Nonlinear Fore(Back)casting and Innovation Filtering for Causal-Noncausal (S)VAR Models

C. Gourieroux*, and J. Jasiak†

This version: June 6, 2023

Abstract

We introduce closed-form formulas of out-of-sample backward and forward predictive densities for forecasting and backcasting of mixed causal-noncausal (Structural) Vector Autoregressive ((S)VAR) models. These nonlinear and time irreversible non-Gaussian VAR processes are shown to satisfy the Markov property in both calendar and reverse time. A post-estimation inference method for assessing the forecast interval uncertainty due to the preliminary estimation step is introduced too. The nonlinear past dependent innovations of a mixed causal-noncausal (S)VAR model are defined and their filtering and identification methods are discussed. Our approach is illustrated by a simulation study and an empirical analysis of real oil prices and real GDP growth rates, as well as an application to cryptocurrency prices.

Keywords: Mixed Causal-Noncausal Process, Predictive Density, Backcasting, Time Reversibility, Filtering, Nonlinear Innovations, Generalized Covariance (GCov) Estimator, Cryptocurrency, Oil Price.

*University of Toronto, Toulouse School of Economics and CREST, e-mail: *Christian.Gourieroux@ensae.fr*

†York University, e-mail: *jasiakj@yorku.ca*

The second author acknowledges financial support of the Natural Sciences and Engineering Council of Canada (NSERC). We thank the editor and anonymous referees for helpful comments.

1 Introduction

There has been a growing interest in applications involving the mixed causal-noncausal Vector Autoregressive (VAR) models [3] and their theoretical properties [Gourieroux, Jasiak (2017),(2022), Davis, Song (2020), Lanne, Luoto (2016),(2021), Swensen (2022)]. In applied research, the stationary mixed VAR models can replicate local trends, spikes and bubbles characterizing, e.g. the commodity prices, which are interpreted as nonstationary features in the traditional literature on causal, i.e. past dependent processes. In terms of theoretical properties, the main differences between the mixed VAR's and the traditional causal VAR's are in the assumptions concerning the eigenvalues of the autoregressive matrix coefficients and the errors of the model. More specifically, in the mixed model the roots of the autoregressive characteristic equation can be in modulus either greater or smaller than 1, as opposed to being only greater than 1 in the causal VARs. While the errors of a causal VAR are assumed to be normally distributed, the errors of a mixed VAR need to be non-Gaussian for the identification of causal and noncausal dynamics.

The traditional Gaussian Maximum Likelihood (ML) and Least Squares estimators are flawed in applications to mixed VAR's, as these methods are constrained to the Gaussian and causal dynamics. Under parametric distributional assumptions, the ML estimators based on non-Gaussian likelihood functions can be used for the mixed VAR models, as shown in Davis, Song (2020). When the error distribution is left unspecified, the Generalized Covariance (GCov) estimator is available, providing consistent, asymptotically normally distributed and semi-parametrically efficient parameters estimates in one step [Gourieroux, Jasiak (2022)].

There is an important difference between the errors of a causal and mixed VAR models, other than non-Gaussianity. Unlike the errors of causal VARs, those of mixed VAR's are not uncorrelated / independent from the past values of the process. Hence, they cannot be interpreted as innovations. For this reason, innovation-based inference, such as impulse response analysis and variance decomposition routinely conducted in the causal VAR models cannot be applied to the mixed VAR models. So far, the nonlinear causal innovations of a mixed VAR model have not been defined in the literature.

Another difficulty arises with the fore- and backcasting of the mixed VAR mod-

[3] Referred to as mixed VAR models, henceforth.

els. A mixed VAR model with non-Gaussian errors cannot be forecast directly from its conditional expectation, unlike the traditional causal VAR model. For out-of-sample (oos) forecasting of mixed causal-noncausal VAR processes, there exist in the literature a simulation-based method of forecasting given in Nyberg, Saikkonen (2014) and a Bayesian method of Lanne, Luoto (2016), which are valid for a constraint multiplicative mixed VAR representation. An operational causal predictive density in closed-form for the multivariate mixed VAR models has been deemed infeasible in Fries, Zakoian (2019) [4].

Moreover, the backcasting, which is straightforward for the time reversible linear Gaussian causal VAR processes, is complicated in this context as well. The mixed VAR does not satisfy the assumption of Gaussianity ensuring the time reversibility, which underlies the commonly used time series and machine learning methods of backcasting [Twumasi, Twumasi (2022)]. A closed-form formula of (oos) backcasting for mixed VAR models has not been introduced in the literature yet.

This paper addresses the two problems of i) the lack of closed-form oos predictive density for (oos) forecasting and backcasting and ii) the lack of a definition of nonlinear causal innovations for the mixed VAR model.

We provide closed-form formulas of backward and forward predictive density for out of sample (oos) forecasting and backcasting of mixed causal-noncausal VAR processes. This approach is built upon the results on noncausal processes in Gourieroux, Jasiak (2016). For fore(back)casting, we show that the mixed VAR is a nonlinear Markov process in both calendar and reverse time and it is time-irreversible. The closed-form expression of predictive densities for fore- and backcasting the mixed causal-noncausal VAR processes are given at horizon 1. They can be used sequentially to predict at any horizon. The proposed oos forecasting and backcasting methods are applicable to the structural (S)VAR models, which can be seen as the mixed causal-noncausal VAR models satisfying an additional assumption of cross-sectional error independence. From the predictive densities, we infer the point forecasts and prediction intervals. A post-estimation inference method for assessing the forecast interval uncertainty due to the preliminary estimation step is introduced too. A confidence set of the predicted set is proposed as a post-estimation inference method to capture the effect of the preliminary estimation step on the prediction interval.

[4] "The predictive density is generally not available under closed form", Fries, Zakoian (2019)

The nonlinear causal innovations are discussed in the context of the mixed causal-noncausal SVAR model. The additional cross-sectional independence condition, commonly used in macroeconomics [see, e.g. Guay (2021)], helps identify the autoregressive parameters and a suitable error orthogonalization method in a SVAR model. We show that, in a (S)VAR model, it is possible to define the nonlinear causal innovations by extending the results in Koop, Pesaran, Potter (1996), Potter (2000), and Gourieroux, Jasiak (2005), which are in line with those in Gonzalves et al.(2022). We define the nonlinear causal innovations for a mixed causal-noncausal (S)VAR model, determine their relationship with the model errors and discuss their identification and in-sample filtering.

The oos forecast performance and nonlinear causal innovations analysis are illustrated by simulations and empirical applications. We use the one-step semi-parametric GCov estimators introduced in Gourieroux, Jasiak (2022) for (S)VAR models. This approach to (S)VAR analysis does not require any distributional assumptions on the errors of a (S)VAR model. The semi-parametric estimation method distinguishes our approach from the existing literature employing maximum likelihood-based methods [Davis, Song (2020), Swensen (2022)], which are consistent provided that the distributional assumptions are satisfied.

The paper is organized as follows. Section 2 describes the causal-noncausal VAR model and the structural SVAR model in which the errors depend linearly on the cross-sectionally independent sources. Section 3 presents the closed-form formulas of the multivariate predictive density for forecasting and backcasting. Section 4 defines the nonlinear causal innovations and discusses their identification and filtering. Section 5 reviews the semi-parametric estimation of the mixed (S)VAR model and introduces a filtering algorithm and post-estimation inference on the prediction set. Two empirical applications are provided in Section 6. The first application considers the joint dynamics of oil prices and GDP growth rates. The second application studies a bivariate process of Bitcoin/USD and Ethereum/USD exchange rates. Section 7 concludes. The technical results are given in Appendices 1-3. Appendix 1 contains the proof of the forward and backward predictive density formula. Appendix 2 explains the incompatibility of the multiplicative representation of a causal-noncausal VAR model with the structural SVAR representation. Appendix 3 discusses the (under)-identification of nonlinear innovations in a multivariate framework.

Additional information is provided in the online Appendices B, C and D, which include a discussion of the identification conditions for Independent Component Analysis (ICA), the closed-form expression of the kernel-based semi-nonparametric estimator of predictive density, and additional results for the empirical applications.

2 Mixed Causal-Noncausal Processes

The causal-noncausal VAR(p) model has been studied in Gourieroux, Jasiak (2016),(2017), Davis, Song (2020), Swensen (2022) and Cubbada, Hecq, Voisin (2023).

2.1 The Model

The multivariate causal-noncausal VAR(p) process is defined by:

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t, \tag{2.1}$$

where Y_t is a vector of size m, Φ_j, $j = 1, ..., p$ are matrices of autoregressive coefficients of dimension $m \times m$ and (ε_t) is a sequence of errors, which are serially independent, identically distributed (i.i.d.) random vectors of dimension m with mean zero and finite variance-covariance matrix Σ. Errors (ε_t) are assumed to have a non-Gaussian distribution. Since (ε_t) is not assumed independent of past Ys, this error process cannot be interpreted as an innovation process.

We also assume that the roots of the characteristic equation of the autoregressive polynomial matrix $det(Id - \Phi_1\lambda - \cdots \Phi_p\lambda^p) = 0$ are of modulus either strictly greater, or strictly less than one, i.e. are either outside, or inside the unit circle. Then, there exists a unique strictly stationary solution to the VAR model (2.1).

The strictly stationary solution (Y_t) to model (2.1) can be written as a two-sided moving average in errors ε_t:

$$Y_t = \sum_{j=-\infty}^{+\infty} C_j \varepsilon_{t-j}. \tag{2.2}$$

This is a linear time series, according to the terminology used in Rosenblatt (2012). The matrices of coefficients on the past and future terms of this MA representation are uniquely defined when ε_t is non-Gaussian, which is an identifying assumption. Y_t is said to be causal

in ε_t, if $Y_t = \sum_{j=0}^{+\infty} C_j \varepsilon_{t-j}$, or noncausal in ε_t, if $Y_t = \sum_{j=-\infty}^{-1} C_j \varepsilon_{t-j} = \sum_{j=1}^{+\infty} C_{-j} \varepsilon_{t+j}$, or mixed causal-noncausal, otherwise.

In the presence of a noncausal component, the assumption of strict stationarity of Y_t implies the nonlinear causal dynamics of Y_t with $E(Y_t|\underline{Y_{t-1}})$ being nonlinear in $\underline{Y_{t-1}} = (Y_t, Y_{t-1}, ...)$ and past-dependent conditional heteroscedasticity $V(Y_t|\underline{Y_{t-1}})$. Then, the process (Y_t) can be characterized by a possibly complex conditional distribution of Y_{t+h} given $\underline{Y_t} = (Y_t, Y_{t-1}, ...)$ for $h = 1, 2, ..$, that provides nonlinear oos predictions. The objective of this paper is to derive this conditional distribution under closed form.

A process (Y_t) with nonlinear causal dynamics may display local trends, spikes and bubbles similar to those observed in the time series of commodity (oil) prices, exchange rates, or cryptocurrency prices [Gourieroux, Zakoian (2017), Gourieroux, Jasiak (2017), Gourieroux, Jasiak, Tong (2021)].

2.2 Multiplicative VAR

An alternative multiplicative Mixed Autoregressive (MAR) representation of the mixed causal-noncausal VAR model was proposed by Lanne, Saikkonen (2008),(2013):

$$\Phi(L)\Psi(L^{-1})Y_t = \epsilon_t^*,$$

where the matrices of autoregressive polynomials Φ and Ψ have both roots outside the unit circle and ϵ_t^* is an i.i.d. sequence of errors. Davis, Song (2020), Swensen (2022), Cubbada, Hecq, Voisin (2022), point out that the multiplicative representation of a mixed causal-noncausal VAR model does not always exist. In addition, Appendix 2 shows that the multiplicative representation may not be compatible with the SVAR model given below.

2.3 Structural VAR (SVAR) Model

A structural mixed causal-noncausal VAR model (henceforth referred to as mixed) satisfies an additional assumption of errors (ϵ_t) being linear functions of cross-sectionally independent sources. The Structural SVAR model is defined by:

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + D u_t, \tag{2.3}$$

where the components $(u_{1t}), (u_{2t})..., (u_{mt})$ (called the sources, later in the text) are serially i.i.d. and are also cross-sectionally independent of one another. The mixing matrix D is an invertible $(m \times m)$ matrix [5] The cross-sectional independence condition on u_t is necessary to identify the matrix D in addition to parameters $\Phi_1, ..., \Phi_p$, and the nonlinear causal innovations introduced in Section 4. The SVAR model is a semi-parametric model with the true distribution P_0 depending on the true parameters $\Phi_{10}, ..., \Phi_{p0}, D_0$ and the true functional parameters of source distributions, i.e. the true densities $f_{0,j}, j = 1, ..., m$ of the sources. Alternatively, the SVAR model can have a parametric specification when the distributions of sources are assumed to be $f_{0,j}(u) = f_{0,j}(u, \gamma_{j,0})$, where $\gamma_{j,0}$ is the true value of parameter γ_j.

Identification Issues in SVAR Models

The classical normality-based ML and least squares estimation methods are applicable to the traditional causal SVAR models. They are flawed in application to the mixed (S)VAR models because they are based on the first and second-order moments, i.e. the means, variances and autocovariances/autocorrelations, which is equivalent to assuming Gaussian error distributions. As mentioned earlier, the causal and noncausal dynamics cannot be distinguished at second-order. More precisely,

i) The roots of the characteristic equation of the autoregressive polynomial matrix cannot be distinguished from their reciprocals.
In the mixed VAR model (2.1) and SVAR model (2.3), the assumption of non-Gaussian errors is crucial for the identification of $\Phi_1, ..., \Phi_p$.

ii) Matrix D is not identifiable
In a mixed SVAR model (2.3), and additional identification issues arise when the errors $\varepsilon_t = Du_t$ are Gaussian. Then, if u_t, $t = 1, ..., T$ are standard normal variables, $u_t \sim N(0, Id)$, where Id denotes the identity matrix, we can identify the error distribution $\varepsilon_t \sim N(0, DD') = N(0, \Sigma)$ and then the product DD', but not the matrix D itself.

In a traditional causal SVAR model with Gaussian errors, this identification issue can be solved by imposing identifying restrictions, such as the shock "ordering" implemented through the Cholesky decomposition [Sims (1980), (2002)] and criticized e.g. by Lutkepohl

[5] A causal SVAR model with a noninvertible matrix D (because of the number of observed variables being larger than the number of shocks) is considered in Cordoni, Corsi (2019).

(1990), or by imposing additional restrictions such as the equality restrictions [Rubio-Ramirez, Waggoner , Zha (2010)], sign restrictions [Granziera, Moon, Schorfheide (2018)], or long-run restrictions [see Blanchard, Quah (1989), Cochrane (1994), Section 2.1.5, Faust, Leeper (1997), Leeper et al. (2013)].

It follows from the Independent Component Analysis (ICA) literature [Comon (1994), Erikson, Koivunen (2004)] and the existence of two-sided moving average representation [Chan, Ho, Tong (2006)] that the two identification issues i) and ii) can be solved in the SVAR model simultaneously without imposing the above identifying restrictions if the components $u_{j,t}$ of the source u_t are non-Gaussian and independent of one another [Gourieroux, Monfort, Renne (2017), Velasco, Lobato (2018), Velasco (2022)]. Then, one can identify the autoregressive coefficient matrices $\Phi_1, ..., \Phi_p$ and matrix D. It is also possible to identify non-parametrically (and then parametrically) the distributions of sources $u_{j,t}, j = 1, ..., m$. Moreover, their independence determining the structural characteristic of the VAR model can be tested. The exact conditions on the distributions of the components of u_t, allowing us to solve the identification issues, are reviewed in the on-line Appendix B.

3 Predictive Density

Let us consider a VAR(p) process of dimension m:

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t, \tag{3.1}$$

where the error vectors ε_t are i.i.d. and have a continuous distribution with probability density function g. We assume that the roots of:

$$\det(Id - \Phi_1 z - \cdots - \Phi_p z^p) = 0, \tag{3.2}$$

are not on the unit circle. Then, there exists a unique stationary solution of model (3.1) with a two-sided moving average representation in ε_t.

