Detecting Co-Movements in Non-Causal Time Series*

GIANLUCA CUBADDA, † ALAIN HECQ‡ and SEAN TELG‡

†Dipartimento di Economia e Finanza, Universita’ di Roma “Tor Vergata”, Via Columbia 2, 00133 Roma, Italy (e-mail: gianluca.cubadda@uniroma2.it)
‡Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands (e-mail: a.hecq@maastrichtuniversity.nl; j.telg@maastrichtuniversity.nl)

Abstract

This paper introduces the notion of common non-causal features and proposes tools to detect them in multivariate time series models. We argue that the existence of co-movements might not be detected using the conventional stationary vector autoregressive (VAR) model as the common dynamics are present in the non-causal (i.e. forward-looking) component of the series. We show that the presence of a reduced rank structure allows to identify purely causal and non-causal VAR processes of order $P > 1$ even in the Gaussian likelihood framework. Hence, usual test statistics and canonical correlation analysis can be applied, where either lags or leads are used as instruments to determine whether the common features are present in either the backward- or forward-looking dynamics of the series. The proposed definitions of co-movements are also valid for the mixed causal—non-causal VAR, with the exception that a non-Gaussian maximum likelihood estimator is necessary. This means however that one loses the benefits of the simple tools proposed. An empirical analysis on Brent and West Texas Intermediate oil prices illustrates the findings. No short run co-movements are found in a conventional causal VAR, but they are detected when considering a purely non-causal VAR.

I. Introduction

A ‘feature’ is defined as a dominant characteristic observed in univariate time series. One might be interested in testing whether certain time series contain the same feature. We

JEL Classification numbers: C12, C32, E32.

*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), the Scientific Meeting of the Italian Statistical Society (Palermo, June 2018) 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. Additionally, we are greatly in debt to Lenard Lieb for fruitful discussions and two anonymous referees for valuable comments and suggestions.

© 2018 The Authors. *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
say that a feature is *common* if a linear combination of multiple series no longer has the feature, while it is present in each of the series individually. There are ample examples of common features in the literature: e.g. cointegration, common ARCH, co-breaking and common nonlinearity. In this paper, we focus on the presence of common cyclical features as introduced by Engle and Kozicki (1993) and Vahid and Engle (1993). That is, we are interested in extracting commonalities in the dynamics of multivariate time series models. This is commonly done by imposing short run restrictions, which are known to have direct implications for forecasting accuracy and parameter efficiency as less coefficients have to be estimated. Common features restrictions are also directly implied by economic theories such as the present value model or the permanent income hypothesis (Guillén et al., 2015; Issler and Vahid, 2001). Impulse response functions are collinear when common transmission mechanisms are present and economies with strong nexus generally show statistical similarities in their business cycle fluctuation phases.

This paper contributes to this literature by studying the existence of common cyclical features in economic and financial variables for both causal and non-causal vector autoregressive (VAR) models. In our terminology, the conventional vector autoregressive model popularized by Sims (1980) is called causal, as the variables of interest only depend on their own past values. The non-causal VAR has been investigated in the literature to a much lesser extent and is defined as the analogue model in reverse time (see e.g. Davis and Song, 2012; Lanne and Saikkonen, 2013; Gouriéroux and Jasiak, 2017). In this framework, variables are modelled to depend on their future values. From an empirical point of view, non-causal VAR models are worth investigating as they are simple linear models that are able to generate different dynamics than their causal counterparts. For instance, Gouriéroux and Zakoïan (2016) show that non-causal models are able to mimic speculative bubbles, i.e. processes that experience (locally) a rapid increase or decrease followed by a sudden crash or recovery. A more general model that depends on both past and future values is called a mixed causal–non-causal model and combines the specific characteristics of both the backward- and forward-looking VAR. It is able to generate even richer dynamics and therefore captures characteristics of macroeconomic and financial variables that could previously only be generated using highly nonlinear and complex models (see Hecq, Lieb and Telg, 2016; Gouriéroux and Jasiak, 2017).

Serial correlation common features (SCCF hereafter, see Engle and Kozicki, 1993) is a well-known approach to test for co-movements in an $n$-dimensional stationary time series process $Y_t$ generated by a causal VAR model of order $p$:

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

where $L$ is the lag operator, and $\epsilon_t$ are i.i.d. innovations with mean vector $\mathbb{E}(\epsilon_t) = 0$ and positive definite covariance matrix $\mathbb{E}(\epsilon_t \epsilon_t') = \Omega$. The presence of SCCF implies that each VAR($p$) coefficient matrix can be written as the product of two reduced rank matrices such that

$$(I_n - \delta_1 \beta_1' L - \ldots - \delta_p \beta_p' L^p)Y_t = \epsilon_t,$$

More precisely, a non-causal model is defined as a model that has a unique stationary solution in terms of current and future error terms. For the mixed causal–non-causal model, this is the two-sided moving average representation, i.e. past, current and future disturbances.

© 2018 The Authors. *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd.
where the full rank \( n \times (n - k) \) matrix \( (0 < k < n) \) \( \delta_\perp \) is the orthogonal complement matrix of \( \delta \) such that \( \delta' \delta_\perp = 0 \). It is clear that there exist \( k \) linear combinations \( \delta' Y_t = \delta' \varepsilon_t \) that annihilate the entire dynamics. A reduced rank structure in the matrix \([\Pi_1, \ldots, \Pi_p]\) is typically employed to detect the commonalities among series \( Y_t \). This paper extends this methodology to the case of mixed causal-non-causal VAR models.

Various papers successfully identify the presence of co-movements in the cyclical fluctuations of economic time series (see e.g. Centoni and Cubadda, 2015 and references therein). Macroeconomic variables like the growth rate of consumption, investment and the gross domestic product tend to co-move (Hecq, Palm and Urbain, 2006); economic indicators (industrial production index, unemployment or inflation) for several countries or regions within a country display some co-movements when a certain degree of convergence is reached. In contrast, there are also cases with well-motivated and graphically visible commonalities that are not detected by formal test statistics. We consider such a case in our empirical application examining monthly growth rates of two oil price series, whose graphs seem to move together almost identically. However, usual test statistics, that we review later, strongly reject the presence of co-movements. This leads to the conclusion that an exogenous shock adjusts differently on the two oil price indicators, which is highly unlikely from an economic point of view. One explanation for the lack of detecting co-movements is the fact that cycles do not have to be synchronous. Several less restrictive statistical models have been introduced to account for adjustment delays. Additional reasons have been suggested in the literature, among which the impact of misspecifications: presence of conditional heteroscedasticity (Candelon, Hecq and Verschoor, 2005), impact of seasonal adjustment (Hecq, 1998; Cubadda, 1999), outlying observations and nonlinear phenomena.

