Robust Optimal Tests for Causality in Multivariate Time Series *

Abdessamad Saidi† and Roch Roy‡

Abstract

Here, we derive optimal rank-based tests for noncausality in the sense of Granger between two multivariate time series. Assuming that the global process admits a joint stationary vector autoregressive (VAR) representation with an elliptically symmetric innovation density, both no feedback and one direction causality hypotheses are tested. Using the characterization of noncausality in the VAR context, the local asymptotic normality (LAN) theory described in Le Cam (1986)) allows for constructing locally and asymptotically optimal tests for the null hypothesis of noncausality in one or both directions. These tests are based on multivariate residual ranks and signs (Hallin and Paindaveine, 2004a) and are shown to be asymptotically distribution free under elliptically symmetric innovation densities and invariant with respect to some affine transformations. Local powers and asymptotic relative efficiencies are also derived. Finally, the level, power and robustness (to outliers) of the resulting tests are studied by simulation and are compared to those of Wald test.

KEY WORDS: Granger causality, Elliptical density, Local asymptotic normality, Multivariate autoregressive moving average model, Multivariate ranks and signs, Robustness.

This version: 17 May, 2006

1 Introduction

The concept of causality introduced by Wiener (1956) and Granger (1969) is now a fundamental notion for analyzing dynamic relationships between subsets of the variables of interest. There is a substantial literature on this topic; see for example the reviews of Pierce and Haugh (1977), Newbold (1982), Geweke (1984), Gourriroux and Monfort (1990, Chapter X) and Lütkepohl (1991). The idea behind this concept is that, if a variable X affects a variable Y, the former should help improving the predictions of the latter variable. A formal definition is presented in Section 2. The original

*This work was partially supported by grants from the Natural Science and Engineering Research Council of Canada (NSERC), the Network of Centres of Excellence on The Mathematics of Information Technology and Complex Systems (MITACS) and the Fonds québecois de la recherche sur la nature et les technologies (FQRNT).

†Département de mathématiques et de statistique, Université de Montréal, CP 6128, succursale Centre-ville, Montréal, Québec, H3C 3J7, Canada (e-mail: saidiabd@dms.umontreal.ca).

‡Département de mathématiques et de statistique and Centre de recherches mathématiques, CP 6128, succursale Centre-ville, Montréal, Québec, H3C 3J7, Canada (e-mail: royr@dms.umontreal.ca).

definition of Granger (1969) refers to the predictability of a variable X, one period ahead. It is also called causality in mean. It was extended to vectors of variables, see for example Tjostheim (1981), Lütkepohl (1991), Boudjellaba, Dufour and Roy (1992, 1994). Lütkepohl (1993), Dufour and Renault (1998) proposed definitions of noncausality in terms of nonpredictability at any number of periods ahead.

In causality analysis, there are two main questions. Firstly, the characterization of noncausality in terms of the parameters of the fitted model to the observed series. Secondly, the development of a valid inference theory for the chosen class of models. In the stationary case, necessary and sufficient conditions for noncausality between two vectors are given, for example, in Lütkepohl (1991, Chapter 2) for vector autoregressive (VAR) models, and by Boudjellaba, Dufour and Roy (1992, 1994) for vector autoregressive moving average (VARMA) models. Characterization of noncausality and inference in possibly cointegrated autoregressions were studied, among others, by Dufour and Renault (1998). For testing causality, the classical test criteria (likelihood ratio, scores, Wald) are generally used, see for example Taylor (1989). With finite autoregressions, the necessary and sufficient conditions for noncausality reduce to zero restrictions on the parameters of the model and the asymptotic chi-square distribution of these classical test statistics remains valid in the stationary case. However, with cointegrated systems, these statistics may follow nonstandard asymptotic distributions involving nuisance parameters, see among others Sims, Stock and Watson (1990), Phillips (1991), Toda and Phillips (1993, 1994), Dolado and Lütkepohl (1996), Dufour, Pelletier and Renault (2005).

The purpose of this paper is to investigate the problem of Granger causality testing via the Le Cam Local Asymptotic Normality (LAN) theory (Le Cam, 1986), and to propose nonparametric (the density of the noise is unknown) and optimal (in the Le Cam sense) procedures for testing causality between two multivariate (or univariate) time series $\mathbf{X}_t^{(1)}$ and $\mathbf{X}_t^{(2)}$. The global process $\mathbf{X}_t = ((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T)^T$, (the superscript ^{T} indicates transpose) is assumed to be a stationary VAR(p) process in order to have linear constraints under the null hypothesis of noncausality. The LAN approach, as we shall see, provides parametric optimal tests, that is the tests proposed are valid and are optimal only when the density of the noise is correctly specified. However, rank-based versions of the central sequence related to the LAN approach will be obtained and a new class of tests depending on a score function will be proposed. These new tests are based on multivariate residual ranks and signs and are shown to be asymptotically distribution free under elliptically symmetric innovation densities and invariant with respect to some affine transformations. Moreover, the optimality property is preserved when the score function used is correctly specified. At our knowledge, nobody has yet

taken advantage of the LAN approach for deriving the asymptotic properties of rank-based statistics for testing causality.

LAN for linear time series models was established in the univariate AR case with linear trend by Swensen (1985), in the ARMA case by Kreiss (1987); a multivariate version of these results was given by Garel and Hallin (1995). Still in the univariate case, a more general approach, allowing for nonlinearities, was taken in Hwang and Basawa (1993), Drost, Klaassen, and Werker (1997), Koul and Schick (1996, 1997); see Taniguchi and Kakizawa (2000) for a survey of LAN for time series. The LAN result we need here is a particular case of Garel and Hallin (1995) established in the general context of VARMA models with possibly nonelliptical noise.

Rank-based methods for a long time have been essentially limited to statistical models involving univariate independent observations, a theory which is essentially complete. In the case of multivariate independent observations, many methods based on different sign and rank concepts were proposed, these works belong to three groups. The first one considers componentwise ranks (Puri and Sen, 1971), however they are not affine-invariant. This was the main motivation for the other two groups. The second group is related to spatial signs and ranks concept; see Oja (1999) for a review. The last one relies on the concept of interdirections developed by Randles (1989) and Peters and Randles (1990). For the multivariate location problem under elliptical symmetry, Hallin and Paindaveine (2002a, 2002b) amalgamate local asymptotic normality and robustness features offered by Peters and Randles (1990)'s signs and ranks. They developed optimal tests based on the concept of interdirections and pseudo-Mahalanobis distances computed with respect to an estimator of the scatter matrix.

The statistical theory of rank tests for univariate stationary time series analysis has a long history, see Hallin and Puri (1992) for a review. The first unified framework in this area was taken by Hallin and Puri (1994) where they proposed an optimal rank-based approach to hypothesis testing in the analysis of linear models with ARMA error terms. In the multivariate case, optimal rank-based tests in stationary VARMA time series were developed for two interesting problems: testing multivariate elliptical white noise against VARMA dependence (Hallin and Paindaveine, 2002c) and testing the adequacy of an elliptical VARMA model (Hallin and Paindaveine, 2004a). Hallin and Paindaveine (2005) developed locally asymptotically optimal tests for affine invariant linear hypotheses in the general linear model with VARMA errors under elliptical innovation densities. A characterization of the collection of null hypotheses that are invariant under the group of affine transformations was also given for the general linear model with VARMA errors, (see, Hallin and Paindaveine, 2003). Among other applications of those tests, we mention the Durbin-Watson problem (testing independence against

autocorrelated noise in a linear model) and the problem of testing the order of a VAR model, see Hallin and Paindaveine (2004b). The approach we are adopting in the present paper is in the same spirit. We combine robustness, invariance and optimality concerns. However, the null hypothesis of interest here is not affine invariant. Indeed, the null hypothesis of no feedback in the VAR model is only invariant with respect to the group of block-diagonal-affine transformations and the problem of noncausality directions is invariant under upper or lower block triangular affine transformations depending on the direction to be tested.

Besides their efficiency properties, rank tests enjoy robustness features. Such features are very desirable in the multivariate time series context where outliers are difficult to detect. Outliers in time series can occur for various reasons, measurement errors or equipment failure, etc. (see, e.g., Martin and Yohai, 1985; Rousseeuw and Leroy, 1987; and Tsay, Peña and Pankratz, 2000). They can create serious problems in the determination of causality direction among variables. Clearly, if the causality inference is erroneous, the forecasting errors may be seriously inflated and their interpretation may be misleading.

The paper is organized as follows. In Section 2, we first recall the characterization of Granger noncausality in VAR models. After having presented some technical assumptions on the elliptical density, the LAN property in stationary VAR models under an elliptical density f is established. In Section 3, we derive the locally asymptotically most stringent test for testing causality between two multivariate time series. The form of this test regrettably implies that its validity is in general limited to the innovation density f for which it is optimal. This density being unspecified in applications, such tests are of little practical interest. The Gaussian case, is a remarkable exception; Gaussian parametric tests are valid irrespective of the true underlying density. When the density is non-Gaussian, the corresponding test is then call "pseudo-Gaussian". Section 4 is devoted to the description of our rank-based test statistics, and to the derivation of their asymptotic distributions under both the null hypothesis and a sequence of local alternatives. Their asymptotic relative efficiencies with respect to the pseudo-Gaussian test are also obtained. Since the proofs are rather long and technical, they are relegated to the Appendix.

The particular case of testing for no feedback in the bivariate VAR(1) model is considered in Section 5, where a numerical investigation was conducted to analyze the level, power and robustness of our new tests and also of the Wald test. Two estimators of the noise covariance matrix were employed: the usual residual covariance matrix and the robust estimator proposed by Tyler (1987). Combined with four score functions (constant, Spearman, Laplace and van der Waerden), it leads

to eight different rank-based tests. When there are no outliers, the level of all the tests considered (Wald, pseudo-Gaussian and the eight rank-based tests) is very well controlled with series of length 100 and 200. Under the alternative of causality (in one direction or the other), the Wald and pseudo-Gaussian tests have similar power. In general, the rank-based tests are slightly less powerful but in all the situations considered, there is always a rank-based test which is almost as powerful as Wald and pseudo-Gaussian tests. In the presence of observation or innovation outliers, both Wald and pseudo-Gaussian tests are severely affected. With innovation outliers, the levels of all rank-based tests are very well controlled. However, with observation outliers, the nonparametric tests are still biased. In general, they overreject and the bias is more important when using the empirical covariance matrix estimator.

A word on notation. Boldface throughout denote vectors and matrices; the superscript ^{T} indicates transpose; vec$\mathbf{A}$ as usual stands for the vector resulting from stacking the columns of a matrix $\mathbf{A}$ on top of each other, and $\mathbf{A} \otimes \mathbf{B}$ for the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. For a symmetric positive definite $k \times k$ matrix $\mathbf{P}$, $\mathbf{P}^{\frac{1}{2}}$ is the unique upper-triangular $k \times k$ matrix with positive diagonal elements that satisfies $\mathbf{P} = \left(\mathbf{P}^{\frac{1}{2}}\right)^T \mathbf{P}^{\frac{1}{2}}$. Also, $\mathbf{A} \leq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is non-negative definite.

2 Preliminary results

2.1 Granger-causality in VAR models

Let $\mathbf{X} := \{\mathbf{X}_t = ((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T)^T, t \in \mathbb{Z}\}$ denote a d-variate process partitioned into $\mathbf{X}^{(1)} := \{\mathbf{X}_t^{(1)}, t \in \mathbb{Z}\}$, with values in $\mathbb{R}^{d_1}$, $d_1 \geq 1$, and $\mathbf{X}^{(2)} := \{\mathbf{X}_t^{(2)}, t \in \mathbb{Z}\}$, with values in $\mathbb{R}^{d_2}$, $d_2 \geq 1$, $d_1 + d_2 = d$. Throughout the paper, $\mathbf{X}$ is assumed to be a centered vector autoregressive VAR(p) process, satisfying a stochastic difference equation of the form

$$\mathbf{X}_t - \sum_{j=1}^{p} \mathbf{A}_j \mathbf{X}_{t-j} = \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}, \tag{2.1}$$

where $\mathbf{A}_j$, $j = 1, ..., p$, are $d \times d$ real matrices and $\boldsymbol{\epsilon}_t$ is d-variate white noise process, i.e., a sequence of uncorrelated random vectors with mean zero and with nonsingular covariance matrix.

The partition of $\mathbf{X}$ into $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ induces a partition of the coefficient matrices $\mathbf{A}_j$, $j = 1, ..., p$, into

$$\mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{(11)} & \mathbf{A}_j^{(12)} \\ \mathbf{A}_j^{(21)} & \mathbf{A}_j^{(22)} \end{pmatrix}, \qquad j = 1, ..., p.$$

Denote by

$$\boldsymbol{\theta} := \left(\mathrm{vec}^T \mathbf{A}_1, ..., \mathrm{vec}^T \mathbf{A}_p\right)^T \tag{2.2}$$

the K-dimensional vector of parameters involved in (2.1); note that $K = pd^2$. We assume that the process is causal:

(A1) The roots of the determinant of the autoregressive polynomial associated with (2.1) all lie outside the unit disk, that is,

$$\left| I_d - \sum_{j=1}^{p} \mathbf{A}_j z^j \right| \neq 0, \quad \forall\, |z| \leq 1,\ z \in \mathbb{C}.$$

The subset of parameter values $\boldsymbol{\theta}$ such that Assumption (A1) holds is denoted by $\boldsymbol{\Theta}$. Under Assumption (A1), the autoregressive polynomial is invertible and we write

$$\left(\mathbf{I}_d - \sum_{j=1}^{p} \mathbf{A}_j z^j \right)^{-1} = \sum_{u=0}^{+\infty} \mathbf{G}_u z^u, \forall |z| < 1, z \in \mathbb{C}.$$

The matrix coefficients $\mathbf{G}_u$ are the Green matrices associated with the autoregressive operator and formally, we should write $\mathbf{G}_u(\boldsymbol{\theta})$. However, when there is no possible confusion, we will drop the argument $\boldsymbol{\theta}$ and we will simply write $\mathbf{G}_u$ instead of $\mathbf{G}_u(\boldsymbol{\theta})$.

The definition of causality in the sense of Granger between vectors of variables that we will use here was proposed by Tjostheim (1981). Boudjellaba, Dufour and Roy (1992) present two equivalent formulation of that definition.

Denote by $\mathcal{H}(\mathbf{X}; t)$ the Hilbert space generated by $\{\mathbf{X}_s; s < t\}$. Write

$$\text{Proj}\left(\boldsymbol{\xi} | \mathcal{H}(\mathbf{X}; t)\right) := \left(\text{Proj}\left(\xi_1 | \mathcal{H}(\mathbf{X}; t)\right), \ldots, \text{Proj}\left(\xi_l | \mathcal{H}(\mathbf{X}; t)\right)\right)$$

for the best linear predictor of $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_l)$ based on $\mathcal{H}(\mathbf{X}; t)$ (namely, the orthogonal projection of $\boldsymbol{\xi}$ onto $\mathcal{H}(\mathbf{X}; t)$), and let $\boldsymbol{\Sigma}\left(\boldsymbol{\xi} | \mathcal{H}(\mathbf{X}; t)\right)$ be the covariance matrix of the corresponding prediction error $\boldsymbol{\xi} - \text{Proj}\left(\boldsymbol{\xi} | \mathcal{H}(\mathbf{X}; t)\right)$.

Definition. The process $\mathbf{X}^{(2)}$ *does not Granger cause* $\mathbf{X}^{(1)}$ if $\boldsymbol{\Sigma}(\mathbf{X}^{(1)}(t) | \mathcal{H}(\mathbf{X}; t)) = \boldsymbol{\Sigma}(\mathbf{X}^{(1)}(t) | \mathcal{H}(\mathbf{X}^{(1)}; t))$. Otherwise, $\boldsymbol{\Sigma}(\mathbf{X}^{(1)}(t) | \mathcal{H}(\mathbf{X}; t)) \leq \boldsymbol{\Sigma}(\mathbf{X}^{(1)}(t) | \mathcal{H}(\mathbf{X}^{(1)}; t))$ and we say that $\mathbf{X}^{(2)}$ *Granger causes* $\mathbf{X}^{(1)}$.

If $\mathbf{X}^{(2)}$ *does not cause* $\mathbf{X}^{(1)}$ and if $\mathbf{X}^{(1)}$ *does not cause* $\mathbf{X}^{(2)}$, we say that there is no feedback between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. In the VAR context, the noncausality directions are characterized by linear restrictions on the autoregressive coefficients as described in the following proposition. A proof is given, in Boudjellaba, Dufour and Roy (1992); see also Lütkepohl (1991, Section 2.3).

