Filtering, Prediction and Simulation Methods for Noncausal Processes

Christian Gourieroux *
Joann Jasiak †

This version: November, 2014

Abstract

This paper presents the filtering, prediction and simulation in univariate and multivariate noncausal processes. A closed-form functional estimator of the predictive density of noncausal and mixed processes is introduced for the computation of prediction intervals up to a finite horizon H. A state space representation of a noncausal and mixed multivariate VAR process is derived. It is shown that the state space representation can be obtained from either the partial fraction decomposition, or the real Jordan canonical form. A recursive BHHH algorithm for the maximization of the approximate log-likelihood function is proposed, which also calculates the filtered values of the unobserved causal and noncausal components of the process. The new methods are illustrated by a simulation study involving a univariate noncausal process with infinite variance.

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

JEL number: C14, G32, G23.

*University of Toronto and CREST, *e-mail*: gouriero@ensae.fr.
†York University, Canada, *e-mail*: jasiakj@yorku.ca.
The first author gratefully acknowledges 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), Lanne Luoto (2014), Nyberg, Saikkonen (2014)], to the Standard and Poor (S & P) market index [Gourieroux, Zakoian (2013)], the commodity prices and electronic currency exchange rates [Gourieroux, Hencic (2014)]. 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 quite 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, Chan, Ho, Tong (2006), Breidt, Davis (2008)]. Therefore, for identification purposes, the assumption of normality in models with causal and noncausal components has to be relaxed.

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 constrained multivariate specifications].

The objective of this paper is to provide a closed-form functional estimator of the joint predictive density that yields simulation-based prediction intervals at multiple finite horizons in one step. The proposed method is computationally less demanding than the methods focused on the optimal

(point) predictor approximation. Moreover, it is valid for non-Gaussian processes with heavy tails, including processes with infinite error variances.

The closed-form functional estimator of the predictive density of $(y_{T+1}, \ldots, y_{T+H})$ given the information set $(y_1, \ldots, y_T)$ is based on an extension of the look-ahead density estimator, introduced in Glynn, Hendersen (1998), (2001) and eliminates the time consuming simulations involved in the prediction method proposed by Lanne, Luoto, Saikkonen (2012) or estimation of more than 30 additional "parameters" in the Bayesian method of Lanne, Luomo, Luoto (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 the prediction intervals.

In application to multivariate processes, the proposed prediction method requires the availability of a state space representation. We consider a state representation of a noncausal and mixed vector autoregressive (VAR) process, which is derived by using either the partial fraction decomposition, or the real Jordan canonical form, respectively. The latter approach is a modification of the method introduced in Davis, Song (2012) and based on the complex Jordan canonical form. We show that the two proposed methods are equivalent and can accommodate any unconstrained partition of the roots of the autoregressive polynomial matrix inside and outside the unit circle. In this aspect, our approach differs from Lanne, Saikkonen (2013) and Lanne, Luoto (2014), where the autoregressive polynomial matrix is defined a product of a causal and a noncausal components of predetermined orders.

Our prediction method is based on the unobserved component representation of noncausal and mixed processes that distinguishes the latent causal and noncausal components of y_t, and introduced by Lanne, Saikkonen (2011) for univariate processes. The filtered unobserved components can be obtained along with the approximate maximum log-likelihood (AML) estimates from a recursive Berndt, Hall, Hall, Hausman (BHHH) algorithm for maximizing the AML function. The recursive BHHH algorithm involves inversions of matrices of smaller dimension than, for example, the popular Broyden, Fletcher, Goldfarb, Shanno (BFGS) algorithm.

The paper is organized as follows. Section 2 describes the unobserved "causal" and "noncausal" component representation of a univariate autoregressive process. The nonlinear prediction in univariate processes is discussed in Section 3. In Section 4, we introduce the filtered component-based recursive BHHH algorithm for maximizing the AML function. Section 5 extends the filtering and prediction methods to the multivariate framework. A simu-

lation study is presented in Section 6 to analyse the properties of the methods introduced in Sections 3 and 4 in a noncausal process with infinite variance. Section 7 concludes. Some proofs are gathered in Appendices 1 and 2.

2 The process and its unobserved components

This section describes the filtering of a univariate autoregressive process (y_t) and introduces the associated unobserved "causal" and "noncausal" component representation. It also establishes a deterministic dynamic relationship between the unobserved components and the process y_t, and highlights the equivalence of various filtrations.

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

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

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 $MA(\infty)$ representation 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.

2.1 Definition of unobserved components

Let us consider the univariate 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, process (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 causal with respect to a given filtration if its current value belongs to the associated information set. Conversely, we can reverse the time and characterize a process as 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 decomposition.

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 $d^0 b_2 \leq s-1$ [see e.g. Van der Waerden (1953) for the existence and uniqueness of polynomials b_1 and b_2, and Kung, Tong (1977) for fast algorithms for partial fraction decomposition].

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 ε-causal and ε-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 :

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

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 ε-causal and ε-noncausal components v and u, respectively.

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

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

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

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

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

The equality $\dfrac{1}{(1 - \phi L)(L - \psi)} = \dfrac{1}{1 - \phi\psi}\left(\dfrac{\phi}{1 - \phi L} + \dfrac{1}{L - \psi}\right)$, implies:

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

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:

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

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 the errors ε. The proof given in Appendix 1 is based on the partial fraction decomposition and provides a simplified alternative to the proof in Lanne, Luoto, Saikkonen (2012) [1].

[1]Our proof eliminates the linear transformation $z' = BAy'$, where $z = (v_1, ...v_r, \varepsilon_{r+1}, ..., \varepsilon_{T-s}, u_{T-s+1}, ..., u_T)$, $y = (y_1, ..., y_T)$ and B and A are high dimensional matrix functions of parameters Φ and Ψ [see Lanne, Saikkonen (2011) for the expressions of matrices A and B].

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 simulating a stationary mixed or noncausal process. Let us outline below the steps for simulating a mixed autoregressive process of orders $r = s = 1$ defined as:

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

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 :

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

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

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 partial fraction 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 a closed-form functional 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 method can be applied in one step to obtain the forecasts up to a given 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 a closed form functional estimator of the conditional probability density function (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 [2].

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 $E(Y_{T+h}|y_1, \ldots, y_T), h = 1, \ldots, H$

[2]For causal autoregressive processes, whose innovation ε_t have stable symmetric distributions with fat tails, Stuck (1977) [see also Cline, Brockwell (1985)] 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.

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].[3]

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 rather than only a location parameter of that density.

3.1 Predictive density

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 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 :

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

[3]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)].

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:

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

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

3.2 Closed-form estimator 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. So far, it has been considered that "no closed-form solution exists for (such) a marginal distribution and that numerical methods are required" [see Lanne, Luomo, Luoto (2012), Section 4]. In this section we propose a look-ahead-type of estimator for l_s in closed form that eliminates the numerical procedure.