In this paper, we investigate whether co-movements in time series can be identified when they are modelled by means of a mixed causal-non-causal VAR rather than the conventional purely causal VAR. We show that for purely causal or non-causal 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 non-causal 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 VAR models 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 lead different from zero. However, this important result does not extend to mixed causal–non-causal models. In these cases, estimation methods that do not solely depend on second-order properties have to be employed to detect the presence of common features.

The rest of the paper is organized as follows. Section II summarizes results about the mixed causal–non-causal autoregressive model with common features. Section III introduces our testing strategy for identifying purely causal and non-causal models with SCCF. The finite sample behaviour of reduced rank regressions and GMM orthogonality tests when non-causality is present is studied in Section IV. Section V applies our new
modelling procedure to investigate the presence of commonalities in Brent and WTI price series. Section VI summarizes and concludes. Additional simulation results are collected in an online Appendix.

II. Causal and non-causal models

Mixed causal–non-causal autoregressive models have recently become increasingly popular because of their appealing properties. At a theoretical level, they offer the possibility to rewrite a process with explosive roots in direct time into a process in reverse time with roots outside the unit circle. In the applied econometric literature, there is a growing interest for mixed causal–non-causal models as they (i) might improve forecast performances, (ii) can be solutions to rational expectations models which take forward-looking behaviour into account and (iii) are able to model in a simple way some nonlinear features observed in macroeconomic and financial data (see e.g. Lanne and Saikkonen, 2011; Lanne, Luoto and Saikkonen, 2012; Hencic and Gouriéroux, 2014; Gouriéroux and Zakoïan, 2016; Hecq et al., 2016). In this paper, we focus on the multivariate aspects.

Multivariate models

Consider the mixed vector autoregressive model of order $(r, s)$, denoted as MVAR$(r, s)$. For an $n$−vector time series $Y_t$ the MVAR$(r, s)$ is given by

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

where $\varepsilon_t$ is a zero mean i.i.d. innovation process with positive definite covariance matrix $\mathbb{E}(\varepsilon_t \varepsilon_t') = \Omega$ and $\Pi(L)$ and $\Phi(L^{-1})$ are $n \times n$ polynomial matrices of order $r$ and $s$ respectively. The roots of both $\det[\Pi(z)]$ and $\det[\Phi(z)]$ are located strictly outside the unit circle. It follows that the series $Y_t$ have a strictly stationary solution as a two-sided MA representation

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

where $\Psi(L, L^{-1}) = [\Pi(L)\Phi(L^{-1})]^{-1}$, with $\Psi_j$ converging to zero at a geometric rate as $j \to \pm \infty$. This model appears to be a natural extension of the univariate case. However, differently from the univariate case, representation (1) is not unique since the product of matrices is generally not commutative. That is, the two polynomials matrices in (1) are ordered such that we first have the causal and then the non-causal component. However, one may find an alternative model with this order swapped around, which might represent the processes $Y_t$ as well. This gives rise to a model

$$
\overline{\Phi}(L^{-1})\overline{\Pi}(L)Y_t \equiv (I_n - \overline{\Phi}_1 L^{-1} - \ldots - \overline{\Phi}_s L^{-s})(I_n - \overline{\Pi}_1 L - \ldots - \overline{\Pi}_r L^r)Y_t = \varepsilon_t,
$$

where again both $\det[\overline{\Phi}(z)]$ and $\det[\overline{\Pi}(z)]$ have roots outside the unit circle. Lanne and Saikkonen (2013) provide reasons to consider the polynomial matrices in the order $\Pi(z)\Phi(z)$ based on economic intuition. However, for the sake of completeness, we do not rule out the alternative representation (3). Note that Davis and Song (2012) propose a different way to characterize the MVAR, which does not involve the multiplicative structure of lag and
lead polynomial coefficient matrices from the outset. Even though Gouriéroux and Jasiak (2017) prove that this alternative representation is more general than the multiplicative one, we adhere to the representations in (1) and (3) as they offer a more explicit and natural way to disentangle the ‘causal’ and ‘non-causal’ parts of the process.

As shown by Lanne and Saikkonen (2013), series \( Y_t \) in (1) also admit the following pseudo-causal \( \text{VAR}(p) \) representation

\[
Y_t = \sum_{j=1}^{p} \Lambda_j Y_{t-j} + \xi_t, \tag{4}
\]

where errors \( \xi_t \) are a multivariate white noise but, in general, not \( i.i.d. \) innovations with respect to the past. Representations (1), (3) and (4) have the same spectral density matrix and hence they cannot be distinguished by means of the Gaussian likelihood framework. Similar to the univariate case, it is therefore common to first estimate \( \text{VAR}(p) \) models of the form (4) in order to determine the total autoregressive order \( p \). The \( p \) chosen in this pseudo-causal model is then fixed and all \( \text{MVAR}(r,s) \) models with \( p = r + s \) are estimated in a second step. The model yielding the highest log-likelihood at the estimated parameters constitutes the final model. Consequently, the choice of \( p \) is crucial as it has an impact on the potential models determined in the second step. See Fries and Zakoïan (2017) for an alternative approach to perform model selection using Portmanteau-type tests and extremes clustering.

**Non-causal common cyclical features: definition**

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

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

The definition of common features in the \( \text{MVAR} \) is thus a straightforward extension of the conventional \( \text{VAR} \) case. Our goal is to determine whether such combinations exist and to estimate their number, namely the column rank of \( \delta \). We would like to remind the reader of the distinction made by Vahid and Engle (1993) between the notion common cyclical features that refers to a rank reduction in the dynamics of the \( \text{VAR} \) or the vector error-correction model (VECM) and the expression common cycles. The latter refers to the implication of common cyclical features in the Beveridge–Nelson decomposition of the level of the series into common trends and common cycles (which is not considered here).

**Remark 2.** By premultiplying both sides of equation (1) by \( \delta' \) and in view of the definition of CFs we have that

\[
\delta' \Psi(L, L^{-1}) = \delta', \tag{5}
\]
which indicates that the parameters in the moving average representation of the series $Y_t$ are collinear at any lag/lead different from zero. It is straightforward to check that the reverse implication applies as well.

As discussed in the previous section, it should be noted that the chosen MV AR specification does not need to be unique due to the multiplicative structure of the autoregressive polynomial matrices. The fact that matrices generally do not commute implies the existence of a different model with the roles of the backward- and forward-looking component reversed. For completeness, Definition 1 takes this alternative model into account. We provide below a simple example of how the same full rank matrix $\delta$ can eliminate the common features in both specifications.