Proposition 2.1 *Suppose that the VAR(p) process* $\mathbf{X} = \left((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T\right)^T$ *in (2.1) satisfies Assumption* (A1)*, and that the covariance matrix of* $\{\boldsymbol{\epsilon}_t\}$ *is nonsingular. Then*

(i) $\mathbf{X}^{(1)}$ *does not Granger cause* $\mathbf{X}^{(2)}$ *(*$\mathbf{X}^{(1)} \not\Rightarrow \mathbf{X}^{(2)}$*)* *if and only if* $\mathbf{A}_j^{(21)} = \mathbf{0}\ \forall j = 1, ..., p$;

(ii) $\mathbf{X}^{(2)}$ *does not Granger cause* $\mathbf{X}^{(1)}$ *(*$\mathbf{X}^{(2)} \not\Rightarrow \mathbf{X}^{(1)}$*)* *if and only if* $\mathbf{A}_j^{(12)} = \mathbf{0}\ \forall j = 1, ..., p$;

(iii) there is no feedback or noncausality between $\mathbf{X}^{(1)}$ *and* $\mathbf{X}^{(2)}$ *(*$\mathbf{X}^{(2)} \not\Leftrightarrow \mathbf{X}^{(1)}$*)* *if and only if* $\mathbf{A}_j^{(12)} = \mathbf{A}_j^{(21)} = \mathbf{0}\ \forall j = 1, ..., p$.

Noncausality (in one direction or in both directions) between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ reduces to the hypothesis that the parameter $\boldsymbol{\theta}$ in (2.2) lies in some linear subspace of $\mathbb{R}^K$. The Assumption that the global process be VAR with order at most p is crucial here. Indeed, when the global process is VAR or VMA, noncausality reduces to linear restrictions on the parameter space, which is necessary to construct optimal tests. In the other hand, in the strict VARMA case (both orders are positive), noncausality is characterized by a set of nonlinear constraints on the parameter space, see Boudjellaba, Dufour and Roy (1992).

The null hypothesis under which (iii) holds will be denoted by $\mathcal{H}_0$. In a similar way, the null hypothesis under which (i) or (ii) hold will be denoted, respectively by $\mathcal{H}_0^{(12)}$ and $\mathcal{H}_0^{(21)}$.

It follows from Proposition 2.1 that $\mathcal{H}_0$ takes the form of a set of $2pd_1d_2$ linear restrictions on the parameter value $\boldsymbol{\theta}$. Let $\mathcal{A}$ be the set of all (p)-tuples $(\mathbf{A}_1, \ldots, \mathbf{A}_p)$ of $d \times d$ real matrices of the block-diagonal form $\mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{(11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_j^{(22)} \end{pmatrix}$, $j = 1, ..., p$. Then $\mathcal{H}_0$ holds iff $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, where, $\boldsymbol{\Theta}_0 := \left\{ \boldsymbol{\theta} = \left(\mathrm{vec}^T \mathbf{A}_1, ..., \mathrm{vec}^T \mathbf{A}_p \right)^T \in \boldsymbol{\Theta} \middle| (\mathbf{A}_1, ..., \mathbf{A}_p) \in \mathcal{A} \right\}$ is the intersection of $\boldsymbol{\Theta}$ with a $(p(d_1^2 + d_2^2))-$dimensional subspace of $\mathbb{R}^K$. In a similar way we can define $\boldsymbol{\Theta}_0^{(12)}$ (resp. $\boldsymbol{\Theta}_0^{(21)}$) such that $\mathcal{H}_0^{(12)}$ (resp. $\mathcal{H}_0^{(21)}$) holds iff $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{(12)}$ (resp. $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{(21)}$).

2.2 Elliptical distributions

In order to construct locally optimal rank-based tests, we restrict ourselves to a class of elliptically symmetric densities. For more details on this class of densities, see Bilodeau and Brenner (1999). The approach we have adopted in the derivation of optimality results is based on Le Cam's asymptotic theory. This requires the model to be uniformly locally asymptotically normal (ULAN). ULAN of course does not hold without a few regularity conditions: finite second-order moments and finite Fisher information of the underlying density of the innovations. Those technical assumptions are taken into account in Assumptions (B1),(B2), and (B3) below in a form that is adapted to the elliptical context.

(B1) Denote by $\boldsymbol{\Sigma}$ a symmetric positive definite $d \times d$ matrix, and by $f : \mathbb{R}_0^+ \to \mathbb{R}^+$ a nonnegative function such that $f > 0$ a.e. and $\int_0^\infty r^{d-1} f(r) dr < \infty$. We will assume throughout that

$\{\boldsymbol{\epsilon}_t, t \in \mathbb{Z}\}$ is a d-variate elliptic strong white noise process with scatter matrix $\boldsymbol{\Sigma}$, i.e., a sequence of independent, identically distributed (iid) random vectors with mean zero, and probability density given by

$$\underline{f}(\mathbf{z}, \boldsymbol{\Sigma}, f) = c_{d,f}(det\boldsymbol{\Sigma})^{-\frac{1}{2}} f(\|\mathbf{z}\|_{\boldsymbol{\Sigma}}), \mathbf{z} \in \mathbb{R}^d.$$

Here, $\|\mathbf{z}\|_{\boldsymbol{\Sigma}}$ denotes the norm of $\mathbf{z}$ in the metric associated with $\boldsymbol{\Sigma}$, i.e. $\|\mathbf{z}\|^2_{\boldsymbol{\Sigma}} = \mathbf{z}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}$. The constant $c_{d,f}$ is the normalization factor $(\kappa_d \mu_{d-1;f})^{-1}$, where ($\kappa_d$ stands for the $(d-1)$-dimensional Lebesgue measure of the unit sphere $\mathcal{S}^{d-1} \in \mathbb{R}^d$, and $\mu_{l;f} = \int_0^\infty r^l f(r)dr$.

Note that $\boldsymbol{\Sigma}$ and f are only identified up to an arbitrary scale factor. This will not be a problem since we just need the multivariate scatter matrix $c\boldsymbol{\Sigma}$ (for some arbitrary $c > 0$), not $\boldsymbol{\Sigma}$ itself. We will denote by $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ the unique upper-triangular $d \times d$ array with positive diagonal elements that satisfies $\boldsymbol{\Sigma}^{-1} = \left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)^T \boldsymbol{\Sigma}^{-\frac{1}{2}}$. With this notation, $\frac{\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\epsilon}_t}{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\epsilon}_t\|}$ are iid, and uniformly distributed over $\mathcal{S}^{d-1}$. Similarly, $\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\epsilon}_t\|$ are iid with probability density function

$$\tilde{f}(r) = (\mu_{d-1;f})^{-1} r^{d-1} f(r) I[r > 0],$$

where I_E denotes the indicator function associated with the Borel set E. The terminology *radial density* will be used for f and $\tilde{f}$ (though only $\tilde{f}$ is a genuine probability density). We denote by $\tilde{F}$ the distribution function associated with $\tilde{f}$.

(B2) We assume that the second-order moments of $\underline{f}$ are finite. A sufficient and necessary condition is given by $\mu_{d+1;f} = \int_0^\infty r^{d+1} f(r)dr < \infty$.

(B3) Let $L^2(\mathbb{R}_0^+, \mu_l)$ the space of all measurable functions $h : \mathbb{R}_0^+ \to \mathbb{R}$ such that $\int_0^\infty [h(r)]^2 r^l dr < \infty$. The square root $f^{\frac{1}{2}}$ of f is in the subspace $W^{1,2}(\mathbb{R}_0^+, \mu_{d-1})$ of $L^2(\mathbb{R}_0^+, \mu_{d-1})$ containing all functions $h : \mathbb{R}_0^+ \to \mathbb{R}$ admitting a weak derivative h' that also belongs to $L^2(\mathbb{R}_0^+, \mu_{d-1})$.

¿From Hallin and Paindaveine (2002a), Assumption (B3) is strictly equivalent to the usual assumption of quadratic mean differentiability of $\underline{f}^{\frac{1}{2}}$, requiring the existence of a square integrable vector $\mathbf{D}\underline{f}^{\frac{1}{2}}$ such that, for all $\mathbf{0} \neq \mathbf{h} \to \mathbf{0}$,

$$\left(\mathbf{h}^T\mathbf{h}\right)^{-1} \int \left(\underline{f}^{\frac{1}{2}}(\mathbf{z}+\mathbf{h}) - \underline{f}^{\frac{1}{2}}(\mathbf{z}) - \mathbf{h}^T\mathbf{D}\underline{f}^{\frac{1}{2}}(\mathbf{z})\right)^2 \mathrm{d}\mathbf{z} \longrightarrow 0, \qquad \text{as} \quad \mathbf{h} \to \mathbf{0}.$$

Assumption (B3) unfortunately is not easy to check for; the following sufficient condition covers most cases of practical interest.

(B3') f is absolutely continuous, with derivative f', and $(f^{\frac{1}{2}})' = \frac{f'}{2f^{1/2}}$ is in $L^2(\mathbb{R}_0^+, \mu_{d-1})$.

Denote $\varphi_f = -2\frac{(f^{1/2})'}{f^{1/2}}$. Assumption (B3) ensures the finiteness of the *radial Fisher information*

$$\mathcal{I}_{d,f} = (\mu_{d-1;f})^{-1} \int_0^\infty [\varphi_f(r)]^2 r^{d-1} f(r) dr.$$

Examples of radial densities f satisfying (B1)-(B3) are $f(r) = \exp(-r^2/2)$ and $f(r) = (1+r^2/\nu)^{-(d+\nu)/2}$, with $\nu > 2$, yielding, respectively, the $d-$variate multinormal distribution and the $d-$variate Student distribution with ν degrees of freedom.

2.3 Local asymptotic normality

The likelihoods we are considering here are conditional likelihoods (conditional upon initial values $(\mathbf{X}_{1-p}, ..., \mathbf{X}_0)$); under Assumption (A1), the influence of these initial values vanishes asymptotically, see for example Toda and Phillips (1993), or Hallin and Werker (1999). Denote by $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}}$ the distribution of $\mathbf{X}^{(N)} := (\mathbf{X}_1, ..., \mathbf{X}_N)$ under the *radial density* f, the scatter matrix $\mathbf{\Sigma}$ and the parameter value $\boldsymbol{\theta}$, conditional on $(\mathbf{X}_{1-p}, ..., \mathbf{X}_0)$. It will be convenient to write $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}, f)$ for the simple hypothesis under which a realization $\mathbf{X}^{(N)} := (\mathbf{X}_1, ..., \mathbf{X}_N)$ is generated by model (2.1) with the *radial density* f, the scatter matrix $\mathbf{\Sigma}$ and the parameter value $\boldsymbol{\theta}$.

Consider two arbitrary sequences of $d \times d$ matrices $\boldsymbol{\gamma}_1^{(N)}, ..., \boldsymbol{\gamma}_p^{(N)}$, and $\tilde{\boldsymbol{\gamma}}_1^{(N)}, ..., \tilde{\boldsymbol{\gamma}}_p^{(N)}$, and let $\boldsymbol{\tau}^{(N)} := \left(\mathrm{vec}^T \boldsymbol{\gamma}_1^{(N)}, ..., \mathrm{vec}^T \boldsymbol{\gamma}_p^{(N)}\right)^T \in \mathbb{R}^K$ and $\tilde{\boldsymbol{\tau}}^{(N)} := \left(\mathrm{vec}^T \tilde{\boldsymbol{\gamma}}_1^{(N)}, ..., \mathrm{vec}^T \tilde{\boldsymbol{\gamma}}_p^{(N)}\right)^T \in \mathbb{R}^K$. We suppose that $\|\boldsymbol{\tau}^{(N)}\|$ and $\|\tilde{\boldsymbol{\tau}}^{(N)}\|$ remain bounded as $N \to \infty$. Whenever $\boldsymbol{\tau}^{(N)}$ is a constant, we write $\boldsymbol{\tau} := \left(\mathrm{vec}^T \boldsymbol{\gamma}_1, ..., \mathrm{vec}^T \boldsymbol{\gamma}_p\right)^T$ instead of $\boldsymbol{\tau}^{(N)}$. Define the local sequences

$$\boldsymbol{\theta}^{(N)} := \left(\mathrm{vec}^T \mathbf{A}_1^{(N)}, ..., \mathrm{vec}^T \mathbf{A}_p^{(N)}\right)^T := \boldsymbol{\theta} + N^{-1/2} \boldsymbol{\tau}^{(N)}, \tag{2.3}$$

$$\tilde{\boldsymbol{\theta}}^{(N)} := \left(\mathrm{vec}^T \tilde{\mathbf{A}}_1^{(N)}, ..., \mathrm{vec}^T \tilde{\mathbf{A}}_p^{(N)}\right)^T := \boldsymbol{\theta}^{(N)} + N^{-1/2} \tilde{\boldsymbol{\tau}}^{(N)}. \tag{2.4}$$

The logarithm of the likelihood ratio for $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}^{(N)}}$ against $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\tilde{\boldsymbol{\theta}}^{(N)}}$ takes the form

$$\Lambda^{(N)}_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}\left(\mathbf{X}^{(N)}\right) := \log\left(\frac{\mathrm{dP}^{(N)}_{\mathbf{\Sigma},f,\tilde{\boldsymbol{\theta}}^{(N)}}}{\mathrm{dP}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}^{(N)}}}\right) = \sum_{t=1}^N \log\left(f(\mathbf{e}_t^{(N)}(\tilde{\boldsymbol{\theta}}^{(N)}))\big/ f(\mathbf{e}_t^{(N)}(\boldsymbol{\theta}^{(N)}))\right), \tag{2.5}$$

where, $\mathbf{e}_t^{(N)}(\boldsymbol{\theta}^{(N)}) := \mathbf{X}_t - \sum_{j=1}^p \mathbf{A}_j^{(N)} \mathbf{X}_{t-j}$, and $\mathbf{e}_t^{(N)}(\tilde{\boldsymbol{\theta}}^{(N)}) := \mathbf{X}_t - \sum_{j=1}^p \tilde{\mathbf{A}}_j^{(N)} \mathbf{X}_{t-j}$.

Now, let $\mathbf{e}_t^{(N)}(\boldsymbol{\theta})$ be the residual under $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}}$,

$$\mathbf{e}_t^{(N)}(\boldsymbol{\theta}) := \mathbf{X}_t - \sum_{j=1}^p \mathbf{A}_j \mathbf{X}_{t-j}, \tag{2.6}$$

that we decompose into $\mathbf{e}_t^{(N)}(\boldsymbol{\theta}) = d_t^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}) \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma})$, where $d_t^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}) := \|\mathbf{e}_t^{(N)}(\boldsymbol{\theta})\|_{\mathbf{\Sigma}}$ and $\mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}) := \mathbf{\Sigma}^{-\frac{1}{2}} \mathbf{e}_t^{(N)}(\boldsymbol{\theta}) / d_t^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma})$. As in Garel and Hallin (1995), we define the residual $\underline{f}$-cross

covariance matrix at lag i as $\mathbf{\Lambda}^{(N)}_{i,\mathbf{\Sigma},\underline{f}}(\boldsymbol{\theta}) := (N-i)^{-1}\sum_{t=i+1}^{N}\varphi_{\underline{f}}(\mathbf{e}^{(N)}_t(\boldsymbol{\theta}))\left(\mathbf{e}^{(N)}_{t-i}(\boldsymbol{\theta})\right)^T$, where $\varphi_{\underline{f}} = -2\frac{(\underline{f}^{1/2})'}{\underline{f}^{1/2}}$. Due to the elliptical structure of $\underline{f}$, these cross-covariance matrices take the form

$$\mathbf{\Lambda}^{(N)}_{i,\mathbf{\Sigma},f}(\boldsymbol{\theta}) := (N-i)^{-1}\left(\mathbf{\Sigma}^{-\frac{1}{2}}\right)^T\left(\sum_{t=i+1}^{N}\varphi_f(d^{(N)}_t(\boldsymbol{\theta},\mathbf{\Sigma}))d^{(N)}_{t-i}(\boldsymbol{\theta},\mathbf{\Sigma})\mathbf{U}^{(N)}_t(\boldsymbol{\theta},\mathbf{\Sigma})\left(\mathbf{U}^{(N)}_{t-i}(\boldsymbol{\theta},\mathbf{\Sigma})\right)^T\right)\left(\mathbf{\Sigma}^{\frac{1}{2}}\right)^T. \tag{2.7}$$

Denote the vector of all cross-covariance matrices by

$$\mathbf{S}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}) = \left((N-1)^{\frac{1}{2}}\left(\mathrm{vec}\mathbf{\Lambda}^{(\mathrm{N})}_{1,\mathbf{\Sigma},\mathrm{f}}(\boldsymbol{\theta})\right)^T, ..., (N-i)^{\frac{1}{2}}\left(\mathrm{vec}\mathbf{\Lambda}^{(\mathrm{N})}_{\mathrm{i},\mathbf{\Sigma},\mathrm{f}}(\boldsymbol{\theta})\right)^T, ..., \left(\mathrm{vec}\mathbf{\Lambda}^{(\mathrm{N})}_{\mathrm{N}-1,\mathbf{\Sigma},\mathrm{f}}(\boldsymbol{\theta})\right)^T\right)^T.$$

Finally, let $\mathbf{M}^{(s)}(\boldsymbol{\theta})$ be the following sequence of $pd^2 \times (s-1)d^2$ dimensional matrices associated with the sequence $\{\mathbf{G}_u(\boldsymbol{\theta})\}$

$$\mathbf{M}^{(s)}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{G}_0(\boldsymbol{\theta})\otimes\mathbf{I}_d & \mathbf{G}_1(\boldsymbol{\theta})\otimes\mathbf{I}_d & ... & \mathbf{G}_{s-2}(\boldsymbol{\theta})\otimes\mathbf{I}_d \\ \mathbf{0} & \mathbf{G}_0(\boldsymbol{\theta})\otimes\mathbf{I}_d & ... & \mathbf{G}_{s-3}(\boldsymbol{\theta})\otimes\mathbf{I}_d \\ & & & \vdots \\ \mathbf{0} & & & \mathbf{G}_{s-p-1}(\boldsymbol{\theta})\otimes\mathbf{I}_d \end{pmatrix}. \tag{2.8}$$

We are now ready to state the ULAN property which is the main result of this section. It is a particular case of the ULAN property in the very general context of a multivariate general linear model with VARMA errors established by Garel and Hallin (1995).

