Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2014-15

Filtering and Prediction in Noncausal processes

C. GOURIÉROUX[1]
J. JASIAK[2]

April, 2014

1 CREST and University of Toronto, Canada. Email : christian.gourieroux@ensae.fr
2 York University, Canada. Email : jasiakj@yorku.ca

Filtering and Prediction in Noncausal Processes

Christian Gourieroux *
Joann Jasiak †

this version: April, 2014

This paper revisits the filtering and prediction in noncausal and mixed autoregressive processes and provides a simple alternative set of methods that are valid for processes with infinite variances. The prediction method provides complete predictive densities and prediction intervals at any finite horizon H, for univariate and multivariate processes. It is based on an unobserved component representation of noncausal processes. The filtering procedure for the unobserved components is provided along with a a simple back-forecasting estimator for the parameters of noncausal and mixed models and a simulation algorithm for noncausal and mixed autoregressive processes. The approach is illustrated by simulations.

Keywords : Noncausal Process, Nonlinear Prediction, Filtering, Look-Ahead Estimator, Speculative Bubble, Technical Analysis.

JEL number: C14, G32, G23.

*University of Toronto and CREST, *e-mail*: gouriero@ensae.fr.
†York University, Canada, *e-mail*: jasiakj@yorku.ca.
The authors gratefully acknowledge financial support of the chair ACPR:Regulation and Systemic Risks, and of the Global Risk Institute.

1 Introduction

In recent literature, there has been a growing number of applications involving economic time series models with causal and noncausal components. The time series modelled as noncausal processes range from the macroeconomic data [Lanne, Saikkonen (2011), Davis, Song (2012), Chen Choi, Escanciano (2012)], to the Standard and Poor (S & P) index prices [Gourieroux, Zakoian (2013)], the commodity prices and electronic currency exchange rates [Gourieroux, Hencic (2013)]. The empirical results reported in the literature suggest that the traditional Box-Jenkins methodology that restricts the temporal dependence in linear autoregressive processes to the past only, has been often found insufficient.

While the empirical literature on noncausal and mixed processes is rather recent, the theoretical results have been long established since the seventies and the eighties [see e.g. Davis, Resnick (1986), or Rosenblatt (2000) for a general presentation of such processes]. Despite that, over the last four decades, applied research remained strongly influenced by the Box-Jenkins methodology, and the normality assumption that underlies the quasi maximum likelihood estimation method was used as an integral part of that methodology. Because the forward- and backward-looking dynamics of a Gaussian time series are not distinguishable, the forward-looking components were disregarded and the past conditioning in ARMA models became a standard practice. Formally, the problem of non-identifiability of the backward and forward-looking autoregressive polynomials in Gaussian time series was formulated by Rosenblatt [see Rosenblatt (2000) [Th 1.3.1]]. Therefore, for identification purposes, the assumption of normality in models with causal and noncausal components has to be relaxed.

The noncausal and mixed processes possess an infinite moving average representation:

$$y_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}, \tag{1.1}$$

where (ε_t) is a sequence of i.i.d. variables and $a_i, i = -\infty, \ldots, +\infty$ a two-sided sequence of moving average coefficients [see e.g. Brockwell, Davis (1991)]. Process y_t exists almost surely under rather weak conditions, such as $E(|\varepsilon_t|^\delta) < \infty$, for $\delta > 0$ and $\sum_{i=-\infty}^{+\infty} |a_i|^\delta < \infty$. The above $MA(\infty)$ represen-

tation differs from the moving average processes considered in the standard Box, Jenkins approach in the following aspects:

i) The moving average is two-sided, including a "causal" component $\sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$, function of the current and lagged values of the error, as well as a "noncausal" component $\sum_{i=-\infty}^{-1} a_i \varepsilon_{t-i}$.

ii) The error term is a strong white noise, instead of a weak white noise.

iii) The distribution of the error term can have fat tails. In particular, ε_t can have infinite variance, or even infinite expectation.

The existing methods of prediction for noncausal processes focus on point prediction of processes with finite means. The predictions are based on conditional expectations, which have no closed-form expressions and need to be approximated by simulations [see Lanne, Luoto, Saikkonen (2012) for univariate processes and Lanne, Saikkonen (2013) for some multivariate processes].

The objective of this paper is to provide a simpler prediction method that yields complete predictive densities and prediction intervals at multiple finite horizons in one step. The proposed method requires less simulations than the methods focused on the optimal (point) predictor approximation and is computationally less demanding. Especially, it is valid for non-Gaussian processes with heavy tails, including processes with infinite error variances. The prediction method for univariate processes is extended to multivariate processes in order to obtain a comprehensive approach to the analysis of noncausal processes.

Our approach exploits the unobserved component representation of noncausal and mixed processes that distinguished the latent causal and noncausal components of y_t. This decomposition has been introduced by Lanne, Saikkonen (2011) for univariate processes. We extend it to the multivariate framework. The filtering procedure for the unobserved components of noncausal and mixed processes is provided in the paper along with a simulation algorithm. In addition, the filtering algorithm itself is used in a consistent and asymptotically efficient parameter estimation method: The back-forecasting BHHH algorithm is introduced as a time-efficient alternative to the approximated maximum likelihood method.

The paper is organized as follows. Section 2 describes the filtering method and introduces the unobserved "causal" and "noncausal" component repre-

sentation of y_t. It establishes a deterministic dynamic relationship between the unobserved components, the process y_t and the errors and highlights the equivalence of various filtrations.

The nonlinear prediction in univariate processes is discussed in Section 3. We first derive a closed-form formula of the consistent estimator of the predictive density of $(y_{T+1}, \ldots, y_{T+H})$ given the information $(y_1, \ldots, y_T)$. This estimator is based on an extension of the look-ahead density estimator, introduced in Glynn, Hendersen (1998), (2001) which circumvents the use of simulations. This eliminates the time consuming simulations involved in the prediction method proposed by Lanne, Luoto, Saikkonen (2012). Next, drawings in the predictive density are performed by using a Sampling Importance Resampling (SIR) method [Gelfand, Smith (1990)a,b], which serve to generate future paths of y and prediction intervals.

In Section 4, we show as an additional outcome, the filtered component-based back-forecasting algorithm that leads to consistent and asymptotically efficient estimation of the parameters of the autoregressive causal and non-causal polynomials and error distribution.

Section 5 extends the filtering and prediction methods to the multivariate framework. A simulation study is presented in Section 6 to analyse the properties of the filtering and prediction methods introduced in Sections 3 and 4. Section 7 concludes.

Some proofs and a description of the SIR method are gathered in appendices.

2 The process and its unobserved components

2.1 Definition of unobserved components

Let us consider the autoregressive process defined as:

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t, \tag{2.1}$$

where the error terms are independent, identically distributed, such that $E(|\varepsilon_t|^\delta) < \infty$, for $\delta > 0, \Phi$ and Ψ are two polynomials of degrees r and s, respectively, with roots strictly outside the unit circle and such that $\Phi(0) = \Psi(0) = 1$. The polynomials Φ and Ψ can be inverted and the process can be rewritten as :

$$y_t = \frac{1}{\Phi(L)} \frac{1}{\Psi(L^{-1})} \varepsilon_t, \tag{2.2}$$

where $1/\Phi(L)$ [resp.$1/\Psi(L^{-1})$] is an infinite series in L (resp. L^{-1}), and the equality in (2.2) holds almost surely [see e.g. Brockwell, Davis (1991), Prop. 13.3.1]. Hence, y_t admits an infinite two sided moving average representation, which is the unique strictly stationary solution of recursive equation (2.1).

Following Lanne, Saikkonen (2011), and Lanne, Luoto, Saikkonen (2012), we consider the following unobserved "causal" and "noncausal" components of process y_t:

$$u_t \equiv \Phi(L) y_t \leftrightarrow \Psi(L^{-1}) u_t = \varepsilon_t, \tag{2.3}$$

and

$$v_t \equiv \Psi(L^{-1}) y_t \leftrightarrow \Phi(L) v_t = \varepsilon_t. \tag{2.4}$$

Let us now consider the filtrations generated by the error term $\underline{\varepsilon_t} = \{\varepsilon_\tau, \tau \leq t\}$, and the filtrations generated by the observations $\underline{y_t} = \{y_\tau, \tau \leq t\}$, respectively. A process is said to be causal with respect to a given filtration if its current value belongs to the associated information set. Conversely, we can reverse the time so that a process is said to be noncausal if its current value depends on a future value of the process that generated the filtration. Equations (2.3)-(2.4) lead to the following proposition :

Proposition 1 : i) u_t is ε-noncausal and y-causal.
ii) v_t is ε-causal and y-noncausal.

Process u (resp. v) is called the noncausal (resp. causal) component with respect to the filtration associated with the error ε.

2.2 Filtering

Suppose that we observe y over a period of length T and denote by $(y_1, \ldots, y_T)$ the observed sequence. The values of unobserved components u and v and errors ε can be computed from a set of observations $(y_1, \ldots, y_T)$ as follows. (i) From equation (2.1) for $t = r, \ldots, T-s$, we obtain the values $\varepsilon_{r+1}, \ldots, \varepsilon_{T-s}$ as functions of $(y_1, \ldots, y_T)$.

(ii) From equation (2.3) : $u_t = \Phi(L)y_t, t = r+1, \ldots, T$, we obtain $u_{r+1}, \ldots, u_T$.

(iii) From equation (2.4) : $v_t = \Psi(L^{-1})y_t, t = 1, \ldots, T-s$, we obtain $v_1, \ldots, v_{T-s}$.

When an additional observation y_{T+1} becomes available, the set of unobserved components can be updated by computing $\varepsilon_{T-s+1}, u_{T+1}$ and v_{T-s+1}.

2.3 Recovering process y from unobserved components

Conversely, the observable process y can be recovered from the ε-causal and ε-noncausal components. Below, we show two methods of recovering y, which are based on partial fraction decompositions.

i) $(u, v)-$ **causal representation of** y

We have :

$$\frac{1}{\Phi(L)\Psi(L^{-1})} = \frac{L^s}{\Phi(L)[L^s\Psi(L^{-1})]},$$

where the denominator is a polynomial in L. That polynomial can be rewritten as:

$$\frac{1}{\Phi(L)[L^s\Psi(L^{-1})]} \equiv \frac{b_1(L)}{\Phi(L)} + \frac{b_2(L)}{L^s\Psi(L^{-1})},$$

where the degree of polynomial b_1 is $d^0 b_1 \leq r-1$, and the degree of polynomial b_2 is $d^0 b_2 \leq s-1$.

It follows that:

$$\begin{aligned}\frac{1}{\Phi(L)\Psi(L^{-1})} &= L^s\left[\frac{b_1(L)}{\Phi(L)} + \frac{b_2(L)}{L^s\Psi(L^{-1})}\right] \\ &= \frac{L^s b_1(L)}{\Phi(L)} + \frac{b_2(L)}{\Psi(L^{-1})},\end{aligned}$$

and

$$y_t = \left[\frac{L^s b_1(L)}{\Phi(L)} + \frac{b_2(L)}{\Psi(L^{-1})}\right]\varepsilon_t.$$

Next, we use expressions (2.3)-(2.4) that define u and v and get:

$$y_t = L^s b_1(L)v_t + b_2(L)u_t. \tag{2.5}$$

Equation (2.5) provides a representation of y_t as a linear function of current and lagged $\varepsilon-$ causal and $\varepsilon-$ noncausal components v and u, respectively.

ii) $(u,v)-$**noncausal representation of** y

Alternatively, we can use the partial fraction decomposition with polynomials in L^{-1}. We get :

$$\begin{aligned} y_t &= \frac{L^{-r}}{[L^{-r}\Phi(L)]\Psi(L^{-1})}\varepsilon_t \\ &= L^{-r}\left[\frac{b_1^*(L^{-1})}{L^{-r}\Phi(L)} + \frac{b_2^*(L^{-1})}{\Psi(L^{-1})}\right]\varepsilon_t, \text{ with } d^0 b_1^* \leq r-1, d^0 b_2^* \leq s-1, \\ &= \frac{b_1^*(L^{-1})}{\Phi(L)}\varepsilon_t + \frac{L^{-r}b_2^*(L^{-1})}{\Psi(L^{-1})}\varepsilon_t. \end{aligned}$$

where d^0 denotes the degree of a polynomial. Next, using the definitions of u and v in (2.3)-(2.4), we get:

$$y_t = b_1^*(L^{-1})v_t + L^{-r}b_2^*(L^{-1})u_t, \tag{2.6}$$

which is a representation of process y as a linear function of its current and future $\varepsilon-$causal and $\varepsilon-$ noncausal components v and u, respectively.