**Example 3.** Consider the following specification for the causal and non-causal polynomials:

\[
\Pi(L) = I_2 - \begin{bmatrix} 0.6 & 0.5 \\ 1.2 & 1 \end{bmatrix} L + \begin{bmatrix} 0.3 & 0.25 \\ 0.6 & 0.5 \end{bmatrix} L^2,
\]

\[
\Phi(L^{-1}) = I_2 - \begin{bmatrix} 0.45 & 0.375 \\ 0.9 & 0.75 \end{bmatrix} L^{-1} + \begin{bmatrix} 0.15 & 0.125 \\ 0.3 & 0.25 \end{bmatrix} L^{-2},
\]

such that both $\det(\Pi(z))$ and $\det(\Phi(z))$ have roots outside the unit circle and $\delta \Pi(L) = \delta' \Phi(L^{-1}) = \delta'$ for $\delta' = [1, -0.5]$. Now, the product $\Pi(L)\Phi(L^{-1})$ gives

\[
\begin{bmatrix} 0.3 & 0.25 \\ 0.6 & 0.5 \end{bmatrix} L^2 - \begin{bmatrix} 0.96 & 0.8 \\ 1.92 & 1.6 \end{bmatrix} L + \begin{bmatrix} 1.84 & 0.7 \\ 1.68 & 2.4 \end{bmatrix} - \begin{bmatrix} 0.69 & 0.575 \\ 1.38 & 1.15 \end{bmatrix} L^{-1}
\]

\[
+ \begin{bmatrix} 0.15 & 0.125 \\ 0.3 & 0.25 \end{bmatrix} L^{-2},
\]

which is the same as $\bar{\Phi}(L^{-1})\bar{\Pi}(L)$ if we define $\bar{\Pi}(L) = \Pi(L)$ and $\bar{\Phi}(L^{-1}) = \Phi(L^{-1})$. In both models, $\delta'$ annihilates the dynamics. Note that, even in cases when the product $\Pi(L)\Phi(L^{-1})$ is not commutative, it is generally still possible to find $\bar{\Phi}(L^{-1})$ and $\bar{\Pi}(L)$ such that their product $\bar{\Phi}(L^{-1})\bar{\Pi}(L)$ is the same.

**Co-movements in purely causal or non-causal 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 non-causal MVAR($0, s$) model. These are interesting cases as we have to look at linear combinations of series that are orthogonal to respectively the past or the future of the variables. We show that usual common cyclical feature test statistics (e.g. canonical correlations) can be applied to non-causal models as well. We prove that the presence of commonalities in non-causal dynamics is not preserved when reversing time (i.e. into a causal framework), except for the case of the MVAR(0,1) model. Consequently, detecting a reduced rank structure in one direction directly identifies the dynamics of the whole underlying system even under Gaussianity. Finally, we show that the mixed model necessitates a non-Gaussian likelihood framework even when a reduced rank structure is present. Accordingly, our findings for pure models cannot be extended to mixed models.
In a sense, this makes the applicability of the results somewhat restrictive. However, to the best of our knowledge, the existing literature has solely focused on conventional (causal) VAR models, which makes the identification of causal from non-causal co-movements in purely causal or non-causal VAR models (in a straightforward manner) important for applied research.

Definition 4. In purely causal [non-causal] models with \( \Phi(L^{-1}) = I_n \) [\( \Pi(L) = I_n \)], series \( Y_t \) have \( k \) causal [non-causal] common features, CCFs [NCCFs] henceforth, if there exists an \( n \times k \) \( (0 < k < n) \) full rank matrix \( \delta_c [\delta_{nc}] \) such that \( \delta_c \Pi(L) = 0 \) \( [\delta_{nc} \Phi(L^{-1}) = 0] \). The common cyclical feature model in the purely causal [non-causal] case implies \( \mathbb{E}(\delta'_c Y_t | Y_{t-1}, Y_{t-2}, \ldots) = 0 \) \( [\mathbb{E}(\delta'_{nc} Y_t | Y_{t+1}, Y_{t+2}, \ldots) = 0] \).

Remark 5. The usual SCCF in a VAR(\( r \)) is obviously equivalent to the CCF in an MVAR(\( r, 0 \)).

We start the analysis by concentrating on the cases in which data are generated by a purely non-causal MVAR(\( 0, s \)) model, i.e. \( s > 0 \). Assuming that the matrix \( \delta_{nc} \) exists (we skip the indexes \( nc \) and \( c \) to simplify notations when no confusion is possible). The non-causal VAR model then reads as

\[
Y_t = \sum_{j=1}^{s} \Phi_j Y_{t+j} + \varepsilon_t = \delta_\perp \sum_{j=1}^{s} A_j Y_{t+j} + \varepsilon_t, \tag{6}
\]

where \( \delta_\perp \) is an \( n \times (n-k) \) full rank matrix such that \( \delta'_c \delta = 0 \) and at least one of the \( n \times (n-k) \) matrices \( A_j \), for \( j = 1, \ldots, s \), has full rank. We are now in the position to state and prove the main result of this section.

Proposition 6. If series \( Y_t \) are generated by a stationary MVAR(\( 0, s \)) [MVAR(\( r, 0 \))] with \( s > 1 \) [\( r > 1 \)], the existence of \( k \) common feature relationships does not imply the presence of any CF in the pseudo-causal [pseudo-non-causal] representation of series \( Y_t \).

Proof. We start by defining autocovariances \( \Gamma_y(-j) = \mathbb{E}(Y_t Y_{t+j}') \) and noting that

\[
\delta' \Gamma_y(-j) = \delta' \mathbb{E}(Y_t Y_{t+j}') = \delta' \mathbb{E}(\varepsilon_t Y_{t+j}') = 0, \quad \forall j > 0,
\]

Hence, it follows that we can factorize these autocovariances as the product

\[
\Gamma_y(-j) = \delta_\perp Y_y'(-j), \quad \forall j > 0,
\]

where \( Y_y(-j) \) is an \( (n-k) \times n \) matrix. We now rewrite the non-causal VAR in (6) under reduced rank restrictions as

\[
Y_t = \Phi' X_t + \varepsilon_t = \delta_\perp A' X_t + \varepsilon_t, \tag{7}
\]

where \( X_t' = [Y_{t+1}', \ldots, Y_{t+s}'] \), \( \Phi' = [\Phi_1', \ldots, \Phi_s'] \) and \( A' = [A_1', \ldots, A_s'] \). The coefficient matrix \( \Phi \) is linked to the autocorrelation matrix function \( \Gamma_y(-j) \) through the relation

\[
\Phi' = \mathbb{E}(Y_t X'_t)[\mathbb{E}(X_t X'_t)]^{-1} = [\Gamma_y(-1), \ldots, \Gamma_y(-s)][\Gamma_x(0)]^{-1} = \delta_\perp [Y_y'(-1), \ldots, Y_y'(-s)][\Gamma_x(0)]^{-1}, \tag{8}
\]
where

