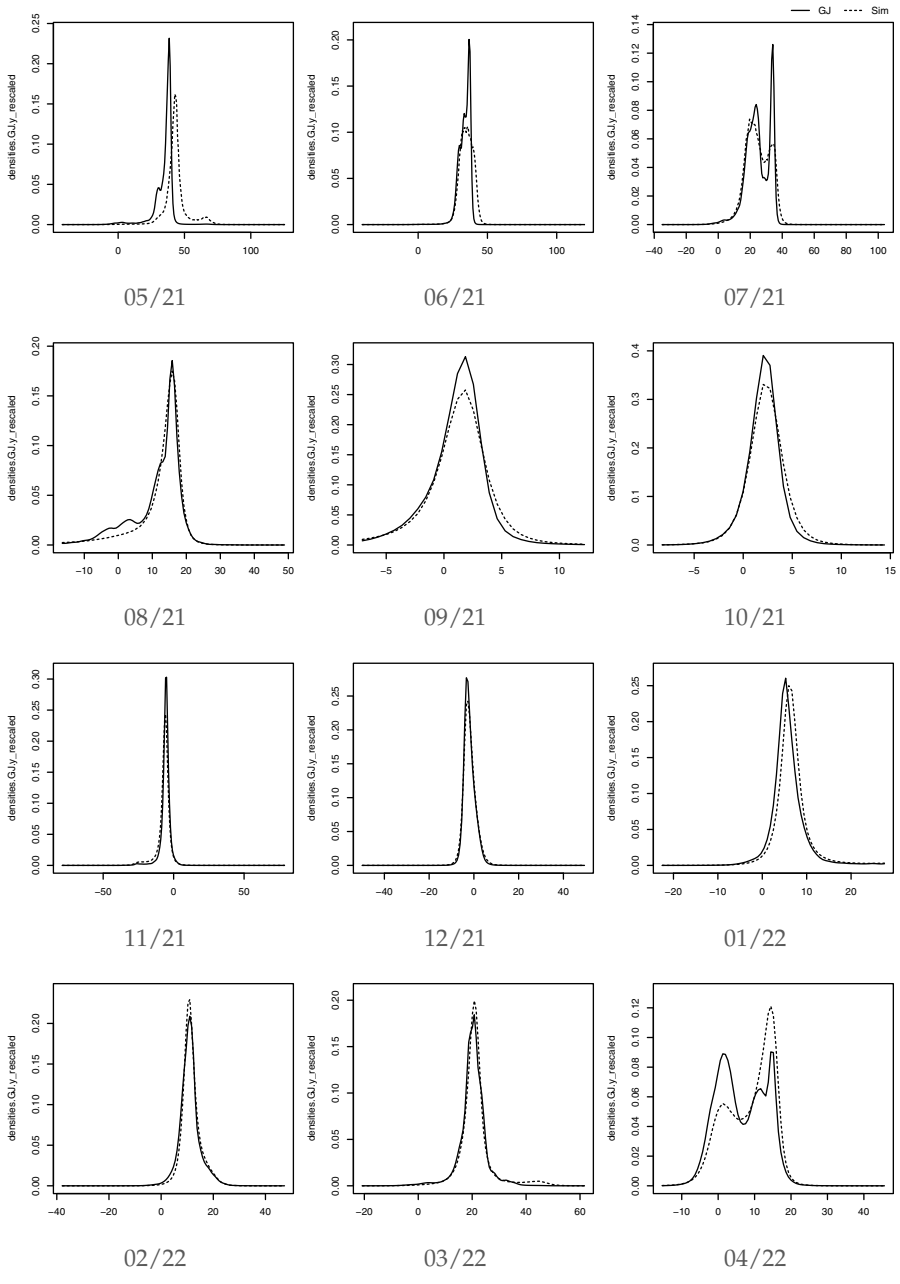


Figure 5.9: Density forecasts of detrended metals price index

We can notice that the densities across the two approaches differ the most around the peaks of locally explosive episodes. Indeed, they predominantly differ between May and July 2021 and in April 2022. The results are therefore in line with the findings of Section 5.4 for such sample size and t-distribution. First, for a $t(1.42)$-distributed process, the deviations from the estimated trend even at the peaks is not that extreme, which implies less discrepancy between the two methods. Then, when the detrended index is close to central values and thus when the predictive densities are unimodal, both approaches yield almost identical densities.

5.6 Conclusion

This chapters investigates the possibility to derive closed-form theoretical conditional density functions of Cauchy-distributed mixed causal-noncausal processes with two leads. While assuming a Student's t-distribution offers more flexibility than Cauchy, it entails the absence of closed-form for the conditional distribution of the process. Investigating the Cauchy case would allow to evaluate the benchmark performance of the methods aiming at approximating the predictive density. We could not obtain all parts of the conditional distribution and were therefore not able to obtain a closed-form. We present the marginal distribution of the MAR$(0, 2)$ and leave the rest to future research. We then propose a prediction approach which builds on the two existing approximation methods. It approximates the conditional distribution using simulations and does not depend on the sample size. In a simulation study, we compare and analyse its results with the ones of a sample-based counterpart, based on different distributions and parameters. We also analyse their respective sensitivity to the number of simulations employed and the sample size. We then illustrate both approaches by forecasting the metals price index which is identified as an MAR$(0, 2)$ process.

6

Detecting common bubbles in multivariate mixed causal-noncausal models

Adapted from: Gianluca Cubadda, Alain Hecq, and Elisa Voisin (2022). *Detecting common bubbles in multivariate mixed causal-noncausal models*. DOI: 10.48550/ARXIV.2207.11557.

Abstract

This chapter proposes methods to investigate whether the bubble patterns observed in individual series are common to various series. We detect the non-linear dynamics using the recent mixed causal and noncausal models. Both a likelihood ratio test and information criteria are investigated, the former having better performances in our Monte Carlo simulations. Implementing our approach on three commodity prices we do not find evidence of commonalities although some series look very similar.

6.1 Introduction

Economic and financial time series exhibit many distinctive characteristics among which the presence of serial correlation, some seasonality, stochastic or deterministic trends, time varying volatility, non-linearities. However, in multivariate analyses, namely when one investigates relationships between variables, it is frequent to observe that one or more of these features that were detected in individual series are common to several variables. We talk about common features when such features are annihilated with some suitable combinations. The most famous example is probably cointegration, that is the presence of common stochastic trends (Engle and Granger, 1987). Other forms of co-movements (Engle and Kozicki, 1993) have also been studied, giving rise to developments around the notions of common cyclical features (Vahid and Engle, 1993), common deterministic seasonality (Engle and Hylleberg, 1996), common volatility (Engle and Susmel, 1993), co-breaking (Hendry and Massmann, 2007), etc. Recognizing these common feature structures presents numerous advantages from an economic perspective (e.g. the whole literature on the existence of long-run relationships). There are also several implications for statistical modeling. Imposing the commonalities helps to reduce the number of parameters that must be estimated. That potentially leads to efficiency gains and to improvements in forecasts accuracy (Issler and Vahid, 2001). The factor structure underlying common features can also be used to forecast a set of time series using only the forecasts of the common component and the estimated loadings.

Building on such a common features approach, we propose in this chapter to detect the presence of common bubbles in stationary time series. Intuitively, the idea is to detect in univariate time series bubble patterns and to investigate whether those bubbles would be common to a set of assets. In the affirmative, a portfolio composed of those series would not have such a non-linear local explosive characteristic.

There are several ways to capture bubbles in the data. We rely on mixed causal-noncausal models (denoted MAR(r, s) hereafter), namely autoregressive time series that depend on both r lags and s leads. There is indeed a recent interest in the properties of noncausal processes associated with a blooming of applications on commodity prices, inflation or cryptocurrency series as well as the developments around the notion of non-fundamental shocks. See Hecq and Voisin (2022) for the references therein. We choose to consider mixed causal and noncausal models as they might also be used for forecasting. This is not necessarily the case with other approaches aiming at identifying bubble phases. In a MAR(r, s) framework, in the presence of common bubbles among a set of time series, several variables can be forecasted with one of the series (and their loadings).

Cubadda, Hecq, and Telg (2019) extend the canonical correlation framework of Vahid and Engle (1993) from purely causal vector autoregressive models (namely the traditional serial correlation common feature approach within a VAR) to purely noncausal VARs (a VAR with leads only). They show that more commonalities emerge when we also look at VARs in reverse time. The tests statistics they developed do not generally work for mixed models though. Consequently, we extend their work and we propose a likelihood ratio test that compares the unrestricted multivariate vector mixed causal-noncausal model (hereafter VMAR(r, s)) with its restricted version in which reduced ranks are imposed on the lead polynomial matrix, which is our notion of common bubbles. Indeed, this is equivalent to require that there exists a linear combination of variables with bubbles that does not possess the bubble feature. See for instance Cubadda and Hecq (2022) for a recent survey on reduced rank techniques for common features. We also consider the use of information criteria as an alternative strategy.

The rest of this chapter is as follows: in Section 6.2 we set up the notations for multivariate mixed causal and noncausal models. Contrarily

to the univariate case, two distinct multivariate multiplicative representations lead to the same linearized form. They consequently also have the same likelihood but with different lag-lead polynomial matrices. We advocate to use the VMAR with the lead polynomial first as the alternative matrix polynomial structure does not allow to discover the presence of common bubbles even when they are present. We explain how to implement the likelihood ratio test that we introduce in this paper. Section 6.3 investigates, using Monte Carlo simulations, the small sample properties of our strategy for a bivariate and three dimensional systems both under the null of common bubbles and the alternative of no rank reductions. Section 6.4 illustrates our approach on three commodity prices. The presence of common bubbles is rejected in every bivariate and three dimensional systems although from the graphs, series looked rather similar. Section 6.5 concludes.