Below, we illustrate the two representations for a mixed causal-noncausal autoregressive process of order (1,1), called MAR(1,1):.

Example 1 : Let us assume $r = s = 1$ and consider the mixed autoregressive model MAR(1,1):

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

with $|\varphi| < 1$, $|\psi| < 1$, or

$$y_t = \frac{L}{(1-\varphi L)(L-\psi)}\varepsilon_t.$$

The equality $\frac{1}{(1-\varphi L)(L-\psi)} = \frac{1}{1-\varphi\psi}\left(\frac{\varphi}{1-\varphi L} - \frac{1}{L-\psi}\right)$, implies:

$$\begin{aligned} y_t &= \frac{1}{1-\varphi\psi}\left(\frac{\varphi L}{1-\varphi L} + \frac{1}{1-\psi L^{-1}}\right)\varepsilon_t \\ &= \frac{1}{1-\varphi\psi}(\varphi v_{t-1} + u_t). \end{aligned} \tag{2.7}$$

In this representation, y_t is a linear function of the first lag of the ε-causal component v and of the current value of the ε-noncausal component u. We also have:

$$y_t = \frac{1}{1-\varphi\psi}(v_t + \psi u_{t+1}). \tag{2.8}$$

In the above representation, y_t is a linear function of the current value of the ε-causal component v and of the first lag of the ε-noncausal component u.

2.4 Equivalence of information sets

The following proposition establishes the equivalence of various information sets that contain the unobserved components u and v and errors ε [see Lanne, Luoto, Saikkonen (2012), or Appendix 1 for a shorter proof].

Proposition 2 : The following information sets are equivalent

i) $(y_1, \ldots, y_T)$;

ii) $(y_1, \ldots, y_r, u_{r+1}, \ldots, u_T)$;

iii) $(v_1 \ldots v_{T-s}, y_{T-s+1}, \ldots, y_T)$;

iv) $(y_1, \ldots, y_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$;

v) $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, y_{T-s+1}, \ldots, y_T)$;

vi) $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$.

The equivalence between $(y_1, \ldots, y_T)$ and $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{t-1}, u_{T-s+1}, ..., u_T)$ is of special importance, as the following three sets of variables $(v_1, \ldots, v_r)$, $(\varepsilon_{r+1}, \ldots, \varepsilon_{T-s})$, and $(u_{T-s+1}, \ldots u_T)$ are independent. Intuitively $(v_1, \ldots, v_r)$ [resp. $(u_{T-s+1}, \ldots, u_T)$] are the initial (resp. terminal) conditions that determine the path of process y over the period $\{1, \ldots, T\}$.

2.5 Simulation of a mixed causal-noncausal process

The results of the previous sections can be used for simulation of stationary causal-noncausal process. Let us outline below the steps for simulating a mixed autoregressive process of orders $r = s = 1$ MAR(1,1) defined as:

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

The simulation steps are as follows:

step 1: Simulate a long path of i.i.d. errors ϵ_t^s.

step 2: Use formulas (2.3)-(2.4) to simulate the paths of the ε-causal and ε-noncausal components :

$$u_t^s = \epsilon_t^s + \psi u_{t+1}^s,\ t = 1, ..., 2T,$$

$$v_t^s = \epsilon_t^s + \phi v_{t-1}^s,\ t = -T, ..., T,$$

starting from a far terminal condition [resp. far initial condition] $u_{2T}^s = u_0$, say [resp. $v_{-T}^s = v_0$].

step 3: Obtain the simulated values of process y from the representation given in (2.5) [or (2.6)]:

$$y_t^s = \frac{1}{1 - \phi\psi}(u_t^s + \phi v_{t-1}^s),\ t = 1, ..., T.$$

3 Prediction

Let us now consider the (nonlinear) prediction of future values of process y. We proceed in three steps as follows. First, we show that the prediction of future values of y is equivalent to the prediction of the future values of the ε-noncausal component u. Next, we introduce an estimator of the joint density of s consecutive future values of u given the past. In order to generate y_{T+1} or future paths $y_{T+1}, ..., y_{T+H}$, we use a Sampling Importance Resampling (SIR) method to draw the future values of the noncausal component from the estimated predictive density and then deduce the simulated y by applying formula (2.3). The proposed method can be applied in one step to obtain the forecast up to a horizon $H > 1$.

More specifically, the prediction of interest is the future path $(y_{T+1}, \ldots, y_{T+H})$, given the observations $(y_1, \ldots, y_T)$. It is accomplished by deriving (or estimating) the conditional p.d.f. $l(y_{T+1}, \ldots, y_{T+H} \mid y_1, \ldots, y_T)$, and next by drawing in that density function in order to obtain the conditional quantiles and prediction intervals.

This approach differs from the traditional point prediction, which will not be followed for the following reasons:

i) For second-order stationary processes, the optimal predictor is the conditional expectation $E(Y_{T+h} \mid y_1, \ldots, y_T), h = 1, \ldots, H$. However, this conditional expectation may not exist in non-Gaussian processes if the errors (and therefore the $Y_t's$) have fat tails. [1]

ii) For noncausal and non-Gaussian processes with finite variances, the formula of the optimal predictor is complicated, in general. While, for Gaussian processes, the conditional expectation $E(Y_{T+h}|y_1, \ldots, y_T), h = 1, \ldots, H$ is a linear function of the observed values, when the error term is not Gaussian, the best predictor $E(Y_{T+h}|y_1, \ldots, y_T)$ is a nonlinear function of the observed values [see e.g. Rosenblatt (2000), Theorem 5.3.1].[2]

Therefore, it is difficult to define and study point prediction in noncausal models that are non-Gaussian and include processes with fat tails. Below, we introduce a prediction method that provides a complete predictive density and not only at a location parameter of that density.

3.1 Equivalence of predictions

In Section 2.4, we showed that the information set $(y_1, \ldots, y_T)$ is equivalent to the information set $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$.

Therefore, the information contained in $(y_1, \ldots, y_{T+H})$ is equivalent to the information in $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T+H-s}, u_{T+H-s+1}, \ldots, u_{T+H})$, and it

[1] For causal autoregressive processes, whose innovation ε_t have stable symmetric distributions with fat tails, Stuck (1977) proposed a linear predictor $\hat{Y}_{T+h}$ defined by minimizing $E(|Y_{T+h} - \hat{Y}_{T+h}|^\alpha)$, where α is the tail index of the stable distribution. These predictions could be computed recursively from an extended Kalman filter. The nonlinear predictors of causal processes computed from minimizing the $\alpha-$ distance measure turned out to be accurate in a non Gaussian framework.

[2] When the error term is integrable with infinite variance, the prediction is linear in the past value only for errors with symmetric stable distributions [see Cambanis, Fakhre-Zakeri (1994)].

is also equivalent to that in $(v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, u_{T-s+1}, \ldots, u_{T+H})$, because $\Psi(L^{-1})u_t = \varepsilon_t, t = T-s+1, \ldots, T+H-s$ by formula (2.3).

Thus, instead of predicting the future value of y, we can equivalently predict the future value of the ε-noncausal component u, by finding the conditional p.d.f :

$$\begin{aligned} & l(u_{T+1}, \ldots, u_{T+H} | y_1, ..., y_T) \\ &= l(u_{T+1}, \ldots, u_{T+H} | v_1, \ldots, v_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, u_{T-s+1}, \ldots, u_T) \\ &= l(u_{T+1}, \ldots, u_{T+H} | u_{T-s+1}, \ldots, u_T), \end{aligned} \tag{3.1}$$

given that $(u_{T-s+1}, \ldots, u_{T+H})$ are independent of the unobserved components $v_1, \ldots, v_r$ and errors $\varepsilon_{r+1}, \ldots, \varepsilon_{T-s}$ in the information set.

The conditional p.d.f. in (3.1) can also be written as :

$$\begin{aligned} & l(u_{T+1}, \ldots, u_{T+H} | u_{T-s+1}, \ldots, u_T) \\ &= \frac{l(u_{T-s+1}, \ldots, u_T, u_{T+1}, \ldots, u_{T+H})}{l_s(u_{T-s+1}, \ldots, u_T)} \\ &= \frac{l(u_{T-s+1}, \ldots, u_{T+H-s} | u_{T+H-s+1}, \ldots, u_{T+H})}{l_s(u_{T-s+1}, \ldots, u_T)} l_s(u_{T+H-s+1}, \ldots, u_{T+H}), \end{aligned} \tag{3.2}$$

where l_s denotes the stationary density of s consecutive values of the ε-noncausal component u, denoted by $u_{\tau-s+1}, \ldots, u_\tau$. The above conditional density given the last s future states is known when polynomials Φ, Ψ and the p.d.f. g of the error ε are known too (see Section 3.2).

Example 2 : If $s = 1$, we get $u_t - \psi u_{t+1} = \varepsilon_t$. The conditional density in the numerator of (3.2) is :

$$\begin{aligned} & l(u_T, \ldots, u_{T+H-1} | u_{T+H}) \\ &= l(u_T | u_{T+1}) l(u_{T+1} | u_{T+2}) \ldots l(u_{T+H-1} | u_{T+H}) \\ &= g(u_T - \psi u_{T+1}) g(u_{T+1} - \psi u_{T+2}) \ldots g(u_{T+H-1} - \psi u_{T+H}). \end{aligned}$$

Example 3 : If $s = 2$, we get $u_t - \psi_1 u_{t+1} - \psi_2 u_{t+2} = \varepsilon_t$. The conditional density is:

$$\begin{aligned} & l(u_{T-1}, \ldots, u_{T+H-2} | u_{T+H-1}, u_{T+H}) \\ &= l(u_{T-1}|u_T, u_{T+1}) \ldots l(u_{T+H-2}|u_{T+H-1}, u_{T+H}) \\ &= g(u_{T-1} - \psi_1 u_T - \psi_2 u_{T+1}) \ldots g(u_{T+H-2} - \psi_1 u_{T+H-1} - \psi_2 u_{T+H}). \end{aligned}$$

Example 4 : In special cases, the predictive density of the ε-noncausal component u admits a closed form. For example, when the error ε follows a Cauchy distribution, the predictive density is:

$$l(u_{T+1}|u_T) = \frac{1}{\pi} \frac{1}{1 + (u_T - \psi u_{T+1})^2} \frac{1 + (1 - |\psi|)^2 u_T^2}{1 + (1 - |\psi|)^2 u_{T+1}^2}, \tag{3.3}$$

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

3.2 Estimation of the predictive density

The predictive density of interest in formula (3.2) is generally unknown for two reasons . First the error density g and the coefficients of autoregressive polynomials Φ, Ψ are usually unknown. Second, the stationary density l_s of s consecutive future values of u is unknown too, and it may be difficult to derive l_s from the transition p.d.f. of the ε-noncausal component u. Therefore, later in this section we propose a look-ahead-type of estimator for l_s.

The first difficulty is easily solved, when the p.d.f g of error ε is parametrized by a parameter denoted by θ. The autoregressive parameters in Φ, Ψ, and θ can be estimated consistently and asymptotically efficiently by the approximated maximum likelihood method [see e.g. Breidt et al. (1991), Lanne, Saikkonen (2011)], or by using a filter-based estimation method introduced in Section 4. Then, the conditional density in the numerator of formula (3.2) can be approximated by replacing the true current and past $u_t, t \leq T$, by their filtered values, computed from the observations on y [see section 2.2]

and $\hat{\Phi}$ [3]. Let us assume $H \geq s$. Then, the approximated conditional density in the numerator of formula (3.2) is:

$$\begin{aligned} & l(u_{T-s+1}, \ldots, u_{T+H-s} | u_{T-H-s+1}, \ldots, u_{T+H}) \\ & \quad \simeq \quad \hat{l}(\hat{u}_{T-s+1}, \ldots, \hat{u}_T, u_{T+1}, \ldots, u_{T+H-s} | u_{T-H-s+1}, \ldots, u_{T+H}), \text{ say.} \end{aligned} \tag{3.4}$$

Let us now focus on the second difficulty concerning the estimation of the unknown stationary p.d.f. l_s. By the Iterated Expectation Theorem, it follows that :

$$\begin{aligned} & l_s(u^*_{\tau-s+1}, \ldots, u^*_\tau) \\ & \quad = \quad E[l(u^*_{\tau-s+1}, \ldots, u^*_\tau | U_{\tau+1}, \ldots, U_{\tau+s})] \\ & \quad = \quad E\{g(u^*_{\tau-s+1} - \psi_1 u^*_{\tau-s+2} \ldots - \psi_s U_{\tau+1}) \ldots g(u^*_\tau - \psi_1 U_{\tau+1} \ldots - \psi_s U_{\tau+s})\}, \end{aligned}$$

where the expectation is taken with respect to the joint density of $(U_{\tau+1}, \ldots, U_{\tau+s})$. It is consistently approximated by :

$$\begin{aligned} & l_s(u^*_{\tau-s+1}, \ldots, u^*_\tau) \\ \simeq \quad & \frac{1}{T-s} \sum_{t=1}^{T-s} \{g(u^*_{\tau-s+1} - \psi_1 u^*_{\tau-s+2} \ldots - \psi_s u_t) \ldots g(u^*_\tau - \psi_1 u_{t+1} \ldots - \psi_s u_{t+s})\}, \end{aligned}$$

and, after replacing g by $\hat{g}$, ψ_j by $\hat{\psi}_j$, and the current and past ε-noncausal components u by their filtered values, we get:

$$\begin{aligned} \hat{l}_s(u^*_{\tau-s+1}, \ldots, u^*_\tau) = \frac{1}{T-s} \sum_{t=1}^{T-s} \quad & \Big\{ \hat{g}(u^*_{\tau-s+1} - \hat{\psi}_1 u^*_{\tau-s+2} \ldots - \hat{\psi}_s \hat{u}_t) \\ & \ldots \hat{g}(u^*_\tau - \hat{\psi}_1 \hat{u}_{t+1} \ldots - \hat{\psi}_s \hat{u}_{t+1}) \Big\}, \end{aligned} \tag{3.5}$$

[3] As shown in Example 3, the estimates of g and Ψ are also needed at this stage, although they don't appear explicitely in formula (3.2).

