Generalized Covariance Estimator

Gourieroux*, C. and J. Jasiak†

August 23, 2022

Abstract

We consider a class of semi-parametric dynamic models with independent identically distributed errors, including the nonlinear mixed causal-noncausal Vector Autoregressive (VAR), Double-Autoregressive (DAR) and stochastic volatility models. To estimate the parameters characterizing the (nonlinear) serial dependence, we introduce a generic Generalized Covariance (GCov) estimator, which minimizes a residual-based multivariate portmanteau statistic. In comparison to the standard methods of moments, the GCov estimator has an interpretable objective function, circumvents the inversion of high-dimensional matrices, and achieves semi-parametric efficiency in one step. We derive the asymptotic properties of the GCov estimator and show its semi-parametric efficiency. We also prove that the associated residual-based portmanteau statistic is asymptotically chi-square distributed. The finite sample performance of the GCov estimator is illustrated in a simulation study. The estimator is then applied to a dynamic model of commodity futures.

Keywords: Semi-Parametric Estimator, Generalized Covariance Estimator, Portmanteau Statistic, Continuously Updating GMM, Canonical Correlation, Mixed Causal-Noncausal Process, Commodities.

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

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

1 Introduction

We consider a class of semi-parametric dynamic models with independent, identically distributed (i.i.d.) error terms. This class includes univariate and multivariate models, such as the Double-Autoregressive (DAR), stochastic volatility, and mixed causal-noncausal Vector Autoregressive (VAR) models. The i.i.d. assumption on the errors is used to define the Generalized Covariance (GCov) estimator of the parameters of interest, which minimizes a residual-based multivariate portmanteau statistic. The GCov is a one-step estimator that is shown to be consistent, asymptotically normally distributed, and semi-parametrically efficient.

The GCov estimator can be viewed as an alternative to the Continuously Updating Generalized Method of Moments (CUGMM) estimator [Hansen, Heaton, Yaron (1996)]. Unlike a GMM estimator, which is based on non-central moments, the GCov relies on central moments, or equivalently, both non-central moments and products of non-central moments. The expression of the asymptotic variance of the GCov estimator is relatively simple because of a convenient standardization of the minimization criterion. In comparison to the CUGMM, the GCov estimator does not require the inversion of matrices of high dimension [see e.g. Altonji and Segal (1996) for the finite sample problems arising from the inversion of the covariance matrix].

The GCov estimator can be applied to multivariate or univariate nonlinear dynamic models. In the latter case, the objective function to be minimized can be computed from a univariate series, or from several (nonlinear) transformations of that series obtained, for example, by discretizing the state space. For testing the hypothesis of i.i.d. errors, we consider the residual-based multivariate portmanteau statistic. We show that this statistic asymptotically follows a chi-square distribution with the degrees of freedom adjusted for the number of identifiable parameters. This extends the well-known result established for linear causal dynamic VAR-type models to a nonlinear dynamic framework.

The following notation is used. For any $m \times n$ matrix A whose jth column is a_j, $j =$

$1, ..., n$, $vec(A)$ will denote the column vector of dimension mn defined as $vec(A) = (a_1' ..., a_j', ..., a_n')'$, where the prime denotes transposition. For any two matrices $A \equiv (a_{ij})$ and B, the Kronecker product $(A \otimes B)$ is the block matrix having $a_{ij}B$ for its (i, j)th block.

The paper is organized as follows. Section 2 recalls the interpretation and asymptotic distribution of the multivariate portmanteau statistic. Section 3 defines the GCov estimator and provides examples of semi-parametric dynamic models, which can be estimated by the GCov estimator. We also discuss the identifiability of the parameters of interest characterizing the serial dependence. Section 4 presents the results on the consistency and asymptotic normality of the GCov estimator. We discuss the simplification of the sandwich formula of asymptotic variance and the differences between the GCov estimator and alternative method of moments estimators with respect to the computational feasibility and choice of instruments. The residual-based multivariate portmanteau statistic and its distribution are examined in Section 5. In Section 6, the generic GCov approach is extended to models with errors that are both serially and cross-sectionally independent, with a special emphasis on the causal/noncausal SVAR models. The finite-sample performance of the GCov estimator is illustrated in Section 7 in a simulation study of mixed causal-noncausal models and in an application to a dynamic model of commodity futures. Section 8 concludes. Proofs and asymptotic expansions are gathered in Appendices 1 to 4 in Sections 2.1-2.4 of Supplemental Material, which also contains additional examples, tables, figures, and references.

2 The Portmanteau Statistic

Let us consider a univariate strictly stationary time series (y_t) with finite fourth-order moments. The test of the null hypothesis $H_0 = \{\gamma(h) = 0, h = 1, ..., H\}$, with $\gamma(h) = Cov(y_t, y_{t-h})$ is commonly based on the test statistic:

$$\xi(H) = T \sum_{h=1}^{H} \hat{\rho}(h)^2 = T \sum_{h=1}^{H} \frac{\hat{\gamma}(h)^2}{\hat{\gamma}(0)^2}, \tag{1}$$

where $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are the sample autocovariance and autocorrelation of order h, respectively.

Under the i.i.d. assumption on (y_t) and the existence of fourth-order moments, this statistic asymptotically follows a chi-square distribution $\chi^2(H)$ with H degrees of freedom [see, Box, Pierce (1970)]. The aim of this Section is to review the analogue of this statistic for strictly stationary time series of higher dimensions.

Let us now consider a multivariate strictly stationary time series (Y_t) of dimension K with finite fourth-order moments. Even though the assumption of existence of the fourth-order moment is needed to derive the asymptotic distribution, that distribution does not depend on the fourth-order cumulant, or more generally, on the distribution of Y_t [see Bartlett (1946) p.30]. Then, the multivariate analogue of statistic (1) for testing the null hypothesis $H_0 = \{\Gamma(h) = 0, h = 1, ..., H\}$, where $\Gamma(h) = cov(Y_t, Y_{t-h})$ denotes the autocovariance matrix of order h, is:

$$\xi(H) = T \sum_{h=1}^{H} Tr[\hat{R}^2(h)], \tag{2}$$

where $\hat{R}^2(h)$ is the sample analogue of the multivariate R-square defined by:

$$R^2(h) = \Gamma(h)\Gamma(0)^{-1}\Gamma(h)'\Gamma(0)^{-1}. \tag{3}$$

Since

$$\hat{R}^2(h) = \hat{\Gamma}(0)^{1/2}[\hat{\Gamma}(0)^{-1/2}\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}\hat{\Gamma}(h)'\hat{\Gamma}(0)^{-1/2}]\hat{\Gamma}(0)^{-1/2},$$

where $\hat{\Gamma}(h)$ denotes the sample autocovariance matrix, the sample R-square is equivalent, up to a change of basis, to the matrix within brackets, which is symmetric and positive-definite. Therefore, it is diagonalisable with a trace equal to the sum of its eigenvalues that are the squares of the canonical correlations between Y_t and Y_{t-h}, denoted by $\hat{\rho}_j^2(h), j = 1, ..., K$

[Hotelling (1936)]. Hence:

$$\xi(H) = T \sum_{h=1}^{H} Tr[\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}\hat{\Gamma}(h)'\hat{\Gamma}(0)^{-1}] = T \sum_{h=1}^{H} [\sum_{j=1}^{K} \hat{\rho}_j(h)^2]. \tag{4}$$

Under the assumption of independently and identically distributed (i.i.d.) process (Y_t) with finite fourth order moments, this statistic follows asymptotically a chi-square distribution $\chi^2(KH)$ [see, e.g. Robinson (1973), Chitturi (1976), Anderson (1999), Section 7, Anderson (2002), Section 5].

Remark 1: In the literature, there exist alternative test statistics that are asymptotically equivalent to the test statistic $\xi(H)$ under the null. For example, we can consider the causal VAR(H) model:

$$Y_t = \alpha + B_1 Y_{t-1} + \cdots + B_H Y_{t-H} + u_t, \tag{5}$$

and apply the Frisch-Waugh-Lovell theorem to obtain the statistic:

$$\tilde{\xi}(H) = T\ Tr[\hat{\Gamma}^*(1)\hat{\Gamma}^*(0)^{-1}\hat{\Gamma}^*(1)'\hat{\Gamma}^*(0)^{-1}], \tag{6}$$

where $\Gamma^*(1) = Cov(Y_t, \underline{Y_{t-1}}), \Gamma^*(0) = V(\underline{Y_{t-1}})$ and $\underline{Y_{t-1}} = (Y'_{t-1}, ..., Y'_{t-H})'$.

The main difference between matrices Γ and Γ^* is their dimension: Γ is the covariance between one current and one lagged value of the series, while Γ^* is the covariance between one current and multiple past values of the series. Under the i.i.d. assumption on u_t and under the null hypothesis $H_0 : \{B_1 = \cdots = B_H = 0\}$, the explanatory variables in (5) are (asymptotically) uncorrelated, as well as the autoregressive coefficients $\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}$ [Mann, Wald (1943), Chitturi (1974), eq. (2.9)]. This allows us to use H canonical correlations of smaller dimension K given in (4) instead of a canonical correlation of dimension KH given in (6). We derive the asymptotic distribution of the statistic under the null in Appendix 1 (Supplemental Material).

