Centre de recherche en économie et statistique

Série des Documents de Travail

n° 2015-02

Semi-Parametric Estimation of Noncausal Vector Autoregression

C.GOURIEROUX[1]
J.JASIAK[2]

[1] CREST and University of Toronto

[2] York University

Semi-Parametric Estimation of Noncausal Vector Autoregression

C., GOURIEROUX [1] and J., JASIAK [2]

May 2015

[1]CREST and University of Toronto.
[2]York University

Semi-Parametric Estimation of Noncausal Vector Autoregression

Abstract

This paper introduces consistent semi-parametric estimation methods for mixed causal/noncausal multivariate non-Gaussian processes.

We show that in the VAR(1) model, the second-order identification is feasible to some limited extent, contrary to the common belief that non-Gaussian processes are not second-order identifiable. In general, in the mixed VAR(1) it is possible to distinguish the mixed processes with different numbers of causal and noncausal components. For detecting the causal and noncausal components, a semi-parametric exploratory analysis is proposed. It includes a semi-parametric estimation method that does not require any distributional assumptions on the errors. For direct estimation of the matrix of autoregressive coefficients of a VAR(1), we use the generalized covariance estimator. Although this estimator is not fully efficient, it provides the estimates in one single optimization while the MLE requires a number of optimizations, which is equal to the number of all possible causal dimensions. The methods are illustrated by a simulation study.

Keywords : Multivariate Noncausal Process, Identification, Semi-Parametric Estimation, Speculative Bubble.

1 Introduction

The analysis of stationary linear time series with a two-sided moving average representation involving independent and identically distributed (i.i.d.) errors can be found in the early classical time series literature [see e.g. Hannan (1973)]. A particular characteristic of non-Gaussian stationary linear time series is that its two-sided moving average representation that includes the present, past and future shocks can be distinguished from a one-sided moving average representation involving the current and lagged values only. In Gaussian processes, those two representations cannot be distinguished and the future shocks that determine the impulse response functions cannot be identified. Nevertheless, the Gaussian processes have been given particular attention in the time series literature, especially since the mid-seventies, in the framework of the Box-Jenkins method of analysis. That methodology. offers a straighforward approach to the identification and estimation of time series from the moments of order up to two only, allowing to accommodate the past dependence in Gaussian processes but leaving out of the scope the noncausal and nonlinear effects. Accordingly, the class of VARMA (ARMA) models that has remained the benchmark specification for linear processes over the last 30 years, can capture only the causal effets of the past on the present and disregards any noncausal effects of the future on the present value. Even if the data are non-Gaussian, the noncausal effects are left out as the Box-Jenkins method relies on the Gaussian pseudo-maximum likelihood estimation and suffers from the aforementioned lack of identification. Consequently, the causal and noncausal dynamics cannot be distinguished, leading the researcher to the erroneous conclusion about the absence of noncausal effects. Moreover, given that the Gaussian pseudo-maximum likelihood estimators may be inconsistent when applied to non-Gaussian data, the inference based on those estimators, including the impulse response analysis, may be unreliable [see, Gourieroux, Monfort (2015)].

Recently, due to the wide-spread use of non-Gaussian processes in Economics and Finance, the interest in noncausal effects in stationary linear time series has grown considerably. Empirical evidence from applications involving multivariate economic and financial time series with non-Gaussian distributions suggests that estimation of causal VAR models by the Gaussian maximum likelihood (ML) in line with the Box-Jenkins procedure, leads to mispecifications. For example, an estimated autoregressive matrix of a non-Gaussian VAR(1) may have eigenvalues of modulus significantly larger than

1 [see e.g. Lanne, Saikkonen (2013) for such estimations, and Gourieroux, Jasiak (2014)b for the analysis of causal misspecification]. Another empirically established fact is the presence of nonlinearities in financial data. In the time series literature, there exist theoretical results that point to the equivalence of noncausal linear dynamics with causal nonlinear patterns [see Rosenblatt (2000)]. Accordingly, a linear noncausal model can represent a possibly complex nonlinear dynamics observed from the calendar time perspectives. In particular, Gourieroux, Zakoian (2014) have shown that a noncausal linear autoregression can be used to model speculative bubbles in processes with fat tails.

There are also valid economic arguments that suggest the presence of noncausal components in the strong structural economic VARMA models. For example, noncausal components can result from a leading effect of a fiscal policy [Leeper, Walker, Yang (2013)], or from the rational expectations introduced in multivariate dynamic models [see e.g. Gourieroux, Monfort (2015a) for general discussions].

The estimation method that has been used in recent literature on mixed causal/noncausal processes is the approximate maximum likelihood (ML), which assumes a parametric specification of the error distribution [3]. The ML method is consistent, provided that the error distribution is properly specified, but generally inconsistent, otherwise. The objective of this paper is to improve the methodology in this respect and to introduce semi-parametric estimation methods for the mixed causal/noncausal dynamics. For expository purpose, the paper is focused on the mixed VAR(1) model, which is the benchmark model in the literature on multivariate linear processes.

We show that second-order identification, that is, the identification of the process from moments up to order two, which include the auto- and cross-covariances, is feasible to some limited extent in the VAR(1) model, contrary to the wide-spread belief in the lack of second-order identification of non-Gaussian linear processes. We claim that, in general, it is possible to distinguish the mixed processes with different values of n_1, n_2, where n_1 and n_2 denote the causal and the noncausal dimensions, respectively, and we propose a semi-parametric exploratory analysis for detecting the causal and noncausal dimensions of the VAR process.

[3] See Breidt et al. (1991), Rosenblatt (2000), Chapter 8, for the asymptotic properties of the ML estimators in the univariate case, Davis, Song (2012), Lanne, Saikkonen (2013), for the multivariate case.

For direct estimation of the autoregressive matrix of coefficients, we propose a generalized covariance estimator. Although this estimator is not fully efficient, it provides the estimates in one single optimization while the ML requires multiple optimizations. More precisely, when the distribution of the errors is parametric, the lack of full efficiency of the generalized covariance estimator is compensated by its numerical simplicity. The ML estimator requires $n+1$ optimizations, which is equal to the number of all possible causal dimensions of the process, while the generalized covariance estimator is obtained in a single optimization only.

The paper is organized as follows. In Section 2, we derive and discuss the causal and noncausal components of a mixed VAR(1) process, along with their second-order properties. Section 3 covers the second-order identification of the mixed VAR(1) process. Section 4 introduces a semi-parametric method of exploratory analysis for estimation of the causal and noncausal dimensions and of the associated causal and noncausal components of a mixed causal/noncausal process. Section 5 develops the semi-parametric estimation method from the covariance-based moment conditions. An application to a set of simulated data is presented in Section 6 to show how to implement the exploratory analysis and the covariance estimator in practice. Section 7 concludes. Proofs are gathered in Appendices.

2 The mixed Vector Autoregressive Process of order 1

Let us consider a mixed VAR(1) process, that is a n-dimensional strictly stationary process (Y_t) satisfying the recursive system :

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \tag{2.1}$$

where (ε_t) is a sequence of i.i.d. n-dimensional random vectors (i.e. a strong white noise), and Φ a (n,n) matrix. We assume that :

Assumption A.1 : ε_t is square integrable with zero mean $E(\varepsilon_t) = 0$, and variance-covariance matrix $V(\varepsilon_t) = \Sigma$.

Assumption A.2 : The eigenvalues of matrix Φ are of modulus different from 1.

Below, it is shown that Assumption A.2 ensures the existence and uniqueness of a stationary solution to recursive equation (2.1). Moreover this solution admits a two-sided moving average representation.
As ε_t is not assumed to be independent of the lagged values of the process $Y_{t-1}, Y_{t-2}, \ldots$, system (2.1) is not necessarily "causal".

2.1 Causal and noncausal components of Y

To prove the existence of a stationary solution of (2.1) and derive the two-sided moving average representation of process Y, we first find its representation in terms of the causal and noncausal components. We follow the approach introduced in Davis, Song (2012) and Gourieroux, Jasiak (2014)a.

Proposition 1 : Let us denote n_1 (resp. $n_2 = n - n_1$) the number of eigenvalues of Φ with modulus strictly smaller than 1 (resp. strictly larger than 1). There exist an invertible (n, n) matrix A, and two square matrices J_1 with dimension (n_1, n_1), J_2 with dimension (n_2, n_2) with all eigenvalues of J_1 (resp. J_2) with their modulus strictly smaller than 1 (resp. larger than 1) such that :

$$
\begin{aligned}
Y_t &= A_1 Y_{1,t}^* + A_2 Y_{2,t}^*, && (2.2)\\
Y_{1,t}^* &= J_1 Y_{1,t-1}^* + \varepsilon_{1,t}^*, Y_{2,t}^* = J_2 Y_{2,t-1}^* + \varepsilon_{2,t}^*, && (2.3)\\
\varepsilon_{1,t}^* &= A^1 \varepsilon_t, \varepsilon_{2,t}^* = A^2 \varepsilon_t, && (2.4)
\end{aligned}
$$

where A_1, A_2 (resp A^1, A^2) are the blocks in the decomposition of matrix A as :

$A = (A_1, A_2)$ [resp. in the decomposition of A^{-1} as $A^{-1} = \begin{pmatrix} A^1 \\ A^2 \end{pmatrix}$].

Proof : It is always possible to decompose matrix Φ as :

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

by considering its real Jordan canonical form and gathering in J_1 (resp. J_2) all the subblocks associated with the eigenvalues with modulus smaller than 1 (resp. larger than 1) [see Gourieroux, Jasiak (2014)a]. Let us premultiply both sides of equation (2.1) by matrix A^{-1} and denote :

$Y_t^* = \begin{pmatrix} Y_{1,t}^* \\ Y_{2,t}^* \end{pmatrix} \equiv A^{-1}Y_t, \varepsilon_t^* = \begin{pmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{pmatrix} \equiv A^{-1}\varepsilon_t$. We get :

$$Y_t^* = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} Y_{t-1}^* + \varepsilon_t^*,$$

or equivalently :

$$Y_{j,t}^* = J_j Y_{j,t-1}^* + \varepsilon_{j,t}^*, j = 1, 2,$$

that are the recursive equations (2.3).

Moreover, equation $Y_t = AY_t^*$, is equivalent to :

$$Y_t = A_1 Y_{1,t}^* + A_2 Y_{2,t}^*,$$

that is the decomposition in equation (2.2). This proves Proposition 1.

QED

Let us examine the recursive equations (2.3). Since all eigenvalues of J_1 are of modulus strictly less than 1, the recursive equation :

$$Y_{1,t}^* = J_1 Y_{1,t-1}^* + \varepsilon_{1,t}^*,$$

is causal and, by recursive backward substitutions, we derive the causal one-sided moving average representation of $Y_{1,t}^*$ as :

$$Y_{1,t}^* = \sum_{h=0}^{\infty} J_1^h \varepsilon_{1,t-h}^* \equiv (Id - J_1 L)^{-1} \varepsilon_{1,t}^*, \tag{2.5}$$

where :

$$(Id - J_1 L)^{-1} = \sum_{h=0}^{\infty} J_1^h L^h. \tag{2.6}$$

and L denotes the lag operator.

The second recursive equation : $Y_{2,t}^* = J_2 Y_{2,t-1}^* + \varepsilon_{2,t}^*$ requires a different treatment, as the eigenvalues of J_2 are of modulus strictly larger than 1. This recursive equation can be rewritten as :

$$Y_{2,t}^* = J_2^{-1} Y_{2,t+1}^* - J_2^{-1} \varepsilon_{2,t+1}^*, \tag{2.7}$$

or by recursive substitution:

$$Y_{2,t}^* = -\sum_{h=1}^{\infty} J_2^{-h} \varepsilon_{2,t+h}^* \equiv (Id - J_2 L)^{-1} \varepsilon_{2,t}^*, \tag{2.8}$$

where :

$$(Id - J_2 L)^{-1} = -\sum_{h=1}^{\infty} J_2^{-h} L^{-h}. \tag{2.9}$$

By comparing formulas (2.6) and (2.9), we see that the expression of the inverse of $(Id - J_j L)$ depends on the modulus of the eigenvalues being greater, or smaller than 1.

This leads us to the following Corollaries :

Corollary 1 : There exists a two-sided moving average solution of (2.1), that is,

$$\begin{aligned} Y_t &= [A_1 (Id - J_1 L)^{-1} A^1 + A_2 (Id - J_2 L)^{-1} A^2] \varepsilon_t \\ &= [A_1 \sum_{h=0}^{\infty} J_1^h L^h A^1 - A_2 \sum_{h=1}^{\infty} J_2^{-h} L^{-h} A^2] \varepsilon_t. \end{aligned}$$

Corollary 2 : i) $(Y_{1,t}^*)$ and $(Y_{2,t}^*)$ are purely causal and noncausal processes, respectively. They can be interpreted as the causal and noncausal components of process (Y_t); ii) These components are deterministic functions of (Y_t) since : $Y_{j,t}^* = A^j Y_t, j = 1, 2$.

Corollary 3 : We can always write:

$$(Id - \Phi L)^{-1} = \Xi_1(L) \Xi_2(L^{-1}),$$

where the roots of $det\, \Xi_1(z) = 0$ and $det\, \Xi_2(z) = 0$ are of modulus greater than 1.