The approach used here to estimate the stationary p.d.f. l_s is a kind of look-ahead estimator suggested by Glynn, Hendersen (1998), (2001) [see also Garibotti (2004)]. The main difference is that the simulation step involved in the look-ahead approach has been eliminated by using the (asymptotically) stationary filtered values of the ε-noncausal component u.

We can now substitute $\hat{l}$ from (3.4) and $\hat{l}_s$ given in the last expression into the predictive density of interest. The consistent estimator of the predictive density in (3.2) is:

$$\begin{aligned} & \frac{\hat{l}(\hat{u}_{T-s+1},\ldots,\hat{u}_T,u_{T+1},\ldots u_{T+H}|u_{T+H-s+1},\ldots,u_{T+H})\hat{l}_s(u_{T+H-s+1},\ldots,u_{T+H})}{\hat{l}_s(\hat{u}_{T-s+1},\ldots,\hat{u}_T)} \\ \equiv\ & \hat{\Pi}(u_{T+1},\ldots,u_{T+H}|\hat{u}_{T-s+1},\ldots,\hat{u}_T), \text{say.} \end{aligned} \tag{3.6}$$

Let us illustrate the predictive density $\hat{\Pi}$ at horizon H for a noncausal process of order 1.

Example 5 : If $s=1$, we get :

$$\begin{aligned} & \hat{\Pi}(u_{T+1},\ldots,u_{T+H}|\hat{u}_T) \\ \equiv\ & \frac{\hat{g}(\hat{u}_T-\hat{\psi}u_{T+1})\hat{g}(u_{T+1}-\hat{\psi}u_{T+2})\ldots\hat{g}(u_{T+H-1}-\hat{\psi}u_{T+H})\sum_{t=1}^{T-1}\hat{g}(u_{T+H}-\hat{\psi}\hat{u}_t)}{\sum_{t=1}^{T-1}\hat{g}(\hat{u}_T-\hat{\psi}\hat{u}_t)}. \end{aligned} \tag{3.7}$$

Expression (3.6) provides a consistent approximation of the predictive density of interest $\hat{\Pi}$. The look-ahead method used to estimate the stationary p.d.f. l_s allows us to avoid simulations and considerably simplifies the methodology, which otherwise is computationally cumbersome [see, Lanne, Luoto, Saikkonen (2012)].

3.3 Prediction of future y

The approximate predictive density $\hat{\Pi}$ of the ε-noncausal component u in (3.6) can be used to generate the future values or future paths of the observ-

able process y and its unobservable causal and noncausal components over a given horizon H.

The approach consists of four steps outlined below. The future values of y are computed from the future values of u that are drawn in $\hat{\Pi}$ by applying a SIR method (see Appendix 2 for a description of the method). More specifically, the procedure is the following:

Step 1 : Use data $(y_1, \ldots, y_T)$ to compute the filtered values of in-sample unobserved components u:

$$\hat{\varepsilon}_{r+1}, \ldots, \hat{\varepsilon}_{T-s},$$

$$\hat{v}_1, \ldots, \hat{v}_{T-s},$$

$$\hat{u}_{r+1}, \ldots, \hat{u}_T.$$

Step 2 : Compute the approximated predictive density $\hat{\Pi}$.

Step 3: Use the SIR method to simulate future u's: $u^s_{T+1}, \ldots, u^s_{T+H}$.

Step 4: Use the recursive formulas (2.2)-(2.4) to compute the future values $y^s_{T+1}, \ldots, y^s_{T+H}$, $\hat{\varepsilon}_{T-s+1}, \ldots, \hat{\varepsilon}_{T+H-s}$, $\hat{v}_{T-s+1}, \ldots, \hat{v}_{T+H-s}$.

One can draw a large number of future paths of length H in order to obtain a complete term structure of predictive densities and prediction intervals from $T+1$ up to $T+H$. From a practical point of view, it is important to choose a computationally convenient forecast horizon H. Indeed,

i) choosing $H \geq s$ is advantageous as the expression of $\hat{\Pi}$ gets simplified.

ii) It is computationally less demanding to apply the SIR approach in one step at horizon $H = 10$, say, than to apply the method recursively 10 times at horizon 1.

iii) Drawing entire future paths of length H provides the term structure of prediction intervals.

4 Estimation based on filtered components

Let us now discuss the parameter estimation in noncausal and mixed autoregressive processes. In recent literature, mixed autoregressive models have been estimated by the approximated maximum likelihood (AML) method [see e.g. Breidt et al. (1991), Lanne, Saikkonen (2011), (2013) [4]]. The AML estimators of parameters θ, Φ, Ψ are defined as:

$$(\hat{\theta}, \hat{\Phi}, \hat{\Psi}) = \arg\max_{\theta,\Phi,\Psi} \sum_{t=r+1}^{T-s} g(\Phi(L)\Psi(L^{-1})y_t; \theta). \tag{4.1}$$

The filtered values of unobserved components of y provide us with an alternative estimation procedure that involves a sequence of back- and forecasts. That alternative estimation method may be less computationally demanding than the AML and is based on the updating of the filtered values of unobserved components until the numerical convergence of the parameter estimates is achieved. The asymptotic properties of the associated fixed point estimator, especially its efficiency, follow from the deterministic one-to-one relationships between the observation on y and on each error ε and unobserved components u and v.

4.1 Likelihood-based fixed point estimator

Let us consider the p^{th} step of the algorithm that maximizes the log-likelihood function in the AML procedure. Let also $\Phi^{(p)}, \ldots, \Psi^{(p)}, \theta^{(p)}$ denote the values of the unknown parameters at step p, and

$$\hat{\varepsilon}^{(p)}_{r+1}, \ldots, \hat{\varepsilon}^{(p)}_{T-s}, \hat{v}^{(p)}_1, \ldots, \hat{v}^{(p)}_{T-s}, \hat{u}^{(p)}_{r+1}, \ldots, \hat{u}^{(p)}_T,$$

denote the filtered components computed with the parameter etimates $\Phi^{(p)}, \Psi^{(p)}, \theta^{(p)}$.

Step 1 : Updating θ

The filtered value of error ε is used to update the estimator of the parameter (vector) θ in error density $g(\theta)$:

$$\theta^{(p+1)} = \arg\max_{\theta} \sum_{t=r+1}^{T-s} \log g(\hat{\varepsilon}^{(p)}_t; \theta).$$

[4] see also Davis, Song (2012) for the multivariate framework.

Step 2 : Updating Φ

Given that $\Phi(L)v_t = \varepsilon_t$ [see (2.4)], the filtered component v is substituted into this causal autoregressive model of ε_t in order to update the estimator of Φ:

$$\Phi^{(p+1)} = \arg\max_{\Phi} \sum_{t=1}^{T-s} \log g[\Phi(L)\hat{v}_t^{(p)}; \theta^{(p)}].$$

Step 3 : Updating Ψ

By analogy, given that $\Psi(L^{-1})u_t = \varepsilon_t$ the filtered component u is substituted into the non-causal autoregressive model of ε_t in order to update the estimator of Ψ:

$$\Psi^{(p+1)} = \arg\max_{\psi} \sum_{t=r+1}^{T} \log g[\Psi(L^{-1})\hat{u}_t^{(p)}; \theta^{(p)}].$$

This likelihood-based approach provides efficient estimates of the parameters (see the discussion in Section 4.2).

4.2 The Berndt, Hall, Hall, Hausman (BHHH) algorithm

The back-forecasting method in Section 4.1 requires more optimizations of the log-likelihood function than the AML approach, although with respect to smaller numbers of parameters. From a numerical point of view, these optimizations can be facilitated by using a BHHH algorithm described below [see Berndt et al. (1974)]. The algorithm consists of the following steps:

Step 1 : Updating θ
Parameter (vector) θ at step $p+1$ is :

$$\theta^{(p+1)} = \theta^{(p)} + \left[\sum_{t=r+1}^{T-s} \frac{\partial \log g}{\partial \theta}(\hat{\varepsilon}_t^{(p)}, \theta^{(p)}) \frac{\partial \log g}{\partial \theta'}(\hat{\varepsilon}_t^{(p)}, \theta^{(p)}) \right]^{-1} \sum_{t=r+1}^{T-s} \frac{\partial \log g}{\partial \theta}(\hat{\varepsilon}_t^{(p)}, \theta^{(p)}).$$

Step 2 : Updating Φ

We have :

$$\Phi^{(p+1)} = \Phi^{(p)} - \left\{ \sum_{t=r+1}^{T-s} [\hat{v}_{t-1}^{(p)}, \ldots, \hat{v}_{t-r}^{(p)}]' [\hat{v}_{t-1}^{(p)}, \ldots, \hat{v}_{t-r}^{(p)}] \left(\frac{\partial \log g}{\partial \varepsilon} (\hat{\varepsilon}_t^{(p)}, \theta^{(p)})^2 \right\}^{-1} \right.$$

$$\left. \sum_{t=r+1}^{T-s} \left\{ [\hat{v}_{t-1}^{(p)}, \ldots, \hat{v}_{t-r}^{(p)}]' \frac{\partial \log g}{\partial \varepsilon} (\hat{\varepsilon}_t^{(p)}, \theta^{(p)}) \right\} \right.$$

Step 3 : Updating Ψ

We have :

$$\Psi^{(p+1)} = \Psi^{(p)} - \left\{ \sum_{t=r+1}^{T-s} [\hat{u}_{t+1}^{(p)}, \ldots, \hat{u}_{t+s}^{(p)}]' [\hat{u}_{t+1}^{(p)}, \ldots, \hat{u}_{t+s}^{(p)}] \left(\frac{\partial \log g}{\partial \varepsilon} (\hat{\varepsilon}_t^{(p)}, \theta^{(p)}) \right)^2 \right\}^{-1}$$

$$\left. \sum_{t=r+1}^{T-s} \left\{ [\hat{u}_{t+1}^{(p)}, \ldots, \hat{u}_{t+s}^{(p)}]' \frac{\partial \log g}{\partial \varepsilon} (\hat{\varepsilon}_t^{(p)}, \theta^{(p)}) \right\} \right.$$

The above BHHH algorithm converges numerically to the limiting values $\theta^{(\infty)}, \Phi^{(\infty)}, \Psi^{(\infty)}$, say. These limiting values are the solutions of the AML likelihood equations. Hence, the BHHH algorithm applied separately to each of the parameters provides an AML estimate.

The adjustment terms in steps 2 and 3 can be interpreted as regression coefficients. For example, the adjustment term in step 2 is the regression coefficient in a regression of a vector of ones as the dependent variable on the following regressor vector:

$$\left[\hat{v}_t^{(p)} \frac{\partial \log g}{\partial \varepsilon} (\hat{\varepsilon}_t, \theta^{(p)}), \ldots, \hat{v}_{t-r}^{(p)} \frac{\partial \log g}{\partial \varepsilon} (\hat{\varepsilon}_t^{(p)}, \theta^{(p)}) \right]'.$$

The BHHH algorithm given above differs from the standard BHHH algorithm that is applied jointly to all parameters. While the standard algorithm relies on the approximation of the complete information matrix:

$$\begin{pmatrix} \hat{I}_{\theta\theta} & \hat{I}_{\theta\Phi} & \hat{I}_{\theta\Psi} \\ \hat{I}_{\Phi\theta} & \hat{I}_{\Phi\Phi} & \hat{I}_{\Phi\Psi} \\ \hat{I}_{\Psi\theta} & \hat{I}_{\Psi\Phi} & \hat{I}_{\Psi\Psi} \end{pmatrix},$$

say, our approach sets to zero the off-diagonal blocks of the information matrix.