3 The Generic GCov Approach

We introduce a class of semi-parametric multivariate nonlinear dynamic models with i.i.d. errors and propose parameter estimators, which minimize the statistic $L_T(\theta) = \sum_{h=1}^{H} Tr\hat{R}^2(h;\theta)$ evaluated from nonlinear transformations $g(Y_t, ..., Y_{t-h}; \theta) = g(\tilde{Y}_t; \theta)$ of an observed process with unknown parameters θ. The minimizer of $L_T(\theta)$ is called the Generalized Covariance (GCov) estimator.

3.1 Semi-Parametric Model

Let us consider a strictly stationary process (Y_t) satisfying a semi-parametric model of type:

$$g(\tilde{Y}_t; \theta) = u_t, \tag{7}$$

where g is a known function, $\tilde{Y}_t = (Y_t, Y_{t-1}, \ldots, Y_{t-L})$, L is a non-negative integer, (u_t) is an i.i.d. sequence, not necessarily with zero mean, and θ is an unknown parameter vector. We assume that the model is well-specified and the true value of parameter θ is θ_0. Model (7) does not imply a nonlinear causal autoregressive specification of order L for process (Y_t) because the dimension of Y_t can be strictly larger (resp. smaller) than the dimension of u_t. Hence, model (7) is not directly invertible with respect to Y_t. Moreover, u_t is not assumed to be independent of $\tilde{Y}_{t-1}$. Therefore, the information generated by the current and lagged values of Y_t does not necessarily coincide with the information generated by the current and lagged values of u_t. The errors u_t are not necessarily interpretable as either causal, or non-causal innovations.

In the multivariate framework, model (7) can include nonlinear simultaneity effects, and as a simultaneous system may admit several stationary dynamic equilibria (Y_t). The GCov estimator remains valid regardless of the equilibria observed. A nonlinear structural VAR model, commonly studied in the literature and illustrated below, admits the semi-parametric representation (7).

While the second-order assumptions on the errors can be sufficient to define the GCov estimator and some of its asymptotic properties, the i.i.d. assumption is needed when various transformations of the series are considered and the difference between the parametric and semi-parametric efficiency is discussed [3]. The i.i.d. assumption is also needed for structural interpretations of nonlinear impulse responses [see Gourieroux, Jasiak (2005),(2022), Gourieroux, Monfort, Renne (2017), Sims (2021)]. In addition, it is commonly used in the recent literature on portmanteau tests [see Hoga (2021), Section 2].

The GCov estimator is applicable to a variety of univariate or multivariate models written in semi-parametric representations (7). The examples are given below:

Example 1: Double Autoregressive and Stochastic Volatility Models

i) A general model encompassing the Double-Autoregressive (DAR) model [Ling (2004)] can be written as:

$$y_t = m(y_{t-1}; \theta_1) + \theta_3 \sigma(y_{t-1}; \theta_2) + \sigma(y_{t-1}; \theta_2) u_t.$$

We assume that the regularity conditions on functions m, σ ensuring the existence of a stationary solution are satisfied. Moreover, the initial value y_0, is assumed drawn in the stationary distribution. The above model extends the standard ARCH-M model by allowing for nonlinear drift and volatility functions. Its semi-parametric representation (7) is:

$$g(\tilde{y}_t; \theta) = \frac{y_t - m(y_{t-1}; \theta_1) - \theta_3 \sigma(y_{t-1}; \theta_2)}{\sigma(y_{t-1}; \theta_2)}, \text{ with } \tilde{y}_t = (y_t, y_{t-1}).$$

ii) The stochastic volatility model below includes volatility shocks and ensures the coherency of its discrete and continuous time representation [Nelson (1990)]:

$$y_t = \alpha + \beta y_{t-1} + \gamma \sigma_t^2 + \sigma_t u_t, \;\; \log \sigma_t^2 = a + b \log \sigma_{t-1}^2 + v_t,$$

where the bivariate errors (u_t, v_t) are i.i.d. sequences. This is an exponential ARCH-M model with stochastic volatility and nonlinear simultaneity in (y_t, σ_t). We assume satisfied

[3] In this respect, a weaker assumption of martingale difference sequence (mds) used, for example, in Velasco (2022) would be inadequate, as it is not invariant to nonlinear tranformations.

the sufficient regularity conditions of stationarity, such as $|\beta| < 1, |b| < 1$, and independence of u_t and v_t. We also suppose that the initial values y_0 and σ_0 are drawn in the stationary distribution. Let the bivariate process be denoted by $Y_t = (y_t, \sigma_t)'$. Then:

$$g(\tilde{Y}_t; \theta) = [(y_t - \alpha - \beta y_{t-1} - \gamma \sigma_t^2)/\sigma_t, \;\; \log \sigma_t^2 - a - b \log \sigma_{t-1}^2],$$

is the semi-parametric representation (7) of this model with a bidimensional error $U_t = (u_t, v_t)'$. In practice, the returns y_t are observed and σ_t is replaced by either the implied or realized volatility computed from high frequency data as a proxy for σ_t.

Example 2: Causal-Noncausal VAR Model

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

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

where $\theta = [vec\Phi_1', ..., vec\Phi_p']'$ and error u_t is a multivariate non-Gaussian i.i.d. process with finite fourth order moments. We assume that the roots of the characteristic equation $det(Id - \Phi_1 \lambda - \cdots \Phi_p \lambda^p) = 0$ are of modulus either strictly greater, or smaller than one. Then, there exists a unique (strictly) stationary solution (Y_t) with a two-sided $MA(\infty)$ representation, which satisfies model (7) with:

$$g_t(\tilde{Y}_t, \theta) = Y_t - \Phi_1 Y_{t-1} - \cdots - \Phi_p Y_{t-p} = u_t.$$

The causal-noncausal VAR(p) model has been studied in Gourieroux, Jasiak (2016),(2017) and Davis, Song (2020). The error u_t cannot be interpreted as an innovation. Moreover, even if function g is linear in the current and lagged values of Y_t, the assumption of strict stationarity of Y_t implies the nonlinear causal dynamics of Y_t with predictions $E(Y_t|\underline{Y_{t-1}})$ nonlinear in $\underline{Y_{t-1}} = (Y_t, Y_{t-1}, ...)$ and past-conditional heteroscedasticity $V(Y_t|\underline{Y_{t-1}})$.

Model (7) can be transformed into a system of higher dimension by considering nonlinear transformations of u_t. Let us introduce J nonlinear transformations $a_1, ..., a_J$. Then we have:

$$\begin{aligned} a_j[g(\tilde{y}_t; \theta)] &= a_j(u_t), \; j = 1, ...J, \\ \text{or, equivalently} \;\; a[g(\tilde{y}_t; \theta)] &= a(u_t) = v_t, \end{aligned} \tag{8}$$

where the transformed process (v_t) is also an i.i.d. process.

The examples of commonly stacked transformations are the financial returns y_t and their squares: $Y_t = (y_t, y_t^2)$ [see Wooldridge (1991), Li, Mak (1994), Velasco, Wang (2015)], returns and their absolute values $Y_t = (y_t, |y_t|)$ [see Pena, Rodriguez (2006)], or return signs and their squares $Y_t = (sign(y_t), y_t^2)$, to separate the volatility dynamics from the bid-ask bounce effect.

3.2 The GCov Estimator

The GCov estimator of θ in model (7) is defined as:

$$\hat{\theta}_T(H) = Argmin_\theta \sum_{h=1}^{H} Tr[\hat{R}^2(h, \theta)], \tag{9}$$

where

$$\hat{R}^2(h, \theta) = \hat{\Gamma}(h; \theta)\hat{\Gamma}(0, \theta)^{-1}\hat{\Gamma}(h; \theta)'\hat{\Gamma}(0; \theta)^{-1}, \tag{10}$$

and $\hat{\Gamma}(h; \theta)$ is the sample covariance between $g(\tilde{Y}_t; \theta)$ and $g(\tilde{Y}_{t-h}; \theta)$.

Suppose that we observe $Y_1, \ldots, Y_T$. Then, the sample autocovariances of $g(\tilde{Y}_t; \theta)$ are computed from $t = L + H + 1$ up to T. These sample autocovariances have to be divided by T instead of $(T - H - L)$ to ensure that the sequence of multivariate sample covariances remains positive semi-definite.

The GCov estimator has no closed-form expression, except in special cases such as the unconstrained causal VAR model when the GCov and OLS estimates of parameters Φ_j are equivalent.

Remark 2: If (u_t) has no finite fourth-order moment, model (7) can be replaced by a transformed model (8), so that the transformed errors $a_j(u_t)$ have finite fourth-order moments. Then, the GCov estimator depends on the selected number of lags H and selected transformation a.

3.3 Discretization of the State Space

Let us consider a strictly stationary process (y_t) and a discretization of the state space defined by a partition: $A_k,\ k = 1, ..., K+1$. We introduce the indicator functions $Y_{k,t},\ k = 1, ..., K+1$, such that $Y_{k,t} = 1$, if $y_t \in A_k$, and $Y_{k,t} = 0$, otherwise. Then, the transformed variables, $Y_{k,t}$, have finite moments of any order, even if the moments of (y_t) do not exist. Since $\sum_{k=1}^{K+1} Y_{k,t} = 1,\ \forall t$, we only consider the K first components to define Y_t. Let us denote the K-dimensional vector with components $p_k = P[y_t \in A_k]$ by p and the $K \times K$ matrix with elements $p_{kl}(h) = P[y_t \in A_k, y_{t-h} \in A_l]$ by $P(h)$. We have:

$$\Gamma(h) = P(h) - pp',\ \Gamma(0) = diag p - pp' \text{ and } \Gamma(0)^{-1} = diag(P^{-1}) - \frac{ee'}{1-e'p},$$

where e is a vector of ones of dimension K. Then, it is easy to check that:

$$\sum_{h=1}^{H} Tr\, R^2(h) = \sum_{h=1}^{H} [\sum_{k=1}^{K+1} \sum_{l=1}^{K+1} \frac{(p_{kl}(h) - p_k p_l)^2}{p_k p_l}] = \sum_{h=1}^{H} \chi^{*2}(h), \tag{11}$$

where $\chi^{*2}(h)$ is the chi-square measure of (in)dependence between Y_t and Y_{t-h}. Thus, the GCov estimator minimizes a measure of pairwise (in)dependence (see Section 7). This state discretization can be used to reach the parametric efficiency of the GCov (see Appendix 4).