Proposition 2.2 *Suppose that Assumptions* (A1), (B1), (B2) and (B3) *are satisfied. Let* $\boldsymbol{\theta} \in \mathbf{\Theta}_0$, $\boldsymbol{\theta}^{(N)}$ *and* $\tilde{\boldsymbol{\theta}}^{(N)}$ *as defined in* (2.3) *and* (2.4), *respectively. Then,*

$$\Lambda^{(N)}_{\tilde{\boldsymbol{\theta}}^{(N)}/\boldsymbol{\theta}^{(N)}}\left(\mathbf{X}^{(N)}\right) = (\tilde{\boldsymbol{\tau}}^{(N)})^T\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}^{(N)}) - \frac{1}{2}(\tilde{\boldsymbol{\tau}}^{(N)})^T\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\tilde{\boldsymbol{\tau}}^{(N)} + o_{\mathrm{P}}(1),$$

under $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}^{(N)}}$, *as* $N\to\infty$, *with the central sequence*

$$\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}) := \begin{pmatrix} \sum_{i=1}^{N-1}(N-i)^{-\frac{1}{2}}(\mathbf{G}_{i-1}(\boldsymbol{\theta})\otimes I_d)\mathrm{vec}\mathbf{\Lambda}^{(\mathrm{N})}_{\mathrm{i},\mathbf{\Sigma},\mathrm{f}}(\boldsymbol{\theta}) \\ \vdots \\ \sum_{i=p}^{N-1}(N-i)^{-\frac{1}{2}}(\mathbf{G}_{i-p}(\boldsymbol{\theta})\otimes I_d)\mathrm{vec}\mathbf{\Lambda}^{(\mathrm{N})}_{\mathrm{i},\mathbf{\Sigma},\mathrm{f}}(\boldsymbol{\theta}) \end{pmatrix} = \mathbf{M}^{(N)}(\boldsymbol{\theta})\mathbf{S}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}), \tag{2.9}$$

and the information matrix

$$\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}) := \xi_d(f)\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}, \tag{2.10}$$

where $\mathbf{M}^{(N)}(\boldsymbol{\theta})$ *is defined in (2.8), the constant* $\xi_d(f) = \frac{\mathcal{I}_{d,f}\mu_{d+1;f}}{d^2\mu_{d-1;f}}$ *and*

$$\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}} = \lim_{N\to+\infty}\mathbf{M}^{(N)}(\boldsymbol{\theta})\left[\mathbf{I}_{N-1}\otimes\left(\mathbf{\Sigma}\otimes\mathbf{\Sigma}^{-1}\right)\right]\left(\mathbf{M}^{(N)}(\boldsymbol{\theta})\right)^T = \left(\sum_{j=\max(i,i')}^{+\infty}\mathbf{G}_{j-i}\mathbf{\Sigma}\mathbf{G}^T_{j-i'}\otimes\mathbf{\Sigma}^{-1}\right)_{i,i'=1,...p}. \tag{2.11}$$

Moreover, as $N \to \infty$, $\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$ is asymptotically normal, with mean $\mathbf{0}$ and covariance $\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$ under $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}}$.

3 Optimal parametric tests for noncausality

3.1 Weak convergence of statistical experiments

Local asymptotic normality (LAN) at $\boldsymbol{\theta} \in \mathbf{\Theta}_0$ implies the weak convergence of the sequence of local experiments (localized at $\boldsymbol{\theta}$) $\mathcal{E}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}) := \{\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}, \boldsymbol{\tau} \in \mathbb{R}^K\}$ to the K-dimensional Gaussian shift experiment

$$\mathcal{E}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}) := \left\{\mathcal{N}\left(\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\,\boldsymbol{\tau},\, \underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\right), \boldsymbol{\tau} \in \mathbb{R}^K\right\}.$$

This convergence implies that all power functions that are implementable from the sequence $\mathcal{E}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$ converge, as $N \to \infty$, pointwise in $\boldsymbol{\tau}$ but uniformly with respect to the set of all possible testing procedures, to the power functions that are implementable in the limit Gaussian experiment $\mathcal{E}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$. Conversely, all risk functions associated with $\mathcal{E}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$ can be obtained as limits of sequences of risk functions associated with $\mathcal{E}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$. Denoting by $\mathbf{\Delta}$ the (K-dimensional) observation in $\mathcal{E}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$, it follows that, if a test $\phi(\mathbf{\Delta})$ enjoys some exact optimality property in the Gaussian experiment $\mathcal{E}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$, then the sequence $\phi\left(\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\right)$ inherits, locally and asymptotically, the same optimality properties in the sequence of experiments $\mathcal{E}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$—see, e.g., Le Cam (1986, Section 11.9).

3.2 Locally asymptotically most stringent test

Denote by $\mathbf{Q}$ a $K \times (K-r)$ matrix of maximal rank $K-r$, and by $\mathcal{M}(\mathbf{Q})$ the linear subspace of $\mathbb{R}^K$ spanned by the columns of $\mathbf{Q}$. The null hypothesis $\mathcal{H}_0 : \boldsymbol{\tau} \in \mathcal{M}(\mathbf{Q})$ is equivalent to $\mathcal{H}_0 : \underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\boldsymbol{\tau} \in \mathcal{M}(\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\,\mathbf{Q})$, a set of linear constraints on the location parameter of the Gaussian shift experiment $\mathcal{E}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$. The most stringent α-level test for this problem, consists in rejecting $\mathcal{H}_0$ whenever

$$\mathbf{\Delta}^T\left[(\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}))^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\mathbf{\Delta} > \chi^2_{r,1-\alpha}, \tag{3.1}$$

where $\chi^2_{r,1-\alpha}$ denotes the $(1-\alpha)$-quantile of a chi-square variable with r degrees of freedom. A locally asymptotically most stringent (at $\boldsymbol{\theta}$) test (see Hallin and Werker, 1999, Section 4.3) thus is obtained by substituting $\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$ for $\mathbf{\Delta}$ in (3.1).

In view of Proposition 2.1, the null hypothesis $\mathcal{H}_0 = \bigcup_{\boldsymbol{\theta}\in\mathbf{\Theta}_0}\bigcup_{\mathbf{\Sigma}}\bigcup_f \mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$ (here and in the sequel, union on $\mathbf{\Sigma}$ is taken over the set of symmetric positive definite $d \times d$ matrices, and union on f is taken over the set of all possible nonvanishing radial densities such that Assumptions (B1)-(B2)-(B3) hold) of noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ takes the form $\mathcal{H}_0 : \mathbf{Q}_\perp^T\boldsymbol{\theta} = \mathbf{0}$, with $\mathbf{Q}_\perp^T :=$

$\mathbf{I}_p \otimes \begin{pmatrix} \mathbf{L}_{d_1 d_2 \times d^2} \\ \mathbf{S}_{d_1 d_2 \times d^2} \end{pmatrix}$, where

$$\mathbf{L} = \left(\begin{pmatrix} \mathbf{I}_{d_1} \\ \mathbf{0}_{d_2 \times d_1} \end{pmatrix}^T \otimes \begin{pmatrix} \mathbf{0}_{d_2 \times d_1} & \mathbf{I}_{d_2} \end{pmatrix} \right) \quad \text{and} \quad \mathbf{S} = \left(\begin{pmatrix} \mathbf{0}_{d_1 \times d_2} \\ \mathbf{I}_{d_2} \end{pmatrix}^T \otimes \begin{pmatrix} \mathbf{I}_{d_1} & \mathbf{0}_{d_1 \times d_2} \end{pmatrix} \right). \tag{3.2}$$

An alternative form for $\mathcal{H}_0$ is $\mathcal{H}_0 : \boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$, with $\mathbf{Q} := \mathbf{I}_p \otimes \left(\mathbf{U}_{d^2 \times d_1 d_1}, \mathbf{V}_{d^2 \times d_2 d_2}\right)$ where

$$\mathbf{U} = \left(\begin{pmatrix} \mathbf{I}_{d_1} \\ \mathbf{0}_{d_2 \times d_1} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{I}_{d_1} \\ \mathbf{0}_{d_2 \times d_1} \end{pmatrix} \right) \quad \text{and} \quad \mathbf{V} = \left(\begin{pmatrix} \mathbf{0}_{d_1 \times d_2} \\ \mathbf{I}_{d_2} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{0}_{d_1 \times d_2} \\ \mathbf{I}_{d_2} \end{pmatrix} \right). \tag{3.3}$$

Referring to (3.1), a sequence of locally (at $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$) asymptotically most stringent α-level tests for the null hypothesis $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$ of noncausality with fixed f and $\boldsymbol{\Sigma}$, is $\underline{\phi}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}) := I\left[\underline{Q}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}) > \chi^2_{r,1-\alpha}\right]$, with $r = 2pd_1d_2$ and

$$\underline{Q}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}) := \left(\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})\right)^T \left[\left(\underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})\right)^{-1} - \mathbf{Q}\left(\mathbf{Q}^T \underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}) \mathbf{Q}\right)^{-1} \mathbf{Q}^T\right] \underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}), \tag{3.4}$$

where $\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})$ and $\underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})$ are defined in (2.9) and (2.10), respectively. The test (3.4) however is of little practical use as long as it explicitly depends on partially unspecified parameter value $\boldsymbol{\theta}$ under the null. In order to construct a sequence of locally (at any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$) asymptotically most stringent α-level tests for the null hypothesis $\mathcal{H}_0(\boldsymbol{\Sigma}, f) = \bigcup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, let us assume that a sequence of estimators $\hat{\boldsymbol{\theta}}^{(N)}$ is available, with the following properties:

(C1) (i) $\hat{\boldsymbol{\theta}}^{(N)} \in \mathcal{M}(\mathbf{Q})$;

(ii) (root-N consistency) for all $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$ and $\epsilon > 0$, there exist $b(\boldsymbol{\theta}, \epsilon)$ and $N(\boldsymbol{\theta}, \epsilon)$ such that

$$\mathrm{P}^{(N)}_{\boldsymbol{\Sigma},f,\boldsymbol{\theta}}\left[\left\|\sqrt{N}\left(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}\right)\right\| > b(\boldsymbol{\theta}, \epsilon)\right] < \epsilon \quad \text{for all} \quad N \geq N(\boldsymbol{\theta}, \epsilon);$$

(iii) $\hat{\boldsymbol{\theta}}^{(N)}$ is *locally asymptotically discrete*, that is, for all $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$ and $c > 0$, there exists $J = J(\boldsymbol{\theta}; c)$ such that the number of possible values of $\hat{\boldsymbol{\theta}}^{(N)}$ in balls of the form $\left\{\mathbf{t} \in \mathbb{R}^K : \left\|\sqrt{N}(\mathbf{t} - \boldsymbol{\theta})\right\| \leq c\right\}$ is bounded by J, uniformly as N tends to infinity.

The local discreteness Assumption (C1)-(iii) is a purely technical requirement, with little practical implications as, for fixed size, any estimate can be considered part of a locally asymptotically discrete sequence. The assumption (C1)-(i) and (C1)-(ii) does not cause any additional difficulty: any sequence of unconstrained $\sqrt{N}$-consistent estimators indeed can be turned into a constrained and $\sqrt{N}$-consistent one by means of a simple projection onto $\mathcal{M}(\mathbf{Q})$. The assumption of $\sqrt{N}$-consistency of unconstrained estimators is satisfied by all classical estimators (Yule-Walker, least squares, maximum likelihood,...).

It is a classical result (see, e.g., Le Cam, 1986, Chapter 11) that, under ULAN (which entails the asymptotic linearity of $\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$) and Assumption (C1), substituting $\hat{\boldsymbol{\theta}}^{(N)}$ for $\boldsymbol{\theta}$ in (3.4) has no influence on the asymptotic behavior of $\underline{\phi}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$, hence on its local asymptotic optimality. Thus the LAN property straightforwardly allows for building locally and asymptotically optimal testing procedures, under fixed $\mathbf{\Sigma}$ and f. The scatter matrix $\mathbf{\Sigma}$ is unknown and consequently plays the role of a nuisance parameter. However, we can replace it by $\hat{\mathbf{\Sigma}}$, a consistent estimator of $a\mathbf{\Sigma}$ for some fixed $a > 0$, and the resulting procedure allows for testing the parametric null hypothesis $\mathcal{H}_0(f) = \bigcup_{\boldsymbol{\theta}\in\mathbf{\Theta}_0}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$.

For simplicity of notation, we will use hereafter $\mathbf{\Omega}_f$ and $\mathbf{\Delta}_f$ instead of $\underline{\mathbf{\Omega}}_{\hat{\mathbf{\Sigma}},f}(\hat{\boldsymbol{\theta}}^{(N)})$ and $\underline{\mathbf{\Delta}}^{(N)}_{\hat{\mathbf{\Sigma}},f}(\hat{\boldsymbol{\theta}}^{(N)})$. The sequence of tests, $\phi_f := I\left[Q_f > \chi^2_{r,1-\alpha}\right]$, where

$$Q_f := \underline{Q}_{\hat{\mathbf{\Sigma}},f}(\hat{\boldsymbol{\theta}}^{(N)}) = \mathbf{\Delta}_f^T\left[\mathbf{\Omega}_f^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{\Omega}_f\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\mathbf{\Delta}_f, \tag{3.5}$$

is locally asymptotically most stringent α-level tests for $\mathcal{H}_0(f)$ against $\bigcup_{\boldsymbol{\theta}\notin\mathbf{\Theta}_0}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$. The procedure is of course highly parametric, since, in general, it is only valid if the underlying radial density f is known. The power of this test against $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}$ satisfies

$$\lim_{N\to+\infty}\mathrm{E}_{\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}}\left[\phi_{\mathbf{\Sigma},f}(\boldsymbol{\theta})\right] = 1 - F^r_{\chi^2}\left(\chi^2_{r,1-\alpha};\psi^2_f\left(\boldsymbol{\tau},\boldsymbol{\theta},\mathbf{\Sigma}\right)\right),$$

where $\psi^2_f = \psi^2_f\left(\boldsymbol{\tau},\boldsymbol{\theta},\mathbf{\Sigma}\right) := \xi_d(f)\delta_{\boldsymbol{\theta},\mathbf{\Sigma},\boldsymbol{\tau}}$ with $\xi_d(f)$ being defined in Proposition (2.2),

$$\delta_{\boldsymbol{\theta},\mathbf{\Sigma},\boldsymbol{\tau}} = \boldsymbol{\tau}^T\left[\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}} - \mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}\mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}\right]\boldsymbol{\tau}, \tag{3.6}$$

and $F^r_{\chi^2}\left(.;\psi^2\right)$ denotes the distribution function of the non central chi-square variable with r degrees of freedom and noncentrality parameter ψ^2.

