Detecting Co-Movements in Noncausal Time Series

Cubadda, Gianluca and Hecq, Alain and Telg, Sean

Universita' di Roma "Tor Vergata", Dipartimento di Economia e Finanza, Maastricht University, School of Business and Economics, Department of Quantitative Economics

2 March 2017

Detecting Co-Movements in Noncausal Time Series

Gianluca Cubadda* **Alain Hecq†** **Sean Telg‡**

March 2, 2017

Abstract

This paper introduces the notion of common noncausal features and proposes tools for detecting the presence of co-movements in economic and financial time series subject to phenomena such as asymmetric cycles and speculative bubbles. For purely causal or noncausal vector autoregressive models with more than one lag, the presence of a reduced rank structure allows to identify causal from noncausal systems using the usual Gaussian likelihood framework. This result cannot be extended to mixed causal-noncausal models, and an approximate maximum likelihood estimator assuming non-Gaussian disturbances is needed for this case. We find common bubbles in both commodity prices and price indicators.

Keywords: mixed causal-noncausal process, common features, vector autoregressive models, commodity prices, common bubbles.

JEL codes: C12, C32, E32.

---

*Universita' di Roma "Tor Vergata", Dipartimento di Economia e Finanza. Email: gianluca.cubadda@uniroma2.it.

†Corresponding author: Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands. Email: a.hecq@maastrichtuniversity.nl.

‡Maastricht University, School of Business and Economics, Department of Quantitative Economics. Email: j.telg@maastrichtuniversity.nl

Previous versions of this paper have been presented at conferences of the Society for Nonlinear Dynamics and Econometrics (Oslo, March 2015), the International Association for Applied Econometrics-IAAE (Tessaloniki, June 2015), the Netherlands Econometric Study Group (Maastricht, May 2015), the International Symposium on Forecasting (Santander, June 2016), the Computational and Financial Econometrics (Sevilla, December 2016), the Brazilian Econometric Society (Iguassu, December 2016) as well as at seminars at the University of Balear Islands (Palma, November 2015), SKKU (Seoul, January 2016), Aix-Marseille (Marseille, March 2016), KU Leuven (Leuven, March 2016), FGV (Rio de Janeiro, May 2016). We thank the participants for helpful comments and suggestions. We would also like to thank Lenard Lieb for stimulating and fruitful discussions.

1 Motivation

This paper studies the existence of common cyclical features in economic and financial variables by considering both causal and noncausal vector autoregressive (VAR) models. A feature is a dominant characteristic observed in univariate time series. Examples of such features are the presence of a stochastic trend, a dynamic pattern, a conditional time varying volatility process, seasonality, structural breaks, nonlinearities, etc. A feature is said to be common if a linear combination of the series no longer has the feature even though each of the series individually has it. One will then talk about cointegration, common cyclical features, common ARCH, co-breaking, common seasonality, and co-jumps to name a few. In this article, we focus on common cyclical features as introduced by Engle and Kozicki (1993) and Vahid and Engle (1993). That is, we are interested in the existence of commonalities in the dynamics of time series. Those restrictions have implications for forecasting accuracy, parameter efficiency as well as for the interpretation of business cycle co-movements and the evaluation of economic theories (Guillén, Hecq, Issler and Saraiva, 2015; Issler and Vahid, 2001).

Although co-movements in the cyclical fluctuations of many economic time series seem theoretically well motivated and even graphically visible, it often happens that the presence of common dynamics is rejected by formal test statistics. The first obvious reason stems from the fact that cycles do not have to be synchronous and therefore several statistical models have been introduced to account for adjustment delays.[1] Several additional reasons have been suggested in the literature, among which the impact of misspecifications: presence of conditional heteroskedasticity (Candelon, Hecq and Verschoor, 2005), seasonal adjustment (Hecq, 1998; Cubadda, 1999), outlying observations and nonlinear phenomena.

In this paper we propose a novel explanation for not being able to detect common cyclical patterns as often as expected. In fact, one of the most popular approaches to test for serial correlation common features, SCCF henceforth, is to consider the VAR model of order p of a stationary $n-$vector time series Y_t, denoted as VAR(p)

$$\Pi(L)Y_t \equiv (I_n - \Pi_1 L - ... - \Pi_p L^p)Y_t = \varepsilon_t,$$

where L is the lag operator, and ε_t are *i.i.d.* innovations with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Omega$ (positive

[1] Those extensions include the polynomial serial correlation common feature specification (Cubadda and Hecq, 2001), the weak form common feature model (Hecq, Palm and Urbain, 2006) and the codependent cycle approach (Vahid and Engle, 1997). See Cubadda (2007) for a unifying framework of the various forms of common features.

definite). Under the existence of SCCF, the VAR(p) can be written as

$$(I_n - \delta_\perp \beta_1' L - ... - \delta_\perp \beta_p' L^p) Y_t = \varepsilon_t,$$

where $\delta_\perp$ is a full rank $n \times (n-k)$ matrix ($0 < k < n$) such that there exist k linear combinations $\delta' Y_t = \delta' \varepsilon_t$ that annihilate the entire dynamics (where δ' is such that $\delta' \delta_\perp = 0$). Hence, a reduced rank regression approach on the matrix $[\Pi_1, ..., \Pi_p]$ is typically used to detect the commonalities among series Y_t. However, the linear VAR model cannot capture several important characteristics of macroeconomic and financial variables, such as asymmetric cycles or bubbles phenomena. Noncausal VARs, i.e. VAR models with leads of series Y_t in place of their lags or, more generally, mixed causal-noncausal VARs, which are models where both lags and leads of Y_t are present in the right-hand side (see Lanne and Saikkonen, 2013), are able to generate much richer dynamics than usual VARs.[2]

Causal and noncausal VARs are often hardly visually distinguishable (e.g., see the graphs in Section 4) and non-Gaussian distributions are needed to discriminate between those models by estimation techniques like Maximum Likelihood (ML). We show in this paper that for purely causal or noncausal VAR models with more than one lag or lead, the presence of a reduced rank structure in the VAR coefficient matrices allows to identify causal from noncausal systems even using the Gaussian framework. Using either lags or leads within a canonical correlation analysis or a Generalized Method of Moments (GMM) approach, the reduced rank restrictions help to identify the correct model. This result stems from the fact that, except for VARs with either one lag or one lead only, the existence of SCCF implies that the autocorrelation matrices of series Y_t have a common left null space for either any lag or any lead different from zero. However, this important observation does not extend to mixed causal-noncausal models.

The rest of the paper is organized as follows. Section 2 summarizes results about the mixed causal-noncausal autoregressive model with common features. Section 3 introduces our testing strategy for identifying purely causal and noncausal models with SCCF. The finite sample behavior of reduced rank regressions and GMM orthogonality tests when noncausality is present is studied in Section 4. Section 5 applies our new modeling procedure to investigate the presence of common bubbles in price index variables as well as in commodity prices. Section 6 summarizes and concludes.

[2]More precisely, we define a noncausal model as a model which has a unique stationary solution in terms of current and future error terms. For the mixed causal-noncausal model, this is the two-sided MA representation, i.e., past, current and future disturbances.