4 Asymptotic Properties of the GCov Estimator

This Section presents the asymptotic properties of the GCov estimator, which minimizes the objective function $L_T(\theta) = \sum_{h=1}^{H} Tr\, \hat{R}^2(h; \theta)$ evaluated from a sample of size T where $\hat{R}^2(h; \theta)$ is computed from the transformations $g(\tilde{Y}_t; \theta) = u_t$ given in (7). The asymptotic expansions and asymptotic variance formulas are derived in Appendix 2 (Supplemental Material).

4.1 Consistency and Identification

Under the strict stationarity of process (Y_t) and the existence of second-order moments of $g(\tilde{Y}_t, \theta)$, the sample autocovariances $\hat{\Gamma}(h; \theta)$ tend to their theoretical counterparts $\Gamma(h; \theta)$, when $T \to \infty$, and, if $\Gamma(0;\theta)$ is invertible for $\theta \in \Theta$, then $L_T(\theta)$ tends to:

$$L_\infty(\theta) = \sum_{h=1}^{H} Tr\,[R^2(h;\theta)].$$

If model (7) is well-specified, we have $L_\infty(\theta_0) = 0$, where the true value is the minimizer $\theta_0 = Argmin_\theta L_\infty(\theta)$. By applying the standard Jennrich's equicontinuity argument [Andrews (1992)], we get the consistency of the GCov estimator under an identification condition. More precisely, let us introduce the following set of assumptions:

Assumption A1:

i) Θ is a compact set in a separable metric space with a non-empty interior $\bar{\Theta}$.

ii) The model is well-specified with a true value $\theta_0 \in \Theta$.

iii) The matrices $\Gamma(h,\theta), h = 1, ..., H$ are continuous functions of θ on $\tilde{\Theta}$ and $\Gamma(0, \theta_0)$ is invertible.

iv) The process (Y_t) is strictly stationary[4], geometrically ergodic with a continuous invariant distribution.

v) $g(\tilde{Y};\theta)$ exists and is measurable with respect to $\tilde{Y}$ and continuous with respect to θ, $\theta \in \tilde{\Theta}$.

vi) The components $g_j(\tilde{Y}_t;\theta_0), j = 1, ..., J$ of $g(\tilde{Y}_t;\theta_0)$ are square integrable.

vii) The transformations $g_j(\tilde{Y}_t;\theta),\ j = 1, ..., J$ satisfy the following equicontinuity condition:

$$sup_{j=1,...,J}|g_j^2(\tilde{Y}_t;\theta) - g_j^2(\tilde{Y}_t;\tilde{\theta})| \leq \mathcal{B}(\tilde{Y}_t)h[d(\theta,\tilde{\theta})], \theta, \tilde{\theta} \in \Theta,$$

where $d(\theta,\tilde{\theta})$ is the distance on the metric parameter space, $h(d)$ is a function that tends to 0 when d tends to 0, and $\mathcal{B}(\tilde{Y}_t)$ is a sequence of non-negative variables, such that $sup_T \frac{1}{T}\sum_{t=1}^{T} E[\mathcal{B}(\tilde{Y}_t)] < \infty$.

This set of assumptions is sufficient to define the objective function, i.e. to ensure the invertibility of $\hat{\Gamma}(0;\theta)$ for $\theta \in \Theta$ and large T, and to prove the existence and consistency of

[4]This implies that $g(\tilde{Y}_t;\theta)$ is strictly stationary for any transformation g and value of parameter θ, and so are the true errors $u_t = g(\tilde{Y}_t;\theta_0)$.

the proposed estimator. The equicontinuity condition corresponds to Assumption W_LIP in Andrews (1992), p 248. This is a stochastic Lipschitz condition on $g_j(Y;\theta)$.

Proposition 1: *Under Assumption A1, if* $\theta = \theta_0$ *is the unique solution that minimizes the limiting objective function* $L_\infty(\theta)$ *on* Θ*, then there exists a measurable GCov estimator* $\hat{\theta}_T = Argmin_\Theta L_T(\theta)$ *that is weakly consistent of* θ_0*.*

The uniqueness condition of the solution of the limiting optimization is an identification condition. It implies that some parameters might not be consistently approximated by a GCov estimator. The parameters characterizing the (nonlinear) serial dependence are generally identifiable, but not the parameters that only affect the marginal distribution of u_t as shown below.

Proposition 2 : *Drift and scale parameters are not identifiable.*

Proof :

Model (7) may include drift and scale parameters. In such a case, the function g can be reparametrized to get

$$g(\tilde{Y}_t;\theta) = C[\tilde{g}(\tilde{Y}_t;\alpha) - m],$$

with $\theta = (\alpha, m, C)$ and α are the other parameters. Since $Tr\, R^2$ is invariant with respect to affine transformations, m and C are not identifiable. Then, the GCov approach has to be applied to $\tilde{g}(\tilde{Y}_t;\alpha)$ with parameter α corresponding to the identification restriction $m = 0, C = Id$.

This identification issue is addressed later on in a comparison of GCov with the GMM estimator.

4.2 Asymptotic Normality and Semi-Parametric Efficiency

Let us now introduce the following additional assumption:

Assumption A.2 :

i) The true value θ_0 is in the interior $\tilde{\Theta}$ of Θ.

ii) The function $g(\tilde{Y};\theta)$ is twice continuously differentiable with respect to θ, θ in $\bar{\Theta}$.

iii) $g(\tilde{Y};\theta_0)$ and $\frac{\partial g(\tilde{Y};\theta_0)}{\partial \theta'}$ have finite fourth-order moments, $\frac{\partial^2 g_j(Y;\theta_0)}{\partial\theta\partial\theta'}$, $j = 1,...,J$ have finite second-order moments.

The existence of moments, i.e. the tail condition, concerns the errors u_t, not the observed variable Y_t itself, as shown in the stochastic volatility example.

Under Assumptions A.1 and A.2, the GCov estimator satisfies the first-order conditions (FOC):

$$\frac{\partial L_T(\hat{\theta}_T)}{\partial \theta} = 0 \iff \sum_{h=1}^{H} \frac{\partial\, Tr\, \hat{R}^2(h;\hat{\theta}_T)}{\partial \theta} = 0.$$

Appendix A.2.1. provides the expressions of the derivatives. We have:

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

for $j = 1,...,J = dim\theta$ and $\hat{\tilde{R}}^2 = \hat{\Gamma}(0;\hat{\theta}_T)^{-1}\hat{\Gamma}(1;\hat{\theta}_T)\hat{\Gamma}(0;\hat{\theta}_T)^{-1}\hat{\Gamma}(1;\hat{\theta}_T)$.

Remark 3: When $K = 1$, $R^2(h;\theta) = \rho^2(h;\theta)$ is a scalar and the FOC become:

$$\sum_{h=1}^{H} [\hat{\rho}(h;\theta)^2 [\frac{dlog\hat{\gamma}(h;\theta)}{d\theta} - \frac{dlog\hat{\gamma}(0;\theta)}{d\theta}] = 0.$$

For ease of exposition, let us consider H=1. The expansion of the first-order conditions in a neighborhood of θ_0 is:

$$\begin{aligned}\frac{dL_T(\hat{\theta}_T)}{d\theta} &= 0,\\ \text{that implies } \sqrt{T}\frac{dL_T(\theta_0)}{d\theta} &+ \frac{d^2 L_T(\theta_0)}{d\theta d\theta'}\sqrt{T}(\hat{\theta}_T - \theta_0) = o_p(1),\end{aligned}$$

where $o_p(1)$ is negligible in probability. When T tends to infinity, we have:

$$\begin{aligned} \sqrt{T}\frac{dL_T(\theta_0)}{d\theta} &= A(\theta_0)\sqrt{T}vec\hat{\Gamma}(1;\theta_0)' + o_p(1), \\ \text{and} \quad \frac{d^2L_T(\theta_0)}{d\theta d\theta'} &= \frac{d^2L_\infty(\theta_0)}{d\theta' d\theta} \equiv -J(\theta_0) + o_p(1), \end{aligned}$$

where $J(\theta_0)$ and $A(\theta_0)$ are given in Appendix A.2.2. We deduce the following lemma:

Lemma 1: *Under Assumptions A.1, A.2 and $H = 1$ and if $\partial vec\Gamma(1;\theta_0)/\partial\theta'$ is of full column rank, then the GCov estimator converges at speed $1/\sqrt{T}$ and we have:*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = J(\theta_0)^{-1}A(\theta_0)\sqrt{T}vec\hat{\Gamma}(1;\theta_0)' + o_p(1),$$

where $J(\theta_0)$ is invertible (a local identification condition).