Proof: We have

$$\begin{aligned} (Id - \Phi L)^{-1} &= (A_1, A_2) \begin{bmatrix} (Id - J_1 L)^{-1} & 0 \\ 0 & (Id - J_2 L)^{-1} \end{bmatrix} \begin{pmatrix} A^1 \\ A^2 \end{pmatrix} \\ &= (A_1, A_2) \begin{pmatrix} (Id - J_1 L)^{-1} & 0 \\ 0 & Id \end{pmatrix} \begin{pmatrix} Id & 0 \\ 0 & (Id - J_2 L)^{-1} \end{pmatrix} \begin{pmatrix} A^1 \\ A^2 \end{pmatrix} \end{aligned}$$

Therefore, $(Id-\Phi L)^{-1} = \Xi_1(L)\Xi_2(L^{-1})$, where $\Xi_1(L) = A \begin{pmatrix} (Id - J_1 L)^{-1} & 0 \\ 0 & Id \end{pmatrix}$ and $\Xi_2(L^{-1}) = \begin{pmatrix} Id & 0 \\ 0 & (Id - J_2 L)^{-1} \end{pmatrix} A^{-1}$.

QED

Note that the possibility to decompose $(Id - \Phi L)^{-1}$ into the causal and noncausal lag operator is equivalent to the decomposition of $(Id - \Phi L)$ proposed in Lanne, Saikkonen (2013) in the special case of a VAR(1) process.

2.2 The autocovariance function

Let us denote by $\Gamma(h) = Cov(Y_t, Y_{t-h})$ the matrix autocovariance of Y at order h, by $\Gamma^*(h) = \begin{bmatrix} \Gamma^*_{1,1}(h) & \Gamma^*_{1,2}(h) \\ \Gamma^*_{2,1}(h) & \Gamma^*_{2,2}(h) \end{bmatrix} = Cov(Y^*_t, Y^*_{t-h})$ the autocovariance of Y^* with its appropriate block decomposition. The expressions of autocovariances $\Gamma^*(h)$ are derived in Appendix 1. We have the following Proposition :

Proposition 2 :

$$\Gamma(h) = A_1\Gamma^*_{1,1}(h)A_1' + A_1\Gamma^*_{1,2}(h)A_2' + A_2\Gamma^*_{2,1}(h)A_1' + A_2\Gamma^*_{2,2}(h)A_2',$$

where :

$$\begin{aligned}
\Gamma^*_{1,1}(h) &= J_1^{|h|}\Gamma^*_{1,1}(0), \text{ for } h \geq 0, = \Gamma^*_{1,1}(0)(J_1')^{|h|}, \text{ for } h \leq 0, \\
\Gamma^*_{2,2}(h) &= \Gamma^*_{2,2}(0)(J_2')^{-|h|}, \text{ for } h \geq 0, = J_2^{-|h|}\Gamma^*_{2,2}(0), \text{ for } h \leq 0, \\
\Gamma^*_{1,2}(h) &= -[\Sigma^*_{1,2}(J_2')^{-h} + J_1\Sigma^*_{1,2}(J_2')^{-h+1} + \ldots + J_1^{h-1}\Sigma^*_{1,2}(J_2')^{-1}], \text{ for } h > 0, \\
&= 0, \text{ for } h \leq 0, \\
\Gamma^*_{2,1}(h) &= [\Gamma^*_{1,2}(h)]', \text{ where } \Sigma^*_{1,2} = Cov(\varepsilon^*_{1,t}, \varepsilon^*_{2,t}).
\end{aligned}$$

3 Identification

There are two different identification problems encountered in the mixed VAR(1) process.

First, the decomposition in formula (2.2) : $Y_t = A_1 Y^*_{1,t} + A_2 Y^*_{2,t}$ involves both causal and noncausal "factors". As it is common in factor models, the factors are defined up to invertible linear transformations and permutations. This is a static identification problem.

Second, it can be difficult to distinguish the causal and noncausal components of the process, especially if the analysis is based on the first and second-order moments of the observable process only. This is the dynamic identification problem. The two identification issues are discussed below.

3.1 Standardization of the factors

The decomposition of matrix Φ into $\Phi = A \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1}$ is not unique as well as the associated decomposition of Y into $Y_t = A_1 Y^*_{1,t} + A_2 Y^*_{2,t}$.

Indeed, let us consider a block diagonal invertible matrix $\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix}$, where Q_j is a square (n_j, n_j) matrix $j = 1, 2$. We can write :

$$\begin{aligned} \Phi &= A \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \begin{pmatrix} Q_1^{-1} & 0 \\ 0 & Q_2^{-1} \end{pmatrix} \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} \begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix} \begin{pmatrix} Q_1^{-1} & 0 \\ 0 & Q_2^{-1} \end{pmatrix} A^{-1} \\ &= \tilde{A} \begin{pmatrix} \tilde{J}_1 & 0 \\ 0 & \tilde{J}_2 \end{pmatrix} \tilde{A}^{-1}, \end{aligned}$$

where $\tilde{A}_j = A_j Q_j, \tilde{J}_j = Q_j^{-1} J_j Q_j, \tilde{A}^j = Q_j^{-1} A^j, j = 1, 2.$

Hence, we also have:

$$Y_t = A_1 Y^*_{1,t} + A_2 Y^*_{2,t} = \tilde{A}_1 \tilde{Y}_{1,t} + \tilde{A}_2 \tilde{Y}_{2,t},$$

where : $\tilde{Y}_{j,t} = \tilde{A}^j Y_t = Q_j^{-1} Y^*_{j,t}, j = 1, 2.$

The causal and noncausal "factors" are defined up to an invertible transform. However, the causal and noncausal vector spaces generated by these components, or equivalently the projectors on these spaces, are of more relevance than the components themselves. In the literature on mixed processes the decomposition has been (partly) normalized by selecting for $\begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}$ the Jordan form of matrix Φ [i.e. the complex Jordan form in Davis, Song

(2012), the real Jordan form in Gourieroux, Jasiak (2014)a]. It seems more appropriate for our purpose to introduce the standardization by means of matrix A^{-1} (see Section 4.1).

Let us consider $Q_j = [A^j(A^j)']^{1/2}$. We get : $\tilde{A}^j(\tilde{A}^j)' = Q_j^{-1}A^j(A^j)'Q_j^{-1} = Id$. This implies the following standardization :

Proposition 3 : We can standardize matrix A^{-1} so that :

$$A^j(A^j)' = Id, j = 1, 2, \tag{3.1}$$

that is, the rows of A^j form a system of orthonormal vectors.

Other standardizations are possible by changing the metric. For instance, we may standardize matrix A^{-1} as follows :

$$A^j\Gamma(0)(A^j)' = Id, j = 1, 2, \tag{3.2}$$

to account for the variance of the observed process.

3.2 The difficulty in identifying the causal dimension

It is known in the literature [see e.g. Chan, Ho, Tong (2006)] that the moving average coefficients of a two-sided moving average process are identifiable (up to the effect of $V(\varepsilon_t) = \Sigma$), if the error distribution has no Gaussian features [4]. Conversely, if the $\varepsilon_t's$ are Gaussian, the dynamics of a two-sided moving average process cannot be distinguished from the dynamics of a pure causal one-sided moving average. This makes us believe that we will likely encounter a dynamic identification problem if the statistical inference relies on only the first and second-order moments of the observable process.

Surprisingly, it is not necessarily the case. Indeed the identification issue is usually considered in an unconstrained framework, where $Y_t = \sum_{j=-\infty}^{+\infty} B_j\varepsilon_{t-j} = \sum_{j=0}^{\infty} B_j\varepsilon_{t-j} + \sum_{j=1}^{\infty} B_{-j}\varepsilon_{t+j}$. This corresponds to a decomposition into a causal component $\tilde{Y}_{1,t} = \sum_{j=0}^{\infty} B_j\varepsilon_{t-j}$ and a noncausal component $\tilde{Y}_{2,t} = \sum_{j=1}^{\infty} B_{-j}\varepsilon_{t+j}$,

[4] More precisely, at most one component of the noise can be Gaussian.

both of dimension n. The decomposition derived in Proposition 1 shows that assumption (2.1) of mixed VAR(1) process imposes additional restrictions on the dimensions of the causal and noncausal spaces, n_1 and n_2, which sum up to n, and on the patterns of the moving average coefficients (see Corollary 1). Those restrictions may facilitate the identification of the causal dimension.[5]

Let us discuss this point for a bidimensional process $n = 2$.

i) First, if (Y_t) is a pure causal VAR(1) process, we get :

$$\Gamma(h) = \Phi^{|h|}\Gamma(0), \text{ if } h \geq 0, = \Gamma(0)(\Phi')^{|h|}, \text{ if } h \leq 0. \tag{3.3}$$

Similarly, if (Y_t) is a pure noncausal VAR(1) process, we get :

$$\Gamma(h) = \Gamma(0)(\Phi')^{-|h|}, \text{ if } h \geq 0, = \Phi^{-|h|}\Gamma(0), \text{ if } h \leq 0. \tag{3.4}$$

By comparing the series of autocovariances, we easily derive the following result :

Proposition 4 : At second-order any purely causal VAR(1) process with autoregressive matrix Φ admits also a purely noncausal VAR(1) representation with matrix $\tilde{\Phi} = \Gamma(0)^{-1}(\Phi')^{-1}\Gamma(0)$.

Thus it is not possible to distinguish a pure causal representation from a pure noncausal representation from the knowledge of autocovariances only. Note that Proposition 4 is valid for any value of n, not for $n = 2$ only.

ii) Let us now consider a mixed causal/noncausal process with $n_1 = n_2 = 1$. The autocovariances of the causal and noncausal components are (see Appendix 1 for the cross autocovariances) :

[5]This possibility is the analogue of a similar result in independent component analysis. The sources can be identified at second-order, if they are serially correlated with distinct spectra (while being cross-sectionally uncorrelated). This identification result is the basis of second-order estimation methods of the sources, such as AMUSE [Tong et al. (1990)], or SOBI [Belouchrani et al. (1997)].

$$\gamma^*_{1,1}(h) = \frac{\sigma^*_{1,1}}{1-J_1^2}J_1^{|h|}, \gamma^*_{2,2}(h) = \frac{\sigma^*_{2,2}}{J_2^2-1}J_2^{-|h|}, \quad (3.5)$$

$$\gamma^*_{1,2}(h) = \frac{\sigma^*_{1,2}}{1-J_1J_2}(J_1^h - J_2^{-h}), \text{ if } h>0, =0, \text{ if } h\leq 0. \quad (3.6)$$

This leads us to the following Proposition (see Appendix 2) :

Proposition 5 : The mixed causal/noncausal process admits a pure causal representation at second-order if and only if $\gamma^*_{12}(0) = 0 \Leftrightarrow \sigma^*_{12} = 0$.

Let us discuss Proposition 5. If $\sigma^*_{12} = 0$, the "causal" and "noncausal" components are independent. Then, we can transform each of them into one another by replacing an eigenvalue of Φ by its reciprocal. In this case the process admits four representations, which cannot be distinguished, that are one purely causal with the eigenvalues of Φ equal to J_1 and $1/J_2$; one purely noncausal with eigenvalues $1/J_1, J_2$; two mixed causal/noncausal with the pair of eigenvalues being either (J_1, J_2), or $(1/J_1, 1/J_2)$. When $\sigma^*_{12} \neq 0$, the mixed process cannot be written as either a pure causal, or a pure noncausal process.

3.3 Identification of the causal dimension

Let us first show that it is not possible to disentangle a model whose causal and noncausal dimensions are $(n_1,\ n_2 = n - n_1)$ from a model whose dimensions are $(n_2 = n - n_1,\ n_1)$, given the knowledge of the autocovariance functions $\Gamma(h), h$ varying, only.

Proposition 6: At second-order a mixed model with characteristics $[(n_1, A_1, J_1), (n_2, A_2, J_2)]$ cannot be distinguished from a model with characteristics $[(n_2, A_2, J_2^{-1}), (n_1, A_1, J_1^{-1})]$.

Proof : Let us consider the two-sided moving average representation of (Y_t) given in Corollary 1. We get :

$$\begin{aligned} Y_t &= \{A_1(Id - J_1L)^{-1}A^1 + A_2(Id - J_2L)^{-1}A^2\}\varepsilon_t \\ &= A_1(Id - J_1L)^{-1}\varepsilon^*_{1,t} + A_2(Id - J_2L)^{-1}\varepsilon^*_{2,t}. \end{aligned}$$

This representation can be equivalently written as :

$$Y_t = -A_1(Id - J_1^{-1}L^{-1})^{-1}J_1^{-1}L^{-1}\varepsilon_{1,t}^* - A_2(Id - J_2^{-1}L^{-1})^{-1}J_2^{-1}L^{-1}\varepsilon_{2,t}^*.$$

Let us now define the process ε^* in reverse time as :

$$\tilde{\varepsilon}_t = \varepsilon_{-t}^*.$$

We get :

$$\begin{aligned} Y_t &= A_1(Id - J_1^{-1}L)^{-1}J^{-1}L\tilde{\varepsilon}_{1,t} - A_2(Id - J_2^{-1}L)^{-1}J_2^{-1}L\tilde{\varepsilon}_{2,t} \\ &= A_1(Id - J_1^{-1}L)^{-1}\tilde{\varepsilon}_{1,t}^* + A_2(Id - J_2^{-1}L)^{-1}\tilde{\varepsilon}_{2,t}^*, \end{aligned}$$

where $\tilde{\varepsilon}_{j,t}^* = -J_j^{-1}\tilde{\varepsilon}_{j,t-1}, j = 1, 2$. This is another mixed causal/noncausal representation of process (Y_t) in which the causal component corresponds to (n_2, A_2, J_2^{-1}) and the noncausal component to (n_1, A_1, J_1^{-1}).

QED

Thus, we can replace the eigenvalues by their reciprocals, as long as it is done simultaneously for all causal (resp. noncausal) components.
Proposition 7 extends Proposition 4 to the mixed causal/noncausal processes.

