Generalized Covariance Estimator
Supplemental Material

Gourieroux*, C. and J. Jasiak†

August 23, 2022

1 Introduction

This supplemental material to "Generalized Covariance Estimator" is organized as follows. Section 2 contains theoretical Appendices 1, 2, 3 and 4 in Sections 2.1 up to 2.4. Appendix 1 (Section 2.1) illustrates the SUR interpretation of the VAR(1) model. It is used to derive the multivariate portmanteau statistic as a Lagrange multiplier test statistic. This interpretation allows us to find the asymptotic distribution of sample autocovariances. The closed-form asymptotic expansions underlying the proof of Proposition 3 are given in Appendices 2 and 3 (Sections 2.2 and 2.3). Appendix 4 (Section 2.4) discusses the possibility of reaching a parametric efficiency bound for a copula density through a sequence of state discretizations. Section 3 provides additional simulation results illustrating the finite sample distribution of the GCov estimator and the effect of lag H. Section 4 provides additional empirical results for the analysis of commodity prices. We provide summary statistics of the series and diagnostics tools. We also report the estimation results for the VAR(3) model. In Section 5, we review various constrained causal VAR models existing in the literature on multivariate analysis and explain why a GCov approach can be used as an alternative to the OLS.

*University of Toronto, Toulouse School of Economics and CREST,
e-mail: *Christian.Gourieroux@ENSAE.fr.*
†York University, e-mail: *jasiakj@yorku.ca.*

2 Appendices

2.1 APPENDIX 1: The SUR Interpretation

Let us consider a stationary process satisfying the causal VAR(1) representation:

$$Y_t = \alpha + BY_{t-1} + u_t,$$

where u_t is a multivariate i.i.d. process with $E(u_t) = 0$, $V(u_t) = \Sigma$, and Σ is invertible. We derive the asymptotic distribution of the OLS estimator of B under the null hypothesis $H_0 : B = 0$ and the closed-form expression of the Lagrange Multiplier (LM) test statistic.

i) GLS estimator

This is an example of a SUR model with identical regressors $X_t = Y_{t-1}$ in all equations. Then, the GLS estimator of B is obtained by applying the OLS equation by equation. We have:

$$\hat{B} = \hat{\Gamma}(1)\hat{\Gamma}(0)^{-1} \iff \hat{B}' = \hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)'.$$

Moreover, we have asymptotically

$$\sqrt{T}[vec(\hat{B}') - vecB'] \approx N[0, \Sigma \otimes \Gamma(0)^{-1}], \tag{a.1}$$

where the $\otimes$ denotes the Kronecker product [see e.g. Chitturi (1974), eq. (1.13)]. In particular, under the null hypothesis $H_0 : (\Gamma(1) = 0) \equiv (B = 0)$, we have $\Sigma = \Gamma(0)$ and

$$\sqrt{T}vec(B') \sim N(0, \Gamma(0) \otimes [\Gamma(0)^{-1}]). \tag{a.2}$$

ii) Lagrange Multiplier test

It follows that the Lagrange Multiplier (LM) test statistic for testing $H_0 : (B = 0) = (\Gamma(1) = 0)$ is:

$$\xi(1) = Tvec[\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)']'[\hat{\Gamma}(0)^{-1} \otimes \hat{\Gamma}(0)]vec[\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)']$$

where the asymptotic covariance matrix of $vec(\hat{B}')$ is estimated under the null hypothesis [see Hosking (1981a,b)]. It is also equal to:

$$
\begin{aligned}
\xi(1) &= T vec[\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)']'[\hat{\Gamma}(0)^{-1/2} \otimes \hat{\Gamma}(0)^{1/2}][\hat{\Gamma}(0)^{-1/2} \otimes \hat{\Gamma}(0)^{1/2}] vec[\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)'] \\
&= T vec[\hat{\Gamma}(0)^{-1/2}\hat{\Gamma}(1)'\hat{\Gamma}(0)^{-1/2}]' vec[\hat{\Gamma}(0)^{-1/2}\hat{\Gamma}(1)'\hat{\Gamma}(0)^{-1/2}],
\end{aligned}
$$

where we use the equality: $vec(ABC) = (C' \otimes A)vecB$ [see e.g. Lemma 4.3.1 in Horn, Johnson (1999), or Magnus, Neudecker (2019), ch. 18, p. 440-441]. Moreover, we have $[vecC]'[vecC] = Tr\ CC'$. Therefore,

$$
\begin{aligned}
\xi(1) &= T\ Tr[\hat{\Gamma}(0)^{-1/2}\hat{\Gamma}(1)'\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)\hat{\Gamma}(0)^{-1/2}] \\
&= T\ Tr[\hat{\Gamma}(1)'\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)\hat{\Gamma}(0)^{-1}] \\
&= T\ Tr\hat{R}^2(1). \qquad \text{(a.3)}
\end{aligned}
$$

Thus, the LM statistic is a function of the sample $\hat{R}^2(1)$ of the VAR(1) model.

iii) Asymptotic behavior of $\hat{\Gamma}(1)$

From (a.2) we easily deduce the asymptotic distribution of $\sqrt{T} vec[\hat{\Gamma}(1)']$ [see also Chitturi (1976), Hannan(1976)]. Indeed, we have:

$$
\sqrt{T}\hat{\Gamma}(1)' = \hat{\Gamma}(0)\sqrt{T}\hat{B}' \approx \Gamma(0)\sqrt{T}\hat{B}',
$$

and then

$$
vec[\sqrt{T}\hat{\Gamma}(1)'] = vec[\Gamma(0)\sqrt{T}\hat{B}'] = [Id \otimes \Gamma(0)]vec(\sqrt{T}\hat{B}').
$$

It follows that:

$$
\begin{aligned}
vec[\sqrt{T}\hat{\Gamma}(1)'] &\sim N[0, [Id \otimes \Gamma(0)][\Gamma(0) \otimes \Gamma(0)^{-1}][Id \otimes \Gamma(0)]] \\
&= N[0, \Gamma(0) \otimes \Gamma(0)].
\end{aligned}
$$

Although the above discussion concerned $h = 1$, i.e. $\hat{\Gamma}(1)$, it is also valid for $X_t = Y_{t-h}$, or $X_t = (Y'_{t-1}, ..., Y'_{t-H})'$. Then, it is easy to see that under H_0, vectors $vec[\sqrt{T}\hat{\Gamma}_T(h)]$, $h = 1, ..., H$ are asymptotically independent and have the same Gaussian distribution $N[0, \Gamma(0) \otimes \Gamma(0)]$.