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 by the approximated maximum likelihood method [see e.g. Breidt et al. (1991), Lanne, Saikkonen (2011)], possibly by using the recursive BHHH algorithm 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 $\hat{\Phi}$ [4]. Let us assume $H \geq s$. Then, the conditional density in the numerator of formula (3.2) can be approximated by:

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

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 the value of the joint density of $U_{\tau-s+1}, ..., U_\tau$ at any given date τ and any lagged sequence of values $u^*_{\tau-s+1}, \ldots, u^*_\tau$ is given by:

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

[4] As shown in Example 3, the estimates of g and Ψ are also needed at this stage, even though they do not appear explicitly in formula (3.2).

where the expectation is taken with respect to the joint density of $(U_{\tau+1}, \ldots, U_{\tau+s})$. Let us now approximate this theoretical expectation by its sample-based counterpart. We get:

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

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 obtain:

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

The method 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 (3.5) into the predictive density of interest, the latter one being evaluated at two sequences u of length H, starting at times $\tau = T - s + 1$ and $\tau = T + H - s + 1$, respectively.

Proposition 3 : A functional closed-form estimator of the predictive density is:

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

where $\hat{l}_s$ is defined in equation (3.5). Under standard regularity conditions this functional estimator is consistent [5].

[5] As we focus on the filtering and prediction, the regularity conditions that ensure the uniform convergence of stochastic function $\hat{\Pi}$ on any bounded sets of values $(u_{T+1}, ..., u_{T+H})$ is out of the scope of this paper [see, e.g. Hansen (2004), (2008)].

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 :

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

Expression (3.6) provides a closed-form consistent functional estimator 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 either the simulations of long paths of future $\varepsilon_{T+1}, ..., \varepsilon_{T+M}$ to recover the future vectors $u_{T+1}, ..., u_{T+H}$ [Lanne, Luoto, Saikkonen (2012)], or the estimation of long sequences $\varepsilon_{T+1}, ..., \varepsilon_{T+M}$ as additional unknown "parameters" in a Bayesian framework [Lanne, Luomo, Luoto (2012)]. These two approaches can be compared to our method as follows: Lanne, Luoto, Saikkonen (2012) and Lanne, Luomo, Luoto (2012) approximate numerically the predictive density from a large number of simulations [see, e.g. footnote 11 in Lanne, Luomo, Luoto (2012)] while our method relies on a closed-form formula of the estimated predictive density given in Proposition 3. Moreover, their approximation to the predictive density is based on a truncation $u_T \approx \sum_{j=1}^{M} \beta_j \varepsilon_{T+j}$. If the noncausal polynomial has a root close to unity, the value of M has to be large enough to avoid a significant truncation bias [see, footnote 6 in Lanne, Luomo, Luoto (2012)]. The closed-form formula given in Proposition 3 eliminates such a potential truncation bias.

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 observable 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 Rubin (1988), Geldfand, Smith (1992)], or alternatively a Metropolis-Hasting algorithm. More specifically, the procedure is as follows:

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: $u^s_{T+1}, \ldots, u^s_{T+H}$. In the SIR approach, the first set of simulations is generated from a misspecified instrumental model and the possible bias due to misspecification is eliminated by resampling.
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 it simplifies the expression of the closed-form functional estimator $\hat{\Pi}$.

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 Filtered components-based algorithm

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), and Davis, Song (2012) for the multivariate framework]. 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 can be used in a new algorithm for computing the AML estimates and based on the updating of the filtered values of unobserved components until the numerical convergence of the parameter estimates is achieved.

4.1 The AML estimator as a fixed point

Before introducing the algorithm, let us first describe the fixed point interpretation of the AML estimates. Instead of optimizing the approximate log-likelihood jointly with respect to all parameters θ, Φ, Ψ, the AML estimator is defined as the limit of successive optimizations with respect to each parameter separately.

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

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

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 $g(\theta)$:

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

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 fixed point of the recursive likelihood-based approach provides the AML estimates of the parameters (see the discussion in Section 4.2).

4.2 The recursive BHHH algorithm

The back-forecasting method in Section 4.1 requires more optimizations of the log-likelihood function than the direct maximization of the log-likelihood, although with respect to smaller numbers of parameters. From a numerical point of view, these successive optimizations can be facilitated by using a recursive Berndt, Hall, Hall, Hausman (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 :

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

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

Step 3 : Updating Ψ

We have :

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

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

The above recursive BHHH algorithm converges numerically to the limiting values that are the solutions of the AML likelihood equations.

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 recursive 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. It takes a similar number of iterations to get the numerical convergence, but each iteration involves inversions of matrices of smaller dimensions. By the same argument, the proposed recursive BHHH argument is computationally less demanding than the popular Broyden, Fletcher, Goldfarb, Shanno (BFGS) algorithm.

Moreover, this recursive algorithm yields not only the parameter estimates, but also the filtered components. It 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 extended to mixed autoregressive moving-average processes or mixed vector autoregressive processes.

As noted in Lanne, Saikkonen (2013) [see also Lanne, Luoto (2014)], the state representation involving the causal and noncausal components derived in Section 2 for univariate processes is also valid in the multivariate framework for a constrained VAR model such that

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

where $det(\Pi(z)) \neq 0$ and $det(\Psi(z)) \neq 0$, for $|z| < 1$. However, as noted in Davis, Song (2012), such a decomposition of the matrix polynomial is very restrictive. For example, for a VAR(1) model, it implies that the model is either purely causal, or purely noncausal, as these are the two only cases when the roots of the characteristic equation $det\Phi(z) = 0$ are either all of modulus less than 1, or all of modulus greater than 1. This restriction eliminates the possibility of having a mixed causal - noncausal VAR(1) model with the number of noncausal roots strictly between 1 and n, where n is the number of component series.

In this section, we introduce an unconstrained mixed VAR(1) model and show how the filtering and prediction methods developed in Sections 2.3, 2.4 can be extended to the multivariate framework. First, we derive a state representation of the mixed process from a partial fraction decomposition and next from a real Jordan canonical form [see, Davis, Song (2012)].

5.1 The causal and noncausal components by partial fraction decomposition

Any mixed VAR model 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. Therefore, we consider the following stationary VAR(1) process y_t of dimension n:

$$(Id - \Phi L)y_t = \varepsilon_t, \tag{5.1}$$

$$\det(Id - \Phi L) \quad = \quad \phi(L)\psi(L), \tag{5.2}$$

where (ε_t) is a strong white noise of dimension n, and Φ is a $n \times n$ matrix. The determinant $det(Id - \Phi(L))$ 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 operators, with roots strictly outside and inside the unit circle, respectively, and of the following degrees: $d^0\phi = r, d^0\psi = s, \phi(0) = \psi(0) = 1$.

In order to write the (two-sided) multivariate moving average representation of y, we introduce the adjugate matrix $(Id - \Phi L)^*$ of $(Id - \Phi L)$, that is the transpose of the matrix of cofactors, such that :

$$(Id - \Phi L)(Id - \Phi L)^* = \det(Id - \Phi L)Id. \tag{5.3}$$

The stationary, multivariate two-sided moving average representation of process y is :

$$y_t = \frac{(Id - \Phi L)^*}{\phi(L)\psi(L)}\varepsilon_t, \tag{5.4}$$

where $1/\phi(L)$ [resp. $1/\psi(L)$] denotes the one-sided convergent (scalar) series in L [resp. L^{-1}], which is the inverse of $\phi(L)$ [resp. of $\psi(L)$] [6].