2 Causal and noncausal models

Mixed causal-noncausal models have recently become increasingly popular because of their appealing properties. On a theoretical level, they offer the possibility to rewrite a process with explosive roots in direct time into as process in reverse time with roots outside the unit circle. In the applied econometric literature, there is a growing interest because mixed causal-noncausal models (*i*) might improve forecast performances, (*ii*) can be solutions to rational expectations models which take forward-looking behavior into account and (*iii*) are able to model nonlinear features in macroeconomic and financial data. A more extensive overview and the appropriate references can be found in Hecq, Lieb and Telg (2016A), Lanne, Luoto and Saikkonen (2012), Lanne and Saikkonen (2011A, 2011B), Hencic and Gouriéroux (2014), Gouriéroux and Zakoïan (2016).

2.1 Univariate models

The mixed causal-noncausal autoregressive process for a scalar series y_t, denoted as MAR(r, s), can be represented as

$$\pi(L)\phi(L^{-1})y_t \equiv (1 - \pi_1 L - ... - \pi_r L^r)(1 - \phi_1 L^{-1} - ... - \phi_s L^{-s})y_t = \epsilon_t, \tag{1}$$

where ϵ_t are scalar *i.i.d.* innovations. Notice that the backshift operator L creates lags when raised to positive powers and leads when raised to negative powers, i.e., $L^j y_t = y_{t-j}$ and $L^{-j} y_t = y_{t+j}$. When $\phi_1 = ... = \phi_s = 0$, the process y_t is a purely causal MAR$(r, 0)$, also known as the conventional causal AR(r) process

$$\pi(L)y_t = \epsilon_t,$$

while the process is a purely noncausal MAR$(0, s)$ model

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

when $\pi_1 = ... = \pi_r = 0$.

In order to ensure stationarity, the roots of both polynomials $\pi(L)$ and $\phi(L^{-1})$ are assumed to lie outside the unit circle. Additionally, the identifiability of the causal and the noncausal part is established by assuming that the error terms ϵ_t are non-Gaussian with $\mathbb{E}(|\epsilon_t|^\delta) < \infty$ for $\delta > 0$ (Breidt, Davis, Lii and Rosenblatt, 1991). Most papers by Lanne, Saikkonen and coauthors use a Student's t distribution as an alternative for the Gaussian distribution while Gouriéroux and coauthors rely on α-stable distributions (in practice, mostly Cauchy or mixtures of Cauchy and Gaussian).

It is well known that one can define a pseudo-causal AR(p) representation of (1) with $p = r + s$ lags[3]

$$a(L)y_t = \epsilon_t^*, \tag{2}$$

where ϵ_t^* is a white noise error, and the autoregressive polynomial $a(L)$ has all roots outside the unit circle (see, Brockwell and Davis, 1991; Hecq et al, 2016A). Although not autocorrelated, the error term ϵ_t^* is generally not *i.i.d.* anymore. Gouriéroux and Zakoïan (2016) exploit this property to discriminate between purely causal and noncausal AR(1) processes.

2.2 Multivariate models with common features

Consider now a mixed vector autoregressive model of order (r, s), denoted as MVAR(r, s), of a second-order stationary $n-$vector time series Y_t such that

$$\Pi(L)\Phi(L^{-1})Y_t \equiv (I_n - \Pi_1 L - ... - \Pi_r L^r)(I_n - \Phi_1 L^{-1} - ... - \Phi_s L^{-s})Y_t = \varepsilon_t, \tag{3}$$

where $\Pi(L)$ and $\Phi(L^{-1})$ are $n \times n$ polynomial matrices with both $\det\{\Pi(L)\} = 0$ and $\det\{\Phi(L^{-1})\} = 0$ having solutions outside the unit circle. The error term ε_t is an *i.i.d.* innovation process with $\mathbb{E}(\varepsilon_t) = 0$ and $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Omega$ (positive definite). From (3) it follows that the series Y_t have a strictly stationary solution as a two-sided MA representation

$$Y_t = \Psi(L, L^{-1})\varepsilon_t = \sum_{j=-\infty}^{\infty} \Psi_j \varepsilon_{t-j}, \tag{4}$$

where $\Psi(L, L^{-1}) \equiv [\Pi(L)\Phi(L^{-1})]^{-1}$, with $\Psi_0 = I_n$ and Ψ_j converging to zero at a geometric rate as $j \to \pm\infty$.

[3]Lanne and Saikkonen (2011A) propose to use usual information criteria to first get the pseudo-causal AR(p). Hencic and Gouriéroux (2014) propose to look at the empirical ACF which is consistent even in the infinite variance case to determine p.

To extend the SCCF property to the MVAR(r, s) case, we search for the existence of a full rank matrix δ such that $\mathbb{E}(\delta' Y_t | Y_{t+s}, ..., Y_{t+1}, Y_{t-1}, ..., Y_{t-r}) = 0$. This gives rise to the following definition.

Definition 1. *Series Y_t have k common features, CFs henceforth, if there exists a $n \times k$ $(0 < k < n)$ full rank matrix δ such that $\delta' \Pi(L) \Phi(L^{-1}) = \delta'$ or, equivalently, $\delta' Y_t = \delta' \varepsilon_t$. We also have that $\delta' \tilde{\Phi}(L^{-1}) \tilde{\Pi}(L) = \delta'$ although the product of matrices is not commutative.*[4]

Remark 2. *By premultiplying both sides of Equation (3) by δ' and in view of the definition of CFs we have that*

$$\delta' \Psi(L, L^{-1}) = \delta',$$

which tells us that the bilateral linear impulse response function of series Y_t is collinear at any lag/lead different from zero. It is easy to check that the reverse implication applies as well.

2.3 Co-movements in purely causal or noncausal models

We first focus on the two polar cases in which data are generated by either a purely causal MVAR$(r, 0)$ or a purely noncausal MVAR$(0, s)$ model. Then we illustrate the properties of the usual common cyclical feature test statistics, e.g., canonical correlations, for noncausal models. Finally, we show that the mixed model necessitates to use a non-Gaussian likelihood framework even when a reduced rank structure is present.

Definition 3. *In purely causal [noncausal] models with $\Phi(L^{-1}) = I_n$ $[\Pi(L) = I_n]$, series Y_t have k causal [noncausal] common features, CCFs [NCCFs] henceforth, if there exists a $n \times k$ $(0 < k < n)$ full rank matrix δ_c $[\delta_{nc}]$ such that $\delta_c' \Pi(L) = 0$ $[\delta_{nc}' \Phi(L^{-1}) = 0]$. The common cyclical feature model in the purely causal [noncausal] case implies $\mathbb{E}(\delta_c' Y_t | Y_{t-1}, Y_{t-2}, ...) = 0$ $[\mathbb{E}(\delta_{nc}' Y_t | Y_{t+1}, Y_{t+2}, ...) = 0]$. The $n \times (n-k)$ orthogonal complement of matrix δ_c $[\delta_{nc}]$ is denoted as $\delta_{c\perp}$ $[\delta_{nc\perp}]$.*

Remark 4. *The usual SCCF in a VAR(r) is obviously equivalent to the CCF in a MVAR$(r, 0)$.*

We start the analysis by concentrating on the cases in which data are generated by a purely noncausal MVAR$(0, s)$ model. Assuming that the matrix δ_{nc} exists (we skip the indexes nc and c to simplify

[4]Notice that, differently from the univariate case, representation (3) is not unique. Since the polynomial matrices are generally not commutative, one may alternatively use the model $\tilde{\Phi}(L^{-1}) \tilde{\Pi}(L) Y_t \equiv (I_n - \tilde{\Phi}_1 L^{-1} - ... - \tilde{\Phi}_s L^{-s})(I_n - \tilde{\Pi}_1 L - ... - \tilde{\Pi}_r L^r) Y_t = \varepsilon_t$. Gouriéroux and Jasiak (2015) have a different respresentation for the MVAR which does not involve the multiplicative structure of two coefficient matrices.

notations when no confusion is possible). The noncausal VAR model then reads as

$$Y_t = \sum_{j=1}^{s} \Phi_j Y_{t+j} + \varepsilon_t = \delta_\perp \sum_{j=1}^{s} A'_j Y_{t+j} + \varepsilon_t, \tag{5}$$

where $\delta'_\perp \delta = 0$ and at least one of the $n \times (n-k)$ matrices A_j, for $j = 1, ..., s$, has full rank. As shown by Lanne and Saikkonen (2013), series Y_t admit a causal autoregressive representation MVAR$(s, 0)$ with the same spectral density as (5) and with errors that are multivariate white noise but not innovations with respect to the past. When no CFs exist, this implies that the noncausal and the pseudo-causal representation cannot be distinguished by means of the Gaussian likelihood framework.