2.2 APPENDIX 2: Asymptotic Expansions

As the objective function, i.e. the portmanteau statistic, is a sum of similar terms in h, we provide the derivatives and expansions for $H = 1$, given that the extension to any $H > 1$ is straightforward.

A.2.1 First-Order Derivative of $Tr\, R^2(1;\theta)$

Without loss of generality, we consider lag $h = 1$. At lag $h = 1$, we have

$$Tr\, R^2(1;\theta) = Tr[\Gamma(1;\theta)\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0,\theta)^{-1}].$$

The first-order partial derivatives are computed by considering the differential:

$$d\, Tr\, R^2(1;\theta) = \sum_{j=1}^{J} \frac{\partial Tr\, R^2(1;\theta)}{\partial \theta_j} d\theta_j.$$

We have:

$$\begin{aligned} d\, Tr\, R^2(1;\theta) &= Tr[d\Gamma(1;\theta)\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0,\theta)^{-1}] \\ &+ Tr[\Gamma(1;\theta)d[\Gamma(0;\theta)^{-1}]\Gamma(1;\theta)'\Gamma(0,\theta)^{-1}] \\ &+ Tr[\Gamma(1;\theta)\Gamma(0;\theta)^{-1}d\Gamma(1;\theta)'\Gamma(0,\theta)^{-1}] \\ &+ Tr[\Gamma(1;\theta)\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'d[\Gamma(0,\theta)^{-1}]], \end{aligned}$$

because $d[A(\theta)B(\theta)] = dA(\theta)B(\theta) + A(\theta)dB(\theta)$ and $d\, Tr[A(\theta) + B(\theta)] = Tr\, [dA(\theta) + dB(\theta)]$. In addition, we know that $Tr\,(AB) = Tr\,(BA), Tr\, A' = Tr\, A$. Hence,

$$\begin{aligned} d\, Tr\, [R^2(1;\theta)] &= Tr\, [\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}d\Gamma(1;\theta)] \\ &+ Tr\, [\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}\Gamma(1;\theta)d[\Gamma(0;\theta)^{-1}]] \\ &+ Tr\, [d\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}\Gamma(1;\theta)\Gamma(0;\theta)^{-1}] \\ &+ Tr\, [\Gamma(1;\theta)\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'d[\Gamma(0;\theta)^{-1}] \\ &= 2Tr\, [\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}d\Gamma(1;\theta)] \\ &+ Tr\, \left[[\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}\Gamma(1;\theta) + \Gamma(1;\theta)\Gamma(0;\theta)^{-1}\Gamma(1;\theta)']d[\Gamma(0;\theta)^{-1}]\right]. \end{aligned}$$

We have:

$$d[A(\theta)^{-1}] = -A(\theta)^{-1}dA(\theta)A(\theta)^{-1}.$$

Next, let us substitute this expression into the formula of $d\,Tr\,[R^2(1;\theta)]$:

$$\begin{aligned} d\,Tr\,R^2(1;\theta) &= 2\,Tr\,[\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}d\Gamma(1;\theta)] \\ &- Tr\,\left[\Gamma(0;\theta)^{-1}[\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}\Gamma(1;\theta)+\Gamma(1;\theta)\Gamma(0;\theta)^{-1}\Gamma(1;\theta)']\Gamma(0;\theta)^{-1}d\Gamma(0;\theta)\right] \\ &= 2Tr\,[\Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}d\Gamma(1;\theta)] \\ &- Tr\,\left[[\tilde{R}^2(1;\theta)\Gamma(0;\theta)^{-1}+\Gamma(0;\theta)^{-1}R^2(1;\theta)]d\Gamma(0;\theta)\right], \end{aligned} \tag{a.4}$$

with

$$\tilde{R}^2(1;\theta) = \Gamma(0;\theta)^{-1}\Gamma(1;\theta)'\Gamma(0;\theta)^{-1}\Gamma(1;\theta). \tag{a.5}$$

A.2.2. First-Order Conditions (FOC)

The First-Order Conditions are:

$$\begin{aligned} &\frac{\partial Tr\,\hat{R}^2(1;\theta_j)}{\partial\theta_j} = 0,\ j=1,...,J=dim\theta \\ \iff\quad & 2\,Tr[\hat{\Gamma}(0;\theta)^{-1}\hat{\Gamma}(1;\theta)'\hat{\Gamma}(0;\theta)^{-1}\frac{\partial\hat{\Gamma}(1;\theta)}{\partial\theta_j}] \\ & -Tr\{\,[\hat{\tilde{R}}^2(1;\theta)\hat{\Gamma}(0;\theta)^{-1}+\hat{\Gamma}(0;\theta)^{-1}\hat{R}^2(1,\theta)]\frac{\partial\hat{\Gamma}(0;\theta)}{\partial\theta_j}\} = 0, \\ & j=1,...,J=dim\theta. \end{aligned} \tag{a.6}$$

A.2.3. Second–Order Expansion

We can derive the first-order conditions (FOC) in the neighborhood of $\hat{\theta}\approx\theta_0$. For some matrix function $A(\theta)$ of θ, we consider the differential defined by:

$$dA(\theta) = \sum_{j=1}^{J}\frac{\partial A(\theta)}{\partial\theta_j}d\theta_j,$$

with $d\theta_j = \hat{\theta}_j - \theta_{j0}$. From (a,6), we get:

$$\begin{aligned} dFOC_j(\theta) &= 2Tr\left[\frac{\partial\hat{\Gamma}(1,\theta)}{\partial\theta_j}\, d[\hat{\Gamma}(0;\theta)^{-1}\hat{\Gamma}(1;\theta)'\hat{\Gamma}(0;\theta)^{-1}]\right] \\ &+ 2\left[\hat{\Gamma}(0;\theta)^{-1}\hat{\Gamma}(1;\theta)'\hat{\Gamma}(0;\theta)^{-1} d[\frac{\partial\hat{\Gamma}(1;\theta)}{\partial\theta_j}]\right] \\ &- Tr\left[\hat{\tilde{R}}^2(1;\theta)\hat{\Gamma}(0;\theta)^{-1} + \hat{\Gamma}(0;\theta)^{-1}\hat{R}^2(1;\theta)\right] d[\frac{\partial\hat{\Gamma}(0;\theta)}{\partial\theta_j}] \\ &- Tr\,\{[\frac{\partial\hat{\Gamma}(0,\theta)}{\partial\theta_j}]\, d[\hat{\tilde{R}}^2(1;\theta)\hat{\Gamma}(0;\theta)^{-1} + \hat{\Gamma}(0;\theta)^{-1}\hat{R}^2(1;\theta)]\}. \end{aligned} \tag{a.7}$$

In the above expression the terms without differential have to be evaluated at $\theta = \theta_0$.