Similarly, the null hypothesis $\mathcal{H}^{(12)}_0$ under which $\mathbf{X}^{(1)}$ does not Granger cause $\mathbf{X}^{(2)}$ takes the form $\mathcal{H}^{(12)}_0 : (\mathbf{Q}^{(12)}_\perp)^T\boldsymbol{\theta} = \mathbf{0}$, with $(\mathbf{Q}^{(12)}_\perp)^T := \mathbf{I}_p \otimes \mathbf{L}_{d_1d_2\times d^2}$, where $\mathbf{L}$ is defined in (3.2). An alternative form for $\mathcal{H}^{(12)}_0$ is $\mathcal{H}^{(12)}_0 : \boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q}^{(12)})$, with $\mathbf{Q}^{(12)} := \mathbf{I}_p \otimes \left(\mathbf{U}_{d^2\times d_1d_1}, \tilde{\mathbf{V}}_{d^2\times dd_2}\right)$, where $\mathbf{U}$ is defined in (3.3) and $\tilde{\mathbf{V}} = \left(\begin{pmatrix}\mathbf{0}_{d_1\times d_2}\\ \mathbf{I}_{d_2}\end{pmatrix}\otimes\mathbf{I}_d\right)$. Locally asymptotically optimal tests for the null hypothesis $\mathcal{H}^{(12)}_0(f) = \bigcup_{\boldsymbol{\theta}\in\mathbf{\Theta}^{(12)}_0}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$. is obtained by substituting $\mathbf{Q}^{(12)}$ for $\mathbf{Q}$ in Assumption (C1) and in equation (3.5). A sequence of locally (at any $\boldsymbol{\theta}\in\mathbf{\Theta}^{(12)}_0$) asymptotically most stringent α-level tests for the null hypothesis $\mathcal{H}^{(12)}_0(f)$ is then given by $\phi^{(12)}_f := I\left[Q^{(12)}_f > \chi^2_{r_1,1-\alpha}\right]$, where

$$Q^{(12)}_f = \mathbf{\Delta}_f^T\left[\mathbf{\Omega}_f^{-1} - \mathbf{Q}^{(12)}\left((\mathbf{Q}^{(12)})^T\mathbf{\Omega}_f\mathbf{Q}^{(12)}\right)^{-1}(\mathbf{Q}^{(12)})^T\right]\mathbf{\Delta}_f, \tag{3.7}$$

with $r_1 = pd_1d_2$. The power of this test against local alternatives, $\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}$ $(\boldsymbol{\theta}\in\mathbf{\Theta}^{(12)}_0)$, satisfies

$$\lim_{N\to+\infty}\mathrm{E}_{\mathrm{P}^{(N)}_{\mathbf{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}}\left[\phi^{(12)}_f\right] = 1 - F^{r_1}_{\chi^2}\left(\chi^2_{r_1,1-\alpha};\psi^2_{1,f}\left(\boldsymbol{\tau},\boldsymbol{\theta},\mathbf{\Sigma}\right)\right),$$

where $\psi_{1,f}^2 = \psi_{1,f}^2(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma}) := \xi_d(f)\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}^{(12)}$ with

$$\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}^{(12)} = \boldsymbol{\tau}^T\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}} - \mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}^{(12)}\left((\mathbf{Q}^{(12)})^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}^{(12)}\right)^{-1}(\mathbf{Q}^{(12)})^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\right]\boldsymbol{\tau}. \tag{3.8}$$

Similarly, the null hypothesis $\mathcal{H}_0^{(21)}$ under which $\mathbf{X}^{(2)}$ does not Granger cause $\mathbf{X}^{(1)}$ takes the form $\mathcal{H}_0^{(21)} : (\mathbf{Q}_\perp^{(21)})^T\boldsymbol{\theta} = \mathbf{0}$, with $(\mathbf{Q}_\perp^{(21)})^T := \mathbf{I}_p \otimes \mathbf{S}_{d_1d_2\times d^2}$ where $\mathbf{S}$ is defined in (3.2). An alternative form for $\mathcal{H}_0^{(21)}$ is $\mathcal{H}_0^{(21)} : \boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q}^{(21)})$, with $\mathbf{Q}^{(21)} := \mathbf{I}_p \otimes \left(\tilde{\mathbf{U}}_{d^2\times dd_1}, \mathbf{V}_{d^2\times d_2d_2}\right)$, where $\mathbf{V}$ is defined in (3.3) and $\tilde{\mathbf{U}} = \left(\begin{pmatrix}\mathbf{I}_{d_1}\\ \mathbf{0}_{d_2\times d_1}\end{pmatrix}\otimes\mathbf{I}_d\right)$. A sequence of locally (at any $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{(21)}$) asymptotically most stringent α-level tests for $\mathcal{H}_0^{(21)}(f) = \bigcup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{(21)}}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$ is $\phi_f^{(21)} := I\left[Q_f^{(21)} > \chi^2_{r_2,1-\alpha}\right]$, where

$$Q_f^{(21)} = \boldsymbol{\Delta}_f^T\left[\boldsymbol{\Omega}_f^{-1} - \mathbf{Q}^{(21)}\left((\mathbf{Q}^{(21)})^T\boldsymbol{\Omega}_f\mathbf{Q}^{(21)}\right)^{-1}(\mathbf{Q}^{(21)})^T\right]\boldsymbol{\Delta}_f, \tag{3.9}$$

with $r_2 = pd_1d_2$. The power of this test against local alternatives, $\mathrm{P}^{(N)}_{\boldsymbol{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}$ $(\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{(21)})$, satisfies

$$\lim_{N\to+\infty}\mathrm{E}_{\mathrm{P}^{(N)}_{\boldsymbol{\Sigma},f,\boldsymbol{\theta}+\frac{1}{\sqrt{N}}\boldsymbol{\tau}}}\left[\phi_f^{(21)}\right] = 1 - F^{r_2}_{\chi^2}\left(\chi^2_{r_2,1-\alpha};\psi^2_{2,f}(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma})\right),$$

where $\psi_{2,f}^2 = \psi_{2,f}^2(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma}) := \xi_d(f)\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}^{(21)}$ with

$$\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}^{(21)} = \boldsymbol{\tau}^T\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}} - \mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}^{(21)}\left((\mathbf{Q}^{(21)})^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}^{(21)}\right)^{-1}(\mathbf{Q}^{(21)})^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\right]\boldsymbol{\tau}. \tag{3.10}$$

3.3 Pseudo-Gaussian tests

A fatal shortcoming of the optimal tests (3.5), (3.7) and (3.9) described in Section 3.2 is that their validity, in general, is limited to the innovation density f. In practice, f is never specified and if the true density is g rather than f, in general, the tests ϕ_f, $\phi_f^{(12)}$ and $\phi_f^{(21)}$ are not asymptotically valid since their asymptotic levels might be different from α. Therefore, these optimal tests are of little practical value. Fortunately, the Gaussian case $\mathcal{N} = exp(-r^2/2)$, is a remarkable exception. The Gaussian central sequence is

$$\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta}) = \mathbf{M}^{(N)}(\boldsymbol{\theta})\left((N-1)^{\frac{1}{2}}\left(\mathrm{vec}\boldsymbol{\Lambda}^{(N)}_{1,\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta})\right)^T, ..., \left(\mathrm{vec}\boldsymbol{\Lambda}^{(N)}_{N-1,\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta})\right)^T\right)^T,$$

where $\boldsymbol{\Lambda}^{(N)}_{i,\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta}) = (N-i)^{-1}\boldsymbol{\Sigma}^{-1}\sum_{t=i+1}^{N}\mathbf{e}_t^{(N)}(\boldsymbol{\theta})\left(\mathbf{e}_{t-i}^{(N)}(\boldsymbol{\theta})\right)^T$. Substituting the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_E = N^{-1}\sum_{t=1}^{N}\mathbf{e}_t^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\left(\mathbf{e}_t^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\right)^T$ for the scatter matrix $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\theta}}^{(N)}$ for $\boldsymbol{\theta}$, the central sequence takes the form

$$\boldsymbol{\Delta}_{\mathcal{N}} = \underline{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\Sigma}}_E,\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)}) = \mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\left((N-1)^{\frac{1}{2}}\left(\mathrm{vec}\boldsymbol{\Lambda}^{(N)}_{1,\hat{\boldsymbol{\Sigma}}_E,\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)})\right)^T, ..., \left(\mathrm{vec}\boldsymbol{\Lambda}^{(N)}_{N-1,\hat{\boldsymbol{\Sigma}}_E,\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)})\right)^T\right)^T.$$

On the other hand, the Gaussian *information matrix* is $\underline{\mathbf{\Omega}}_{\mathbf{\Sigma},\mathcal{N}}(\boldsymbol{\theta}) = \mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}$ and a consistent estimator (under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}, f)$) is

$$\mathbf{\Omega}_{\mathcal{N}} = \underline{\mathbf{\Omega}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)}) = \mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\mathbf{\Sigma}}_E} = \mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}) \left[\mathbf{I}_{N-1} \otimes (\hat{\mathbf{\Sigma}}_E \otimes \hat{\mathbf{\Sigma}}_E^{-1})\right] \left(\mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\right)^T.$$

Now, using $\mathbf{\Delta}_{\mathcal{N}}$ and $\mathbf{\Omega}_{\mathcal{N}}$ instead of $\mathbf{\Delta}_f$ and $\mathbf{\Omega}_f$ in (3.5), (3.7) and (3.9), we obtain the Gaussian parametric tests $\phi_{\mathcal{N}}$, $\phi_{\mathcal{N}}^{(12)}$ and $\phi_{\mathcal{N}}^{(21)}$ and their corresponding statistics $Q_{\mathcal{N}}$, $Q_{\mathcal{N}}^{(12)}$ and $Q_{\mathcal{N}}^{(21)}$. The Gaussian parametric tests, are valid irrespective of the true underlying density f, provided that second order moments are finite. Therefore, in the sequel, Gaussian tests will be called pseudo-Gaussian tests and we will concentrate on this pseudo-Gaussian version.

The following Theorem gives their asymptotic distribution-freeness, as well as their local powers and optimality properties. The proof is given in the Appendix. These results allow for computing asymptotic relative efficiencies. Indeed, the Gaussian test will serve as a benchmark in Section 4.3.

Theorem 3.1 *Assume that Assumptions (A1), (B1), (B2), and (C1) hold. Consider the sequence of tests $\phi_{\mathcal{N}}$ (resp. $\phi_{\mathcal{N}}^{(12)}$ or $\phi_{\mathcal{N}}^{(12)}$) that rejects the null hypothesis $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$) whenever $Q_{\mathcal{N}}$ (resp. $Q_{\mathcal{N}}^{(12)}$ or $Q_{\mathcal{N}}^{(21)}$) exceeds the $1-\alpha$ quantile $\chi^2_{r,1-\alpha}$ of a chi-square variable with $r = 2pd_1d_2$ (resp $r_1 = pd_1d_2$ or $r_2 = pd_1d_2$) degrees of freedom. Then*

(i) $Q_{\mathcal{N}}$ (resp. $Q_{\mathcal{N}}^{(12)}$ or $Q_{\mathcal{N}}^{(21)}$) is asymptotically chi-square with $r = 2pd_1d_2$ (resp. $r_1 = pd_1d_2$ or $r_2 = pd_1d_2$) degrees of freedom under $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$).

(ii) $Q_{\mathcal{N}}$ (resp. $Q_{\mathcal{N}}^{(12)}$ or $Q_{\mathcal{N}}^{(21)}$) is noncentral chi-square with $r = 2pd_1d_2$ (resp. $r_1 = pd_1d_2$ or $r_2 = pd_1d_2$) degrees of freedom and with noncentrality parameter $\psi^2_{\mathcal{N},f} = \psi^2_{\mathcal{N},f}(\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{\Sigma}) = \omega_d(f)\delta_{\boldsymbol{\theta},\mathbf{\Sigma},\boldsymbol{\tau}}$, where $\omega_d(f) = \frac{1}{d^2}\left(\int_0^1 \tilde{F}^{-1}(u)\varphi_f \circ \tilde{F}^{-1}(u)du\right)^2$, (resp. $\psi^2_{1,\mathcal{N},f} = \psi^2_{1,\mathcal{N},f}(\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{\Sigma}) = \omega_d(f)\delta^{(12)}_{\boldsymbol{\theta},\mathbf{\Sigma},\boldsymbol{\tau}}$, and $\psi^2_{2,\mathcal{N},f} = \psi^2_{2,\mathcal{N},f}(\boldsymbol{\tau}, \boldsymbol{\theta}, \mathbf{\Sigma}) := \omega_d(f)\delta^{(21)}_{\boldsymbol{\theta},\mathbf{\Sigma},\boldsymbol{\tau}}$), under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \mathbf{\Sigma}, f)$.

(iii) The sequence of tests $\phi_{\mathcal{N}}$ (resp. $\phi_{\mathcal{N}}^{(12)}$ or $\phi_{\mathcal{N}}^{(12)}$) are locally asymptotically most stringent for $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$) against $\bigcup_{\boldsymbol{\theta}\notin\mathbf{\Theta}_0}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}, \mathcal{N})$ (resp. $\bigcup_{\boldsymbol{\theta}\notin\mathbf{\Theta}_0^{(12)}}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}, \mathcal{N})$ or $\bigcup_{\boldsymbol{\theta}\notin\mathbf{\Theta}_0^{(21)}}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta}, \mathbf{\Sigma}, \mathcal{N})$).

4 Optimal nonparametric tests for noncausality

The tests defined in (3.5), (3.7) and (3.9) are valid only when the density of the noise is correctly specified. In this Section, a rank-based version of the central sequence will be obtained and a family of nonparametric tests will be defined. These new tests are based on multivariate residual ranks and

signs. In the sequel, we focus on testing for noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ ($\mathcal{H}_0 : \mathbf{X}^{(2)} \not\Leftarrow \mathbf{X}^{(1)}$). Testing for causality directions is achieved by replacing the matrix $\mathbf{Q}$ by $\mathbf{Q}^{(12)}$ or $\mathbf{Q}^{(21)}$ (depending on which direction $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$ is to be tested).

4.1 Multivariate signs and ranks

The generalized cross-covariances (2.7) are measurable with respect to the spherical distances between the residuals (2.6) and the origin in $\mathbb{R}^d$, $d_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) := \|\mathbf{e}_t^{(N)}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}$, and the standardized residuals $\mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) := \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{e}_t^{(N)}(\boldsymbol{\theta})/d_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. Under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, the residuals $\mathbf{U}_1^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$,...,$\mathbf{U}_N^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ are i.i.d., and uniformly distributed over the unit sphere $\mathcal{S}^{d-1} \in \mathbb{R}^d$, hence generalizing the traditional concept of signs: we henceforth call them *multivariate signs*. The distances, $d_1^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}), ..., d_N^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, are i.i.d. over the real line, with density function $\tilde{f}(r) = (\mu_{d-1;f})^{-1} f^{d-1} f(r) I[r > 0]$; their ranks, denoted $R_1(\boldsymbol{\theta}, \boldsymbol{\Sigma}), ..., R_N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, thus have the same distribution-freeness and maximal invariance properties as those of the absolute values of any univariate symmetrically distributed sample: we henceforth call them *multivariate ranks*. For $d = 1$, *multivariate ranks and signs* reduce to the ranks of absolute values and traditional signs, respectively.

For each $\boldsymbol{\Sigma}$ and N consider the group of *continuous monotone radial transformations* $\Im_{\boldsymbol{\Sigma}}^{(N)} = \left\{\Im_{\boldsymbol{\Sigma},g}^{(N)}\right\}$, acting on $(\mathbb{R}^d)^N$ and characterized by

$$\Im_{\boldsymbol{\Sigma},g}^{(N)}\left(\mathbf{e}_1^{(N)}(\boldsymbol{\theta}), ..., \mathbf{e}_N^{(N)}(\boldsymbol{\theta})\right) := \left(g(d_1^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}))\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}_1^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}), ..., g(d_N^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}))\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}_N^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})\right)$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a continuous monotone increasing function such that $g(0) = 0$ and $\lim_{r\to+\infty} g(r) = 0$. The group $\Im_{\boldsymbol{\Sigma}}^{(N)}$ is a generating group for $\bigcup_f \mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$ where the union is taken over the set of all possible nonvanishing radial densities. Along with signs $\left(\mathbf{U}_1^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}), ..., \mathbf{U}_N^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})\right)$, the ranks $R_t(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, $t = 1, ..., N$, are a maximal invariant for the group $\Im_{\boldsymbol{\Sigma}}^{(N)}$ of continuous monotone radial transformations.