We wish to examine the implications of the presence of such CFs in the pseudo-causal VAR representation of the system. In order to do that, we start by defining autocovariances $\Gamma_y(-j) = \mathbb{E}(Y_t Y'_{t+j})$ and note that

$$\delta' \Gamma_y(-j) = \delta' \mathbb{E}(Y_t Y'_{t+j}) = \delta' \mathbb{E}(\varepsilon_t Y'_{t+j}) = 0, \quad \forall j > 0,$$

as $\delta' Y_t = \delta' \varepsilon_t$. Hence, it follows that we can factorize these autocovariances as the product

$$\Gamma_y(-j) = \delta_\perp \Upsilon'_y(-j), \quad \forall j > 0,$$

where $\Upsilon_y(-j)$ is $(n-k) \times n$ matrix. Let us now rewrite the noncausal VAR in (5) under reduced rank restrictions as

$$Y_t = \Phi' X_t + \varepsilon_t = \delta_\perp A' X_t + \varepsilon_t, \tag{6}$$

where $X'_t = [Y'_{t+1}, ..., Y'_{t+s}]$, $\Phi' = [\Phi_1, ..., \Phi_s]$ and $A' = [A'_1, ..., A'_s]$. The coefficient matrix Φ is linked to the autocorrelation matrix function $\Gamma_y(-j)$ through the relation

$$\begin{aligned} \Phi' &= \mathbb{E}(Y_t X'_t)[\mathbb{E}(X_t X'_t)]^{-1} = [\Gamma_y(-1), ..., \Gamma_y(-s)][\Gamma_x(0)]^{-1} \\ &= \delta_\perp [\Upsilon'_y(-1), ..., \Upsilon'_y(-s)][\Gamma_x(0)]^{-1}, \end{aligned} \tag{7}$$

where

$$\Gamma_x(0) = \begin{bmatrix} \Gamma_y(0) & \Gamma_y(-1) & \cdots & \Gamma_y(-s+1) \\ \Gamma_y(1) & \Gamma_y(0) & \cdots & \Gamma_y(-s+2) \\ \vdots & & & \vdots \\ \Gamma_y(s-1) & \Gamma_y(s-2) & \cdots & \Gamma_y(0) \end{bmatrix},$$

and hence $\delta'\Phi' = 0$. Let us now write the pseudo-causal representation of series Y_t as follows

$$Y_t = \sum_{j=1}^{s} \widetilde{\Phi}_j Y_{t-j} + \widetilde{\varepsilon}_t = \widetilde{\Phi}' Z_t + \widetilde{\varepsilon}_t, \tag{8}$$

where $Z_t' = [Y_{t-1}', ..., Y_{t-s}']$ and $\widetilde{\varepsilon}_t$ is a multivariate white noise process such that $\mathrm{E}(\widetilde{\varepsilon}_t Y_{t-j}') = 0$ for any $j > 0$. The coefficient matrix $\widetilde{\Phi}' = [\widetilde{\Phi}_1, ..., \widetilde{\Phi}_s]$ of the pseudo-causal VAR is linked to the autocorrelation matrix function of series Y_t through the relation

$$\begin{aligned} \widetilde{\Phi}' &= \mathrm{E}(Y_t Z_t')[\mathrm{E}(Z_t Z_t')]^{-1} \\ &= [\Gamma_y(1), ..., \Gamma_y(s)][\Gamma_z(0)]^{-1} = [\Upsilon_y(1)\delta_\perp', ..., \Upsilon_y(s)\delta_\perp'][\Gamma_z(0)]^{-1}, \end{aligned} \tag{9}$$

where

$$\Gamma_z(0) = \begin{bmatrix} \Gamma_y(0) & \Gamma_y(1) & \cdots & \Gamma_y(s-1) \\ \Gamma_y(-1) & \Gamma_y(0) & \cdots & \Gamma_y(s-2) \\ \vdots & & & \vdots \\ \Gamma_y(-s+1) & \Gamma_y(-s+2) & \cdots & \Gamma_y(0) \end{bmatrix},$$

which implies that $\delta'\widetilde{\Phi}' \neq 0$. In view of (9), we conclude that the pseudo-causal VAR in (8) generally does not have any (pseudo) CFs. It is easy to see that the reverse is also true, i.e., the noncausal representation of a causal VAR with CFs does not generally exhibit the (pseudo) CFs.

A special case occurs when $p = 1$. Equations (7) and (9) simplify to

$$\begin{aligned} \Phi' &= \Gamma_y(-1)[\Gamma_y(0)]^{-1} = \delta_\perp \Upsilon_y'(-1)[\Gamma_x(0)]^{-1}, \\ \widetilde{\Phi}' &= \Gamma_y(1)[\Gamma_y(0)]^{-1} = \Upsilon_y(1)\delta_\perp'[\Gamma_z(0)]^{-1}, \end{aligned}$$

from which we see that the existence of k CFs in a MVAR$(0,1)$ implies the presence of k CFs in the pseudo-causal representation of series Y_t. It is easy to check that the reverse is true as well. We summarize the above results in the following proposition.

Proposition 5. *If series Y_t are generated by a second-order stationary $MVAR(r,0)$ [$MVAR(0,s)$] with $r>1$ [$s>1$], the existence of k CFs does not imply the presence of any CF in the pseudo-noncausal [pseudo-causal] representation of series Y_t.*

With regard to statistical inference, a relevant implication of the above proposition is that, except for the case of VAR models with either one lead or lag, the Gaussian ML approach is able to identify the unique representation of series Y_t that exhibit the common features. As is well known, the Gaussian likelihood analysis of the reduced rank regression model (6) is based on canonical correlations between Y_t and X_t, denoted as $\text{CanCor}(Y_t, X_t)$; see e.g. Johansen (2008) and the references therein.

Remark 6. *In case of the $MVAR(1,0)$ and $MVAR(0,1)$ model, it is impossible to discriminate between causal and noncausal CFs using the Gaussian ML analysis. Indeed, by the assumption of stationarity and noting that autocorrelation is a symmetric measure, it follows that*

$$\text{CanCor}(Y_t, Y_{t+1}) = \text{CanCor}(Y_{t-1}, Y_t) = \text{CanCor}(Y_t, Y_{t-1}).$$

Consequently, even if there exists a NCCF, we will also detect a CCF using canonical correlations. However, although the eigenvalues from the solutions of $\text{CanCor}(Y_t, Y_{t+1})$ and $\text{CanCor}(Y_t, Y_{t-1})$ are identical, the eigenvectors, and hence the relationships that link series, will be different. This might explain why "strange" estimates of common cycle relationships are sometimes obtained in empirical works.