The expressions involving differential term $d[.]$ are linear with respect to $\hat{\theta} - \theta_0$ and then of order $1/\sqrt{T}$. Moreover, for $\theta = \theta_0$, we have $\Gamma(0;\theta_0)$ invertible, $\Gamma(1;\theta_0) = 0$. Therefore $\hat{\Gamma}(0;\theta_0)$ is of order 1 and $\hat{\Gamma}(1;\theta_0)$ of order $1/\sqrt{T}$. We deduce that the second and third components of the right hand side of the above equation (a.7) evaluated at $\theta = \theta_0$ are negligible with respect to the other components. Then, the right hand side of equation (a.7) can be replaced by:

$$\begin{aligned} &2Tr\left[\frac{\partial\hat{\Gamma}(1,\theta_0)}{\partial\theta_j}\, d[\hat{\Gamma}(0;\theta)^{-1}\hat{\Gamma}(1;\theta)'\hat{\Gamma}(0;\theta)^{-1}]\right] \\ -\ &Tr\left[\frac{\partial\Gamma(0,\theta_0)}{\partial\theta_j} d[\hat{\Gamma}(0;\theta)^{-1}\hat{\Gamma}(1,\theta)'\hat{\Gamma}(0;\theta)^{-1}\hat{\Gamma}(1,\theta) + \hat{\Gamma}(1,\theta)\hat{\Gamma}(0,\theta)^{-1}\hat{\Gamma}(1,\theta)'\hat{\Gamma}(0;\theta)^{-1}]\right]. \end{aligned}$$

i) **The matrix** $J(\theta)$

Next, we use $d[A(\theta)B(\theta)] = [dA(\theta)]B(\theta) + A(\theta)[dB(\theta)]$. Let us consider the limiting FOC, when the sample autocovariances are replaced by their theoretical counterparts and use

$\Gamma(1;\theta_0) = 0$. We get:

$$
\begin{aligned}
dFOC_j(\theta) &= 2Tr\,[\frac{\partial\Gamma(1;\theta_0)}{\partial\theta_j}\,\Gamma(0;\theta_0)^{-1}d\Gamma(1;\theta)'\Gamma(0;\theta_0)^{-1}] \\
&- \;\; Tr[\frac{\partial\Gamma(0,\theta_0)}{\partial\theta_j}[\Gamma(0;\theta_0)^{-1}d\Gamma(1,\theta)'\Gamma(0;\theta_0)^{-1}\Gamma(1,\theta_0) \\
&+ \;\; \Gamma(0,\theta_0)^{-1}\Gamma(1,\theta_0)'\Gamma(0,\theta_0)^{-1}d\Gamma(1,\theta) + d\Gamma(1,\theta)'\Gamma(0;\theta_0)^{-1}\Gamma(1,\theta_0)'\Gamma(0,\theta_0)^{-1} \\
&+ \;\; \Gamma(1,\theta_0)\Gamma(0;\theta_0)^{-1}d\Gamma(1,\theta)'\Gamma(0;\theta_0)^{-1}]] \\
&= \;\; 2\,Tr\,[\frac{\partial\Gamma(1;\theta_0)}{\partial\theta_j}\Gamma(0;\theta_0)^{-1}d\Gamma(1;\theta)'\Gamma(0;\theta_0)^{-1}].
\end{aligned}
$$

The matrix $J(\theta_0)$ of second-order derivatives has elements (j,k) such that:

$$
-J_{jk}(\theta_0) = 2Tr\,\left[\frac{\partial\Gamma(1;\theta_0)}{\partial\theta_j}\Gamma(0;\theta_0)^{-1}\frac{\partial\Gamma(1;\theta_0)'}{\partial\theta_k}\Gamma(0;\theta_0)^{-1}\right],\;\forall j,k,
$$

or, equivalently:

$$
J(\theta_0) = -2\frac{\partial vec\Gamma(1;\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0)^{-1}\otimes\Gamma(0;\theta_0)^{-1}]\frac{\partial vec\Gamma(1;\theta_0)}{\partial\theta'}. \tag{a.8}
$$

ii) **Asymptotic equivalence for** $\sqrt{T}(\hat\theta_T - \theta_0)$ (written for $H = 1$)
Hence we have: $\sqrt{T}[\hat\theta_T - \theta_0] = J(\theta_0)^{-1}\sqrt{T}\frac{dL_T(\theta)}{d\theta} = J(\theta_0)^{-1}\sqrt{T}X(\hat\Gamma) + o_p(1)$, where $o_p(1)$ is negligible in probability and it follows from (a.6) that:

$$
\begin{aligned}
X_j(\hat\Gamma) &= 2Tr[\hat\Gamma(0;\theta_0)^{-1}\hat\Gamma(1,\theta_0)'\hat\Gamma(0;\theta_0)^{-1}\frac{\partial\hat\Gamma(1,\theta_0)}{\partial\theta_j}] \\
&- \;\; Tr\,\{[\hat{\hat{R}}^2(1;\theta_0)\hat\Gamma(0;\theta_0)^{-1} + \hat\Gamma(0;\theta_0)^{-1}\hat R^2(1;\theta_0)]\frac{\partial\hat\Gamma(1,\theta_0)}{\partial\theta_j}\},
\end{aligned}
$$

It is normally distributed as $\hat\Gamma \approx \Gamma$ with the Central Limit Theorem that holds for the sample autocovariances. We have $X_j(\hat\Gamma) - X_j(\Gamma) = dX_j(\Gamma)$. We can disregard all terms including $\Gamma(1,\theta_0)$ as $\Gamma(1;\theta_0) = 0$. Hence,

$$
\begin{aligned}
\sqrt{T}X_j(\hat\Gamma) &= 2Tr\,[\Gamma(0;\theta_0)^{-1}\sqrt{T}\hat\Gamma(1;\theta_0)\Gamma(0;\theta_0)^{-1}\frac{\partial\Gamma'(1;\theta_0)}{\partial\theta_j}] + o_p(1) \\
&= 2Tr\,[\Gamma(0;\theta_0)^{-1}\frac{\partial\Gamma'(1;\theta_0)}{\partial\theta_j}\Gamma(0;\theta_0)^{-1}\sqrt{T}\hat\Gamma(1;\theta_0)] + o_p(1),
\end{aligned}
$$

by applying $Tr\, A' = Tr\, A$ and $Tr(AB) = Tr(BA)$.