Under the same assumptions $\sqrt{T}vec\hat{\Gamma}'(1;\theta_0)$ is asymptotically normally distributed with mean zero. It follows that:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N[0, J(\theta_0)^{-1}A(\theta_0)V_{asy}[\sqrt{T}vec\hat{\Gamma}(1;\theta_0)']A(\theta_0)'J(\theta_0)^{-1}].$$

We get a sandwich expression $J(\theta_0)^{-1}I(\theta_0)J(\theta_0)^{-1}$ of the asymptotic variance-covariance matrix of the GCov estimator, with $I(\theta_0) = A(\theta_0)V_{asy}[\sqrt{T}vec\hat{\Gamma}(1;\theta_0)']A(\theta_0)$. In our framework, the formula of the asymptotic variance-covariance matrix of the GCov estimator can be simplified further (see Appendix A.2.3) because matrices $J(\theta_0)$ and $I(\theta_0)$ are proportional. From Appendix 1, it follows that this result can be extended to any H, by using the fact that the sample autocovariances computed from the i.i.d. errors are such that:

i) The vectors $vec\sqrt{T}\hat{\Gamma}_T(h),\ h = 1, ..., H$ are asymptotically independent.

ii) Their common asymptotic variance is $\Gamma(0) \otimes \Gamma(0)$.

Proposition 3: *Under Assumptions A.1, A.2 and if $Rk\left[\frac{\partial vec\Gamma(1,\theta_0)'}{\partial\theta}, ..., \frac{\partial vec\Gamma(H,\theta_0)'}{\partial\theta}\right] = dim\,\theta$, we have:*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N[0, \Omega(\theta_0)^{-1}],$$

where

$$\Omega(\theta_0) = \sum_{h=1}^{H}[\frac{\partial vec\Gamma(h,\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]\frac{\partial vec\Gamma(h,\theta_0)}{\partial\theta'}].$$

The rank condition ensures that the H first autocovariances are locally fully informative about θ_0. It implies the invertibility of $\Omega(\theta_0)$ and is a local identification condition. Proposition 3 is in particular valid for $K = 1$.

Corollary 1: *In the univariate framework with* $K = 1$*, we get:*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N\left\{0, \gamma(0;\theta_0)^2[\sum_{h=1}^{H} \frac{\partial\gamma(h;\theta_0)}{\partial\theta}\frac{\partial\gamma(h;\theta_0)}{\partial\theta'}]^{-1}\right\}.$$

The simplification in the sandwich formula due to matrix $I(\theta_0)$ being proportional to $J(\theta_0)$ with the proportionality factor given in the proof in Appendix A.2.3, reveals the semi-parametric efficiency of the GCov estimator and justifies the term "Generalized". This semi-parametric efficiency is reached by the GCov estimator in a single optimization and is a consequence of the adequate choice of weights $\hat{\Gamma}(0)^{-1}$ in the objective function [see the asymptotic behavior of $\hat{\Gamma}(h)^{-1}$ in Appendix 1 and the discussion in Section 4.3.1]. More precisely, it would have been possible to define the weighted autocovariance estimators, Cov estimators, say, obtained by minimizing the objective function such as:

$$[vec\hat{\Gamma}(1,\theta), ..., vec\hat{\Gamma}(h,\theta)]'W[vec\hat{\Gamma}(1,\theta)', ..., vec\hat{\Gamma}(h,\theta)']',$$

where W is a symmetric positive semi-definite weighting matrix. The GCov estimator has the smallest asymptotic variance-covariance matrix in the class of such Cov estimators.

4.3 Comparison with GMM Approaches

4.3.1 Computational Issues

The GCov estimator is a continuously updating estimator based on **central moments**:

$$Vec\hat{\Gamma}(h), h = 1, ...H,$$

since the vectors $\sqrt{T}\text{Vec}\hat{\Gamma}(h), h = 1, ...H$ are asymptotically independent with the same asymptotic variance-covariance matrix $\Gamma(0) \otimes \Gamma(0)$, where $\otimes$ denotes the Kronecker product of matrices (see Appendix 1).

Let us now compare the GCov generic approach to the generic Continuously Updating GMM (CUGMM). To clarify the differences between these two approaches, let us consider the following model:

$$y_t = a(y_{t-1};\theta) + b(y_{t-1};\theta)u_t,$$

where errors $u_t, t = 1, ..., T$ are i.i.d.. The parameter θ is assumed to characterize the serial dependence. Under this specification, the errors are non-central and not standardized.

This dynamic model can be rewritten as:

$$y_t = a(y_{t-1};\theta) + mb(y_{t-1};\theta) + \sigma b(y_{t-1};\theta)v_t, \tag{12}$$

where errors $v_t, t = 1, ..., T$ are now i.i.d. with $E(v_t) = 0, Var(v_t) = 1$. In the above specification (12), additional parameters m, σ have been introduced in order to formulate the non-central moment conditions.

Let us now consider some transformations of u_t such as u_t and u_t^2, say. Then, the central moments in the GCov estimator of the first specification are: $cov(u_t, u_{t-h}), cov(u_t^2, u_{t-h}), cov(u_t, u_{t-h}^2)$, $cov(u_t^2, u_{t-h}^2)$ for $h = 1, ..., H$. These moments involve the following non-central moments of errors v_t: $E(v_t), E(v_t^2), E(v_t v_{t-h}), E(v_t^2 v_{t-h}), E(v_t v_{t-h}^2)$, $E(v_t^2 v_{t-h}^2)$.

The generic CUGMM estimator depends on the semi-parametric specification, the set of selected moments, and the selected estimator of the optimal weighting matrix. In our framework, it is natural to apply it to the semiparametric model (12) with the extended parameter vector $\beta = (m, \sigma, \theta')'$, the non-central moments $\phi(\tilde{y}_t;\beta)$ with $\tilde{y}_t = (y_t, y_{t-1})$ corresponding to the 6 moments in v_t given above, where v_t is replaced by $v_t(\beta)$:

$$v_t(\beta) = \frac{y_t - a(y_{t-1};\theta) - mb(y_{t-1};\theta)}{\sigma b(y_{t-1};\theta)}.$$

Hence, the CUGMM estimator is the minimizer of:

$$[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\phi(\tilde{y}_t;\beta)]'[\widehat{Var}(\beta)]^{-1}[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\phi(\tilde{y}_t;\beta)],$$

where $\widehat{Var}(\beta)$ is a consistent estimator of the asymptotic variance of $[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi(\tilde{y}_t;\beta)]$.

In brief, the differences between the generic GCov and CUGMM estimators are the following:

a) The GCov estimator is based on central moments instead of non-central moments.

b) The GCov estimator focuses on the subset of parameters characterizing the (nonlinear) serial dependence, whereas additional parameters have to be introduced before applying an associated CUGMM estimator.

c) The objective function of GCov requires the inversion of matrix $\Gamma(0)$ of dimension $(p \times p)$, where p is the number of nonlinear transformations of u_t (p=2 in our example). In comparison, the objective function of CUGMM requires the inversion of a matrix of dimension $p + Hp(p+1)$ (the first p for the marginal expectations and p(p+1) for the cross-expectation at lag h, h=1,...,H). For example, for p=4, H=5, the GCov requires inverting a (4×4) matrix, while the CUGMM requires estimating and inverting a (104×104) matrix, which is quite large and may yield imprecise results [Altonji, Segal (1996)].

A common feature of the GCOV and CUGMM is that both estimators are invariant to bijective reparametrization. In the special case of linear dynamic models, Velasco and Lobato (2018) (VL) introduce a two-step minimum distance approach, which is a two-step GMM-type estimator based on the third and fourth cumulant spectral density for univariate models. Their moment estimator is not optimally weighted in the first step and its asymptotic variance has a sandwich formula [see e.g. Theorem 3 in VL (2018)]. The semi-parametric efficiency is reached in the second-step when the optimal weights are estimated. In the first step, this estimator requires a computation of an objective function with multiple large sums [see e.g. eq. 12 in VL (2018) and the definition of $\sum_{j=1}^{T-1}$, page 564]. The second step involves the inversion of a large matrix, especially in a multivariate case [see Theorem 4 and eq. 23 in VL (2018)]. An extension introduced in Velasco (2022) is based on the notion of generalized spectral densities [Hong (1999)] and developed in a univariate linear dynamic framework. The asymptotic variance of the first step estimator given in Theorems 2 and 5 shows that the semi-parametric efficiency is not achieved. In general, the finite sample properties of such two-step estimators tend to be very sensitive to the choice of estimated optimal weights used in the second step. Also, unlike the GCov and CUGMM, this estimator is not invariant to bijective reparametrization.

By focusing on serial dependence parameters and using central moments in each transformation, we gain a significant simplification of the expression of the objective function. We also obtain estimates which require the inversion of matrices of smaller dimension than in the CUGMM estimator. It is also important to clarify that the GCov estimator of θ is not a CUGMM estimator concentrated with respect to m, σ.

4.3.2 Instrumental Variables

Another difference between the generic GMM and GCov estimators is the choice of instruments. In the generic GMM approach based on the conditional moments, the instrumental variables are observed. In our framework of Section 4.3.1, the underlying instruments are lagged $v_t(\beta_0)'$s, which are unobserved and depend on the unknown parameter values. These instruments are estimated jointly with the parameters of the model by the GCov estimator.