Finally note that this recursive algorithm is easily applicable to stream on-line data, and allows for continuous updating of the parameter estimates and filtered components.

5 Filtering and prediction in multivariate processes

The filtering and prediction methods introduced in the previous sections for univariate processes can be easily extended to mixed autoregressive moving-average processes or mixed vector autoregressive processes. These models can be rewritten as a Vector Autoregressive model of order 1 (VAR(1)) when the current and lagged values of the process are stacked in a vector of a larger dimension.

In recent literature, one finds two different approaches to the specification of noncausal multivariate autoregressive models, that reveals the difficulty in finding the appropriate state variables that would facilitate the prediction and estimation. For example, in order to generalize the results known from univariate processes, Lanne, Saikkonen (2013) assume that the noncausal VAR model, not necessarily of order 1, can be written as:

$$\Pi(L)\Phi(L^{-1})y_t = \epsilon_t,$$

where $det(\Pi(z)) \neq 0$ and $det(\Phi(z)) \neq 0$, for $|z| < 1$. As noted in Davis, Song (2012) such restrictions are strong and moreover the matrix lag polynomials do not commute, in general. Therefore, the above specification differs generally from the following one:

$$\Phi^*(L^{-1})\Pi^*(L)y_t = \epsilon_t.$$

Davis, Song (2012) consider an unrestricted noncausal VAR(1) model and propose to use the Jordan canonical form of an appropriate matrix in order

to circumvent the problem of state variables for the approximate maximum likelihood estimation.

In this section, we introduce a mixed vector autoregressive MVAR model and show how the filtering and prediction developed in Sections 2.3, 2.4 are easily extended to this multivariate framework.

5.1 The MVAR(r,s) model

Let us consider the following stationary vector autoregressive model for process y_t of dimension n:

$$\begin{aligned} \Phi(L)y_t &= \varepsilon_t^*, & (5.1)\\ \det \Phi(L) &= c\Phi^*(L)\Psi^*(L^{-1})L^s, & (5.2) \end{aligned}$$

where (ε_t^*) is a strong white noise of dimension n, and $\Phi(L)$ is a $n \times n$ is an autoregressive polynomial matrix in L of order 1. Its determinant $det\Phi(L)$ admits r roots strictly outside the unit circle and $s = n-r$ roots strictly inside the unit circle. Accordingly, Φ^* and Ψ^* are (scalar) polynomials in the lag and lead operators L and L^{-1}, respectively, with roots strictly outside the unit circle, and of the following degrees: $d^0\Phi^* = r, d^0\Psi^* = s, \Phi^*(0) = \Psi^*(0) = 1$. In addition, c is a non-zero scalar, $c \neq 0$.

In order to write the (two-sided) multivariate moving average representation of y, we introduce the adjoint matrix $\tilde{\Phi}(L)$ of $\Phi(L)$, that is the transpose of the matrix of cofactors, such that :

$$\Phi(L)\tilde{\Phi}(L) = \det \Phi(L) Id_n. \quad (5.3)$$

The adjoint matrix polynomial in L is also of order 1, The stationary, multivariate two-sided moving average representation of process y is :

$$y_t = \frac{1}{c}\frac{\tilde{\Phi}(L)}{\Phi^*(L)\Psi^*(L^{-1})}\varepsilon_t, \quad (5.4)$$

or

$$y_t = \frac{1}{\Phi^*(L)\Psi^*(L^{-1})}\eta_t, \quad (5.5)$$

where

$$\varepsilon_t = \varepsilon^*_{t+s}, \quad \eta_t = \frac{1}{c}\tilde{\Phi}(L)\varepsilon_t, \tag{5.6}$$

and where $1/\Phi^*(L)$, [resp. $1/\Psi^*(L^{-1})$] denotes the convergent (scalar) series in L [resp. L^{-1}], which is an inverse of $\Phi^*(L)$ [resp. of $\Psi^*(L^{-1})$].

Let us clarify the change of variables in (5.6). The change of the white noise from ε^*_t to ε_t is needed for the normalization of the noise and of the causal and noncausal lag polynomials Φ^*, Ψ^* to make them compatible with the univariate analysis presented in previous sections [5].

Equation (5.4) provides the two-sided multivariate moving average representation of y_t in terms of the strong white noise ε, while equation (5.5) is a two-sided multivariate moving average representation of y_t in terms of process η, which is itself a Vector Moving Average process of order 1 [VMA (1)].

5.2 Filtering

By analogy to equations (2.3)-(2.4), we define the η-causal and η-noncausal unobserved components and denote those processes by (u_t) and (v_t), respectively:

$$u_t = \Phi^*(L)y_t \Leftrightarrow \Psi^*(L^{-1})u_t = \eta_t, \tag{5.7}$$
$$v_t = \Psi^*(L^{-1})y_t \Leftrightarrow \Phi^*(L)v_t = \eta_t. \tag{5.8}$$

The η-causal component v and η-noncausal component u are processes of dimension n.
Along the lines of Section 2.4, we discuss the equivalence of information sets containing the unobserved components u, v and η.

Proposition 3 : The following information sets are equivalent:

i) $(y_1, \ldots, y_T)$;

ii) $(v_1, \ldots, v_r, \eta_{r+1}, \ldots, \eta_{T-s}, u_{T-s+1}, \ldots, u_T)$;

[5] This normalization is also required for the appropriate derivation of the approximated likelihood function. Note that this normalization requires the knowledge of the causal and noncausal autoregressive orders r and s.

iii) $(v_1, \ldots, v_r, \varepsilon_r, \varepsilon_{r+1}, \ldots, \varepsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$.

Proof : To prove the equivalence between i) and ii), use the same arguments as in the proof of Proposition 2 (see Appendix 1). However, in the multivariete setup, expression ii) of the information set is not of direct use, because the terms $\eta_{r+1}, \ldots, \eta_{T-s}, u_{T-s+1}, \ldots, u_T$ are dependent. The equivalence between i), ii) and information set iii) can be established by observing that $(\eta_{r+1}, \ldots, \eta_{T-s}) \subset (\varepsilon_r, \ldots, \varepsilon_{T-s})$ by (5.6), and that $(\varepsilon_r, \ldots, \varepsilon_{T-s}) \subset (y_1, \ldots, y_T)$, by (5.4).

QED

Information set iii), which contains $T+1$ vectors of unobserved components u, v, and white noise ε is more convenient for the filtering purposes than information set ii). The reason is that by (5.6-5.7) u_t is a linear combination of $\varepsilon_{t-1}, \varepsilon_t, \ldots..$ Therefore $(v_1, \ldots, v_r, \varepsilon_r, \ldots, \varepsilon_{T-s-1})$ and $(\varepsilon_{T-s}, u_{T-s+1}, \ldots, u_T)$ are independent.

Suppose that we have a sample of observations $(y_1, \ldots, y_T)$. The unobserved components can be filtered as follows:

$v_1, \ldots, v_{T-s}$, by the first equation of (5.8),

$u_{T-s+1}, \ldots, u_T$, by the first equation of (5.7),

$\eta_{r+1}, \ldots, \eta_{T-s}$, by the second equation of (5.7),

$\varepsilon_r, \ldots, \varepsilon_T$. by equation (5.6).

5.3 Prediction

The prediction problem is solved along the lines of Section 3. We observe that predicting y_{T+1} [resp. $y_{T+1}, ..., y_{T+H}$],is equivalent to the prediction of u_{T+1} [resp. $u_{T+1}, ..., u_{T+H}$] given $y_1, ..., y_T$.

For expository purpose, let us consider short term prediction at horizon 1: $H = 1$. The predictive density of u_{T+1} is :

$$
\begin{aligned}
& l(u_{T+1}|y_1,\ldots,y_T) \\
= \; & l(u_{T+1}|v_1,\ldots,v_r,\varepsilon_r,\varepsilon_{r+1},\ldots,\varepsilon_{T-s},u_{T-s+1},\ldots,u_T) \\
= \; & l(u_{T+1}|\varepsilon_{T-s},u_{T-s+1},\ldots,u_T), \text{by the independence property.}
\end{aligned}
$$

The only difference with the univariate case is the presence of n-dimensional ε_{T-s} in the conditioning set. In other words, the state variable summarizing the future of y is enlarged and includes not only s past $u's$, but also an ε. Finally, the predictive density of y_{T+1} given $y_1, ..., y_T$ is obtained by replacing u_{T+1} by $\Phi^*(L)y_{T+1}$ by the first equation of (5.7), since the associated Jacobian is equal to $\Phi^*(0) = 1$.

6 Simulation Study

To illustrate the implementation of the filtering, prediction and estimation methods presented in the previous sections above, we perform a simulation study based on the data generating process described below. It is meant to replicate the dynamics observed in time series such as the commodity prices, Bitcoin/USD exchange rates and S&P 500 returns. In practice, these processes can be modelled as stationary noncausal and mixed processes that display short-lived explosive patterns, called bubbles. A bubble is characterized by a phase of slow or moderate growth, which is followed by a sudden drop, and can be replicated in practice by assuming that the errors of the noncausal model are Cauchy distributed [see Gourieroux, Zakoian (2013)].

6.1 The Data Generating Process

The process examined is a mixed causal-noncausal autoregressive process of orders $r = 1$, $s = 1$:

$$(1 - \varphi L)(1 - \psi L^{-1})y_t = \varepsilon_t, \tag{6.1}$$

with Cauchy errors :

$$g_\sigma(\varepsilon) = \frac{1}{\sigma\pi}\frac{1}{1+\varepsilon^2/\sigma^2} = \frac{\sigma}{\pi}\frac{1}{\sigma^2+\epsilon^2}. \tag{6.2}$$

We consider four sets of parameter values, in which σ and ϕ are constant while parameter ψ takes on four different values:

$$\sigma = 1, \varphi = 0.3, \psi = 0, 0.3, 0.5, 0.9.$$

Figure 1 shows the simulated trajectory of length $T = 200$ for the four sets of parameter values given above and with the same vector of simulated Cauchy-distributed errors.

[Insert Figure 1 : Simulated Paths]

We observe several positive and negative bubbles of different durations and magnitudes. We also find that the larger the noncausal autoregressive parameter ψ, the longer and larger the increasing phase of the bubble.

6.2 Parameter estimation

For each DGP defined in (6.1-6.2), the parameters φ, ψ, σ are estimated by the approximated maximum likelihood and by the BHHH algorithm of Section 4.2. The estimated values of the parameters are given in Table 1.

Table 1: AML and BHHH Estimates of ϕ, ψ and σ

	Parameter	standard error	t-ratio
	Maximum likelihood		
$\psi = 0.0$	0.004	0.015	0.285
$\phi = 0.3$	0.281	0.017	15.646
$\sigma = 1$	0.877	0.094	8.721
$\psi = 0.3$	0.303	0.014	21.428
$\phi = 0.3$	0.281	0.017	15.800
$\sigma = 1$	0.878	0.095	8.611
$\psi = 0.5$	0.505	0.012	39.558
$\phi = 0.3$	0.280	0.018	15.318
$\sigma = 1$	0.878	0.097	8.467
$\psi = 0.9$	0.902	0.006	131.649
$\phi = 0.3$	0.280	0.018	15.058
$\sigma = 1$	0.881	0.097	9.041
	BHHH		
$\psi = 0.0$	0.004	0.018	0.226
$\phi = 0.3$	0.281	0.016	16.725
$\sigma = 1$	0.877	0.086	10.101
$\psi = 0.3$	0.303	0.016	18.068
$\phi = 0.3$	0.281	0.016	16.742
$\sigma = 1$	0.878	0.087	10.091
$\psi = 0.5$	0.504	0.014	34.341
$\phi = 0.3$	0.280	0.016	16.732
$\sigma = 1$	0.879	0.087	10.048
$\psi = 0.9$	0.902	0.006	139.427
$\phi = 0.3$	0.280	0.016	16.635
$\sigma = 1$	0.881	0.087	10.057

The estimation results from the two methods are close and seem to be quite accurate. The accuracy of parameter ϕ is almost independent of the value of parameter ψ, which controls the length and size of the bubble. The variance of ψ decreases when that parameter approaches the unit root.