6.2 Multivariate mixed causal-noncausal models

Recall that a univariate MAR(r, s) model is constructed as follows,

$$(1 - \phi_1 L - \ldots - \phi_r L^r)(1 - \psi_1 L^{-1} - \ldots - \psi_s L^{-s})y_t = e_t,$$

where all coefficients are scalars and thus the model is commutative,

$$(1 - \psi_1 L^{-1} - \ldots - \psi_s L^{-s})(1 - \phi_1 L - \ldots - \phi_r L^r)y_t = e_t.$$

That is, either of the two representations will yield the same coefficients. L^r is the lag operator such that $L^r y_t = y_{t-r}$ and L^{-s} is the lead operator such that $L^{-s} y_t = y_{t+s}$.

Let us now consider Y_t, an N dimensional stationary process. We assume for notation simplicity that deterministic elements such as the intercept or seasonal dummies have been subtracted. Analogously to the univariate case, a multivariate mixed causal-noncausal model with r lags and s leads, denoted VMAR(r, s), is defined in its multiplicative

forms as follows,

$$\Psi(L^{-1})\Phi(L)Y_t = \varepsilon_t, \tag{6.1}$$

$$\bar{\Phi}(L)\bar{\Psi}(L^{-1})Y_t = \bar{\varepsilon}_t. \tag{6.2}$$

where,

$$\Psi(L^{-1})\Phi(L) = (I_N - \Psi_1 L^{-1} - \ldots - \Psi_s L^{-s})(I_N - \Phi_1 L^1 - \ldots - \Phi_s L^r)$$
$$\bar{\Psi}(L^{-1})\bar{\Psi}(L) = (I_N - \bar{\Psi}_1 L^{-1} - \ldots - \bar{\Psi}_s L^{-s})(I_N - \bar{\Psi}_1 L^1 - \ldots - \bar{\Psi}_s L^r).$$

To simplify the analyses, ε_t and $\bar{\varepsilon}_t$ follow multivariate Student's *t*-distributions. We can consider other distributions than the Student as long as they are non Gaussian. This is indeed the condition that allows for the distinction of the causal, noncausal or mixed specifications. Both models (6.1) and (6.2) are equivalent but contrary to univariate MAR models, they are two distinct representations of the same process, given the non commutativity property of the matrix product. This means that the lag polynomial matrices $\Phi(L)$ and $\bar{\Phi}(L)$, though of the same order *r*, have different values for coefficient matrices. The same observation applies to the lead polynomials $\Psi(L^{-1})$ and $\bar{\Psi}(L^{-1})$ which are of the same order *s*. We assume that the roots of the determinants of each polynomial matrices $\Psi(L^{-1}), \Phi(L)$, $\bar{\Phi}(L), \bar{\Psi}(L^{-1})$ are outside the unit circle to fulfill the stationarity condition. Furthermore, we will show later that the distribution of the errors ε_t and $\bar{\varepsilon}$ have identical degrees of freedom λ but different scale matrices. We denote them by Σ and $\bar{\Sigma}$ respectively, both being symmetric positive definite matrices.

Let us further denote $A(L)$ and $\bar{A}(L)$ the expanded products[1] of the

[1]This is the restricted linear form that is used in the maximum likelihood estimation. Gourieroux and Jasiak (2017) have proposed an alternative approach based on roots inside and outside the unit circle of an autoregressive polynomial.

lag and lead polynomials in the two models with

$$\Psi(L^{-1})\Phi(L) \equiv A(L) = \sum_{j=-s}^{r} A_j L^j \quad \rightarrow \quad A(L)Y_t = \varepsilon_t,$$
$$\bar{\Phi}(L)\bar{\Psi}(L^{-1}) \equiv \bar{A}(L) = \sum_{j=-s}^{r} \bar{A}_j L^j \quad \rightarrow \quad \bar{A}(L)Y_t = \bar{\varepsilon}_t.$$

The general forms of the expansion of the lead and lag polynomials for each representation are

$$\begin{aligned}
A(L) &\equiv \sum_{j=-s}^{r} A_j L^j = \\
&I + \sum_{i=1}^{\substack{min \\ \{r,s\}}} \Psi_i \Phi_i - \sum_{i=1}^{r} \left(\Phi_i - \sum_{\substack{\forall\{l,m\} \\ s.t. \\ l-k=i}} \Psi_l \Phi_m \right) L^i - \sum_{j=1}^{s} \left(\Psi_j - \sum_{\substack{\forall\{l,m\} \\ s.t. \\ m-l=j}} \Psi_l \Phi_m \right) L^{-j}, \\
\bar{A}(L) &\equiv \sum_{j=-s}^{r} \bar{A}_j L^j = \\
&I + \sum_{i=1}^{\substack{min \\ \{r,s\}}} \bar{\Phi}_i \bar{\Psi}_i - \sum_{i=1}^{r} \left(\bar{\Phi}_i - \sum_{\substack{\forall\{l,m\} \\ s.t. \\ m-l=i}} \bar{\Phi}_m \bar{\Psi}_l \right) L^i - \sum_{j=1}^{s} \left(\bar{\Psi}_j - \sum_{\substack{\forall\{l,m\} \\ s.t. \\ m-l=j}} \bar{\Phi}_m \bar{\Psi}_l \right) L^{-j},
\end{aligned} \tag{6.3}$$

with $1 \leq l \leq s$ and $1 \leq m \leq r$. This shows that both multiplicative representations yield the exact same additive form,

$$\underbrace{B(L)}_{\substack{A_0^{-1}A(L) \\ = \\ \bar{A}_0^{-1}\bar{A}(L)}} Y_t = \underbrace{\eta_t}_{\substack{A_0^{-1}\varepsilon_t \\ = \\ \bar{A}_0^{-1}\bar{\varepsilon}_t}}, \tag{6.4}$$

where η_t follows a multivariate Student-t distribution with degrees of freedom λ – analogously to ε_t and $\bar{\varepsilon}_t$ from representations (6.1) and (6.2) – and with a scale matrix $\Omega = A_0^{-1}\Sigma(A_0^{-1})' = \bar{A}_0^{-1}\bar{\Sigma}(\bar{A}_0^{-1})'$. The lag polynomial in (6.4) is the following,

$$B(L) = I - \sum_{i=1}^{r} B_i L^i - \sum_{j=1}^{s} B_{-j} L^{-j}. \tag{6.5}$$

The example of derivations of the coefficients of a VMAR(2,2) is given in Section 6.2.1.

Contrary to the univariate case, with commutative multiplicative form and therefore a unique solution, a multivariate VMAR(r, s) processes has two distinct representations. The multivariate process can be estimated with either of the multiplicative representations (6.1) and (6.2). While the coefficient matrices will differ, both representations will correspond to the same expanded form of the model (6.4). This makes however the interpretation of the lag and lead coefficient matrices in the multiplicative forms more intricate. Lanne and Saikkonen (2013) advocate for the use of one or the other representation depending on the analysis performed; one representation might be easier to employ for certain inquiries.

6.2.1 Common bubbles in VMAR(*r*,*s*)

Now that we have set up the notations of the unrestricted multivariate mixed model we consider additional restrictions coming from commonalities in the lead polynomial matrix. Indeed it is the lead component that induces some non-linearities similar to the bubble pattern (Gouriéroux and Zakoïan, 2013).Although the focus in this chapter is on common bubbles, our approach can be easily extended the investigation to commonalities in the causal part or in both the lag and the lead components.

Definition 1. *The N dimensional process Y_t displays common bubbles (hereafter CB) if there exists a matrix δ of dimension $N \times k$, with $0 < k < N$, such that, $\delta' B_{-j} = 0$ for $j = 1, \ldots, s$, where the coefficient matrix B_{-j} is a matrix of the expanded lag polynomial* (6.5)*. This implies that the coefficient matrices B_{-j} can be decomposed as $\delta_\perp \beta_j'$ where $\delta_\perp$ is the $N \times (N-k)$ orthogonal complement of δ' such that $\delta' \delta_\perp = 0$ and β_j' is a matrix with dimension $(N-k) \times N$.*

Let us start from the example $r = s = 2$. The coefficient matrices of the leads in the additive representation (6.4) with reduced rank restrictions are

$$\begin{aligned} B_{-1} &= A_0^{-1}(\Psi_1 - \Psi_2\Phi_1) & B_{-2} &= A_0^{-1}\Psi_2 \\ &= \bar{A}_0^{-1}(\bar{\Psi}_1 - \bar{\Phi}_1\bar{\Psi}_2) & &= \bar{A}_0^{-1}\bar{\Psi}_2 \\ &= \delta_\perp \beta_1', & &= \delta_\perp \beta_2', \end{aligned}$$

where the matrices A_0 and $\bar{A}_0$ have been derived from the expanded lag polynomials (6.3) with

$$\begin{aligned} A_0 &= (I_N + \Psi_1\Phi_1 + \Psi_2\Phi_2) \\ \bar{A}_0 &= (I_N + \bar{\Phi}_1\bar{\Psi}_1 + \bar{\Phi}_2\bar{\Psi}_2). \end{aligned}$$

Hence, the matrix δ' of dimension $k \times N$, with $0 < k < N$ annihilates the forward looking dynamics

$$\delta' B_{-1} = \delta' B_{-2} = 0.$$

This implies that, for the second lead coefficients,

$$\delta' B_{-2} = \delta' A_0^{-1}\Psi_2 = \delta' \bar{A}_0^{-1}\bar{\Psi}_2 = 0.$$

Since $\delta' A_0^{-1}$ cannot be equal to zero, it implies that $\delta' A_0^{-1} = \gamma'$ (resp. $\delta' \bar{A}_0^{-1} = \bar{\gamma}'$) , where γ (resp. $\bar{\gamma}$) is some $N \times k$ dimensional matrix and thus $\gamma' \Psi_2 = 0$ (resp. $\bar{\gamma}' \bar{\Psi}_2 = 0$). Hence, both Ψ_2 and $\bar{\Psi}_2$ must have rank $N - k$, but potentially different left null spaces (see also

Cubadda, Hecq, and Telg, 2019).