Since $Tr\,(A'B) = (vecA)'(vecB)$ and $vec(A \otimes C) = (C' \otimes A)$, we deduce:

$$\begin{aligned}\sqrt{T}X_j(\hat{\Gamma}) &\approx 2vec[\Gamma(0;\theta_0)^{-1}\frac{\partial\Gamma(1;\theta_0)}{\partial\theta_j}\Gamma(0;\theta_0)^{-1}]'vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)] \\ &= A_j(\theta)vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)],\ j = 1, ..., J = dim\theta.\end{aligned}$$

or,

$$\sqrt{T}X(\hat{\Gamma}) = A(\theta)vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)],$$

where

$$A(\theta) = 2\frac{\partial vec\Gamma(1;\theta)'}{\partial\theta}\left[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}\right]. \tag{a.9}$$

From the remark in Appendix 1 (Section 2.1), it follows that:

$$vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)] \approx N[0, \Gamma(0;\theta_0) \otimes \Gamma(0;\theta_0)].$$

Therefore $\sqrt{T}X(\hat{\Gamma})$ is asymptotically normally distributed, with mean zero and asymptotic variance:

$$V_{asy}[\sqrt{T}X(\hat{\Gamma})] = -2J(\theta_0). \tag{a.10}$$

From (a.10), and the absence of correlation between the matrices $\hat{\Gamma}(h)$ at different lags, we deduce the simplification in deriving the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ in Corollary 1.

2.3 APPENDIX 3: Expansion of the Multivariate Portmanteau Statistic

(i) Let us consider the asymptotic expansion of:

$$TL_T(\theta_0) = T\,Tr\,[\hat{\Gamma}(1;\theta_0)\hat{\Gamma}(0;\theta_0)^{-1}\hat{\Gamma}(1;\theta_0)'\hat{\Gamma}(0;\theta_0)^{-1}],$$

written for $H = 1$. Since $Tr(C'C) = [vec\,C]'[vec\,C]$, we have:

$$TL_T(\theta_0) = Tvec[\hat{\Gamma}(1;\theta_0)'\hat{\Gamma}(0;\theta_0)^{-1}]'vec[\hat{\Gamma}(1;\theta_0)'\hat{\Gamma}(0;\theta_0)^{-1}].$$

Moreover, since $vec(ABC) = (C' \otimes A)vecB$, we get:

$$TL_T(\theta_0) \;\approx\; [\sqrt{T}vec\hat{\Gamma}'(1;\theta_0)]'[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}][\sqrt{T}vec\hat{\Gamma}'(1;\theta_0)].$$

(ii) Then, let us consider the asymptotic expansion of the standardized residual-based portmanteau statistic. By using (a.8), (a.9) and the above result, we get:

$$TL_T(\hat{\theta}_T) \approx vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)']'\,\Pi\,vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)'], \tag{a.11}$$

with

$$\begin{aligned}\Pi &= \Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1} \\ &- [\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]\frac{\partial vec\Gamma(1;\theta_0)}{\partial\theta'} \\ &\quad \left\{\frac{\partial vec\Gamma(1;\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]\frac{\partial vec\Gamma(1;\theta_0)}{\partial\theta}\right\}^{-1} \\ &\quad \frac{\partial vec\Gamma(1;\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}].\end{aligned} \tag{a.12}$$

(iii) The condition

$$\Pi\, V_{asy}[\sqrt{T}\,vec\hat{\Gamma}(1;\theta_0)']\,\Pi = \Pi,$$

with $V_{asy}\sqrt{T}\,vec[\hat{\Gamma}(1;\theta_0)'] = \Gamma(0;\theta_0) \otimes \Gamma(0;\theta_0)$ is verified. It is due to an interpretation in terms of orthogonal projection of $vec[\sqrt{T}\hat{\Gamma}'(1;\theta_0)']$ on the $\partial vec\Gamma(1;\theta_0)'/\partial\theta$ for the scalar product associated with $\Gamma(0;\theta_0) \otimes \Gamma(0;\theta_0)$.

2.4 APPENDIX 4: Semi-Parametric versus Parametric Efficiency

The GCov approach can be applied to stacked nonlinear transformations of the series: $Y_t = a(y_t)$, say. Let us consider $H = 1$ for expository purpose. Then, the asymptotic variance-covariance matrix of $\hat{\theta}_T$ depends on the choice of function a, including its dimension, and $\Omega(\theta_0; a)$ depends on a. Thus, we can expect that there exist a transformation a^*, say, such that:

$$a^* = Argmax_a \Omega(\theta_0, a).$$

Let us assume that process (y_t) is a univariate Markov process and consider a grid $A_k = [a_k, a_{k+1}]$ in Subsection 3.3. When the grid becomes very fine and the number of elements in the grid increases, we use more and more autocovariance conditions (see Section 8, iii)) and the chi-square measure becomes equivalent to the Kullback-Leibler measure between the copula density and the uniform density. Then, the GCov estimator is a maximum likelihood estimator based on the parametric copula. Therefore, as expected $\Omega^*(\theta_0) = Max_a \Omega(\theta_0, a)$ is the information matrix corresponding to this copula-based ML approach. In other words, when a varies, we reach a parametric efficiency bound for the parameters characterizing the copula. This is the property of adaptive estimation that follows from the i.i.d. assumption on the error terms [see e.g. Hall, Welsh (1983) for adaptive estimation].

3 Simulation

We explore the GCov estimation of a VAR(1) model with parameters $\phi_{11} = 0.9, \phi_{12} = -0.3, \phi_{2,1} = 0, \phi_{22} = 1.2$ and eigenvalues of matrix Φ equal to 0.9 and 1.2.

3.1 Long simulated path

Figure 1 shows the path of 1000 realizations of this process with t(4) distributed errors to illustrate the occurrence of spikes and bubbles in the trajectory of a stationary process.