Can we always find observable instrumental variables for the dynamic models used in practice? To find out, let us consider the mixed causal-noncausal models for macroeconomic applications (Lanne and Saikkonen (2011) report that mixed causal and noncausal dynamics were evidenced in 242 out of 343 macroeconomic and financial time series). No standard method of moments with observed instruments is available in the literature on mixed causal-noncausal models. The reason is that error u_t in model (7) cannot be interpreted as an innovation (either causal, or noncausal) of the observed process (y_t) [see Gourieroux, Jasiak (2022) for innovations and impulse response functions in a mixed VAR model]. Therefore, there does not exist a known function of the observed trajectory of (y_t) that could be used as an instrumental variable. Hence, there is no possibility of applying the standard GMM estimator based on the conditional moments (as shown in Lanne, Saikkonen (2011) the use of lagged values of y_t as instruments in this framework leads to inconsistent GMM and 2SLS estimators). This explains why these models have been commonly estimated by the maximum likelihood in a parametric framework under the risk of misspecification, or by a covariance-based approaches in a semi-parametric framework [see Gourieroux, Jasiak (2017) for mixed multivariate causal-noncausal models], or in the frequency domain by calibration based on cumulant spectral density [Velasco, Lobato (2018), Velasco (2022)].

5 Residual-Based Multivariate Portmanteau Statistic

The asymptotic expansions in Section 4 can be used to derive the asymptotic properties of the residual-based multivariate portmanteau test statistic when θ is replaced by the GCov estimator. The test statistic is:

$$\hat{\xi}_T(H) = TL_T(\hat{\theta}_T).$$

Let us consider the second-order expansion of $L_T(\theta_0)$, when θ_0 is close to $\hat{\theta}_T$. As $\frac{dL_T(\hat{\theta}_T)}{d\theta} = 0$, we get:

$$L_T(\theta_0) = L_T(\hat{\theta}_T) + \frac{1}{2}(\hat{\theta}_T - \theta_0)' \frac{d^2 L_T(\hat{\theta}_T)}{d\theta d\theta'} (\hat{\theta}_T - \theta_0) + o_p(1),$$

and

$$TL_T(\hat{\theta}_T) = TL_T(\theta_0) - \frac{1}{2}\sqrt{T}(\hat{\theta}_T - \theta_0)' \frac{d^2 L_\infty(\theta_0)}{d\theta d\theta'} \sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1).$$

The closed-form expansions of $TL_T(\theta_0)$ and $TL_T(\hat{\theta}_T)$ are derived in Appendix 3 (Supplemental Material). We have:

$$TL_T(\hat{\theta}_T) = vec\,[\sqrt{T}\hat{\Gamma}(1;\theta_0)']'\,\Pi\,vec\,[\sqrt{T}\hat{\Gamma}(1;\theta_0)'] + o_p(1),$$

where the expression of Π is given in Appendix 3, eq. (a.12). As matrix Π is positive definite, we can apply R.201 in Gourieroux, Monfort (1995). It follows that if $\Pi V_{asy}[\sqrt{T}vec\hat{\Gamma}(1;\theta_0)']\Pi = \Pi$, then statistic $TL_T(\hat{\theta})$ follows asymptotically a χ^2 distribution with the degrees of freedom equal to $K^2H - dim\theta$. We check the validity of this condition in Appendix 3. as well as the rank of matrix Π. We get the following result:

Proposition 4: *Under the Assumptions of Proposition 3, the statistic $TL_T(\hat{\theta}_T)$ follows asymptotically the chi-square distribution:* $\chi^2(K^2H - dim\theta)$.

To clarify the proof in Appendix 3, let us describe more precisely the case $K = 1$ and any H. The expansion of $TL_T(\hat{\theta}_T)$ becomes:

$$TL_T(\hat{\theta}_T) \sim [\sqrt{T}\hat{\gamma}(\theta_0)]' \frac{Id - P}{\gamma(0,\theta_0)^2} [\sqrt{T}\hat{\gamma}(\theta_0)],$$

with $\hat{\gamma}(\theta_0) = [\hat{\gamma}(1;\theta_0), ..., \hat{\gamma}(H;\theta_0)]'$ and $P = Z(Z'Z)^{-1}Z'$, where $Z = \partial\gamma(\theta_0)/\partial\theta'$. As matrix P is an orthogonal projector and $V[\sqrt{T}\hat{\gamma}(\theta_0)] = \gamma(0;\theta_0)^2 Id \equiv \Sigma$, we find that: $\Pi\Sigma\Pi = \Pi$. Therefore, we obtain the following corollary:

Corollary 2: *For $K = 1$, the residual-based portmanteau statistic asymptotically follows a chi-square distribution with the degrees of freedom equal to H less the rank of the matrix:*

$$\frac{\partial\gamma(\theta_0)'}{\partial\theta} = \left[\frac{\partial\gamma(1;\theta_0)}{\partial\theta}, ..., \frac{\partial\gamma(H;\theta_0)}{\partial\theta}\right]$$

When the Jacobian is of full column rank: $rk\frac{\partial\gamma(\theta_0)}{\partial\theta'} = dim\theta$, the adjustment is equal to the number of estimated parameters. This case arises when θ is identifiable by the CGov approach. This has been implicitly assumed when writing the inverse of matrix $J(\theta_0)$.

The result in Proposition 4 is well-established for the unconstrained (causal) VAR model, when the autoregressive parameters are estimated by the (unconstrained) OLS [see Chitturi (1974), Hosking (1980), Li, McLeod (1981) for the multivariate framework]. We have extended this result to a larger class of nonlinear causal/ noncausal dynamic models. It is a consequence of the appropriate choice of the objective function and estimation method as well as, indirectly, the semi-parametric efficiency of the GCov estimator discussed in Section 4.

i) If model (7) is causal and the GCov estimator of θ is replaced by another estimator such as a quasi-maximum likelihood (QML) estimator based on the (pseudo) student distribution of errors u_t or a nonlinear least squares estimator, the sandwich formula will not simplify and the associated residual-based portmanteau statistic will not be asymptotically chi-square distributed. For example, it may follow a mixture of chi-square distributions [see, Francq, Roy, Zakoian(2005), Theorem 3, for a special case].

ii) Some modifications of the objective function $L_T(\theta)$ have been proposed in the literature. However, the estimators obtained by maximizing these modified objective functions are not semi-parametrically efficient. For example, for a univariate y_t, Pena, Rodriguez (2002), (2006), Lin, McLeod (2006) replace $\sum_{h=1}^{H} Tr\,[\hat{R}^2(h)] = \sum_{h=1}^{H} \hat{\rho}^2(h)$, by the log-determinant of the Toeplitz matrix:

$$log\,det \begin{bmatrix} 1 & \hat{\rho}(1) & \cdots & & \hat{\rho}(H) \\ \hat{\rho}(1) & \ddots & & & \vdots \\ \vdots & & & & \\ & & & & \hat{\rho}(1) \\ \hat{\rho}(H) & & & \hat{\rho}(1) & 1 \end{bmatrix}.$$

An extension to a multivariate framework has been considered in Mahdi, McLeod (2012) [see also Fisher, Gallagher (2012)]. By changing the objective function $L_T(\theta)$, one can change the asymptotic distribution of the residual-based portmanteau statistic. This is also true for objective functions such as $\sum_{h=1}^{H} Tr\,[\Gamma(h)\Gamma(h)']$ [see Lam, Yao (2012)], and

$$\sum_{h=1}^{H} Tr[[diag\,\gamma(0)]^{-1}\Gamma(h)[diag\,\gamma(0)]^{-1}\Gamma(h)'],$$

where $diag\,\gamma(0)$ is the diagonal matrix with terms $\gamma_{jj}(0)$ on the main diagonal [see, Gourieroux, Jasiak (2017), Forrester, Zhang (2020)].

iii) By selecting an appropriate GCov estimator, we also avoid the adjustment of the portmanteau statistic by recursive projections of estimated autocorrelations that can be numerically demanding in the multivariate framework [see, Velasco, Wang (2015) in a special case of univariate framework].

6 Serially and Cross-Sectionally Independent Errors

This Section discusses the GCov estimator applied to model (7) under an additional assumption of cross-sectional independence of errors.

6.1 Model with Cross-Sectionally Independent Errors

Let us consider model (7) with cross-sectionally independent errors:

$$g_j(\tilde{Y}_t;\theta) = u_{j,t}, \quad j = 1, ..., J, \tag{13}$$

where errors $u_{j,t}, t = 1, ..., T$ are serially independent and $(u_{j,t}),\ j = 1, ..., J$ are independent of each other. Model (13) is a more structural version of model (7) due to the additional cross-sectional independence.

Example 3: Structural VAR (SVAR) Model

Model (13) can represent a causal VAR model written as:

$$g(\tilde{Y}_t; \Phi_0, ..., \Phi_p) = \Phi_0 Y_t - \Phi_1 Y_{t-1} - \cdots - \Phi_p Y_{t-p} = u_t,$$

with independent $u_{j,t}$, i.e. by including the simultaneity characteristics on the right hand side of the system. We assume that Φ_0 is invertible and the roots of $\det(\Phi_0 - \Phi_1 z .. - \Phi_p z^p) = 0$ are not on the unit circle.

If the error vector is Gaussian, parameters $\Phi_i, i = 0, ..., p$ are not identifiable or, more precisely, are identifiable up to a unitary matrix. This leads to additional structural identifying restrictions, such as long-run restrictions, zero restrictions, or sign restrictions proposed in the growing literature. However, this identification problem is entirely solved without additional restrictions if at most one component $(u_{j,t})$ is Gaussian [Comon (1994), Gourieroux, Monfort, Renne (2017) and the references therein].

While a Gaussian QML estimator cannot be used to estimate consistently all the parameters due to the lack of identification under Gaussianity, the covariance-based approaches can be implemented by exploiting the cross-sectional independence [see e.g. Gourieroux, Monfort, Renne (2017), (2020), Lanne, Luoto (2021)] and cross-moments of nonlinear transformations of the components $u_{j,t}$.