Unfortunately, the *multivariate ranks and signs* can not be computed from the residuals $\mathbf{e}_t^{(N)} = \mathbf{e}_t^{(N)}(\boldsymbol{\theta}^{(N)})$ since they depend on the unknown scatter matrix $\boldsymbol{\Sigma}$. Under finite second-order moments Assumption (B2), a natural root-N consistent candidate for estimating $\boldsymbol{\Sigma}$ is the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_E = N^{-1}\sum_{t=1}^{N}\mathbf{e}_t^{(N)}\left(\mathbf{e}_t^{(N)}\right)^T$. However, the robustness properties of the empirical covariance matrix are rather poor. More generally, we assume that $\boldsymbol{\Sigma}$ is estimated by some $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\mathbf{e}_1^{(N)}, ..., \mathbf{e}_N^{(N)})$ such that

(D1) There exist a positive real constant a such that $\sqrt{N}(\hat{\boldsymbol{\Sigma}} - a\boldsymbol{\Sigma}) = O_p(1)$ as $N \to +\infty$ and $\hat{\boldsymbol{\Sigma}}$ is invariant under permutations and reflections (with respect to the origin in $\mathbb{R}^d$) of the $\mathbf{e}_t^{(N)}(\boldsymbol{\theta})$'s.

The corresponding distances from the origin $\hat{d}_t^{(N)}(\boldsymbol{\theta}) := d_t^{(N)}(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}})$ will be called *pseudo-Mahalanobis distances* and the corresponding signs $\hat{\mathbf{U}}_t^{(N)}(\boldsymbol{\theta}) := \mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}})$, the *pseudo-Mahalanobis signs*. Similarly, the *pseudo-Mahalanobis ranks* $\hat{R}_t(\boldsymbol{\theta}) = R_t(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}})$ are defined as the ranks of the *pseudo-Mahalanobis distances*. The terminology *Mahalanobis distances, signs and ranks* is used when $\hat{\boldsymbol{\Sigma}}$ is the empirical covariance matrix.

The parameter value $\boldsymbol{\theta}$ is partially unspecified under the null hypothesis (the *alignment* problem) and has to be substituted by a sequence of estimators $\hat{\boldsymbol{\theta}}^{(N)}$ such that Assumption (C1) holds. The corresponding signs and ranks $\tilde{\mathbf{U}}_t := \hat{\mathbf{U}}_t^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})$ and $\tilde{R}_t =: \hat{R}_t(\hat{\boldsymbol{\theta}}^{(N)})$ will be called *aligned* signs and ranks.

The null hypothesis $\mathcal{H}_0$ under which $\mathbf{X}^{(2)} \not\Rightarrow \mathbf{X}^{(1)}$ is invariant under block-diagonal-affine transformations, in the sense that for $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, $\mathbf{I}_p \otimes \left((\mathbf{B}^{-1})^T \otimes \mathbf{B} \right) \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, for all full rank matrices $\mathbf{B}$ of the form $\mathbf{B} = \begin{pmatrix} \mathbf{B}^{(11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{(22)} \end{pmatrix}$, where the dimensions of $\mathbf{B}^{(11)}$ and $\mathbf{B}^{(22)}$ are respectively, $d_1 \times d_1$ and $d_2 \times d_2$. This also means that no feedback between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, implies no feedback between the transformed processes $\mathbf{B}^{(11)}\mathbf{X}^{(1)}$ and $\mathbf{B}^{(22)}\mathbf{X}^{(2)}$: we apply a block-diagonal-affine transformation to the observations $\mathbf{X}_t$, i.e. $\mathbf{x} \to \mathbf{B}\mathbf{x}$, with $\mathbf{B} = \begin{pmatrix} \mathbf{B}^{(11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{(22)} \end{pmatrix}$. Since the testing problem is invariant under block-diagonal-affine transformations, classical invariance arguments in such situations suggest considering testing procedures that are invariant with respect to this group of transformations. In order to obtain invariant procedures for testing no feedback between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, the following equivariance properties of $\hat{\boldsymbol{\theta}}^{(N)}$ and $\hat{\boldsymbol{\Sigma}}$ are needed. Given an arbitrary $d \times d$ full rank matrix $\mathbf{C}$, denote by $\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{C})$ the value of $\hat{\boldsymbol{\theta}}^{(N)}$ computed from the transformed sample $\mathbf{C}\mathbf{X}_1, ..., \mathbf{C}\mathbf{X}_N$, and $\hat{\boldsymbol{\Sigma}}(\mathbf{C})$ the value of $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\mathbf{e}_1^{(N)}, ..., \mathbf{e}_N^{(N)})$ obtained from the transformed residuals $\mathbf{C}\mathbf{e}_1^{(N)}, ..., \mathbf{C}\mathbf{e}_N^{(N)}$.

(C2) The constrained estimator $\hat{\boldsymbol{\theta}}^{(N)} \in \mathcal{M}(\mathbf{Q})$ is block-diagonal-affine-equivariant, this means that for any block-diagonal full rank matrix $\mathbf{B}$, we have $\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}) = \mathbf{I}_p \otimes ((\mathbf{B}^{-1})^T \otimes \mathbf{B})\hat{\boldsymbol{\theta}}^{(N)}$. This is also equivalent to $\hat{\mathbf{A}}_j(\mathbf{B}) = \mathbf{B}\hat{\mathbf{A}}_j\mathbf{B}^{-1}$ for all $j = 1, ..., p$, where $\hat{\mathbf{A}}_j(\mathbf{B})$ is the value of $\hat{\mathbf{A}}_j$ obtained from the transformed sample $\mathbf{B}\mathbf{X}_1, ..., \mathbf{B}\mathbf{X}_N$.

(D2) The estimator $\hat{\boldsymbol{\Sigma}}$ in Assumption (D1) is block-diagonal-quasi-affine-equivariant, i.e., for any N, for all block-diagonal full rank matrix $\mathbf{B}$, $\hat{\boldsymbol{\Sigma}}(\mathbf{B}) = \hat{\boldsymbol{\Sigma}}(\mathbf{B}\mathbf{e}_1^{(N)}, ..., \mathbf{B}\mathbf{e}_N^{(N)}) = k\mathbf{B}\hat{\boldsymbol{\Sigma}}\mathbf{B}^T$, where k denotes some positive scalar that may depend on $\mathbf{B}$ and $(\mathbf{e}_1^{(N)}, ..., \mathbf{e}_N^{(N)})$.

Assumption (C2) implies that under block-diagonal-affine transformations, $\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}) \in \mathcal{M}(\mathbf{Q})$. Further, the corresponding Green matrices are such that $\mathbf{G}_u(\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B})) = \mathbf{B}\mathbf{G}_u(\hat{\boldsymbol{\theta}}^{(N)})\mathbf{B}^{-1}$ for any integer u.

Any sequence of unconstrained affine equivariant estimators can be turned into a sequence of block-diagonal-quasi-affine-equivariant constrained estimators by means of a simple projection onto $\mathcal{M}(\mathbf{Q})$.

Under Assumption (C2), the pseudo-Gaussian procedure $\phi_{\mathcal{N}}$ described in Section 3.1 is invariant under the group of block-diagonal transformations (in the sense that, the value of the test statistic obtained from the transformed sample $\mathbf{BX}_1, ..., \mathbf{BX}_N$ is the same for all block-diagonal full rank matrices $\mathbf{B}$). They are of course distribution-free. However, they are not even asymptotically invariant under continuous monotone radial transformations (since they are not measurable with respect to the maximal invariant).

Assumption (D2) is satisfied by the empirical covariance matrix which is affine-equivariant ($\hat{\mathbf{\Sigma}}_E(\mathbf{B}) = \mathbf{B}\hat{\mathbf{\Sigma}}_E\mathbf{B}^T$ for all full rank matrix $\mathbf{B}$). However, a more robust and quasi-affine-equivariant (then satisfying also (D2)) estimator could be used, such as the one proposed by Tyler (1987). Tyler's scatter matrix estimator is defined by $\hat{\mathbf{\Sigma}}_T = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}$, where for any N-uple of d-dimensional vectors of residuals $\mathbf{e}^{(N)} = (\mathbf{e}_1^{(N)}, ..., \mathbf{e}_N^{(N)})$, $\mathbf{C} := \mathbf{C}(\mathbf{e}^{(N)})$ is the unique (for $N > d(d-1)$) upper triangular $d \times d$ matrix with positive diagonal elements and with one in the upper left corner that satisfies

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{Ce}_i^{(N)}}{\left\|\mathbf{Ce}_i^{(N)}\right\|}\right)\left(\frac{\mathbf{Ce}_i^{(N)}}{\left\|\mathbf{Ce}_i^{(N)}\right\|}\right)^T = \frac{1}{d}\mathbf{I}_d.$$

When testing for causality direction ($\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$), the equivariance property of ($\hat{\boldsymbol{\theta}}^{(N)}$ and $\hat{\mathbf{\Sigma}}$) should be compatible with the null hypothesis to be tested. Indeed, $\mathcal{H}_0^{(12)} : \mathbf{X}^{(1)} \not\Rightarrow \mathbf{X}^{(2)}$ is invariant under block-upper-triangular-affine transformations, i.e., the group of affine transformations $\mathbf{x} \to \mathbf{Bx}$, where $\mathbf{B}$ is a $d \times d$ full rank matrix of the form $\mathbf{B} = \begin{pmatrix} \mathbf{B}^{(11)} & \mathbf{B}^{(12)} \\ \mathbf{0} & \mathbf{B}^{(22)} \end{pmatrix}$. Similarly, $\mathcal{H}_0^{(21)} : \mathbf{X}^{(2)} \not\Rightarrow \mathbf{X}^{(1)}$ is invariant under block-lower-triangular-affine transformations, i.e., the group of affine transformations $\mathbf{x} \to \mathbf{Bx}$, where $\mathbf{B}$ is a $d \times d$ full rank matrix $d \times d$ of the form $\mathbf{B} = \begin{pmatrix} \mathbf{B}^{(11)} & \mathbf{0} \\ \mathbf{B}^{(21)} & \mathbf{B}^{(22)} \end{pmatrix}$. Now, if $\mathcal{H}_0^{(12)} : \mathbf{X}^{(1)} \not\Rightarrow \mathbf{X}^{(2)}$ is the null hypothesis of interest, the equivariance of $\hat{\boldsymbol{\theta}}^{(N)}$ and $\hat{\mathbf{\Sigma}}$ under block-upper-triangular-affine transformations will be needed in Assumptions (C2) and (D2) to obtain invariant procedures. However, the equivariance property of $\hat{\boldsymbol{\theta}}^{(N)}$ is not satisfied by the usual estimators (constrained estimators as well as by means of a simple projection onto $\mathcal{M}(\mathbf{Q}^{(12)})$ of unconstrained estimators). So, for the problem of testing causality direction $\mathcal{H}_0^{(12)}$ (resp. $\mathcal{H}_0^{(12)}$), we are not able to construct procedures that are invariant under block-upper-triangular (resp. block-lower-triangular) affine transformations.

4.2 Optimal rank-based tests

The nonparametric (signed rank) J-score versions of the cross-covariance matrices (2.7) are

$$\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta}) := (N-i)^{-1}\left(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}}\right)^T\left(\sum_{t=i+1}^{N} J_1\left(\frac{\hat{R}_t(\boldsymbol{\theta})}{N+1}\right) J_2\left(\frac{\hat{R}_{t-i}(\boldsymbol{\theta})}{N+1}\right)\hat{\mathbf{U}}_t^{(N)}(\boldsymbol{\theta})\left(\hat{\mathbf{U}}_{t-i}^{(N)}(\boldsymbol{\theta})\right)^T\right)\left(\hat{\mathbf{\Sigma}}^{\frac{1}{2}}\right)^T, \tag{4.1}$$

where the scores functions J_1 and J_2 satisfy the following assumption:

(E1) The score functions J_l: $]0,1[\to\mathbb{R}$, $l=1,2$, are continuous differences of two monotone increasing functions and satisfy $\mathrm{E}[J_l^2(U)] = \int_0^1 J_l^2(u)du < \infty$ $(l=1,2)$.

The scores functions yielding locally and asymptotically optimal procedures, as we shall see, are of the form $J_1 = \varphi_{f_*}\circ\tilde{F}_*^{-1}$ and $J_2 = \tilde{F}_*^{-1}$ for some radial density f_*. Therefore, Assumption (E1) becomes an assumption on f_* which is the following.

(E1') The radial density f_* is such that φ_{f_*}, is the continuous difference of two monotone increasing functions, $\mu_{d+1;f_*} = \int_0^\infty r^{d+1}f_*(r)dr < \infty$, and $\int_0^\infty[\varphi_{f_*}(r)]^2 r^{d-1}f_*(r)dr < \infty$.

The simplest scores are the constant ones $(J_1(u) = J_2(u) = 1)$ that yield multivariate sign cross-covariance matrices. The linear scores $(J_1(u) = J_2(u) = u)$ yield cross-covariances of the Spearman type. The score functions yielding asymptotically locally optimal procedures, under radial density f_*, are $J_1(u) = \varphi_{f_*}\circ\tilde{F}_*^{-1}(u)$ and $J_2(u) = \tilde{F}_*^{-1}(u)$. The most familiar example is that of the van der Waerden scores, associated with the normal radial density $(f_* = exp(-r^2/2))$ yielding the van der Waerden cross-covariance matrices

$$\hat{\mathbf{\Gamma}}_{i,W}^{(N)}(\boldsymbol{\theta}) := (N-i)^{-1}\left(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}}\right)^T\left(\sum_{t=i+1}^{N}\sqrt{\Psi_d^{-1}\left(\frac{\hat{R}_t(\boldsymbol{\theta})}{N+1}\right)}\sqrt{\Psi_d^{-1}\left(\frac{\hat{R}_{t-i}(\boldsymbol{\theta})}{N+1}\right)}\hat{\mathbf{U}}_t^{(N)}(\boldsymbol{\theta})\left(\hat{\mathbf{U}}_{t-i}^{(N)}(\boldsymbol{\theta})\right)^T\right)\left(\hat{\mathbf{\Sigma}}^{\frac{1}{2}}\right)^T,$$

where $\Psi_d(.)$ stands for the chi-square distribution function with d degrees of freedom. Another classical example is the scores associated with the double-exponential radial density $(f_* = exp(-r))$ yielding the Laplace scores $J_1(u) = 1$ and $J_2(u) = \kappa_d^{-1}(u)$, where $\kappa_d(u) = \frac{\Gamma_d(u)}{\Gamma(d)}$, with $\Gamma_d(u) = \int_0^u r^{d-1}exp(-r)dr$ and $\Gamma(d) = \int_0^{+\infty} r^{d-1}exp(-r)dr = (d-1)\times ... \times 1 = (d-1)!$, stands for the incomplete gamma function.