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

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

$$
Y_t = \sum_{j=1}^s \Lambda_j Y_{t-j} + \xi_t = \Lambda' Z_t + \xi_t,
$$

where $Z_t' = [Y_{t-1}', \ldots, Y_{t-s}']$. Since $E(\xi_t Y_{t-j}) = 0$ for any $j > 0$, the coefficient matrix $\Lambda' = [\Lambda_1, \ldots, \Lambda_s]$ of the pseudo-causal VAR is linked to the autocorrelation matrix function of series $Y_t$ through the relation

$$
\Lambda' = E(Y_t Z_t')E(Z_t Z_t')^{-1} = [\Gamma_\gamma(1), \ldots, \Gamma_\gamma(s)]E(\gamma_x(0))^{-1}
$$

$$
= [\Gamma_\gamma(-1)', \ldots, \Gamma_\gamma(-s')]E(\gamma_x(0))^{-1} = [\gamma_x(1)\delta'\perp, \ldots, \gamma_x(s)\delta'\perp]E(\gamma_x(0))^{-1},
$$

where

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

which implies that matrices in $\Lambda'$ do not have a common left null space and thus that $\delta'\Lambda' \neq 0$. In view of (10), we conclude that the pseudo-causal VAR in (9) generally does not have any (pseudo) CFs. It is straightforward to see that the reverse is also true, i.e. the non-causal representation of a causal VAR with CFs generally does not exhibit (pseudo) CFs. However, a special case occurs when $p = 1$. Indeed, equations (8) and (10) simplify to

$$
\Phi' = \gamma_x(-1)[\gamma_x(0)]^{-1} = \delta' \gamma_x(-1)[\gamma_x(0)]^{-1},
$$

$$
\Lambda' = \gamma_x(1)[\gamma_x(0)]^{-1} = \gamma_x(1)\delta' \gamma_x(0)]^{-1},
$$

from which we see that $\delta'\Phi' = 0$ leads to $Y_x'(1)\Lambda' = 0$. Hence, the existence of $k$ CFs in an MVAR(0, 1) implies the presence of $k$ CFs in the pseudo-causal representation of series $Y_t$ although the CF vectors will be different in the two representations. It is easy to check that the reverse is true as well.

**Remark 7.** With regard to statistical inference, a relevant implication of Proposition 6 is that, except for the case of VAR models with either one lead or one lag, the Gaussian ML approach is able to identify the unique representation of series $Y_t$ that exhibit the common features. It is well-known that the Gaussian likelihood analysis of the reduced rank regression model (7) is based on canonical correlations between $Y_t$ and $X_t$, denoted as CanCor$(Y_t, X_t)$; see e.g. Johansen (2008) and the references therein.

**Remark 8.** In case of the MVAR(1, 0) and the MVAR(0, 1) models, it is impossible to discriminate between causal and non-causal 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 non-causal common feature vector, we will also detect a causal common feature relationship using canonical correlations and we cannot identify whether the system is an MV AR(0,1) or an MV AR(1,0). 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 cyclical feature relationships are sometimes obtained in empirical works.

**Example 9.** We illustrate by a numerical example the relationship between the matrices $\Phi$ for the non-causal VAR and $\Lambda$ for the causal VAR in a bivariate MV AR(0,2) with a reduced rank structure given by

$$\Phi' = [\Phi_1, \Phi_2]$$

$$= \begin{bmatrix}
0.5 & 0.2 \\
0.5 & 0.2
\end{bmatrix},
\begin{bmatrix}
-0.35 & 0.25 \\
-0.35 & 0.25
\end{bmatrix}$$

$$= \begin{bmatrix}
1 & [0.5 & 0.2] \\
1 & [-0.35 & 0.25]
\end{bmatrix}$$

with $\delta' = [1, -1]$ such that $\delta' \Phi = 0$. These are the coefficient matrices used in our simulation study. Denote

$$\mathbb{E}(\varepsilon_t \varepsilon_t') = \Omega = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix}, \tilde{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 \\ I_n & 0_n \end{bmatrix}, \tilde{\Omega} = \begin{bmatrix} \Omega & 0_n \\ 0_n & 0_n \end{bmatrix}.$$

Basic algebra can be used to link the autocovariances with matrices $\Phi_1$ and $\Phi_2$ (see Tsay, 2014, p. 38-39):

$$[I_{(2n)^2} - \tilde{\Phi} \otimes \tilde{\Phi}'] \text{vec}(\Gamma_z(0)) = \text{vec}(\tilde{\Omega}),$$

from which we get $\Gamma_z(0)$ and $\Gamma_z(-1)$ after extracting $\Gamma_z(0)$ by inversion. With the numerical values we consider, this leads to (rounded at 4 digits):

$$\Gamma_z(0) = \begin{bmatrix} 13.5902 & 6.5902 \\ 6.5902 & 5.5902 \end{bmatrix},$$

$$\Gamma_z(-1) = \begin{bmatrix} 6.5347 & 2.8347 \\ 6.5347 & 2.8347 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 6.5347 & 2.8347 \end{bmatrix},$$

$$\Gamma_z(-2) = \begin{bmatrix} 1.4653 & 1.0753 \\ 1.4653 & 1.0753 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1.4653 & 1.0753 \end{bmatrix},$$

for the MV AR(0,2). We have obtained $\Gamma_z(-2)$ using $[\Gamma_z(-1), \Gamma_z(-2)] = \Phi' \Gamma_z(0)$ (Tsay, 2014, p. 38). Those values imply causal VAR coefficient matrices

$$\Lambda' = [Y_z(1)\delta_z', \ldots, Y_z(s)\delta_z'] [\Gamma_z(0)]^{-1}.$$
In particular, we see that $\Lambda'$ is full-rank, and hence cannot be annihilated by $\delta'$, as

$$
\Lambda' = [\Lambda_1, \Lambda_2] = \begin{bmatrix}
6.5347 & 6.5347 \\
2.8347 & 2.8347
\end{bmatrix}
\begin{bmatrix}
1.4653 & 1.4653 \\
1.0753 & 1.0753
\end{bmatrix} [\Gamma_z(0)]^{-1}.
$$

**Co-movements in mixed models**

Now consider the general MVAR($r$, $s$) case where both $r$ and $s$ are larger than zero. Also in this case, series $Y_t$ admit a pseudo-causal VAR($p$) representation (4), where $p = r + s$ (Lanne and Saikkonen, 2013). Our main result on the general case is given by the following proposition.