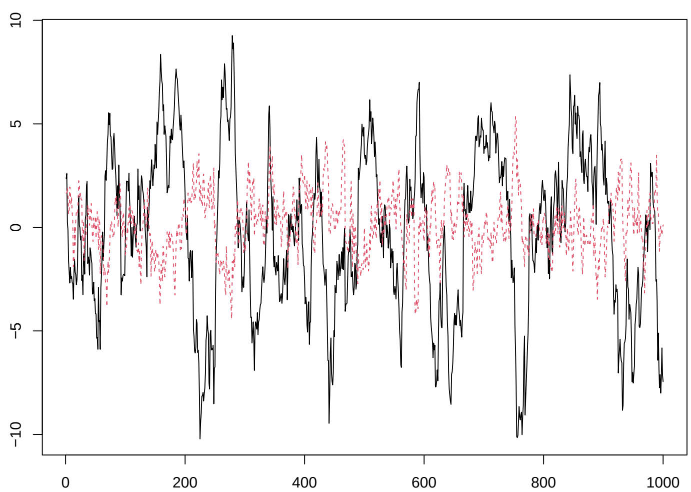


Figure 1: DGP with t(4) distributed errors, T=1000

3.2 Additional statistics

3.2.1 ACF

Figure 2 presents the ACF function computed from a sample of length T=200 of the DGP displayed in Figure 1 with t(4) distributed errors corresponding to the process in Section 7.1.

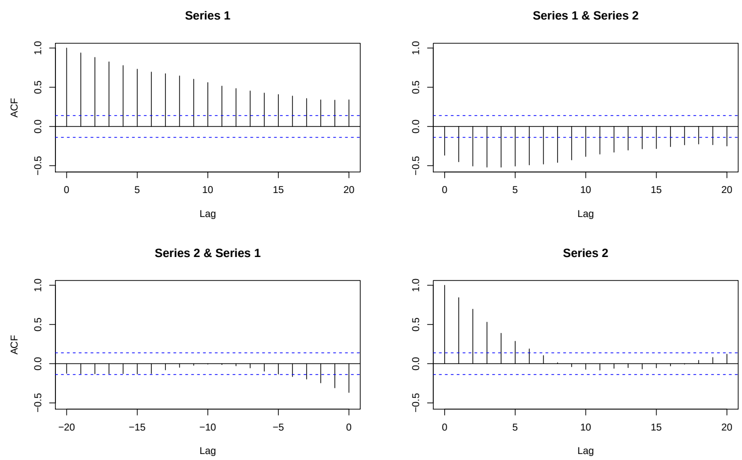


Figure 2: ACF, DGP with t(4) distributed errors

We observe a long range of persistence in the autocorrelations of $y_{1,t}$ and persistent cross-correlations of $y_{1,t}$ and $y_{2,t}$. Even though the process is stationary, it features local trends that can be mistakenly interpreted as global trends when the series is observed over a short time.

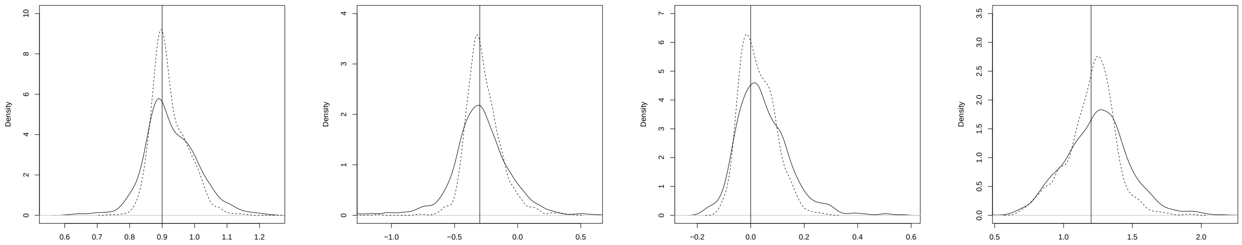
3.2.2 Finite sample distribution of GCov estimator

Figure 3 presents and compares the densities of GCov estimators computed from the process with t(4) distributed errors, and based on power transformations with exponents 1,2,3 and 4, number of lags H=10 and sample sizes T=200 and T=400.
The densities of GCov estimators for processes with t(6) distributed errors are displayed below.

The densities based on t(4) distributed errors have longer tails. The densities based on the sample size T=400 are closer to the normal density than those based on T=200.

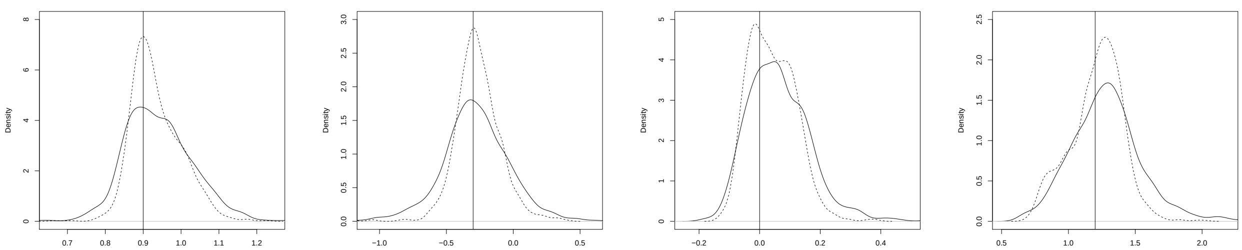
3.2.3 Effect of H

Next, we examine the effect of lag H by replicating 1000 times the DGP with t(4) distributed errors (kurtosis infinity) estimated from 2 and 4 transforms and the sample size T=400.



(a) $\hat{\phi}_1$ (b) $\hat{\phi}_{12}$ (c) $\hat{\phi}_{21}$ (d) $\hat{\phi}_{22}$

Figure 3: Densities of GCov estimators, close to 1 eigenvalues, t(4) distributed errors: T=200 (solid line), T=400 (dashed line)



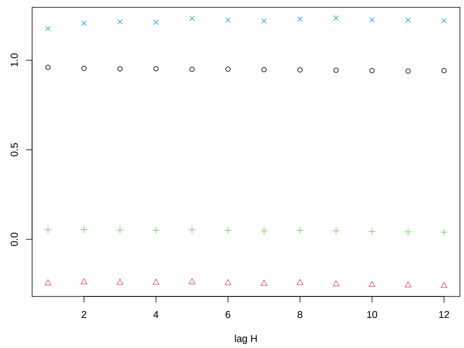
(a) $\hat{\phi}_{11}$ (b) $\hat{\phi}_{12}$ (c) $\hat{\phi}_{21}$ (d) $\hat{\phi}_{22}$

Figure 4: Densities of GCov estimators: T=200 (solid line), T=400 (dashed line) errors with finite kurtosis

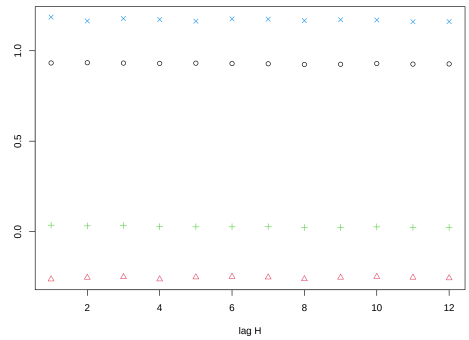
Figure 5 shows the sample means of GCov estimators plotted as a function of lag $H, H = 1, ..., 12$.