Proposition 7 : The product of causal and noncausal dimensions $n_1(n-n_1)$ is generically identifiable at second-order, that is, except on a set of Σ of measure zero.

Proof : From Proposition 2, it follows that the dimension of the vector space generated by matrices $\Gamma(h), h < 0$, is generically equal to $n_1^2 + (n - n_1)^2 = n^2 - 2n_1(n - n_1)$. This quantity is identifiable as well as $n_1(n - n_1)$. This proves the result.

QED

We conclude that, for a mixed VAR(1) process, the second-order identification is feasible to some limited extent. Indeed, for such a process of dimension n, we are able to identify at second-order the couples of representations corresponding to causal/noncausal dimensions : $\{(n, 0), (0, n)\}, \{(n-1, 1), (1, n-1)\}, \ldots$, and the associated causal/noncausal directions and dynamics. The next section describes how this result is used in the semi-parametric exploratory analysis of mixed causal/noncausal processes.

4 Semi-Parametric Exploratory Analysis

Despite the lack of identification at second-order, a semi-parametric exploratory analysis is available and based on the structure of autocovariances of a mixed VAR(1) process, discussed in Section 3. More precisely, for a mixed process with causal characteristics (n_1, A^1) and noncausal characteristics (n_2, A^2), we know that (see Proposition 2) :

$$\begin{aligned}\Gamma^*_{1,2}(h) &= Cov(Y^*_{1,t}, Y^*_{2,t-h}) \\ &= Cov(A^1 Y_t, A^2 Y_{t-h}) \\ &= A^1\Gamma(h)(A^2)' \\ &= 0 \quad \text{, for } h \leq 0. \end{aligned} \tag{4.1}$$

The above covariance conditions are used in the exploratory analysis that consists of the following steps:
step 1 estimate A^1, A^2 for a given set $(n_1, n_2 = n - n_1)$ [see Gourieroux, Monfort (2014), Section 4, for the definition and properties of Covariance estimators].
step 2 identify the product $n_1(n - n_1)$;
step 3 identify n_1 itself;
step 4 analyze the dynamics of the causal and noncausal components.

4.1 Search for the causal and noncausal directions

Let us consider a mixed process with $1 \geq n_1 \geq n - 1$. Below, we discuss the identification of the causal and noncausal directions from covariance restrictions (4.1).

Let us fix the causal/noncausal dimensions as $n_1, n_2 = n - n_1$. Then the causal/noncausal directions can be estimated as the solutions of the following constrained minimization:

$$(\hat{A}^1, \hat{A}^2) = \arg\min_{A^1, A^2} \sum_{h=0}^{-H} ||A^1\hat{\Gamma}(h)(A^2)'||^2, \tag{4.2}$$

$$\text{s.t.} \quad A^1\hat{\Gamma}(0)(A^1)' = Id_{n_1}, A^2\hat{\Gamma}(0)(A^2)' = Id_{n-n_1},$$

where $||C||^2 = Tr(CC')$, $\hat{\Gamma}(h)$ is the empirical counterpart of $\Gamma(h)$, $H, H > 0$, a sufficiently large lag, and the standardization has been written on matrix

A^{-1} directly (see Section 3.1). The value of the objective function at the optimum is denoted by $\hat{L}(n_1, n-n_1)$.

The constrained minimization (4.2) is similar to the canonical correlation analysis. We are looking for the linear transformations A^1, A^2, which are the least correlated at any nonpositive lag [6]. Therefore, one can replace the global optimization of objective function (4.2) by a recursive optimization [see e.g. Ilmonen et al. (2012) for such recursive optimization in independent component analysis, and Appendix 3].

The optimization of objective function (4.2) involves n^2 arguments, that are the elements of matrix A^{-1}, subject to $\frac{n_1(n_1+1)}{2} + \frac{(n-n_1)(n-n_1+1)}{2}$ constraints; thus the number of functionally independent arguments is equal to :

$$2n_1(n-n_1) + \frac{n_1(n_1-1)}{2} + \frac{(n-n_1)(n-n_1-1)}{2}. \tag{4.3}$$

In standard economic applications the dimension n is rather small, and the global optimization of objective function (4.2) is easily performed (see also the first-order conditions in Section 4.2). We provide in Table 1 below the number of independent arguments.

Table 1 : Number of independent arguments

	Causal dimension n_1	
Size n	$n_1 = 1$	$n_1 = 2$
2	2	/
3	5	/
4	9	10
5	14	16

There are no numerical outcomes for $n_1 = 0$, which is a degenerate case. For $n_1 > n/2$ the outcomes are symmetric.

[6] The objective function in (4.2) has an analogue in Second-Order Blind Identification (SOBI) of jointly uncorrelated, but serially correlated sources [see Belouchrani et al. (1997)]. The criterion is of the type $\sum_{h=1}^{h} ||A\hat{\Gamma}(h)A'||^2$, s.t. $A\hat{\Gamma}(0)A' = Id$.

For each lag h, we get $n_1(n-n_1)$ quadratic functions of A^1, A^2 to be minimized by means of the norm $||\quad||^2$, and in the global optimization $(H+1)n_1(n-n_1)$ such quadratic elements. Thus there is a minimum value of H to select for given size n and causal dimension n_1 in order to have a unique minimizer of objective function (4.2). The order condition, that is the minimum H, is given in Table 2.

Table 2 : Order Condition, i.e. Minimum Value of H

	Causal dimension n_1		
Size n	$n_1 = 1$	$n_1 = 2$	uniform in n_1
2	1	/	1
3	2	/	2
4	2	2	2
5	3	2	3

In model (2.1) and under Assumptions A1-A2, the sample autocovariances $\hat{\Gamma}(h), h = 0, \ldots, -H$ converge a.s. to their theoretical counterpart $\Gamma(h)$, when the number of observations T tends to infinity. Thus the solutions $\hat{A}^1, \hat{A}^2$ that minimize the objective function (4.2) will converge a.s. to the solutions of the associated asymptotic problem where the empirical $\hat{\Gamma}(h)$ are replaced by their theoretical counterparts $\Gamma(h)$. If n_1 is equal to the true causal dimension $n_{1,0}$, we know that, generically, the asymptotic objective function is minimized for the true causal and noncausal directions. Thus $\hat{A}^j$ is a consistent estimator of $A_0^j, j = 1, 2$. If n_1 is equal to $n - n_{1,0}$, the minimum of the asymptotic objective function is also equal to zero. If n_1 is different from $n_{1,0}$ and $n - n_{1,0}$, the minimum of the asymptotic objective function is generically strictly positive.

4.2 First-order conditions

Let us now derive the first-order conditions (FOC) to the constrained minimization of objective function (4.2). Below, it is shown how to eliminate the estimated Lagrange multipliers from the FOC in order to obtain a system defining $\hat{A}^1, \hat{A}^2$ only. The row vectors of matrix A^1 (resp. A^2) are denoted by $a_i^1, i = 1, \ldots, n_1$ (resp. $a_k^2, k = 1, \ldots, n_2$) .
The objective function to be minimized is :

$$\min_{a_i^1, a_k^2} \sum_{h=0}^{-H} \sum_i \sum_k (a_i^{1'} \hat{\Gamma}(h) a_k^2)^2, \tag{4.4}$$

$$\text{s.t. } a_i^{1'} \hat{\Gamma}(0) a_i^1 = 1, \forall i, a_i^{1'} \hat{\Gamma}(0) a_j^1 = 0, \forall i < j, \tag{4.5}$$

$$a_k^{2'} \hat{\Gamma}(0) a_k^2 = 1, \forall k, a_k^{2'} \hat{\Gamma}(0) a_l^2 = 0, \forall k < l. \tag{4.6}$$

Let us introduce the Lagrange multipliers : $\lambda_{i,i}/2, \lambda_{i,j}, i < j, \mu_{k,k}/2, \mu_{k,l}, k < l$, associated with the orthonormality restrictions. The FOC for the optimization of the Lagrangian are :

$$\frac{\partial \mathcal{L}}{\partial a_i^1} = 0 \Leftrightarrow \Sigma_h \Sigma_k (\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2) \hat{\Gamma}(h) \hat{a}_k^2 - \hat{\lambda}_{i,i} \hat{\Gamma}(0) \hat{a}_i^1 - \Sigma_{j,j>i} \hat{\lambda}_{i,j} \hat{\Gamma}(0) \hat{a}_j^1 = 0, \forall i,$$

$$\frac{\partial \mathcal{L}}{\partial a_k^2} = 0 \Leftrightarrow \Sigma_h \Sigma_i (\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2) \hat{\Gamma}(h)' \hat{a}_i^1 - \hat{\mu}_{k,k} \hat{\Gamma}(0) \hat{a}_k^2 - \Sigma_{l:l>k} \hat{\mu}_{k,l} \Gamma(0) \hat{a}_l^2 = 0, \forall k.$$

The Lagrange multipliers can be eliminated from the FOC by premultiplying the FOC by $\hat{a}_j^1, j < i$, and by $\hat{a}_l^2, l < k$, respectively. We get the following set of FOC for the estimates of A^1, A^2, only :

$$\begin{cases} \Sigma_h \Sigma_k [(\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2)(\hat{a}_j^{1'} \hat{\Gamma}(h) \hat{a}_k^2)] = 0, \forall j < i, \\ \Sigma_h \Sigma_i [(\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2)(\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_l^2)] = 0, \forall l < k, \\ \hat{a}_i^{1'} \hat{\Gamma}(0) \hat{a}_i^1 = 1, \forall i, \hat{a}_i^{1'} \hat{\Gamma}(0) \hat{a}_j^1 = 0, \forall i < j, \\ \hat{a}_k^{2'} \hat{\Gamma}(0) \hat{a}_k^2 = 1, \forall k, \hat{a}_k^{2'} \hat{\Gamma}(0) \hat{a}_l^2 = 0, \forall k < l. \end{cases} \tag{4.7}$$

4.3 Identification of the causal and noncausal dimensions

In practice, the nonlinear dynamics are expected to be of a rather small dimension, i.e. the noncausal dimension is expected to be small. That dimension can be found as follows:

Step 1 : Analysis of the pure causal process

Regress Y_t on Y_{t-1}, in the framework of the Seemingly Unrelated Regression (SUR), estimate Φ by the OLS and compute the associated residuals $\hat{\varepsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$. Next, plot a nonlinear multivariate ACF of these residuals, and check if these ACF are not significant. If they are not significant, the process is purely causal. Otherwise, go to step 2.

Step 2 : Analysis of the mixed process with dimension $(n-1, 1)$

Apply the estimation method of A^1, A^2 introduced in Section 4.1. Next, compute the estimated causal and noncausal components : $\hat{Y}^*_{1,t} = \hat{A}^1 Y_t, \hat{Y}^*_{2,t} = \hat{A}^2 Y_t$. Regress $\hat{Y}^*_{1,t}$ on $\hat{Y}^*_{1,t-1}$ to find $\hat{J}_1$ and the associated residuals $\hat{\varepsilon}^*_{1,t}$ [resp. $\hat{Y}^*_{2,t}$ on $\hat{Y}^*_{2,t+1}$ to find $-\hat{J}_2^{-1}$]. Estimate matrix Φ as $\hat{\Phi} = \hat{A} \begin{pmatrix} \hat{J}_1 & 0 \\ 0 & \hat{J}_2 \end{pmatrix} \hat{A}^{-1}$, and derive the mixed residuals $\hat{\varepsilon} = Y_t - \hat{\Phi} Y_{t-1}$. Plot the nonlinear ACF of $(\hat{\varepsilon}_t)$. If the nonlinear ACF are not significant, the process is mixed $(n-1, 1)$. Otherwise, go to the next step for a mixed process $(n-2, 2)$, and so on.

The exploratory analysis is mainly based on second-order methods, except for the analysis of nonlinear autocorrelograms, which requires nonlinear methods and relies on the serial independence of the error terms.

We have mentioned in the introduction the importance of noncausal components for modeling the speculative bubbles, in processes with fat-tailed errors $\varepsilon^*_{2,t}$. The presence of fat tails is not compatible with the existence of second-order moments of the error term. Nevertheless, the procedure described above relies on the sample autocovariances $\hat{\Gamma}(h)$ and not on the theoretical $\Gamma(h)$ themselves. It is known that $\hat{\Gamma}(h)$ can preserve the consistency and asymptotic distributional properties, even in the presence of fat tail errors, such as errors with stable distributions [see e.g. Davis, Resnick (1986)]. Therefore, one can expect that the method proposed above will provide consistent estimators of A^1, A^2, provided that the standardization $A^{j\prime} \hat{\Gamma}(0) A^j = Id$ is used in order to control the possibly different speeds of convergence of the elements of $\hat{\Gamma}$.

4.4 Increasing the lag(s)

We have shown above that it is possible to estimate matrix Φ in model (2.1) as:

$$\hat{\Phi} = \hat{A}_1 \hat{J}_1 \hat{A}^{1'} + \hat{A}_2 \hat{J}_2 \hat{A}^{2'}. \tag{4.8}$$

However, matrix $\hat{\Phi}$ as well as model (2.1) have no economic interpretation contrary to decomposition (2.2) into causal and noncausal components. That decomposition can be used to extend the approach described above to processes of higher autoregressive orders, or to the moving average processes. More precisely, the approach described in Section 4.1 can also be used to identify the causal and noncausal directions in models of the type :

$$Y_t = A_1 Y_{1,t}^* + A_2 Y_{2,t}^*, \tag{4.9}$$

where $A = (A_1, A_2)$ is invertible,

$$\begin{cases} \Phi_1(L) Y_{1,t}^* = \varepsilon_{1,t}^*, \\ \\ \Phi_2(L^{-1}) Y_{2,t}^* = \varepsilon_{2,t+1}^*, \end{cases} \tag{4.10}$$

where ε_t^* is a strong white noise and Φ_1, Φ_2 polynomials such that the roots of $\det \Phi_j(z) = 0, j = 1, 2$, are of modulus strictly larger than 1.