**Proposition 10.** If series $Y_t$ are generated by a 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-causal [pseudo-non-causal] representation of series $Y_t$.

**Proof.** In view of equation (2), we see that when both $r$ and $s$ are positive, the auto-correlation matrix function of series $Y_t$ reads

$$
\Gamma_y(j) = \sum_{i=-\infty}^{\infty} \Psi_{i+j} \Omega \Psi_i' = \Psi_0 \Omega \Psi_{-j}' + \sum_{\forall i \neq -j} \Psi_i \Omega \Psi_{i-j}',
$$

which, in combination with equation (5), implies that

$$
\delta' \Gamma_y(j) = \delta' \Omega \Psi_{-j}'.
$$

Hence, the coefficient matrix of the corresponding pseudo-causal VAR, i.e.

$$
\Lambda' = [\Lambda_1, \Lambda_2, \ldots, \Lambda_p] = [\Gamma_y(1), \ldots, \Gamma_y(p)][\Gamma_z(0)]^{-1},
$$

will generally be such that $\delta' \Lambda' \neq 0$. It is straightforward to see that the same conclusion applies to the pseudo-non-causal representation as well.

**Remark 11.** Notice that a simple solution to detecting CFs in mixed models such as 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

$$
Y_t = \delta C' U_t + \Theta_0^{-1} \epsilon_t,
$$

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.

To sum up, we have shown that the presence of a reduced rank structure allows to identify purely causal and non-causal systems even under Gaussianity (for lag/lead order higher than one), since the true data generating process has either a backward- or forward-looking MA representation. This implies that causal processes are independent of future errors, while non-causal processes are independent of past errors. As the common features are present either in the causal or non-causal part, one can always find suitable instruments (i.e. either
the lags or the leads) to draw inference based on the first two moments. In a mixed causal–non-causal model, the processes \( Y_t \) possess a two-sided moving average representation and thus such instruments cannot be found. Whereas the detection of commonalities in mixed causal–non-causal models is clearly an interesting topic, it is outside the scope of this paper. The reason for this is that there is a price to pay in case the researcher wants to go ahead with the mixed model. Under the alternative, one would need to identify a pseudo-causal VAR\((p)\) and then find the MVAR\((r,s)\) that maximizes the log-likelihood of a particular non-Gaussian distribution. Imposing the reduced rank restriction to perform either a Likelihood Ratio (LR) test or a Wald test, requires the derivation of their asymptotic distributions and an evaluation of their finite sample properties.

III. Common feature test statistics

In order to test for common cyclical features in either purely causal or non-causal models, we first determine the lag order \( p \) in the pseudo-causal VAR model by using information criteria. Even in the presence of a non-causal component, we obtain a consistent estimate of the pseudo-causal order. In Section IV, we investigate the performances of LR and GMM tests for the presence of co-movements both in purely causal and non-causal models.\(^3\) We propose to perform all tests twice: once with lags and once with leads as instruments. Since for \( P > 1 \), a reduced rank can only be present in either the causal or the non-causal dynamics, the proposed tests for detecting common features also offer a useful way to identify whether the model is causal or non-causal.

LR tests

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

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

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

\[
\text{CanCor} \{Y_t, W_t | D_t\}. 
\]

The LR test, which has the null hypothesis that there exist at least \( k \) common feature vectors, is based on the statistic

\[
\tilde{\xi}_{LR} = -(T - p) \sum_{i=1}^{k} \ln(1 - \hat{\lambda}_i), \quad k = 1, 2, \ldots, n,
\]

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

\(^3\) An alternative testing procedure to detect common features, even in high-dimensional systems, is based on partial least squares (Cubadda and Hecq, 2011). However, as a first step, we focus on classical testing procedures only in this section.
We split the time series $Y_t$ into two subsets as $Y_t = [y_{1t}, y_{(2:n)t}]$ in which $y_{1t}$ denotes the dependent variable, while $y_{(2:n)t} = [y_{2t}, y_{3t}, \ldots, y_{nt}]$ collects the $n-1$ series out of the $n$ variables. A GMM test for a single common feature can now be conducted. First, we write $\delta = [1, -x']'$ and estimate the $(n-1) \times 1$ vector $x$ by the instrumental variable (IV) estimator using $W_t$ as instruments. The IV estimator of $x$ is given by

$$
\hat{x}_{IV} = (y'_{(2:n)} W (W' W)^{-1} W' y_{(2:n)})^{-1} (y'_{(2:n)} W (W' W)^{-1} W' y_1),
$$

where $[y_1, y_{(2:n)}] = Y$ and $W$ indicate the matrices of $T - p$ observations of variables $Y_t$ and $W_t$, respectively, after the linear influence of the deterministic terms $D_t$ has been removed. 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

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

where $u = y_1 - y_{(2:n)} \hat{x}_{IV}$ and $\hat{\sigma}_u^2 = u' u / (T - p)$.

The tests presented so far embed the assumption of homoskedasticity. In a Monte Carlo study, Candelon et al. (2005) illustrate that $J_1(\tilde{z}_{IV})$ has large size distortions in the presence of GARCH disturbances. Therefore, we use the White heteroscedasticity robust estimator (see, e.g. Hamilton, 1994):

$$
\hat{x}_{IV-W} = (y'_{(2:n)} W (W' B W)^{-1} W' y_{(2:n)})^{-1} (y'_{(2:n)} W (W' B W)^{-1} W' y_1),
$$

where $B = \text{diag}(u_1^2, u_2^2, \ldots, u_T^2)$ and $u_t = y_{1t} - y_{(2:n)t} \hat{x}_{IV}$. Then we run a $J$-test robust to heteroscedasticity using the statistic

$$
J_2(\tilde{z}_{IV-W}) = (u^* W) (W' B W)^{-1} (W' u^*),
$$

where $u^* = y_1 - y_{(2:n)} \hat{x}_{IV-W}$.\footnote{Alternative approaches for handling heteroscedastic disturbances have been evaluated in Hecq and Issler (2012).} Under the null hypothesis, both $J_1(\tilde{z}_{IV})$ and $J_2(\tilde{z}_{IV-W})$ follow asymptotically a $\chi^2(v)$ distribution with $v = knp - k(n - k)$ degrees of freedom. For example, this reduces to a $\chi^2(2p - 1)$ in the bivariate case with $k = 1$. 