When 2 transforms are used, one may prefer to select lags H greater than 2, while when 4 transforms are used, the GCov estimators seem equivalent at any lag H.
Figure 6 shows the variances of estimators plotted as a function of lag $H, H = 1, ..., 12$ for the DGP with errors characterized by infinite kurtosis.

We observe that when 2 transforms are used, higher lags H offer greater chances of obtaining estimators with lower variances. In the case of 4 transforms, lower lags H are associated with lower variances of estimators.

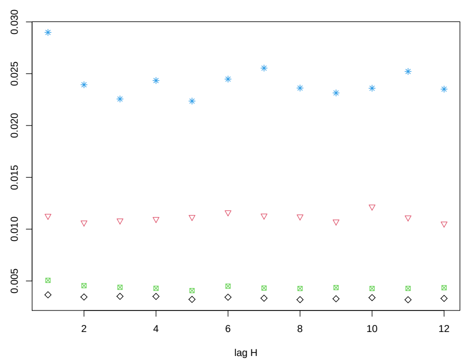


(a) 2 transforms

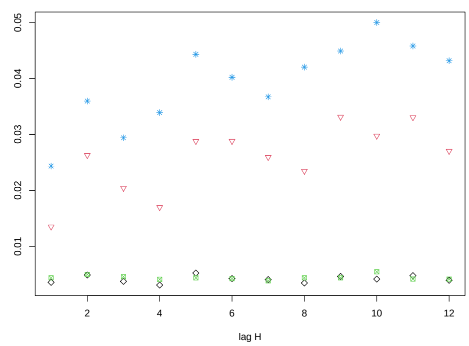


(b) 4 transforms

Figure 5: means of VAR(1) estimators: square for $\hat{\phi}_{11}$, circle for $\hat{\phi}_{12}$, triangle for $\hat{\phi}_{21}$, cross for $\hat{\phi}_{22}$



(a) 2 transforms



(b) 4 transforms

Figure 6: variances of VAR(1) estimators: diamond for $v\hat{a}r(\hat{\phi})_{11}$, square for $v\hat{a}r(\hat{\phi})_{12}$, triangle for $v\hat{a}r(\hat{\phi})_{21}$, star for $v\hat{a}r(\hat{\phi})_{22}$

4 Application to Commodity Prices

This section provides additional results for the empirical application presented in Section 7.2 of the paper. We provide the summary statistics and discuss the selection of H in GCov estimation. Next, the alternative OLS-based methods, the diagnostics for the mixed VAR(1) model and the mixed VAR(3) model are covered.

4.1 Summary statistics

The summary statistics of the wheat and corn futures are given in Table 1:

Table 1. Summary Statistics

series	q(0%)	q(100%)	q(50%)	mean	var	skew	kurtosis
wheat	371.50	539.25	428.75	433.91	769.73	1.07	4.42
corn	329.50	392.25	357.75	358.79	162.21	0.22	2.42

The summary statistics indicate that the series are not normally distributed. Moreover, the wheat and corn futures are characterized by long-range persistence. The serial correlation of the wheat and corn futures is illustrated by the autocorrelation function (ACF) plotted in Figure 7.

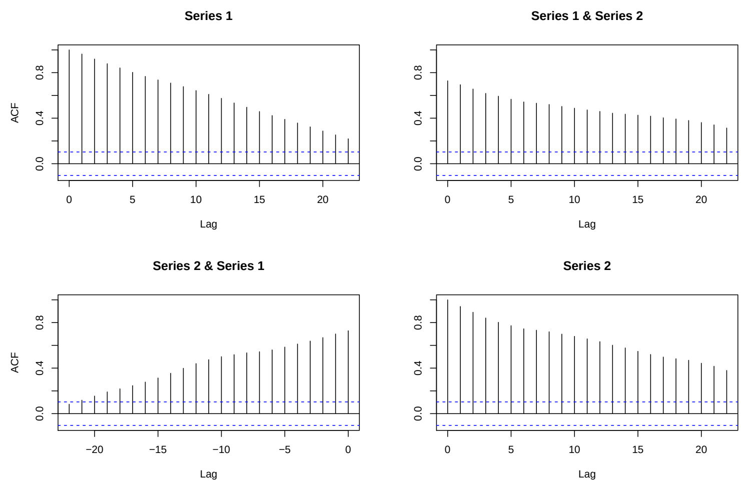


Figure 7: ACF series: wheat (Series 1) and corn (Series 2)

The ACF shows slowly decaying autocorrelations in the wheat and corn prices and persistent cross-correlations.

4.2 Selection of order H

We compute the GCov estimators of the series at various lags H. The results are presented in Figure 8.

We observe that the GCov estimators of VAR(1) parameters display some variation at short lags and become stable at lags 5 and higher. To determine the optimal lag choice, we plot the residual variances in Figure 9.

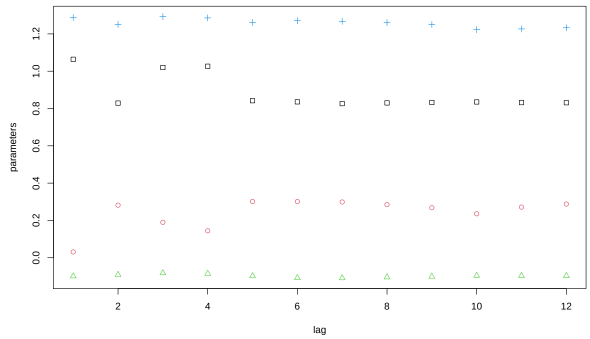


Figure 8: VAR(1) estimators: square for $\hat{\phi}_{11}$, circle for $\hat{\phi}_{12}$, triangle for $\hat{\phi}_{21}$, cross for $\hat{\phi}_{22}$