For the first lead coefficient of representation (6.1),

$$\begin{aligned}\delta' B_{-1} &= \delta' A_0^{-1}(\Psi_1 - \Psi_2\Phi_1) \\ &= \gamma'(\Psi_1 - \Psi_2\Phi_1) \\ &= \gamma'\Psi_1 = 0\end{aligned}$$

which implies that Ψ_1 and Ψ_2 must have the same left null space. For the alternative representation in (6.2),

$$\begin{aligned}\delta' B_{-1} &= \delta' \bar{A}_0^{-1}(\bar{\Psi}_1 - \bar{\Phi}_1\bar{\Psi}_2) \\ &= \bar{\gamma}'(\bar{\Psi}_1 - \bar{\Phi}_1\bar{\Psi}_2) = 0\end{aligned}$$

which implies that $\bar{\gamma}'\bar{\Psi}_1 = \bar{\gamma}'\bar{\Phi}_1\bar{\Psi}_2$ and that $\bar{\Psi}_1$ might not necessarily even be a reduced-rank matrix.

Since Ψ_1 and Ψ_2 of representation (6.1) must have the same left null space and keeping in mind that $\delta' A_0^{-1} = \gamma'$ we have,

$$\delta' = \gamma' A_0 = \gamma'(I_N + \Psi_1\Phi_1 + \Psi_2\Phi_2) = \gamma',$$

which shows that Ψ_1 and Ψ_2 have the same left null space as B_{-1} and B_{-2}.

For representation (6.2), only the second lead coefficient matrix must have reduced rank. The lead coefficients of the two representations in

the presence of CB are the following,

$$\Psi_1 = A_0\delta_\perp(\beta_1' + \beta_2'\Phi_1) = (I_N + \Psi_1\Phi_1 + \Psi_2\Phi_2)\delta_\perp(\beta_1' + \beta_2'\Phi_1),$$
$$\Psi_2 = A_0\delta_\perp\beta_2' = (I_N + \Psi_1\Phi_1 + \Psi_2\Phi_2)\delta_\perp\beta_2',$$

$$\begin{aligned}\bar{\Psi}_1 &= \bar{A}_0\delta_\perp\beta_1' + \bar{\Phi}_1\bar{A}_0\delta_\perp\beta_2' \\ &= (I_N + \bar{\Phi}_1\bar{\Psi}_1 + \bar{\Phi}_2\bar{\Psi}_2)\delta_\perp\beta_1' + \bar{\Phi}_1(I_N + \bar{\Phi}_1\bar{\Psi}_1 + \bar{\Phi}_2\bar{\Psi}_2)\delta_\perp\beta_2', \\ \bar{\Psi}_2 &= \bar{A}_0\delta_\perp\beta_2' = (I_N + \bar{\Phi}_1\bar{\Psi}_1 + \bar{\Phi}_2\bar{\Psi}_2)\delta_\perp\beta_2'.\end{aligned}$$

While it is clear that Ψ_1, Ψ_2, B_{-1} and B_{-2}, having the same left null space, satisfies the condition for the presence of a CB, we can see that pre-multiplying $\bar{\Psi}_1$ and $\bar{\Psi}_2$ by δ' will not annihilate the dynamics. This due the matrix multiplication structure of their components, as revealed by the right-hand sides of the equations. For instance, from the equation of $\bar{\Psi}_2$,$\delta'(I_N + \bar{\Phi}_1\bar{\Psi}_1 + \bar{\Phi}_2\bar{\Psi}_2)$ would have to simplify to δ' to annihilate the dynamics, however, this would require restrictions on the lag coefficients.

Overall, it is easy to see that the same conclusion applies for any VMAR(*r*,*s*). In the presence of CB, all lead matrices of representation (6.1) must have the same left null space. Hence, for the investigation of CB this representation is the most appropriate and straightforward to allow for tests.

6.2.2 Testing for common bubbles

In an $N-$dimensional VMAR(r, s), our definition of common bubbles implies that there exists a full-rank $N \times k$ matrix δ, with $0 < k < N$ that annihilates the noncausal dynamics of the multivariate process (6.4), i.e. $\delta' B_{-j} = 0$ for all $j = 1, \ldots, s$. It therefore entails that $\delta' Y_t$ is a purely causal k-dimensional process. We showed above that in the presence of CB all lead matrices estimated from the multiplicative representation

(6.1) must have the same left null space as the coefficients B_{-j}. They can thus be decomposed as

$$\Psi_j = \delta_\perp \Gamma_j', \qquad \text{with} \quad j = 1, \ldots, s, \tag{6.6}$$

with $\delta_\perp$ the aforementioned $N \times (N-k)$ matrix such that $\delta'\delta_\perp = 0$ and Γ_j' a $(N-k) \times N$ full-rank matrix.

We suggest a likelihood ratio test comparing the likelihood value of the unrestricted model (6.1) with the likelihood value of the restricted model

$$(I_N - \delta_\perp \Gamma_1' L^{-1} - \cdots - \delta_\perp \Gamma_s' L^{-s})(I_N - \Phi_1 L - \cdots - \Phi_r L^r) Y_t = \varepsilon_t, \tag{6.7}$$

where the coefficient matrices are as defined in (6.6). Since $\delta_\perp$ has dimension $N \times (N-k)$ with $0 < k < N$, there are $N-1$ possible reduced-rank model to consider, for all possible k. Furthermore, the matrix δ' is normalized, such that $\delta' = [I_k, \ \delta^*]$ and thus only involves $k \times (N-k)$ free parameters in δ^*. Hence, the test statistic based on the log-likelihood values,

$$LR_k = -2[\ln(L_0^{(k)}) - \ln(\hat{L})], \tag{6.8}$$

follows, under the null of CB, a χ^2_ρ distribution, with $\rho = k^2 - Nk(1-s)$, against the alternative of a full rank.

An alternative is to employ information criteria as a selection method between the restricted and unrestricted specifications,

$$BIC_k = K \ln(T) - 2 \ln(\hat{L}), \tag{6.9}$$

$$AIC_k = 2K - 2 \ln(\hat{L}), \tag{6.10}$$

with K the number of coefficients estimated in the lag and lead matrices of the model – for the unrestricted model, $K_u = N^2(r+s)$

and for the restricted models, $K_{r_k} = K_u - (k^2 - Nk(1-s)) = K_u - \rho$.

Alternatively, one can test the null hypothesis that the lead coefficient matrices have rank $0 < (N-k) < N$ against the alternative that they do have commonalities, however that they have a larger rank $(N-k) < (N-l) < N$. In such case, the difference in the number of estimated coefficients $\rho = N(1-s)(k-l) + l^2 - k^2$ is smaller than against the alternative of full rank.

6.3 Monte Carlo analysis

We investigate using Monte Carlo simulations the performance of our strategies to detect common bubbles in bivariate and trivariate VMAR(1,1) models. We consider two sample sizes ($T = 500$ and 1000) and two different degrees of freedom of the error term with very leptokurtic distributions, namely $\lambda = 3$ and 1.5, to respectively consider a finite and infinite variance case. We employ lead coefficient matrices with and without reduced rank to analyse the detection of the correct model under the null of common bubbles and under the alternative of no such co-movements. The coefficients employed in the bivariate settings are displayed in Table 6.1.

Results, based on 3 000 replications for each combination of parameters, are reported in Table 6.4.[2] All entries are the frequency of correctly detected model. That is, under the null of a CB, we report the proportion of correctly detected CB, and under the alternative of no CB, we report the proportion of correctly rejected CB. We hence perform the test $H_0 : rank(\Psi) = 1$ against the alternative that the rank is 2. The LR tests are performed at a 95% confidence level. The information criteria detect a CB when the IC of the restricted model is

[2]Estimating multivariate causal-noncausal models using maximum likelihood presents high sensitivity to starting values. We do not investigate this matter here and therefore employ true values as starting values in the estimations.

Table 6.1: Monte Carlo parameters for bivariate VMAR(1,1)

$$\Phi = \begin{bmatrix} 0.5 & 0.1 \\ 0.2 & 0.3 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 4 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$T = \{500,\ 1\,000\}$$

$$\lambda = \{1.5,\ 3\}$$

$$\Psi = \begin{cases} \begin{bmatrix} 0.3 & 0.25 \\ 0.6 & 0.5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 0.3 & 0.25 \end{bmatrix} & (H_0 : \text{CB}) \\ \begin{bmatrix} 0.1 & 0.4 \\ 0.6 & 0.5 \end{bmatrix} & (H_1 : \text{no CB}) \end{cases}$$

lower than the one of the unrestricted model.

We can notice that the frequency of Type I errors of the LR test increases when the variance of the errors becomes infinite. It however does not significantly decrease when the sample size gets larger. With finite variance ($\lambda = 3$) the LR test has an appropriate size of around 5.5% and it increases to around 8.6% when the degrees of freedom of the errors distribution reach 1.5. Under the alternative, the LR test has a power of at least 99.9% across all parameters combinations implying that it almost never detects a CB when there are none.