Equation (5.4) provides the two-sided multivariate moving average representation of y_t in terms of the strong white noise ε.

Next, we apply the partial fraction decomposition to each element of matrix $\dfrac{(Id - \Phi L)^*}{\phi(L)\psi(L)}$ and get:

$$\frac{(Id - \Phi L)^*}{\phi(L)\psi(L)} = \frac{B_1(L)}{\phi(L)} + \frac{B_2(L)}{\psi(L)}, \tag{5.5}$$

where $B_1(L)$ [resp. $B_2(L)$] are matrix polynomials of degree $s-1$ [resp. $r-1$]. Hence, the multivariate process Y_t can be written as:

$$Y_t = v_t + u_t, \tag{5.6}$$

[6] Let us recall that $\psi(L) = \prod_{i=1}^{s}(1 - \lambda_i L)$, where the eigenvalues λ_i of Φ are of modulus greater than 1. We have : $\psi(L) = \prod_{i=1}^{s}\lambda_i L^s(-1)^s\prod_{i=1}^{s}(1 - \frac{1}{\lambda_i}L^{-1})$. This polynomial in L is invertible. Its inverse is equal to $1/\psi(L) = (\prod_{i=1}^{s}\lambda_i)^{-1}L^{-s}(-1)^s\prod_{i=1}^{s}(1 - \frac{1}{\lambda_i}L^{-1})^{-1}$, where $(1 - \frac{1}{\lambda_i}L^{-1})^{-1} = \sum_{h=0}^{\infty}\lambda_i^{-h}L^{-h}$. Thus $1/\psi(L)$ is a one-sided series in L^{-1} starting from $h = s$.

where $v_t = \frac{B_1(L)}{\phi(L)}\varepsilon_t$ is a ε-causal process, and $u_t = \frac{B_2(L)}{\psi(L)}$ is a ε-noncausal process. This extends formula 2.7 [up to a scalar and a shift in time of v].

5.2 Another state space representation?

A multivariate dynamic model has a multiplicity of state space representations. Let us now consider the state space representation introduced by Davis and Song (2012) to derive a closed-form expression of the approximate log-likelihood function [Davis, Song (2012), Section 4.1]. The state variables are derived from the Jordan representation of the autoregressive matrix Φ:

$$\Phi = A \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1},$$

where J_1 (resp. J_2) can be written as:

$$J_1 = \begin{pmatrix} \lambda_1 & 0/1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0/1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \lambda_n \end{pmatrix}$$

with the eigenvalues of modulus less than 1 on the diagonal (resp. larger than 1), elements equal to either 0 or 1 on the diagonal first above, and where A is a matrix of change of basis. Davis and Song show that the information in $(Y_1, ..., Y_T)$ is equivalent to the information in $[(\tilde{Y}_{1,1}, \tilde{\varepsilon}_{2,1}), \varepsilon_2, ..., \varepsilon_{T-1}, (\tilde{Y}_{2,T}, \tilde{\varepsilon}_{1,T})$, where $\tilde{Y}_t = A^{-1}Y_t$, (resp. $\tilde{\varepsilon}_t = A^{-1}\varepsilon_t$), $\tilde{Y}_{1,t}$ (resp. $\tilde{Y}_{2,t}$) is the subvector of the first s components of $\tilde{Y}_t$ (resp. the last $n-s$ components of $\tilde{Y}_t$) and $\tilde{\varepsilon}_{1,t}$ (resp. $\tilde{\varepsilon}_{2,t}$) is the subvector of the first s components of $\tilde{\varepsilon}_t$ (resp. the last $n-s$ components of $\tilde{\varepsilon}_t$).

However, this state space representation is inconvenient for simulation and prediction purposes. Indeed, matrices J_1, J_2, A as well as the state vectors $\tilde{Y}_{1,t}$, $\tilde{Y}_{2,t}$ are complex, in general [7].

Below, we propose to derive a real Jordan canonical form, by putting together the vectors of the new basis associated with a complex eigenvalue

[7] The complex valued components are inconvenient for deriving the approximate log-likelihood function. For example, in Davis, Song (2012), that derivation is based on a change of variables by the Jacobian formula applied to complex analytic functions, whereas real transformations are easier to interpret.

and its conjugate [see, e.g. Perko (2001)] in order to derive a real state space representation.

Let us focus on the block of a Jordan matrix corresponding to a complex eigenvalue λ and its conjugate $\bar{\lambda}$. Such a block can be written as :

$$J(\lambda) = \begin{pmatrix} \lambda & 0/1 & 0 & \cdots & 0 \\ 0 & \lambda & 0/1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \bar{\lambda} & 0/1 \\ 0 & \cdots & \cdots & \cdots & \bar{\lambda} \end{pmatrix},$$

where the first m elements of the diagonal above the main diagonal are equal to zero, and the next $M - m$ elements are equal to one. M is the algebraic multiplicity order of the eigenvalue and m its the geometric multiplicity order, that is the dimension of the null space of $\Phi - \lambda Id$. Let us denote by e_i, $i = 1, ..., M$ the complex vectors of the change of basis associated with λ. By construction, they are such that:

$$\begin{cases} \Phi e_i &= \lambda e_i, & i = 1, ..., m, \\ \Phi e_i &= \lambda e_i + e_{i-1}, & i = m+1, ..., M. \end{cases}$$

As matrix Φ has real elements, we find that:

$$\begin{cases} \Phi \bar{e}_i &= \bar{\lambda} \bar{e}_i, & i = 1, ..., m, \\ \Phi \bar{e}_i &= \bar{\lambda} \bar{e}_i + \bar{e}_{i-1}, & i = m+1, ..., M. \end{cases}$$

For deriving the real Jordan canonical form, we select $\bar{e}_i, i = 1, ..., M$ as the complex vectors of the new basis associated with eigenvalue $\bar{\lambda}$.

Then, instead of using the complex basis $(e_1, ..., e_M, \bar{e}_1, ..., \bar{e}_M)$,we use the real basis $[Re(e_i), Im(e_i), i = 1, ..., M]$, where $Re(e_i)$ [resp. $Im(e_i)$] is the vector whose components are the real (resp. imaginary) parts of the components of e_i. Let us now write matrix Φ restricted to this new basis. We get:

$$\begin{aligned} \Phi Re(e_i) &= \frac{1}{2}\Phi(e_i + \bar{e}_i) \\ &= \frac{1}{2}(\lambda e_i + \bar{\lambda}\bar{e}_i) \\ &= \frac{1}{2}\{\lambda[Re(e_i) + iIm(e_i)] + \bar{\lambda}[Re(e_i) - iIm(e_i)]\} \end{aligned}$$

$$
\begin{aligned}
&= Re(e_i)\frac{1}{2}(\lambda+\bar{\lambda}) + Im(e_i)\frac{i}{2}(\lambda-\bar{\lambda}) \\
&= Re(e_i)Re(\lambda) - Im(e_i)Im(\lambda), \text{ if } 1 \leq i \leq m.
\end{aligned}
$$