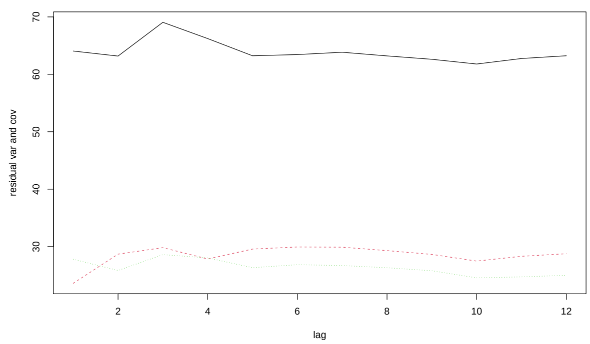
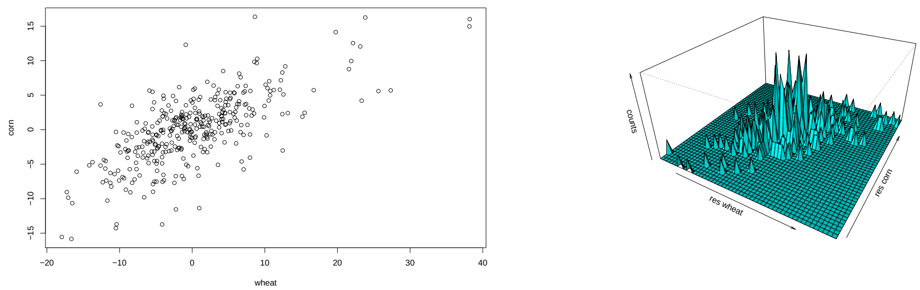


Figure 9: Residual variances and covariance as a function of H: solid line: wheat residual variance, dashed line: corn residual variance, dotted line: covariance

We observe that both residual variances are minimized at lag H=10 that justifies our choice of parameters estimates reported in Section 7.2.

4.3 Mixed VAR(1) : Diagnostics

Let us now consider the mixed VAR(1) model estimated by the GCov estimator and discuss the model diagnostics. The joint sample density of residuals is shown in Figure 10.



(a) residual: scatter (b) residual: histogram

Figure 10: Residuals: joint sample density

Figure 10 confirms the presence of nonlinear tail dependence due to some extreme values of residuals.

Let us now examine the serial dependence of u_t. We provide the ACF and cross-ACF functions of residuals in Figure 11 and of squared residuals in Figure 12. The serial dependence in the observed bivariate series is mostly accommodated by the mixed causal-noncausal VAR(1) model:

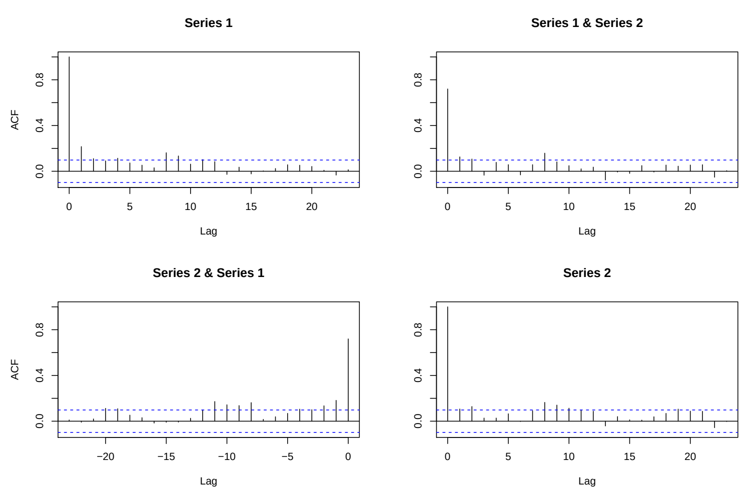


Figure 11: ACF residuals VAR(1): wheat (Series 1) and corn soy(Series 2)

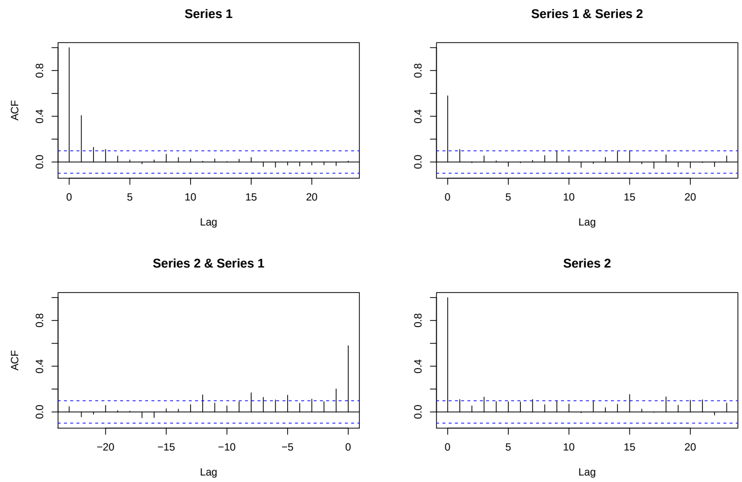


Figure 12: ACF squared residuals VAR(1): res1 (wheat) square (Series 1) and res 2 (corn) square (Series 2)

The residuals still display some slightly significant autocorrelations and cross-correlations.

The squared residuals display some slightly significant serial correlation as well. To accommodate the remaining serial correlation, we estimate the VAR(3) model below.

4.4 Mixed VAR(3) Model

4.4.1 Estimation

The mixed VAR(3) model is estimated by the GCov from the autocovariances of errors and squared errors with lag $H = 10$. The estimated model is as follows:

$$\hat{y}_t = \begin{bmatrix} 2.314 & -1.168 \\ 0.071 & 1.011 \end{bmatrix} y_{t-1} + \begin{bmatrix} -0.786 & 0.344 \\ -0.146 & -0.100 \end{bmatrix} y_{t-2} + \begin{bmatrix} -0.357 & 0.522 \\ 0.115 & -0.051 \end{bmatrix} y_{t-3}.$$

The roots of $det(Id - \Phi_1 z - \Phi_2 z^2 - \Phi_3 z^3) = 0$ are equal to the reciprocals of eigenvalues of the autoregressive matrix of a stacked VAR(1) representation of the process. They are given in Table A1:

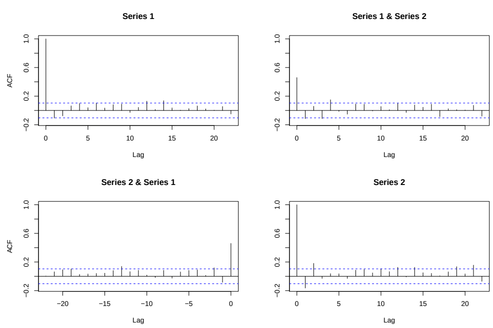
Table A1. Estimation of mixed VAR(3) model: eigenvalues

eigenvalues
1.7045
-0.3592
0.1161 + 0.2759i
0.1161 - 0.2759i
0.9125
0.8348

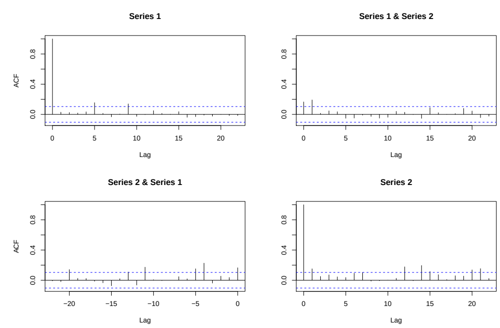
We find that the roots are both inside and outside the unit circle with the causal order equal to 5 and noncausal order equal to 1.