Regarding the model selection using information criteria, results show that BIC outperforms AIC. Under the null of a CB, BIC selects the correct model specification in 98.9% of the cases with finite variance and a sample size of 500. The frequency increases to 99.3% when the sample size increases to 1 000. AIC on the other hand selects the

Table 6.2: MC results for N=2

	$\lambda = 3$					
	T=500			T=1000		
DGP	LR test	BIC	AIC	LR test	BIC	AIC
With CB (rank 1)	0.946	0.989	0.838	0.944	0.993	0.834
Without CB (rank 2)	0.999	0.994	1.000	1.000	1.000	1.000
	$\lambda = 1.5$					
	T=500			T=1000		
DGP	LR test	BIC	AIC	LR test	BIC	AIC
With CB (rank 1)	0.913	0.968	0.779	0.914	0.977	0.783
Without CB (rank 2)	0.999	0.999	0.999	1.000	1.000	1.000

Based on 3000 iterations. All results are the frequencies of correctly detected models. The LR test is performed at a 95% confidence level. For the IC, the favoured model is the one with the lowest IC value. The ranks refer to the rank of the lead coefficient in the DGP.

correct model in only 83.8% of the cases and does not increase with the sample size. The frequency of correctly selected model decreases for both when in the infinite variance case, but more drastically for AIC, which decreases to around 78%. BIC still selects the correct model for 96.8% of the cases with a sample size $T = 500$, and the frequency increases to 97.7% for $T = 1\,000$. Under the alternative of no CB however, both IC correctly select the unrestricted specification in more than 99.4% across all parameters combinations.

We now turn to the trivariate case. Now, in the presence of a CB, the rank of the lead coefficient matrix can be either 1 or 2. We thus consider the two possible CB structures. The parameters of the data generating processes are displayed in Table 6.4.

We evaluate our approach with 1 500 replications with each of the parameters combinations. Under the null of a CB we test the correct CB specification against the alternative of the unrestricted full rank

Table 6.3: Monte Carlo parameters for trivariate VMAR(1,1)

$$\Phi = \begin{bmatrix} 0.5 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.1 \\ 0.1 & 0.4 & 0.6 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 4 \end{bmatrix}$$

$$T = \{500,\ 1\,000\}$$

$$\lambda = \{1.5,\ 3\}$$

$$\Psi = \begin{cases} \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ 0.2 & 0.3 & 0.4 \\ 0.7 & 0.35 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0.5 \end{bmatrix} \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ 0.2 & 0.3 & 0.4 \end{bmatrix} & (H_0: \text{ 1 CB feature}) \\ \begin{bmatrix} 0.15 & 0.25 & 0.4 \\ 0.3 & 0.5 & 0.8 \\ 0.075 & 0.125 & 0.2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0.5 \end{bmatrix} \begin{bmatrix} 0.15 & 0.25 & 0.4 \end{bmatrix} & (H_0: \text{ 2 CB features}) \\ \begin{bmatrix} 0.3 & 0.2 & 0.1 \\ 0.2 & 0.5 & 0.4 \\ 0.7 & 0.125 & 0.2 \end{bmatrix} & (H_1: \text{ no CB feature}) \end{cases}$$

model. Under the alternative of no CB we test for each of the CB specifications.[3] Table 6.4 reports the frequencies of correctly detected models either with the LR test or with model selection using the information criteria. Analogously to the bivariate case, the LR tests are performed at a 95% confidence level and the information criteria detect a CB when the IC of the restricted model is lower than the one of the unrestricted model.

We can notice that the size of the LR test when the true rank of the

[3] Results for other tests, such as 1 vs 2 when the true rank is 2 for instance, are available upon requests.

Table 6.4: MC results for N=3

		$\lambda = 3$					
		T=500			T=1000		
$rank(\Psi)$	Rank test	LR	BIC	AIC	LR	BIC	AIC
2	2 vs 3	0.944	0.984	0.817	0.951	0.992	0.843
1	1 vs 3	0.919	1.000	0.871	0.933	1.000	0.883
3	2 vs 3	0.695	0.481	0.855	0.932	0.802	0.970
	1 vs 3	1.000	1.000	1.000	1.000	1.000	1.000

		$\lambda = 1.5$					
		T=500			T=1000		
$rank(\Psi)$	Rank test	LR	BIC	AIC	LR	BIC	AIC
2	2 vs 3	0.915	0.972	0.775	0.907	0.978	0.776
1	1 vs 3	0.857	0.998	0.774	0.860	0.999	0.783
3	2 vs 3	0.997	0.994	0.999	1.000	1.000	1.000
	1 vs 3	1.000	1.000	1.000	1.000	1.000	1.000

Based on 1 500 iterations. All results are the frequencies of correctly detected models. The LR test is performed at a 95% confidence level. For the IC, the favoured model is the one with the lowest IC value. The ranks refer to the rank of the lead coefficient. $rank(\Psi)$ is the rank of the lead coefficient matrix in the DGP.

lead coefficient matrix is 2 is similar to the bivariate case. With a finite variance errors distribution the size of the LR test is around 5% and it increases to around 9% when the variance is infinite ($\lambda = 1.5$). We can see that the size of the test decreases in the more restrictive CB specification, when the rank of the matrix is 1. For the finite variance cases the size decreases to 91.9% when $T = 500$ and to 93.3% when $T = 1\,000$. The correctly detected model frequency decreases further to 86% in the infinite variance case. Under the alternative of no CB, with finite variance and a sample size of $T = 500$, the LR test wrongly detects a bubble (*2 vs 3*) in 30.5% of the cases, however this frequency

decreases to 6.8% when the sample size increases to 1 000. Hence, it seems that with a smaller sample size and the finite variance of the errors distribution, estimating 8 coefficients in the lead matrix instead of 9 in the unrestricted model still provide a good enough fit to not be rejected by the test. The power of the test for all other model specification is above 99.7%.[4]

When it comes to model selection using information criteria, BIC outperforms AIC in each of the settings to detect common bubbles. BIC correctly select a model with CB in more than 97.2% of the cases across all model specifications and the frequencies increase with the sample size and the amount of restricted coefficients. Indeed, it correctly selects a restricted model with a coefficient matrix of rank 1 in at least 99.8% of the cases. Whereas AIC selects the correct restricted model in less than 88.3% and the frequency decreases with the sample size, the variance of the errors and when the rank of the restricted matrix is closer to full rank. Hence for the infinite variance case with a sample size $T = 500$, its frequency of correctly selected CB model is around 77.5% for each of the CB specification. Under the alternative of no CB, we observe the same pattern as for the LR test. In the finite variance case, both information criteria over select a restricted model with a matrix of rank 2. For a sample size of 500, BIC selects the restricted model in 51.9% of the cases, though it decreases to 19.8% when the sample size increases to 1 000. AIC on the other hand only selects the restricted model in 14.5% with $T = 500$ and it even decreases to 3% with $T = 1\,000$. For all other model specification both IC select the correct model in at least 99.4% of the cases.

Overall, the size of the LR test seems to converge to 5% in the finite

[4]Something to take into account is that under the null of no CB, in the estimations of the restricted models with coefficient matrix of rank 2 or 1, the likelihood function might not have reached the global maximum due to the starting values issue. This could imply an overestimation of the frequencies displayed in the *2 vs 3* and *1 vs 3* when the true rank is 3.

variance cases when the sample size increases. In the infinite variance cases, the size is around 5 percentage points lower and seems to be less affected by the sample size. The power of the test is above 93% in all model specifications except with $\lambda = 3$ and $T = 500$, with a restricted model that has only 1 coefficient less to estimate than the unrestricted model (*2 vs 3*). For the model selection using information criteria, BIC overall outperforms AIC in correctly detecting a CB, but also tends to detect a CB more often than AIC when there is none in the *2 vs 3* case with $\lambda = 3$.[5]

6.4 Common bubbles in commodity indices?

We illustrate our strategies to test for common bubbles in mixed causal-noncausal processes on three commodity price indices: food and beverage, industrial inputs[6] and fuel (energy)[7]. The sample of 362 data points ranges from January 1992 to January 2022.[8] We can see from graphs (a) of Figures 6.1 and 6.2, which respectively shows the series in levels and logs, that the indices seem to follow similar trends. Long-lasting increases and crashes often happen at the same time. This could potentially suggest the presence of common bubbles between the series. Following the work of Hecq and Voisin (2022), we detrend all series using the Hodrick-Prescott filter (hereafter HP filter). Although this approach to get stationary time series has been strongly criticized, in particular for the investigation of business cycles, Hecq and Voisin (2022) show that it is a convenient strategy to preserve the bubble features. They also show in a Monte Carlo simulation that this is the filter that preserves the best the identification of the MAR(r, s)

[5]Note that Hannan-Quin information criterion $HQC = 2Kln(ln(T)) - 2\ln(\hat{L})$ performs exactly in between BIC and AIC both under the null and under the alternative. We thus omit it to save space but results are available upon request.

[6]Includes agricultural raw materials which includes timber, cotton, wool, rubber and hides.

[7]Includes crude oil, natural gas, coal and propane.

[8]All the data is retrieved from the IMF database. They are price indices with base year 2016.

model. Giancaterini, Hecq, and Morana (2022) reinforce the same conclusion using analytical arguments.

Detrended series are displayed on graphs (b) of the two Figures. It can indeed be seen that the dynamics inherent to mixed causal-noncausal processes mentioned above are preserved. The crashes occurring during the financial crisis of 2007 and the COVID-19 pandemic in 2020, while being of different magnitude, happened at the same time on all three series. Furthermore, long lasting increases such as the one before the financial crash, the recovery around 2009 or after 2020 are also present in all three index prices.

We first analyze the series individually. We estimate pseudo causal autoregressive models to identify the order of autocorrelation in each of the detrended series (both in levels and logs). All models that we identify using BIC end up to be AR(2) processes. The normality of the errors is rejected for all series: values of the Jarque-Bera statistics range between 48 and 253 for the 6 series. The next step is to identify MAR(r, s) models for all r and s subject to the constraint $p = r + s = 2$, namely MAR(2,0), MAR(1,1) or MAR(0,2). Based on the maximum likelihood estimator with Student's t-distributed error term, the best fitting model for all six series is a MAR(1,1) model.