By stacking the present and lagged values of process (Y_t), model (3.1) can be rewritten as a $n = mp$ multidimensional VAR(1) model:

$$\begin{pmatrix} Y_t \\ \tilde{Y}_{t-1} \end{pmatrix} = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = \Psi \begin{pmatrix} Y_{t-1} \\ \tilde{Y}_{t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}, \tag{3.3}$$

with

$$\Psi = \begin{pmatrix} \Phi_1 & \cdots & \cdots & \Phi_p \\ Id & 0 & \cdots & \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & Id & 0 \end{pmatrix}. \tag{3.4}$$

The eigenvalues of the autoregressive matrix Ψ are the reciprocals of the roots of the characteristic equation (3.2).

Matrix Ψ has a real Jordan representation:

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

where J_1 (resp. J_2) are real $(n_1 \times n_1)$ (resp. $(n_2 \times n_2)$) matrices where $n_2 = n - n_1$ with all eigenvalues of modulus strictly less than 1 [resp. strictly larger than 1] and A is a $(n \times n)$ invertible matrix [see Perko (2001), Gourieroux, Jasiak (2016), Section 5.2, Davis, Song (2020), Remark 2.1. for real Jordan representations]. Then, equation (3.3) can be rewritten as:

$$A^{-1} \begin{pmatrix} Y_t \\ \tilde{Y}_{t-1} \end{pmatrix} = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1} \begin{pmatrix} Y_{t-1} \\ \tilde{Y}_{t-2} \end{pmatrix} + A^{-1} \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}. \tag{3.5}$$

Let us introduce a block decomposition of A^{-1}:

$$A^{-1} = \begin{pmatrix} A^1 \\ A^2 \end{pmatrix}, \tag{3.6}$$

where A^1 is of dimension $(n_1 \times n)$ and the transformed variables:

$$Z_t = \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix} = A^{-1} \begin{pmatrix} Y_t \\ \tilde{Y}_{t-1} \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} = A^{-1} \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}. \tag{3.7}$$

Then, we get two sets of latent components of process Y_t, written as the following functions of current and lagged values of (Y_t):

$$\begin{aligned} Z_{1,t} &= J_1 Z_{1,t-1} + \eta_{1,t}, \\ Z_{2,t} &= J_2 Z_{2,t-1} + \eta_{2,t}. \end{aligned} \tag{3.8}$$

The first set of equations in system (3.8) defines a causal VAR(1) model. Thus, process $(Z_{1,t})$ has the causal MA(∞) representation:

$$Z_{1,t} = \sum_{j=0}^{+\infty} J_1 \eta_{1,t-j}, \tag{3.9}$$

with $\eta_{1,t}$ as the causal innovation of $Z_{1,t}$.

The second set of equations in system (3.8) needs to be inverted to obtain a MA representation in matrices with eigenvalues of modulus strictly less than 1. We get:

$$\begin{aligned} Z_{2,t} &= J_2^{-1} Z_{2,t+1} - J_2^{-1} \eta_{2,t+1}, \\ &= -\sum_{j=0}^{+\infty} [J_2^{-j-1} \eta_{2,t+j+1}]. \end{aligned} \tag{3.10}$$

Therefore, $(Z_{2,t})$ is a noncausal process with a one-sided moving average representation in future values $\varepsilon_{t+1}, \varepsilon_{t+2}, ...,$. This noncausal component is also a linear function of future values of $\eta_{2,t}$ and has nonlinear dynamics in calendar time. Variable η_{2t} is the noncausal innovation of Z_{2t}.

The expression of the predictive density of Y_{T+1} given $\underline{Y}_T = (Y_T, Y_{T-1}, ...)$ is given below and derived in Appendix 1.

Proposition 1: The conditional probability density function (pdf) of Y_{T+1} given $\underline{Y}_T$ is:

$$l(y|\underline{Y}_T) = \frac{l_2\left[A^2 \begin{pmatrix} y \\ \tilde{Y}_T \end{pmatrix}\right]}{l_2\left[A^2 \begin{pmatrix} Y_T \\ \tilde{Y}_{T-1} \end{pmatrix}\right]} \, |\det J_2| \, g(y - \Phi_1 Y_T - \cdots - \Phi_p Y_{T-p+1}), \tag{3.11}$$

where $l_2(z_2)$ is the stationary pdf of $Z_{2,t} = A^2 \begin{pmatrix} Y_t \\ \tilde{Y}_{t-1} \end{pmatrix}$, if $n_2 \geq 1$. In the pure noncausal process, $n_2 = 0$, we have:

$$l(y|\underline{Y}_T) = g(y - \Phi_1 Y_T - \cdots - \Phi_p Y_{T-p+1}).$$

In the special case of VAR(1) process with $p = 1$, the predictive density becomes:

$$l(y|\underline{Y}_T) = l(y|Y_T) = \frac{l_2(A^2 y)}{l_2(A^2 Y_T)} \,|\det J_2|\, g(y - \Phi Y_T). \tag{3.12}$$

Proof: See Appendix 1.

This predictive density is constrained by the VAR representation (3.1). It is a semi-parametric function of parameters $\Phi_1, ..., \Phi_p$, determining A^2 and J_2, and of functional parameters g, l_2. These parameters are identifiable in the SVAR model (2.3) with non-Gaussian sources.

In the SVAR model (2.3), the density $g(\varepsilon)$ can be determined from the joint density $f(u) = \prod_{j=1}^{m} f_j(u_j)$ of the sources by the Jacobian formula. We have:

$$g(\varepsilon) = (1/|det\, D|) f_0(D^{-1}\varepsilon).$$

Corollary 1: Markov Property The mixed causal-noncausal process is a Markov process of order p in calendar time for $n_2 \geq 1$.

This corollary extends to linear mixed causal-noncausal processes of any autoregressive order, the result of Cambanis, Fakhre-Zakeri (1994), who show that a linear pure noncausal autoregressive process of order 1 is a causal Markov process of order 1.

Since the mixed VAR process of order p is Markov of order p in both calendar and reverse time, we can derive the closed-form expression of the backward predictive density in reverse time for backcasting.

Corollary 2: Backcasting Let us consider a mixed VAR(1) model. The backward predictive density of Y_{T-1} given Y_T is:

$$l_B(y|Y_T) = \frac{l_1(A^1 y)}{l_1(A^1 y_T)} |det\ J_2|\ g(Y_T - \Phi y),$$

where l_1 is the stationary density of z_{1t} and g is the density of ϵ.
Proof: See Appendix 1.

By considering jointly Proposition 1 and Corollary 2, we can see that the mixed VAR models have a nonlinear dynamic structure that allows for extending to nonlinear framework the standard Kalman filter for linear Gaussian processes.

There exists a multiplicity of real Jordan representations and of matrices A built from "extended" eigenspaces. However, $\det J_2 = \prod_{j=1}^{n_2} \lambda_j$, where $|\lambda_j| > 1$, $j = 1, ..., n_2$, is independent of the real Jordan representation. Similarly, the noncausal component Z_2 is defined up to a linear invertible transformation. Since the Jacobian is the same for the numerator and denominator of the ratio $\frac{l_2\left[A^2\begin{pmatrix} y \\ \tilde{Y}_T \end{pmatrix}\right]}{l_2\left[A^2\begin{pmatrix} Y_T \\ \tilde{Y}_{T-1} \end{pmatrix}\right]}$, it has no effect on the ratio. Thus, the expression of $l(y|\underline{Y}_T)$ does not depend on the selected real Jordan representation.

In the VAR(1) model, we have $dim\, Z_t = dim\, Y_t$ and $Y_t = AZ_t$. Therefore, each component of Y_t is a linear combination of the latent causal and noncausal components $Z_{1,t}, Z_{2,t}$, respectively. When $p = 2$, there exist more latent causal and non-causal components.

A mixed VAR process has nonlinear causal dynamics, which is captured by the predictive density. In this nonlinear and non-Gaussian framework, the predictive density provides the oos prediction intervals at various horizons, replacing the linear pointwise predictions and prediction intervals.

The closed-form expression of the predictive density is available at horizon 1. It can be written at horizon h as a multivariate integral over h future values of the process and estimated in practice as follows: The closed-form predictive density at horizon 1 allows us to perform drawings of future values of the process, by using the Sampling Importance Resampling (SIR) method to transform the predictive density into a multivariate cumulative distribution function (c.d.f)[Smith, Gelfand (1992), Tanner (1993)] (see Section 6.1.4 for an illustration). To estimate the predictive density at horizon $h > 1$ we need S independent future paths of the process. Each path $s = 1, ..., S$ is obtained by forecasting sequentially $Y^s_{T+1}|Y_T$, followed by $Y^s_{T+2}|Y^s_{T+1}, ..., Y^s_{T+h}|Y^s_{T+h-1}$, from the predictive densities. This provides a drawing $Y^s_{T+1}, ..., Y^s_{T+h}$ of a future path s. By replicating independently for $s = 1, ..., S$, we get $Y^s_{T+h},\ s = 1, ..., S$ and can use their sample distribution as an estimator of the predictive distribution at horizon h.

4 Nonlinear Causal Innovations

Let us recall that the SVAR models (2.3) are VAR models (2.1) satisfying the additional assumption of errors being linear functions of cross-sectionally independent sources. In macroeconomics and finance, the inference on the traditional causal SVAR model commonly includes the impulse response functions (IRF) and variance decompositions, which are important tools for the economists and financial policy makers. This inference needs to be consistent with the structural interpretation of the model. More specifically, the "structural" IRFs have to be based on shocks applied to the "structural" nonlinear causal innovations that need to be identified. These causal innovations have to satisfy both the serial and cross-sectional independence conditions [Gourieroux, Jasiak (2005), Gourieroux, Monfort, Renne (2017), Gonzalves, Herrera, Kilian, Pesavento (2022)]. As mentioned earlier, errors ε_t of mixed VAR models (or the sources u_t of mixed SVAR models) cannot be interpreted as innovations and then used to define the shocks and IRFs. We discuss below the filtering and identification of nonlinear causal innovations in the framework of mixed SVAR models.

4.1 Definition of Nonlinear Innovations

Several types of disturbances can be defined from the mixed causal-noncausal (S)VAR models, such as the model errors (ε_t), (sources (u_t)) and the causal and noncausal latent component errors $(\eta_{1,t})$ and $(\eta_{2,t})$, respectively. These disturbances have closed form expressions in the parameters and observations and can be easily filtered. Among them, only $(\eta_{1,t})$ can be interpreted as causal innovations, being the causal innovations of $Z_{1,t} = A^1 Y_t$.

The remaining errors $(\eta_{2,t})$ and components of (ε_t) (sources (u_t)) do not have causal interpretation. Therefore, they should not be used directly to define and examine the impulse response functions [Davis, Song (2020), Figure 1.7].

The nonlinear causal innovations are defined below for any pure noncausal or mixed (S)VAR model.

Definition 1: A nonlinear causal innovation is a process (v_t) of dimension m such that:

i) the vectors v_t are serially i.i.d.

ii) the strictly stationary process (Y_t) can be written in a nonlinear autoregressive form:

$$Y_t = a(\tilde{Y}_{t-1}, v_t). \tag{4.1}$$

It is easy to see that this condition is equivalent to the Markov of order p property of (Y_t) with a continuous distribution, and can be applied to the mixed model (3.1) by Corollary 1. The nonlinear causal innovations follow from the Volterra representation of a multivariate strictly stationary process, providing the basis of nonlinear impulse response functions [Potter (2000), Gourieroux, Jasiak (2005)]. If function a is invertible with respect to v, then v_t is a nonlinear function of Y_t given its past.

The Markov dynamics has to be compatible with the (S)VAR model. Therefore, the Markov transition probability derived in Proposition 1 is constrained and depends on the matrix parameters $\Phi, (D)$ and the common distribution of independent errors ε_t (sources u_t). The following section discusses the identification of nonlinear causal innovations, referred to as nonlinear innovations for simplicity.

4.2 Identification of Nonlinear Innovations

The nonlinear innovations are not defined in a unique way, even under an additional cross-sectional independence condition of components $(v_{1,t}), ..., (v_{m,t})$, which is required to appropriately define the notion of a shock to one innovation $v_{2,t}$, say, which has no effect on the remaining components of v_t. This identification issue is examined in Appendix 3 in a functional framework. In particular, we get the following Proposition:

Proposition 2:

In a VAR(1) process, the dimension of under-identification in the functional space of nonlinear innovations is finite and equal to $2m$.

Therefore, the identification of nonlinear innovations and then of structural shocks is an issue in the mixed (S)VAR models, similarly to the linear causal Gaussian VAR models. However, by using the insights from the traditional causal VAR, it is easy to derive the innovations of a mixed (S)VAR by extending the Sim's approach of shock "ordering" to nonlinear dynamic framework. For example, for $m = 2$, we can suppose that the first variable to be shocked is Y_1. Then, we consider the conditional cumulative

distribution function (cdf) F_1 of $Y_{1,t}$ given Y_{t-1} and define $v_{1,t} = \Phi^{-1}F_1(y_{1,t}|y_{t-1})$, where Φ is the cdf of the standard Normal distribution. Then $v_{1,t}$ is a standard Gaussian white noise. Next, we consider the conditional cdf $F_{2|1}$ of $Y_{2,t}$ given $Y_{1,t}$, Y_{t-1} and define $v_{2,t} = \Phi^{-1}F_{2|1}(y_{2,t}|y_{1,t}, y_{t-1})$. This provides us another standard Gaussian white noise, which is independent of $(v_{1,t})$, and as an outcome of the procedure, we get a causal innovation v_t with independent components.

Proposition 2 implies that there exists a multiplicity of Gaussian nonlinear innovations that are not computed recursively, in general (see Appendix 3).

Additional restrictions could be introduced either on parameters $\Phi_i, i = 1, ..., p$, (D), or on the error (source) distributions to identify the Gaussian innovations and eliminate their multiplicity.

Nevertheless, the applications of mixed models indicate that the noncausal order is often equal to 1, i.e. $n_2 = 1$ [see e.g. Gourieroux, Jasiak (2017), Hecq, Lieb, Telg (2016), and Section 6.3 of this paper]. This is likely due to the type of nonlinear dynamics generated by noncausal roots capturing the speculative bubbles and local trends. For example, in macroeconomic models, the noncausal component can be related to speculative bubbles in oil prices impacting jointly the GDP and other macroeconomic variables (see the illustration in Section 6.2).

In practice it is more insightful to examine the consequences of a structural shock to the latent noncausal component (Z_2), which determines the nonlinear dynamics of the model rather than to the observed variables. In the bivariate SVAR model, this can be done by applying shock "ordering" to the pure causal and noncausal components (Z_t) and selecting Z_2 as the first variable to be shocked. Using obvious notation, we can proceed as follows:

$$v_{2,t}(Z_t) = \Phi^{-1}[F_2(Z_{2,t}|\underline{Y}_{t-1})] = \Phi^{-1}[F_2(Z_{2,t}|Z_{t-1})], \tag{4.2}$$

where F_2 is the conditional cumulative distribution function (c.d.f) of $Z_{2,t}$. From equation (a.5), it follows that the conditional density of $Z_{2,t}$ given Z_{t-1} has a closed form given by:

$$l(z_{2,t}|z_{t-1}) = \frac{l_2(z_{2,t})}{l_2(z_{2,t-1})} \, |det \, J_2| \, g_{\eta_2}(z_{2,t} - J_2 \, z_{2,t-1}), \tag{4.3}$$

where g_{η_2} is the marginal density of η_2. The c.d.f. F_2 is the integral of $l(z_{2,t}|z_{t-1})$ over the

set of admissible values of $z_{2,t}$.