6.3 Filtering

The filtered causal and noncausal components and errors are calculated from the simulated y and AML parameter estimates. Figure 2 shows their trajectories based on the DGP with $\psi = 0.9$

[Insert Figure 2 : Trajectories of the components and errors]

The series of filtered ϵ is close to a series of independent drawings in the Cauchy distribution. The filtered ϵ series reveals the dates of extreme values of Cauchy distributed errors. Recall that process y_t can be represented in terms of its unobserved ϵ-causal and ϵ-noncausal components as:

$$y_t = \frac{1}{1-\phi\psi}(u_t - \phi v_{t-1}).$$

Both unobserved components contribute to the formation of bubbles. Component u with parameter ψ determines the increasing phase of the bubble while component v and parameter ϕ determine the bubble burst. Both components u and v have AR(1) representations in reverse and direct times, respectively. The high value of parameter $\psi = 0.9$ explains strong persistence in reverse time of the ϵ-noncausal process u.

Figure 3 provides the sample ACF of these filtered components.

[Insert Figure 3 : ACF of the components]

The asymptotic distribution of the sample ACF of these components differs from the standard Gaussian distribution derived for processes with finite variance. The confidence bounds have to be adjusted for Cauchy errors, as the rates of convergence of autocorrelations estimators differ and the limiting distribution involves ratios of symmetric stable distributions [see e.g. Davis, Resnick (1986)]. The adjusted bounds are given in Figure 3 and are significantly larger than the standard ones. We observe that the hypothesis of a strong white noise cannot be rejected for the filtered $\hat{\epsilon}$ component, while significant dependence is displayed by the ϵ-causal v and ϵ-noncausal u components.

6.4 Predictive densities

a) Short-term prediction

The predictive density of ϵ-noncausal component u is estimated by the formula given in Example 4, Section 3.2. That density is shifted by a constant to become the predictive density of process y. Let us first consider short horizons $H = 1, 2$ and the set of parameter values $\varphi = 0.3, \psi = 0.9, \sigma = 1$. The last in-sample values used for the prediction are: $y_T = 16.67, y_{T-1} = 14.27, u_T = 12.39$. They correspond to the last values of the simulated trajectory of y for $T = 200$ and the fourth set pf parameters given in Figure 1.

[Insert Figure 4 : Predictive density at horizon 1, $y_T = 16.67, u_T = 12.39$.]

The short term predictive density is peaked around the last observed value $y_T = 16.67$ and has a long left tail. For example, the probability of an increase of y between dates T and $T + 1$ is equal to 0.59 and is significantly larger than 0.5.

It is also possible to display the joint predictive density for y_{T+1}, y_{T+2}. We show that bivariate density in Figure 5 and its contour plot in Figure 6 [6].

[Insert Figure 5 : Joint predictive density at horizon 2, $y_T = 16.67, u_T = 12.39$.]

[Insert Figure 6: Contour plot of the predictive density at horizon 2]

The joint predictive density is rather concentrated around $(y_{T+1}, y_{T+2}) = (y_T, y_T) = (16.67, 16.67)$ and the lines of the contour plot differ considerably from ellipsoids. They resemble contour plots encountered in the joint analysis of extreme events [see e.g. Balkema, Embrechts, Nolde (2013)]. We observe two risk directions with left tails fatter than right tails in each direction. Moreover, these directions are affine as the observations are defined by linear dynamic equations.

The predictive densities were based on $y_T = 16.67, u_T = 12.39$, that are the last observed and filtered value of the simulated series and its unobserved component, respectively. If the predictive density were computed at different dates and values y, u, different shapes of predictive densities would be obtained. This dependence is illustrated in Figure 7.

[Insert Figure 7 : Predictive densities for different states.]

[6] It is estimated by the approach described in Section 3.2, even though it is also available in closed-form (see Example 4).

The location of predictive densities depends on y_T, while the influence of u_T is more complex. Figure 7 shows that u_T can create left or right asymmetries.

b) Future pattern recognition

The joint predictive density at horizon 2 can be used to study the likelihood of different types of future dynamics of y. The simulated data is such that $y_T = 16.67 > y_{T-1} = 14.27$, which corresponds to a recent increase in y. From the increasing pattern observed at the end of the simulated series, one can infer that the prediction is likely performed at the beginning of a bubble. Therefore, it is interesting to examine if that increase will continue in the future, or will be followed by a downturn at a future date and what would be the magnitude of that downturn. The various possible future scenarios at horizon 2 are displayed in Figure 8.

[Insert Figure 8 : Predicted patterns.]

The plane in y_{T+1} and y_{T+2} is divided into eight semi-orthants centered at y_T, y_T. The future pattern depends on the semi-orthant, which characterizes the future increase or decrease of y_{T+1}, y_{T+2}, and their positioning with respect to the last observed value.

In order to provide more insights on the future patterns, we compute the probabilities of selected future scenarios, given $y_T = 16.67, y_{T-1} = 14.27$

Table 2: Probabilities of future patterns

pattern	probability
$y_T < y_{T-1} < y_{T+1} < y_{T+2}$	0.176
$y_{T-1} < y_T$, $y_T > y_{T+1} > y_{T+2}$	0.132
$y_{T-1} < y_T < y_{T+1}$, $y_{T+2} < y_{T+1}$	0.520
$y_{T-1} < y_T < y_{T+1}$, $y_{T+2} < y_{T+1}$, $y_{T+2} < y_{T-1}$	0.008

If the future values y_{T+1}, y_{T+2} were "uniformly" distributed on $(-\infty, \infty)^2$, the first three probabilities would be equal to 1/8=0.125 (corresponding to a semi-orthant) and 3/8=0.375 (corresponding to three semi-orthants). The probability of a continuing increase over two next periods, that is of a downturn after date $T+3$ is equal to 0.176 and significantly above the benchmark of 0.125. The probability of a downturn at $T+2$ is especially large and equal to 0.520 for a benchmark of 0.375. To see if that downturn is sharp,

we provide in the fourth row of Table 2 the probability of y returning to the value prior to two consecutive increases. The probability is small and equal to 0.008, hence the downturn cannot be sharp.

c) Medium-term prediction

When the prediction horizon H is larger than 3, the closed form expression of the joint predictive density cannot be used directly. The contour plot is not available and the predictive density of y_{T+10} alone, say, cannot be found. The reason is that it would require integrating out all the intermediate values of $y_{T+1}, ..., y_{T+9}$, which is numerically difficult. Even at a short horizon $H = 1, 2$, it may be difficult to derive the prediction intervals from the closed-form formula of the predictive density. A closed form of that density does not ensure the existence of a closed-form cdf or a closed-form quantile function. Instead, a sampling-importance resampling approach described in Section 2 can be used. To do that, we exploit the recursive relationship between the y_t and the u_t and simulate a set of future paths of u. Given that $u_t - \psi u_{t+1} = \epsilon_t$, we know that (u_t) is a Markov process of order 1 in reverse time. Therefore it is also a Markov process in calendar time,but with nonlinear dynamics [see Gourieroux, Zakoian (2013)]. As the instrumental misspecified model in the sampling step of the SIR, we use a Gaussian AR(1) model:

$$\tilde{u}_t = \hat{\rho}\tilde{u}_{t-1} + \hat{\sigma}\tilde{\epsilon}_t,$$

where $\tilde{\epsilon}_t \sim IIN(0, 1)$ and $\hat{\rho}$ [resp. $\hat{\sigma}$] is the sample autocorrelation of order 1 [resp. residual variance] computed from the series $\hat{u}_t, t = 1, ..., T$. This auxiliary model is clearly misspecified as the true dynamics of u_t is not linear, and its first-order moment does not exist. Nevertheless, the resampling scheme will correct for the misspecification error. The sample sizes for sampling and resampling are $S = 2000$ and $S^* = 5000$, respectively.

Figures 9 and 10 display the term structures of predictive density and the 95% prediction interval, respectively. The selected horizons are $H = 9$ and $H = 10$, respectively.

[Insert Figure 9: Term structure of predictive density]

[Insert Figure 10: Term structure of prediction interval]

The increase of the width of the prediction interval with regard to the horizon is not surprising. However, it is interesting to compare the pattern of

the upper quantile with the pattern of the upper quantile of a causal process with a unit root (ψ is close to 1) and errors with finite variance. Such a standard upper quantile changes at the rate of $\sqrt{H}$ and is a concave function of time. In our framework, the pattern is convex. Indeed, the probability of a bubble increases with H and the prediction interval accounts also for the sustainability of the bubble due to the large value of ψ.

7 Concluding remarks

This paper revisited the filtering, prediction, simulation and estimation in mixed causal/noncausal autoregressive processes. A new prediction method was proposed along with the back-forecasting algorithms for simulation of mixed processes and for asymptotically efficient estimation of the causal and noncausal dynamic parameters.

The proposed methods simplify those introduced in the literature on non-linear prediction of the process with causal and noncausal components.

Moreover, the proposed univariate methods of filtering, prediction and estimation were extended to the multivariate framework and illustrated by a simulation study.

References

[1] Balkema, G., Embrechts, P., and N., Nolde (2013) : "The Shape of Asymptotic Dependence", in Springer Proceedings in Mathematics and Statistics, Special Volume: "Prokhorov and Contemporary Probability Theory", eds. A. Shirayev, S. Varadhan, and E. Presman, 33, 43-67.

[2] Berndt, E., Hall, B., Hall, R., and J., Hausman (1974) : "Estimation and Inference in Nonlinear Structural Models", Annals of Economic and Social Measurement, 3, 653-665.

[3] Braun, A., Li, H., and J., Stachurski (2012) : "Generalized Look-Ahead Methods for Computing Stationary Densities", Mathematics of Operations Research, 37, 489-500.

[4] Breidt, F., Davis, R., Li, K., and M., Rosenblatt (1991) : "Maximum Likelihood Estimation for Noncausal Autoregressive Processes", Journal of Multivariate Analysis, 36, 175-198.

[5] Brockwell, P., and R., Davis (1991) : "Time Series : Theory and Methods", 2nd edition, Springer Verlag, New-York.

[6] Cambanis, S., and I., Fakhre-Zakari (1994): "On Prediction of Heavy-Tailed Autoregressive Sequences : Forward Versus Reversed Time", Theory Probab. Appl., 39, 217-233.

[7] Chen, B., Choi, J., and J.C., Escanciano (2012) : "Testing for Fundamental Vector Moving Average Representations", DP Indiana University.

[8] Davis, R., and S., Resnick (1986) : "Limit Theory for the Sample Covariance and Correlation Function of Moving Averages", Ann. Stat., 14, 533-558.

[9] Davis, R., and L., Song (2012) : "Noncausal Vector AR Processes with Application to Economic Time Series", DP Columbia University.

[10] Efron, B. (1982) : "The Bootstrap, Jacknife and Other Resampling Plans", Philadelphia Society of Industrial and Applied Mathematics.

[11] Garibotti, G. (2004) : "Estimation of the Stationary Distribution of Markov Chains", PhD Dissertation, University of Massachussets.

[12] Gelfand, A., and A., Smith (1990)a : "Sampling Based Approaches to Calculating Marginal Densities", J. Amer. Statist. Assoc., 85, 398-409.

[13] Gelfand, A., and A., Smith (1990)b : "Bayesian Statistics Without Tears : A Sampling Resampling Perspective", 46, 84-88.

[14] Glynn, P., and S., Hendersen (1998) : "Estimation of Stationary Densities of Markov Chains", Mederos, D., Watson, E., Carson, J., and M., Manivannan eds., Proc. 1998, Winter Simulation, Conf IEIE, Piscataway, NJ.

[15] Glynn, P., and S., Hendersen (2001) : "Computing Densities for Markov Chains via Simulation", Mathematics for Operations Research, 26, 375-400.

[16] Gourieroux, C., and A. Hencic (2014): "Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate", York University DP.

[17] Gourieroux, C., and J.M., Zakoïan (2013) : "Explosive Bubble Modelling by Noncausal Processes", CREST DP.

[18] Lanne, M., Luoto, J., and P., Saikkonen (2012) : "Optimal Forecasting of Noncausal Autoregressive Time Series", International Journal of Forecasting, 28, 623-631.

[19] Lanne, M., and P., Saikkonen (2011) : "Noncausal Autoregressions for Economic Time Series", Journal of Time Series Econometrics, De Gruyter, 3, 1-32.

[20] Lanne, M., and P., Saikkonen (2013) : "Noncausal Vector Autoregressive", Econometric Theory, 29, 447-481.

[21] Rosenblatt, M. (2000) : "Gaussian and Non-Gaussian Linear Time Series and Random Fields", New-York, Springer Verlag.

[22] Rubin, D. (1988) : "Using the SIR Algorithm to Simulate Posterior Distribution", in Bayesian Statistics, 3, eds. Bernando, J., De Groot, M., Lindley, D., and A., Smith, Oxford University Press.

[23] Stuck, B. (1977) : "Minimum Error Dispersion Linear Filtering of Scalar Symmetric Stable Distributions", IEEE Transactions on Automatic Control, 23, 507-509.