Similarly, we have:

$$
\begin{aligned}
\Phi Re(e_i) &= Re(e_i)Re(\lambda) - Im(e_i)Im(\lambda) + Re(e_{i-1}), \text{ if } m+1 \leq i \leq M, \\
\Phi Im(e_i) &= Re(e_i)Im(\lambda) + Im(e_i)Re(\lambda), \text{ if } 1 \leq i \leq m, \\
\Phi Im(e_i) &= Re(e_i)Im(\lambda) + Im(e_i)Re(\lambda) + Im(e_{i-1}), \text{ if } m+1 \leq i \leq M,
\end{aligned}
$$

By using the new basis of real vectors $[Re(e_i), Im(e_i), i = 1, ..., m]$, the initial complex Jordan matrix $J(\lambda)$ can be replaced by the following real Jordan matrix $RJ(\lambda)$ given by:

$$
RJ(\lambda) = \begin{pmatrix}
R(a,b) & 0_2 & 0_2 & \cdots & 0_2 \\
0 & R(a,b) & 0_2 & \cdots & 0_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \vdots & R(a,b) & Id_2 \\
0_2 & \cdots & \cdots & \cdots & R(a,b)
\end{pmatrix}
$$

with the following 2×2 matrix on the main block-diagonal

$$
R(a,b) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \text{ where } \lambda = a + ib,
$$

either the 2×2 matrix of zeros, denoted by 0_2, or identity matrix Id_2, on the block diagonal first above and blocks of zeros everywhere else.

The subsection below shows that a real state space representation, similar to the one in Davis, Song (2012), can be obtained by replacing the complex Jordan forms by their real Jordan counterparts and also in the derivation of the approximate log-likelihood as $detRJ(\lambda) = detJ(\lambda)$. In practice, the true autoregressive matrix Φ is estimated by $\hat{\Phi}$ and the probability that $\hat{\Phi}$ has a double root is equal to zero. In this case $RJ(\lambda)$ is reduced to $R(a,b)$.

5.3 Relationship between the real Jordan canonical form and the partial fraction decomposition

The two approaches described in Sections 5.1 and 5.2 are closely related. Let us denote by $\Phi = AJA^{-1}$ the real canonical form of Φ where

$$J = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix},$$

and J_1 (resp. J_2) contains all real blocks corresponding to the eigenvalues of modulus less than 1 (resp. greater than 1).

Let us decompose accordingly $A = (A_1, A_2)$ and $A^{-1} = (A^{1'}, A^{2'})'$. We obtain the following result:

Proposition 4 :
The operator $Id - \Phi L$ is invertible and its inverse is given by:

$$(Id - \Phi L)^{-1} = \frac{A_1(Id - J_1L)^* A^{1'}}{det(Id - J_1L)} + \frac{A_2(Id - J_2L)^* A^{2'}}{det(Id - J_2L)}$$

where * denotes the adjugate matrix.

Proof: see Appendix 2.

The elements of $(Id - J_1L)^*$ are polynomials of a degree less or equal to the degree of $det(Id - J_1L)$ less one. The equality in Proposition 5 is simply the multivariate partial fraction decomposition of:

$$(Id - \Phi L)^{-1} = \frac{(Id - \Phi L)^*}{det(Id - \Phi L)} = \frac{(Id - \Phi L)^*}{det(Id - J_1L)det(Id - J_2L)}.$$

Thus the multivariate partial fraction decomposition of $(Id - \Phi L)^{-1}$ follows from the real Jordan canonical form.

Note that the real Jordan canonical form is not unique, but all forms lead to the same partial fraction decomposition. Proposition 4 can be used to show the link between the state variables u and v derived in Section 5.1 and the state variables $\tilde{Y}_1, \tilde{Y}_2$.

Corollary : We have $v_t = A_1\tilde{Y}_{1,t}, u_t = A_2\tilde{Y}_{2,t}$.

Consequently, all the methods derived in the univariate framework are also valid for multivariate processes, especially the prediction method based on a consistent functional estimator of the joint predictive density, outlined in Section 3.

6 Simulation Study

To illustrate the implementation of the filtering, prediction and simulation methods, we perform a Monte Carlo study based on the data generating process described below. It is intended 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 followed by a sudden drop, and can be replicated in practice by assuming that the errors of the noncausal model have fat tails [see Gourieroux, Zakoian (2013)]. We assume that the errors are Cauchy distributed in order to get both fat tails of the marginal densities of simulated processes and a closed-form of the approximate likelihood function.

6.1 The Data Generating Process

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

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

with Cauchy errors :

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

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

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

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

[Insert Figure 1 : Simulated paths with four sets of ϕ and ψ]

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 AML estimates of parameters ϕ, ψ, σ are computed by the standard BFGS algorithm and by the recursive BHHH algorithm of Section 4.2. The estimated values of the parameters are given in Table 1. As the limiting distribution of the AML estimators of ψ and ϕ [8] is intractable due to Cauchy errors [see, Andrews, Calder, Davis (2009), Th. 3.2], the standard errors are computed by bootstrap [Davis, Wu (1997), Andrews, Calder, Davis (2009) Th. 3.4]. The values of the t-ratios are also given as measures of the relative accuracies of the estimators.

[8] The AML estimator of σ is asymptotically normal.

Table 1: AML Estimates of ϕ, ψ and σ

	Parameter	standard error	t-ratio
	standard BFGS		
$\psi = 0.0$	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
	recursive 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 results from the two algorithms are close. The accuracy of the estimator of parameter ϕ is almost independent of the value of parameter ψ, which controls the length and size of the bubble. The variance of $\hat{\psi}$ 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 the trajec-

tories of filtered errors ε and components u and v based on the DGP with $\phi = 0.3, \psi = 0.9$.

[Insert Figure 2 : Filtered component series from process with $\phi = 0.3, \psi = 0.9$]

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:

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

The unobserved components u and v 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 the strong persistence in reverse time of the ε-noncausal process u. The sample autocorrelations of u are statistically significant up to lag 10 and take the following values:

Table 2: Sample autocorrelations of u

lag	1	2	3	4	5	6	7	8	9	10
acf	0.92	0.83	0.75	0.67	0.58	0.52	0.45	0.39	0.31	0.24

As the errors have fat tails, the asymptotic distribution of the sample autocorrelations differs from the standard Gaussian distribution. The confidence bounds for testing the significance are obtained from the simulated limiting distribution given in Davis, Resnick (1986). The adjusted upper bound is 0.23 and is greater than the normality-based upper bound of 0.14.

6.4 Predictive densities

a) Short-term prediction

The predictive density of the ε-noncausal component u of the process with parameters $\phi = 0.3, \psi = 0.9$ is estimated by the formula given in Example 4, Section 3.1. That density is shifted by a constant equal to ϕy_T to become the predictive density of process y. Let us first consider predictions of future

values of that process at short horizons $H = 1, 2$. 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 $T = 200$ with parameters $\phi = 0.3, \psi = 0.9$ and its last value of filtered component u. The predictive density at horizon 1 of process (y_t) with parameters $\phi = 0.3, \psi = 0.9$ is plotted in Figure 3 below:

[Insert Figure 3: 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.

It is also possible to display the joint predictive density at horizon 2 for y_{T+1}, y_{T+2} conditioned on $y_T = 16.67, u_T = 12.39$. That bivariate density is shown in Figure 4.