5 Nonlinear Covariance Estimators

This section introduces a direct consistent estimation of the matrix of autoregressive coefficients in the VAR model based on sample nonlinear autocovariances. Given the estimate $\hat{\Phi}$ of Φ, the real Jordan canonical form involving $\hat{A}, \hat{J}_1, \hat{J}_2$ can also be found as well as the causal and noncausal components of the process.

The advantage of this approach is that is allows us to assess the accuracy of the estimator, unlike the exploratory analysis described in Section 4, where the outcomes are determined by visual inspection of the nonlinear ACF.

5.1 Estimation of Φ

By considering the restrictions $Cov(A^1 Y_t, A^2 Y_{t-h}) = 0$ for $h \leq 0$, based on second-order moments only, we are unable to disentangle the mixed models $(n_1, n - n_1)$ and $(n - n_1, n_1)$. However, as long as the error terms ε_t are serially independent, there exist other covariance based conditions that can

be used. More precisely let us consider scalar functions of n_1 and $n - n_1$ arguments denoted by c_1, c_2, respectively. We have :

$$Cov[c_1(A^1 Y_t), c_2(A^2 Y_{t-h})] = 0, \text{ for any } c_1, c_2, \text{ and } h \leq 0. \tag{4.11}$$

From a theoretical point of view, this extended set of covariance restrictions can be used not only to identify the causal and noncausal dimensions, but also to improve the accuracy of the estimators of A^1, A^2 due to appropriately selected weights [see the generalized covariance estimators in Gourieroux, Monfort (2014)].

The practical choice of transformations c_1, c_2 is not clear. One can choose quadratic c_1 and linear c_2 to capture the absence of leverage effect at lag $h, h \leq 0$, or quadratic c_1 and quadratic c_2 to capture the absence of volatility persistence at lag $h, h \leq 0$, or even cubic , quartic, or higher order c_1 and c_2 as proposed for instance to estimate "autoregressive" schemes for mixed unidimensional processes in Gassiat (1990), and Rosenblatt (2000), Section 8.7.

Recall that in the minimization of objective function (4.2), all covariances are given equal weights as they are intuitively of the same order of magnitude. That is no longer the case if the third or fourth order (cross) moments are included.

From a theoretical point of view, there exist optimal weights for various covariances, along the lines of the generalized method of moments. However, from a practical point of view, unequal optimal weights can be numerically cumbersome to implement, due to a rather large dimension of the weighting matrix [see,e.g. Gourieroux, Monfort (2015b) for the optimal weights in an i.i.d. framework). There exist a simple way to circumvent that problem, by replacing autocovariances by autocorrelations. More precisely, let us introduce a set of functions $a_k, k = 1, ..., K$, and the autocorrelations $\hat{\rho}_{j,k}(h, \Phi) = Corr[a_j(Y_t - \Phi Y_{t-1}), a_k(Y_{t-h} - \Phi Y_{t-h-1})]$. A generalized covariance estimator is a weighted covariance estimator, defined as the minimizer of the following objective function:

$$\hat{\Phi} = \sum_{j=1}^{K} \sum_{k=1}^{K} [\sum_{h=1}^{H} \hat{\rho}_{j,k}^2(h, \Phi)], \tag{4.12}$$

where H is the highest selected lag and the theoretical autocorrelations are replaced by their sample counterparts.

This estimator and its accuracy depend on the choice of functions $a_k, k = 1, ..., K$ and the maximum lag H.

5.1.1 Estimation of the Error Distribution

After identifying the causal and noncausal dimensions, the true error terms ε_t, or their transformed versions ε_t^*, are consistently estimated by the associated residuals $\hat{\epsilon}_t, \hat{\epsilon}_t^*$, respectively. Next, these residuals can be used to estimate the unknown joint density f^*, say, of the ε_t^*, for instance by kernel smoothing. These functional estimators of the densities can be used for two different purposes. First, $\hat{f}$ can be used to simulate trajectories of the mixed process and to compute by bootstrap the accuracy of the generalized covariance estimator (see, Section 6). Second, it is possible to reestimate the mixed causal/noncausal model by maximum likelihood after substituting this kernel estimator $\hat{f}^*$ to the true f^* [see e.g. Gassiat (1993) for the asymptotic properties of such an approach].

6 Illustration

As an illustration, we consider below the application of the exploratory analysis and the generalized covariance estimator, introduced in Sections 4 and 5, respectively, to a set of simulated data.

6.1 The simulated data

Let us consider a bivariate process $n = 2$ of causal and noncausal dimensions equal to 1: $n_1 = n - n_1 = 1$. The following parameter values are fixed: $J_1 = 0.7, J_2 = 2$,

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, A^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The errors $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ are such that $\epsilon_{1,t}, \epsilon_{2,t}$ are drawn independently in the same t-student distribution with the degree of freedom $\nu = 4$, zero mean and variance equal to $\nu/(\nu - 2)$.
The autoregressive matrix is equal to:

$$\Phi = A \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1} = \begin{pmatrix} 0.7 & -1.3 \\ 0.0 & 2.0 \end{pmatrix}.$$

We draw $T = 1000$ values of the errors $\epsilon_t^s, t = 1, ..., T$, and compute the simulated transformed errors $\epsilon_t^{*s} = A^{-1}\epsilon_t^s$, the values of the causal component $Y_{1,t}^{*s} = J_1 Y_{1,t-1}^{*s} + \epsilon_{1,t}^{*s}, t = 1, ..., T$, with initial value $Y_{1,0}^{*s} = 0$, the values of the noncausal component: $Y_{2,t}^{*s} = 1/J_2 Y_{2,t+1}^{*s} - 1/J_2 \epsilon_{2,t+1}^{*s}, t = 1, ..., T$, with terminal value $Y_{2,T+1}^{*s} = 0$. Next, we compute the values of the series $Y_t^s = AY_t^{*s}, t = 1, ..., T$. They are related to the observed process as follows: $Y_{1,t} = Y_{1,t}^*$, $Y_{2,t} = Y_{1,t}^* + Y_{2,t}^* \iff Y_{1,t}^* = Y_{1,t}$, $Y_{2,t}^* = Y_{2,t} - Y_{1,t}$. Hence, the first component of Y_t is purely causal and its second component is a mixture of a causal and a noncausal process.
Figure 1 shows the path of the two components of the observed process Y^s and Figure 2 shows its autocorrelation function.

[Figure 1: Simulated Y_t]

The first observed component has mean -0.147 and variance 8.296 and the second component has mean 0.028 and variance 0.633. Their contemporaneous correlation is -0.221.

[Figure 2: Autocorrelation Function of Y_t]

The simulated data display multiple peaks in the trajectory due to the fat tails of the errors. Indeed, for the selected value of parameter $\nu = 4$, the kurtosis of the errors does not exist. The marginal and cross autocorrelations are significant up to lag 10 with exponential decay rates determined by the values of $J_1 = 0.7$ and $1/J_2 = 0.5$.
Let us now consider the auto- and cross-correlations of the causal and noncausal components of Y_t^*.

[Figure 3: Autocorrelation Function of Y_t^*]

The cross-correlations are almost all not significant in the South-West panel of Figure 3. This illustrates the condition: $\Gamma_{1,2}^*(h) = 0$, for $h \leq 0$, derived in Proposition 2, and used in the exploratory analysis below.

6.2 The exploratory analysis

The first objective of the exploratory analysis is to determine the causal/noncausal dimensions of the observed process. Hence, we consider the various possible combinations that are:
$(n_1, n - n_1) = (2,0)$ for a pure causal process,
$(n_1, n - n_1) = (0,2)$ for a pure noncausal process,

$(n_1, n-n_1) = (1,1)$ for a mixed causal/noncausal process.
In the two pure cases, parameter matrix Φ can be directly estimated from a Seemingly Unrelated Regression (SUR) model of Y_t with appropriately chosen lags or leads of Y_t as the right hand side variables. The mixed case is analyzed according to the method introduced in Section 4.1.

Pure causal model

In the pure causal model, matrix Φ is estimated by the Ordinary Least Squares (OLS) from the Seemingly Unrelated Regression of Y_t on Y_{t-1}. We get the estimated matrix $\hat{\Phi} = \begin{pmatrix} 0.721 & -1.272 \\ 0.008 & 0.473 \end{pmatrix}$ with eigenvalues $\lambda_1 = 0.67$ and $\lambda_2 = 0.52$. As expected, the eigenvalues are close to $J_1 = 0.7$ and $1/J_2 = 0.5$. The explosive root is captured by the estimate of its stationary counterpart.

It is clear that the pure causal model is **misspecified** as the second row of matrix $\hat{\Phi}$ is very different from the second row of the true matrix Φ. In practice, the true matrix Φ is unknown and such a misspecification will be detected from the analysis of the causal residuals.

Figure 4 below displays the ACF of the SUR-based causal residuals: $\hat{\epsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$.

[Figure 4: Autocorrelation Function of Causal Residuals]

The correlations are not significant, which implies that the causal residuals can be considered as weak white noises. The misspecification cannot be detected from the second-order properties of the residuals. Let us now consider the acf computed from the squared causal residuals.

[Figure 5: Autocorrelation Function of Squared Causal Residuals]

We observe nonsignificant autocorrelations in the South-East panel. This implies that causal errors ϵ_t are not serially independent. Thus, the pure causal dynamics is rejected.

Pure noncausal process

A similar approach is used for the pure noncausal process. More precisely, the initial model $Y_t = \Phi Y_{t-1} + \epsilon_t$ is transformed into its forward-looking representation $Y_t = \Phi^{-1} Y_{t+1} - \Phi^{-1} \epsilon_{t+1}$. Therefore, matrix Φ^{-1} can be estimated by the OLS in the SUR regression of Y_t on Y_{t+1}.
The estimated autoregressive coefficient is $\hat{\Phi}^{-1} = \begin{pmatrix} 0.823 & 0.392 \\ -0.117 & 0.368 \end{pmatrix}$. Its

inverse provides the estimate of Φ: $\hat{\Phi} = \begin{pmatrix} 1.054 & -1.122 \\ 0.336 & 2.356 \end{pmatrix}$. The eigenvalues of $\hat{\Phi}$ are $\lambda_1 = 1.492$ and $\lambda_2 = 1.917$ and are close to $1/J_1 = 1.428$ and $J_2 = 2.0$. The autocorrelation functions of the noncausal residuals and their squares are provided in Figures 6 and 7.

[Figure 6: Autocorrelation Function of Noncausal Residuals]

[Figure 7: Autocorrelation Function of Squared Noncausal Residuals]

The noncausal residuals satisfy the weak white noise condition. However, the pure noncausal specification is rejected due to significant squared autocorrelations of the squared residuals in the top panels.

Mixed process

In the mixed case, we estimate the rows A^1, A^2 of matrix A^{-1}, by minimizing the objective function (4.2). This constrained minimization involves the autocovariances up to lag $H = 4$ and yields the estimated matrix $\hat{A}^{-1} = \begin{pmatrix} 0.356 & 0.293 \\ -0.008 & 1.250 \end{pmatrix}$. Let us now compare the matrices $\hat{A}^{-1}$ and A^{-1}. We know that A^1, A^2 are defined up to some invertible transform (as well as the associated causal and noncausal components) (see the discussion in Section 3.1). Therefore in our framework, we verify if the first rows of $\hat{A}^{-1}$ and A^{-1} (resp. the second rows) are close to being proportional. The cosine between the row vectors are:

$cos_1 = 0.995$, $cos_2 = 0.999$, showing a quasi-proportionality.

Given these estimates, we can compute the fitted components:
$\hat{Y}^*_{1,t} = \hat{A}^1 Y_t$, $\hat{Y}^*_{2,t} = \hat{A}^2 Y_t$, by Corollary 2. Figure 8 displays the scatterplots of $(\hat{Y}^*_{j,t}, Y^*_{j,t}), j = 1, 2$.

[Figure 8: Scatterplots of Fitted and True Components]

The true and fitted components satisfy a quasi-linear relationship, which is compatible with the definition of these components up to a multiplicative scalar. The R^2 of the associated regressions are $R^2_1 = 0.997$ and $R^2_2 = 0.999$, respectively. Let us now consider the auto- and cross-correlations of $\hat{Y}^*_t$.

[Figure 9: Autocorrelation Function of $\hat{Y}^*_t$]

As expected, the autocorrelations of $\hat{Y}^*_t$ in the South-West panel are almost nonsignificant.

The regression coefficient obtained by regressing $\hat{Y}^*_{1,t}$ on $\hat{Y}^*_{1,t-1}$ [resp. $\hat{Y}^*_{2,t}$ on $\hat{Y}^*_{2,t+1}$] provide the estimated values $\hat{J}_1 = 0.725$ and $1/\hat{J}_2 = 0.472$. Given these and the previously estimated matrix $\hat{A}^{-1}$ the estimated Φ matrix is

$$\hat{\Phi} = \begin{pmatrix} 0.732 & -1.141 \\ -0.008 & 2.111 \end{pmatrix}$$

Next, we compute the mixed causal-noncausal residuals as:

$\hat{\epsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$,

and display the ACF of the mixed residuals and of the squared mixed residuals in Figures 10 and 11.

[Figure 10: Autocorrelation Function of Mixed Residuals]

[Figure 11: Autocorrelation Function of the Squared Mixed Residuals]

All autocorrelations are nonsignificant and the mixed causal/noncausal model is not rejected.