2.4 Co-movements in mixed models

Let us now consider the general case of a $\text{MVAR}(r,s)$ where both r and s are larger than zero. As shown by Lanne and Saikkonen (2013), series Y_t admit a pseudo-causal $\text{MVAR}(p,0)$ representation, where $p=r+s$, with the same spectral density as (3) and with errors that are multivariate white noise but not innovations with respect to the past. In view of (4), we see that when both r and s are positive, the autocorrelation matrix function of series Y_t reads

$$\Gamma_y(j) = \sum_{i=-\infty}^{\infty} \Psi_{i+j} \Omega \Psi_i' = \Omega \Psi_{-j}' + \sum_{\forall i \neq -j} \Psi_i \Omega \Psi_{i-j}',$$

which implies that the coefficient matrix of the corresponding pseudo-causal VAR

$$[\widetilde{\Phi}_1, ..., \widetilde{\Phi}_p] = [\Gamma_y(1), ..., \Gamma_y(p)][\Gamma_z(0)]^{-1},$$

will generally be full rank. It is easy to see that the same conclusion applies to the pseudo-noncausal representation as well. We summarize the above results in the following proposition.

Proposition 7. *If series Y_t are generated by a second-order stationary $MVAR(r, s)$ with $r > 0$ and $s > 0$, the existence of k CFs does not imply the presence of any CFs in the pseudo-noncausal [pseudo-causal] representation of series Y_t.*

Notice that a simple solution to detecting CFs in mixed models such as $\text{CanCor}\,(Y_t, U_t)$, where $U_t' = [X_t', Z_t']$, cannot be adopted. The reason is that in the associated reduced rank regression model

$$Y_t = \delta_\perp C' U_t + \Theta_0^{-1} \varepsilon_t,$$

where C is a full rank $np \times (n - k)$ matrix and $\Theta_0 \equiv \sum_{j=-\min(r,s)}^{\min(r,s)} \Pi_j \Phi_j$, the regressors are correlated with the errors.

3 Test statistics

In order to test for common cyclical features in purely causal or noncausal models, we first determine the lag order p in standard VAR models by means of various information criteria, see Section 4 for details. Even in the presence of a noncausal component, we obtain an estimate of the pseudo-causal order $p = s$. Having fixed p, we perform both Likelihood Ratio (LR) and GMM tests to detect the presence of co-movements both in purely causal and noncausal models.

3.1 LR tests

The Gaussian LR test is based on the partial canonical correlations between Y_t and W_t, where

$$W_t = \begin{cases} [Y_{t-1}', ..., Y_{t-p}']' & \text{for the causal VAR,} \\ [Y_{t+1}', ..., Y_{t+p}']' & \text{for the noncausal VAR,} \end{cases}$$

having concentrated out the effects of the deterministic terms D_t (e.g., intercepts and dummies). This procedure is denoted as

$$\text{CanCor}\left\{Y_t, W_t \mid D_t\right\}, \tag{10}$$

The LR test for null hypothesis that there exists at least k common feature vectors is based on the statistic

$$\xi_{LR} = -T\sum_{i=1}^{k} \ln(1-\hat{\lambda}_i), \qquad k = 1, 2, ..., n,$$

where $\hat{\lambda}_i$ is the i-th smallest squared canonical correlations computed from (10). More specifically, $\hat{\lambda}_i$ is the i-th smallest eigenvalue of

$$\hat{\Sigma}_{WW}^{-1/2}\hat{\Sigma}_{WY}\hat{\Sigma}_{YY}^{-1}\hat{\Sigma}_{YW}\hat{\Sigma}_{WW}^{-1/2},$$

where $\hat{\Sigma}_{AB}$ denotes the sample covariance matrix of two vector time series A_t and B_t. Under the null hypothesis, ξ_{LR} follows asymptotically a $\chi^2(\nu)$ distribution with $\nu = knp - k(n-k)$.

3.2 GMM tests

When the time series Y_t is bivariate to simplify notation, i.e., $Y_t' = [y_{1t}, y_{2t}]$, a GMM single equation common feature test can be conducted. First, we write $\delta = [1, -\alpha]'$ and estimate α by the instrumental variable (IV) estimator using W_t as instruments. The IV estimator of α is given by

$$\hat{\alpha}_{IV} = \left(\mathbf{y}_2'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}_2\right)^{-1}\left(\mathbf{X}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}_1\right),$$

where $[\mathbf{y}_1, \mathbf{y}_2] \equiv \mathbf{Y}$ and $\mathbf{W}$ respectively indicate the matrices of $T-p$ observations of variables Y_t and W_t after having removed the linear influence of the deterministic terms D_t. Second, we run an overidentification J-test (Hansen, 1982) for the validity of the orthogonality conditions, i.e. the existence of a common feature relationship, using the statistic

$$J_1(\widehat{\alpha}_{IV}) = (\mathbf{u}'\mathbf{W})(\hat{\sigma}_u^2\mathbf{W}'\mathbf{W})^{-1}(\mathbf{W}'\mathbf{u}),$$

where $\mathbf{u} = \mathbf{y}_1 - \mathbf{y}_2\widehat{\alpha}_{IV}$ and $\hat{\sigma}_u^2 = \mathbf{u}'\mathbf{u}/(T-p)$.

The tests presented so far embed the assumption of homoscedasticity. Candelon et al (2005) have illustrated in a Monte Carlo exercise that $J_1(\widehat{\alpha}_{IV})$ has large size distortions in the presence of GARCH disturbances. Therefore, we also use the White heteroskedasticity robust estimator (see, e.g., Hamilton, 1994):

$$\hat{\alpha}_{IV-W} = \left(\mathbf{y}_2'\mathbf{W}(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}_2\right)^{-1}\left(\mathbf{y}_2'\mathbf{W}(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1}\mathbf{W}'\mathbf{y}_1\right), \tag{11}$$

where $\mathbf{B} = \text{diag}(u_1^2, u_2^2, ...u_T^2)$ and $u_t = y_{1t} - \hat{\alpha}_{IV}' y_{2t}$. Then we run a J-test robust to heteroskedasticity using the statistic

$$J_2(\hat{\alpha}_{IV-W}) = (\mathbf{u}^{*\prime}\mathbf{W})(\mathbf{W}'\mathbf{B}\mathbf{W})^{-1}(\mathbf{W}'\mathbf{u}^*). \tag{12}$$

where $\mathbf{u}^* = \mathbf{y}_1 - \mathbf{y}_2\hat{\alpha}_{IV-W}$.[5] Under the null hypothesis, both $J_1(\widehat{\alpha}_{IV})$ and $J_2(\widehat{\alpha}_{IV-W})$ follow asymptotically a $\chi^2(2p-1)$ in the bivariate case with $k=1$.