[Insert Figure 4: Contour plot of joint predictive density at horizon 2, $y_T = 16.67, u_T = 12.39$]

The joint predictive density is peaked and resembles a density 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 determined by linear dynamic equations.

b) Future pattern recognition

The joint predictive density at horizon 2 can be used to study the likelihood of various future dynamics of process y with parameters $\phi = 0.3, \psi = 0.9$. 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 origin is 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 of what magnitude. The various possible future scenarios at horizon 2 are displayed in Figure 5.

[Insert Figure 5 : Predicted patterns.]

The plane in y_{T+1} and y_{T+2} with the support of the bivariate density plotted in Figure 4, is divided into eight semi-orthants centered at $y_T = 16.67, y_{T-1} = 14.27$. 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 $y_T = 16.67, y_{T-1} = 14.27$, which are shown in Table 3.

Table 3: 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 3 the probability of y returning to the value prior to the two consecutive increases. This probability is small and equal to 0.008, hence the downturn cannot be sharp.

c) Medium-term prediction

When the prediction horizon H is equal or larger than 3, the closed form expression of the joint estimated predictive density (3.3) cannot be displayed graphically. However, the closed form of the estimator of the predictive density can be used jointly with the sampling-importance-resampling (SIR) method to simulate future paths.

To apply the SIR, we exploit the recursive relationship between the y_t and the u_t and simulate a set of future paths of u. Given that $u_t - \psi u_{t+1} = \varepsilon_t$, 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:

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

where $\tilde{\varepsilon}_t \sim IIN(0,1)$ 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 instrumental 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 corrects for misspecification of the instrumental model [9].

Figures 6 and 7 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 6: Term structure of predictive density]

[Insert Figure 7: Term structure of prediction interval]

The increase of the width of the prediction interval with respect to the horizon is not surprising. However, it is interesting to compare the pattern of the upper bound with the pattern of the upper bound of a causal process with a unit root (ψ is close to 1) and errors with zero mean and finite variance, usually proposed to capture permanent explosive behavior. Such a standard upper bound changes at the rate of $\sqrt{H}$ and is a concave function of time [10]. In our framework, there are transitory explosive behaviors and 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 and simulation methods for mixed causal/noncausal autoregressive processes. We derived a closed-form estimator of joint predictive densities of future values of the process at multiple horizons, which can be used to obtain prediction intervals for future values of the processes, in contrast to recent literature on noncausal processes that has been focused on point forecasts.

[9] The sample sizes for the sampling and resampling are $S = 2000$ and $S^* = 5000$, respectively.

[10] In the case of a unit root model that may be encountered in financial applications, the prediction error will be $\sum_{h=1}^{H} \varepsilon_{t+h}$ that is such that $1/\sqrt{H}\sum_{h=1}^{H} \varepsilon_{t+h} \sim N(0, \sigma^2)$ with $\sigma^2 = var\varepsilon_t$.

The proposed univariate methods of filtering, prediction and simulation were extended to the multivariate framework. We derived a state space representation of the mixed VAR process. We showed that the derivation itself can be based on either the partial fraction decomposition or the real Jordan canonical form. The state space representation allows us to simulate multivariate mixed casaul/noncausal process, which can next be used in simulation-based estimation methods, such as the indirect inference or the simulated maximum likelihood. The simulations can also serve to study the finite sample distributional properties of the AML estimators by bootstrap.

References

[1] Andrews, B., Calder, M. and R. Davis (2009): " Maximum Likelihood Estimation for α-Stable Autoregressive Processes", Annals of Statistics, 37, 1946-1982.

[2] 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.

[3] 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.

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

[5] Breidt, F. and R. Davis (2008): Time Reversibility, Identifiability and Independence of Innovations for Stationary Time Series", Journal of Time Series Analysis, 13, 377-390.

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

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

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

[9] Chan, K., Ho, L. and H. Tong (2006): "A Note on Time Reversibility of Multivariate Linear Processes", Biometrika, 93, 221-227.

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

[11] Cline, D. and P. Brockwell (1985): "Linear Prediction of ARMA Processes with Infinite Variances", Stochastic Processes and Their Applications, 19, 281-296.

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

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

[14] Davis, R. and A. Wu (1997): "Bootstrapping M-estimates in Regression and Autoregression with Infinite Variances", Statistica Sinica, 7, 1135-1154.

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

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

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

[18] 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.

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

[20] Gourieroux, C. and A. Hencic (2014): "Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate" in "Econometrics of Risk", Series: Studies in Computational Intelligence, Springer

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

[22] Hansen, B. (2008): "Uniform Convergence Rates for Kernel Estimators with Dependent Data", Econometric Theory, 24, 726-748.

[23] Hansen, B. (2004): "Uniform Convergence Rates for Nonparametric Estimators", D.P. University of Wisconsin.

[24] Kung, H. and D. Tong (1977): "Fast Algorithms for Partial Fraction Decomposition", SIAM J. COMPUT., 6, 582-593.

[25] Lanne, M. and J. Luoto (2014): "Noncausal Bayesian Vector Autoregression", CREATES Research Paper 2014-7.

[26] Lanne, M., Luoma, A. and J. Luoto (2012): " Bayesian Model selection and Forecasting in Noncausal Autoregressive Models", Journal of Applied Econometrics, 27, 812-830.

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

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

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

[30] Nyberg, H. and P. Saikkonen (2014): "Forecasting with a Noncausal VAR Model", Computational Statistics and Data Analysis, 76, 536-555.

[31] Perko, L. (2001): " Differential Equations and Dynamical Systems", Springer.

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

[33] 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.

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

[35] Van der Waerden, B. (1953): "Modern Algebra", Vol. 1, Frederic Ungar, New-York.