The exploratory analysis outlined above is an important preliminary step prior to applying more sophisticated estimation methods. It provides values of J_1, J_2, A, Φ, which can be used to initiate the algorithms for computing more efficient semi-parametric estimators, such as the covariance estimator discussed in the next section.

6.3 Covariance estimators

Let us now illustrate the use of the generalized covariance (GC) estimators obtained by maximizing selected linear and nonlinear autocorrelations. The residuals are denoted by $\epsilon_t(\Phi) = Y_t - \Phi Y_{t-1}$, as they depend on the unknown matrix parameter Φ. The following set of four functions of the errors is considered: $a_1(\epsilon) = \epsilon_1, a_2(\epsilon) = \epsilon_2$, $a_3(\epsilon) = \epsilon_1^2, a_4(\epsilon) = \epsilon_2^2$. From the observed process, we compute the following series:

$$(\epsilon_{1,t}(\Phi), \epsilon_{2,t}(\Phi), \epsilon^2_{1,t}(\Phi), \epsilon^2_{2,t}(\Phi)) = (Z_{1,t}(\Phi), Z_{2,t}(\Phi), Z_{3,t}(\Phi), Z_{4,t}(\Phi)),\ t = 1, ..., T.$$

Let $\hat{\rho}_{j,k}(h, \Phi)$ denote the sample autocorrelation between $Z_{j,t}(\Phi)$ and $Z_{k,t-h}(\Phi)$, $j, h = 1, ..., 4$. These include the standard serial correlation, the correlation of the squares, correlations between the residual series and the squared residual series.

The GC estimate is obtained from the following minimization based on the associated portmanteau statistic up to lag $H = 10$.

$$\hat{\Phi} = Argmin_{\Phi} \sum_{j=1}^{4} \sum_{k=1}^{4} \sum_{h=1}^{10} [\hat{\rho}_{j,k}(h, \Phi)]^2 \tag{4.13}$$

The estimated autoregressive matrix is

$$\hat{\Phi} = \begin{pmatrix} 0.7246 & -1.4525 \\ -0.0302 & 1.9939 \end{pmatrix}$$

with eigenvalues $\lambda_1 = 0.690$, $\lambda_2 = 2.027$ close to the true values $J_1 = 0.7$ and $J_2 = 2.0$.The standard errors of $\hat{\Phi}$ are obtained by bootstrap and are equal to 0.023, 0.308, for the elements of the first row, and to 0.009, 0.120 for the elements of the second row.

After estimating Φ, we compute the GC residuals $\hat{\epsilon} = Y_t - \hat{\Phi} Y_{t-1}$ in order to approximate the density of ϵ_t. We plot in Figures 12 and 13 the smoothed empirical densities of $\hat{\epsilon}_{1,t}$ and $\hat{\epsilon}_{2,t}$ and compare them to the true t-Student density of the errors.

[Figure 12: Empirical Density of GC Residuals and True Errors (causal component)]

[Figure 13: Empirical Density of GC Residuals and True Errors (non-causal component)]

The standard kernel-based density estimators, produced by S+ by using a default bandwidth overlap, indicates that the empirical densities of the GC errors and of the true errors are very close. Indeed, the extreme values of the GC errors are very close to the extremes of the true errors, as the maxima are 7.555 and 7.579, respectively and the minima are -11.183 and -11.277, respectively. The means of the GC and true errors are -0.006 and -0.013 and their standard arrors are 1.437 and 1.430, respectively, and are very close too.

Both residual series are serially uncorrelated. Their contemporaneous correlation is statistically significant and equal to -0.095, as compared to the standard asymptotic normality-based critical value of 0.0632.

7 Concluding remarks

This paper examined the problem of second-order identification in mixed causal/noncausal linear processes and revealed that in the mixed VAR(1)

process the second-order identification is available to some limited extent. That is sufficient to ensure the feasibility of semi-parametric estimation based on the autocovariance function. We propose a semi-parametric method of exploratory analysis that allows to detect the causal and noncausal dimensions of a multivariate VAR(1) process. For direct estimation of a VAR model with causal and noncausal components, a generalized covariance estimator is also introduced. It provides the estimation of $\nu_1, A_1, A_2, J_1, J_2$ in a single optimization. When the distribution of the errors is parametric, the lack of full efficiency of the generalized covariance estimator is compensated by its numerical simplicity. The ML estimator requires $n+1$ optimizations, which is equal to the number of all possible causal dimensions of the process, while the generalized covariance estimator takes one single optimization only.

REFERENCES

Belouchrani, A., Abed-Meraim, K., K., Cardoso, J., and E., Moulines (1997) : "A Blind Source Separation Technique Using Second-Order Statistics", IEEE Trans. on Signal Processing, 45, 434-444.

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

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

Davis, R., and S., Resnick (1986) : "Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages", The Annals of Statistics, 14, 533-558.

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

Gassiat, E. (1990) : "Estimation Semi-Paramétrique d'un Modéle Autorégressif Stationnaire Multiindice Non Nécessairement Causal", Ann. Inst. Henri Poincaré, 26, 181-205.

Gassiat, E. (1993) : "Adaptive Estimation in Noncausal Stationary AR Processes", Annals of Statistics, 21, 2022-2042.

Gourieroux, C., and J., Jasiak (2001) : "Nonlinear Autocorrelograms : An Application to Intertrade Durations", Journal of Time Series Analysis, 1-28.

Gourieroux, C., and J., Jasiak (2014)a : "Filtering, Prediction, and Simulation Methods for Noncausal Processes", CREST DP.

Gourieroux, C., and J., Jasiak (2014)b : "Misspecification of Causal and Noncausal Orders in Autoregressive Processes", CREST DP.

Gourieroux, C., and A., Monfort (2015a) : "Revisiting Identification and Estimation in Structural VARMA Models", CREST DP.

Gourieroux, C., and A., Monfort (2015b) : "Statistical Inference for Independent Component Analysis", CREST DP.

Gourieroux, C., and J.M., Zakoian (2014) : "Explosive Bubble Modelling by Noncausal Process", CREST DP.

Hannan, E. (1973) : "The Asymptotic Theory of Linear Time Series Models", Journal of Applied Probability, 10, 130-145.

Hyvarinen, A., Karhunen, J., and E., Oja (2001) : "Independent Component Analysis", New-York, Wiley.

Ilmonen, P., Nordhausen, K., Oja, H., and E., Ollila (2012) : "On Asymptotics of ICA Estimators and Their Performance Indices",DP, Universite Libre de Bruxelles.

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

Leeper, E., Walker, T., and S., Yang (2013) : "Fiscal Foresight and Information Flows", Econometrica, 81, 1115-1145.

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

Tong, L., Soon, V., Huang, Y., and R., Liu (1990) : "Amuse : A New Blind Identification Algorithm", in Proc. IEEE ISCAS, 1784-1787, New-Orleans, May.

Appendix 1

Proof of Proposition 2.

We derive the expression $\Gamma^*_{1,2}(h)$ for $h > 0$.

We have :

$$\begin{aligned}
\Gamma^*_{1,2}(h) &= Cov(Y^*_{1,t}, Y^*_{2,t-h}) \\
&= -Cov(\varepsilon^*_{1,t} + J_1\varepsilon^*_{1,t-1} + \ldots + J_1^{h-1}\varepsilon^*_{1,t-h+1}, J_2^{-h}\varepsilon^*_{2,t} + \ldots + J_2^{-1}\varepsilon^*_{2,t-h+1}) \\
&= -[\Sigma^*_{1,2}(J_2')^{-h} + J_1\Sigma^*_{1,2}(J_2')^{-h+1} + \ldots + J_1^{h-1}\Sigma^*_{1,2}(J_2')^{-1}], \qquad \text{(a.1)}
\end{aligned}$$

by using the serial independence of ε^*_t and $\Sigma^*_{1,2} = Cov(\varepsilon^*_{1,t}, \varepsilon^*_{2,t}), \forall t$.

In particular, if $n = 2, n_1 = n_2 = 1$, then J_1, J_2 and $\Sigma^*_{1,2}$ are scalars and we get :

$$\begin{aligned}
\gamma^*_{1,2}(h) &= -\sigma^*_{1,2}[J_2^{-h} + J_1J_2^{-(h-1)} + \ldots + J_1^{h-1}J_2^{-1}] \\
&= -\sigma^*_{1,2}J_2^{-h}[1 + (J_1J_2) + \ldots + (J_1J_2)^{h-1}] \\
&= -\sigma^*_{1,2}J_2^{-h}\frac{[1 - (J_1J_2)^h]}{1 - J_1J_2} \\
&= \frac{\sigma^*_{1,2}}{1 - J_1J_2}[J_1^h - (J_2^{-h})], \text{ if } h > 0. \qquad \text{(a.2)}
\end{aligned}$$

Appendix 2

Condition for distinguishing a mixed process from a pure process, $n = 2$

As the identification of process (Y_t) and of a one-to-one linear transformation of (Y_t) are equivalent, we assume that the mixed causal/noncausal process is (Y_t^*) itself, with the autocovariances given in (3.5)-(3.6).

By Proposition 4, we consider a pure causal process, without loss of generality. Let us denote such a process as : $Y_t = \Phi Y_{t-1} + \varepsilon_t$, where the modulus of the eigenvalues of Φ are strictly smaller than 1, and analyze the conditions ensuring that the associated autocovariances $\Gamma(h)$ coincide with the autocovariances given in (3.5)-(3.6). Let us assume for expository purpose that $J_1 \neq 1/J_2$. Then matrix Φ' is diagonalizable with eigenvalues J_1 and $1/J_2$, and we can write :

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

where $C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$ is the matrix, whose columns provide eigenvectors of Φ'. For $h \leq 0$, we have from (3.3):

$$\Gamma(h) = \Gamma(0)(\Phi')^{|h|} = \Gamma(0) C^{-1} \begin{pmatrix} J_1^{|h|} & 0 \\ 0 & J_2^{-|h|} \end{pmatrix} C.$$

Let us now consider the implications of the conditions $\gamma_{1,2}^*(h) = 0$, if $h \leq 0$. We get :

$$(1,0)\Gamma(h) \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0, \forall h, h \leq 0,$$

$$\Leftrightarrow \quad (1,0)\Gamma(0) C^{-1} \begin{pmatrix} J_1^{|h|} & 0 \\ 0 & J_2^{-|h|} \end{pmatrix} C \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0, \forall h \leq 0.$$

$$\Leftrightarrow \quad (d_{11}, d_{12}) \begin{pmatrix} J_1^{|h|} & 0 \\ 0 & J_2^{-|h|} \end{pmatrix} \begin{pmatrix} c_{12} \\ c_{22} \end{pmatrix} = 0, \forall h \leq 0.$$

$$(\text{where } (d_{11}, d_{12}) = (1,0)\Gamma(0)C^{-1})$$

$\Leftrightarrow d_{11}c_{12}J_1^{|h|} + d_{12}c_{22}J_2^{-|h|} = 0, \forall h \leq 0.$

$\Leftrightarrow$ either $c_{12} = 0$ and $d_{12} = 0$, or $c_{22} = 0$ and $d_{11} = 0$,

(since the vectors $(d_{11}, d_{12})'$ and $(c_{12}, c_{22})'$ are non zero vectors and the sequences $J_1^{|h|}$ and $J_2^{-|h|}$ are linearly independent).

a) Let us consider the case $c_{12} = 0$. Then c_{22} can be standardized to 1.

We have :

$C = \begin{pmatrix} c_{11} & 0 \\ c_{21} & 1 \end{pmatrix}, \quad C^{-1} = \frac{1}{c_{11}} \begin{pmatrix} 1 & 0 \\ -c_{21} & c_{11} \end{pmatrix},$

$(1,0)\Gamma(0)C^{-1} = \frac{1}{c_{11}}(\gamma_{11}(0), \gamma_{12}(0)) \begin{pmatrix} 1 & 0 \\ -c_{21} & c_{11} \end{pmatrix},$

and $d_{12} = \gamma_{12}(0)$.

Thus the condition $d_{12} = 0$ is equivalent to the condition $\gamma_{12}(0) = 0$.

b) Similarly, if $c_{22} = 0$, we can fix $c_{12} = 1$. We have :

$$C = \begin{pmatrix} c_{11} & 1 \\ c_{21} & 0 \end{pmatrix}, C^{-1} = -\frac{1}{c_{21}} \begin{pmatrix} 0 & -1 \\ -c_{21} & c_{11} \end{pmatrix},$$

$$(1,0)\Gamma(0)C^{-1} = -\frac{1}{c_{21}}[\gamma_{11}(0), \gamma_{12}(0)] \begin{pmatrix} 0 & -1 \\ -c_{21} & c_{11} \end{pmatrix},$$

and $d_{11} = \gamma_{12}(0)$. The condition $d_{11} = 0$ is equivalent to the condition $\gamma_{12}(0) = 0$.

Appendix 3

Numerical Considerations

A.3.1 Recursive optimisation

The global constrained optimisation (4.2) can be replaced by recursive optimizations, with a tradeoff between the number of optimisations and the dimensions of the optimizers [see Hyvarinen et al. (2001), Gourieroux, Monfort (2015) for the recursive analogue in independent component analysis]. For instance for $n_1 = n_2 = 2$, consistent estimators of the causal and non-causal directions can be derived in two optimizations, first with respect to the first rows of A^1 and A^2, then with respect to their second rows as follows :

First optimization :

$$(\hat{a}_1^1, \hat{a}_1^2) = \arg\min_{a_1^1, a_1^2} \sum_{h=0}^{-H} (a_1^1 \hat{\Gamma}(h) a_1^{2'})^2,$$

$$\text{s.t. } a_1^1 \hat{\Gamma}(0)(a_1^1)' = 1, a_1^2 \hat{\Gamma}(0)(a_1^2)' = 1,$$

where a_1^1 and a_1^2 have dimensions $(1, n)$.