4 Monte Carlo results

4.1 The pseudo-causal VAR(p) order

We first investigate by Monte Carlo simulations the frequency with which various information criteria select the correct order p of the pseudo-causal VAR(p). The data generating process is a MVAR(1,2) $(1-\Pi_1 L)(1-\Phi_1 L^{-1} - \Phi_2 L^{-2})Y_t = \varepsilon_t$, with the following coefficient matrices

$$\Pi_1 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.7 \end{bmatrix}, \tag{13}$$

$$\Phi_1 = \begin{bmatrix} 0.5 & 0.2 \\ 0.5 & 0.2 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -0.35 & 0.25 \\ -0.35 & 0.25 \end{bmatrix}. \tag{14}$$

Note that the presence of reduced rank matrices in the noncausal part (14) has no implications for the order of the pseudo-causal VAR representation. We consider two relatively small sample sizes, $T = \{100, 250\}$.[6] For ε_t, we consider the following correlated distributions in addition to the multivariate $N(0,1)$: (*i*) Student's $t(\nu) = \zeta_t \odot (\nu/\chi^2(\nu))^{1/2}$ and (*ii*) Laplace $= \zeta_t \odot (-\ln(U[0,1]))^{1/2}$, where $\odot$ denotes

[5] Alternative approaches for handling heteroskedastic disturbances have been evaluated in Hecq and Issler (2012).
[6] We simulate an additional 50×2 observations (both sides) burn in period to initialize processes.

the Hadamar product. The ζ_t are simulated according to

$$\zeta_t \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}\right). \tag{15}$$

For (i) we consider the following values for the degrees of freedom $\nu = \{10, 3, 1\}$, where $\nu = 1$ corresponds to the Standard Cauchy case. Each distribution has its own seed. The four information criteria are AIC, BIC, HQ and the Takeuchi information criterion (TIC).[7] We let them find their minimum for $p = 0$ to $p_{\max} = 8$.

Table 1 reports the results over 10,000 replications. Overall, information criteria determine the pseudo-causal VAR order rather well for $T = 250$ (in particular HQ). As usual BIC detects a proportion of models with too parsimonious dynamics. This is clearly a problem for the identification of mixed causal-noncausal models because r lags and s leads are going to be estimated within that p. AIC tends to overestimate the number of lags. The TIC does not perform very well by construction when the kurtosis does not exist, e.g., the $t(3)$. The frequency with which the correct model is found is approaching 100% when $T > 250$ for instance (except for the Cauchy).

4.2 Testing for common cyclical features

The previous subsection has illustrated that usual information criteria can guide us towards the right choice for the pseudo-causal VAR order p. Let us now focus on common cyclical feature test statistics using the tests reviewed above. For this study, we restrict our attention to two different data generating processes: a purely causal and purely noncausal VAR. More precisely, we consider a noncausal bivariate MVAR(0,2) with a reduced rank in coefficient matrices such as (14). For the causal MVAR(2,0) we have the same parameters in the first two lags, i.e., $\Pi_1 = \Phi_1$ and $\Pi_2 = \Phi_2$. Under the null hypothesis, there is a CCF [NCCF] relationship with a cofeature vector $\delta' = [1, -1]$. It should be noted that although we

[7] The TIC has been proposed by Takeuchi (1976) as an alternative to the AIC as it does not rely on the assumption that the true model is included in the set of investigated specifications. He obtained the following criterion

$$\text{TIC} = \ln(\widehat{\Sigma}) + 2\widehat{\eta}/T,$$

where

$$\widehat{\eta} = \sum_{t=1}^{T} \widehat{\varepsilon}_t' \widehat{\Sigma}^{-1} \widehat{\varepsilon}_t h_t + \frac{1}{2}\left[T^{-1} \sum_{t=1}^{T} \left(\widehat{\varepsilon}_t' \widehat{\Sigma}^{-1} \widehat{\varepsilon}\right)^2 - n(n+2)\right],$$

$h_t = X_t' \left(\sum_{i=1}^{t} X_t X_t'\right)^{-1} X_t$, $\widehat{\varepsilon}_t$ are the OLS residuals, and $\widehat{\Sigma}$ is the residual covariance matrix (see Yanagihara, 2006, for further details).

Normal	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$\mathbf{p = 3}$	**26.07**	**56.01**	**67.30**	**37.56**	**82.17**	**96.15**	**86.98**	**94.38**
$p < 3$	73.83	42.01	17.39	61.96	17.80	2.68	0.26	5.16
$t(10)$	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$\mathbf{p = 3}$	**26.38**	**55.38**	**68.00**	**38.07**	**81.67**	**96.25**	**86.57**	**93.67**
$p < 3$	73.55	42.69	17.33	61.47	18.32	2.69	0.26	5.78
$t(3)$	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$\mathbf{p = 3}$	**27.25**	**55.29**	**67.03**	**37.20**	**80.22**	**95.05**	**86.49**	**81.89**
$p < 3$	72.55	42.82	18.20	59.35	19.63	3.67	0.62	16.24
$t(1)$	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$\mathbf{p = 3}$	**41.77**	**55.46**	**61.44**	**26.29**	**71.21**	**78.76**	**77.80**	**35.28**
$p < 3$	53.87	35.27	19.83	57.15	24.76	13.25	6.62	53.60
Laplace	$T = 100$				$T = 250$			
	BIC	HQ	AIC	TIC	BIC	HQ	AIC	TIC
$\mathbf{p = 3}$	**29.03**	**57.88**	**69.15**	**39.87**	**82.56**	**96.10**	**87.07**	**92.21**
$p < 3$	70.91	40.30	16.20	55.55	17.44	2.81	0.30	7.22

Table 1: Frequencies with which ICs find the pseudo-causal VAR(3) for a MVAR(1,2) DGP

have strong co-movements between series, it is impossible to figure out graphically whether we have a causal or a noncausal pattern. To illustrate this point, Figures 1 and 2 compare causal and noncausal VAR(2) models with SCCF restrictions for different distributions.

We can now report the frequencies with which the three common feature test statistics that we propose reject the null that a combination of series is orthogonal to the past for the causal model or to the future for the noncausal model. At a 5% significance level the null follows a $\chi^2(3)$ under normality for a VAR(2) with one SCCF vector. An entry in the vicinity of 5% denotes an absence of size distortion. Results are based on 10,000 replications. Tables 2 and 3 reveal that, regardless of the distribution chosen to simulate processes, both canonical correlation and GMM tests do not suffer from severe size distortions. To obtain this result the "correct"[8] set of instruments must be used when performing those tests, namely orthogonality conditions to $\{Y_{t-1}, Y_{t-2}\}$ $[\{Y_{t+1}, Y_{t+2}\}]$ for the causal [noncausal] model. This result is very encouraging because it means that purely noncausal common features can be easily found by reverting the timeline of the instruments. The only test that faces a problem is the robust GMM when

[8]Note that we solely mean lags for the causal model and leads for the noncausal model. We do not comment on under- or overspecification of these lag and lead orders.

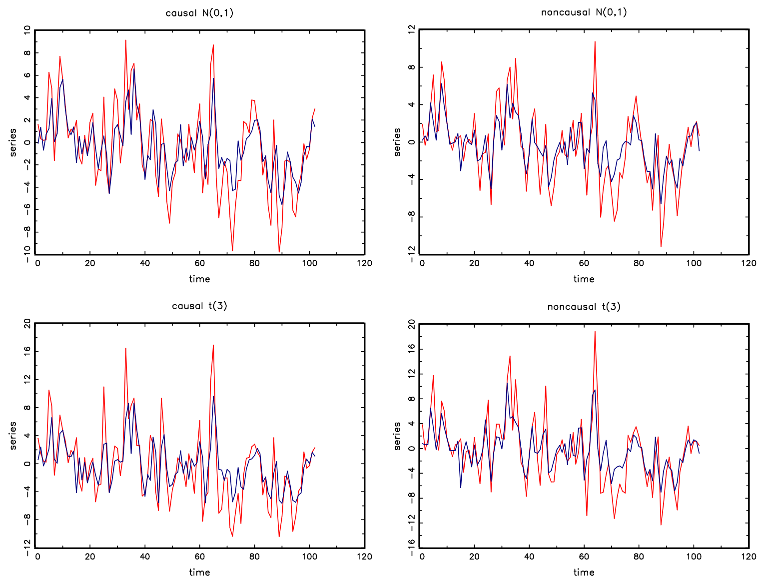

Figure 1: Causal and noncausal VAR(2) with SCCF - N(0,1) and t(3)