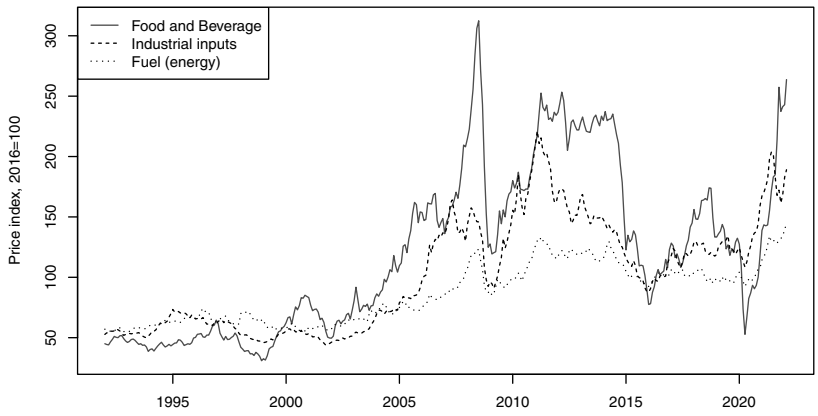
The estimated models are shown in Table 6.5.[9] For comparison purposes with the trivariate case shown later, we display both the coefficients estimated from the multiplicative from

$$(1 - \phi L)(1 - \psi L^{-1})y_t = \varepsilon_t, \qquad \text{with} \quad \varepsilon_t \sim t(\lambda), \tag{6.11}$$

but also the coefficients b_1 and b_2 of the expanded form obtained after

[9] We use different starting values in the estimation to account for the bimodality of the coefficients (see Bec, Nielsen, and Saïdi, 2020a, for more details).

(a) Indices in levels



(b) HP-detrended levels

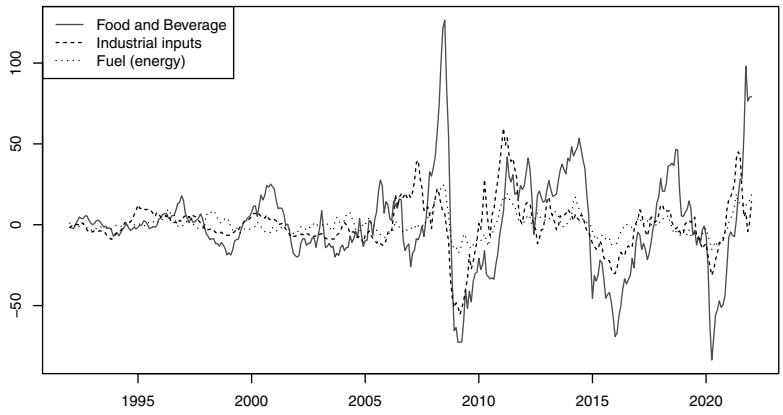
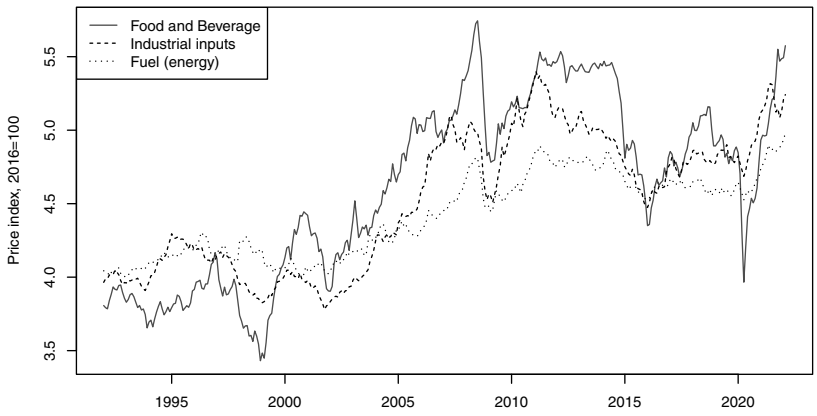


Figure 6.1: Price indices in levels

estimations,

$$\begin{aligned} y_t &= \frac{\phi}{1+\phi\psi} y_{t-1} + \frac{\psi}{1+\phi\psi} y_{t+1} + \varepsilon_t^* \\ &= b_1 y_{t-1} + b_2 y_{t+1} + \varepsilon_t^*. \end{aligned} \tag{6.12}$$

(a) Indices in logs



(b) HP-detrended logs

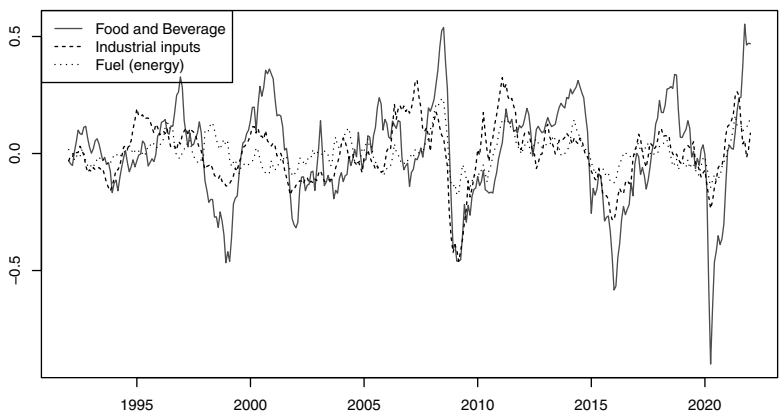


Figure 6.2: Price indices in logs

It emerges that "food and beverage" as well as "industrial inputs" are both mostly forward looking with lead coefficients close to 0.85 and lag coefficients around 0.4. On the opposite, the "fuel index" appears more backward looking with coefficients inverted. Except for the level

of industrial inputs, all models have error terms with finite variance, and as expected, one obtains lower variance for the logs of the series. The similar dynamics between food and beverages and industrial inputs could indicate commonalities. The same conclusions can be drawn from the rescaled expanded coefficients.

Table 6.5: Estimated coefficients on univariate MAR(1,1) models

Variable	Estimated coefficients				
	Multiplicative			Expanded	
	ϕ	ψ	λ	b_1	b_2
Food and Beverage	0.38	0.85	3.70	0.29	0.64
log(Food and Beverage)	0.34	0.86	5.47	0.26	0.67
Industrial inputs	0.43	0.87	1.66	0.31	0.63
log(Industrial inputs)	0.42	0.89	4.62	0.31	0.65
Fuel (energy)	0.87	0.44	2.20	0.63	0.32
log(Fuel)	0.83	0.48	4.95	0.59	0.34

The coefficients in the multiplicative form are the estimated coefficients from equation (6.11). The expanded coefficients are the ones obtained after expanding the multiplicative from like in (6.12).

For the multivariate investigations, we analyze both bivariate and trivariate systems. Similarly to the univariate estimation, the strategy consists in first estimating the pseudo lag order p using a standard VAR(p) for the six bivariate combinations (three in levels and three in logs) and the two trivariate models. Using BIC, all VARs are identified as VAR(2). There are starting values issues when estimating VMARs by maximum likelihood, meaning that we often reach local maxima. To avoid this, we used a large range of starting values to estimate VMAR(1,1) with multivariate Student's t-distributed errors and we keep the estimated model with the highest likelihood value.[10]

[10]We fixed the starting values for the correlation matrix Σ and the degrees of freedom

The estimated models are shown below in Table 6.6. We employ representation (6.1) for the estimation but the coefficients displayed are those of the additive form (6.4), which are independent of the representation used for the estimations in the following form,

$$Y_t = B_1 Y_{t-1} + B_{-1} Y_{t+1} + \eta_t,$$

where η_t follows a multivariate Student's t-distribution with λ degrees of freedom and correlation matrix Ω.

Comparing to the expanded coefficients b_1 and b_2 of the univariate models in 6.5, the directions and magnitudes of the dynamics have been preserved in the multivariate models estimations. From the off-diagonal coefficients of the bivariate models, we notice that 'Food' is impacting both 'Indus' and 'Fuel' with the lag and the lead, with coefficients magnitude between 0.11 and 0.47 for the levels. However in the other direction, the magnitude of the coefficients does not exceed 0.05 for the lag of 'Fuel' on 'Food'. 'Fuel' slightly impacts 'Indus' with coefficients of magnitude around 0.1. These dynamics can also be observed in the trivariate model.

To perform the common bubble tests we estimated VMAR models with restrictions on the lead coefficients matrix as shown in (6.7).[11] In the trivariate settings the LR test and the information criteria compares the unrestricted model where the lead matrix has full rank with both CB specifications, namely imposing rank 2 or rank 1 to the lead coefficient matrix.

λ and performed 100 MLEs based on random lead and lag coefficient matrices fulfilling stationary conditions.

[11] We also used 100 combinations of starting values to make sure we obtain the best fitting models.