6.2 The GCov Estimator under Serial and Cross-Sectional Independence

Let the matrix of sample covariances between $g_i(\tilde{Y}_t;\theta)$ and $g_j(\tilde{Y}_{t-1};\theta)$ be denoted by $\hat{\Gamma}_{ij}(h\ \theta)$. Then, the GCov estimator can be defined as:

$$\hat{\theta}_T(H) = Argmin_\theta \sum_i \sum_j \sum_{h=0}^{H} Tr[\hat{R}^2_{ij}(h;\theta)], \tag{14}$$

where

$$\hat{R}^2_{ij}(h;\theta) = \hat{\Gamma}_{ij}(h;\theta)\hat{\Gamma}_{jj}(0;\theta)^{-1}\hat{\Gamma}_{ji}(h;\theta)'\hat{\Gamma}_{ii}(0;\theta). \tag{15}$$

The terms $\hat{R}^2_{ii}(0,\theta)$ need to be removed from the sum because $\hat{R}^2_{ii}(0,\theta) = Id$ does not depend on θ.

Then, the results of the previous sections are easily extended, yielding the following Proposition, which is the analogue of Proposition 3.

Proposition 5: *Under the Assumptions of Proposition 3 and if processes* $(u_{j,t})$, $j = 1, ..., J$ *are independent:*

i) The GCov estimator $\hat{\theta}_T(H)$ *defined in (14)-(15) is consistent of* θ_0.

ii) $\sqrt{T}(\hat{\theta}_T(H) - \theta_0) \sim N[0, \Omega(\theta_0)^{-1}]$,

with $\Omega(\theta_0) = \sum_i \sum_j \sum_{h=0}^{H} \frac{\partial vec\Gamma_{ij}(h;\theta_0)'}{\partial\theta}[\Gamma_{ii}(0;\theta_0)^{-1} \otimes \Gamma_{jj}(0;\theta_0)^{-1}]\frac{\partial vec\Gamma_{ij}(h;\theta_0)}{\partial\theta'}$.

iv) The estimator is semi-parametrically efficient.

In an application to a SVAR model, Gourieroux, Monfort and Renne (2017) use a weighted covariance-based estimator with non-optimal weights. Indeed the first-order conditions of the constrained Pseudo-Maximum Likelihood are covariance conditions (See Proposition 1 in Gourieroux, Monfort, Renne (2017)). Also, Guay (2021) and Lanne, Luoto (2021) apply the moment methods based on the cross moments of second, third, and fourth orders without centering these moments, which causes the computational issues mentioned in Section 4.3.1. In contrast, the centering of moments is required for the GCov estimation.

Proposition 5 shows that the semi-parametric efficiency is reached with a closed-form objective function. The objective function to be minimized can be decomposed into two components. The first component includes the terms with $h = 0$, and focuses on the cross-sectional dependence of errors. The second component includes the sum of terms with $h = 1, ..., H$ and captures their serial dependence. That second component corresponds to the objective function of Section 5 since, under the independence of processes $(u_{j,t})$, $\Gamma(0, \theta) = diag(\Gamma_{ii}(0; \theta))$.

From the computational perspectives, the matrices to be inverted in the objective function appearing in (14)-(15) are of the same dimensions as those in the model without cross-sectional independence, and depend only on the number of transformations of each u_j.

Example 3: (continued:)

In the application to SVAR models, there is no direct relationship between the second component of the objective function evaluated at the solution of (14) and the residual-based portmanteau statistic of Section 5. The identifiable subsets of parameters of the SVAR model differ in these two optimizations, being $\Phi_0^{-1}\Phi_1, ..., \Phi_0^{-1}\Phi_p$ in Section 5, and $\Phi_i,\ i = 0, ..., p$ (up to scalar scale factors) in this Section.

7 Simulation Study and Application to Commodities

To illustrate the relevance and performance of the GCov estimator, we examine an application to commodity prices. The trajectories of commodity prices feature local trends due to speculative bubbles and may not display the global trends observed in other price processes. Because a non-causal model can capture and reproduce the dynamics of bubbles, we estimate a semi-parametric mixed causal-noncausal VAR representation of a bivariate series of commodity prices.

The first part of this Section reports simulation results for the causal-noncausal VAR model estimated in the time domain by the GCov. The second part presents the estimation results. Additional results are available in the Supplemental Material.

7.1 Mixed VAR(1) Model

We consider a bivariate causal-noncausal VAR(1) model $Y_t = \Phi_1 Y_{t-1} + \epsilon_t$ with $g_t(\tilde{Y}_t, \theta) = Y_t - \Phi Y_{t-1}$, and $\epsilon_t, t = 1, ..., T$, an i.i.d. non-Gaussian bivariate process. We study the distributional properties of the GCov estimators when matrix Φ has eigenvalues outside and inside the unit circle. In this simulation study we address: 1) the effects of error distribution by considering t-student distributed errors with the degrees of freedom parameter $\nu = 4$ and $\nu = 6$ characterized by infinite and finite kurtosis respectively, and the Laplace and Uniform distributions; 2) the effects of sample size, by examining the sample sizes T=200 and T=400; 3) the effects of the number of transformations in autocovariance conditions; 4) the effect of lag H.

The matrix Φ has elements: $\phi_{1,1} = 0.9, \phi_{1,2} = -0.3, \phi_{2,1} = 0, \phi_{2,2} = 1.2$ with eigenvalues equal to 0.9 and 1.2. The component $Y_{1,t}$ is a mixture of causal and noncausal dynamics, while component $Y_{2,t}$ is pure noncausal. The simulation of the stationary mixed VAR(1) process is performed according to the method outlined in Gourieroux and Jasiak (2016), (2017). All results are based on 1000 replications and assume identity variance matrix of the errors.

A sample of length 200 of this data generating process (DGP henceforth) with t(4) distributed errors, is displayed in Figure 1 (a path of 1000 realizations is given in Figure 1, Section 3 of Supplemental Material). This DGP violates the assumption of the existence of the 4th order moment but, when the errors follow a fat tailed distribution (e.g. a stable distribution) it is still possible to derive the asymptotic properties of sample autocovariances [Davis, Resnick (1986)] and show the consistency of the GCov estimator. Then, the speed of convergence and asymptotic distribution of the estimator are different than those given in Proposition 3.

Table 1 summarizes the means and variances of GCov estimators computed from 1000 replications based on four power transformations with exponents 1,2,3 and 4 and H=10. We observe that biases and variances of the GCov estimators diminish with the sample size. To examine the effect of the number of transformations, we repeat the experiment with two transformations. Table 2 summarizes the moments of GCov estimators computed from 1000 replications based on power transformations with exponents 1 and 2, and H=10. By comparing Tables 1 and 2, we observe a trade-off between the bias and variance as the estimators based on 4 transformations have smaller biases, but slightly higher variances.

7.2 Application to wheat and corn futures

7.2.1 The series

We consider a bivariate series of 360 daily observations on CBOT closing prices of wheat and corn futures in US Dollars recorded between October 18, 2016 and March 29, 2018 (Data available on https://ca.finance.yahoo.com, wheat futures: ticker ZW=F and corn futures, ticker ZC=F).

The dynamics of both series and their sample densities are displayed in Figure 2. The top panel of Figure 2 shows the two series plotted over time. The series do not display a global trend or other global and long-lasting explosive patterns. Instead, we observe local trends, bubbles, and spikes. The series move in parallel and their spikes often occur simultaneously. The most pronounced among them is a common bubble on July 17, 2017 when the wheat futures exceeded 535 US Dollars. The absence of global explosive patterns associated with the unit root dynamics suggests we can assume stationarity and explore the sample densities of the series. The bottom panels of Figure 2 show the kernel-smoothed density estimators of the wheat and corn series. The wheat futures have a peaked symmetric density with a long right tail. The density of corn futures is asymmetric in the center and has short tails. The standard normality tests reject the null hypothesis of normality in both series. The wheat and corn futures are contemporaneously correlated with the contemporaneous correlation of 0.7285. The basic statistics summarizing their marginal distributions are given in Table 1, Section 4 of Supplemental Material.

We explore the fit of a VAR(1) model estimated by the GCov and OLS estimators to wheat and corn futures. Recall that the GCov estimator can detect any type of autoregressive roots, whereas an OLS estimator by construction provides roots outside the unit circle.

7.2.2 Estimation of VAR(1) Model

The VAR(1) model is fitted to the demeaned bivariate price series $Y_t = (\text{corn}, \text{wheat})'$, t=1,...,360 with a 2×2 matrix Φ of autoregressive coefficients $\phi_{ij}, i, j = 1, 2$, and error ϵ_t assumed to be a bivariate i.i.d. white noise. The parameters are estimated by the GCov estimator using the sample autocovariances of residuals, squared residuals, and powers 3 and 4 of residuals with the lag length H=10. The choice of lag H and implementation of the GCov estimator are illustrated in Section 4.2 (Figure 8) of Supplemental Material.

The parameter estimates along with their estimated standard errors computed from the formula in Proposition 1 are given in the left panel of Table 3. The matrix $\hat{\Phi}$ estimated by the GCov has eigenvalues 0.9063 and 1.1510 located inside and outside the unit circle, respectively. We obtain a mixed (causal-noncausal) autoregressive model. The normalized eigenvectors associated with these eigenvalues are [3.3049, 1] and [0.7441, 1]. The eigenvalue larger than 1 is due to "speculative" bubbles arising simultaneously in the two series. From the second eigenvector, it follows that the portfolio, which is the least sensitive to bubbles, contains the wheat and corn futures with allocations 1 and -0.7441, respectively.