Second optimization :

The next optimization is performed to find the second rows of A^1, A^2, denoted by a_2^1, a_2^2, once the first-rows have been derived. The estimators are :

$$(\hat{a}_2^1, \hat{a}_2^2) = \arg\min_{a_2^1, a_2^2} \sum_{h=0}^{-H} || \begin{pmatrix} \hat{a}_1^1 \\ a_2^1 \end{pmatrix} \hat{\Gamma}(h) \begin{pmatrix} \hat{a}_1^2 \\ a_2^2 \end{pmatrix}' ||^2,$$

$$\text{s.t. } \hat{a}_2^1 \hat{\Gamma}(0)(\hat{a}_2^1)' = 1, \hat{a}_2^2 \hat{\Gamma}(0)(\hat{a}_2^2)' = 1,$$

$$\hat{a}_2^1 \hat{\Gamma}(0)(\hat{a}_1^1)' = 0, \hat{a}_2^2 \hat{\Gamma}(0)(\hat{a}_1^2)' = 0.$$

Appendix 4

Identification of a mixed process for $n = 2$

Let us consider the mixed process (Y_t^*) with $n_1 = n_2 = 1$ and the autocovariances given in (3.5)-(3.6).
i) We know from Proposition 6 that this process has the same second-order properties as another mixed process, where the first component is noncausal associated with the inverse J_1^{-1}, and the second one is causal associated with J_2^{-1}.
ii) Let us now consider another mixed representation of process (Y_t^*) with the matrix A^{-1} defining the new causal and noncausal components.
The autocovariance of the new causal component is given by:

$$\begin{aligned}\tilde{\gamma}_{1,1}(h) &= (a^{11}, a^{12})\Gamma(h)\begin{pmatrix} a^{11} \\ a^{12} \end{pmatrix} \\ &= (a^{11})^2 \frac{\sigma_{11}^*}{1-J_1^2} J_1^{|h|} + (a^{12})^2 \frac{\sigma_{22}^*}{J_2^2-1} J_2^{-|h|} \\ &\quad + \frac{a^{11}a^{12}}{1-J_1J_2}\sigma_{12}^*(J_1^{|h|} - J_2^{-|h|}).\end{aligned}$$

This autocovariance has the exponential form (3.5) only if it is proportional to either $J_1^{|h|}$ or $J_2^{-|h|}$. Without loss of generality, we can consider the case when it is proportional to $J_1^{|h|}$ (by applying part i) above).
We have:

$$\tilde{\gamma}_{1,1}(h) = \gamma_{1,1} J_1^{|h|} \iff a^{12}\left[a^{12}\frac{\sigma_{22}^*}{J_2^2-1} - \frac{a^{11}}{1-J_1J_2}\sigma_{12}^*\right] = 0.$$

When $a^{12} = 0$, we get the initial causal component $(Y_{1,t}^*)$ (defined up to a multiplicative scalar). Otherwise, we can assume $\tilde{a}^{12} = 1$ and obtain:

$$\tilde{a}^{11} = \frac{\sigma_{22}^*}{J_2^2-1}\frac{1-J_1J_2}{\sigma_{12}^*}, \qquad \text{(a.1)}$$

whenever $\sigma_{12}^* \neq 0$.
By symmetry, we can consider the autocovariance of the new second component, that is,

$$\tilde{\gamma}_{2,2}(h) = (a^{21}, a^{22})\Gamma(h) \begin{pmatrix} a^{21} \\ a^{22} \end{pmatrix}.$$

It is proportional to $J_2^{-|h|}$, if and only if, either $a^{21} = 0$, or $\tilde{a}^{21} = 1$ and

$$\tilde{a}^{22} = \frac{\sigma_{11}^*}{1 - J_1^2} \frac{1 - J_1 J_2}{\sigma_{12}^*}. \tag{a.2}$$

Finally, the cross-covariance between the new components is

$$\tilde{\gamma}_{1,2}(h) = (a^{11}, a^{12})\Gamma(h) \begin{pmatrix} a^{21} \\ a^{12} \end{pmatrix}.$$

This cross-covariance has to be zero for $h < 0$. This condition is equivalent to:

$$\begin{aligned}
&\tilde{\gamma}_{1,2}(h) = 0 \\
&\iff \quad \tilde{a}^{11}\tilde{a}^{21} \frac{\sigma_{11}^*}{1 - J_1^2} J_1^{|h|} + \tilde{a}^{12}\tilde{a}^{22} \frac{\sigma_{22}^*}{J_2^2 - 1} J_2^{-|h|} \\
&\qquad + \frac{\tilde{a}^{12}\tilde{a}^{21}}{1 - J_1 J_2} \sigma_{12}^* (J_1^{|h|} - J_2^{-|h|}) = 0, \ \ \forall h \leq 0 \\
&\iff \quad \begin{cases} \tilde{a}^{11} \frac{\sigma_{11}^*}{1 - J_1^2} + \frac{\sigma_{12}^*}{1 - J_1 J_2} = 0, \\ \\ \tilde{a}^{22} \frac{\sigma_{22}^*}{J_2^2 - 1} + \frac{\sigma_{12}^*}{1 - J_1 J_2} = 0 \end{cases}
\end{aligned} \tag{a.3}$$

We note that the conditions (a.3) are not compatible with the conditions (a.1)-(a.2). To summarize, if $\sigma_{12}^* \neq 0$, the only identification problem for the mixed process is the change of eigenvalues into their inverses discussed in part i).