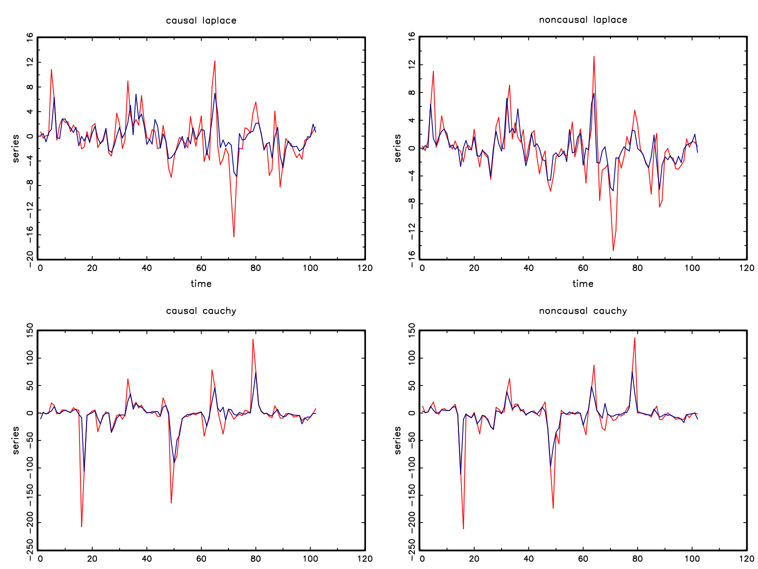

Figure 2: Causal and noncausal VAR(2) with SCCF - Cauchy and Laplace

	$T = 100$					
DGP/instruments	$\{Y_{t-1}, Y_{t-2}\}$			$\{Y_{t+1}, Y_{t+2}\}$		
Causal VAR(2)	Cancor	GMM	GMM^W	Cancor	GMM	GMM^W
$N(0,1)$	5.64	4.91	4.63	23.52	29.45	26.64
$t(10)$	5.37	4.82	4.58	23.47	29.52	25.40
$t(3)$	5.50	5.02	3.80	21.55	27.77	18.13
Cauchy	5.68	5.42	0.97	23.06	29.38	5.38
Laplace	5.83	5.10	4.30	27.67	38.70	31.51
	$T = 500$					
	$\{Y_{t-1}, Y_{t-2}\}$			$\{Y_{t+1}, Y_{t+2}\}$		
Causal VAR(2)	Cancor	GMM	GMM^W	Cancor	GMM	GMM^W
$N(0,1)$	5.16	5.08	4.56	91.84	92.54	91.99
$t(10)$	4.98	4.84	4.78	91.90	92.54	91.41
$t(3)$	4.84	4.72	4.15	89.17	89.89	78.32
Cauchy	4.76	4.73	1.00	67.46	69.22	12.61
Laplace	5.27	5.20	4.98	95.62	96.36	95.18

Table 2: Empirical size/power for the causal VAR(2) DGP

the Cauchy distribution is implemented. This is not surprising because in a heteroscedasticity robust correction à la White, it is assumed that the unconditional variance exists. In practice, the squares of the residuals cause problems in the Cauchy case when large "outlying" observations are introduced via the $\mathbf{B} = \text{diag}(u_1^2, u_2^2, ...u_T^2)$ matrix.

When the wrong set of instruments is used, namely only past variables for the noncausal model or the future variables for the causal specification, the presence of co-movements measured by a reduced rank in the dynamics of the systems is rejected. The rejection frequencies increase with the sample size, indicating that we are under the alternative indeed. Because we cannot distinguish whether co-movements are present in a causal or a noncausal framework by solely graphical means, we recommend to first determine the VAR order p (potentially the pseudo-causal model) in empirical works and then to apply the common cyclical feature test statistics in both directions.

As an additional identification tool, we notice that one should reject the null hypothesis that the errors are *i.i.d.* if a pseudo-causal model is estimated when the data are generated by a noncausal model. Consequently, we can run the regression

$$\hat{\varepsilon}_t = \mu + \Xi\hat{\varepsilon}_{t-1}^2 + u_t,$$

			$T = 100$			
DGP/instruments	$\{Y_{t-1}, Y_{t-2}\}$			$\{Y_{t+1}, Y_{t+2}\}$		
Noncausal VAR(2)	Cancor	GMM	GMM^W	Cancor	GMM	GMM^W
$N(0,1)$	22.92	28.79	25.74	5.58	5.00	4.73
$t(10)$	22.86	28.88	24.50	5.74	5.05	4.46
$t(3)$	22.06	28.39	18.39	5.39	4.83	3.62
Cauchy	23.10	29.86	5.20	5.65	5.47	1.01
Laplace	27.02	38.69	30.94	5.45	4.86	4.23
			$T = 500$			
	$\{Y_{t-1}, Y_{t-2}\}$			$\{Y_{t+1}, Y_{t+2}\}$		
Noncausal VAR(2)	Cancor	GMM	GMM^W	Cancor	GMM	GMM^W
$N(0,1)$	91.39	92.06	89.55	5.13	4.94	5.02
$t(10)$	91.16	91.81	91.07	5.01	4.88	4.28
$t(3)$	89.64	90.43	78.77	5.11	4.98	4.17
Cauchy	67.93	69.55	12.72	4.55	4.53	1.00
Laplace	95.72	96.43	95.31	4.99	4.86	4.63

Table 3: Empirical size/power for the noncausal VAR(2) DGP

where $\hat{\varepsilon}_t$ are the residuals from the pseudo-causal VAR(p) model, and test for the null hypothesis H_0 : $\text{vec}(\Xi) = 0$. We investigate the behavior of a Wald test that is based on the statistic

$$\xi_W^{iid} = \text{vec}(\hat{\Xi})'(\widehat{\Sigma}_{uu} \otimes (\mathbf{W}'\mathbf{W})^{-1})^{-1}\text{vec}(\hat{\Xi}),$$

where $\widehat{\Sigma}_{uu}$ is the sample covariance matrix of the disturbance term u_t and $\mathbf{W}$ is the matrix of observations of the demeaned regressor $\hat{\varepsilon}^2_{t-1}$. We can also consider a LR test, which is based on the canonical correlations between $\hat{\varepsilon}_t$ and $\hat{\varepsilon}^2_{t-1}$. The LR test statistic is given by

$$\xi_{LR}^{iid} = -T\sum_{i=1}^{m} \ln(1 - \hat{\lambda}_i),$$

where $\hat{\lambda}_i$ is the i-th smallest squared canonical correlation coming from CanCor $\left\{\hat{\varepsilon}_t, \hat{\varepsilon}^2_{t-1} \mid 1\right\}$. Under the null hypothesis, both ξ_W^{iid} and ξ_{LR}^{iid} are asymptotically distributed as $\chi^2(n^2)$.