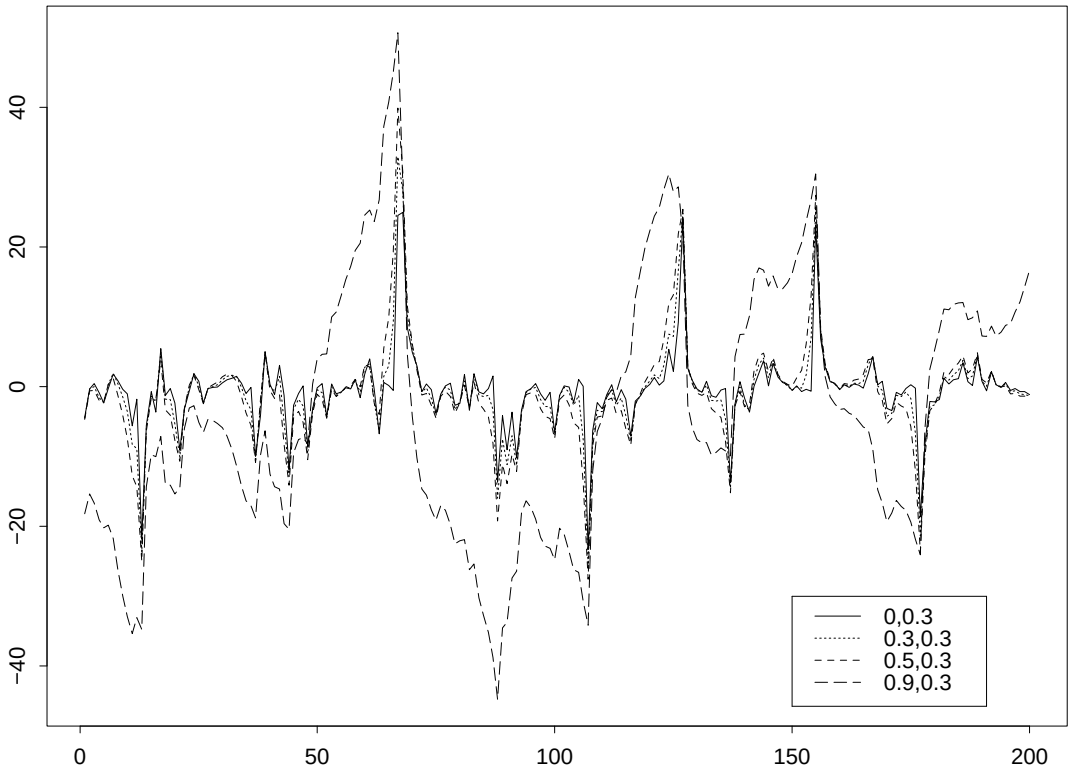

Figure 1: Simulated paths

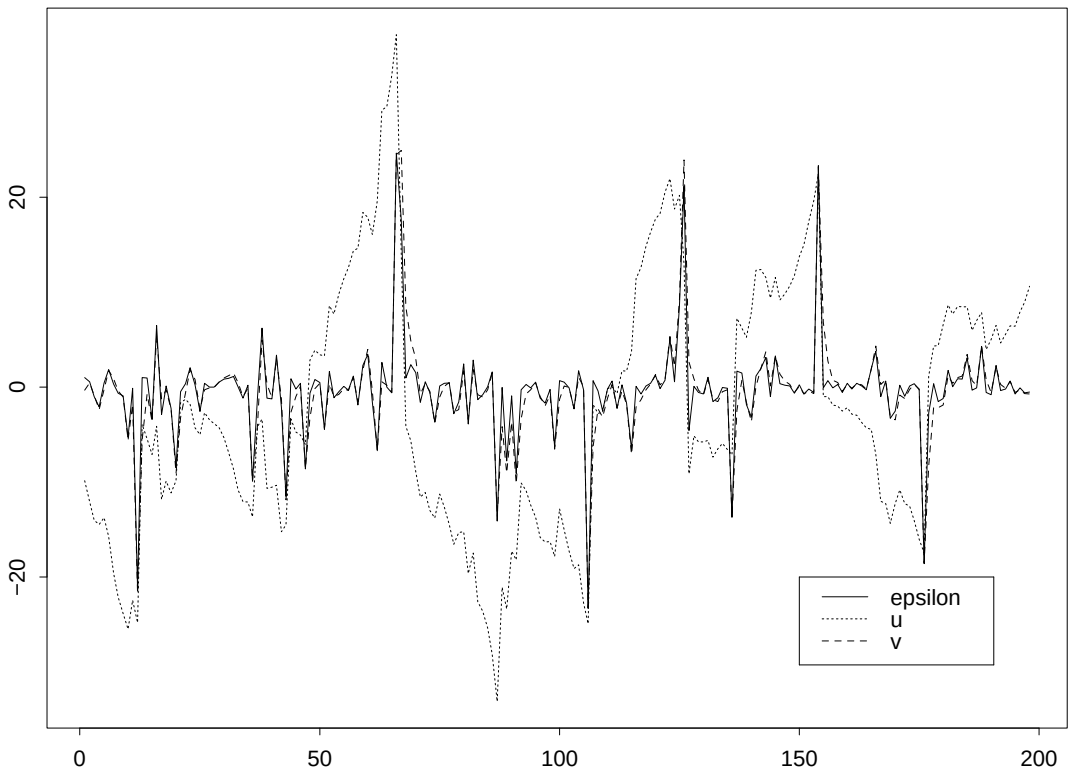

Figure 2: Component series

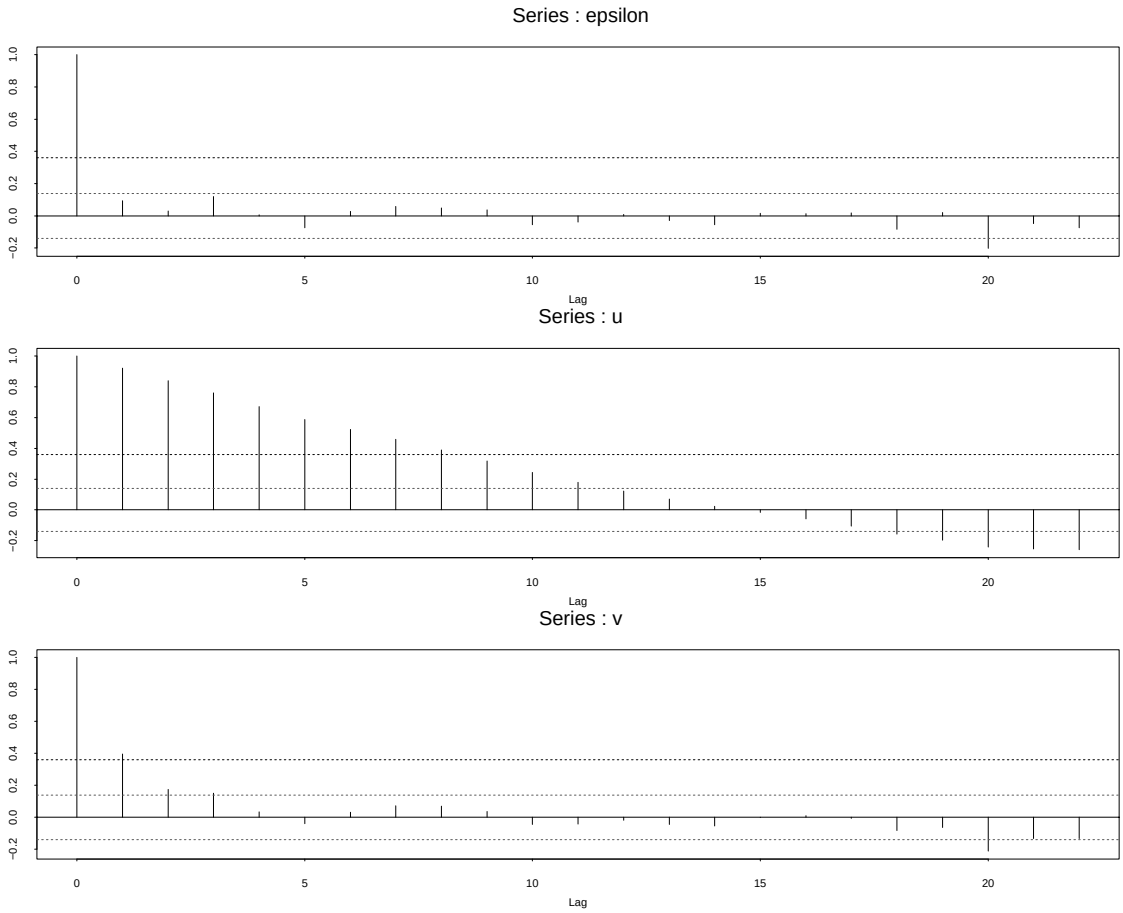

Figure 3: ACF of component series

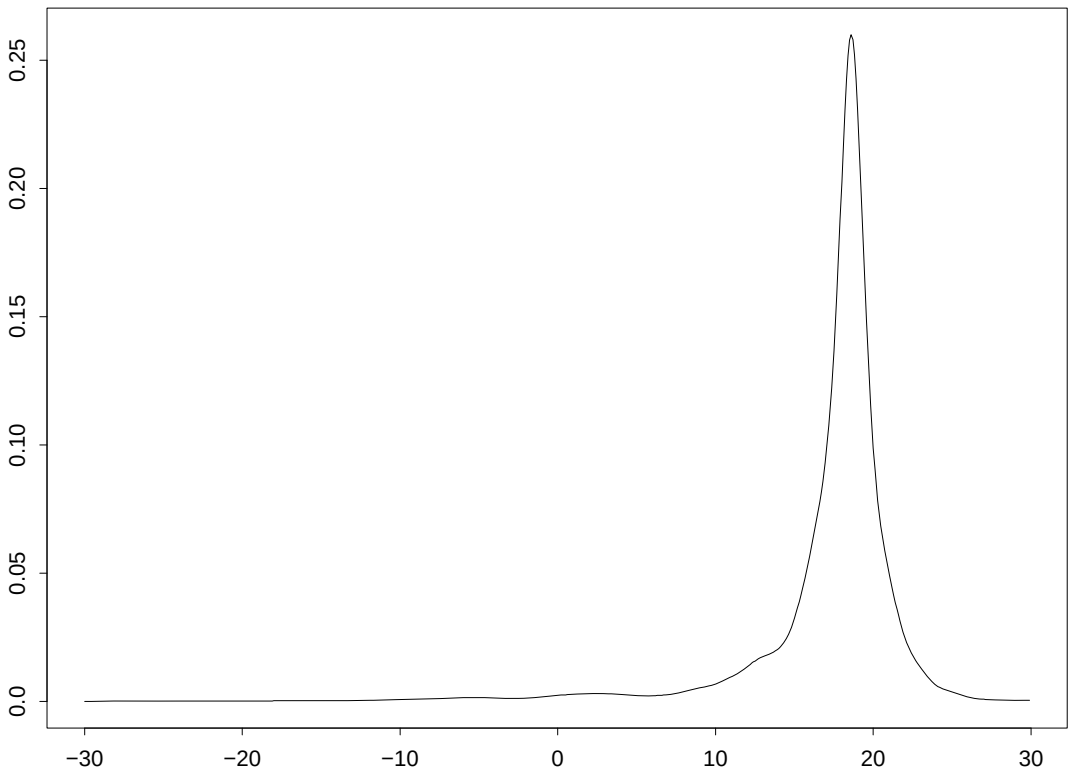

Figure 4: Predictive density at horizon 1, $y_T = 16.67, u_T = 12.39$

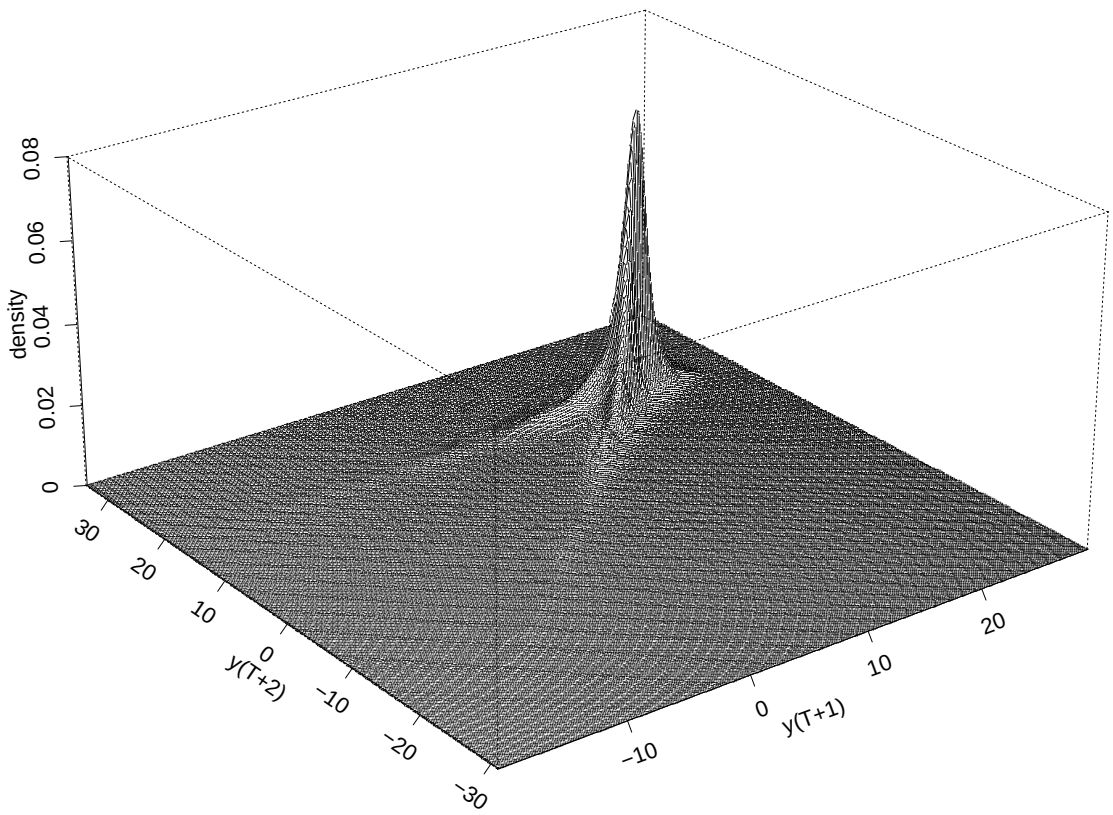

Figure 5: Joint predictive density at horizon 2, $y_T = 16.67, u_T = 12.39$

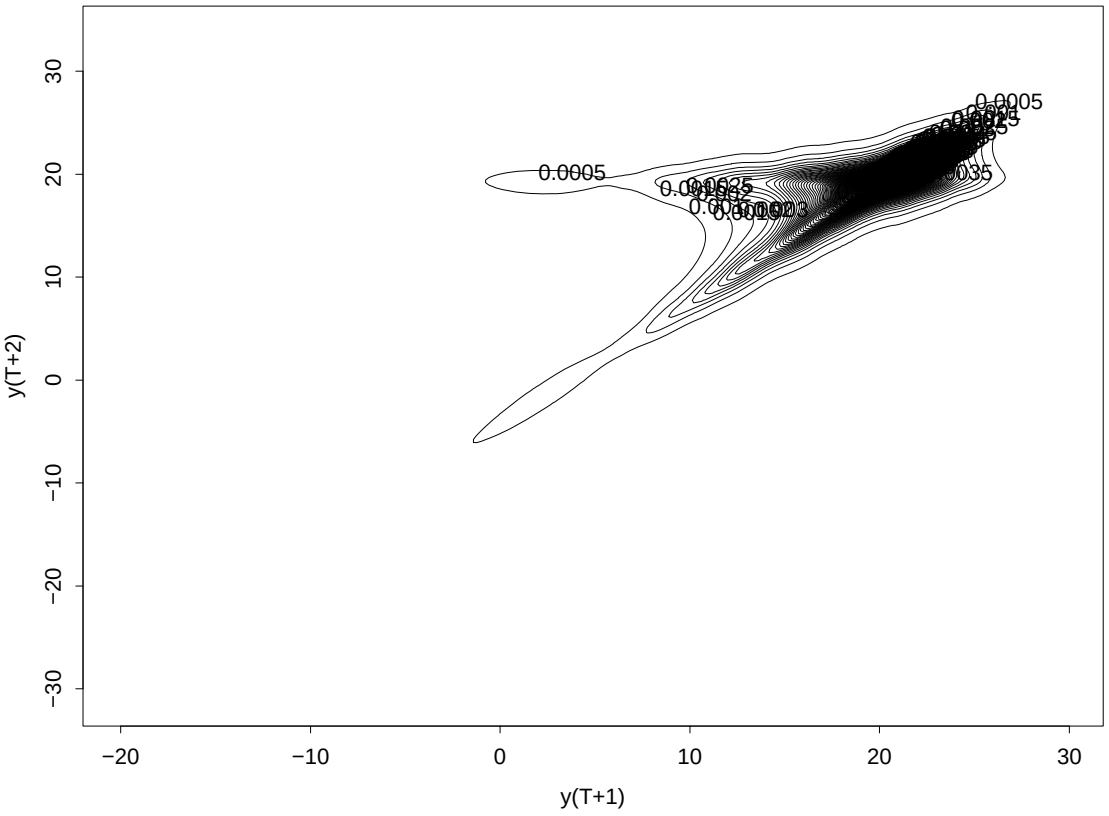

Figure 6: Contour plot, predictive density at horizon 2

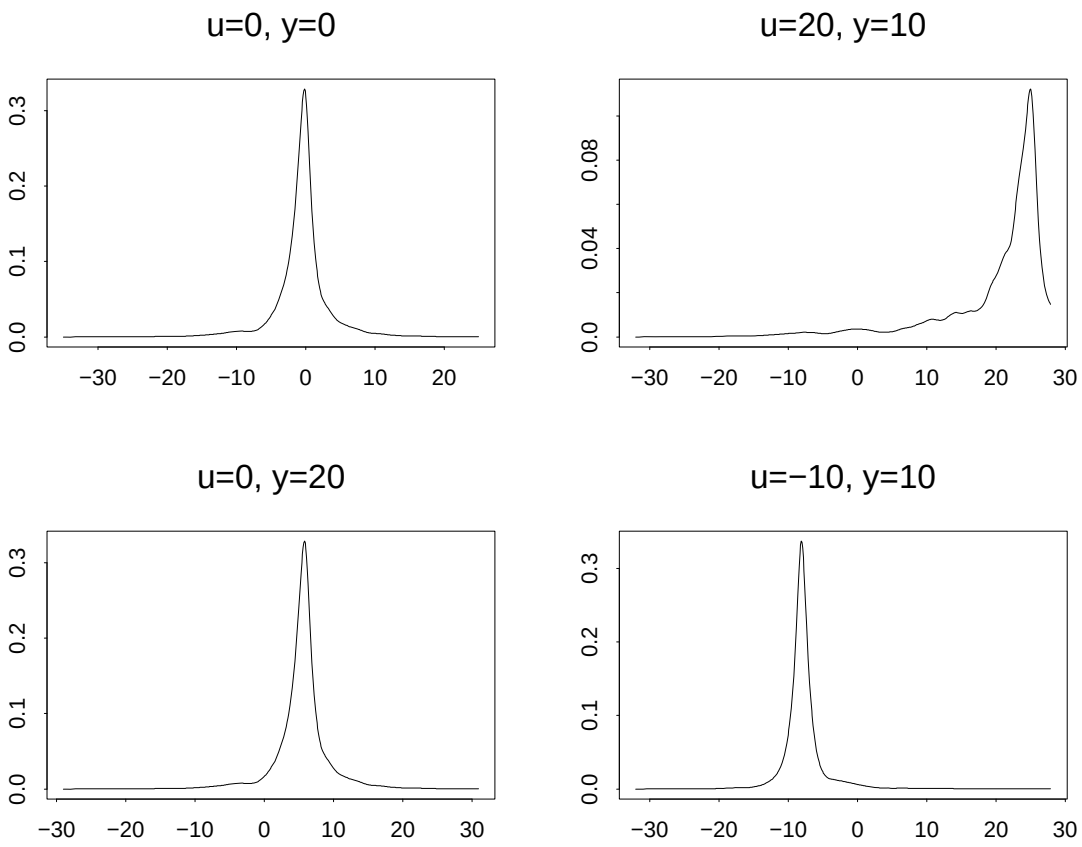

Figure 7: Predictive densities for different states

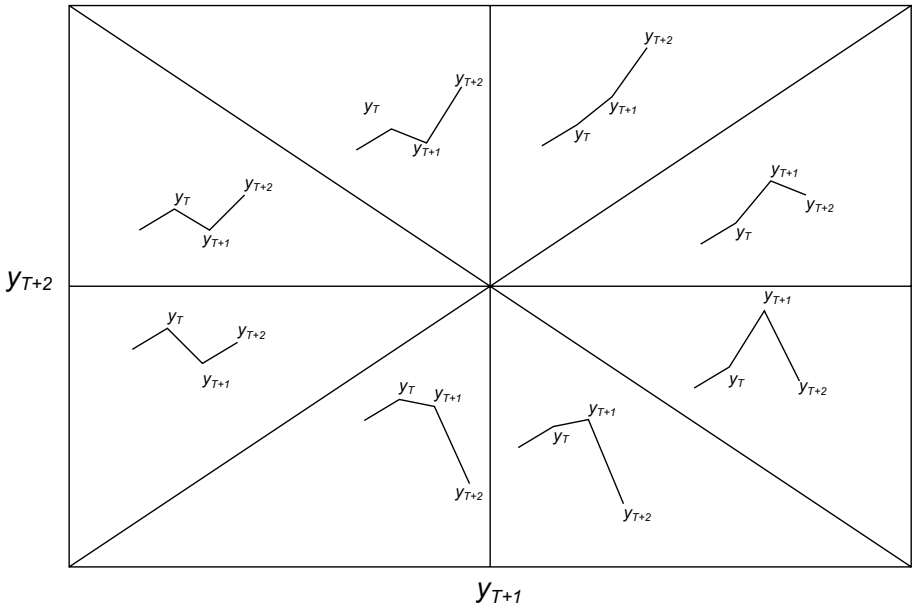

Figure 8: Predicted patterns

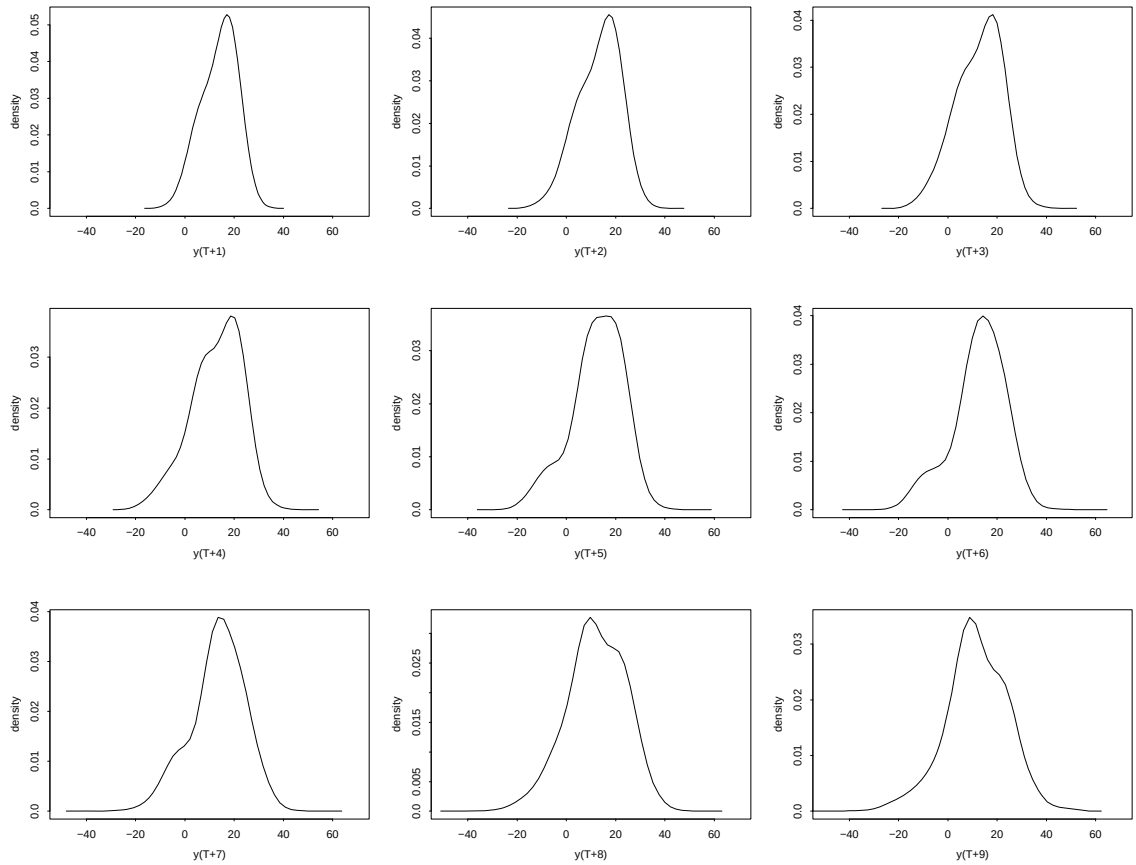

Figure 9: Term structure of predictive density

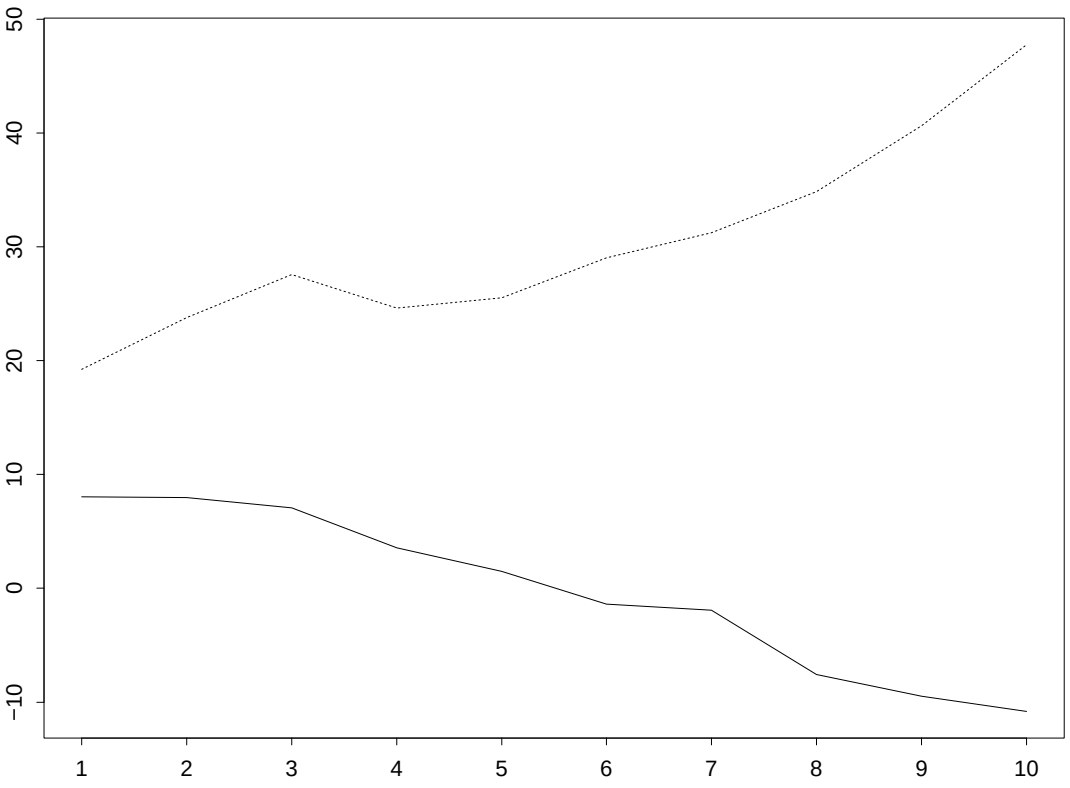

Figure 10: Term structure of prediction interval

A P P E N D I X 1

Proof of Proposition 2

i) is equivalent to ii) by applying recursively the first equality of (2.3).

i) is equivalent to iii) by applying recursively the first equality of (2.4).

ii) is equivalent to iv) by applying recursively the second equality of (2.3).

iii) is equivalent to v) by applying recursively the second equality of (2.4).

iv) is equivalent to vi) by applying the $(u - v)$ noncausal decomposition (2.6).

v) is equivalent to vi) by applying the $(u - v)$ causal decomposition (2.5).

A P P E N D I X 2

The Sampling Importance Resampling (SIR) Method

This approach has been introduced in Rubin (1988) and its were properties studied in Geldfand, Smith (1992). It is a weighted variant of the bootstrap resampling procedure [Efron (1982)]. The aim of the procedure is to draw independent values in a distribution whose density f is known, but the quantile function difficult to compute. The method requires an instrumental (or importance) distribution with known density g, in which it is easy to draw.

The steps are the following ones :

step 1 : Sampling

First draw independent values $X^s, s = 1, \ldots, S$, in distribution G.

step 2 : Importance Resampling

Then draw independent values $\tilde{Y}^1, \ldots, \tilde{Y}^S$ in the simulated set $\{X^1, \ldots, X^S\}$, with weights $f(X^s)/g(X^s), s = 1, \ldots, S$.

We have :

$$E[a(\tilde{Y}^s)] = EE[a(\tilde{Y}^s)|X^1, \ldots, X^S]$$

$$= E\left\{\left[\sum_{s=1}^{S} \frac{f(X^s)}{g(X^s)} a(X^s)\right] / \left[\sum_{s=1}^{S} \frac{f(X^s)}{g(X^s)}\right]\right\}$$

$$= E\left\{\frac{1}{S}\sum_{s=1}^{S}\left[\frac{f(X^s)}{g(X^s)} a(X^s)\right] / \frac{1}{S}\sum_{s=1}^{S}\left[\frac{f(X^s)}{g(X^s)}\right]\right\},$$

for any integrable function a.

If the number S of replications is large, we get :

$$
\begin{aligned}
& E[a(\tilde{Y}^s)] \\
\simeq\ & E\left\{E\left[\frac{f(X)}{g(X)}a(X)\right]/E\left[\frac{f(X)}{g(X)}\right]\right\} \\
=\ & E\left[\frac{f(X)}{g(X)}a(X)\right]/E\left[\frac{f(X)}{g(X)}\right] \\
=\ & E[a(Y)].
\end{aligned}
$$