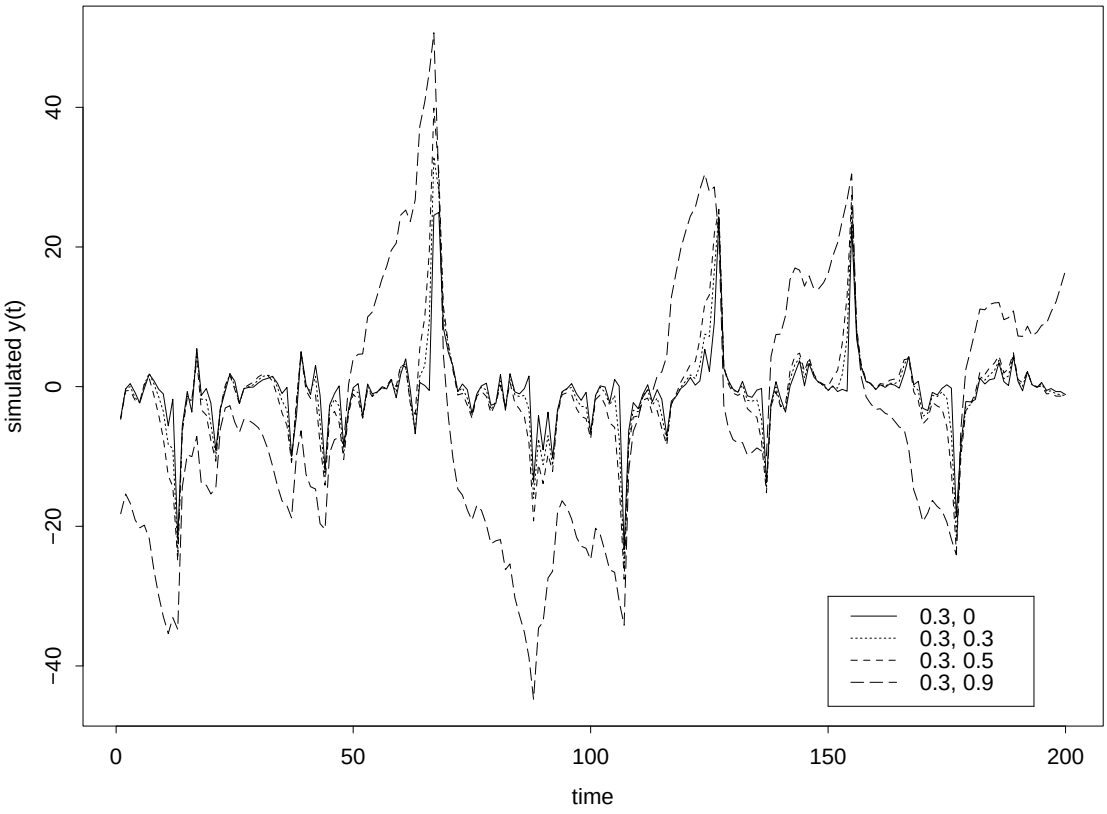

Figure 1: Simulated paths with four sets of ϕ and ψ

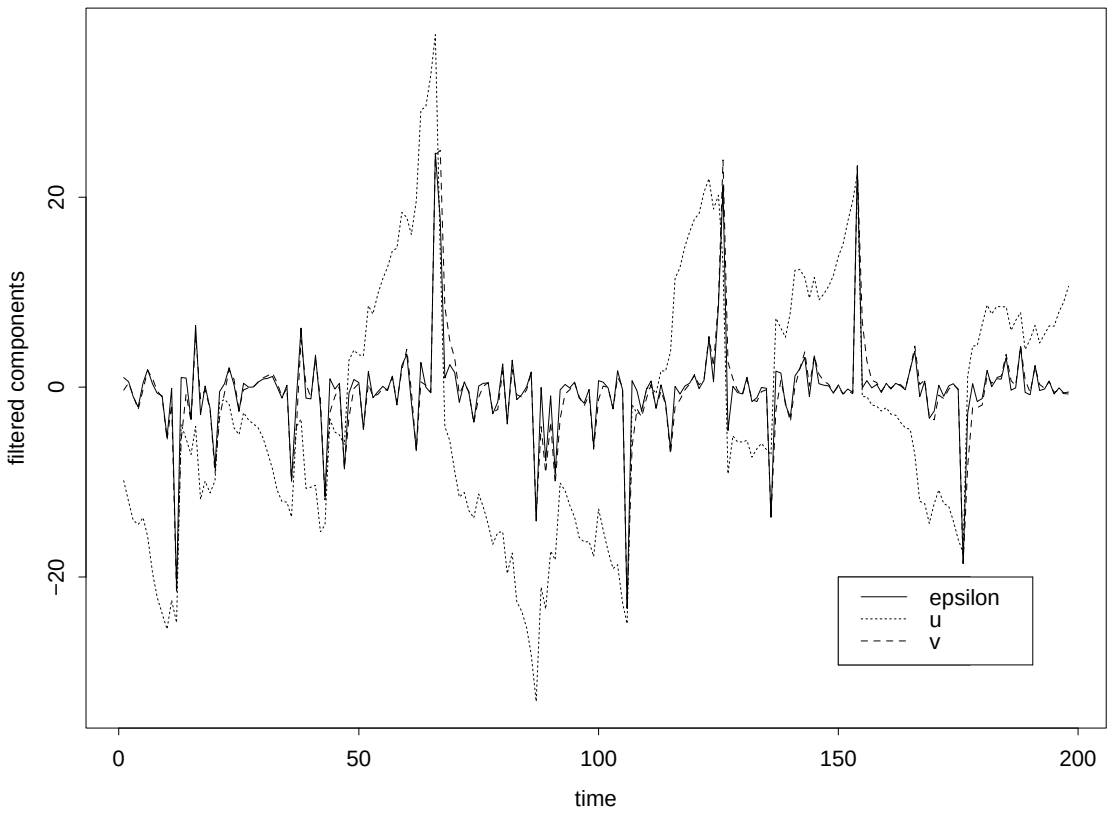

Figure 2: Filtered component series from process with $\phi = 0.3, \psi = 0.9$