In order to evaluate the extent to which the above tests are able to determine whether we have a causal or a noncausal VAR, we use the same DGPs (either causal and noncausal) as before, then we estimate a (pseudo-)causal VAR(2) model and use both the statistics ξ_W^{iid} and ξ_{LR}^{iid}. The results can be found in Table 4. We obtain the well-known result that, under Gaussianity, causal and noncausal models

	$T = 100$		$T = 500$	
Causal VAR(2)	*Wald*	*LR*	*Wald*	*LR*
$N(0, 1)$	6.23	5.50	5.18	5.08
$t(10)$	5.89	5.16	5.14	4.96
$t(3)$	5.68	5.14	5.32	5.22
Cauchy	4.76	4.52	3.25	3.22
Laplace	5.66	5.05	5.51	5.29
	$T = 100$		$T = 500$	
Noncausal VAR(2)	*Wald*	*LR*	*Wald*	*LR*
$N(0, 1)$	5.71	4.85	5.14	5.00
$t(10)$	10.68	9.56	13.22	12.85
$t(3)$	42.67	40.58	75.35	75.13
Cauchy	81.20	80.20	98.57	98.56
Laplace	25.24	23.57	33.51	33.04

Table 4: Empirical size/power for the causal vs. noncausal VARs

cannot be distinguished. The rejection frequencies are in the vicinity of 5% for both sample sizes and both tests. When the model is causal we do not reject the null of *i.i.d.*-ness using both the Wald and the LR approaches. The rejection frequencies are all around 5% regardless of the distribution considered. The only exception is the Cauchy distribution which is slightly undersized. It is likely that other critical values should be considered in this case. When the true process is a noncausal VAR, both the Wald and the LR are able to reject the *i.i.d.* null hypothesis in large samples when the departure from the Normal distribution is large. The (non-size-adjusted) power is pretty low for $T = 100$, however with rejection frequencies of about 25% and around 33% with $T = 500$ for the Laplace. The $t(10)$ induces low power, which is expected as this specification is the closest to Gaussianity. The power is relatively high for the Cauchy and the $t(3)$. Alternative tests are worth investigating, but this is out of the scope of this paper.

5 Empirical illustrations

5.1 Price indicators

We first consider 20 price series from "The Department Store Inventory Price Index" dataset. Those data are based on inventory weighted price indices of goods carried by department stores. The Bureau of Labor Statistics (BLS) maintained these statistics until the end of 2013. We work with monthly data from January 1980 to December 2013. Indices are available for the following 23 items: Piece goods, Domestics and draperies, Women's and children's shoes, Men's shoes, Infants' wear and furniture, Women's underwear, Women's and girls' hosiery, Women's and girls' accessories, Women's outerwear & girls' wear, Men's clothing, Men's furnishings, Boys' wear, Jewelry and silverware, Notions, Toilet articles & drugs, Furniture and bedding, Floor coverings, Housewares, Major appliances, Radios and television sets. We do not include the last three items of the dataset in our analysis (Recreation & education, Home improvements, Automotive accessories) because those series only start in 1986.

Common cyclical feature test statistics have severe size distortions and low power when a wrong model for the seasonality is considered (Hecq, 1998; Cubadda, 1999). We refrain from using standard seasonal adjustment methods such as X13-ARIMA and TRAMO/SEATS, as they directly affect (partial) autocorrelation functions which might create spurious causal and noncausal dynamics (see Hecq, Telg and Lieb, 2016B). We assume for this application that there is only one unit root at the zero frequency and that seasonality is deterministic. This means that we take the monthly growth rates $\Delta \ln P_{i,t}$ $(i = 1, ..., 20)$ in deviation from twelve deterministic monthly dummy variables. We construct 190 bivariate models from the 20 residual series. We have the following lag lengths for each of the 190 pairs of VARs: $p = 0$ (31%), $p = 1$ (5%), $p = 2$ (37%), $p = 3$ (5%), $p \geq 4$ (21%). There are consequently 59 pairs that are multivariate white noise and for which we should find both a causal and a noncausal common feature relationship. We use $p = 4$ and test for CCF and NCCF using the LR test. We do not reject the null of 74 causal common feature relationships and 88 noncausal ones. This means that on the top of 59 multivariate white noise we have 15 additional causal SCCF relationship and 30 noncausal relationships. Obviously the impact of the assumption of the stochastic versus the deterministic seasonality is an important issue to investigate. We do not identify mixed models as we have seen (Section 2.4) that we would not have detected the presence of common cycles if the processes where mixed causal-noncausal.

5.2 Commodity prices co-movements

For the second application we consider raw monthly commodity prices observed during the period January 1992 to October 2016, i.e., 298 observations, released by the IMF.[9] These are benchmark prices which are representative of the global market. They are determined by the largest exporter of a given commodity. IMF releases many different individual commodity prices but we only focus on the following three indices:

[9]IMF Primary Commodity Prices, see http://www.imf.org/external/np/res/commod/index.aspx.

- RAWM: Agricultural Raw Materials Index, 2005 = 100, includes Timber, Cotton, Wool, Rubber, and Hides Price Indices. IMF name: PRAWM;
- GAS: Natural Gas, Indonesian Liquefied Natural Gas in Japan, US Dollar per Million Metric British Thermal Unit. IMF name: PNGASJP;
- OIL: Crude Oil (Petroleum), Price index, 2005 = 100, simple average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh. IMF name: POILAPSP.

We consider the vector of series

$$Y_t = (\Delta \ln RAWM_t, \Delta \ln GAS_t, \Delta \ln OIL_t)',$$

where the monthly growth rates of the series are taken as we do not reject the zero frequency unit root in all three cases. The resulting series are displayed in Figure 3.

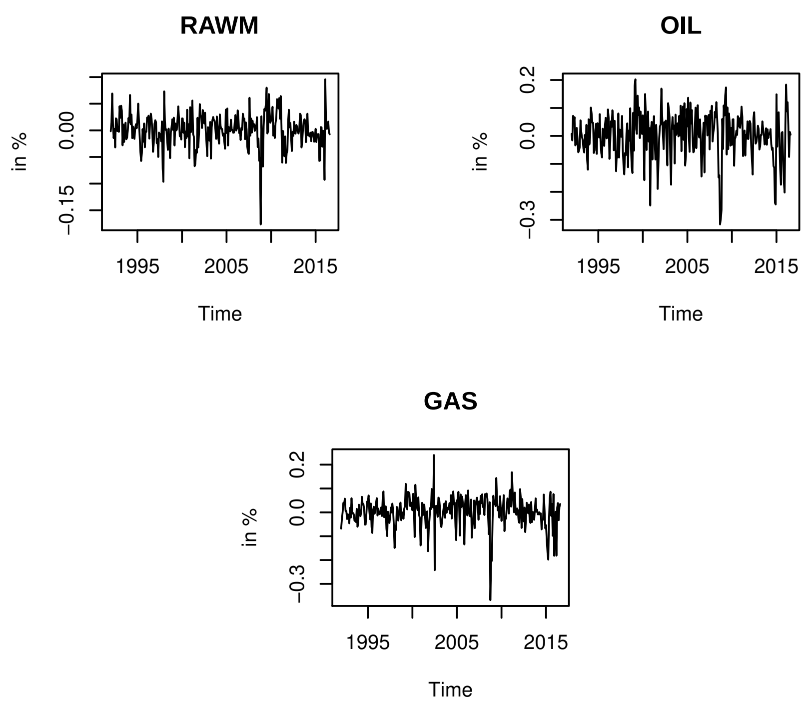

Figure 3: Growth rates of Raw Materials, Crude Oil and Natural Gas

The pseudo-causal VAR order is detected to be $p = 12$ according to AIC in combination with cross correlogram analysis and seasonal dummies are included because they are jointly significant according to both ordinary (s.e.$_{ols}$) and heteroscedasticity robust (s.e.$_{hcse}$) F-tests. To look at the robustness of our results we have also considered Y_t^{SA} where series are taken in deviation from a regression on seasonal dummies. There are signs of non-normality according to the Jarque-Bera (JB) test. Table 5 displays p-values of the above tests.