We deduce that the simulated values $\tilde{Y}^1, \ldots, \tilde{Y}^S$ are drawn in distribution F, when $S \to \infty$.

Moreover, since $E[a(\tilde{Y}_s)|X^1, ..., X^S] \approx E[a(\tilde{Y}_s)]$ for S large, these simulated values are asymptotically independent.

The sample size can be as large as desired and the distribution can be multivariate. As usual in Monte-Carlo integration, the more g resembles f, the smaller the number of replications S needed to get accurate results. We can also use different simulation lengths S and S^* in the sampling and resampling steps.

A P P E N D I X 3

Backward/-Forward Linear Regressions

It seems natural to apply the methods described in Sections 4.1-4.2 after replacing the true density g of the errors by a Gaussian pseudo-density. This could simplify the algorithm, but this approach inherits the identification problem encountered in the Gaussian case, even if we alternate backward and forward regressions. We first describe the algorithm and then discuss the notions of pseudo-autocorrelation and pseudo-Yule-Walker equations.

A.3.1 The algorithm

The algorithm is as follows and focuses on parameters Φ and Ψ.

Step 1 : Updating Φ

The autoregressive parameters $\varphi_1, \ldots, \varphi_r$ are estimated at step p by regressing $\hat{v}_t^{(p)}$ on $\hat{v}_{t-1}^{(p)}, \ldots, \hat{v}_{t-r}^{(p)}$.

Step 2 : Updating Ψ

The autoregressive parameters $\psi_1, \ldots, \psi_s$ are obtained by regressing

$$\hat{u}_t^{(p)} \text{ on } \hat{u}_{t+1}^{(p)}, \ldots, \hat{u}_{t+s}^{(p)}.$$

Compared to the approach in Section 4.1, the advantage is to avoid the nonlinear optimisation at each step, since the OLS estimators have simple analytical expressions.

If T is large enough, this approach ensures that the estimated roots of lag-polynomials are well located outside the unit circle. The mixed nature of the process is taken into account by this recursive sequence of backward and forward regressions.

Let us now consider an example to see if this algorithmic approach is consistent, or if it inherits the identification problem existing in the Gaussian case.

Example : Let us consider the case $r = s = 1$ and denote $\varphi^{(p)}$ and $\psi^{(p)}$ the p^{th} step values of the parameters. We first compute :

$$\hat{u}_t^{(p)} = y_t - \varphi^{(p)} y_{t-1}, t = 2, \ldots, T,$$

$$\text{and} \quad \hat{v}_t^{(p)} = y_t - \psi^{(p)} y_{t+1}, t = 1, \ldots, T-1.$$

Then we get :

$$\varphi^{(p+1)} = \sum_{t=2}^{T-1} \hat{v}_t^{(p)} \hat{v}_{t-1}^{(p)} / \sum_{t=2}^{T-1} [\hat{v}_{t-1}^{(p)}]^2,$$

$$\psi^{(p+1)} = \sum_{t=2}^{T-1} \hat{u}_t^{(p)} \hat{u}_{t+1}^{(p)} / \sum_{t=2}^{T-1} [\hat{u}_{t+1}^{(p)}]^2.$$

These recursive equations can be rewritten as functions of the sample ACF of the observed process y. Let us denote $\hat{\rho}(1)$ and $\hat{\rho}(2)$ the first and second-order sample autocorrelations, we get :

$$\varphi^{(p+1)} = \frac{-\psi^{(p)} + (1 + \psi^{(p)2})\hat{\rho}(1) - \psi^{(p)}\hat{\rho}(2)}{1 + \psi^{(p)2} - 2\psi^{(p)}\hat{\rho}(1)}, \quad \text{a.1}$$