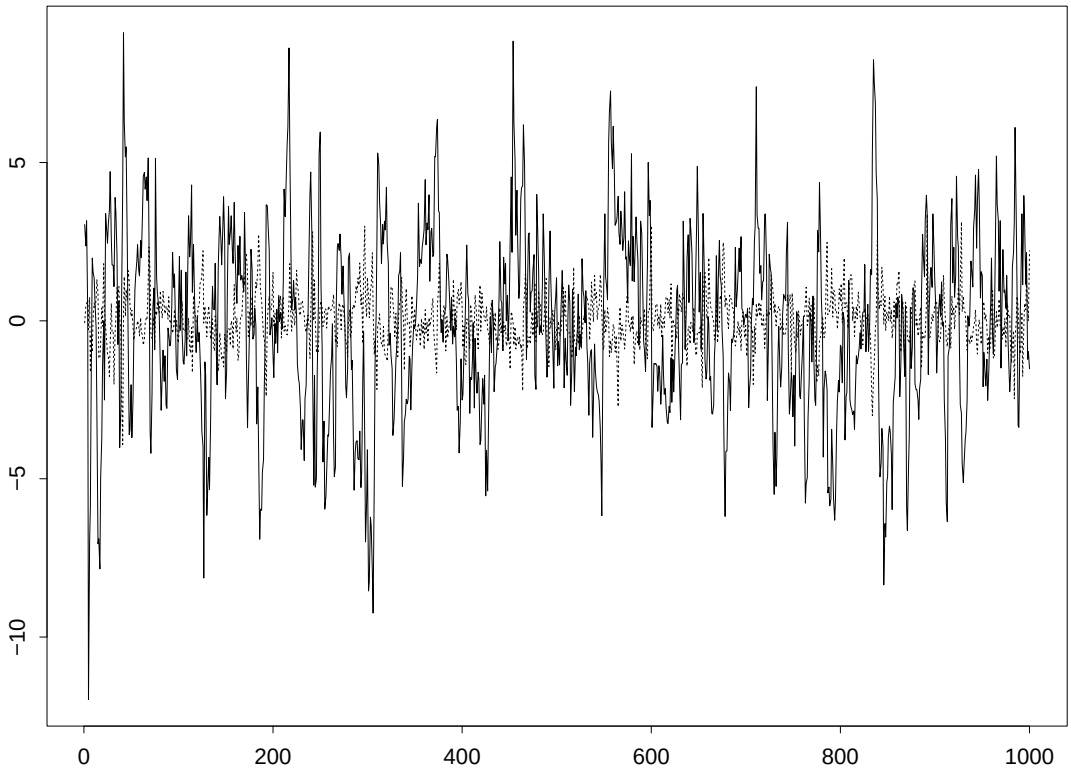

Figure 1: Simulated Y_t

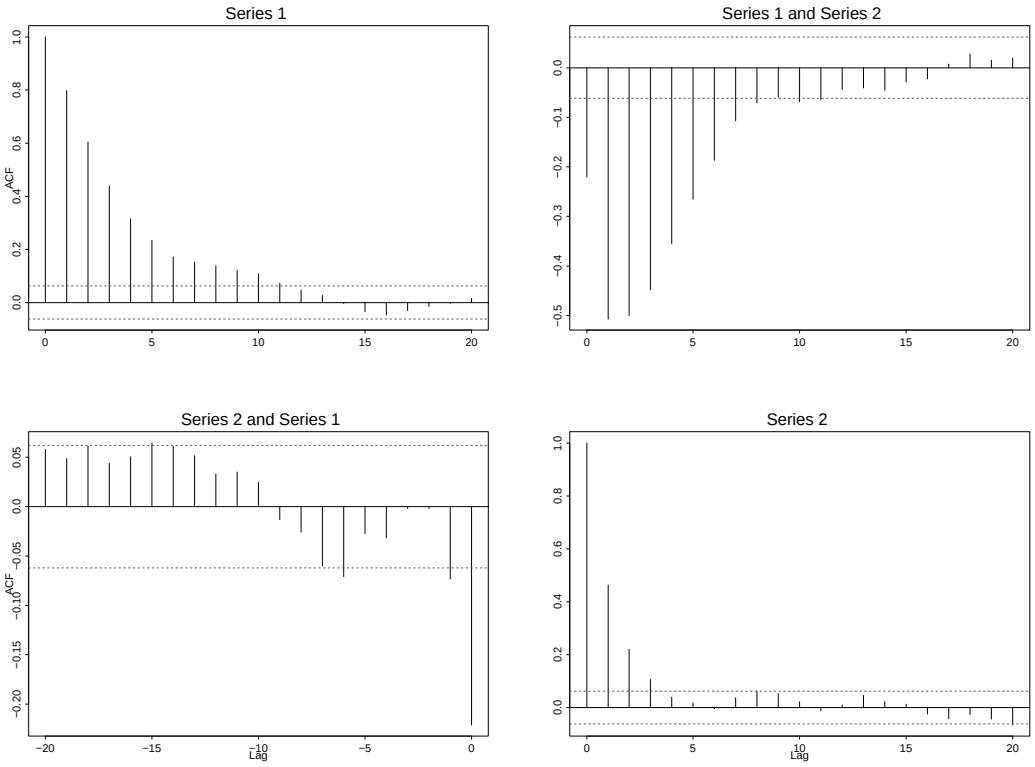

Figure 2: Autocorrelation Function of Y_t

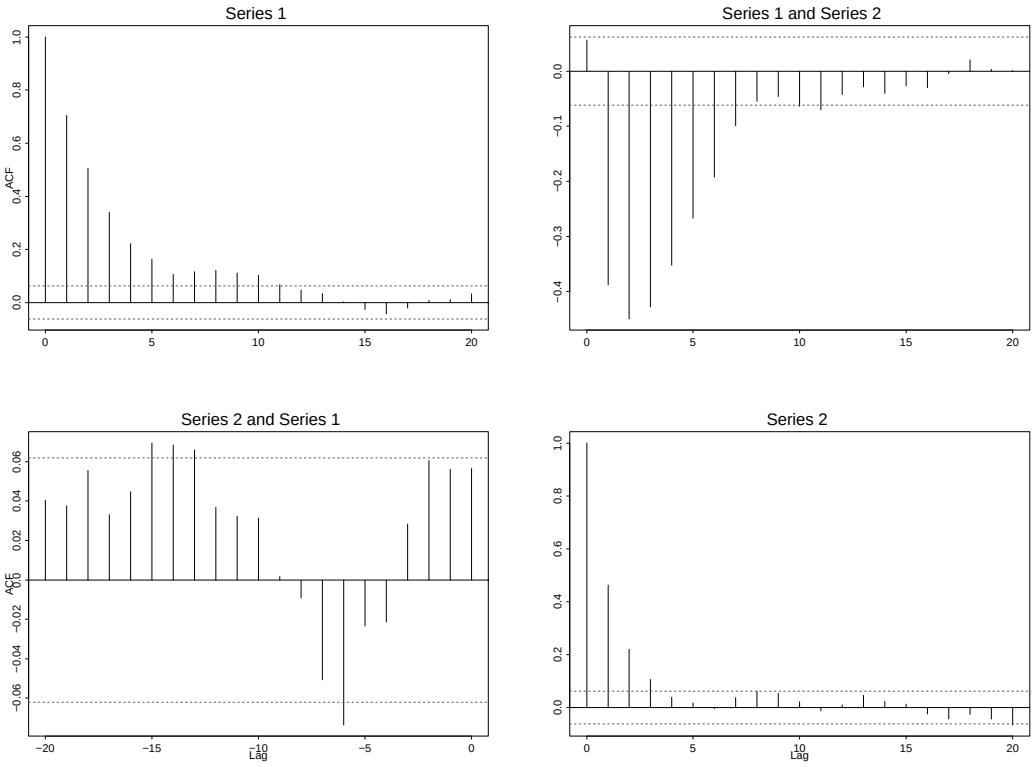

Figure 3: Autocorrelation Function of Y_t^*

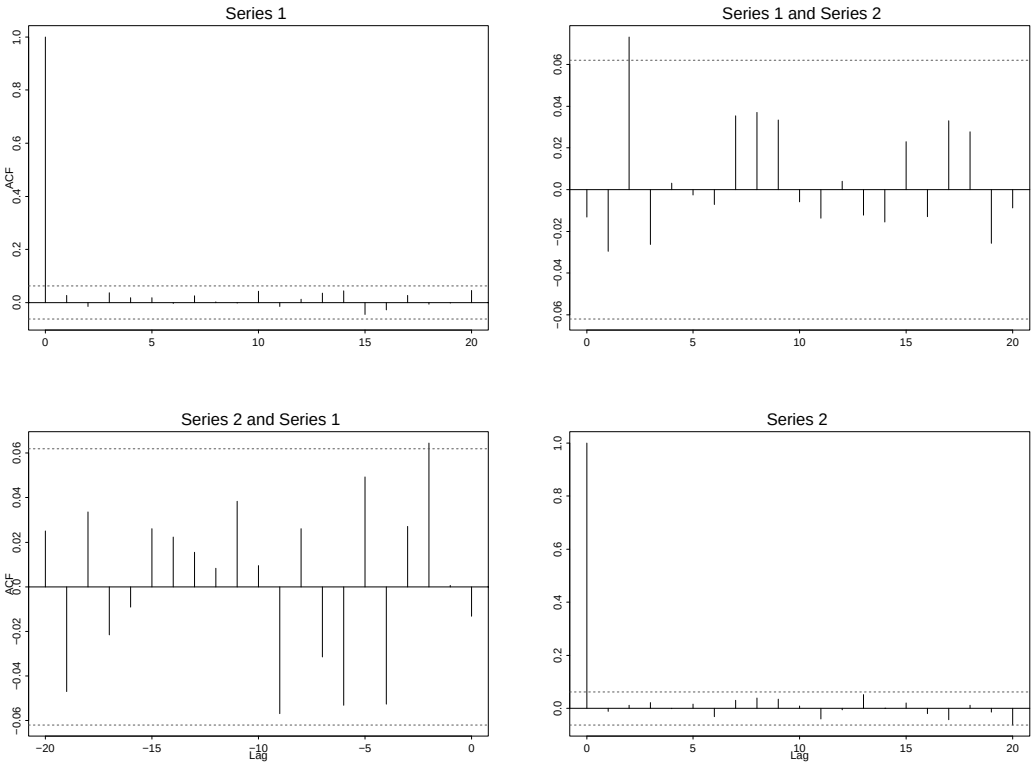

Figure 4: Autocorrelation Function of Causal Residuals

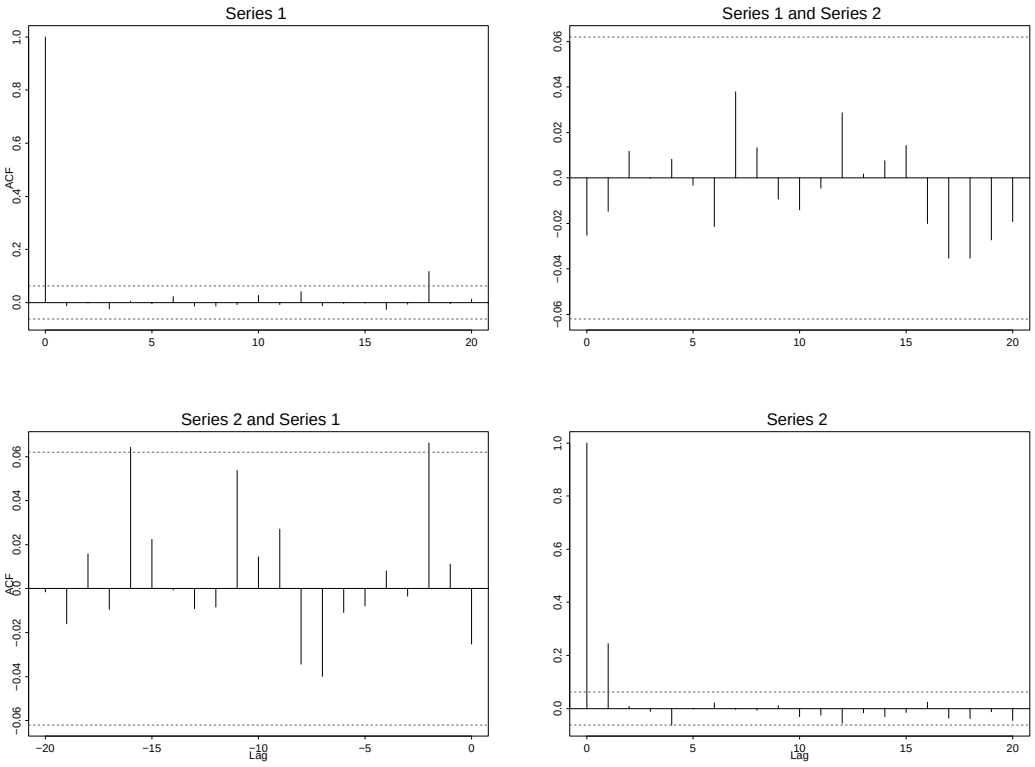

Figure 5: Autocorrelation Function of Squared Causal Residuals

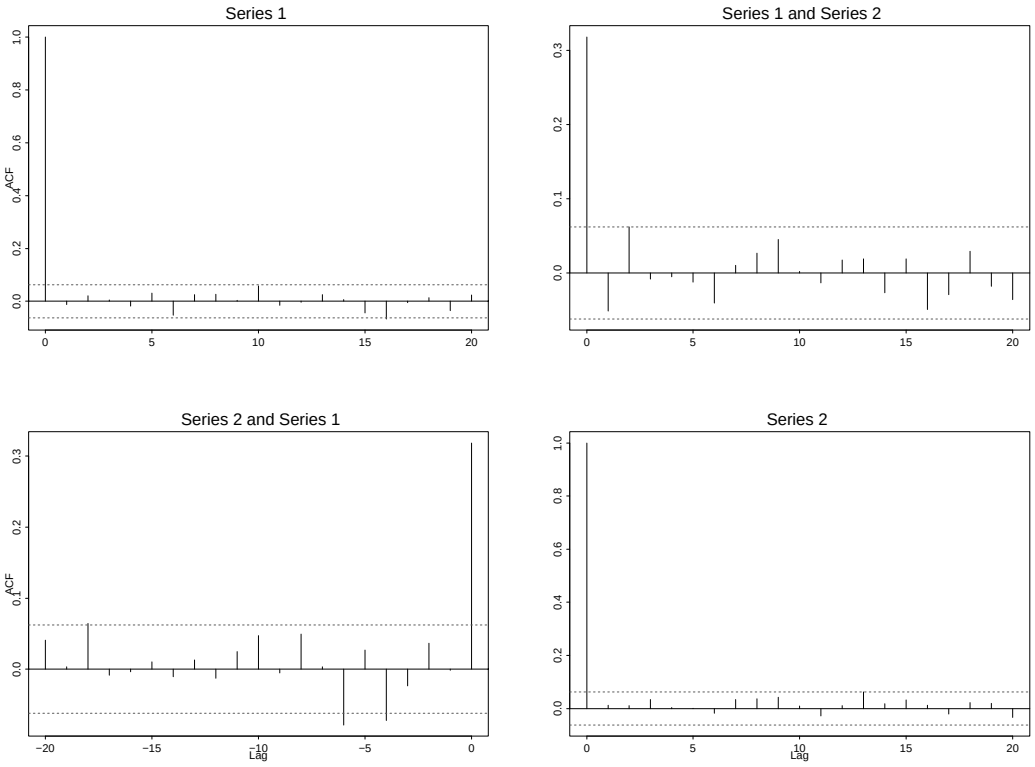

Figure 6: Autocorrelation Function of Noncausal Residuals

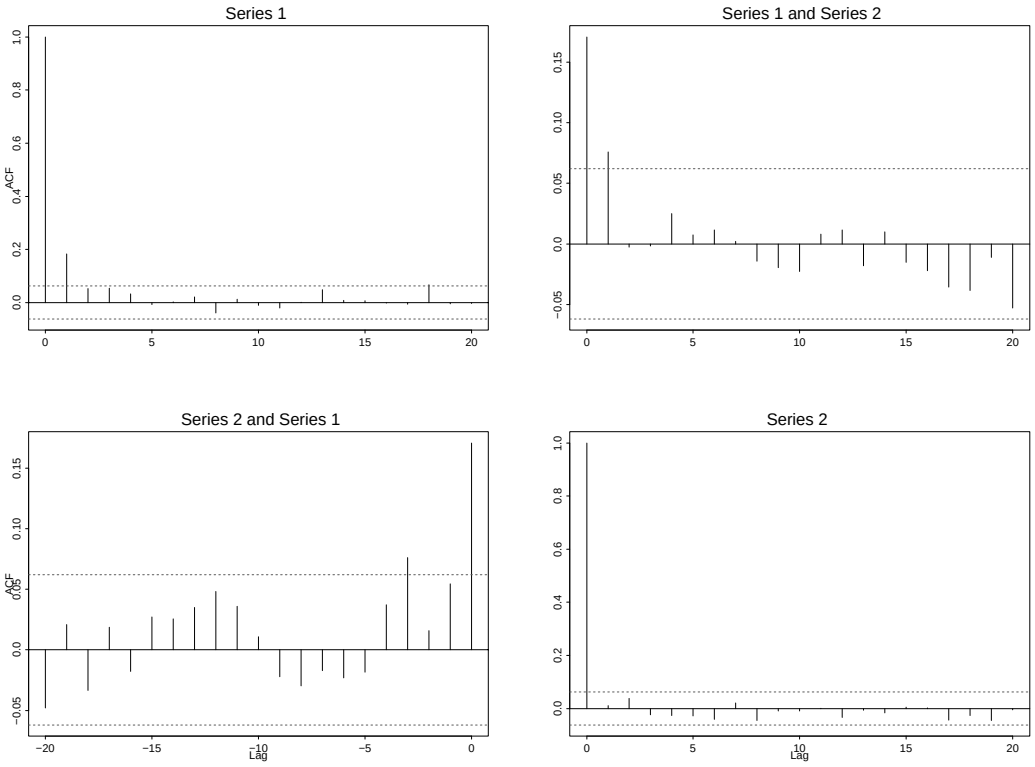

Figure 7: Autocorrelation Function of Squared Noncausal Residuals

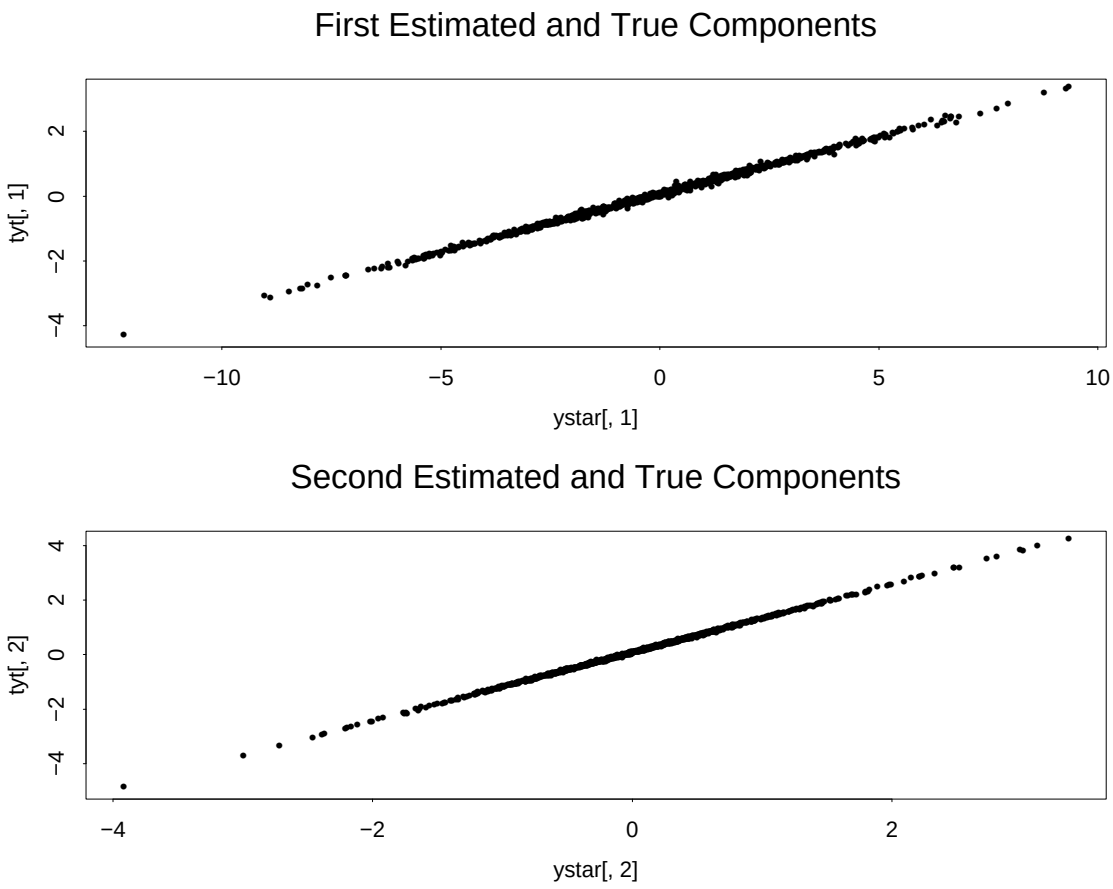