Table 6.6: Estimated coefficients on the multivariate VMAR(1,1) models

B_1	B_{-1}	Ω	λ
Food and Indus			
$\begin{bmatrix} 0.28 & 0.01 \\ 0.26 & 0.27 \end{bmatrix}$	$\begin{bmatrix} 0.65 & -0.02 \\ -0.11 & 0.65 \end{bmatrix}$	$\begin{bmatrix} 1.32 & 0.16 \\ 0.16 & 3.35 \end{bmatrix}$	2.49
Food and Fuel			
$\begin{bmatrix} 0.35 & 0.05 \\ 0.47 & 0.52 \end{bmatrix}$	$\begin{bmatrix} 0.55 & -0.04 \\ -0.40 & 0.40 \end{bmatrix}$	$\begin{bmatrix} 1.42 & 0.87 \\ 0.87 & 12.90 \end{bmatrix}$	3.01
Indus and Fuel			
$\begin{bmatrix} 0.29 & 0.01 \\ -0.11 & 0.47 \end{bmatrix}$	$\begin{bmatrix} 0.63 & 0.03 \\ 0.09 & 0.48 \end{bmatrix}$	$\begin{bmatrix} 2.22 & 1.50 \\ 1.50 & 7.17 \end{bmatrix}$	1.67
Food, Indus and Fuel			
$\begin{bmatrix} 0.27 & 0.01 & 0.03 \\ 0.27 & 0.25 & 0.02 \\ 0.30 & -0.10 & 0.56 \end{bmatrix}$	$\begin{bmatrix} 0.64 & -0.02 & -0.02 \\ -0.16 & 0.66 & -0.01 \\ -0.27 & 0.12 & 0.37 \end{bmatrix}$	$\begin{bmatrix} 1.34 & 0.12 & 0.55 \\ 0.12 & 3.29 & 2.20 \\ 0.55 & 2.20 & 10.02 \end{bmatrix}$	2.28
B_1	B_{-1}	$10^3\Omega$	λ
Food and Indus			
$\begin{bmatrix} 0.25 & 0.01 \\ 0.22 & 0.24 \end{bmatrix}$	$\begin{bmatrix} 0.69 & -0.03 \\ -0.12 & 0.70 \end{bmatrix}$	$\begin{bmatrix} 0.27 & 0.03 \\ 0.03 & 0.46 \end{bmatrix}$	6.30
Food and Fuel			
$\begin{bmatrix} 0.25 & 0.02 \\ 0.16 & 0.38 \end{bmatrix}$	$\begin{bmatrix} 0.67 & -0.02 \\ -0.14 & 0.55 \end{bmatrix}$	$\begin{bmatrix} 0.25 & 0.06 \\ 0.06 & 1.21 \end{bmatrix}$	5.23
Indus and Fuel			
$\begin{bmatrix} 0.26 & 0.04 \\ -0.09 & 0.56 \end{bmatrix}$	$\begin{bmatrix} 0.67 & -0.01 \\ 0.09 & 0.37 \end{bmatrix}$	$\begin{bmatrix} 0.42 & 0.24 \\ 0.24 & 1.19 \end{bmatrix}$	4.77
Food, Indus and Fuel			
$\begin{bmatrix} 0.88 & -0.17 & -0.02 \\ -0.04 & 0.27 & 0.07 \\ -0.04 & 0.06 & 0.58 \end{bmatrix}$	$\begin{bmatrix} 0.21 & 0.15 & 0.00 \\ 0.13 & 0.76 & -0.08 \\ 0.02 & 0.05 & 0.33 \end{bmatrix}$	$\begin{bmatrix} 0.32 & 0.02 & 0.07 \\ 0.02 & 0.51 & 0.26 \\ 0.07 & 0.26 & 1.35 \end{bmatrix}$	6.15

The results are shown in Table 6.7. The LR column displays the LR test statistic and the IC columns are the difference in the IC values of the restricted and the unrestricted models.

Table 6.7: Common bubble detection on multivariate combinations of the variables

Levels						
Food	Indus	Fuel	Rank test	LR	BIC	AIC
■	■		1 vs 2	25.93	20.04	23.93
■		■	1 vs 2	59.96	54.07	57.96
	■	■	1 vs 2	70.49	64.59	68.49
■	■	■	2 vs 3	16.26	10.37	14.26
			1 vs 3	88.12	64.55	80.12
Logs						
Food	Indus	Fuel	Rank test	LR	BIC	AIC
■	■		1 vs 2	16.04	10.15	14.04
■		■	1 vs 2	34.36	28.47	32.36
	■	■	1 vs 2	46.05	40.16	44.05
■	■	■	2 vs 3	15.81	9.92	13.81
			1 vs 3	75.01	51.44	67.01

LR is the likelihood ratio test statistic. For the bivariate models the critical value of the LR test at 95% confidence level is 3.41. For the trivariate models, the critical values are 3.841 and 9.488 for *2 vs 3* and *1 vs 3* respectively. The column BIC and AIC show the difference between the restricted and unrestricted information criteria.

Looking at the LR tests, the null hypothesis of a common bubble in the bivariate and trivariate models is rejected for all combination of variables at a confidence level of 95%. All information criteria also indicate a better fit for the models without commonalities since all values are positive. Even for the trivariate cases *2*

vs 3, no bubble is detected even though in the simulations exercise, the test and information criteria over-detected a CB for such sample size and degrees of freedom. Hence, while the series seem to follow similar pattern in the locally explosive episodes throughout the time period, we do not find significant indication of commonalities in their forward looking components.

6.5 Conclusion

This chapter proposes methods to investigate whether the bubble patterns observed in individual series are common to various series. We detect the non-linear dynamics using the recent mixed causal and noncausal models. The lead component of the model allows to capture, for instance, locally explosive episodes in a parsimonious and strictly stationary setting. We hence employ multivariate mixed causal-noncausal models and apply restrictions to the lead coefficients matrices to test for the presence of commonalities in the forward looking components of the series. We propose a likelihood ratio (LR) test to test for the presence of a common bubble. In a simulation study, we investigate the accuracy of the common bubbles detection using the LR test as well as by model selection using information criteria. Then, implementing our approach on three commodity prices we do not find evidence of commonalities despite the similarities between the series. Our definition of common bubbles requires that all noncausal matrices span the same left null space. A natural extension to our approach would be to relax that hypothesis to investigate non synchronous common bubbles, allowing for some adjustment delays along the lines of Cubadda and Hecq (2001).

7

Conclusions and discussion

This thesis explores the modelling and forecasting of time series with mixed causal-noncausal autoregressive (MAR) models. Those models employ not only past values of the process but also future values. They are parsimonious and allow for non-linear dynamics within a strictly stationary setting. That is, they can capture for instance locally explosive episodes, which are often observed in commodity prices, inflation rates, stock prices or cryptocurrencies. While a process depending on its own future values might seem counter-intuitive, some variables do have anticipative characteristics and MAR models offer a large flexibility in modelling stationary series that exhibit nonlinear dynamics. The model is rather simple and proved to provide a better fit than purely causal models in many applications.

Coming back to the example given in the introduction of this thesis, a long lasting increase in various commodity prices will lead to a persistent increase in inflation. Hence, the more information policy makers, investors or applied researchers have regarding the potential downturn of the bubble, the more adequately they can act. From a

more financial perspective for instance, investing in a stock which price persistently increases offers great opportunity of profits but also considerable risks. Understanding the extent to which this bubble is likely to go on or suddenly drop can help better perceive the incurred risks. These examples illustrate the importance of not only being able to model such processes but also to forecast them accurately.

Chapter 2 to 4 focus on univariate MAR models with a unique lead and unrestricted amount of lags, namely MAR(r,1). Indeed, a unique lead is sufficient to capture non-linear patterns such as bubbles. Chapter 2 investigates the existing forecasting methods of MAR processes and shows that as a series deviates from central values, its predictive density splits and becomes bimodal, indicating either a further increase or a drop. The information provided by the bimodality of the distribution would not be captured by point forecasts of by purely backward-looking AR models.

MAR models can be employed in various areas of applications and for distinct purposes, as illustrated by Chapters 3 and 4. Chapter 3, after investigating a common issue that applied researchers face, that is the detrending of non-stationary series, employs MAR models to predict probabilities of crash in oil prices. It analyses the COVID-19 pandemic outbreak, a period in which prices were more volatile and predicts probabilities of crash of different magnitudes. Chapter 3 on the other hand proposes the use of the probabilities obtained from MAR models to construct a short-term credibility index of Central banks' inflation targeting system. While during periods of increasing inflation, the targeting system is logically not credible, the index can also be used during rather stable period when inflation is close to the target by providing short-term risks of exiting the tolerance bounds.

Probabilities are sometimes difficult to interpret. For instance, when forecasting the probabilities of a crash, from which threshold should

we consider it to be too risky? Chapter 4 suggests the use of receiver operating characteristic (ROC) curves to help determine the adequate threshold, which will depend on the aversion of false positives for a given inquiry. We illustrate their use on the credibility of the Central bank inflation targeting system, but it can naturally be employed in other analyses, such as the probabilities of crash mentioned above.

The thesis also investigates different settings of MAR models for which the literature is still at this date limited. Chapter 5 allows for more leads and pave the way for further research regarding theoretical findings of the predictive density. Various commodity prices in many applications in the literature are identified as MAR processes with more than one lead, emphasising the importance of having more theoretical understanding of general MAR models. The chapter also suggest a new method for forecasting MAR(r, s) processes which, based on the researcher's inquiry and the sample size of the data available, is a suitable alternative to the existing ones.

As illustrated by the oil prices in Chapter 3 or the three commodity indices in Chapter 6, it is common to observe bubble-like patterns at the same time on various time series. The detection of commonalities improves the parsimony of autoregressive multivariate models for which the number of coefficients increases exponentially with the number of variables and lags. Chapter 6 therefore investigates the detection of, what we define as common bubbles, commonalities in the forward looking component of MAR processes, in a multivariate setting. While we have not found yet an example of the detection of such commonalities in practice, we propose a likelihood ratio test and the use of information criteria, to detect the presence of common bubbles.

Overall, this thesis covers different frameworks and features of mixed causal-noncausal models ranging from the forecast of univariate pro-

cesses to the detection of commonalities in multivariate settings. The wide range of applications of MAR models demonstrate their versatility and their ability to capture non-linear dynamics that are often observed in economic time series. While the interest in MAR models is continually increasing in the literature, these models are still rather new and there are a lot of features and extensions that have not yet been explored. Besides the inquiries of Chapter 5 that could be further investigated, extensions of MAR models could for instance include more complex dynamic structures. Multivariate MAR models could also be used in the construction of portfolios. In line with Chapter 6, different structures of commonalities could be considered, to construct portfolios in which the bubble patterns observed in the individual stocks disappear, hedging against the risks that each of them carries.

Bibliography

Alquist, Ron and Lutz Kilian (2010). "What do we learn from the price of crude oil futures?" In: *Journal of Applied econometrics* 25.4, pp. 539–573.