Next, we can append it with:

$$v_{1,t}(Z) = \Phi^{-1}[F_{1|2}(Z_{1,t}|Z_{2,t}, Z_{t-1})], \tag{4.4}$$

where $F_{1|2}$ is the conditional cumulative distribution function of $Z_{1,t}$ given $Z_{2,t}, Z_{t-1}$. From equation (a.5), we get the closed-form expression of the conditional density:

$$l(z_{1,t}|z_{2,t}, z_{t-1}) = l(z_t|z_{t-1})/l(z_{2,t}|z_{t-1}) = \frac{g_\eta(z_{1,t} - J_1 z_{1,t-1},\ z_{2,t} - J_2 z_{2,t-1})}{g_{\eta_2}(z_{2,t} - J_2 z_{2,t-1})}. \tag{4.5}$$

The c.d.f. $F_{1|2}$ is the integral of $l(z_{1,t}|z_{2,t}, z_{t-1})$ over the set of admissible values of $z_{1,t}$.

Since the choice of the variable to be shocked, i.e. Z_2 defines the selected "ordering" it is easy to check that $v_{1,t}(Z)$ differs in general from the causal innovation $\eta_{1,t}$ to the latent causal component $Z_{1,t}$.

Under the parameter identification condition of Section 2 and given a specific choice of a Gaussian nonlinear innovation, the nonlinear residuals $\hat{v}_{1,t}, \hat{v}_{2,t}$ can be computed after replacing the conditional cdfs by their consistent nonparametric estimates [6]. Next, these residuals can be examined in residual plots and used for computing specification test statistics.

5 Post-Estimation Inference

The parameters of a (S)VAR model need to be estimated before the forecasts and nonlinear causal innovations are computed. The first part of this Section reviews the estimation methods that exist in the literature and describes the prediction/filtering algorithm leading to the estimated nonlinear causal innovations. Next, we introduce a post-estimation inference method for assessing the prediction interval uncertainty due to the preliminary estimation step.

5.1 Estimation and Filtering

The mixed (S)VAR model can be estimated by the maximum likelihood method based on an assumed parametric distribution of ε_t, or u_t [see Breidt et al. (1991), Lanne, Saikkonen

[6] or parametric estimates if the error (source) distribution is parametric.

(2010), (2013), Davis, Song (2020), Bec et al. (2020)]. This approach yields consistent estimators provided that the parametric distributional assumption is valid.

Alternatively, the mixed (S)VAR model can be consistently estimated without any parametric assumptions on the distribution of the sources by using the (Generalized) Covariance (GCov) estimator [Gourieroux, Jasiak (2022)]. The GCov estimator is consistent, asymptotically normally distributed and semi-parametrically efficient [7]. An alternative covariance estimator has been used in Gourieroux, Monfort, Renne (2017), (2020), and Gourieroux, Jasiak (2017). In addition, minimum distance estimators based on the cumulant spectral density of order 3 and 4 have been proposed in Velasco, Lobato (2019) and Velasco (2022). A two-step application of the continuum method of moments is developed in Starck (2023). The method of moment estimators provide consistent estimators only in applications to the SVAR models, i.e. to the VAR models satisfying the additional cross-sectional independence condition [see, Guay (2021)], and are inconsistent otherwise [Lanne, Luoto (2021)].

The prediction and filtering methods can be applied in a parametric or semi-parametric framework. In the semi-parametric framework, it can be applied along the following lines:
step 1. Apply the GCov estimator based on zero auto-covariance conditions of nonlinear error functions to obtain the estimators of matrices of autoregressive coefficients $\hat{\Phi}_1, ..., \hat{\Phi}_p$.
step 2. Use the $\hat{\Phi}_i, i = 1, ..., p$, estimates to compute the roots of the lag-polynomial and more generally an estimated Jordan representation: $\hat{A}, \hat{J}_1, \hat{J}_2$.
step 3. Compute the approximated model errors using the estimates obtained in Step 1: $\hat{\varepsilon}_t = Y_t - \hat{\Phi}_1 Y_{t-1} - ... - \hat{\Phi}_p Y_{t-p}$.
step 4. Compute $\hat{Z}_t = \hat{A}^{-1} \begin{pmatrix} Y_t \\ \tilde{Y}_{t-1} \end{pmatrix}$, $\hat{\eta}_t = \hat{A}^{-1} \begin{pmatrix} \hat{\varepsilon}_t \\ 0 \end{pmatrix}$.
step 5. An Independent Component Analysis (ICA) analysis can be conducted by applying the GCov estimator based on cross-sectional zero covariance conditions of nonlinear functions of $\hat{\varepsilon}_t$ to obtain in two-steps a consistent estimate $\hat{D}$ of matrix D, and next the approximations $\hat{u}_t$ of sources u_t [8].
step 6. The following densities can be estimated by kernel estimators applied to the approximated series:

[7] See Gourieroux, Jasiak (2022) for regularity conditions.

[8] The matrices $\Phi_i, i = 1, ..., p$ and D can be also estimated in one step from a Gcov combining the serial and cross-sectional zero covariance conditions.

- the density g of ε_t can be estimated from $\hat{\varepsilon}_t,\ t = 1, ..., T$;

- the density l_2 of $Z_{2,t}$ can be estimated from $\hat{Z}_{2,t},\ t = 1, ..., T$;

- the densities of errors $u_{j,t},\ j = 1, ..., m$ can be estimated from the empirical distribution of $\hat{u}_{j,t},\ j = 1, ..., m,\ t = 1, ..., T$;

step 7. The predictive density can be estimated by replacing in the formula of Proposition 2 l_2 by $\hat{l}_2$, A^2 by $\hat{A}^2$, and also J_2 by $\hat{J}_2$, g by $\hat{g}$, and $\Phi_1, ..., \Phi_p$ by $\hat{\Phi}_1, ..., \hat{\Phi}_p$. The mode (median) of the predictive density provides the point forecasts and the quantiles of the predictive density can be used to obtain prediction intervals at horizon 1.

step 8. The nonlinear causal residuals $\hat{v}_{2,t}(Z)$ can be computed from the estimated distributions as $\hat{v}_{2,t}(Z) = \Phi^{-1}(\hat{F}_{2,T}(\hat{Z}_{2,t}|\hat{Z}_{t-1}))$ by applying the formula of predictive density (4.3) with l_2 replaced by $\hat{l}_2$ and g_{η_2} replaced by $\hat{g}_{\eta_2}$, i.e. the empirical density of $\hat{\eta}_{2,t},\ t = 1, ..., T$.

step 9. Next, they can be appended by the causal innovations independent of $\hat{v}_{2,t}(Z)$ such as

$\hat{v}_{1,t} = \Phi^{-1}(\hat{F}_{1|2,T}(\hat{Z}_{1,t}|\hat{Z}_{2,t}, \hat{Z}_{t-1}))$.

These causal innovations can be computed from the predictive density formula (4.5) with g_{η_2} and g_η replaced by their empirical counterparts.

5.2 Uncertainty of the Estimated Prediction Set

The estimated model parameters and residuals $\hat{\varepsilon}_t$ can be used to build oos forecast intervals conditional on given values of the last observation in the sample, called the conditional prediction interval. The estimation error associated with the scalar and functional parameters has an effect on the uncertainty of the conditional prediction interval. For ease of exposition, let us consider the VAR(1) model, forecast horizon $h = 1$ and the future value of the first component series $Y_{1,T+1}$ to be forecast at date T out of sample (oos) given $Y_T = (y_{1,T}, y_{2,T})'$. Then, the true prediction interval at level $1 - \alpha_1$ for $Y_{1,T+1}$ is:

$$PI(y, \alpha_1) = [Q_l(y, \alpha_1; P_0),\ Q_u(y, \alpha_1; P_0)], \tag{5.1}$$

where P_0 is the true predictive density function of $Y_{1,T+1}$ conditional on $Y_T = y$ and derived from the joint multivariate predictive density $l(y|\underline{Y}_T)$ (see, Proposition 1 for the closed-form expression of the predictive density). Let Q_l and Q_l denote the true $\alpha_1/2$ and

$1-\alpha_1/2$ conditional quantiles of P_0, respectively. Recall that under the semi-parametric approach, the true distribution P_0 is characterized by Φ_0 and g_0. By using the expression of the prediction interval in a Gaussian framework, the asymptotically valid prediction interval for $Y_{1,T+1}$ can be equivalently written as:

$$PI(y,\alpha_1) = [m(y,\alpha_1;P_0) \pm \Phi^{-1}(\alpha_1/2)\,\sigma(y,\alpha_1;P_0)], \tag{5.2}$$

with [9]

$$m(y,\alpha_1;P_0) = 0.5[Q_l(y,\alpha_1 P_0) + Q_u(y,\alpha_1;P_0)], \tag{5.3}$$

$$\sigma(y,\alpha_1;P_0) = -[1/(2\Phi^{-1}(\alpha_1/2)][Q_u(y,\alpha_1;P_0) - Q_l(y,\alpha_1;P_0)]. \tag{5.4}$$

This representation resembling the traditional Gaussian approach can be used even if the conditional density function of (Y_t) is not Gaussian. In particular, the functions $y \to m(y,\alpha_1;P_0), y \to \sigma(y,\alpha_1;P_0)$ are nonlinear in this case, in general.

In our framework, the unknown marginal predictive density function P_0 can be consistently estimated from its expression (3.11) as $\hat{P}$, given the estimated matrix of autoregressive parameters $\hat{\Phi}$ and the residuals $\hat{\varepsilon}_t$ obtained from the semi-parametric GCov estimator that provides the nonparametric estimator $\hat{g}$. Then, we can compute an estimated prediction interval of $Y_{1,T+1}$:

$$\widehat{PI}(y,\alpha_1) = [m(y,\alpha_1;\hat{P}) \pm \Phi^{-1}(\alpha_1/2)\,\sigma(y,\alpha_1;\hat{P})] = [Q_l(y,\alpha_1;\hat{P}), Q_u(y,\alpha_1;\hat{P})]. \tag{5.5}$$

where $Q_l(y,\alpha_1;\hat{P})$ and $Q_u(y,\alpha_1;\hat{P})$ are the $\alpha_1/2$ and $1-\alpha_1/2$ conditional quantiles of the estimated predictive density $\hat{P}$.

This estimated prediction interval (5.5) is consistent of the true prediction interval (5.2) when the number of observations tends to infinity. We need to distinguish:

a) the true prediction interval $PI(y,\alpha_1)$ of $Y_{1,T+1}$ that satisfies:

$$P_0[Y_{1,T+1} \in PI(y,\alpha_1) | Y_T = y] = 1-\alpha_1, \ \ \forall y.$$

[9] Note that $\Phi^{-1}(\alpha_1/2)$ is negative.

By construction, the true prediction interval has the correct conditional coverage probability of $1-\alpha_1$. However, it depends on the unknown true distribution, or equivalently on Φ_0, g_0.

b) The estimated prediction interval $\widehat{PI}(y,\alpha_1)$ that does not satisfy the conditional coverage condition:

$$P_0[Y_{1,T+1} \in \widehat{PI}(y,\alpha_1)|Y_T = y] \neq 1-\alpha_1, \ \ \forall y.$$

due to the estimation errors.

Since this estimated prediction interval is random, its asymptotic distribution can be approximated by bootstrap, which is applied by replicating the trajectory of the process by backcasting. More precisely, given $\hat{\Phi}$, $\hat{g}$ and the residuals, we can generate by backcasting the artificial data $Y_t^s, \ t = 1, ..., T$ such that, $Y_T^s = Y_T, \forall s = 1, ..., S$, i.e. with the same terminal condition $Y_T = y$ for all the bootstrapped samples, i.e. paths replicated by backcasting (see, Corollary 2 for the closed-form expression of the backward predictive distribution). The backcasting can be performed as the following sequence of one-step backcasts: starting from Y_T, we backcast Y_{T-1}^s, conditional on Y_T, next we backcast Y_{T-2}^s conditional on Y_{T-1}^s, and so on. By replicating the backcasted path S times, we end up generating S bootstrapped series $Y_t^s, t = 1, ..., T$ of length T equal to the length of the initial series and with the same terminal value Y_T. Next, from each replicated path $(Y_t^s, t = 1, ..., T)$ we estimate the model parameters Φ^s and $\hat{g}^s$, $s = 1, ..., S$. That allows us for computing at Y_T the predictive density estimator $\hat{P}^s, s = 1, ..., S$ of Y_{T+1} given Y_T and S new prediction intervals of $Y_{1,T+1}$ from each of the replicated paths.

The bootstrapped prediction interval obtained from a replicated path is:

$$\widehat{PI}^s(y,\alpha_1) = [m(y,\alpha_1;\hat{P}^s) \pm \Phi^{-1}(\alpha_1/2)\sigma(y,\alpha_1;\hat{P}^s)], \tag{5.6}$$

where $\hat{P}^s$ is the semi-parametric estimate of P_0 from the artificial path $(Y_t^s), t = 1, ..., T$. Given $\hat{P}^s, s = 1, ..., S$ of Y_{T+1} this bootstrapped PI of $Y_{1,T+1}$ given Y_T can be replicated independently $s = 1, ..., S$ times.

The components of the prediction interval (5.6) are denoted by:

$$\hat{m}^s(y,\alpha_1) = m(y,\alpha_1;\hat{P}^s), \ \ \hat{\sigma}^s(y,\alpha_1) = \sigma(y,\alpha_1;\hat{P}^s), \ s = 1, ..., S. \tag{5.7}$$

For large S, the joint sample distribution of $[\hat{m}^s(y,\alpha_1), \hat{\sigma}^s(y,\alpha_1)]$ provides an approximation of the distribution of $[m(y,\hat{P}), \sigma(y,\hat{P})]$, when T is large.

5.3 Confidence Interval of the Prediction Interval

The estimated prediction interval $\widehat{PI}(y,\alpha_1)$ is a pointwise estimator of $PI(y,\alpha_1)$. Let us now extend the method of pointwise estimation to confidence set estimation. Since there does not exist a total ordering on intervals, we first constrain the confidence set to be of the form:

$$\widehat{PI}(y,\alpha_1,q) = [m(y,\alpha_1;\hat{P}) \pm q\,\sigma(y,\alpha_1;\hat{P})]. \tag{5.8}$$

This confidence set has the following coverage probability of the true prediction interval:

$$\begin{aligned}
\Pi_0(y,\alpha_1;q) &= P_0[\widehat{PI}(y,\alpha_1,q) \supset PI(y,\alpha_1)|Y_T = y] \\
&= P_0[m(y,\alpha_1;\hat{P}) - q\,\sigma(y,\alpha_1;\hat{P}) < m(y,\alpha_1;P_0) + \Phi^{-1}(\alpha_1/2)\,\sigma(y,\alpha_1;P_0), \\
&\quad m(y,\alpha_1;\hat{P}) + q\,\sigma(y,\alpha_1;\hat{P}) > \\
&\quad m(y,\alpha_1;P_0) - \Phi^{-1}(\alpha_1/2)\,\sigma(y,\alpha_1;P_0)|Y_T = y],
\end{aligned} \tag{5.9}$$

because $\Phi^{-1}(\alpha_1/2) < 0$. For $\alpha_2 \in (0,1)$, possibly different from α_1, there exists a value $q_0(y,\alpha_1;\alpha_2)$ such that:

$$\Pi_0[y,\alpha_1;q_0(y,\alpha_1;\alpha_2)] = 1-\alpha_2. \tag{5.10}$$

Asymptotically, we get the $1-\alpha_2$ coverage probability of the true conditional prediction interval, although the true predictive density P_0 and the true distribution of $\hat{P}$ remain unknown.