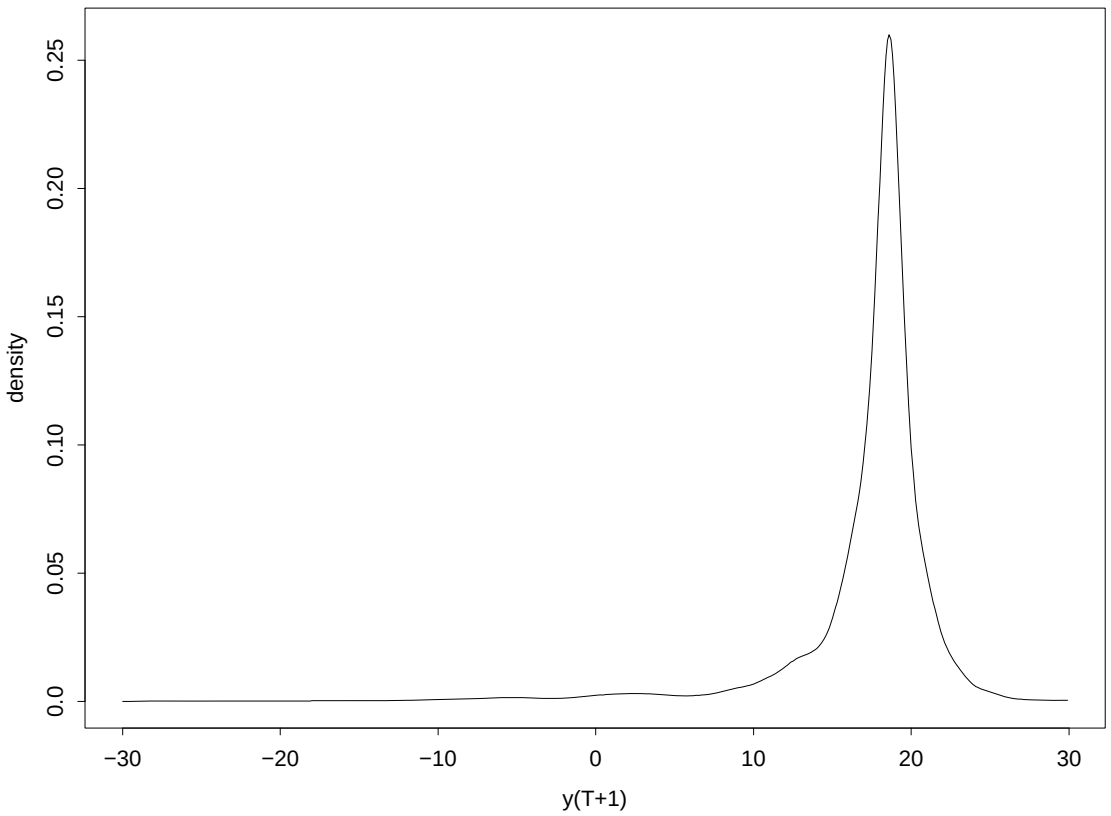

Figure 3: Predictive density at horizon 1, $y_T = 16.67, u_T = 12.39$

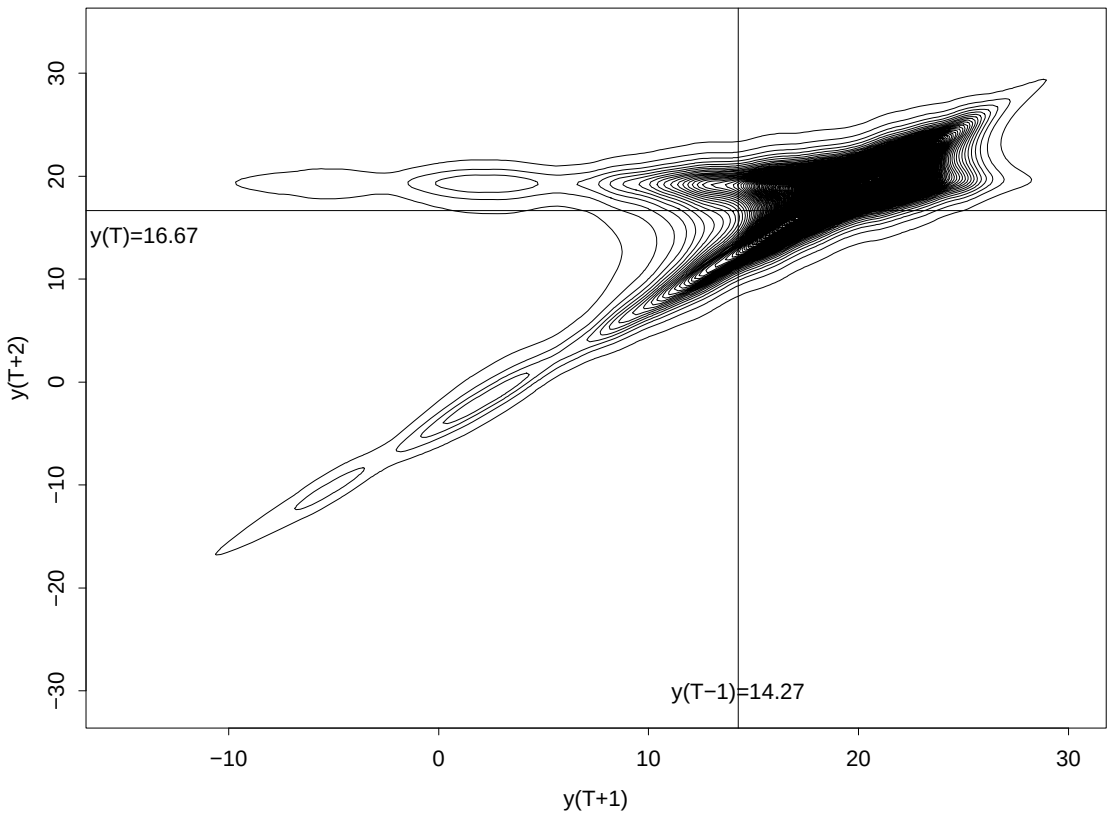

Figure 4: Contour plot of joint predictive density at horizon 2, $y_T = 16.67, u_T = 12.39$