Alquist, Ron, Lutz Kilian, and Robert J Vigfusson (2013). "Forecasting the price of oil". In: *Handbook of economic forecasting*. Vol. 2. Elsevier, pp. 427–507.

Andrews, Beth, Richard Davis, and Jay Breidt (2006). "Maximum likelihood estimation for all-pass time series models". In: *Journal of Multivariate Analysis* 97.7, pp. 1638–1659.

Backus, David K and Patrick J Kehoe (1992). "International evidence on the historical properties of business cycles". In: *The American Economic Review*, pp. 864–888.

Baumeister, Christiane and Lutz Kilian (2016). "Forty years of oil price fluctuations: Why the price of oil may still surprise us". In: *Journal of Economic Perspectives* 30.1, pp. 139–60.

Bec, Frédérique, Heino Bohn Nielsen, and Sarra Saïdi (2020a). "Mixed Causal–Noncausal Autoregressions: Bimodality Issues in Estimation and Unit Root Testing 1". In: *Oxford Bulletin of Economics and Statistics* 82.6, pp. 1413–1428.

— (2020b). "Mixed causal–noncausal autoregressions: Bimodality issues in estimation and unit root testing 1". In: *Oxford Bulletin of Economics and Statistics* 82.6, pp. 1413–1428.

Bertelsen, Kristoffer Pons (2019). "Comparing Tests for Identification of Bubbles". In: *Department of Economics and Business Economics, Aarhus University*.

Blinder, Alan S (2000). "Central-bank credibility: why do we care? How do we build it?" In: *American economic review* 90.5, pp. 1421–1431.

Bomfim, Antulio N and Glenn D Rudebusch (2000). "Opportunistic and deliberate disinflation under imperfect credibility". In: *Journal of Money, Credit and Banking*, pp. 707–721.

Breidt, F Jay et al. (1991). "Maximum likelihood estimation for noncausal autoregressive processes". In: *Journal of Multivariate Analysis* 36.2, pp. 175–198.

Brooks, Chris, Marcel Prokopczuk, and Yingying Wu (2015). "Booms and busts in commodity markets: bubbles or fundamentals?" In: *Journal of Futures Markets* 35.10, pp. 916–938.

Campbell, John Y and Robert J Shiller (1987). "Cointegration and tests of present value models". In: *Journal of political economy* 95.5, pp. 1062–1088.

Canova, Fabio (1998). "Detrending and business cycle facts". In: *Journal of monetary economics* 41.3, pp. 475–512.

Cavaliere, Giuseppe, Heino Bohn Nielsen, and Anders Rahbek (2018). "Bootstrapping noncausal autoregressions: with Applications to explosive bubble modeling". In: *Journal of Business and Economic Statistics*, pp. 1–13.

Cecchetti, Stephen G and Stefan Krause (2002). "Central bank structure, policy efficiency, and macroeconomic performance: exploring empirical relationships". In: *Review-Federal Reserve Bank of Saint Louis* 84.4, pp. 47–60.

Cubadda, Gianluca and Alain Hecq (2001). "On non-contemporaneous short-run co-movements". In: *Economics Letters* 73.3, pp. 389–397.

— (2022). "Reduced Rank Regression Models in Economics and Finance". In: *Oxford Research Encyclopedia of Economics and Finance*. DOI: 10.1093/acrefore/9780190625979.013.677.

Cubadda, Gianluca, Alain Hecq, and Sean Telg (2019). "Detecting Co-Movements in Non-Causal Time Series". In: *Oxford Bulletin of Economics and Statistics* 81.3, pp. 697–715.

Cubadda, Gianluca, Alain Hecq, and Elisa Voisin (2022). *Detecting common bubbles in multivariate mixed causal-noncausal models*. DOI: 10.48550/ARXIV.2207.11557.

Diba, Behzad T and Herschel I Grossman (1988). "Explosive rational bubbles in stock prices?" In: *The American Economic Review* 78.3, pp. 520–530.

Dovern, Jonas, Ulrich Fritsche, and Jiri Slacalek (2012). "Disagreement among forecasters in G7 countries". In: *Review of Economics and Statistics* 94.4, pp. 1081–1096.

Engle, Robert F and Clive WJ Granger (1987). "Co-integration and error correction: representation, estimation, and testing". In: *Econometrica: journal of the Econometric Society*, pp. 251–276.

Engle, Robert F and Svend Hylleberg (1996). "Common seasonal features: Global unemployment". In: *Oxford Bulletin of Economics and Statistics* 58.4, pp. 615–630.

Engle, Robert F and Sharon Kozicki (1993). "Testing for common features". In: *Journal of Business and Economic Statistics* 11.4, pp. 369–380.

Engle, Robert F and Raul Susmel (1993). "Common volatility in international equity markets". In: *Journal of Business and Economic Statistics* 11.2, pp. 167–176.

Fries, Sébastien (2018). "Conditional Moments of Anticipative Alpha-Stable processes and the prediction of bubble crash odds". In: *arXiv preprint arXiv:1805.05397*.

— (2021). "Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds". In: *Journal of Business and Economic Statistics*.

Fries, Sébastien and Jean-Michel Zakoïan (2019a). "Mixed causal-noncausal AR processes and the modelling of explosive bubbles". In: *Econometric Theory* 35.6, pp. 1234–1270.

— (2019b). "Mixed causal-noncausal AR processes and the modelling of explosive bubbles". In: *Econometric Theory* 35.6, pp. 1234–1270.

Gali, Jordi, Mark Gertler, and J David Lopez-Salido (2005). "Robustness of the estimates of the hybrid New Keynesian Phillips curve". In: *Journal of Monetary Economics* 52.6, pp. 1107–1118.

Giancaterini, Francesco and Alain Hecq (2022). "Inference in mixed causal and noncausal models with generalized Student's t-distributions". In: *Econometrics and Statistics*.

Giancaterini, Francesco, Alain Hecq, and Claudio Morana (2022). "Is climate change time reversible?" In: *arXiv preprint arXiv:2205.07579*.

Gourieroux, C, J Jasiak, and A Monfort (2020). "Stationary bubble equilibria in rational expectation models". In: *Journal of Econometrics* 218.2, pp. 714–735.

Gourieroux, Christian, Andrew Hencic, and Joann Jasiak (2021). "Forecast performance and bubble analysis in noncausal MAR(1, 1) processes". In: *Journal of Forecasting* 40.2, pp. 301–326.

Gourieroux, Christian and Joann Jasiak (2017). "Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation". In: *Journal of Econometrics* 200.1, pp. 118–134.

Gourieroux, Christian, Joann Jasiak, and Michelle Tong (2021). "Convolution-based filtering and forecasting: An application to WTI crude oil prices". In: *Journal of Forecasting* 40.7, pp. 1230–1244.

Gouriéroux, Christian, Andrew Hencic, and Joann Jasiak (2018). "Forecast Performance in Noncausal MAR(1, 1) Processes". In.

Gouriéroux, Christian and Joann Jasiak (2016). "Filtering, prediction and simulation methods for noncausal processes". In: *Journal of Time Series Analysis* 37.3, pp. 405–430.

— (2018). "Misspecification of noncausal order in autoregressive processes". In: *Journal of Econometrics* 205.1, pp. 226–248.

Gouriéroux, Christian, Joann Jasiak, and Alain Monfort (2016). "Stationary Bubble Equilibria in Rational Expectation Models". In: *CREST Working Paper. Paris, France: Centre de Recherche en Economie et Statistique*.

Gouriéroux, Christian and Jean-Michel Zakoïan (2013). "Explosive Bubble Modelling by Noncausal Process". In: *CREST. Paris, France: Centre de Recherche en Economie et Statistique*.

— (2015). "On Uniqueness of Moving Average Representations of Heavy-tailed Stationary Processes". In: *Journal of Time Series Analysis* 36.6, pp. 876–887.

— (2017). "Local explosion modelling by non-causal process". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.3, pp. 737–756.

Hamilton, James D (2018). "Why you should never use the Hodrick-Prescott filter". In: *Review of Economics and Statistics* 100.5, pp. 831–843.

Hecq, Alain, João Victor Issler, and Sean Telg (2020). "Mixed causal–noncausal autoregressions with exogenous regressors". In: *Journal of Applied Econometrics* 35.3, pp. 328–343.

Hecq, Alain, João Victor Issler, and Elisa Voisin (2022). *A short term credibility index for central banks under inflation targeting: an application to Brazil*. DOI: 10.48550/ARXIV.2205.00924.

Hecq, Alain, Lenard Lieb, and Sean Telg (2016). "Identification of Mixed Causal-Noncausal Models in Finite Samples". In: *Annals of Economics and Statistics/Annales d'Économie et de Statistique* 123/124, pp. 307–331.

— (2017a). "Simulation, Estimation and Selection of Mixed Causal-Noncausal Autoregressive Models: The MARX Package". In.

— (2017b). "Simulation, Estimation and Selection of Mixed Causal-Noncausal Autoregressive Models: The MARX Package". In: *Available at SSRN 3015797*.

Hecq, Alain and Li Sun (2021). "Identification of noncausal models by quantile autoregressions". In: *Studies in Nonlinear Dynamics and Econometrics*.

Hecq, Alain, Sean Telg, and Lenard Lieb (2017). "Do seasonal adjustments induce noncausal dynamics in inflation rates?" In: *Econometrics* 5.4, p. 48.

Hecq, Alain and Elisa Voisin (2021). "Forecasting bubbles with mixed causal-noncausal autoregressive models". In: *Econometrics and Statistics* 20, pp. 29–45.

— (2022). "Predicting bubble bursts in oil prices during the COVID-19 pandemic with mixed causal-noncausal models". In: *Advances in Econometrics in honor of Joon Y. Park*. Forthcoming.

Hencic, Andrew and Christian Gouriéroux (2015). "Noncausal autoregressive model in application to bitcoin/USD exchange rates". In: *Econometrics of risk*. Springer, pp. 17–40.