Therefore, equations (5.9) and (5.10) can be replaced by their bootstrapped counterparts obtained from S replicated paths of the series. More precisely, the bootstrapped conditional coverage probability is defined as:

$$\hat{\Pi}^s(y,\alpha_1,q) = \frac{1}{S}\sum_{s=1}^{S}\delta^s,$$

where

$$\delta^s = \begin{cases} 1, & \text{if } \hat{m}^s(y,\alpha_1) - q\,\hat{\sigma}^s(y,\alpha_1) < \hat{m}(y,\alpha_1) + \Phi^{-1}(\alpha_1/2)\,\hat{\sigma}(y,\alpha_1), \\ & \text{and } \hat{m}^s(y,\alpha_1) + q\,\hat{\sigma}^s(y,\alpha_1) > \hat{m}(y,\alpha_1) - \Phi^{-1}(\alpha_1/2)\,\hat{\sigma}(y,\alpha_1), \\ 0, & \text{otherwise.} \end{cases}$$

Then, we consider a solution $\hat{q}(y,\alpha_1,\alpha_2)$ of:

$$\hat{\Pi}^s[y,\alpha_1,\hat{q}(y,\alpha_1,\alpha_2)] = 1-\alpha_2. \tag{5.11}$$

ensuring the $1-\alpha_2$ coverage probability. The bootstrap confidence set of the prediction interval is:

$$\widehat{CSPI}(y,\alpha_1,\alpha_2) = \left\{ m(y,\alpha_1,\hat{P}) \pm \hat{q}(y,\alpha_1,\alpha_2)\sigma(y,\alpha_1,\hat{P}) \right\}. \tag{5.12}$$

This bootstrap confidence set is such that:

$$\lim_{T\to\infty}\lim_{S\to\infty} P_0[\widehat{CSPI}(y,\alpha_1,\alpha_2) \supset PI(y,\alpha_1)|Y_T = y] = 1-\alpha_2,\ \forall P_0. \tag{5.13}$$

Hence, the length of the estimated $\widehat{PI}(y,\alpha_1)$ is enlarged by a factor $\hat{q}(y,\alpha_1,\alpha_2)/|\Phi^{-1}(\alpha_1/2)|$, that depends on the observed value y, in general. In practice, we can choose $\alpha_1 = \alpha_2 = 0.05$ corresponding to the standard levels for prediction intervals and confidence intervals. One could also choose α_1 different from α_2, including $\alpha_1 = 1.0$, which would correspond to the confidence set of point prediction equal to the median of the predictive density.

The analysis of a confidence set of the prediction set is related to research on a confidence set of the identified set in models under partial identification. The partial identification literature considers a parametric model with a parameter γ, say, that is partly identifiable [10]. The objective is to determine either the confidence set under the classical approach, or the credible set under the Bayesian approach. In our framework, we have two types of "parameters", P and $Y_{1,T+1}$, say. The second one is not identifiable, although its conditional distribution is estimated, and plays the role of a conditional prior.

[10] See Imbens, Mansky (2004) for confidence intervals of identified intervals in the framework of partial identification and confidence intervals that asymptotically cover the true interval with a probability larger or equal to $1-\alpha_2$.

6 Illustration

We illustrate the nonlinear forecasts from the causal-noncausal VAR model. The first part of this Section, presents a simulation study that examines the out of sample forecasts from the model. Next, the semi-parametric GCov estimation and forecasts are illustrated in applications to the joint analysis of GDP growth rate and oil prices and a bivariate series of Bitcoin/USD and Ethereum/USD exchange rates.

6.1 Simulation Study

6.1.1 **The Artificial Data Set**

We consider a simulated trajectory of a bivariate causal-noncausal VAR(1) process with the following matrix of autoregressive coefficients:

$$\Phi = \begin{pmatrix} 0.7 & -1.3 \\ 0 & 2 \end{pmatrix},$$

with eigenvalues 0.7 and 2, located inside and outside the unit circle, respectively [11]. The errors follow a bivariate noise with independent components both t-student distributed with ν =4 degrees of freedom, mean zero and variance equal to $\nu/(\nu-2) = 2$. The true DGP is such that $D = Id$ and then ε_t is equal to the sources u_t. The matrix A is as follows:

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

The simulated paths of the series of length 600 are displayed in Figure 1. The solid (black) line represents process Y_{1t} and the dashed (red) line represents process Y_{2t}.

[Insert Figure 1: Trajectory of the Bivariate Causal-Noncausal VAR(1) Process]

Next, we calculate the error series defined as $\varepsilon_t = Y_t - \Phi_1 Y_{t-1}$. The bivariate series is plotted in Figure 2. The sequence of spikes in the noncausal component $Z_{2t} = Y_{2t}$ impacts the component Y_{1t} through the recursive form of matrix Φ.

[Insert Figure 2: Trajectory of Error Processes]

[11] This process was examined and illustrated in Gourieroux, Jasiak (2017).

The solid (black) line represents process ε_1 and the dashed (red) line represents process ε_2.

6.1.2 **Estimated Predictive Density, Point and Interval Forecasts**

Let us now consider forecasts based on the estimated model parameters. The Generalized Covariance (GCov) estimate of matrix Φ is obtained by minimizing the portmanteau statistic computed from the auto- and cross-correlations up to and including lag $H = 10$ of the errors $\varepsilon_t = Y_t - \Phi_1 Y_{t-1}$ and their squared values [see Gourieroux, Jasiak (2022)]. The following estimated autoregressive matrix is obtained:

$$\hat{\Phi} = \begin{pmatrix} 0.724 & -1.452 \\ -0.030 & 1.993 \end{pmatrix},$$

with eigenvalues $\hat{\lambda}_1 = 0.690$, $\hat{\lambda}_2 = 2.027$ close to the true values $\lambda_1 = J_1 = 0.7$ and $\lambda_2 = J_2 = 2.0$. The standard errors of $\hat{\Phi}$ obtained by bootstrap are 0.023, 0.308, for the elements of the first row, and 0.009, 0.120 for the elements of the second row.
After estimating Φ, the GCov estimated errors $\hat{\varepsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$ are computed and used for forecasting.
Matrices A and A^{-1} are identified from the spectral (Jordan) representation of matrix Φ, up to scale factors. The estimated matrix $\hat{A}^{-1}$ computed from the normalized Jordan decomposition of $\hat{\Phi}$ is

$\hat{A}^{-1} = \begin{pmatrix} 0.022 & 0.025 \\ -0.022 & 0.974 \end{pmatrix}$. It corresponds to matrix $A^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ up to scale factors of about 0.02 and 0.97 for each column.

We use these estimates to approximate the causal and noncausal components displayed in Figure 3.

[Insert Figure 3: Trajectory of Components: $\hat{Z}_1$: solid line, $\hat{Z}_2$: dashed line]

Let us now consider the oos nonlinear forecast one step ahead performed at date $T = 590$. At time T=590, the process takes values -3.367 and -0.239. The true values of Y_1 and Y_2 at T+1=591 are -2.260 and -0.331, respectively. The nonlinear forecasts are summarized by the theoretical predictive density that can be used to compute the pointwise predictions of $Y_{1,T+1}$ and $Y_{2,T+1}$. The predictive bivariate density is given in Figure 4 along with the predictive marginal densities of $Y_{1,T+1}$ and $Y_{2,T+1}$, respectively. The predictive density is estimated from formula (3.12) one-step ahead oos by using a

kernel estimator over a grid of 100 values below and above Y_1 and Y_2, with Gaussian kernels and bandwidths $h_2 = 1$ and $h_{11} = s.d.(\varepsilon_1), h_{12} = s.d.(\varepsilon_2)$ (see On-Line Appendix C for kernel density estimators).

[Insert Figure 4: Joint and Marginal Estimated Predictive Densities]

The estimated point forecast are obtained from the mode of the predictive density. The forecast of $Y_{1,591}$ based on the GCov estimated parameters is -2.80 and the point forecast of $Y_{2,591}$ is -0.30. The estimated prediction intervals at level 0.80 determined from the predictive density are as follows: The estimated prediction interval for $Y_{1,591}$ at level 0.80 is [-4.80, -0.80] and the prediction interval for $Y_{2,591}$ is [-2.60, 2.10]. The rationale for choosing level 80% is to ensure a sufficiently large number of observations in the tails for computing the quantiles of predictive density. Both prediction intervals contain the true future values of the process.
Next, we compute oos one-step ahead forecasts based on the sub-sample of 500 observations on Y_t. The series of 100 one-step ahead forecasts along the trajectory is displayed in Figure 5.

[Insert Figure 5: Estimation-Based One-Step Ahead Forecasts]

6.1.3 **Unconditional Coverage of the Estimated Prediction Interval**

We consider two data generating (DGP) processes. The first one is the bivariate VAR(1) illustrated in 6.1.1 with the autoregressive matrix:

$$\Phi = \begin{pmatrix} 0.7 & -1.3 \\ 0.0 & 2.0 \end{pmatrix},$$

eigenvalues $\lambda_1 = 0.7$ and $\lambda_2 = 2.0$ and an identity variance-covariance matrix of errors ε_t. In this experiment, the errors have t-student distributions with 3, 6 and 9 degrees of freedom, respectively. The second process is a bivariate VAR with the autoregressive matrix:

$$\Phi = \begin{pmatrix} 0.9 & -0.3 \\ 0.0 & 1.2 \end{pmatrix}$$

and eigenvalues $\lambda_1 = 0.9, \lambda_2 = 1.2$, which are closer to the unit circle [12]. The errors ε_t

[12]This process was examined and illustrated in Gourieroux, Jasiak (2022).

have an identity variance-covariance matrices and t-student distributions with 3, 6 and 9 degrees of freedom.

The simulated bivariate series are of length 100, 500 and 1000 and each DGP is replicated 500 times. The last observation form each simulated path is set aside. It is forecast and used for forecast coverage assessment. The autoregressive parameters $\hat{\Phi}$ are estimated by the GCov estimator from the errors $\varepsilon_t = Y_t - \Phi_1 Y_{t-1}$, their second, third and fourth powers, and their lags up to $H = 10$. The objective function maximizing algorithm is each time initiated at the starting values 0.1, 0.1, 0.5, 0.5. The oos predictive density given in (3.12) of $y(1000)$ is evaluated, given the parameter estimates, over a grid of 100 possible future values for each of the components series. We use Gaussian kernels and bandwidths $h_2 = s.d.(Y_{2,t})$ $h_{11} = s.d.(\varepsilon_1)$ and $h_{12} = s.d.(\varepsilon_2)$ (see On-Line Appendix C for kernel density estimators). The mode of the predictive density estimated from formula (3.12) with Gaussian kernels provides the point forecast. The oos prediction intervals at horizon 1 are obtained from the 10th and 90th percentiles of the bivariate predictive density. As before, the rationale for choosing level 80% is to ensure a sufficiently large number of observations in the tails for computing the quantiles of predictive density.

The unconditional coverage of the prediction interval is reported in Table 1 below for the two DGPs described above, sample sizes T=500 and T=1000 and t-student error ε_t distributions with 3, 6 and 9 degrees of freedom.

Table 1: Coverage of GCov Estimated Prediction Interval at 80%

	T=100			T=500			T=1000		
VAR(1) with eigenvalues 0.7, 2.0									
component	t(3)	t(6)	t(9)	t(3)	t(6)	t(9)	t(3)	t(6)	t(9)
$y_1(T+1)$	90.0	91.0	92.6	91.4	92.8	92.0	91.4	90.8	93.2
$y_2(T+1)$	92.2	93.8	91.8	93.4	94.0	94.8	95.2	93.2	94.5
VAR(1) with eigenvalues 0.9, 1.2									
component	t(3)	t(6)	t(9)	t(3)	t(6)	t(9)	t(3)	t(6)	t(9)
$y_1(T+1)$	80.2	82.8	87.4	82.2	86.0	84.2	84.6	84.4	88.6
$y_2(T+1)$	90.4	89.4	90.8	90.6	91.6	92.40	91.4	91.0	92.8

We observe that the coverage is either greater or close to the theoretical size of the prediction interval for both DGPs and sample sizes.

6.1.4 **Estimated Prediction Set Uncertainty**

This Section illustrates the conditional prediction interval uncertainty described in Section 5.2.

We consider the simulated trajectory of length T=200 of the bivariate VAR(1) process with autoregressive matrix:

$$\Phi = \begin{pmatrix} 0.9 & -0.3 \\ 0.0 & 1.2 \end{pmatrix}$$

with eigenvalues $\lambda_1 = 0.9, \lambda_2 = 1.2$ and t(6) distributed errors with an identity variance-covariance matrix.

We are interested in the conditional prediction interval out-of-sample of the first component $Y_{1,T+1} = Y_{1,200}$ given the past and current values of the two series.

The VAR(1) model is estimated by the GCov estimator from the observations $t = 1, ..., 199$ with four power transforms of model errors and lag $H = 10$, providing the following estimates of the autoregressive parameters: 0.8931, -0.2146, 0.0180, 1.2797, and the following estimates of eigenvalues: 0.903 and 1.269.

The estimated predictive density of Y_{200} conditional on Y_{199} is computed from formula 3.12 by using kernel smoothed density estimators. More specifically, we employ Gaussian kernels and bandwidths $h_2 = s.d.(Y_2), h_{11} = s.d.(\epsilon_1), h_{12} = s.d(\epsilon_2)$ (see On-Line Appendix C for kernel density estimators). For $\alpha_1 = 0.2$ and $\Phi^{-1}(\alpha_1/2) = -1.28$, the estimated prediction interval of $Y_{1,T+1} = Y_{1,200}$ is

$$\widehat{PI}(y, \alpha_1) = [-6.954, 0.360].$$

It contains the true value of $Y_{1,200} = -0.696$.

In the next step, we replicate the initial paths $S = 50$ times by backcasting from the bivariate terminal condition $Y_{199} = [-1.188, 0.473]$. We use the backcasting formula given in Corollary 2, evaluated from estimated model parameters and kernel density estimators. We employ Gausian kernels and bandwidths $h_1 = s.d.(Y_{1,t}), h_{11} = s.d.(\epsilon_1)$ and $h_{12} = s.d.(\epsilon_2)$.

The randomness of the backcasted paths is generated as follows: i) We first draw 100 values as if the components $Y_{1,T-1}, Y_{2,T-1}$ were independent conditional on Y_T. This is done by inverting the estimated conditional c.d.f.s of $Y_{1,T-1}$ and $Y_{2,T-1}$ given Y_T. This provides the sampling with the importance (misspecified) density; ii) Next, we re-sample

in the set of values obtained in step i) above, with the weights proportional to the ratio of the joint backward predictive density divided by the product of the two marginal backward predictive densities. This procedure adjusts for the omitted cross-sectional dependence in step i).

The parameters Φ and g are re-estimated from each replicated path, and then the values of $\hat{m}^s$, $\hat{\sigma}^s$ are computed from the quantiles of 50 predictive densities of Y_{T+1} conditional on Y_T. They are next used to compute the bootstrap confidence interval at level $\alpha_2 = 0.1$

$$\widehat{CSPI}(y, \alpha_1, \alpha_2) = [-11.839, 5.246],$$

with the solution $\hat{q}(y, \alpha_1, \alpha_2) = 2.99$. The interval $\widehat{CSPI}$ is much larger than the interval $\widehat{PI}$. This result illustrates the importance of taking into account the estimation risk on Φ, g when providing a prediction interval, especially that the estimator of the functional parameter g converges at a lower speed than the parameter estimator. The effect of estimation risk is twofold: i) the length of the interval has almost doubled, ii) the interval became less symmetric with respect to 0.

6.2 Application to Real Oil Prices and Real GDP Growth Rates

We examine the quarterly series of oil prices and US GDP growth rate over the period: Q1 1986 - Q2 2019. The real US GDP growth rate series is calculated from the series of Real Gross Domestic Product, Quarterly, Seasonally Adjusted Annual Rate. It is an on-line data set from the Federal Reserve Economic Data available at at https://fred.stlouisfed.org. [13]

The oil prices are provided on-line by the US Energy Information Administration under the Short-Term Energy Outlook Real and Nominal Prices, March 2023. The selected series is the Quarterly Average Imported Crude Oil Price/barrel, Real Price, deflated by the US Consumer Price CPI index. It approximates "the price of oil paid by US refiners for crude oil purchased from abroad" examined by Kilian, Vigfusson (2017).