Let $\hat{\mathbf{T}}_J^{(N)}(\boldsymbol{\theta})$ be the vector of all J-score cross-covariance matrices

$$\hat{\mathbf{T}}_J^{(N)}(\boldsymbol{\theta}) = \left((N-1)^{\frac{1}{2}}\left(\mathrm{vec}\hat{\mathbf{\Gamma}}_{1,J}^{(N)}(\boldsymbol{\theta})\right)^T, ..., (N-i)^{\frac{1}{2}}\left(\mathrm{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta})\right)^T, ..., \left(\mathrm{vec}\hat{\mathbf{\Gamma}}_{N-1,J}^{(N)}(\boldsymbol{\theta})\right)^T\right)^T$$

and denote $\hat{\mathbf{\Delta}}_J(\boldsymbol{\theta}) := \mathbf{M}^{(N)}(\boldsymbol{\theta})\hat{\mathbf{T}}_J^{(N)}(\boldsymbol{\theta})$. Now, substituting $\hat{\boldsymbol{\theta}}^{(N)}$ for $\boldsymbol{\theta}$, and let us denote

$$\tilde{\mathbf{\Gamma}}_{i,J}^{(N)} = \hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}) = \frac{1}{N-i}\left(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}}\right)^T \left(\sum_{t=i+1}^{N} J_1\left(\frac{\tilde{R}_t}{N+1}\right) J_2\left(\frac{\tilde{R}_{t-i}}{N+1}\right)\tilde{\mathbf{U}}_t\tilde{\mathbf{U}}_{t-i}^T\right)\left(\hat{\mathbf{\Sigma}}^{\frac{1}{2}}\right)^T,$$

the resulting nonparametric (*aligned* ranks and signs) J-score cross-covariance matrix at lag i and by $\tilde{\mathbf{T}}_J^{(N)}$ the vector of *aligned* J-score cross-covariance matrices, i.e.,

$$\tilde{\mathbf{T}}_J^{(N)} = \hat{\mathbf{T}}_J^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}) = \left((N-1)^{\frac{1}{2}}\left(\mathrm{vec}\tilde{\mathbf{\Gamma}}_{1,J}^{(N)}\right)^T, ..., (N-i)^{\frac{1}{2}}\left(\mathrm{vec}\tilde{\mathbf{\Gamma}}_{i,J}^{(N)}\right)^T, ..., \left(\mathrm{vec}\tilde{\mathbf{\Gamma}}_{N-1,J}^{(N)}\right)^T\right)^T.$$

Define the *aligned* J-score version of the central sequence by $\tilde{\mathbf{\Delta}}_J := \mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\tilde{\mathbf{T}}_J^{(N)}$, and let $\mathbf{\Omega}_{\mathbf{\Sigma},J}(\boldsymbol{\theta}) = \frac{1}{d^2}\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}$, where $\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}$ is defined in (2.11), and denote by $\tilde{\mathbf{\Omega}}_J = \mathbf{\Omega}_{\hat{\mathbf{\Sigma}},J}(\hat{\boldsymbol{\theta}}^{(N)})$. Building on the previous notations, invariant optimal rank-based tests for $\mathcal{H}_0$ (noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$) is $\tilde{\phi}_J := I\left[\tilde{Q}_J > \chi^2_{r,1-\alpha}\right]$, where

$$\tilde{Q}_J := \tilde{\mathbf{\Delta}}_J^T\left[\tilde{\mathbf{\Omega}}_J^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\tilde{\mathbf{\Omega}}_J\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\tilde{\mathbf{\Delta}}_J. \tag{4.2}$$

Similarly, optimal rank-based tests for the null hypothesis $\mathcal{H}_0^{(12)}$ under which $\mathbf{X}^{(1)}$ does not Granger cause $\mathbf{X}^{(2)}$ ($\mathbf{X}^{(1)} \not\Rightarrow \mathbf{X}^{(2)}$) is $\tilde{\phi}_J^{(12)} := I\left[\tilde{Q}_J^{(12)} > \chi^2_{r_1,1-\alpha}\right]$, where

$$\tilde{Q}_J^{(12)} = \tilde{\mathbf{\Delta}}_J^T\left[\tilde{\mathbf{\Omega}}_J^{-1} - \mathbf{Q}^{(12)}\left((\mathbf{Q}^{(12)})^T\tilde{\mathbf{\Omega}}_J\mathbf{Q}^{(12)}\right)^{-1}(\mathbf{Q}^{(12)})^T\right]\tilde{\mathbf{\Delta}}_J. \tag{4.3}$$

A sequence of locally asymptotically most stringent α-level tests for the null hypothesis $\mathcal{H}_0^{(21)}$ under which $\mathbf{X}^{(2)}$ does not Granger cause $\mathbf{X}^{(1)}$ ($\mathbf{X}^{(2)} \not\Rightarrow \mathbf{X}^{(1)}$), is given by $\tilde{\phi}_J^{(21)} := I\left[\tilde{Q}_J^{(21)} > \chi^2_{r_1,1-\alpha}\right]$, where

$$\tilde{Q}_J^{(21)} = \tilde{\mathbf{\Delta}}_J^T\left[\tilde{\mathbf{\Omega}}_J^{-1} - \mathbf{Q}^{(21)}\left((\mathbf{Q}^{(21)})^T\tilde{\mathbf{\Omega}}_J\mathbf{Q}^{(21)}\right)^{-1}(\mathbf{Q}^{(21)})^T\right]\tilde{\mathbf{\Delta}}_J. \tag{4.4}$$

The scores functions yielding locally and asymptotically optimal procedures, as we shall see, are of the form $J_1 = \varphi_{f_*} \circ \tilde{F}_*^{-1}$ and $J2 = \tilde{F}_*^{-1}$ for some radial density f_*. The corresponding statistics will be denoted by $\tilde{\phi}_{f_*}$, $\tilde{\phi}_{f_*}^{(12)}$ and $\tilde{\phi}_{f_*}^{(21)}$ instead of $\tilde{\phi}_J$, $\tilde{\phi}_J^{(12)}$, and $\tilde{\phi}_J^{(21)}$. Note that our optimal tests are (locally and asymptotically) most stringent, not uniformly most powerful—so that they can be dominated, for particular alternatives, by their competitors.

In this paper, we have used *pseudo-Mahalanobis signs and ranks*. However, any combination of a concept of multivariate signs (either Mahalanobis signs, pseudo-Mahalanobis signs, or absolute interdirections (Randles, 1989)) with a concept of multivariate ranks (Mahalanobis ranks, pseudo-Mahalanobis ranks, or lift-interdirection ranks (Oja and Paindaveine, 2005)) may be considered and yields the same asymptotic results. However, when absolute interdirections is used with any type of

ranks, the resulting test statistics $\tilde{Q}_J$ will be only asymptotically invariant under block-diagonal affine transformations.

Before stating the main result of this paper, we need some more notations. Let $\eta_d(J,f) = \int_0^1 J(u)\tilde{F}^{-1}(u)du$, $\pi_d(J,f) = \int_0^1 J(u)\varphi_f \circ \tilde{F}^{-1}(u)du$, and $\mathrm{E}[J^2(U)] = \int_0^1 J^2(u)du$, where J denotes a score function defined on $]0,1[$. When J is the score associated with some radial density g ($J_1 = \varphi_g \circ \tilde{G}^{-1}$ and $J_2 = \tilde{G}^{-1}$), we write $\eta_d(g,f)$ and $\pi_d(g,f)$ for $\eta_d(\tilde{G}^{-1},f)$ and $\pi_d(\varphi_g \circ \tilde{G}^{-1},f)$, respectively. Also for simplicity, we write $\eta_d(f)$ and $\pi_d(f)$ for $\eta_d(f,f)$ and $\pi_d(f,f)$.

In the following theorem, we give the optimal testing procedures for noncausality in VAR models, their invariance and distribution freeness features, as well as their local powers and optimality properties. The proof is given in the Appendix.

Theorem 4.1 *Assume that Assumptions (A1), (B1), (B2), (B3'), (C1), (D1), and (E1) hold. Consider the sequence of aligned rank tests $\tilde{\phi}_J$ (resp. $\tilde{\phi}_J^{(12)}$ or $\tilde{\phi}_J^{(21)}$) that rejects the null hypothesis $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$) whenever $\tilde{Q}_J$ (resp. $\tilde{Q}_J^{(12)}$ or $\tilde{Q}_J^{(21)}$) exceeds the $1-\alpha$ quantile $\chi^2_{r,1-\alpha}$ of a chi-square variable with $r = 2pd_1d_2$ (resp $r_1 = pd_1d_2$ or $r_2 = pd_1d_2$) degrees of freedom. Then*

(i) $\tilde{Q}_J$ (resp. $\tilde{Q}_J^{(12)}$ or $\tilde{Q}_J^{(21)}$) is asymptotically chi-square with $r = 2pd_1d_2$ (resp. $r_1 = pd_1d_2$ or $r_2 = pd_1d_2$) degrees of freedom under $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$).

(ii) $\tilde{Q}_J$, (resp. $\tilde{Q}_J^{(12)}$ or $\tilde{Q}_J^{(21)}$) is asymptotically invariant with respect to the group of continuous monotone radial transformations. Further, under the equivariance Assumptions (C2) and (D2), $\tilde{Q}_J$ is block-diagonal-affine-invariant.

(iii) The sequence $\tilde{Q}_J$ (resp. $\tilde{Q}_J^{(12)}$ or $\tilde{Q}_J^{(21)}$) is asymptotically noncentral chi-square with $r = 2pd_1d_2$ (resp. $r_1 = pd_1d_2$ or $r_2 = pd_1d_2$) degrees of freedom and with noncentrality parameter $\tilde{\psi}^2_{J,f} = \tilde{\psi}^2_{J,f}(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma}) = \frac{1}{d^2}\frac{\pi_d^2(J_1,f)\eta_d^2(J_2,f)}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$, where $\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ is given in (3.6) (resp. $\tilde{\psi}^2_{1,J,f} = \tilde{\psi}^2_{1,J,f}(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma}) := \frac{1}{d^2}\frac{\pi_d^2(J_1,f)\eta_d^2(J_2,f)}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\delta^{(12)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$, or $\tilde{\psi}^2_{2,J,f} = \tilde{\psi}^2_{2,J,f}(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma}) := \frac{1}{d^2}\frac{\pi_d^2(J_1,f)\eta_d^2(J_2,f)}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\delta^{(21)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$, where $\delta^{(12)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ and $\delta^{(21)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ are given in (3.8) and (3.10), respectively), under the sequence of local alternatives, $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau},\boldsymbol{\Sigma},f)$, with $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$ (resp. $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q}^{(12)})$ or $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q}^{(21)})$).

(iv) For any radial density f_* satisfying Assumptions (B1), (B2), (B3') and (E1'), the test $\tilde{\phi}_{f_*}$ (resp. $\tilde{\phi}^{(12)}_{f_*}$ or $\tilde{\phi}^{(12)}_{f_*}$) is locally asymptotically most stringent for $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$) against $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f_*)$ (resp. $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0^{(12)}}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f_*)$ or $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0^{(21)}}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f_*)$).

Remark 4.1 *For the problem of testing for causality directions, the tests $\tilde{\phi}^{(12)}_{f_*}$ (resp. $\tilde{\phi}^{(21)}_{f_*}$) are not invariant under block upper (resp. lower) triangular-affine transformations. Indeed, the usual*

estimators do not allow for constructing $\hat{\boldsymbol{\theta}}^{(N)}$ such that Assumption (C2) holds. However, note that if there exist a sequence $\hat{\boldsymbol{\theta}}^{(N)}$ such that (C2) holds, then the tests $\tilde{Q}_J^{(12)}$ (resp. $\tilde{Q}_J^{(21)}$) are block-upper-triangular-affine-invariant (resp. block-lower-triangular-affine-invariant).

Remark 4.2 *Using similar arguments, optimal rank-based tests could be constructed for the problem of testing noncausality when the global process is a vector moving average process with known order VMA(q). Indeed, noncausality (in one direction or in both directions) between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ reduces to the hypothesis that the parameter $\boldsymbol{\theta}$ of interest lies in some linear subspace of $\mathbb{R}^K$. The matrices $\mathbf{Q}$, $\mathbf{Q}^{(12)}$ and $\mathbf{Q}^{(21)}$ do not change. However, the central sequence and the information matrix must be adapted to the VMA(q) context.*

4.3 Asymptotic relative efficiencies

In this Section, we turn to asymptotic relative efficiencies (ARE) of the rank-based tests $\tilde{\phi}_J$ with respect to their Gaussian counterparts $\phi_{\mathcal{N}}$. The powers of the pseudo-Gaussian test will serve as a benchmark for computing the asymptotic relative efficiencies. The distribution of the test statistics $\tilde{Q}_J$, $\tilde{Q}_J^{(12)}$ and $\tilde{Q}_J^{(21)}$ under local alternatives are noncentral chi-square, with noncentrality parameters that depend on the order p of the VAR model, the dimensions d_1 and d_2 of the processes $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, the underlying density f, the perturbation $\boldsymbol{\tau}$, the VAR parameters $\boldsymbol{\theta}$, the scatter matrix $\boldsymbol{\Sigma}$ and on the chosen score function J. On the other hand, the Gaussian counterparts are also noncentral chi-square under local alternatives, with the same degrees of freedom but with different noncentrality parameters. Computing the ratios of the noncentrality parameters in the asymptotic distributions under local alternatives of $\tilde{\phi}_J$ (resp. $\tilde{\phi}_J^{(12)}$ or $\tilde{\phi}_J^{(21)}$) with respect to $\phi_{\mathcal{N}}$ (resp. $\phi_{\mathcal{N}}^{(12)}$ or $\phi_{\mathcal{N}}^{(21)}$) yields the ARE of these tests with respect their parametric Gaussian counterparts. The following result follows from Theorems 3.1 and 4.1 and from the fact that $\omega_d(f) = \frac{1}{d^2}\left(\int_0^1 \tilde{F}^{-1}(u)\varphi_f \circ \tilde{F}^{-1}(u)du\right)^2 = 1$ under Assumptions (B1), (B2) and (B3').

Theorem 4.2 *Assume that Assumptions (A1), (B1), (B2), (B3'), (C1), (D1) and (E1) hold. Then the asymptotic relative efficiency of $\tilde{\phi}_J$ (resp. $\tilde{\phi}_J^{(12)}$ or $\tilde{\phi}_J^{(12)}$) with respect to $\phi_{\mathcal{N}}$ (resp. $\phi_{\mathcal{N}}^{(12)}$ or $\phi_{\mathcal{N}}^{(12)}$), under radial density f, is*

$$\mathrm{ARE}_{d,f}(\tilde{\phi}_J, \phi_{\mathcal{N}}) = \mathrm{ARE}_{d,f}^{(12)}(\tilde{\phi}_J^{(12)}, \phi_{\mathcal{N}}^{(12)}) = \mathrm{ARE}_{d,f}^{(21)}(\tilde{\phi}_J^{(21)}, \phi_{\mathcal{N}}^{(21)}) = \frac{1}{d^2}\frac{\pi_d^2(J_1, f)\eta_d^2(J_2, f)}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}.$$

Note that the asymptotic relative efficiency does not depend on p, $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\tau}$. It depends only on the underlying radial density f, the score functions J_1 and J_2, and the dimensions d_1 and d_2 through

the dimension d of the global process. From Theorem 4.2 and the generalized Chernoff-Savage result obtained in Proposition 6 of Hallin and Paindaveine (2002c), it follows that the asymptotic relative efficiencies of our procedures with respect to the Gaussian procedure ($\mathrm{ARE}_{d,f}(\tilde{\phi}_J, \phi_{\mathcal{N}})$), when the van der Waerden scores are used (i.e., $J_1 = J_2 = \sqrt{\Psi_d^{-1}(u)}$, where Ψ_d stands for the chi-square distribution function with d degrees of freedom) are always larger than or equal to one, irrespective of the radial density f and the dimension of the global process d. The equality holds if and only if f is normal. Moreover, it appears that the advantage of the van der Waerden procedure over the Gaussian procedure, in the case of the multivariate Student density, grows with the dimension of the global process d and with the weight of the tail of the radial density (an ARE value of 1.458 is reached for 4-variate Student density with 3 degrees of freedom).

Another interesting result follows from Theorem 4.2 and the multivariate Hodges-Lehmann result (see Proposition 10 in Hallin and Paindaveine, 2004a). More precisely, the Spearman-type procedure exhibits excellent asymptotic efficiency properties, with respect to the Gaussian procedure, especially for relatively small dimensions d. Indeed, we have that $\underset{f}{inf}\mathrm{ARE}_{d,f}(\tilde{\phi}_{SP}, \phi_{\mathcal{N}})$, where the infimum is taken over all radial densities f satisfying Assumptions (B2) and (B3) and $\tilde{\phi}_{SP}$ stands for our procedure when Spearman-type scores are used, is monotonically decreasing in d and tends to $9/16 = 0.5625$ as $d \to +\infty$. In the case of testing causality between two univariate time series ($d = 2$), this lower bound is equal to 0.913.

5 The bivariate VAR(1) case and some Monte Carlo results

As an illustration, and in order to investigate the finite sample performance (size, power and robustness) of our tests, we conducted a Monte Carlo investigation with the bivariate autoregressive model of order 1 ($p = 1$ and $d = 2$). To ease the presentation, the notation was adapted to this particular context and the global process is denoted by $\mathbf{X}_t = (X_t, Y_t)^T$. It is characterized by the following equation

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} \phi & \gamma_{12} \\ \gamma_{21} & \theta \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix},$$

and the vector of parameters is $\boldsymbol{\theta} = (\phi, \gamma_{21}, \gamma_{12}, \theta)^T$.

Here again we only focus on testing for no feedback between X and Y, i.e., $\gamma_{12} = \gamma_{21} = 0$. We consider four particular cases of the following data generating equation

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} 0.5 & \gamma_{12} \\ \gamma_{21} & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} u_t \\ v_t \end{pmatrix}, \tag{5.1}$$

where the bivariate spherical density of the noise $(u_t, v_t)^T$ is a bivariate Normal or Student distributions with zero mean and $\mathbf{I}_2$ scatter matrix.

- *Experiment A:* $\gamma_{12} = \gamma_{21} = 0$. Under this experiment, there is no feedback between X and Y; this experiment allows for checking the validity of asymptotic distributions under the null;
- *Experiment B:* $\gamma_{12} = 0$ and $\gamma_{21} = 0.1m$, with $m = 1, 2, 3$. Under this alternative, X causes Y and Y does not cause X;
- *Experiment C:* $\gamma_{21} = 0$ and $\gamma_{12} = 0.1m$, with $m = 1, 2, 3$. Under this alternative, Y causes X and X does not cause Y;
- *Experiment D:* $\gamma_{12} = 0.1m$ and $\gamma_{21} = 0.1m$, with $m = 1, 2, 3$. Under this alternative, X causes Y and Ycauses X.