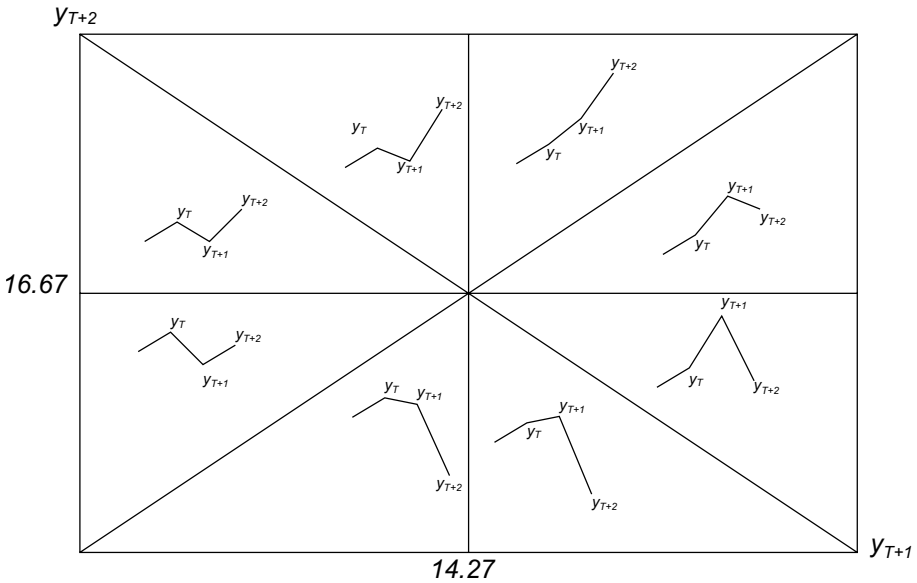

Figure 5: Predicted patterns

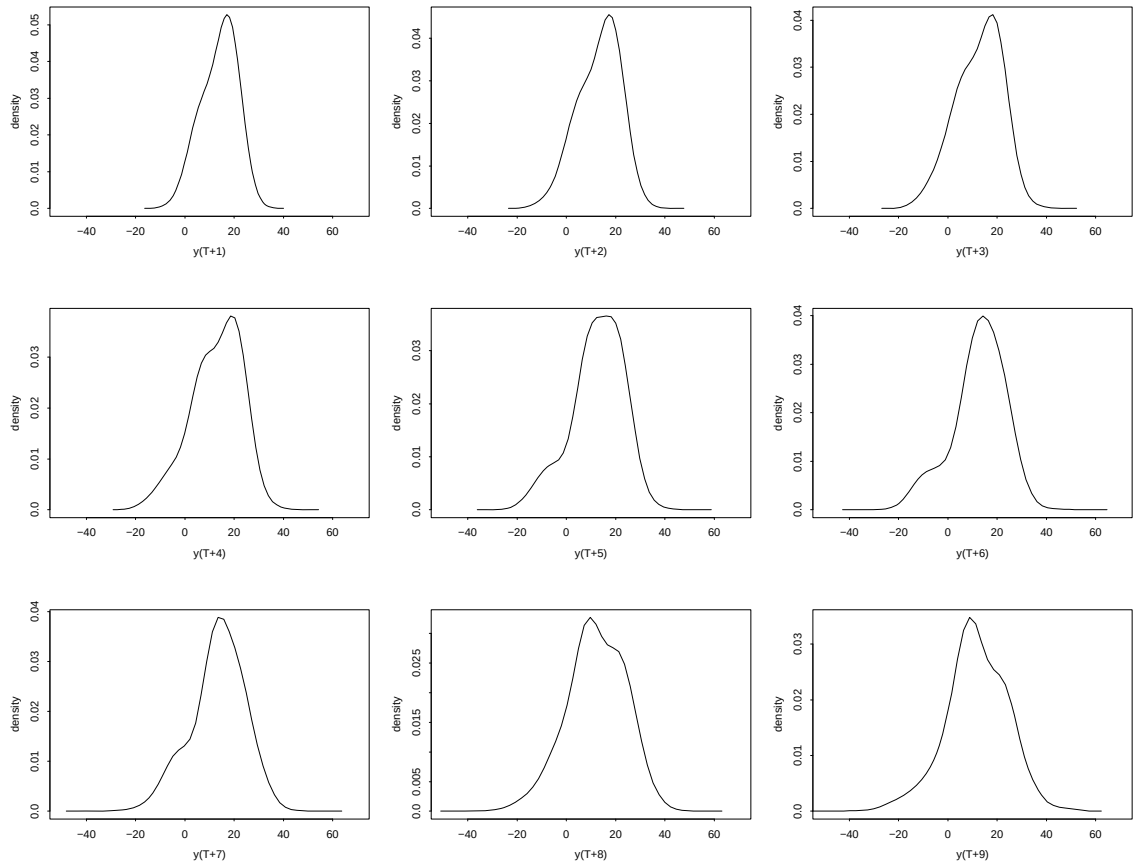

Figure 6: Term structure of predictive density

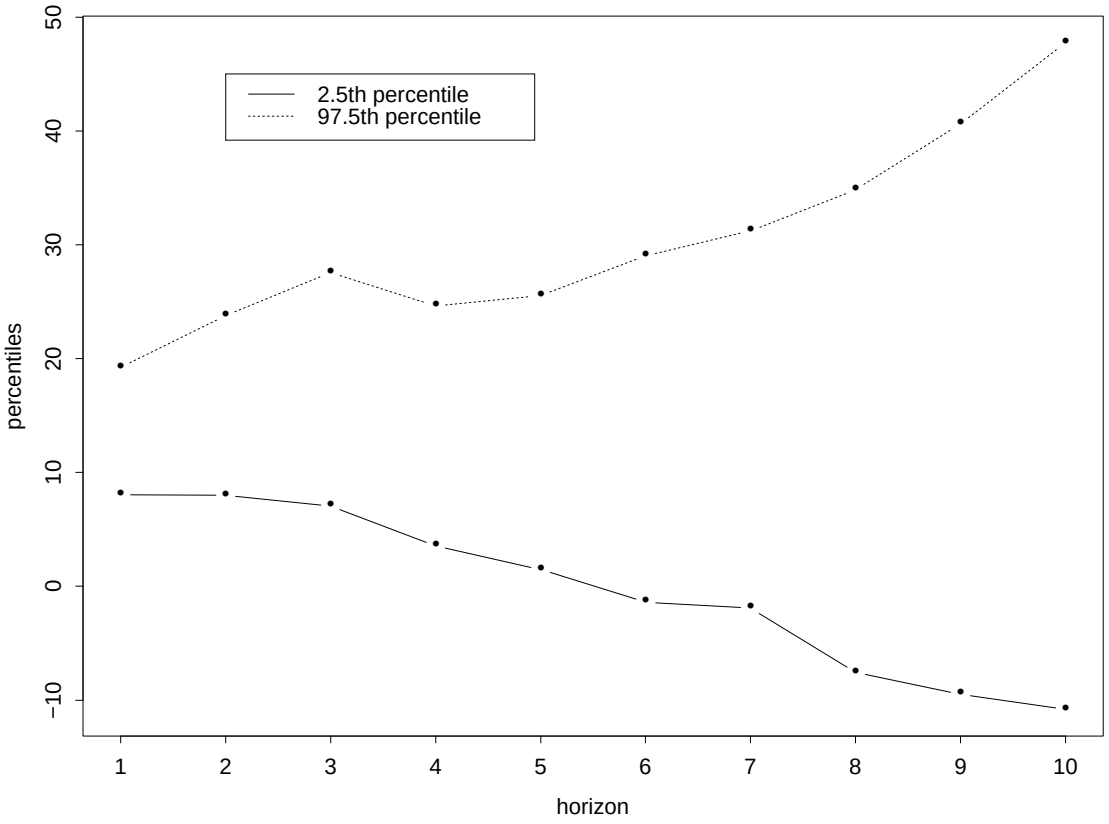

Figure 7: Term structure of prediction interval

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

Proof of Proposition 4

Let us follow Davis, Song (2012) but use the real Jordan canonical form instead of the complex Jordan form. The VAR(1) model can be written as:

$$\begin{aligned} Y_t &= \Phi Y_{t-1} + \epsilon_t \\ &\iff A^{-1}Y_t = JA^{-1}Y_{t-1} + A^{-1}\epsilon_t \\ &\iff \tilde{Y}_t = J\tilde{Y}_{t-1} + \tilde{\epsilon}_t, \end{aligned}$$

where $\tilde{Y}_t = A^{-1}Y_t, \tilde{\epsilon}_t = A^{-1}\epsilon_t$.
This is equivalent to the ϵ-causal and ϵ-noncausal systems:

$$\begin{aligned} \tilde{Y}_{1,t} &= J_1\tilde{Y}_{1,t-1} + \tilde{\epsilon}_{1,t} \\ \tilde{Y}_{2,t} &= J_2\tilde{Y}_{2,t-1} + \tilde{\epsilon}_{2,t} \iff \tilde{Y}_{2,t} = J_2^{-1}\tilde{Y}_{2,t+1} - J_2^{-1}\tilde{\epsilon}_{2,t+1}. \end{aligned}$$

It follows that:

$$\begin{aligned}\tilde{Y}_{1,t} &= (Id - J_1 L)^{-1}\tilde{\epsilon}_{1,t},\\ \tilde{Y}_{2,t} &= -(Id - J_2^{-1}L^{-1})^{-1}J_2^{-1}L^{-1}\tilde{\epsilon}_{2,t},\end{aligned}$$

or formally $\tilde{Y}_{2,t} = (Id - J_2 L)^{-1}\tilde{\epsilon}_{2,t}$. Note that the equality for $\tilde{Y}_{2,t}$ proves the invertibility of the operator $Id - J_2 L$ and provides the expression of its inverse as a series in L^{-1}, as follows:

$$(Id - J_2 L)^{-1} = -\sum_{h=1}^{\infty} J_2^{-h} L^{-h}.$$

Then, the above system is equivalent to:

$$\begin{aligned}\tilde{Y}_{1,t} &= (Id - J_1 L)^{-1} A^{1'}\epsilon_t,\\ \tilde{Y}_{1,t} &= (Id - J_2 L)^{-1} A^{2'}\epsilon_t.\end{aligned}$$

It implies that:

$$\begin{aligned}Y_t &= A_1\tilde{Y}_{1,t} + A_2\tilde{Y}_{2,t}\\ &= A_1(Id - J_1 L)^{-1}A^{1'}\epsilon_t + A_2(Id - J_2 L)^{-1}A^{2'}\epsilon_t\\ &= \left[\frac{A_1(Id - J_1 L)^{*}A^{1'}}{det(Id - J_1 L)} + \frac{A_2(Id - J_2 L)^{*}A^{2'}}{det(Id - J_2 L)}\right]\epsilon_t.\end{aligned}$$

The result follows directly.