[13] Other measures of global real economic activity could be used, especially to distinguish the global business cycle fluctuations from fluctuations on the global commodity markets [See, Kilian (2009), (2019), Kilian, Zhou (2018), Hamilton (2018) for the indexes of real economic activity].

The series of GDP rate and oil price (divided by 10) are displayed in Figure 6. The length of this bivariate series is 134.

We estimate the causal-noncausal VAR(1) from the demeaned series of GDP growth (series 1) and demeaned oil price divided by 10 (series 2). The sample mean of growth is 0.640 and the mean of rescaled oil is 6.348. The estimated VAR(1) model has the autoregressive matrix $\hat{\Phi} = \begin{bmatrix} 0.2987 & -0.0229 \\ -0.1527 & 1.0690 \end{bmatrix}$ with standard deviations of autoregressive coefficients of 0.014, 0.001, 0.032, 0.004. The eigenvalues are 0.294 and 1.073. The densities of estimated errors provided in online Appendix D are non-Gaussian.

The presence of a noncausal root reflects the nonlinear dynamic features corresponding to the sequence of speculative bubbles. The bubbles in oil prices are accommodated by the stationary mixed VAR(1) model of GDP rate and oil price levels.[14] The mixed causal-noncausal modelling is an alternative to the modelling of bubbles by threshold autoregressive models for first differences of oil price logarithms [see e.g. Herrera, Lagalo, Wada (2015), Kilian, Vigfusson (2017)]. The comparison with these alternative nonlinear models is discussed below.

We find that that the approximated errors, their squares and third powers are serially uncorrelated. The variance-covariance matrix is $\hat{\Sigma} = \begin{bmatrix} 0.262 & 0.184 \\ 0.184 & 1.506 \end{bmatrix}$ and contemporaneous correlation is statistically significant equal to 0.28.

We perform out-of-sample (oos) predictions from the mixed VAR(1) at the points indicated in Figure 7 below.

During the selected period, both series displayed several sudden changes, while the oil price was slowly decreasing. More specifically, the oil prices at times T=110, 112, 114, 115, 116 and 118 were 126.84, 119.98, 125.86, 119.44, 91.14 and 71.52. The growth rate spiked briefly and took the values: 0.13, 0.71, 1.28, 1.16, 0.44 and 0.58. These changes are reflected in the values of demeaned growth rate and rescaled and demeaned oil price. From the technical point of view, we chose a period when a fair number of observations is available at the end of the trajectory to evaluate the predictive density reasonably well.

The predictive density is evaluated over a grid of 200 points, equidistant by 0.1 below and above the last observed value, except for the density for T=110, which is evaluated

[14]The analysis of oil prices is often performed on log-differences. This approach could lead to over-differencing if the trend (bubble) is a local explosive patterns in a strictly stationary process.

over a grid of 100 values. We use again Gaussian kernels and bandwidths $h_2 = s.d.(Z_{2,t})$ $h_{11} = s.d.(\varepsilon_1)$ and $h_{12} = s.d.(\varepsilon_2)$ to estimate one-step ahead oos predictive density (3.12) (see On-Line Appendix C for kernel density estimators). The point forecasts are the highest modes of the predictive density, and the forecast intervals are the quantiles of the predictive density. The forecast interval is at level 80% to ensure a sufficient number of observations to estimate the quantiles of the predictive density and to alleviate the effect of bimodality as much as possible.

Table 2 : Prediction Intervals

T	forecast		interval growth		interval oil		true values	
	growth	oil	q11	q12	q21	q22	growth	oil
110	-0.037	6.937	-1.437	0.962	4.337	8.537	-0.501	6.335
112	-0.015	7.333	-1.315	0.784	4.933	8.833	0.070	5.649
114	-0.359	6.221	-1.559	0.540	3.921	7.921	0.643	6.237
115	0.145	6.451	-1.054	1.045	4.051	8.251	0.524	5.595
116	0.121	5.862	-1.078	1.021	3.562	7.562	-0.191	2.765
118	0.077	-0.379	-0.922	1.077	-2.479	1.420	-0.060	0.803

The times of predictive density estimation are indicated in Figure 7.

[Insert Figure 7: Predictive Density Estimation Points]

The estimated predictive density at T=110 is plotted in Figure 8.

[Insert Figure 8: Predictive Density Estimation]

We observe deformations and bimodality associated with the fall in oil prices and spike in the growth rate. The deformations disappear when both processes approach their sample mean of 0 values at time T=118 (see On-Line Appendix D).

The observed deformations in the bivariate predictive density are in line with those observed in the univariate predictive density estimators for noncausal processes introduced in Gourieroux, Jasiak (2016) and examined in Gourieroux, Hencic, Jasiak (2020) and Cubbada, Hecq, Voisin (2023).

Let us now consider the predictions of the last 2 points in the trajectory. They are computed according to the procedure described earlier.

The conditioning last observed values at time T=132 are growth -0.458 and oil 0.268.

Table 3 : Predictions at $T = 132$ and $T = 133$.

T	forecast		interval growth		interval oil		true values	
	growth	oil	q11	q12	q21	q22	growth	oil
133	-0.099	0.340	-1.099	0.800	-1.759	2.140	-0.097	0.257
134	-0.020	0.307	-1.020	0.879	-1.792	2.007	0.032	1.103

We encounter a difficulty in predicting the trend reversal at time T=134. We use the approach outlined in Section 5.2 to estimate the confidence set of the prediction interval for the forecast of the first variable at $T = 134$. It is based on 100 backcasted replications of the series of length 100. The confidence set of prediction interval for the demeaned growth rate at level $1 - \alpha_2 = 0.9$ is: $\widehat{CSPI}(y, \alpha_1, \alpha_2) = [-1.546, 1.405]$. This interval accounts for the estimation risk and is larger than the one in Table 3.

The nonlinear causal innovations are estimated from the kernel density estimator, cumulated and transformed by the quantiles of the standard Gaussian distribution. The density $l(z_{2,t}|z_{t-1})$ is first is estimated from formula 4.3 using Gaussian kernel smoothed estimators of the densities and bandwidths equal to the standard errors of the latent component $z_{2,t}$ and of the residuals. Next, the density is cumulated into a conditional c.d.f and transformed by the quantile function of the standard normal variable, providing $\hat{v}_{2,t}$. The density $l(z_{1,t}|z_{2,t}, z_{t-1})$ is estimated from formula 4.5 by using Gaussian kernels as well and bandwidths equal to the standard errors of model residuals. It is cumulated and transformed by the quantile function of standard normal, to obtain $\hat{v}_{1,t}$.

The values of $\hat{v}_{1,t}$ and $\hat{v}_{2,t}$ are computed from the last 80 observations, to ensure sufficiently many observations for the nonparametric estimation of both densities. $\hat{v}_{1,t}$ and $\hat{v}_{2,t}$ are displayed in Figure 9.

[Insert Figure 9: Nonlinear Causal Innovations]

The instantaneous correlation of nonlinear causal innovations is equal to 0.143 and is not statistically significant. Both nonlinear innovations are not significantly serially correlated.

The On-Line Appendix D provides a comparison of our estimated causal-noncausal model with the threshold autoregressive models that have been used in the literature. These models are VAR models with multiple lags and different types of threshold effects for the oil price [see e.g. Herrera et al. (2015) for a discussion of these thresholds]. This comparison shows that :

i) a better fit is obtained with the mixed VAR model even with a much smaller number of lags. This means that the noncausal root captures the bubbles better than arbitrary thresholds.

ii) The "causal" innovations deduced from the threshold autoregressive representation estimated by OLS are serially dependent and non-Gaussian. This can be seen when we plot the ACF of their squares and third powers. In contrast, these nonlinear ACF's are not significant in the mixed VAR model estimated by the GCov.

iii) The "causal" innovations of the threshold autoregressive model are non-Gaussian and significantly instantaneously correlated. This creates a difficulty when analyzing the effect of a shock on oil price and interpreting the associated impulse response functions, when this dependence is disregarded.

6.3 Application to Cryptocurrency Prices

Let us now consider the bivariate series of Bitcoin and Ethereum prices in US Dollars. The Bitcoin (BTC) and Ethereum (ETH) are the main cryptocurrencies with market capitalizations of about 765 and 364 trillions of Dollars, respectively. The bivariate series of 257 daily adjusted closing BTC/USD and ETH/USD exchange rates recorded between July 21, 2021 and April, 04, 2022 [15] are displayed in Figure 10.

[Insert Figure 10: BTC/USD and ETH/USD exchange rates]

We observe that both series display comovements over time and their dynamics are characterized by spikes and local trends/bubbles.

The bivariate VAR model is estimated from the data rescaled and adjusted by a polynomial function of time of order 2 for the hump-shaped pattern in the middle of the sampling period. The estimation is performed without imposing any distributional assumptions on the errors. We use the semi-parametrically efficient GCov estimator with nonlinear transformations including powers two, three and four of the errors summed up to lag 10. The estimated matrix $\hat{\Phi}$ is:

$$\hat{\Phi} = \begin{pmatrix} -0.0901 & 1.1998 \\ 0.4183 & 0.7724 \end{pmatrix},$$

[15] Data Source: Yahoo Finance Canada https://ca.finance.yahoo.com/

with eigenvalues -0.488 and 1.171. Hence, there is a single noncausal component that captures the codependent successive bubbles observed in Figure 9. The standard errors of $\hat{\Phi}_{1,1}$ and $\hat{\Phi}_{1,2}$ are 0.014 and 0.013. For $\hat{\Phi}_{2,1}$ and $\hat{\Phi}_{2,2}$ the standard errors take values 0.022 and 0.019, respectively. The autocorrelations of the approximated errors and their square indicate that the error is close to a bivariate white noise and the model provides a satisfactory fit as shown in Figures 11, 12.

[Insert Figure 11: ACF on $\hat{\varepsilon}_t$]

[Insert Figure 12: ACF on $\hat{\varepsilon}_t^2$]

The Jordan decomposition of matrix $\hat{\Phi}$ is:

$$\hat{\Phi} = \hat{A}\hat{J}\hat{A}^{-1}$$

with

$\hat{A} = \begin{pmatrix} -3.013 & 0.951 \\ 1 & 1 \end{pmatrix}$, $\hat{J} = \begin{pmatrix} -0.488 & 0 \\ 0 & 1.171 \end{pmatrix}$, $\hat{A}^{-1} = \begin{pmatrix} -0.252 & 0.240 \\ 0.252 & 0.759 \end{pmatrix}$.

We forecast the adjusted closing BTC/USD and ETH/USD exchange rates on April 4, 2022 [16] equal to 46622.67 and 3521.24, respectively. The point forecast of demeaned and rescaled $y_{1,T+1}$ is -740.00, and the forecast of demeaned $y_{2,T+1}$ is 105.00. The forecast interval of y_1 at 0.90% is [-940.00 and -150.00] and contains the true value -844.81. The forecast interval of y_2 at 0.90% is [57.00, 174.00] and contains the true value 126.68. After adjusting for the mean and scale, we obtain forecasts of 46727.0 for Bitcoin and 3499.56 for Ethereum prices, which are off by 105 and 22 Dollars, respectively, outperforming a combined "no-change" forecast based on the previous day values.

7 Concluding Remarks

This paper considers the oos nonlinear forecasting and backcasting in the causal-noncausal mixed (S)VAR model. It introduces the closed-form expression of the causal (past-dependent) predictive distribution for forecasting a mixed (S)VAR model.

A definition of causal (past-dependent) nonlinear innovations for VAR models is also given. Since the causal nonlinear innovations are not uniquely defined, their identification

[16] This last observation is excluded from the estimation.

is examined. As a post-estimation inference method, the confidence set of prediction set is introduced.

The proposed approach is applied to the analysis of the joint dynamic of oil price and GDP growth rate, and a bivariate series of cryptocurrency exchange rates. In both cases a noncausal component is revealed and the nonlinear forecast and confidence intervals are computed.

R E F E R E N C E S

Bec, F., Nielsen, H., and S., Saidi (2020): "Mixed Causal-Noncausal Autoregression: Bimodality Issues in Estimation and Unit Root Testing", Oxford Bulletin of Economics and Statistics, 82, 1413-1428.

Blanchard, O., and D., Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances", American Economic Review, 79, 655-673.

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

Cambanis, S., and I., Fakhre-Zakeri (1994): "On Prediction of Heavy-Tailed Autoregressive Sequences Forward versus Reversed Time", Theory of Probability and its Applications, 39, 217-233.

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

Cochrane, J. (1994): "Shocks", Carnegie-Rochester Conference Series on Public Policy", 41, 295-324.

Comon, P. (1994): "Independent Component Analysis: A New Concept?", Signal Process., 36, 287-314.

Cordoni, F., and F., Corsi (2019): "Identification of Singular and Noisy Structural VAR Models: The Collapsing ICA Approach", D.P. University of Pisa.

Cubadda, G., Hecq, A., and E., Voisin (2023): "Detecting Common Bubbles in Multivariate Mixed Causal-Noncausal Models", CEIS working paper 555.

Davis, R., and L., Song (2020) : "Noncausal Vector AR Processes with Application to Economic Time Series", Journal of Econometrics, 216, 246-267.

Eriksson, J., and V., Koivunen (2004): "Identifiability, Separability and Uniqueness of Linear ICA Models", IEEE Signal Processing Letter, 11, 601-604.

Faust, J., and E., Leeper (1997): "When Do Long Run Identifying Restrictions Give Reliable Results", Journal of Business and Economic Statistics, 15, 345-353.

Fries, S., and J.M., Zakoian (2019): "Mixed Causal-Noncausal Autoregressive Processes", Econometric Theory, 35, 1234-1270.

Funovits, B., and J., Nyholm (2016): "Multivariate All-Pass Time Series Models: Modelling and Estimation Strategies", D.P. University of Helsinki.

Gelfand, A., and A., Smith (1992): "Bayesian Statistics without Tears: A Sampling-Resampling Perspective", Annals of Statistics, 46, 84-88.

Gonzalves, S., Herrera, A., Kilian, L., and E., Pesavento (2021): 'Impulse Response Analysis for Structural Dynamic Models with Nonlinear Regressors", Journal of Econometrics, 225, 107-130.

Gonzalves, S., Herrera, A., Kilian, L., and E., Pesavento (2022): "When Do State-dependent Local Projections Work?", McGill U, working paper.

Gourieroux, C., and A., Hencic (2015) : "Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates", Econometrics of Risk, Studies in Computational Intelligence, 583:17-40

Gourieroux, C., and J., Jasiak (2005) : "Nonlinear Innovations and Impulse Responses with Application to VaR Sensitivity", Annals of Economics and Statistics, 78, 1-31.

Gourieroux, C., and J., Jasiak (2016): "Filtering, Prediction, and Simulation Methods for Noncausal Processes", Journal of Time Series Analysis, 37, 405-430.

Gourieroux, C., and J., Jasiak (2017): "Noncausal Vector Autoregressive Process: Representation, Identification and Semi-Parametric Estimation", Journal of Econometrics, 200, 118-134.

Gourieroux, C., and J., Jasiak (2022): "Generalized Covariance Estimator", Journal of Business and Economic Statistics, forthcoming

Gourieroux, C., Jasiak, J., and M., Tong (2021): "Convolution-Based Filtering and Forecasting: An Application to WTI Crude Oil Prices", Journal of Forecasting, 40, 1230-1244.

Gourieroux, C., Monfort, A., and J.P., Renne (2017) : "Statistical Inference for Independent Component Analysis", Application to Structural VAR Models", Journal of Econometrics, 196, 111-126.

Gourieroux, C., Monfort A., and J.P., Renne (2020): "Identification and Estimation in Nonfundamental Structural VARMA Models", Review of Economic Studies, 87, 1915-1953.