Next, the matrix of autoregressive coefficients is estimated by the OLS, (right panel of Table 3) producing a pure causal autoregressive matrix with close to 1 values of estimated coefficients ϕ_{11} and ϕ_{22} and high t-ratios, typical for unit roots. The statistically non-significant and close to 0 values of $\hat{\phi}_{12}$ and $\hat{\phi}_{21}$ suggest an absence of feedback effects.

These results are spurious and originate from the fact that the OLS estimator provides estimates of $\hat{\Phi}$ restricted to causal dynamics only. Hence, standard inference results based on the OLS estimators need to be interpreted with caution. In our example, the OLS estimators are inconsistent, upper-bounded by 1 but "attracted" by the noncausal eigenvalue, which explains why $\hat{\phi}_{11}$ and $\hat{\phi}_{22}$ are close to 1. The eigenvalues of the OLS estimated matrix $\hat{\Phi}$ are 0.964 and 0.935.

Figure 3 displays the dynamics of residuals $\hat{\epsilon}_t$ of the mixed VAR(1) model estimated by the GCov and their sample densities. The residual series are reasonably close to stationary, except for an episode of increased variance in the middle of the sampling period. The residual sample densities are uni-modal with long tails. The residual variance-covariance matrix is $\hat{\Sigma} = \begin{pmatrix} 61.7948 & 27.4743 \\ 27.4743 & 24.5573 \end{pmatrix}$ and the contemporaneous correlation is 0.705. The joint sample density of residuals is reported in Section 4.3 (Figure 10) of Supplemental Material, where the model diagnostics are also discussed. The mixed VAR(1) model successfully accommodates most of the serial dependence in the data, but there remain some significant residual autocorrelations and cross-correlations. In order to eliminate that remaining serial dependence, we extend the autoregressive order of the model and estimate the mixed VAR(3) model to improve the fit (see, Section 4.4 of Supplemental Material).

8 Concluding Remarks

This paper introduces a semi-parametric estimation approach for a large class of multivariate nonlinear dynamic models with i.i.d. errors. The GCov estimator, obtained by minimizing a multivariate portmanteau criterion, has the property of semi-parametric efficiency, which is achieved in a one-step procedure, and circumvents numerical issues involved in other GMM-type estimators. We have shown that the associated residual-based portmanteau statistic asymptotically follows a chi-square distribution with an adjusted degree of freedom.

Among further extensions are the following ones:

1. The approach is based on the knowledge of the asymptotic behavior of sample autocovariances, which may differ for errors without finite second-order moments. In such a case, the speed of convergence and asymptotic distribution of sample autocovariances and the GCov estimator are modified [see, Davis, Resnick (1986)]. The asymptotic distribution of the (residual-based) portmanteau statistic is altered as well, likely becoming a function of a stable limiting distribution [see e.g. Gourieroux, Zakoian (2017)].

2. A slightly modified GCov approach can also be applied to a model with nonfundamental nonlinear moving average effects as :

$$g(\tilde{Y}_t; \theta) = c(\tilde{u}_t; \beta),$$

where $\tilde{u}_t = (u_t, u_{t-1}, \ldots, u_{t-q})$, say. This specification implies covariance conditions, such as $Cov[g(\tilde{Y}_t; \theta), g(\tilde{Y}_{t-h}; \theta)] = 0$, for $h \geq q$, which can be used to identify the parameter θ, as well as orders L and q. This approach is analogous to the use of Yule-Walker equations for identifying the orders of univariate ARMA processes. It was applied in the first step of the estimation before focusing on parameter β in the second step by Gourieroux, Monfort, Renne (2020).

3. From a theoretical point of view, it would be interesting to view the independence condition as an infinite number of autocovariance conditions. For ease of exposition, let us consider the pairwise independence underlying both the GCov estimator and the approach based on cumulant spectral density [Velasco (2022)]. The condition of $u_t, u_\tau,\ \ t \neq \tau$ is equivalent to the condition $Cov(a(u_t), b(u_{t+h})) = 0$ for any h and pair of square integrable functions $a(.), b(.)$. This infinity is more complex than any analogous condition considered in the literature so far: in the univariate framework, Velasco, Lobato (2018) and Velasco (2022) focus on the lag h dimension (or, equivalently on that frequency) with a finite number of transformations $a(.), b(.)$. The state discretization given in Section 3.3 and Appendix 4 is the first step towards accommodating the multiplicity of $a(.), b(.)$. To our knowledge, the joint analysis of both types of infinity has not yet been done.

Acknowledgments: The authors thank the Associate Editor and two anonymous referees as well as G. Imbens, D. Matteson, A. Hecq, and the participants of CMStatistics 2021 for their helpful comments. The first author acknowledges financial support of the ACPR Chair ”Regulation and Systemic Risk” and the ERC DYSMOIA. The second author acknowledges financial support of the Natural Sciences and Engineering Council of Canada (NSERC).

Conflict of Interest Statement: The authors report there are no competing interests to declare.

Table 1. Effect of sample size and error distribution, 4 transforms

	t(4), kurt $= \infty$				t(6), kurt $< \infty$			
	T=200		T=400		T=200		T=400	
	mean	var	mean	var	mean	var	mean	var
$\hat{\phi}_{11}$	0.927	0.006	0.923	0.003	0.941	0.008	0.934	0.004
$\hat{\phi}_{12}$	-0.291	0.051	-0.270	0.021	-0.285	0.068	-0.264	0.029
$\hat{\phi}_{21}$	0.047	0.009	0.025	0.004	0.063	0.010	0.040	0.005
$\hat{\phi}_{22}$	1.239	0.042	1.200	0.031	1.292	0.084	1.209	0.037
	Laplace(0,1)				Uniform[-1,1]			
	T=200		T=400		T=200		T=400	
	mean	var	mean	var	mean	var	mean	var
$\hat{\phi}_{11}$	0.962	0.007	0.968	0.003	0.952	0.006	0.937	0.002
$\hat{\phi}_{12}$	-0.295	0.058	-0.195	0.027	-0.306	0.0601	-0.230	0.024
$\hat{\phi}_{21}$	0.090	0.009	0.076	0.004	0.091	0.009	0.045	0.003
$\hat{\phi}_{11}$	1.324	0.073	1.245	0.036	1.308	0.069	1.207	0.027

Table 2. Effect of sample size and error distribution, 2 transforms

	t(4), kurt = ∞				t(6), kurt < ∞			
	T=200		T=400		T=200		T=400	
	mean	var	mean	var	mean	var	mean	var
$\hat{\phi}_{11}$	0.940	0.004	0.940	0.003	0.953	0.005	0.950	0.003
$\hat{\phi}_{12}$	-0.251	0.021	-0.254	0.011	-0.243	0.025	-0.243	0.013
$\hat{\phi}_{21}$	0.044	0.006	0.043	0.004	0.059	0.006	0.054	0.004
$\hat{\phi}_{22}$	1.220	0.040	1.224	0.023	1.226	0.044	1.223	0.031
	Laplace(0,1)				Uniform[-1,1]			
	T=200		T=400		T=200		T=400	
	mean	var	mean	var	mean	var	mean	var
$\hat{\phi}_{11}$	0.971	0.005	0.968	0.003	0.968	0.005	0.964	0.0032
$\hat{\phi}_{12}$	-0.213	0.026	-0.209	0.013	-0.219	0.027	-0.207	0.013
$\hat{\phi}_{21}$	0.081	0.006	0.076	0.004	0.082	0.006	0.074	0.003
$\hat{\phi}_{11}$	1.261	0.043	1.249	0.030	1.270	0.044	1.254	0.026

Table 3. Estimation of VAR(1) model

	GCov			OLS		
parameter	estimate	st.err.	t-ratio	estimate	st.err.	t-ratio
ϕ_{11}	0.8351	0.0583	14.3241	0.9750	0.0204	47.590
ϕ_{12}	0.2356	0.1174	2.0068	-0.0326	0.0447	-0.731
ϕ_{21}	-0.0958	0.0469	-2.0426	0.0127	0.0117	1.086
ϕ_{22}	1.2230	0.0787	15.5400	0.9250	0.0255	36.219