TABLE 1

| DGP/instruments | \(\{Y_{t-1}, Y_{t-2}\}\) | \(\{Y_{t+1}, Y_{t+2}\}\) |
|----------------|-----------------|-----------------|
| Causal VAR(2)  | CanCor           | GMM             | GMM\(^{W}\) | CanCor | GMM | GMM\(^{W}\) |
| \(T = 100\)    | 5.64             | 4.91            | 4.63         | 23.52  | 29.45 | 26.64         |
|                | 5.37             | 4.82            | 4.58         | 23.47  | 29.52 | 25.40         |
|                | 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         |
| \(T = 500\)    | 5.16             | 5.08            | 4.56         | 91.84  | 92.54 | 91.99         |
|                | 4.98             | 4.84            | 4.78         | 91.90  | 92.54 | 91.41         |
|                | 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         |

IV. Monte Carlo results

In this section, we investigate by simulations the empirical size and power of the test statistics \(\xi_{LR}, J_1(\hat{\pi}_IV)\) and \(J_2(\hat{\pi}_IV - W)\) in finite samples. We restrict our attention to the either purely causal or purely non-causal VAR. More precisely, we consider the bivariate causal [non-causal] MVAR(2,0) [MVAR(0,2)] with a reduced rank structure, i.e.

\[
\Pi_1[\text{resp. } \Phi_1] = \begin{bmatrix} 0.5 & 0.2 \\ 0.5 & 0.2 \end{bmatrix}, \quad \Pi_2[\text{resp. } \Phi_2] = \begin{bmatrix} -0.35 & 0.25 \\ -0.35 & 0.25 \end{bmatrix}.
\]

We consider two relatively small sample sizes, \(T = \{100, 250\}\).\(^5\) Define the variance-covariance matrix

\[
\Omega = \begin{bmatrix} 10 & 3 \\ 3 & 2 \end{bmatrix},
\]

and let \(\zeta_t \sim N(0, \Omega)\). For disturbances \(\varepsilon_t\), we consider various correlated distributions:

(i) Normal \(N(0, \Omega)\), (ii) Student’s \(t(v) = \zeta_t \odot (\varepsilon/\chi^2(v))^{1/2}\) and (iii) Laplace \(= \zeta_t \odot (-\ln (U[0, 1]))^{1/2}\), where \(\odot\) denotes the Hadamard product. For (ii) we consider the following values for the degrees of freedom \(v = \{10, 3, 1\}\), where \(v = 1\) corresponds to the Standard Cauchy case. Each distribution has its own seed and all reported results are based on 10,000 replications.

Under the null hypothesis, the causal [non-causal] VAR has a CCF [NCCF] relationship with a cofeature vector \(\delta' = [1, -1]\). We remark that, according to the theoretical results in Section II, non-rejection should be interpreted as evidence that the process is causal [non-causal] and that a CCF [NCCF] exist. Consequently, rejection of both tests for the existence of a CCF and an NCCF indicates that there is neither the former nor the latter.

Tables 1 and 2 report the frequencies with which each common feature test rejects the null hypothesis at a 5% significance level. Both tests have asymptotically a \(\chi^2(3)\)

\(^5\)We simulate an additional \(50 \times 2\) observations (both sides) burn in period to initialize processes.
distribution under the null hypothesis. An entry in the vicinity of 5% denotes an absence of size distortion. Regardless of the distribution chosen to simulate processes, neither LR nor GMM tests suffer from severe size distortions. To obtain this result the ‘correct’ 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 [non-causal] model. This means that purely non-causal common features can easily be found by reverting the timeline of the instruments. The only test that faces a problem is the robust GMM when 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 squared residuals cause problems in the Cauchy case when large ‘outlying’ observations are introduced via the \(B = \text{diag}(u_2^2, u_2^2, \ldots, u_T^2)\) matrix. When the wrong set of instruments is used, namely only lags for the non-causal model or leads for the causal specification, the presence of a reduced rank structure is rejected. The rejection frequencies increase with the sample size, indicating that we are indeed under the alternative hypothesis.

According to these findings, a recommendation for empirical works is to first determine the lag order \(p\) and then to apply the common cyclical feature test statistics in both directions. It has to be mentioned that in the proposed two-step procedure, the detection of CFs in either the causal or non-causal dynamics is conditional on the correct determination of \(p\). This step is important and should be performed with caution, as an overidentification of \(p\) decreases the power of the tests, while choosing \(p\) too low might lead to a failure of detecting common features at all. In an online appendix to this paper, we present additional simulation results on the determination of the order \(p\), the performances of the three proposed test statistics when the DGP is of a mixed form, and some misspecification tests for the null hypothesis that the errors are i.i.d. if a pseudo-causal model is estimated when data are generated by a non-causal model.

Note that we solely mean lags for the causal model and leads for the non-causal model. We do not comment on under- or overspecification of these lag and lead orders.
V. Co-movements in oil prices

Now, we illustrate how to use the tools that we propose in this paper in practice. We study the presence of short-run co-movements between two oil price series, that is the price of a barrel of crude oil for Brent in Europe ($Brent_t$) and West Texas Intermediate at delivery in Cushing, Oklahoma ($WTI_t$). Our sample runs from May 1987 to September 2017, which accounts for $T = 365$ monthly observations. The data is available at the U.S. Energy Information Administration. We first consider the monthly growth rates of the series as plotted in the second panel of Figure 1 and we define the bivariate vector

$$\Delta \ln Y_t = [\Delta \ln Brent_t, \Delta \ln WTI_t]' .$$

Using information criteria to determine the VAR order for $\Delta \ln Y_t$, we obtain $p = 1$ with BIC, $p = 2$ with HQ and $p = 4$ for AIC. Table 3 reports the $P$-values of three misspecification tests for the three different VAR models considered. More specifically, we use a multivariate LM test for the null of no autocorrelation from lag 1 to 2, the multivariate version of the Jarque–Bera test as well as the multivariate heteroscedasticity test (no cross terms). Both the information criteria and residual tests statistics can be obtained with several statistical packages. In this study, we use EViews 10.

From the LM test, it emerges that the VAR(1) and to a lesser extent the VAR(2) are misspecified as we reject the null of no autocorrelation at a 5% significance level. In contrast, we fail to reject the null of no autocorrelation at 5% for the VAR(4). From the Jarque–Bera test, we conclude that we reject the null of normally distributed residuals for all three models considered. This provides an additional motivation to distinguish between backward- and forward-looking behaviour. Given that the presence of heteroscedasticity can have an impact on the tests of autocorrelation, we continue the investigation with all three models. In particular, we keep the VAR(1) in the set of considered models, as it helps to illustrate that while possibly finding a reduced rank structure, it is not possible to discriminate between causal and non-causal models (see Remark 7).