Gourieroux, C., and J.M., Zakoian (2017) : "Local Explosion Modelling by Noncausal Process", Journal of the Royal Statistical Society, B, 79, 737-756.

Granzeria, E., Moon, H., and F., Schorfheide (2018): "Inference for VARs Identified with Sign Restrictions", Quantitative Economics, 9, 1087-1121.

Guay, A. (2021): "Identification of Structural Vector Autoregressions Through Higher Unconditional Moments", Journal of Econometrics, 225, 27-45.

Hamilton, J. (2018) : "Measuring Global Economic Activity", DP University of California at San Diego.

Hecq, A. , Lieb, L. and S. Telg (2016): "Identification of Mixed Causal-Noncausal Models in Finite Samples", Annals of Economics and Statistics, 123/124, 307-331.

Hecq, A. and E. Voisin (2023): "Detecting Common Bubbles in Multivariate Mixed Causal-Noncausal Models", working paper, Maastrich University.

Herrera, A., Lagalo, L., and T., Wada (2015) : "Asymmetries in the Response of Economic Activity to Oil Price Increases and Decreases ?", Journal of International Money and

Finance, 50, 108-133.

Imbens, G., and C., Manski (2004): "Confidence Intervals for Partially Identified Parameters", Econometrica, 72, 1845-1857.

Keweloh, S. (2021): "A Generalized Method of Moments Estimator for Structural Vector Autoregressions Based on Higher Moments", Journal of Business and Economic Statistics, 39, 772-782.

Kilian, L. (2009) : "Not Oil Price Shocks are Alike : Disentangling Demand and Supply Shocks in the Crude Oil Market", American Economic Review, 99, 1053-1069.

Kilian, L. (2019) : "Measuring Global Real Economic Activity : Do Recent Critiques Hold up to Scrutiny ?", Economics Letters, 178, 106-110.

Kilian, L., and R., Vigfusson (2017) : "The Role of US Oil Price Shocks in Causing US Recessions", Journal of Money, Credit and Banking, 40, 1747-1776.

Kilian, L., and X., Zhou (2018) : "Modeling Fluctuations in the Global Demand for Commodities", Journal International Money and Finance, 88, 54-78.

Koop, G., Pesaran, H., and S., Potter (1996): "Impulse Response Analysis in Nonlinear Multivariate Models", Journal of Econometrics, 74, 119-147.

Lanne, M., and J., Luoto (2016): "Noncausal Bayesian Vector Autoregression", Journal of Applied Econometrics, 31, 1392-1406.

Lanne, M., and J., Luoto (2021): "GMM Estimation of Non-Gausian Structural Vector Autoregressions", Journal of Business and Economic Statistics, 39, 69-81.

Lanne, M., and P., Saikkonen (2008): "Modelling Expectations with Noncausal Autoregressions", MPRA paper 8411.

Lanne, M., and P., Saikkonen (2010) : "Noncausal Autoregressions for Economic Time Series", Journal of Time Series Econometrics, 3, 1-39.

Lanne, M., and P., Saikkonen (2011) : "GMM Estimators with Non-Causal Instruments", Oxford Bulletin of Economics and Statistics, 71, 581-591.

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.

Lutkepohl, H. (1990): "Asymptotic Distribution of Impulse Response Functions and Forecast Error Variance Decomposition of Vector Autoregressive Models", Review of Economics and Statistics, 72, 116-125.

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

Perko, L. (2001): "Differential Equations and Dynamical Systems", Springer, New York.

Potter, S. (2000): "Nonlinear Impulse Response Functions", Journal of Economic Dynamics and Control, 24, 1425-1426.

Rosenblatt, M. (2012) : "Gaussian and Non-Gaussian Linear Time Series and Random Fields", Springer Verlag.

Rubio-Ramirez, J., Waggoner, D., and T., Zha, (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference", Review of Economic Studies, 77, 665-696.

Sims, C. (1980): "Macroeconomics and Reality", Econometrica, 48, 1-48.

Sims, C. (2002): "Structural VAR's", Econ 513, Time Series Econometrics, Princeton.

Starck, V. (2022): "Estimation of Independent Component Analysis Systems", Working Paper, Brown University.

Swensen, A. (2022): "On Causal and Non-Causal Cointegrated Vector Autoregressive

Time Series", Journal of Time Series Analysis, 42, 178-196.

Tanner, M. (1993): "Tools for Statistical Inference", Springer Series in Statistics, 2nd edition, Springer, New York.

Twumasi, C. and J. Twumasi (2022): "Machine Learning Algorithms for Forecasting and Backcasting Blood Demand Data with Missing Values and Outliers: A Study of Tema General Hospital of Ghana", International Journal of Forecasting, 38, 1258-1277.

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Result from an Agnostic Identification Procedure", Journal of Monetary Economics, 52, 381-419.

Velasco, C. (2022): "Identification and Estimation of Structural VARMA Models Using Higher Order Dynamics", Journal of Business and Economic Statistics, forthcoming.

Velasco, C., and I., Lobato (2018): "Frequency Domain Minimum Distance Inference for Possibly Noninvertible and Noncausal ARMA Models", Annals of Statistics, 46, 555-579.

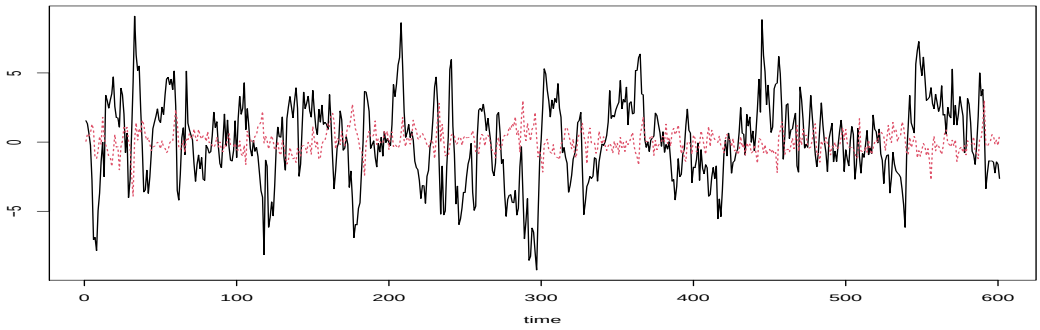


Figure 1: Trajectory of the Bivariate Causal-Noncausal VAR(1) Process: Y_1: solid line, Y_2: dashed line

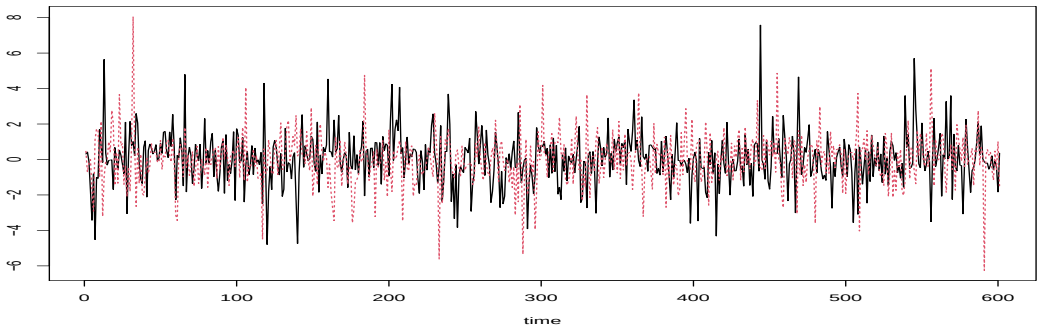


Figure 2: Trajectory of Error Processes: ϵ_1: solid line, ϵ_2: dashed line

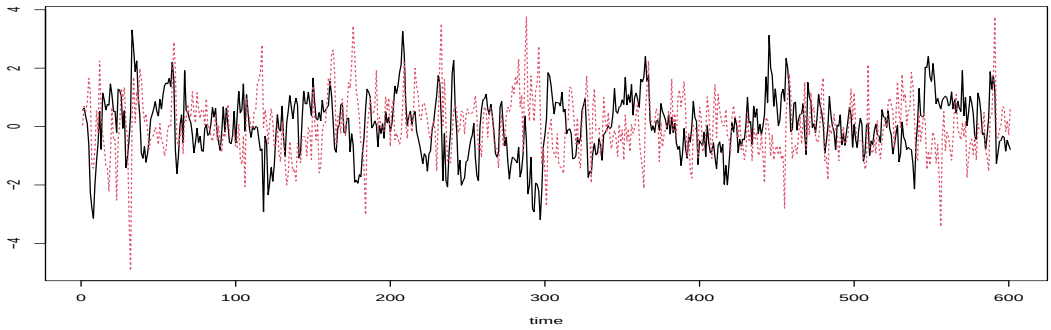
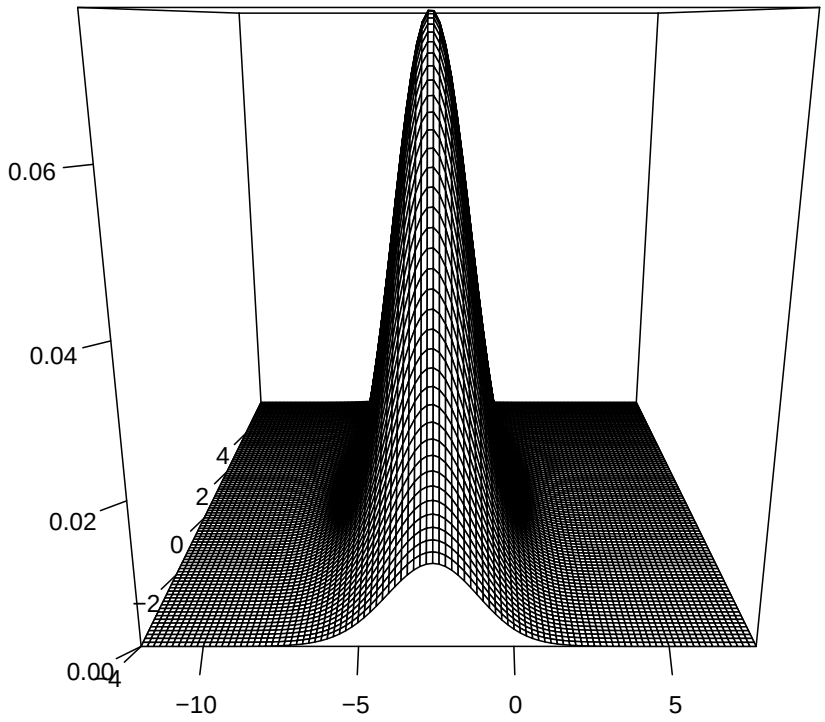
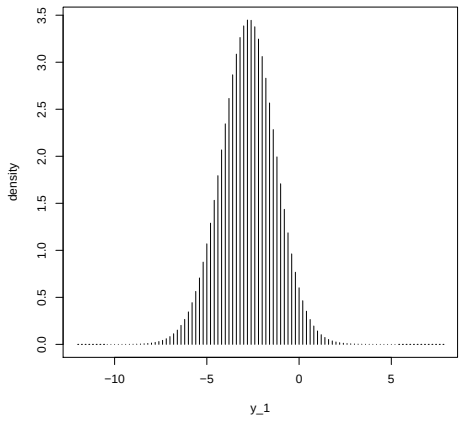


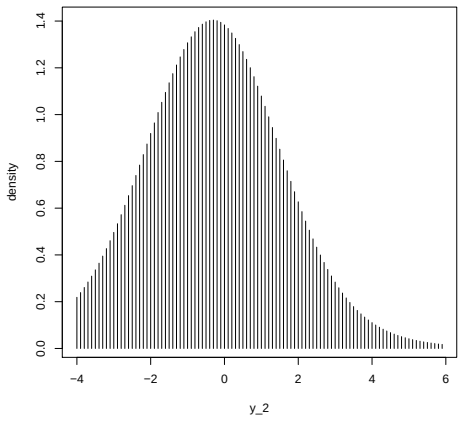
Figure 3: Latent Components: $\hat{Z}_1$: solid line, $\hat{Z}_2$: dashed line



(a) Estimated Joint Predictive Density

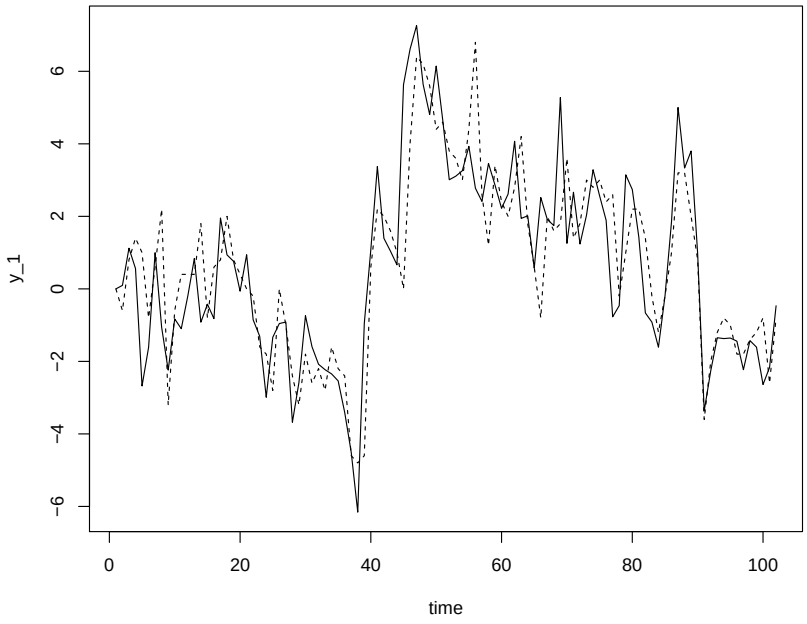


(b) Estimated Marginal Predictive Density of $Y_{1,T+1}$

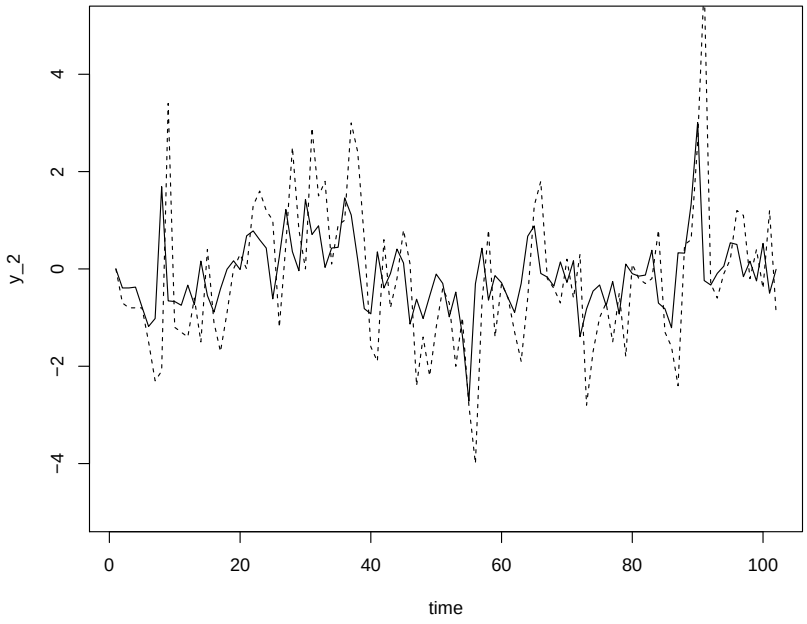


(c) Estimated Marginal Predictive Density of $Y_{2,T+1}$

Figure 4: Estimated Predictive Joint and Marginal Densities



(a) Estimated 1-step-ahead Forecasts of $Y_{1,t}$



(b) Estimated 1-step-ahead Forecasts of $Y_{2,t}$

Figure 5: Estimation-Based oos One-Step Ahead Forecasts true:solid line, forecast:dashed line