For each of these four experiments, and for each of the following standardized (mean zero and identity scatter matrix) four densities: bivariate normal ($\mathcal{N}$) and Student ($\mathcal{T}_\nu$) with $\nu = 3, 6, 9$ degrees of freedom, 1000 replications of a bivariate iid white noise $(u_t, v_t)^T$ of length 300 were generated from the chosen density. These sequences of observations $(u_t, v_t)^T$ were plugged into the various models, yielding 1000 replications, of length 300, of the process $(X_t, Y_t)^T$. Initial values X_0 and Y_0 were put to zero. In order to prevent starting values to affect the stationarity of the generated series, only the subseries of length $N = 100$ (respectively, $N = 200$) resulting from dropping the 200 (respectively, 100) first observations, were considered for the analysis.

¿From a practical point of view, it is natural to inquire about the finite sample properties of the proposed test statistics, in particular their exact level and power whether or not there are outliers in the series under study. For each of these four experiments and four each replication, the following two scenarios were considered.

Scenario 1: No contamination occurred in both generated series.

Scenario 2: Outliers occurred in both generated series. Six type of outliers are considered: $\mathcal{O}_1$, $\mathcal{O}_2$, and $\mathcal{O}_3$ are observation outliers and $\mathcal{I}_1$, $\mathcal{I}_2$, and $\mathcal{I}_3$ are innovation outliers .

$\mathcal{O}_1$ (observation outliers): Outliers occurred in X_t and Y_t: observations X_t and Y_t, were replaced respectively, with $X_t + 20$ and $Y_t + 20$, at $t = 220$, 230.

$\mathcal{O}_2$ (observation outliers): Outliers occurred in X_t and Y_t: we added 20 to observations X_{220}, X_{260}, X_{270}, Y_{220}, Y_{240} and Y_{250}.

$\mathcal{O}_3$ (observation outliers): Outliers occurred in X_t and Y_t: observations X_{220} and Y_{220}, were replaced respectively, with $X_{220}+20$ and $Y_{220}+20$.

$\mathcal{I}_1$ (innovation outliers): Outliers occurred in u_t and v_t: innovations u_t and v_t were replaced respectively, with $5u_t$ and $5v_t$ for $t = 210, 220, 230, 240, 250, 260, 270, 280, 290$.

$\mathcal{I}_2$ (innovation outliers): Outliers occurred in u_t and v_t: innovations u_t and v_t were replaced respectively, with $5u_t$ and $5v_t$ for $t = 210, 220, 230, 240$, and with $-5u_t$ and $-5v_t$ for $t = 250, 260, 270, 280, 290$.

$\mathcal{I}_3$ (innovation outliers): Outliers occurred in u_t and v_t: innovations u_{290} and v_{290}, were replaced respectively, with $u_{290}+10$ and $v_{290}+10$. Innovations u_{211} and u_{251} were replaced respectively, with $20u_{211}$ and $u_{251}+10$. Innovations v_{220} and v_{276} were replaced respectively, with $20v_{220}$ and $v_{276}-10$.

For each scenario and for each of the replications thus obtained, under experiments A through D, the Yule-Walker method yields a sequence of unconstrained $\sqrt{N}$-consistent estimator of $\boldsymbol{\theta}$

$$\tilde{\boldsymbol{\theta}} = \mathrm{vec}\tilde{\mathbf{A}}_1, \quad \text{where} \quad \tilde{\mathbf{A}}_1 = \left(\sum_{t=2}^{N} \mathbf{X}_t \mathbf{X}_{t-1}^T\right)\left(\sum_{t=2}^{N} \mathbf{X}_t \mathbf{X}_t^T\right)^{-1}.$$

The assumption of $\sqrt{N}$-consistency on the unconstrained estimators is satisfied by the other classical estimators (least squares, maximum likelihood,...). This estimator $\tilde{\boldsymbol{\theta}}$ can be turned into a constrained and $\sqrt{N}$-consistent $\hat{\boldsymbol{\theta}}^{(N)}$ by means of a simple projection onto $\mathcal{M}(\mathbf{Q})$, i.e., $\hat{\boldsymbol{\theta}}^{(N)} = \mathbf{Q}\left(\mathbf{Q}^T\mathbf{Q}\right)^{-1}\mathbf{Q}^T\tilde{\boldsymbol{\theta}}$. The matrix Q in this problem is given by $Q^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

¿From the sequences of estimators $\tilde{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^{(N)}$, we computed

- the Wald test statistic Q^* (see Lütkepohl, 1991, Section 3.6; Boudjellaba, Dufour and Roy, 1992);
- the pseudo-Gaussian statistics $Q_{\mathcal{N}}$ of Theorem 3.1,
- the statistics $\tilde{Q}_J^{(E)}$ which correspond to $\tilde{Q}_J$ of Theorem 4.1 obtained from *Mahalanobis signs and ranks* (the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}_E$ is used). Four type of scores are used: constant, Spearman, Laplace and van der Waerden. The corresponding statistics are denoted $\tilde{Q}_S^{(E)}$ (S for Sign test), $\tilde{Q}_{SP}^{(E)}$, $\tilde{Q}_L^{(E)}$, $\tilde{Q}_{vW}^{(E)}$.
- the statistics $\tilde{Q}_J^{(T)}$ which correspond to $\tilde{Q}_J$ obtained from *pseudo-Mahalanobis signs and ranks* (Tyler estimator of the scatter $\hat{\boldsymbol{\Sigma}}_T$ is used). These versions are supposed to allow for a better

control against outliers in the data. Again, four type of scores are used: constant, Spearman, Laplace and van der Waerden. The corresponding statistics are denoted $\tilde{Q}_S^{(T)}$, $\tilde{Q}_{SP}^{(T)}$, $\tilde{Q}_L^{(T)}$, $\tilde{Q}_{vW}^{(T)}$.

For each replication, these statistics were compared with their asymptotic critical values. Rejection frequencies under Scenario 1 are reported in Tables 1, 2, and 3, for two series length ($N = 100$ and $N = 200$), at the nominal α-value 0.05, and for the various densities considered. Under Scenario 2, we narrowed down the analysis to experiments A and D with a Gaussian noise and the series length $N = 100$. Rejection frequencies under Experiment A, Scenario 2, at the nominal α-value 0.05, are reported in Table 4. The results in this table very clearly indicate that under Scenario 2, the levels of the tests are quite far from 0.05. Therefore, to compare the performances of these tests under Experiment D, Scenario 2, we used the empirical critical values obtained from the corresponding 1000 replications generated under Experiment A with the same scenario. The rejection frequencies based on these empirical critical values are reported in Tables 5. The standard error of the empirical levels in Tables 1 and 4 is 0.0069 and at the 5% significance level, we almost reject the hypothesis that the true level is 0.05 if the rejection frequency is outside the interval]0.0365, 0.0635[.

N	f	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
100	N	.064	.058	.063	.063	.048	.057	.064	.058	.054	.055
	$\mathcal{T}_3$	.063	.057	.055	.059	.051	.061	.054	.062	.055	.055
	$\mathcal{T}_6$	.063	.054	.054	.055	.053	.053	.052	.054	.056	.054
	$\mathcal{T}_9$	.057	.051	.048	.057	.050	.049	.041	.055	.050	.045
200	N	.057	.052	.055	.049	.047	.048	.050	.046	.047	.047
	$\mathcal{T}_3$	.055	.056	.048	.042	.050	.044	.054	.043	.048	.048
	$\mathcal{T}_6$	.040	.038	.045	.042	.045	.036	.044	.042	.047	.037
	$\mathcal{T}_9$	.063	.057	.048	.057	.048	.052	.048	.053	.049	.049

Table 1. Rejection frequencies in 1000 replications of Experiment A under Scenario 1 for the Wald test, the Gaussian test, and the optimal rank tests based either on the empirical covariance matrix or on Tyler estimator, using constant, Spearman, Laplace and van der Waerden scores, at the significance level $\alpha = 0.05$, for various densities f of the innovations, and for series lengths $N = 100$ and 200.

Discussion of the level and power under Scenario 1

Rejection frequencies for Experiment A under Scenario 1 are reported in Table 1. For all series lengths and for the various densities of the innovations, the rejection frequencies are all within the 5% significance limits except two values that are between 2 and 3 standard errors from 5%.

Table 2 reports the rejection frequencies (based on the asymptotic critical values), for Experiments B and C, at probability level $\alpha = 0.05$. Inspection of that table reveals an excellent overall performance of all rank-based procedures considered. The figures in that table also indicate that the performance

of the rank tests either based on empirical covariance matrix or on a robustified version given by Tyler estimator are similar. The sign test seems to be the weakest among the nonparametric tests. Under the Gaussian density, Wald test is doing slightly better than the others. However, as N increases, we observe that the rejection frequencies of Wald test become closer to those of the pseudo Gaussian and van der Waerden tests, which confirms the relevance of the asymptotic theory developed in this paper. For instance, under Experiment B with $m = 3$ and $N = 200$, the latter tests (Q^*, $Q_{\mathcal{N}}$, $\tilde{Q}_{vW}^{(E)}$ and $\tilde{Q}_{vW}^{(T)}$) yield the same empirical power of .994. Under the Student $\mathcal{T}_3$ density (except for $N = 100$ and $m = 1$), van der Waerden, Laplace and Spearman tests slightly dominate the Wald test. However, when the degrees of freedom ν increase, Wald test does slightly better and the rejection frequencies become closer to those obtained under a Gaussian density. Similar conclusions can be drawn from Table 3 which reports the rejection frequencies under Experiment D. However, the power of each test is slightly higher than under Experiment B or C, which is not surprising.

Discussion of the level and power under Scenario 2

The rejection frequencies for Experiment A under Scenario 2 are reported in Table 4. The rejection frequencies very clearly show that Wald and Gaussian tests are very sensitive to the presence of outliers, irrespective of their type. Indeed, the latter two tests appear to be seriously biased; their rejection frequencies are either very high (around 0.90) or very low (around 0.01).

The rank tests based either on the empirical covariance matrix or on Tyler estimator are resistant to innovation outliers. Indeed, all the corresponding rejection frequencies are within the 5% significance limits except one (0.036). With observation outliers, the situation is quite different. The tests based on Tyler estimator better resist but we cannot say that the level is satisfactorily controlled since all rejection frequencies except two are outside the 5% significance limits. There is a tendancy to overreject (4 frequencies out of 12 are greater than 0.10). The use of the empirical covariance matrix is clearly inappropriate in that situation since all four tests are strongly biased, especially those based on Spearman, Laplace and van der Waerden scores.

Rejection frequencies based on empirical critical values for Experiment D under Scenario 2 are reported in Table 5. It is immediately seen that with observation outliers, Wald test, the Gaussian test and the rank tests based on the empirical covariance matrix dramatically underreject the null hypothesis, they are uniformly weaker than the rank tests based on Tyler estimator. On the other hand, with innovation outliers, there is at least one rank test whose power is similar to those of Wald and Gaussian tests except in the case $m = 1$ and with $\mathcal{I}_3$-type outliers. In that case, the power of the

Gaussian test is 0.786 whilst the power of the more powerful rank test $\tilde{Q}_{vW}^{(T)}$ is 0.679.

Exp	f	N	m	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
B	$\mathcal{N}$	100	1	.181	.171	.119	.170	.123	.164	.121	.163	.128	.153
			2	.516	.478	.322	.484	.355	.466	.322	.467	.354	.463
			3	.881	.869	.643	.854	.742	.855	.642	.849	.726	.844
		200	1	.296	.294	.195	.286	.226	.283	.192	.283	.232	.280
			2	.833	.820	.595	.818	.690	.816	.593	.815	.684	.811
			3	.994	.994	.928	.996	.959	.994	.926	.996	.960	.994
	$\mathcal{T}_3$	100	1	.207	.188	.140	.190	.186	.189	.145	.190	.197	.195
			2	.528	.501	.389	.546	.545	.541	.393	.555	.568	.545
			3	.815	.796	.716	.854	.858	.852	.727	.859	.866	.851
		200	1	.318	.307	.233	.336	.352	.333	.230	.333	.358	.335
			2	.790	.786	.711	.863	.883	.867	.714	.864	.883	.873
			3	.965	.966	.948	.992	.991	.992	.954	.993	.991	.991
	$\mathcal{T}_6$	100	1	.171	.152	.128	.161	.144	.157	.128	.160	.141	.148
			2	.538	.520	.376	.527	.459	.511	.383	.516	.474	.510
			3	.844	.830	.675	.843	.794	.841	.673	.831	.787	.833
		200	1	.305	.304	.224	.302	.261	.302	.224	.296	.260	.301
			2	.844	.833	.683	.833	.810	.844	.689	.835	.807	.841
			3	.987	.985	.944	.990	.986	.991	.943	.989	.985	.991
	$\mathcal{T}_9$	100	1	.199	.174	.139	.177	.147	.171	.139	.174	.148	.167
			2	.529	.495	.366	.513	.465	.486	.370	.502	.459	.482
			3	.871	.851	.660	.838	.768	.845	.672	.839	.763	.841
		200	1	.311	.296	.216	.306	.275	.299	.221	.306	.268	.298
			2	.816	.813	.653	.812	.768	.817	.651	.808	.762	.817
			3	.990	.990	.937	.987	.979	.991	.930	.987	.977	.989
C	$\mathcal{N}$	100	1	.176	.158	.122	.151	.129	.151	.120	.145	.131	.139
			2	.534	.509	.351	.517	.387	.495	.349	.507	.391	.486
			3	.864	.853	.642	.851	.711	.843	.649	.842	.700	.835
		200	1	.305	.296	.201	.292	.232	.285	.204	.287	.237	.286
			2	.565	.533	.415	.571	.573	.577	.414	.580	.581	.581
			3	.994	.993	.927	.993	.960	.992	.921	.993	.954	.988
	$\mathcal{T}_3$	100	1	.202	.187	.135	.191	.180	.185	.141	.199	.199	.195
			2	.565	.533	.415	.571	.573	.577	.414	.580	.581	.581
			3	.837	.813	.723	.871	.871	.867	.728	.877	.871	.868
		200	1	.310	.310	.223	.332	.360	.336	.235	.336	.368	.336
			2	.811	.800	.755	.873	.897	.871	.760	.880	.904	.877
			3	.964	.967	.958	.993	.997	.993	.958	.994	.998	.994
	$\mathcal{T}_6$	100	1	.202	.185	.138	.187	.172	.182	.139	.185	.173	.176
			2	.531	.505	.390	.532	.483	.522	.398	.536	.484	.518
			3	.874	.857	.684	.862	.823	.863	.676	.854	.814	.850
		200	1	.301	.282	.221	.301	.282	.299	.223	.305	.288	.301
			2	.833	.828	.684	.841	.797	.837	.678	.841	.806	.835
			3	.992	.991	.948	.994	.990	.994	.943	.994	.989	.994
	$\mathcal{T}_9$	100	1	.180	.157	.128	.155	.123	.149	.125	.147	.128	.147
			2	.548	.521	.378	.525	.460	.512	.378	.517	.454	.507
			3	.864	.852	.657	.843	.753	.835	.649	.839	.744	.830
		200	1	.316	.302	.211	.302	.254	.298	.213	.301	.257	.299
			2	.840	.829	.648	.833	.782	.831	.651	.830	.771	.829
			3	.988	.986	.932	.989	.980	.989	.934	.991	.980	.990

Table 2. Rejection frequencies in 1000 replications of Experiments B and C under Scenario 1 , for the Wald test, the Gaussian test, and the optimal rank tests based either on the empirical covariance matrix or on Tyler estimator, using constant, Spearman, Laplace and van der Waerden scores, at significance level $\alpha = 0.05$, for various densities f of the innovations, and for series lengths $N = 100$ and 200.