The last four columns of Table 3 report for both causal and non-causal models the $P$-values of the LR canonical correlation and the robust GMM test statistics. Concerning the VAR(1), i.e. the row associated to the BIC criterion, neither $\xi_{LR}$ nor $J_2(\hat{\theta}_{IV - W})$ rejects the null of a reduced rank. It is also seen that $P$-values for both the causal and the non-causal specification are identical by construction. The eigenvector from canonical correlations gives for the causal relationship $\delta_{c}' = [1, -0.928]$ and $\delta_{nc}' = [1, -1.118]$ for the non-causal counterpart. The GMM relationships are very similar and yield respectively $\delta_{c}' = [1, -0.930]$ and $\delta_{nc}' = [1, -1.102]$. In this case, it is not obvious from those vector coefficients to favor the causal or the non-causal relationship.

For the VAR(2), both tests fail to reject the null of common cyclical features only for the non-causal model. If we consider the VAR(4), there is a common feature vector in the non-causal framework only if we rely on the GMM test. It emerges from simulations in Candelon et al. (2005) that the latter test is more robust to the presence of heteroscedasticity. Since the results in Table 3 suggest heteroscedastic residuals, it is expected that this test is more appropriate in this empirical application. As far as the relationships are concerned, we find very similar eigenvectors in all cases. That is, the VAR(2) yields $\delta_{nc}' = [1, -1.197]$

We do not report results for $J_1(\hat{\theta}_{IV})$ as its outcome is similar to that of $\xi_{LR}$.
for canonical correlations and $\delta'_{nc} = [1, -1.156]$ for GMM. For the VAR(4), the respective relationships are $\delta'_{nc} = [1, -1.167]$ (canonical correlations) and $\delta'_{nc} = [1, -1.118]$ (GMM). From these results we conclude that a purely non-causal VAR has generated the dynamics of $\Delta \ln Y_t$. Note that we identified this process using simple tools that neither require the imposition of a particular non-Gaussian distribution nor estimation of models by complex nonlinear iterative algorithms.

Finally, we consider the co-movements in the log levels of the series. We observe that the two prices display commonalities in the trend component as well as strong co-movements in the bubble feature, except for the period 2011–2014 where the Brent is systematically higher than the WTI. Several economic motivations are advocated to explain this deviation during that period, e.g. the extraction of more expensive oil sands in the United States until the installation of new pipelines for its transport. It is not obvious to determine the order of integration of this type of data. Are series I(1) as we have assumed\(^8\) or are they globally stationary with some temporary explosive roots generating the bubbles? Several authors (Cavaliere, Nielsen and Rahbek, 2018; Hencic and Gouriéroux, 2014; Fries and Zakoïan, 2017) detrend similar series with a polynomial trend function before working with the ‘stationary’ detrended series. While we see the interest of this step as it keeps the bubble pattern intact, this procedure is rather ad hoc. An alternative way to model the oil prices would involve the estimation of a step function with a kind of natural level in prices of roughly 20$ per barrel until 2000 and around 60–80$ per barrel from 2005 onwards. However, the story of a natural price is incorrect now retrospectively. When the barrel reached its peak in June 2008, several people predicted the price to be as high as 200$ the barrel before the end of 2000. This is not what happened. This being said, a polynomial trend of order 4 seems to best capture the trending pattern of our series in levels, a result that is not supported with the stationary features we observe in first differences.

If we assume that the series are I(1), we also fail to reject the presence of a cointegrating relationship within a VAR(3) in levels using Johansen’s approach. The trace test is 28.31 for the null of zero cointegrating vectors, the critical value at a 5% significance level being 15.5. The cointegrating vector is given by $\beta' = [1, -1.102]$. In order to check to what

\(^8\)\textit{P}-values of ADF(1) tests with an intercept and a time trend are respectively 0.190 and 0.180 for the log level of the Brent and the WTI.
TABLE 3

P-values for misspecification and common feature tests

| Criterion | VAR(p) | Misspecification tests | Causal model | Non-causal model |
|-----------|--------|------------------------|--------------|-----------------|
|           |        | LM[1-2] | JB | Hetero | $\xi_{LR}$ | $J_2(\tilde{z}_{IV-W})$ | $\tilde{\xi}_{LR}$ | $J_2(\tilde{z}_{IV-W})$ |
| BIC       | VAR(1) | 0.005 | 0.000 | 0.000 | 0.341 | 0.434 | 0.341 | 0.434 |
| HQ        | VAR(2) | 0.033 | 0.000 | 0.000 | 0.004 | 0.020 | 0.166 | 0.366 |
| AIC       | VAR(4) | 0.096 | 0.000 | 0.000 | 0.001 | 0.009 | 0.015 | 0.124 |

extent our common feature results are robust to the presence of cointegration, we look at the weak form common feature approach proposed by Hecq et al. (2006). That is, we look for a combination in the VECM with 2 lags in our specific case

$$\Delta \ln Y_t = \mu + \alpha \beta' \ln Y_{t-1} + \Theta_1 \Delta \ln Y_{t-1} + \Theta_2 \Delta \ln Y_{t-2} + \nu_t,$$

such that $\delta_c$ only annihilates the short run dynamics. That is, under the null, $\delta'_c(\Delta \ln Y_t - \mu - \alpha \beta' \ln Y_{t-1}) = \delta'_c \nu_t$, which implies that $\delta'_c \Theta_1 = \delta'_c \Theta_2 = 0$. Although the implication of this VECM in reverse time is outside the scope of this paper, we also look at co-movements in

$$\Delta \ln Y_t = \mu^* + \alpha^* \beta' \ln Y_{t+1} + \Theta_1^* \Delta \ln Y_{t+1} + \Theta_2^* \Delta \ln Y_{t+2} + \nu_t^*$$

such that $\delta_{nc}$ only annihilates the lead short run dynamics. Similarly, we have under the null hypothesis that $\delta_{nc}'(\Delta \ln Y_t - \mu^* - \alpha^* \beta' \ln Y_{t+1}) = \delta_{nc}' \nu_t^*$ which implies that $\delta_{nc}' \Theta_1^* = \delta_{nc}' \Theta_2^* = 0$. P-values are respectively 0.004 and 0.020 for the LR and the robust GMM test for the causal model; they are 0.173 and 0.368 for the non-causal specification. Common feature vectors are similar to the ones without cointegration. Hence, the results appear to be robust to the presence of a long run relationship.