Hendry, David F and Michael Massmann (2007). "Co-breaking: Recent advances and a synopsis of the literature". In: *Journal of Business and Economic Statistics* 25.1, pp. 33–51.

Hodrick, Robert J and Edward C Prescott (1997). "Postwar US business cycles: an empirical investigation". In: *Journal of Money, credit, and Banking*, pp. 1–16.

Homm, Ulrich and Jörg Breitung (2012). "Testing for speculative bubbles in stock markets: a comparison of alternative methods". In: *Journal of Financial Econometrics* 10.1, pp. 198–231.

Issler, João Victor and Ana Flávia Soares (2022). "Central Bank credibility and inflation expectations: a microfounded forecasting approach". In: *Macroeconomic Dynamics* In Print. DOI: 10.1017/S1365100522000207.

Issler, João Victor and Farshid Vahid (2001). "Common cycles and the importance of transitory shocks to macroeconomic aggregates". In: *Journal of Monetary Economics* 47.3, pp. 449–475.

Karapanagiotidis, Paul (2014). "Dynamic modeling of commodity futures prices". In: *MPRA Paper 56805, University Library of Munich, Germany.*

Kilian, Lutz (2009). "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market". In: *American Economic Review* 99.3, pp. 1053–69.

Kilian, Lutz and Daniel P Murphy (2014). "The role of inventories and speculative trading in the global market for crude oil". In: *Journal of Applied econometrics* 29.3, pp. 454–478.

Kilian, Lutz and Xiaoqing Zhou (2020a). "Does drawing down the US Strategic Petroleum Reserve help stabilize oil prices?" In: *Journal of Applied Econometrics* 35.6, pp. 673–691.

— (2020b). "The Econometrics of Oil Market VAR Models". In: *CESifo Working Paper Series* 8153.

Lanne, Markku and Jani Luoto (2013). "Autoregression-based estimation of the new Keynesian Phillips curve". In: *Journal of Economic Dynamics and Control* 37.3, pp. 561–570.

Lanne, Markku, Jani Luoto, and Pentti Saikkonen (2012). "Optimal forecasting of noncausal autoregressive time series". In: *International Journal of Forecasting* 28.3, pp. 623–631.

Lanne, Markku, Henri Nyberg, Erkka Saarinen, et al. (2012). "Does noncausality help in forecasting economic time series?" In: *Economics Bulletin*.

Lanne, Markku and Pentti Saikkonen (2011). "Noncausal autoregressions for economic time series". In: *Journal of Time Series Econometrics* 3.3.

— (2013). "Noncausal vector autoregression". In: *Econometric Theory* 29.3, pp. 447–481.

Levieuge, Grégory, Yannick Lucotte, and Sébastien Ringuedé (2018). "Central bank credibility and the expectations channel: Evidence based on a new credibility index". In: *Review of World Economics* 154.3, pp. 493–535.

Lof, Matthijs and Henri Nyberg (2017). "Noncausality and the commodity currency hypothesis". In: *Energy Economics* 65, pp. 424–433.

Mendonça, Helder Ferreira de and Gustavo JG Souza (2007). "Credibilidade do regime de metas para inflação no Brasil". In: *Pesquisa e Planejamento Econômico* 37.2, pp. 247–282.

Mendonça, Helder Ferreira de and Gustavo José de Guimarães Souza (2009). "Inflation targeting credibility and reputation: the consequences for the interest rate". In: *Economic Modelling* 26.6, pp. 1228–1238.

Nason, James M and Gregor W Smith (2008). "Identifying the new Keynesian Phillips curve". In: *Journal of Applied Econometrics* 23.5, pp. 525–551.

Phillips, Peter CB and Zhentao Shi (2019). "Boosting the hodrick-prescott filter". In.

Phillips, Peter CB, Yangru Wu, and Jun Yu (2011). "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?" In: *International economic review* 52.1, pp. 201–226.

Pindyck, Robert S (1993). "The present value model of rational commodity pricing". In: *The Economic Journal* 103, pp. 511–530.

Ravn, Morten O and Harald Uhlig (2002). "On adjusting the Hodrick-Prescott filter for the frequency of observations". In: *Review of economics and statistics* 84.2, pp. 371–376.

Svensson, Lars EO (2000). "How should monetary policy be conducted in an era of price stability?" In: *National bureau of economic research Cambridge, Mass., USA*.

Vahid, Farshid and Robert F Engle (1993). "Common trends and common cycles". In: *Journal of Applied Econometrics*, pp. 341–360.

Impact paragraph

This thesis focuses on one particular type of non-linear dynamics: bubbles. They are defined as episodes of persistent increase followed by a sudden crash. Bubbles are commonly observed in many economic and financial data. Naturally, they come in different magnitudes, some have mild impact while others have dramatic global consequences. Among the most notorious bubbles that have occurred, there are for instance the tulip mania of the 1630s in the Netherlands, the Japanese asset price bubble in the late 1980s, the dot-com bubble of the 1990s or the U.S. housing bubble of the 2000s. The consequential repercussions that bubbles can have accentuate the importance of being able to model and predict such episodes. Bubbles occur regularly and we nowadays observe what seem to be the boom phases of bubbles in various real estate markets across the world for instance as well as surges in inflation rates all around the world. There is therefore a need for simple models that can capture such complex patterns. Indeed, the simpler the model, the easier it is to estimate and the less requirements and uncertainty are added to the forecasts.

Mixed causal-noncausal autoregressive (MAR) models, that are employed in this thesis, are simple models that can capture non-linear dynamics. Within the framework of MAR models, a variable can be explained by its own past and future values. In their simplest form, namely with a unique regressor being the future value of the variable, the model is already able to capture bubbles. Nonetheless, these models are still rather novel and there is thus room for further research and developments which could help understand the complex dynamics of various economic and financial time series.

In the second chapter of this thesis, MAR models are used to predict the probabilities of crash along a bubble in Nickel prices during the extreme episode of the 2007 crisis. Density forecasts, from which we

can derive much more information than from point forecasts, allows to foresee the potential magnitude of an upcoming crash. We find that at the top of the bubble in 2007, although predicted probabilities indicated a large probabilities of a crash, it was suggested that the initial crash would not be larger than 40%, which turned out to be correct. While we present only one empirical example in this chapter, many other commodity prices are characterised by the same dynamics, and the crisis of 2007 generated a bubble in almost all commodity prices.

In the third chapter, oil prices are forecasted during the COVID-19 pandemic outbreak. Although we cannot predict the end of COVID-19, we are able to capture valuable information carried by the past and current dynamics in the data. For instance, we find that after the initial impact of the implementations of the numerous worldwide lockdown, the probabilities of a further decrease in prices were at that point smaller than before, yet that the magnitude of the decrease would be larger. Nowadays, the economic situation with the war in Ukraine, triggering a surge in oil prices and many other commodities stresses the importance of building and developing models that can capture the dynamics of these processes. Indeed, commodity prices are at the core of inflation rates, which are currently surging all around the world and while wars are difficultly predictable, policy makers could employ the information derived from MAR models.

In the fourth chapter of this thesis, we investigate the inflation rate of Brazil, with a focus on a rather stable period. Brazil has undergone periods of hyperinflation in the past and the credibility of its Central Bank is therefore of the utmost importance for the population as persistent deviations from the target could trigger a new hyperinflation episode. We employ the probability forecasts to construct a short-term credibility index of the central bank over time, consisting of the probability that the inflation rate meets the target announced by the central bank. We find that the short-term index, which can be employed as an early warning of exiting the target bounds, is complementary to the long-term indices that are based on people's beliefs over longer period of times. Naturally, this index can be constructed for any country

which central bank applies an inflation targeting approach.

In the fifth chapter we consider metals price index, which presents a different dynamic structure than nickel or oil prices mentioned above as it is identified as an MAR(0,2), namely including 2 leads of the variable. This chapter shows that MAR models allows to capture more complex dynamics with the inclusion of more leads. In this chapter we propose to combine existing forecasting approaches to limit the drawbacks of these methods in certain conditions. We offer this as an alternative, especially when the sample size is too small, which is often the case with quarterly or yearly data sets that are often employed in macroeconomic analyses for instance.

In the sixth chapter of the thesis we investigate commonalities in the forward-looking component (defined as common bubbles) within groups of variables, which paves the way to further research in that direction. Detecting common bubbles could be used to significantly decrease the risks incurred by each of the individual component of the group and drastically limit the consequences of bubble bursts within investment strategies for instance.

The versatility of MAR models implies that they can be of relevance for a wide range of practitioners, from policy makers to investors. Overall, MAR models offer an adequate and easy-to-implement alternative to complex models in the presence of various types of non-linear dynamics in a time series. The current economic situation stresses the importance of building and developing models that can help predict and anticipate potential upcoming bubbles and bursts. This therefore asserts the relevance of analysing, both theoretically and empirically, MAR models.

About the author

Elisa Voisin was born on May 18, 1995, in Paris, France. She graduated high school in 2013, after following the scientific track. She then enrolled at Maastricht University and obtained a B.Sc. in Econometrics and Operations Research in 2016. In the course of her bachelor she spent a semester at Gothenburg University in Sweden. She subsequently pursued a Master in Economic and Financial Research at Maastricht University with a specialisation in Econometrics, finance and monetary economics. During that time she also worked as a research assistant. She obtained her M.Sc. with cum laude in 2018.

In September 2018, she joined the department of Quantitative Economics at Maastricht University as a Ph.D. candidate under the supervision of prof. dr. Alain Hecq. Later on, dr. Ines Wilms joined the supervision. Between 2018 and 2022, Elisa presented her work at international conferences and seminars. She went for three months for a research visit to Tor Vergata, University of Rome in late 2021, to work with prof. dr. Gianluca Cubadda. The results of her research are presented in this thesis, which consists of adaptations of her published articles and working papers.