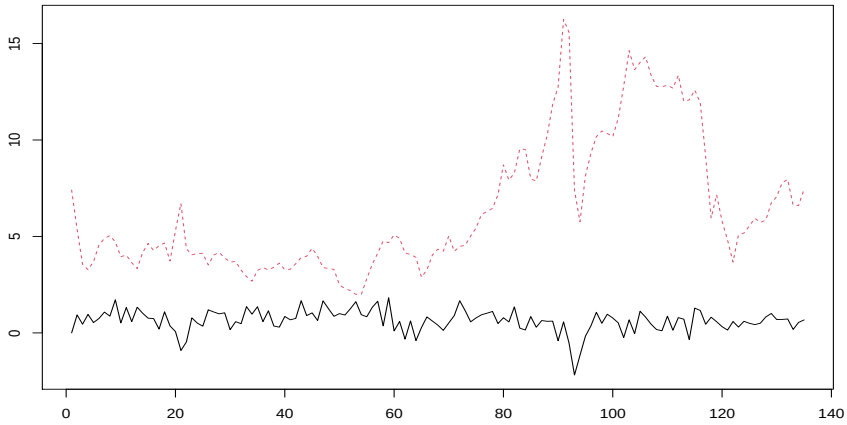


Figure 6: Quarterly Real GDP rate (black) and Oil Price (red), Q1 1986 -Q2 2019

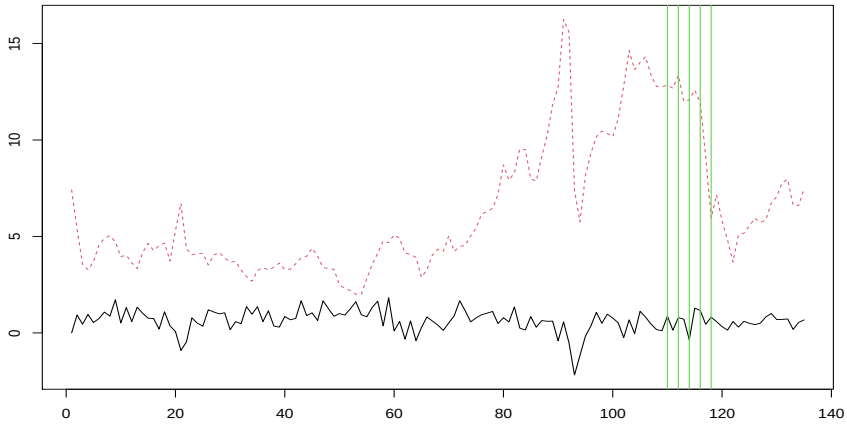


Figure 7: VAR(1) based predictive density estimation points

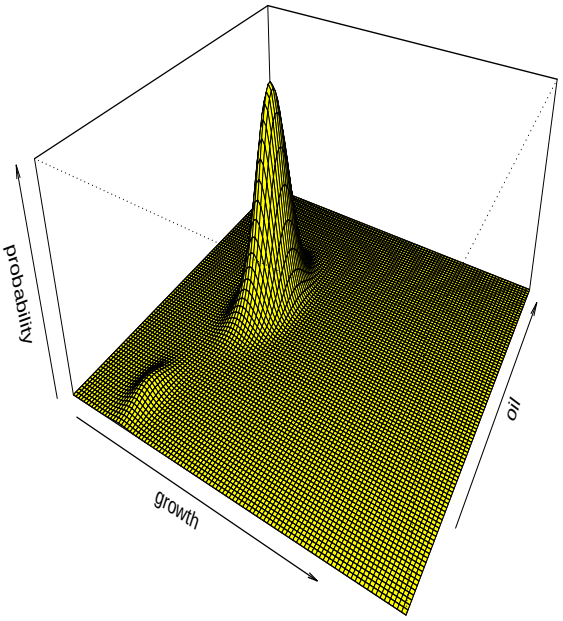
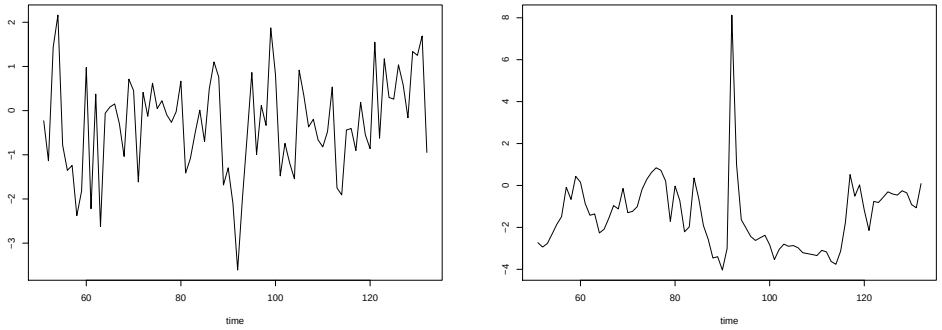


Figure 8: VAR(1) based predictive density estimation, T=110



(a) Nonlinear causal innovations $v_{1,t}$ (b) Nonlinear causal innovations $v_{2,t}$

Figure 9: Nonlinear causal innovations for T=51 to 134

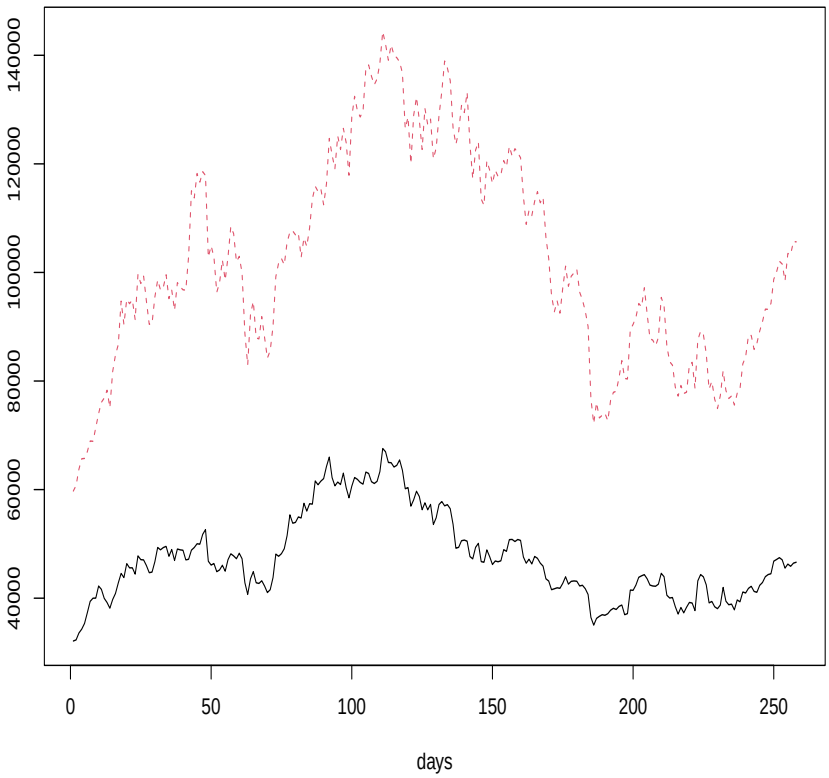


Figure 10: BTC/USD and ETH/USD Daily Closing Rates, July 21, 2021 to April 04, 2022 BTC/USD: solid black line, ETH/USD (times 30): dashed red line

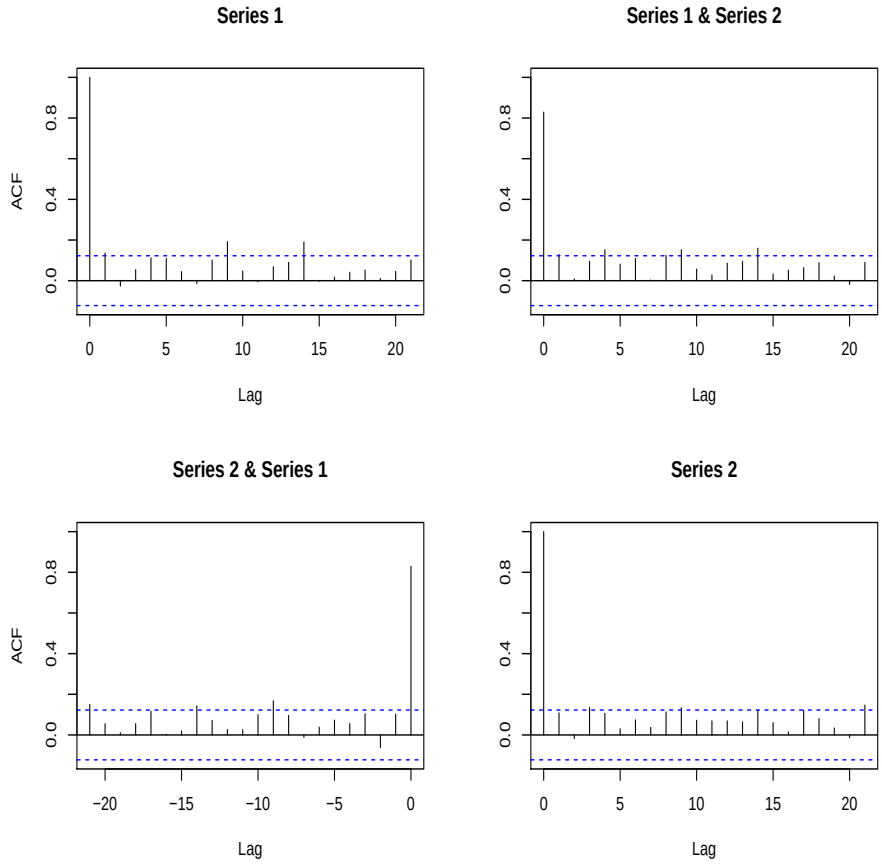


Figure 11: ACF of $\hat{\varepsilon}_t$

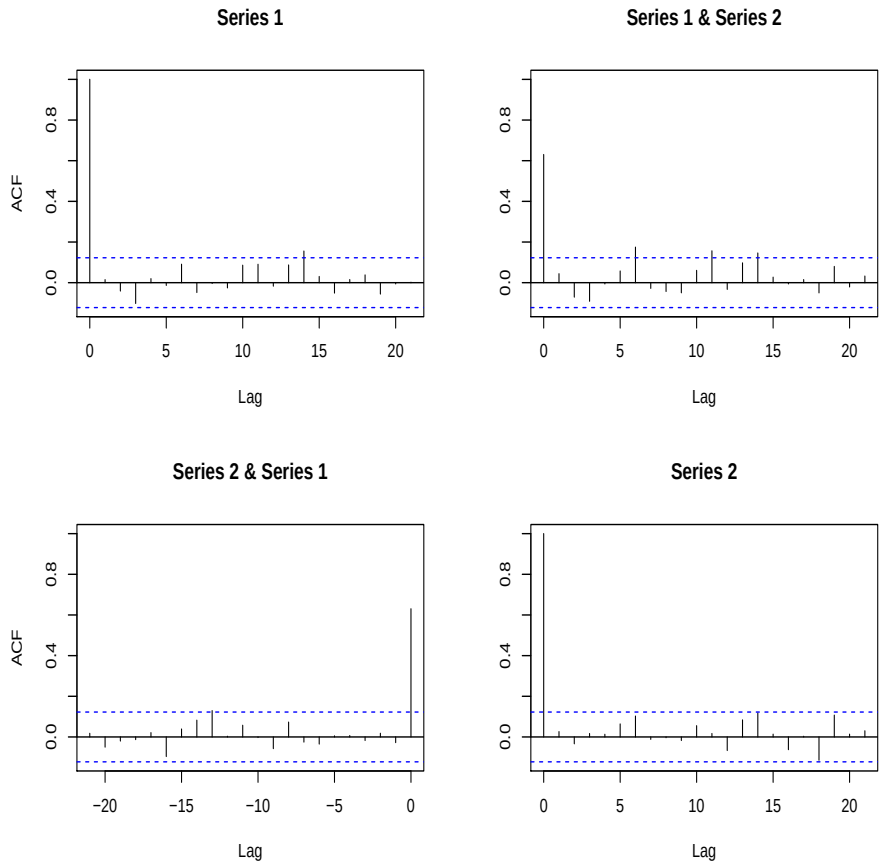


Figure 12: ACF of Squared $\hat{\varepsilon}_t$

APPENDIX 1

Proof of Proposition 1

The proof follows the same lines as in Gourieroux, Jasiak (2016), (2017).

1) Let us consider the information set:

$$I_{T+1} = (Y_1, ..., Y_T, Y_{T+1}).$$

This set is equivalent to the set generated by $(Z_1, Z_2, ..., Z_{T+1})$ and the set generated by $(Z_{1,2}, \eta_3, \eta_4, ..., \eta_T, \eta_{1,T+1}, Z_{2,T}, Z_{2,T+1})$ by using the recursive equations (3.9). Since $(\eta_{1,T+1}, Z_{2,T}, Z_{2,T+1})$ is independent of $\underline{\varepsilon}_T$, we see that the conditional density $l(\eta_{1,T+1}, Z_{2,T}, Z_{2,T+1}|\underline{\varepsilon}_T) = l(\eta_{1,T+1}, Z_{2,T}, Z_{2,T+1})$ is equal to the marginal density.

It follows that the conditional density is:

$$l(\eta_{1,T+1}, Z_{2,T+1}|Z_{2,T}, \varepsilon_T) = l(\eta_{1,T+1}, Z_{2,T+1}|I_T) = l(\eta_{1,T+1}, Z_{2,T+1}|Z_{2,T}). \tag{a.1}$$

The last conditional density needs to be rewritten with a conditioning variable being the future Z_2. From the Bayes theorem, it follows that:

$$l(\eta_{1,T+1}, Z_{2,T+1}|I_T) = \frac{l_2(Z_{2,T+1})}{l_2(Z_{2,T})}\, l(\eta_{1,T+1}, Z_{2,T}|Z_{2,T+1}), \tag{a.2}$$

where l_2 is the marginal density of $Z_{2,t}$.

2) Let us now consider the vector $\eta_t = A^{-1}\begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$. This random vector takes values on the subspace $E = A^{-1}(I\!R^m \times 0^{n-m})$. Its distribution admits a density $g_\eta(\eta_1, \eta_2)$ with respect to the Lebesgue measure on subspace E. Moreover, we have:

$$\eta_{T+1} = \begin{pmatrix} \eta_{1,T+1} \\ Z_{2,T+1} - J_2 Z_{2,T} \end{pmatrix} = \begin{pmatrix} Id & 0 \\ 0 & -J_2 \end{pmatrix} \begin{pmatrix} \eta_{1,T+1} \\ Z_{2,T} \end{pmatrix} + \begin{pmatrix} 0 \\ Z_{2,T+1} \end{pmatrix}. \tag{a.3}$$

Then, conditional on $Z_{2,T+1}$, vector $\begin{pmatrix} \eta_{1,T+1} \\ Z_{2,T} \end{pmatrix}$ takes values in the affine subspace $F = \begin{pmatrix} Id & 0 \\ 0 & -J_2 \end{pmatrix}^{-1} \left[E - \begin{pmatrix} 0 \\ Z_{2,T+1} \end{pmatrix} \right]$ with a density with respect to the Lebesgue measure on F. Since the transformation from η_{T+1} to $\begin{pmatrix} \eta_{1,T+1} \\ Z_{2,T} \end{pmatrix}$ is linear affine invertible, we can apply the Jacobian formula to get:

$$l(\eta_{1,T+1}, Z_{2,T}|Z_{2,T+1}) = |\det J_2|\, g_\eta(\eta_{1,T+1}, Z_{2,T+1} - J_2 Z_{2,T}). \tag{a.4}$$

Then from (a.2), (a.4) and $Z_{1,T+1} = J_1 Z_{1,T} + \eta_{1,T+1}$, it follows that:

$$l(Z_{1,T+1}, Z_{2,T+1}|I_T) = \frac{l_2(Z_{2,T+1})}{l_2(Z_{2,T})}\, |\det J_2|\, g_\eta(Z_{1,T+1} - J_1 Z_{1,T}, Z_{2,T+1} - J_2 Z_{2,T}). \tag{a.5}$$

Let us now derive the predictive density of Y_{T+1} given I_T. We get a succession of affine transformations of variables with values in different affine subspaces (depending on the conditioning set) along the following scheme:

$$\begin{pmatrix} \varepsilon_{T+1} \\ 0 \end{pmatrix} \overset{A^{-1}}{\to} \eta_{T+1} \overset{Id}{\to} \underset{(\text{given } Z_T)}{Z_{T+1}} \overset{A}{\to} \underset{(\text{given } \tilde{Y}_T)}{\begin{pmatrix} Y_{T+1} \\ \tilde{Y}_T \end{pmatrix}}.$$

Then, we can apply three times the Jacobian formula on manifolds. Since $|\det A^{-1}||\det A| = \frac{|\det A|}{|\det A|} = 1$, the Jacobians cancel out and the predictive density becomes:

$$l(y|\underline{Y}_T) = \frac{l_2\left[A^2\begin{pmatrix} y \\ \tilde{Y}_T \end{pmatrix}\right]}{l_2\left[A^2\begin{pmatrix} Y_T \\ \tilde{Y}_{T-1} \end{pmatrix}\right]} |\det J_2|\, g(y - \Phi_1 Y_T - \cdots - \Phi_p Y_{T-p+1}),$$

which yields the formula in Proposition 1.