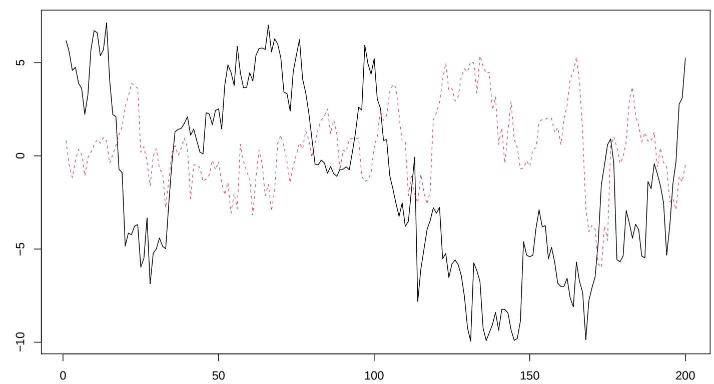
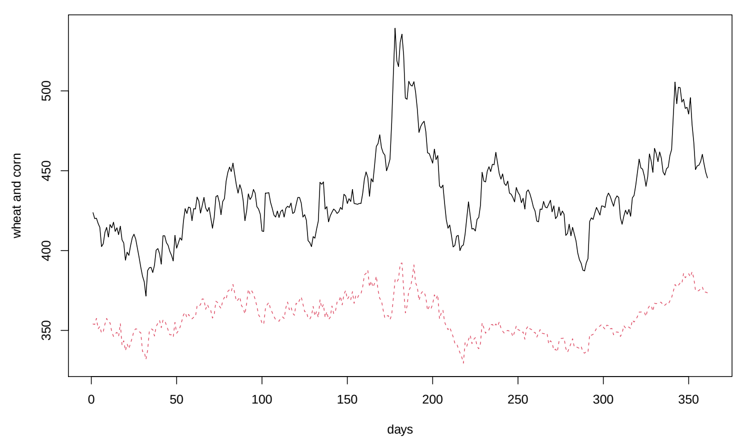
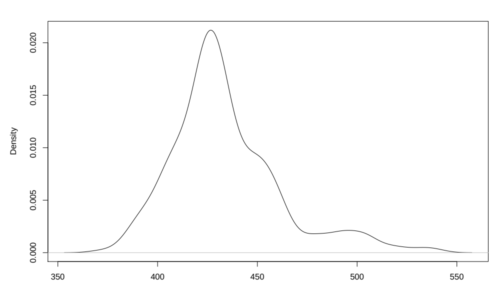


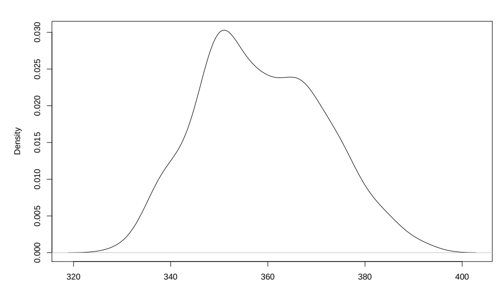
Figure 1: Simulation: DGP, T=200 solid line: $Y_{1,t}$, dashed line: $Y_{1,t}$



(a) wheat (black solid line) and corn (red dashed line)

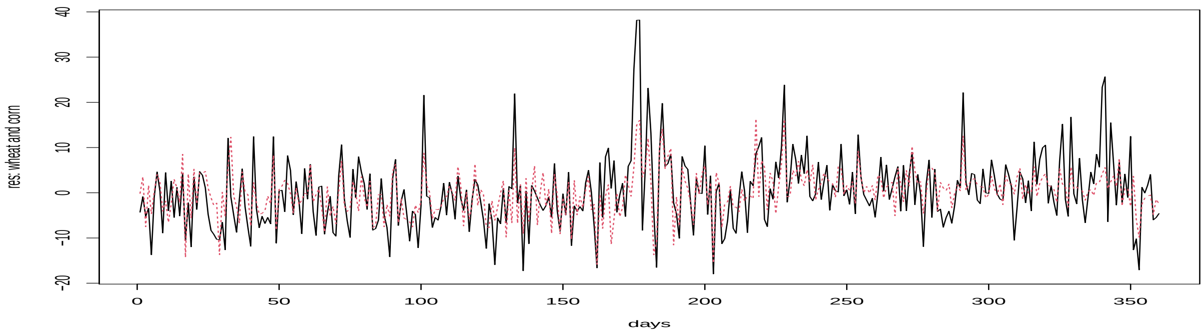


(b) wheat

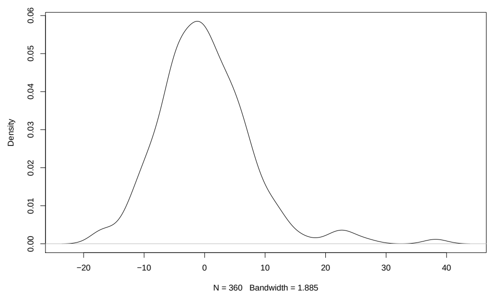


(c) corn

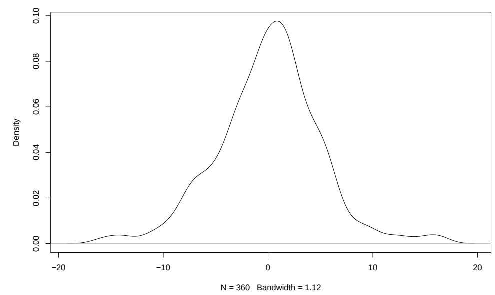
Figure 2: Daily future price series: dynamics and marginal sample densities



(a) residual series: wheat (black solid line) and corn (red dashed line)



(b) residual: wheat



(c) residual: corn

Figure 3: Residuals: dynamics and marginal sample densities

REFERENCES

Altonji, J. and L. Segal (1996): "Small Sample Bias in GMM Estimation of Covariance Structures", Journal of Business and Economic Statistics, 14, 353-366.

Anderson, T. (1999): "Asymptotic Theory for Canonical Correlation Analysis", Journal of Multivariate Analysis, 70, 1-29.

Anderson, T. (2002): "Canonical Correlation Analysis and Reduced Rank Regression in Autoregressive Models", Annals of Statistics, 30, 1134-1154.

Andrews, D. (1992): "Generic Uniform Convergence", Econometric Theory, 8, 241-257.

Bartlett, H. (1946): "On the Theoretical Specifications and Sampling Properties of Autocorrelated Time Series", Supplement to JRSS, 8, 27-41.

Box, G. and D. Pierce (1970): "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", JASA, 65, 1509-1526.

Chitturi, R. (1974): "Distribution of Residual Autocorrelations in Multiple Autoregressive Schemes", JASA, 69, 928-934.

Chitturi, R. (1976): "Distribution of Multivariate White Noise Autocorrelations", JASA, 71, 223-226.

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

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

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

Fisher, T. and C. Gallagher (2012): "New Weighted Portmanteau Statistic for Time Series Goodness of Fit Testing", JASA, 107, 777-787.

Forrester, P. and J. Zhang (2020): "Parametrizing Correlation Matrices", Journal of Multivariate Analysis, 178, 104619.

Francq, C., Roy, R. and J.M. Zakoian (2005): "Diagnostic Checking in ARMA Models with Uncorrelated Errors", JASA, 100, 532-544.

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

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

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

Gourieroux, C., and J., Jasiak (2022) : "Nonlinear Forecasts and Impulse Responses for Causal-Noncausal (S)VAR Models", ArXiv 2205.09922.

Gourieroux, C. and A. Monfort (1995): " Statistics and Econometric Models", Vol 2, Cambridge Univ. Press.

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

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

Gourieroux, C. and J.M. Zakoian (2017): "Local Explosion Modelling by Non-Causal Process", JRSS B, 79, 737-756.

Guay, A. (2021): "Identification of Structural Vector Autoregressions Through Higher Unconditional Moments:, Journal of Econometrics, 225, 27-46.

Hansen, L., Heaton , J. and A. Yaron (1996): "Finite Sample Properties of Some Alternative GMM Estimators", Journal of Business and Economic Statistics, 14, 262-280.

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

Hoga, Y. (2021): "Testing for Serial Extreme Dependence in Time Series Residuals", D.P. University of Duisburg-Essen.

Hong, Y. (1999): "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach", JASA, 94, 1201-1220.

Hosking, J. (1980): "The Multivariate Portmanteau Statistic", JASA, 75, 602-608.

Hotelling, H. (1936) : "Relation Between Two Sets of Variants", Biometrika, 28, 321-377.

Lam, C. and Q. Yao (2012): "Factor Modelling for High Dimensional Time Series", Biometrika, 98, 901-918.

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

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

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

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

Li, W. and A. McLeod (1981): "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models", JRSS B, 43, 231-233.

Li, W. and T. Mak (1994): "On the Squared Residual Autocorrelations in Nonlinear Time Series with Conditional Heteroscedasticity", JTSA, 15, 627-639.

Lin, J. and A. McLeod (2006): "Improved Pena-Rodriguez Portmanteau Test", Computational Statistics and Data Analysis, 51, 1731-1738.

Ling, S. (2004): "Estimation and Testing Stationarity for Double-Autoregressive Models", J.R.S.S., Series B, 66, 63-78.

Mahdi, E. and I. McLeod (2012): "Improved Multivariate Portmanteau Test", Journal of Time Series Analysis", 23, 211-222.

Mann, H. and A. Wald (1943): "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica, 11, 173-220.

Nelson, D. (1990): "ARCH Models as Diffusion Approximations", Journal of Econometrics, 45, 7-38.

Pena, D. and J. Rodriguez (2002): "A Powerful Portmanteau Test of Lack of Fit for Time Series", JASA, 97, 601-610.

Pena, D. and J. Rodriguez (2006): "The Log of the Determinant of the Autocorrelation Matrix for Testing Goodness of Fit in Time Series", Journal of Statistical Planning and Inference, 136, 2706-2718.

Robinson, P. (1973): " Generalized Canonical Analysis for Time Series", Journal of Multivariate Analysis, 3, 141-160.

Sims, C. (2021): "SVAR Identification through Heteroscedasticity with Misspecified Regimes", Princeton, DP.

Velasco, C. (2020) : "Identification and Estimation of Structural VARMA Models Using Higher Order Dynamics", arXiv : 2009.04428v.

Velasco, C. (2022): "Estimation of Time Series Models Using Residual Dependence Measures", Annals of Statistics, forthcoming.

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

Velasco, C., and X. Wang (2015): "A Joint Portmanteau Test for Conditional Mean and Variance Models", Journal of Time Series Analysis, 36, 39-60.

Wooldridge, J. (1991): "On the Application of Robust Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances", Journal of Econometrics, 47, 5-46.