$$\psi^{(p+1)} = \frac{-\varphi^{(p)} + [1 + \varphi^{(p)2}]\hat{\rho}(1) - \varphi^{(p)}\hat{\rho}(2)}{1 + \varphi^{(p)2} - 2\varphi^{(p)}\hat{\rho}(1)} \quad \text{a.2.}$$

These forward/backward OLS estimators are moment estimators based on the first and second-order sample autocorrelations of process y.

The system of equations (a.1)-(a.2) is symmetric in the sequence of approximations of φ and ψ. This explains why the identification problem is not solved by this backward-forward regression approach. Indeed let us assume T large and starting values $\varphi^{(1)}, \psi^{(1)}$ such that $\varphi^{(\infty)}, \psi^{(\infty)}$ are close to the true values of the parameters φ_0, ψ_0, say. Then by selecting as new starting values $\psi^{(1)}$ for ϕ and $\phi^{(1)}$ for ψ, the algorithm will produce values close to ψ_0 for ϕ and ϕ_0 for ψ. Therefore, even if we are sure that this algorithm provides estimates of φ and ψ inside the unit circle, this algorithm does not always properly assign the true values to the causal and noncausal components.

A.3.2 Pseudo-correlation and pseudo Yule-Walker equations for a mixed model of order $r = s = 1$

The computations of the example in Appendix A.3.1 are valid with the true values of φ and ψ, and for large T. It has been proven in Davis,

Resnick (1986) that the sample autocorrelations $\hat{\rho}(1)$ and $\hat{\rho}(2)$ tend to limits $\rho^*(1), \rho^*(2)$, say. These limits cannot be interpreted as autocorrelations, when the second-order moments of y do not exist. These are pseudo autocorrelations. The last system in the example shows how to compute these pseudo autocorrelations in terms of causal and noncausal autoregressive parameters φ and ψ. They are the analogues of the Yule-Walker equations, but written in the mixed case.

When $r = s = 1$, these equations are :

$$\begin{cases} \varphi &= \dfrac{-\psi + (1+\psi^2)\rho^*(1) - \psi\rho^*(2)}{1 + \psi^2 - 2\psi\rho^*(1)}, \\ \\ \Psi &= \dfrac{-\varphi + (1+\varphi^2)\rho^*(1) - \varphi\rho^*(2)}{1 + \varphi^2 - 2\varphi\rho^*(1)} \end{cases}.$$

We deduce a bivariate linear system in $\rho^*(1), \rho^*(2)$, which can be solved to get the pseudo autocorrelations in terms of the autoregressive parameters. Since the problem is symmetric w.r.t. the causal and noncausal components, the expressions of $\rho^*(1), \rho^*(2)$ are symmetric in φ and ψ. We get :

$$\begin{cases} \psi\rho^*(2) - \rho^*(1)[1 + \psi^2 + 2\varphi\psi] = -\psi - \varphi(1+\psi^2), \\ \\ \varphi\rho^*(2) - \rho^*(1)[1 + \varphi^2 + 2\varphi\psi] = -\varphi - \psi(1+\varphi^2). \end{cases}$$

The determinant of this system is equal to : $(\varphi - \psi)(1 + \varphi\psi)$, and thus this system has a unique solution if $\varphi \neq \psi$. This solution is :

$$\rho^*(1) = \frac{\varphi + \psi}{1 + \varphi\psi}, \rho^*(2) = \frac{\varphi^2 + \psi^2 + 2\varphi\psi + \varphi^2\psi^2}{1 + \varphi\psi}.$$