VI. Conclusion

This paper provides a novel way to detect common features in a VAR framework with potentially a lead component. We show that common features that cannot be detected in the causal (i.e. backward-looking) representation might be revealed in the non-causal (i.e. forward-looking) dynamics of the series. However, whereas the difference between data generated from purely causal, purely non-causal and mixed causal—non-causal VAR models 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 test for the potential presence of non-causal co-movements. Using sets of either lags or leads as instruments within a canonical correlation or a GMM framework, we show how additional relationships are discovered between series both in Monte Carlo simulations and in oil price series. Interestingly, we find that the presence of a reduced rank structure in the dynamics of purely causal and non-causal systems with more than one lag/lead permits identification in the usual Gaussian framework. This result cannot be extended to mixed causal–non-causal models, as the instruments are correlated with the error term of the reduced rank model by construction.

© 2018 The Authors. Oxford Bulletin of Economics and Statistics published by Oxford University and John Wiley & Sons Ltd.
References

Breidt, F., Davis, R., Lii, K. and Rosenblatt, M. (1991). ‘Maximum likelihood estimation for noncausal autoregressive processes’, *Journal of Multivariate Analysis*, Vol. 36, pp. 175–198.

Candelon, B., Hecq, A. and Verschoor, W. (2005). ‘Measuring common cyclical features during financial turmoil’, *Journal of International Money and Finance*, Vol. 24, pp. 1317–1334.

Cavaliere, G., Nielsen, H. B. and Rahbek, A. (2018). ‘Bootstrapping noncausal autoregressions: with Applications to explosive bubble modeling’, *Journal of Business and Economic Statistics*, https://doi.org/10.1080/07350015.2018.1448830.

Centoni, M. and Cubadda, G. (2015). ‘Common feature analysis of economic time series: an overview and recent developments’, *Communications for Statistical Applications and Methods*, Vol. 22, pp. 415–434.

Cubadda, G. (1999). ‘Common cycles in seasonal non-stationary time series’, *Journal of Applied Econometrics*, Vol. 14, pp. 273–291.

Cubadda, G. (2007). ‘A unifying framework for analysing common cyclical features in cointegrated time series’, *Computational Statistics & Data Analysis*, Vol. 52, pp. 896–906.

Cavaliere, G. and Nielsen, H. B. and Rahbek, A. (2018). ‘Bootstrapping noncausal autoregressions: with Applications to explosive bubble modeling’, *Journal of Business and Economic Statistics*, https://doi.org/10.1080/07350015.2018.1448830.

Cubadda, G. and Hecq, A. (2001). ‘On non-contemporaneous short-run co-movements’, *Economics Letters*, Vol. 73, pp. 389–397.

Cubadda G. and Hecq, A. (2011). ‘Testing for common autocorrelation in data rich environments’, *Journal of Forecasting*, Vol. 30, pp. 325–335.

Davis, R. and Song, L. (2012). Noncausal Vector AR Processes with Application to Economic Time Series, *Discussion Paper Colombia University*.

Engle, R. F. and Kozicki, S. (1993). ‘Testing for common features (With Comments)’, *Journal of Business and Economic Statistics*, Vol. 11, pp. 369–395.

Fries, S. and Zakoïan, J.-M. (2017). Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles. MPRA Paper 81345, University Library of Munich, Germany.

Gouriéroux, C. and J. Jasiak (2017). ‘Noncausal vector autoregressive processes: representation, identification and semi-parametric estimation’, *Journal of Econometrics*, Vol. 200, pp. 118–134.

Gouriéroux, C. and Zakoïan, J. M. (2016). ‘Local explosion modelling by noncausal process’, *Journal of the Royal Statistical Society, Series B*, Vol. 79, pp. 737–756

Guillén, O., Hecq, A., Issler, J. and Saraiva, D. (2015). ‘Forecasting multivariate time series under present-value model short- and long-run co-movement restrictions’, *International Journal of Forecasting*, Vol. 31, pp. 862–875.

Hamilton, J. (1994). *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hansen, L.P. (1982). ‘Large sample properties of generalized method of moments estimators’, *Econometrica*, Vol. 50, pp. 1029–1054.

Hecq, A. (1998). ‘Does seasonal adjustment induce common cycles?’, *Economics Letters*, Vol. 59, pp. 289–297.

Hecq, A. and Issler, J. V. (2012). A Common-Feature Approach for Testing Present-Value Restrictions with Financial Data, METEOR RM/12/006, Maastricht University.

Hecq, A., Lieb, L. and Telg, S. (2016). ‘Identification of mixed causal-noncausal models in finite samples’, *Annals of Economics and Statistics*, Vol. 123-124, pp. 307–331.

Hecq, A., Palm, F. C. and Urbain, J. P. (2006). ‘Testing for common cyclical features in VAR models with cointegration’, *Journal of Forecasting*, Vol. 132, pp. 117–141.

Henric, A. and Gouriéroux, C. (2014). ‘Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rate’, in Huynh, V.-N., Kreinovich, V., Sriboonchitta, S. and Suriya, K.(eds.), *Econometrics of Risk, Series: Studies in Computational Intelligence*, Basel: Springer International Publishing, pp. 17–40.

Issler, J. and Vahid, F. (2001). ‘Common cycles and the importance of transitory shocks to macroeconomic aggregates’, *Journal of Monetary Economics*, Vol. 47, pp. 449–475.

Johansen, S. (2008). Reduced rank regression, in Durlauf, S. N. and L. E. Blume (eds.) *The New Palgrave Dictionary of Economics* (2nd Edition), Basingstoke: Palgrave Macmillan.

Lanne, M., Luoto J. and Saikkonen, P. (2012). ‘Optimal forecasting of noncausal autoregressive time series’, *International Journal of Forecasting*, Vol. 28, pp. 623–631.
Lanne, M. and Saikkonen, P. (2011). ‘Noncausal autoregressions for economic time series’, Journal of Time Series Econometrics, Vol. 3, pp. 1–32.
Lanne, M. and Saikkonen, P. (2013). ‘Noncausal vector autoregression’, Econometric Theory, Vol. 29, pp. 447–481.
Sims, C. (1980). ‘Macroeconomics and reality’, Econometrica, Vol. 48, pp. 1–48.
Tsay, R. (2014). Multivariate Time Series Analysis: With R and Financial Applications, Wiley, Hoboken, NJ.
Vahid, F. and Engle, R. (1993). ‘Common trends and common cycles’, Journal of Applied Econometrics, Vol. 8, pp. 341–360.
Vahid, F. and Engle, R. (1997). ‘Codependent cycles’, Journal of Econometrics, Vol. 80, pp. 199–221.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1. Frequencies with which ICs find the pseudo-causal VAR(3) for a MVAR(1,2) DGP.
Table S2. Testing for common features in mixed models.
Table S3. Empirical size/power for the causal vs. non-causal VAR model.