Proof of Corollary 2

To keep the notation simple, let us assume a mixed VAR(1) model. Then, from Corollary 1, it follows that (Y_t) as well as (Z_t) are Markov processes of order 1 in both calendar and reverse time. The distribution of process (Z_t) is characterized by the pairwise distribution of (Z_{t-1}, Z_t).

From the proof of Proposition 1, it follows that this joint distribution is:

$$l(z_{t-1}, z_t) = l_1(z_{1,t-1}) l_2(z_{2,t})\ |det\, J_2|\ g_\eta(z_{1,t} - J_1 z_{1,t-1}, z_{2,t} - J_2 z_{2,t-1}).$$

Then, the conditional distribution of Z_{t-1} given $Z_t = z_t$ is:

$$\begin{aligned} l(z_{t-1}|z_t) &= l(z_{t-1}, z_t)/l(z_t) \\ &= l(z_{t-1}, z_t)/[l_1(z_{1,t})\, l_2(z_{2,t})] \text{ ,because } Z_{1,t} \text{ and } Z_{2,t} \text{ are independent,} \\ &= \frac{l_1(z_{1,t-1})}{l_1(z_{1,t})}\ |det\ J_2|\ g_\eta(z_{1,t} - J_1 z_{1,t-1}, z_{2,t} - J_2 z_{2,t-1}). \end{aligned}$$

The result in Corollary 2 follows by applying the transformations: $Y_t = AZ_t,\ \epsilon_t = A\eta_t$.

APPENDIX 2

The Causal-Noncausal Model in Multiplicative Form

The multiplicative causal-noncausal model is:

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

where both autoregressive polynomials have roots outside the unit circle and i.i.d. errors ε_t^*. This model is used in Lanne, Luoto, Saikkonen (2012), Lanne, Saikkonen (2013) and Nyberg, Saikkonen (2014). While the multiplicative representation is equivalent to the general AR(p) model for univariate time series, it is not the case in the multivariate framework.

As pointed out in Davis, Song (2020), p. 247, this decomposition implies restrictions on the autoregressive coefficients $\Phi_1, ..., \Phi_p$ of the past-dependent representation.

Moreover, it is not compatible with the SVAR specification (2.1). To illustrate this problem, let us consider the multiplicative bivariate model:

$$\begin{pmatrix} 1-\phi L & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1-\psi L^{-1} \end{pmatrix} Y_t = \varepsilon_t^*.$$

It follows that:

$$\begin{aligned} Y_{1,t} - \phi Y_{1,t-1} &= \varepsilon_{1,t}^*, \\ Y_{2,t} - \psi Y_{2,t+1} &= \varepsilon_{2,t}^*, \end{aligned}$$

or equivalently:

$$\begin{aligned} Y_{1,t} - \phi Y_{1,t-1} &= \varepsilon_{1,t}^*, \\ Y_{2,t} - \frac{1}{\psi} Y_{2,t-1} &= -\frac{1}{\psi}\varepsilon_{2,t-1}^*, \end{aligned}$$

$$\iff Y_t = \begin{pmatrix} \phi & 0 \\ 0 & \frac{1}{\psi} \end{pmatrix} Y_{t-1} + \varepsilon_t,$$

where $\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}$ with $\varepsilon_{1,t} = \varepsilon_{1,t}^*$ and $\varepsilon_{2,t} = -\frac{1}{\psi}\varepsilon_{2,t-1}^*$.

We observe that, if $\varepsilon_{1,t}^*, \varepsilon_{2,t}^*$ are correlated, then $\varepsilon_{1,t}$ and $\varepsilon_{2,t+1}$ are correlated too. Therefore the condition of i.i.d. errors in the SVAR model (2.1) cannot be satisfied.

This major difficulty is a consequence of a different normalization. For example, if $\Phi(L) = Id - \Phi L$ and $\Psi(L^{-1}) = Id - \Psi L^{-1}$, then the multiplicative model is such that:

$$\Phi(L)\Psi(L^{-1})Y_t = -\Psi Y_{t+1} + (Id + \Phi\Psi)Y_t - \Phi Y_{t-1} = \epsilon_t^*,$$

which cannot be transformed into:

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \varepsilon_t,$$

if matrix Ψ is not invertible.

APPENDIX 3
Identification of Nonlinear Causal Innovations

For ease of exposition, let us consider a bivariate VAR(1) model, which is a Markov process. By analogy to the recursive (i.e. causal) approach for defining the shocks, we start from the first component.

i) Let $F_1[y_1|Y_{T-1}]$ denote the conditional c.d.f. of $Y_{1,T}$ given Y_{T-1} and define:

$$v_{1,T} = F_1[Y_{1,T}|Y_{T-1}]. \tag{a.6}$$

Then, $v_{1,T}$ follows a uniform distribution $U_{[0,1]}$ for any Y_{T-1}. In particular, $v_{1,T}$ is independent of Y_{T-1}.

ii) Let $F_2[y_2|Y_{1,T}, Y_{T-1}]$ denote the conditional c.d.f. of $Y_{2,T}$ given $Y_{1,T}, Y_{T-1}$, and define:

$$v_{2,T} = F_2[Y_{2,T}|Y_{1,T}, Y_{T-1}]. \tag{a.7}$$

It follows that $v_{2,T}$ follows a uniform distribution on [0,1], for any $Y_{1,T}, Y_{T-1}$, or equivalently for any $v_{1,T}, Y_{T-1}$. Therefore, $v_{2,T}$ is independent of $v_{1,T}, Y_{T-1}$.

iii) By inverting equations (a.6)-(a.7), we obtain a nonlinear autoregressive representation: $Y_T = g(Y_{T-1}, v_T)$, where the v_T's are i.i.d. such that $(v_{1,T}), (v_{2,T})$ are independent.

This approach resembles the shock "ordering" used in linear Gaussian models to solve the identification issue. Alternatively, one can use an alternative ordering: $Y_{2,T}$ followed

by $Y_{1,T}$ given $Y_{2,T}$. More generally, for any invertible nonlinear transformation $Y_T^* = c(Y_T)$, the above approach can be applied first to $Y_{1,T}^*$ and next to $Y_{2,T}^*$ conditional on $Y_{1,T}^*$.

Let us now describe in detail all the innovation identification issues. First, we can assume that $v_{1,T}, v_{2,T}$ are i.i.d and independent of one another with uniform distributions on [0,1]. We need to find out if there exists another pair of variables $w_{1,T}, w_{2,T}$, which are independent and uniformly distributed such that:

$$g(Y_{T-1}, w_T) = g(Y_{T-1}, v_T), \quad \forall\, Y_{T-1},$$

or, equivalently, a pair of variables w_T that satisfy a (nonlinear) one-to-one relationship with v_T. Let $w = a(v)$ denote this relationship. We have the following Lemma:

Lemma A.1:

Let us assume that a is continuous, twice differentiable and that the Jacobian matrix $\partial a(v)/\partial v$ has distinct eigenvalues. Then, the components of a are harmonic functions, that is:

$$\frac{\partial^2 a_j(v)}{\partial v_1^2} + \frac{\partial^2 a_j(v)}{\partial v_2^2} = 0, \quad j = 1, 2.$$

Proof:

i) We can apply the Jacobian formula to get the density of w given the density of v. Since both joint densities are uniform, it follows that $|\det \frac{\partial a(v)}{\partial v}| = 1$, $\forall v \in [0,1]^2$.

ii) Let us consider the eigenvalues $\lambda_1(v), \lambda_2(v)$ of the Jacobian matrix $\frac{\partial a(v)}{\partial v}$. The eigenvalues are continuous functions of this matrix, and therefore continuous functions of v (whenever these eigenvalues are different). Then, two cases can be distinguished:

case 1: The eigenvalues are real.

case 2: The eigenvalues are complex conjugates.

iii) In case 1, we have $\lambda_2(v) = 1/\lambda_1(v)$ (or $-1/\lambda_1(v)$), where $\lambda_1(v)$ is less or equal to 1 in absolute value for any v, and then $\lambda_2(v)$ is larger than or equal to 1 in absolute value for any v. Since $a(v) \in [0,1]^2$ for any $v \in [0,1]^2$, it follows that $\lambda_2(v)$ cannot be explosive.

iv) Therefore case 2 is the only relevant one. Let us consider the case $\det \frac{\partial a(v)}{\partial v} = 1$, $\forall v \in [0,1]^2$ (the analysis of $\det \frac{\partial a(v)}{\partial v} = -1$ is similar). Then, the Jacobian matrix is a rotation matrix:

$$\frac{\partial a(v)}{\partial v} = \begin{pmatrix} \frac{\partial a_1(v)}{\partial v_1} & \frac{\partial a_1(v)}{\partial v_2} \\ \frac{\partial a_2(v)}{\partial v_1} & \frac{\partial a_2(v)}{\partial v_2} \end{pmatrix} \equiv \begin{pmatrix} \cos\theta(v) & -\sin\theta(v) \\ \sin\theta(v) & \cos\theta(v) \end{pmatrix}.$$

We deduce that:

$$\begin{aligned} \frac{\partial a_1(v)}{\partial v_1} &= \frac{\partial a_2(v)}{\partial v_2}, \\ \frac{\partial a_1(v)}{\partial v_2} &= -\frac{\partial a_2(v)}{\partial v_1}. \end{aligned} \tag{a.8}$$

Let us differentiate the first equation with respect to v_1 and the second one with respect to v_2. We get:

$$\frac{\partial^2 a_1(v)}{\partial v_1^2} = \frac{\partial^2 a_2(v)}{\partial v_1 \partial v_2} \text{ and } \frac{\partial^2 a_1(v)}{\partial v_2^2} = -\frac{\partial^2 a_2(v)}{\partial v_1 \partial v_2}, \tag{a.9}$$

and by adding these equalities:

$$\frac{\partial^2 a_1(v)}{\partial v_1^2} + \frac{\partial^2 a_1(v)}{\partial v_2^2} = 0. \tag{a.10}$$

Therefore a_1 is a harmonic function that satisfies the Laplace equation (a.10). Similarly, a_2 is also a harmonic function. QED

Harmonic functions are regular functions: they are infinitely differentiable and have series representations that can be differentiated term by term [Axler et al. (2001)]:

$$\begin{aligned} a_1(v) &= \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}(a_{1hk} v_1^k v_2^h), \\ a_2(v) &= \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}(a_{2hk} v_1^k v_2^h). \end{aligned} \tag{a.11}$$

Moreover, these series representations are unique. Then, we can apply the conditions (a.8) to these expansions to derive the constraints on the series coefficients and the link between functions a_1 and a_2.

Let us define:

$$\frac{\partial a_1(v)}{\partial v_1} = \frac{\partial a_2(v)}{\partial v_2} \equiv \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}(c_{hk} v_1^h v_2^k).$$

Then, by integration, we get:

$$\begin{aligned} a_1(v) &\equiv \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}[c_{hk}\frac{v_1^{h+1}}{h+1}v_2^k] + \sum_{k=0}^{\infty} d_{1k}v_2^k, \\ a_2(v) &\equiv \sum_{h=0}^{\infty}\sum_{k=0}^{\infty}[c_{hk}v_1^h\frac{v_2^{k+1}}{k+1}] + \sum_{h=0}^{\infty} d_{2h}v_1^h, \end{aligned}$$

where the second sums on the right hand sides are the integration "constants". Equivalently, we have:

$$\begin{aligned} a_1(v) &\equiv \sum_{k=0}^{\infty} d_{1k}v_2^k + \sum_{h=1}^{\infty}\sum_{k=0}^{\infty}[c_{h-1,k}\frac{v_1^h}{h}v_2^k], \\ a_2(v) &\equiv \sum_{h=0}^{\infty} d_{2h}v_1^h + \sum_{h=0}^{\infty}\sum_{k=1}^{\infty}[c_{h,k-1}v_1^h\frac{v_2^k}{k}]. \end{aligned}$$

Let us now write the second equality in (a.8), i.e.

$$\frac{\partial a_1(v)}{\partial v_2} = -\frac{\partial a_2(v)}{\partial v_1}.$$

This yields:

$$\begin{aligned} \frac{k+1}{h}c_{h-1,k+1} &= -\frac{h+1}{k}c_{h+1,k-1}, \quad h \geq 1, k \geq 1, \\ \frac{1}{h}c_{h-1,1} &= -(h+1)d_{2,h+1} \quad h \geq 1, \\ \frac{1}{h}c_{1,k-1} &= -(k+1)d_{1,k+1} \quad k \geq 1, \\ d_{11} &= -d_{21}. \end{aligned} \tag{a.12}$$

The set of restrictions (a.12) provides information on the dimension of underidentification. As the dimension concerns functional spaces, we describe it from the series expansions (a.11) and the number of independent parameters $a_{1,h,k}, a_{2,h,k}$ with $h + k \leq m$. This number is equal to $(m + 1)(m + 2)/2$.

Proposition A.2

The space of parameters $(a_{1,h,k}, a_{2,h,k},\ h + k \leq m)$ is of dimension $2m$.

Proof:

Let us consider an alternative parametrization of (a.12) with parameters $c_{h,k}, d_{1,h}, d_{2,h}$. The parameters $a_{1,h,k}, a_{2,h,k}$ with $h+k=j$ are linear functions of parameters $c_{h,k}, h+k=j+1, d_{1,j+1}d_{2,j+1}$. The restrictions (a.12) imply a degree of freedom equal to 2 on this subset. The result follows.

QED

Other identification issues can arise if transformation a is not assumed twice continuously differentiable. Let us consider the first component v_1 that follows the uniform distribution on $[0,1]$ and introduce two intervals $[0,c]$ and $[1-c,1]$ with $c<0.5$. Then, the variable w_1 defined by:

$$w_1 = \begin{cases} v_1, \text{ if } v_1 \in (c, 1-c), \\ 2v_1 - 1, \text{ if } v_1 \in (0,c) \cup (1-c,1), \end{cases}$$

also follows the uniform distribution and, similarly to v_1, variable w_1 is independent of $v_2 = w_2$.

Note that this transformation is not monotonous. Therefore, the size δ of a shock to v_1 is difficult to interpret in terms of a magnitude of a shock to w_1.

We conclude that, in a nonlinear dynamic framework, the assumption of independence between the components of v_t is insufficient to identify the structural innovations to be shocked.