4.4.2 Residual ACF

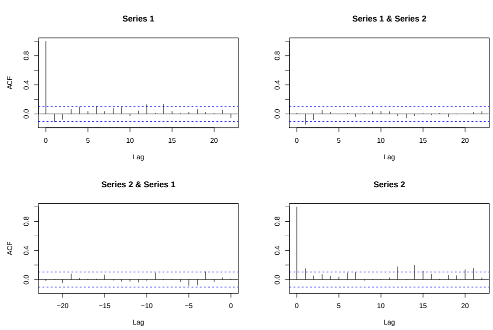
Figure 13 shows the ACF and cross-ACF of the residuals and squared residuals. The ACF's show almost no remaining serial dependence, even at lags greater than the maximum lag $H = 10$ used in the GCov estimation procedure.



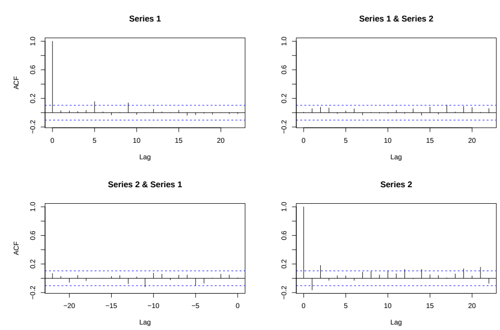
(a) residuals



(b) residuals square



(c) res 1, res 2 square



(d) res 1 square, res 2

Figure 13: ACF: Residuals of VAR(3)

5 Causal VAR Model

This Section links the GCov estimator with the literature on multivariate analysis of constrained causal VAR(p) models. These models are usually estimated by least squares estimators based on autocovariances of u_t with $H = p$, the causal autoregressive order. This analysis is efficient if the errors are Gaussian. Otherwise, a GCov estimator involving additional transforms of u_t can be more adequate to capture the skewness or kurtosis effects in u_t and their propagation through the AR dynamics.

The multivariate VAR(p) process is defined by:

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

where $\theta = [vec\Phi_1', ..., vec\Phi_p']'$.

The (causal) VAR model has been commonly used in the literature on residual-based portmanteau test [see e.g. Hosking (1980), Li, McLeod (1981)].
In the causal VAR specification, parameters $\Phi_1, ..., \Phi_p$ can be constrained by some linear or nonlinear restrictions. This includes in particular:

i) The VAR(1) model with reduced rank of Φ_1 [see, Velu et al. (1986), Ahn, Reinsel (1988), Engle, Kozicki (1993), Reinsel, Velu (1998), Anderson (2002), Lam, Yao (2012)].

ii) The VAR(p) model with common canonical directions at all lags [see Kettenring (1971), Neuenschwander, Flury (1995)].

iii) The VAR(p) processes with subprocesses, which are assumed independent [see, e.g. Haugh (1976), El Himdi, Roy (1997), Yata, Aoshima(2016), Jin, Matteson (2018)], or based on Dynamic Orthogonal Components [Matteson, Tsay (2011)].

iv) The VAR(p) model with causality restrictions [see e.g. Boudjellaba et al. (1994)].

v) The VAR(p) model with a Kronecker structure [Niu et al. (2020)].
and all other structural VAR models.

REFERENCES

Ahn, S., and G., Reinsel (1988): " Nested Reduced Rank Autoregressive Models for Multiple Time Series", JASA, 89, 849-956.

Boudjellaba, H., Dufour, J.M. and R. Roy (1994): "Simplified Conditions for Noncausality Between Vectors in Multivariate ARMA Models", Journal of Econometrics, 63, 271-287.

El Himdi, H. and R. Roy (1997): "Tests for Noncorrelation of Two Multivariate ARMA Time Series", Canadian Journal of Statistics, 25, 233-256.

Engle, R. and S. Kozicki (1993): "Testing for Common Features", Journal of Business and Economic Statistics", 11, 369-395.

Hall, P. and A. Welsh (1985): "Adaptive Estimates of Parameters of Regular Variation", Annals of Statistic, 13, 331-341.

Haugh, L. (1976): "Checking the Independence of Two Covariance-Stationary Time Series: A Univariate Residual Cross-Correlation Approach", JASA, 71, 378-385.

Horn, R. and C. Johnson (1999): "Topics in Matrix Analysis", Cambridge University Press.

Hosking, J. (1981)a: "Equivalent Forms of the Multivariate Portmanteau Statistic", JRSS B, 43, 261-262.

Hosking, J. (1981)b: "Lagrange Multiplier Tests of Multivariate Time Series Models", JRSS B, 43, 219-230.

Jin, Z. and D. Matteson (2018): "Generalizing Distance Covariance to Measure and Test Multivariate Mutual Dependence via Complete and Incomplete V-Statistics", Journal of Multivariate Analysis, 168, 304-322.

Kettenring, J. (1971): "Canonical Analysis of Several Sets of Variables", Biometrika, 58, 433-451.

Magnus, J. and H. Neudecker (2019): " Matrix Differential Calculus with Applications in Statistics and Econometrics", Wiley.

Matteson, D., and R. Tsay (2011): "Dynamic Orthogonal Components for Multivariate Time Series", JASA, 106, 1450-1463.

Neuenschwander, B. and B. Flury (1995): "Common Canonical Variates", Biometrika, 82, 553-560.

Niu, L., Liu, X. and J. Zhao (2020): "Robust Estimators of the Correlation Matrix with Sparse Kronecker Structure for a High Dimensional Matrix Variate", Journal of Multivariate Analysis, 177, Issue C.

Reinsel, G. and R. Velu (1998): "Multivariate Reduced Rank Regressions", Springer, New York.

Velu, R., Reinsel, G. and D. Wichern (1986): "Reduced Rank Models for Multiple Time Series", Biometrika, 73, 109-118.

Yata, K., and M. Aoshima (2016): "High Dimensional Inference on Covariance Structures via the Extended Cross-Data Matrix Methodology", Journal of Multivariate Analysis, 151, 151-166.