Figure 8: Scatterplots of Fitted and True Components

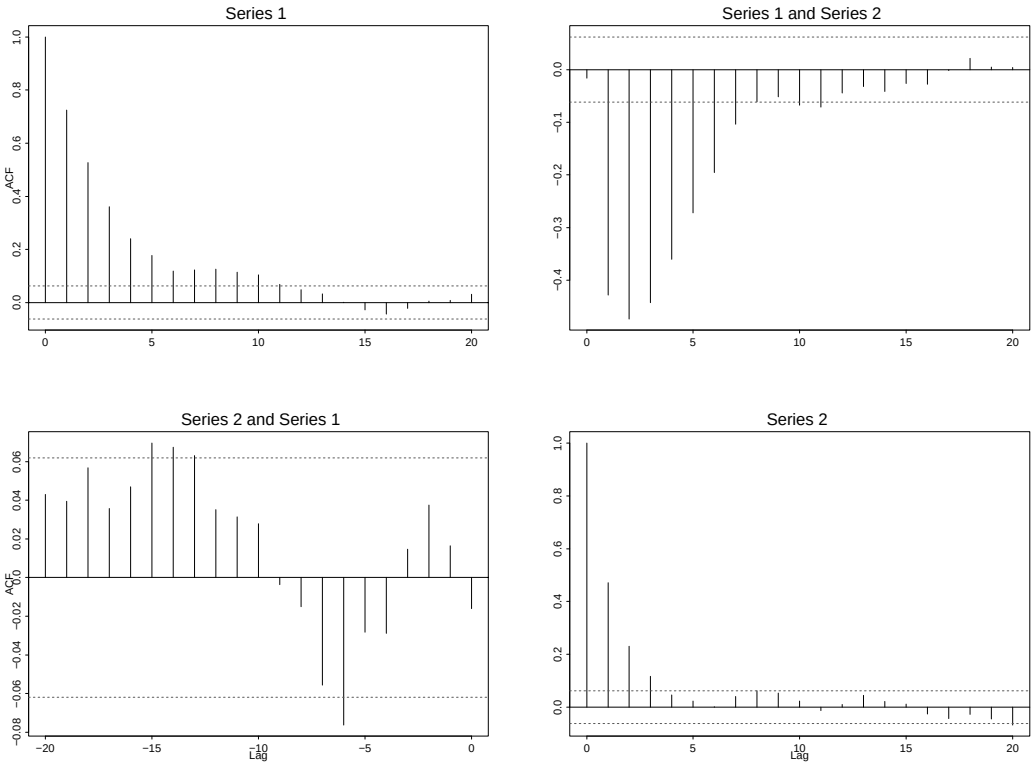

Figure 9: Autocorrelation Function of $\hat{Y}_t^*$

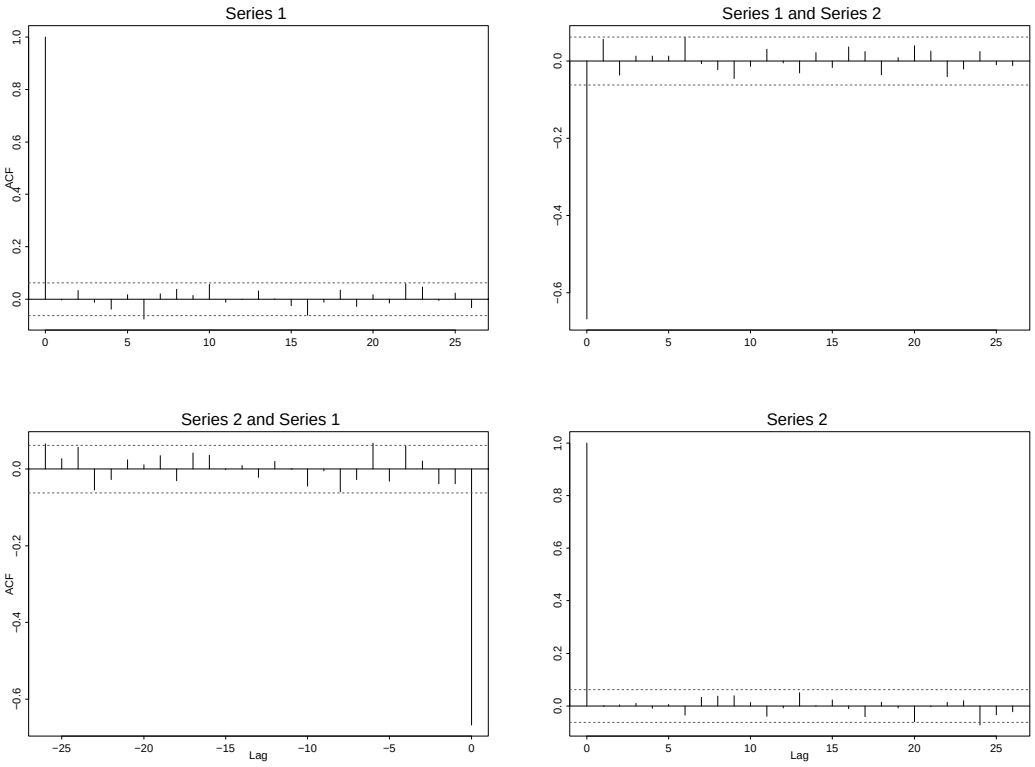

Figure 10: Autocorrelation Function of Mixed Residuals

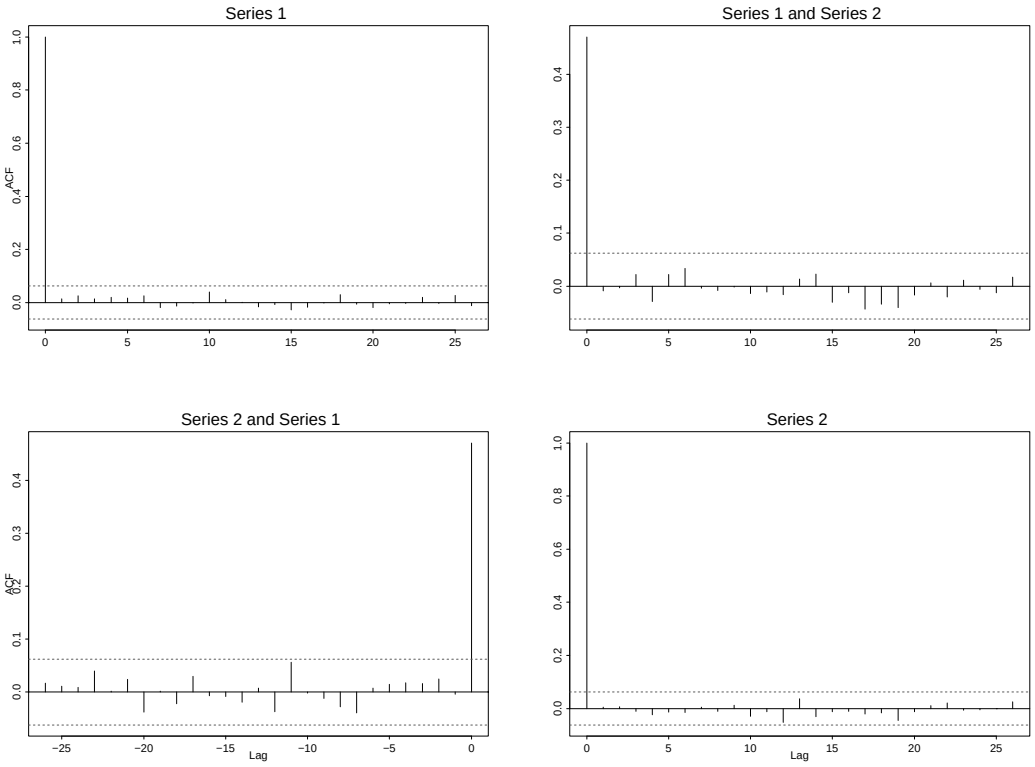

Figure 11: Autocorrelation Function of Squared Mixed Residuals

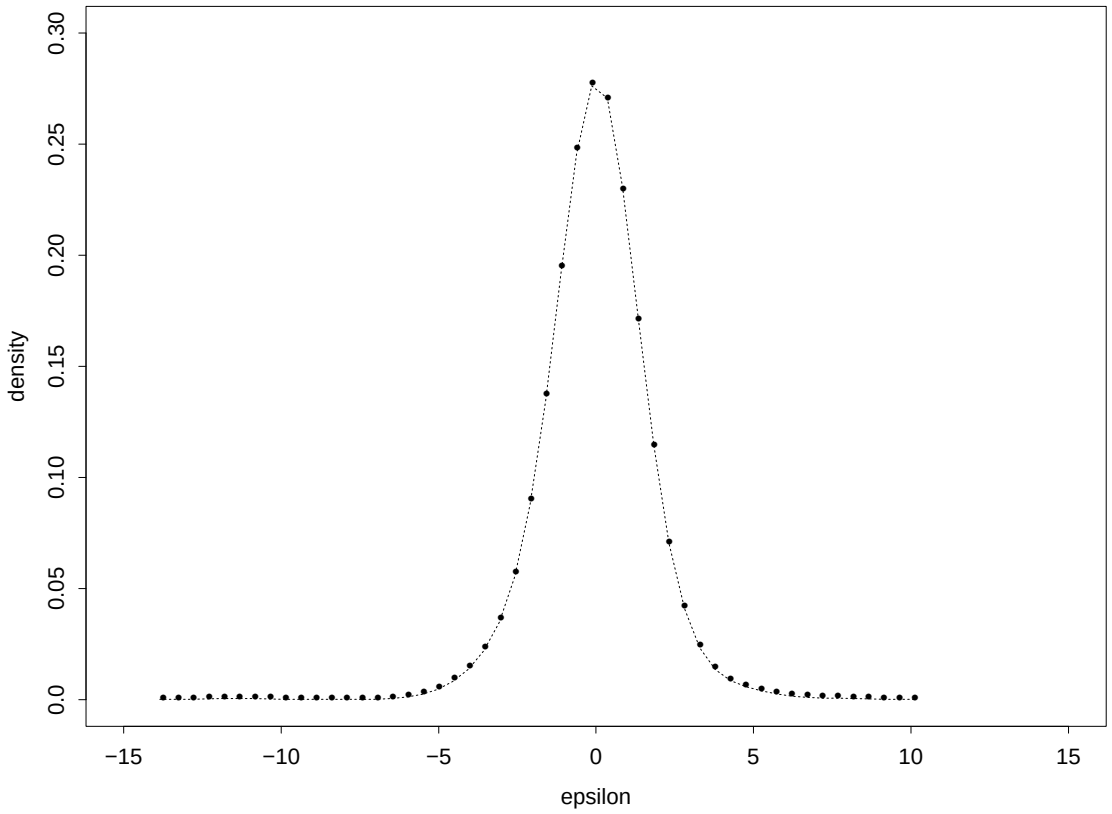

Figure 12: Empirical Density of GC Residuals *dotted line* and True Errors *dot symbols* (causal component)

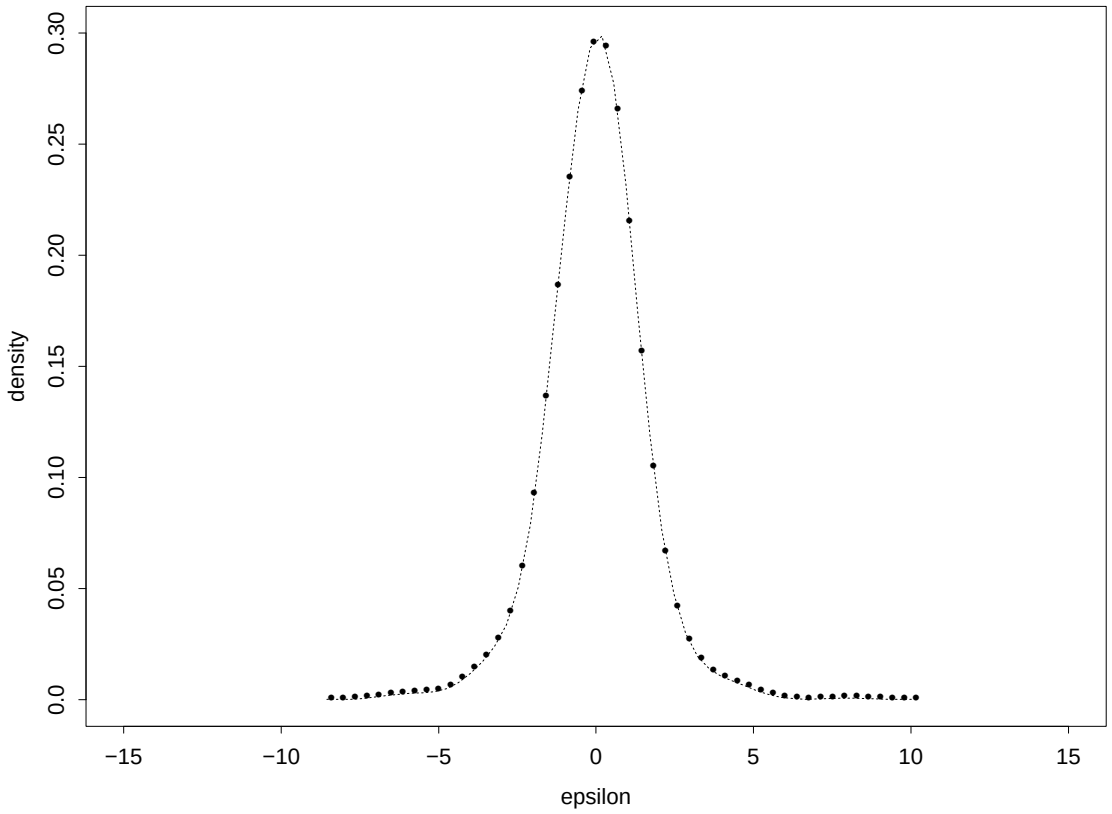

Figure 13: Empirical Density of GC Residuals *dotted line* and True Errors *dot symbols* (noncausal component)