Series	p	JB-test			F-test					
					s.e.$_{ols}$	s.e.$_{hcse}$	s.e.$_{ols}$	s.e.$_{hcse}$	s.e.$_{ols}$	s.e.$_{hcse}$
Y_t	12	0.000	0.000	0.151	0.000	0.001	0.769	0.898	0.329	0.426
Y_t^{SA}	12	0.000	0.000	0.124	-	-	-	-	-	-

Table 5: Summary statistics of the (pseudo-)causal VAR

It is clear from Table 6 that both the canonical correlation test ξ_{LR} and the robust GMM J_2-test do not reject the null of a common feature vector in the noncausal framework while it is rejected (at 5%) in the purely causal VAR. Eigenvalues are also very interesting to investigate. In the causal VAR there is not much difference between the first two smallest eigenvalues. Therefore, it is difficult to claim that the first value is not different from zero while the second one is. This is directly in contrast with the noncausal case as we do observe a large difference between them.

	CCF$_{s=1}$			NCCF$_{s=1}$		
	CanCor$_{\xi_{LR}}$	$\hat{\lambda}_i$	GMM$_{J_2}$	CanCor$_{\xi_{LR}}$	$\hat{\lambda}_i$	GMM$_{J_2}$
Y_t	0.015	(0.172, 0.186, 0.513)	0.031	0.333	(0.121, 0.221, 0.567)	0.090
Y_t^{SA}	0.044	(0.158, 0.196, 0.520)	0.013	0.716	(0.096, 0.272, 0.553)	0.363

Table 6: P-values of tests - causal vs. noncausal VARs

The presence of noncausal common features detected in both applications can imply that there exist some commonalities in the bubble behavior of several variables. It can also be the sign of collinear nonlinear impulse response functions as noncausal linear VARs have a nonlinear causal interpretation (Gourieroux and Zakoïan, 2016).

6 Conclusion

This paper provides a novel explanation for the inability of detecting common features in a VAR framework. We show that common features that cannot be detected in the causal (i.e., backward-looking) representation might be revealed in the noncausal (i.e., forward-looking) dynamics of the series. Whereas the difference between data generated from purely causal, purely noncausal and mixed causal-noncausal VARs is not always graphically visible, it is well-known that the latter two are able to generate much richer dynamics than the conventional causal VAR. This makes it possible to detect for example speculative bubbles and asymmetric cycles which are common to a set of series.

We propose tools to highlight the potential presence of noncausal co-movements. Using both sets of lag and lead instruments within a canonical correlation or a GMM framework, we show how additional relationships are discovered between series both in Monte Carlo simulations and two empirical illustrations. Interestingly, we find that the presence of a reduced rank structure in the dynamics of purely causal and noncausal systems with more than one lag/lead permits identification in the usual Gaussian framework. This result cannot be extended to mixed causal-noncausal models, as the instruments are correlated with the error term of the reduced rank model by construction. We identify common bubbles in both commodity prices and price indicators.

References

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

[2] Brockwell, P. and R. Davis (1991), *Time Series: Theory and Methods*, Springer-Verlag New York, Second Edition.

[3] Candelon, B., Hecq, A. and W. Verschoor (2005), Measuring Common Cyclical Features During Financial Turmoil, *Journal of International Money and Finance*, 24, 1317-1334.

[4] Cubadda, G. (1999), Common cycles in seasonal non-stationary time series, *Journal of Applied Econometrics*, 14, 273-291.

[5] CUBADDA, G. (2007), A unifying framework for analysing common cyclical features in cointegrated time series, *Computational Statistics & Data Analysis*, 52, 896–906.

[6] CUBADDA ,G., AND A. HECQ (2001), On non-contemporaneous short-run co-movements, *Economics Letters*, 73, 389-397.

[7] ENGLE, R. F. AND S. KOZICKI (1993), Testing for Common Features (With Comments), *Journal of Business and Economic Statistics* 11, 369-395.

[8] GOURIÉROUX, C. AND J. JASIAK (2015), Semi-Parametric Estimation of Noncausal Vector Autoregression, *CREST Working Paper, 2015-02*.

[9] GOURIÉROUX, C., AND J.M. ZAKOÏAN (2016), Local Explosion Modelling by Noncausal Process, *Journal of the Royal Statistical Society, Series B*, doi:10.1111/rssb.12193.

[10] GUILLÉN, O., HECQ, A., ISSLER, J. AND D. SARAIVA (2015), Forecasting Multivariate Time Series Under Present-Value Model Short- and Long-Run Co-Movement Restrictions, *International Journal of Forecasting, Elsevier*, 31(3), 862-875.

[11] HAMILTON, J. (1994), *Time Series Analysis*, Princeton University Press.

[12] HANSEN, L.P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50(4), 1029-1054.

[13] HECQ, A. (1998), Does Seasonal Adjustment Induce Common Cycles?, Economics Letters, 59, 289-297.

[14] HECQ, A. AND J.V. ISSLER (2012), A Common-Feature Approach for Testing Present-Value Restrictions with Financial Data, *METEOR RM/12/006*, Maastricht University.

[15] HECQ, A., LIEB, L. AND S. TELG (2016A), Identification of Mixed Causal-Noncausal Models in Finite Samples, *Annals of Economics and Statistics*, 123-124, 307-331.

[16] HECQ, A., TELG, S. AND L. LIEB (2016B), Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?, *MPRA Paper 74922*, University Library of Munich, Germany.

[17] Hecq, A., Palm F.C. and J.P. Urbain (2006), Testing for Common Cyclical Features in VAR Models with Cointegration, *Journal of Econometrics* 132(1), 117-141.

[18] Hencic, A. and C. Gouriéroux (2014), Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate, *Econometrics of Risk, Series: Studies in Computational Intelligence,* Springer International Publishing, 17-40.

[19] Issler, J. and F. Vahid (2001), Common cycles and the importance of transitory shocks to macroeconomic aggregates, *Journal of Monetary Economics*, 47(3), 449-475.

[20] Johansen, S. (2008), Reduced Rank Regression, *The New Palgrave Dictionary of Economics* (2nd Edition), Durlauf, S.N. and L.E. Blume (eds.), doi:10.1057/9780230226203.1410.

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

[22] Lanne, M. and P. Saikkonen (2011A), Noncausal Autoregressions for Economic Time Series, *Journal of Time Series Econometrics*, 3(3), 1-32.

[23] Lanne, M. and P. Saikkonen (2011B), GMM Estimation with Noncausal Instruments, *Oxford Bulletin of Economics and Statistics*, 73(5), 581-592.

[24] Lanne, M. and P. Saikkonen (2013), Noncausal Vector Autoregression, *Econometric Theory*, 29(3), 447-481.

[25] Takeuchi, K. (1976), Distribution of informational statistics and a criterion of model fitting, *Suri-Kagaku (Mathematic Sciences)*, 153, 12-18 (in Japanese).

[26] Vahid, F. and R. Engle (1993), Common Trends and Common Cycles, *Journal of Applied Econometrics*, 8(4), 341-360.

[27] Vahid, F. and R. Engle (1997), Codependent cycles, *Journal of Econometrics*, 80(2), 199-221.

[28] Yanagihara, H. (2006), Corrected version of AIC for selecting multivariate normal linear regression models in a general nonnormal case, *Journal of Multivariate Analysis*, 97, 1070-1089.