Exp	f	N	m	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
D	$\mathcal{N}$	100	1	.305	.290	.178	.285	.223	.282	.180	.286	.229	.277
			2	.840	.834	.648	.826	.722	.825	.647	.816	.717	.817
			3	.995	.993	.960	.993	.983	.994	.956	.991	.983	.992
		200	1	.524	.525	.364	.512	.411	.510	.361	.509	.412	.515
			2	.987	.988	.936	.985	.957	.987	.933	.984	.954	.983
			3	1.00	1.00	.999	1.00	1.00	1.00	.999	1.00	1.00	1.00
	$\mathcal{T}_3$	100	1	.347	.340	.276	.357	.359	.357	.279	.363	.372	.361
			2	.867	.863	.791	.890	.894	.889	.786	.887	.896	.883
			3	.997	.997	.989	.999	.997	.998	.988	.999	.997	.998
		200	1	.545	.547	.457	.617	.670	.626	.456	.627	.685	.635
			2	.990	.987	.977	.995	.997	.995	.977	.994	.996	.995
			3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\mathcal{T}_6$	100	1	.336	.317	.243	.314	.293	.314	.245	.322	.285	.312
			2	.878	.878	.735	.878	.834	.870	.750	.866	.830	.865
			3	.999	.999	.980	.999	.991	.998	.982	.998	.990	.998
		200	1	.552	.548	.420	.557	.547	.554	.425	.553	.547	.549
			2	.992	.991	.955	.993	.987	.993	.956	.993	.989	.993
			3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\mathcal{T}_9$	100	1	.306	.294	.240	.300	.253	.289	.233	.299	.255	.287
			2	.866	.861	.740	.854	.806	.856	.734	.854	.803	.854
			3	.996	.995	.979	.996	.991	.996	.977	.996	.990	.995
		200	1	.528	.526	.421	.537	.478	.536	.417	.536	.476	.537
			2	.991	.988	.944	.989	.983	.990	.945	.989	.985	.990
			3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3. Rejection frequencies in 1000 replications of Experiment D under Scenario 1, for the Wald test, the Gaussian test, and the optimal rank tests based either on the empirical covariance matrix or on Tyler estimator, using constant, Spearman, Laplace and van der Waerden scores, at significance level $\alpha = 0.05$, for various densities f of the innovations, and for series lengths $N = 100$ and 200.

Type	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
$\mathcal{O}_1$	.920	.931	.180	.558	.350	.538	.082	.148	.086	.174
$\mathcal{O}_2$	.000	.000	.121	.253	.146	.213	.063	.074	.041	.068
$\mathcal{O}_3$	.878	.881	.196	.533	.359	.531	.096	.114	.093	.121
$\mathcal{I}_1$	.006	.007	.050	.048	.039	.038	.057	.061	.053	.048
$\mathcal{I}_2$	.016	.013	.051	.044	.038	.036	.063	.057	.054	.053
$\mathcal{I}_3$	.002	.002	.050	.039	.041	.039	.063	.052	.048	.048

Table 4. Rejection frequencies in 1000 replications of Experiment A under Scenario 2, for the Wald test, the Gaussian test, and the optimal rank tests based either on the empirical covariance matrix or on Tyler estimator, using constant, Spearman, Laplace and van der Waerden scores, at significance level $\alpha = 0.05$, with the Gaussian density for the innovations, and $N = 100$.

m	Type	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
1	$\mathcal{O}_1$	.004	.014	.009	.008	.009	.007	.123	.065	.075	.050
	$\mathcal{O}_2$	.042	.041	.041	.009	.015	.007	.127	.111	.106	.102
	$\mathcal{O}_3$	.008	.007	.008	.004	.006	.004	.136	.124	.083	.091
	$\mathcal{I}_1$	.519	.525	.305	.443	.496	.466	.294	.429	.486	.466
	$\mathcal{I}_2$	.467	.479	.273	.465	.503	.482	.255	.429	.450	.452
	$\mathcal{I}_3$	.777	.786	.334	.584	.616	.627	.354	.639	.659	.679
2	$\mathcal{O}_1$	.000	.008	.028	.000	.000	.000	.515	.400	.352	.286
	$\mathcal{O}_2$	.076	.076	.292	.152	.177	.140	.467	.451	.424	.409
	$\mathcal{O}_3$	.000	.000	.036	.000	.000	.000	.589	.590	.486	.501
	$\mathcal{I}_1$	.985	.986	.854	.966	.978	.977	.855	.960	.972	.975
	$\mathcal{I}_2$	.972	.972	.855	.966	.976	.973	.842	.960	.971	.963
	$\mathcal{I}_3$	.999	.999	.917	.994	.988	.995	.924	.996	.991	.998
3	$\mathcal{O}_1$	.000	.003	.203	.006	.009	.005	.920	.862	.826	.771
	$\mathcal{O}_2$	.254	.249	.764	.642	.628	.609	.866	.868	.829	.835
	$\mathcal{O}_3$	.000	.000	.351	.013	.023	.008	.921	.957	.907	.928
	$\mathcal{I}_1$	1.00	1.00	.997	1.00	1.00	1.00	.995	1.00	1.00	1.00
	$\mathcal{I}_2$	1.00	1.00	1.00	.999	.999	1.00	.998	.999	.999	1.00
	$\mathcal{I}_3$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 5. Rejection frequencies (based on the empirical critical values) in 1000 replications of Experiment D under Scenario 2, for the Wald test, the Gaussian test, and the optimal rank tests based either on the empirical covariance matrix or on Tyler estimator, using constant, Spearman, Laplace and van der Waerden scores, at significance level $\alpha = 0.05$, with the Gaussian density for the innovations, and $N = 100$.

6 Conclusion

In this paper, we have introduced a new parametric (with respect to the density of the noise) test and a class of nonparametric tests for checking noncausality between two vectors of variables. The pseudo-Gaussian test is based on the Gaussian density but its validity is established for a general class of elliptically symmetric densities. The nonparametric tests are based on multivariate ranks and signs. The asymptotic properties of the proposed tests are established invoking the general LAN theory developed by Le Cam (1986). All the new tests enjoy some invariance and optimality properties and the nonparametric ones also exhibit some robustness properties with respect to outliers.

In a small Monte Carlo experiment, the finite sample properties (level and power) of the new tests were compared with the classical Wald test in a specific VAR(1) context. Two estimators of the noise covariance matrix were employed: the usual residual covariance matrix and Tyler (1987)'s robust estimator. When there are no outliers, the level of all the tests considered (Wald, pseudo-Gaussian and the eight rank-based tests) is very well controlled with series of length 100 and 200. Under the alternative of causality (in one direction or the other), the Wald and pseudo-Gaussian tests have similar power. In general, the rank-based tests are slightly less powerful but in all the situations considered,

there is always a rank-based test which is almost as powerful as Wald and pseudo-Gaussian tests. In the presence of observation or innovation outliers, both Wald and pseudo-Gaussian tests are severely affected and should not be used in practice. With innovation outliers, the levels of all rank-based tests are very well controlled. However, with observation outliers, the nonparametric tests are still biased. In general, they overreject and the bias is more important when using the empirical covariance matrix estimator.

Here, we supposed that the global process was a finite causal VAR. With similar arguments, optimal rank-based tests can also be constructed when the global process is a VMA since the noncausality constraints are still linear (see Remark 4.2).

7 Appendix

Theorems 3.1 and 4.1 follow from the following Propositions and Lemmas.

Proposition 7.1 *Assume that $\boldsymbol{\theta}$ belongs to $\boldsymbol{\Theta}_0$. Let Assumptions (B1), (D1), and (E1) hold. Then, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, $\forall i$,* as $N \to +\infty$,

$$(N-i)^{1/2}\mathrm{vec}\left(\hat{\boldsymbol{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta}) - \breve{\boldsymbol{\Gamma}}_{i,\boldsymbol{\Sigma},J,f}^{(N)}(\boldsymbol{\theta})\right) = o_p(1),$$

where,

$$\breve{\boldsymbol{\Gamma}}_{i,\boldsymbol{\Sigma},J,f}^{(N)}(\boldsymbol{\theta}) = \frac{1}{N-i}\ \left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)^T \left(\sum_{t=i+1}^{N} J_1\left(\tilde{F}(d_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}))\right) J_2\left(\tilde{F}(d_{t-i}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}))\right)\mathbf{U}_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})(\mathbf{U}_{t-i}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}))^T\right)\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^T.$$

Proof. This result is a particular case of Proposition 2, established in the general context of multivariate general linear model with VARMA errors by Hallin and Paindaveine (2005). □

Lemma 7.1 *Assume that $\boldsymbol{\theta}$ belongs to $\boldsymbol{\Theta}_0$. Let Assumptions (B1), (D1), and (E1) hold. Then, for any integer m, the vector*

$$\left((N-1)^{1/2}\mathrm{vec}\left(\breve{\boldsymbol{\Gamma}}_{1,\boldsymbol{\Sigma},J,f}^{(N)}(\boldsymbol{\theta})\right)^T, ..., (N-m)^{1/2}\mathrm{vec}\left(\breve{\boldsymbol{\Gamma}}_{m,\boldsymbol{\Sigma},J,f}^{(N)}(\boldsymbol{\theta})\right)^T\right)^T \tag{7.1}$$

is asymptotically normal, with mean $\mathbf{0}$ under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$ and with mean

$$\left(\frac{1}{d^2}\pi_d(J_1,f)\eta_d(J_2,f)\left[\mathbf{I}_m \otimes \left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}^{-1}\right)\right]\left(\mathbf{M}^{(m+1)}(\boldsymbol{\theta})\right)^T\boldsymbol{\tau}\right)$$

under $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f)$. Under both hypotheses, the covariance matrix is given by

$$\frac{1}{d^2}\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]\left[\mathbf{I}_m \otimes \left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}^{-1}\right)\right].$$

Proof. The proof follows along the same arguments as in Proposition 3.1 and Proposition 4.3 in Garel and Hallin (1995). A standard application of the classical Hoeffding-Robins central-limit result for m-dependent sequence leads to the asymptotic distribution of (7.1) under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$. The joint distribution, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, of (7.1) and the log-likelihood ratio $\Lambda^{(N)}_{\boldsymbol{\theta}^{(N)}/\boldsymbol{\theta}}\left(\mathbf{X}^{(N)}\right)$ decomposition given in Proposition 2.2, follows also from the same arguments. Application of Le Cam's third Lemma then yields the asymptotic normality under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta}^{(N)}, \boldsymbol{\Sigma}, f)$. The details are left to the reader. □

Lemma 7.2 *Assume that Assumptions (C2) and (D2) hold. Denote by* $\tilde{\boldsymbol{\Gamma}}^{(N)}_{i,J}(\mathbf{B})$ *the statistics* $\tilde{\boldsymbol{\Gamma}}^{(N)}_{i,J}$ *computed from the* N*-tuple* $(\mathbf{BX}_1, ..., \mathbf{BX}_N)$*, where* $\mathbf{B}$ *is a* $d \times d$ *block-diagonal full rank matrix. Then,*

$$\tilde{\boldsymbol{\Gamma}}^{(N)}_{i,J}(\mathbf{B}) = (\mathbf{B}^{-1})^T \tilde{\boldsymbol{\Gamma}}^{(N)}_{i,J} \mathbf{B}^T.$$

Proof. Let $\mathbf{B}$ be $d \times d$ block-diagonal full rank matrix. Assumption (C2) insures that the residuals obtained from the transformed sample $(\mathbf{BX}_1, ..., \mathbf{BX}_N)$ are

$$\mathbf{e}^{(N)}_t(\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B})) = \mathbf{B}\mathbf{e}^{(N)}_t(\hat{\boldsymbol{\theta}}^{(N)}), t = 1, ..., N. \tag{7.2}$$

¿From Assumption (D2), we have

$$\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{B}) = \left(\hat{\boldsymbol{\Sigma}}(\mathbf{B})\right)^{-\frac{1}{2}} = k^{\frac{1}{2}} \mathbf{O} \hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \mathbf{B}^{-1}, \tag{7.3}$$

where $\mathbf{O}$ stands for an orthogonal matrix. Let $\tilde{R}_t(\mathbf{B})$ and $\tilde{\mathbf{U}}_t(\mathbf{B})$, respectively, the aligned ranks and signs computed from the transformed sample $\mathbf{BX}_1, ..., \mathbf{BX}_N$. Now, from (7.2) and (7.3), we can verify that

$$\tilde{R}_t(\mathbf{B}) = \tilde{R}_t \quad \text{and} \quad \tilde{\mathbf{U}}_t(\mathbf{B}) = \mathbf{O}\tilde{\mathbf{U}}_t \tag{7.4}$$

Then, the result directly follows from (7.4) and (7.3). □

Proposition 7.2 *Suppose that Assumptions (A1), (B1), (B2), (B3), (D1), and (E1) hold. Then, under* $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$*, with* $\boldsymbol{\theta}$ *belonging to* $\boldsymbol{\Theta}$*,*

$$\sqrt{N-i}\left\{\text{vec}\hat{\boldsymbol{\Gamma}}^{(N)}_{i,J}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}^{(N)}) - \text{vec}\hat{\boldsymbol{\Gamma}}^{(N)}_{i,J}(\boldsymbol{\theta})\right\} + \frac{\pi_d(J_1, f)\eta_d(J_2, f)}{d^2}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{a}_i(\boldsymbol{\tau}^{(N)}, \boldsymbol{\theta}) = o_p(1),$$

where, $\mathbf{a}_i(\boldsymbol{\tau}, \boldsymbol{\theta}) = \sum_{j=1}^{min(p,i)} (\mathbf{G}_{i-j}(\boldsymbol{\theta}) \otimes I_d)^T \text{vec}\boldsymbol{\gamma}_j.$

Proof. The result is a particular case of an asymptotic linearity property, established in the general context of multivariate general linear model with VARMA errors by Hallin and Paindaveine (2006). □

We only prove Theorems 3.1 and 4.1 for the problem of testing noncausality in both directions. In that case, the test statistics of interest are $Q_{\mathcal{N}}$ or $\tilde{Q}_J$. Proofs are very similar when testing for causality directions with the test statistics $Q_{\mathcal{N}}^{(12)}$ and $Q_{\mathcal{N}}^{(21)}$ or $\tilde{Q}_J^{(12)}$ and $\tilde{Q}_J^{(21)}$. Therefore, we assume that $\boldsymbol{\theta} \in \mathcal{M}(\mathbf{Q})$.

Proof of Theorem 3.1. (i) Note that, $\mathbf{\Lambda}_{i,\boldsymbol{\Sigma},\mathcal{N}}^{(N)}(\boldsymbol{\theta}) = \breve{\mathbf{\Gamma}}_{i,\boldsymbol{\Sigma},J,f}^{(N)}(\boldsymbol{\theta})$, for $J_1 = J_2 = \tilde{F}^{-1}$. Then, using Lemma 7.1, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, the Gaussian central sequence $\underline{\mathbf{\Delta}}_{\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta})$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\left(\frac{\mu_{d+1;f}}{d\mu_{d-1;f}}\right)^2 \mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$. The asymptotic linearity of $\underline{\mathbf{\Delta}}_{\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta})$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, follows from equation (7.8) for $J_1 = J_2 = \tilde{F}^{-1}$:

$$\underline{\mathbf{\Delta}}_{\boldsymbol{\Sigma},\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)}) = \underline{\mathbf{\Delta}}_{\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta}) - \frac{1}{d^2}\eta_d(f)\pi_d(\tilde{F}^{-1}, f)\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}) + o_p(1). \tag{7.5}$$

Since, the empirical covariance matrix $\hat{\mathbf{\Sigma}}_E$ is a consistent estimator for $\mathrm{cov}(\mathbf{e}_t^{(N)}(\boldsymbol{\theta})) = \mathrm{cov}(\boldsymbol{\epsilon}_t) = \frac{\mu_{d+1;f}}{d\mu_{d-1;f}}\mathbf{\Sigma} = \frac{\eta_d(f)}{d}\mathbf{\Sigma}$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, equation (7.5) becomes

$$\mathbf{\Delta}_{\mathcal{N}} = \underline{\mathbf{\Delta}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)}) = \underline{\mathbf{\Delta}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\boldsymbol{\theta}) - \frac{1}{d}\pi_d(\tilde{F}^{-1}, f)\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}) + o_p(1). \tag{7.6}$$

On the other hand, the continuity of $\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ with respect to $\boldsymbol{\theta}$ and $\mathbf{\Sigma}$, and the consistency of $\hat{\boldsymbol{\theta}}^{(N)}$ and $\hat{\mathbf{\Sigma}}_E$, insure that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$,

$$Q_{\mathcal{N}} = \mathbf{\Delta}_{\mathcal{N}}^T \left[(\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}})^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T \right] \mathbf{\Delta}_{\mathcal{N}} + o_p(1).$$

Now, (7.6) and Assumption (C1) imply that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$,

$$Q_{\mathcal{N}} = \left(\underline{\mathbf{\Delta}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\boldsymbol{\theta})\right)^T \left[(\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}})^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T \right] \underline{\mathbf{\Delta}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\boldsymbol{\theta}) + o_p(1). \tag{7.7}$$

Using the fact that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, $\underline{\mathbf{\Delta}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\boldsymbol{\theta})$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$, it follows that, under $\mathcal{H}_0 = \bigcup_{\boldsymbol{\theta}\in\mathbf{\Theta}_0}\bigcup_{\boldsymbol{\Sigma}}\bigcup_f \mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$, $Q_{\mathcal{N}}$ is asymptotically chi-square with $r = 2pd_1d_2$ degrees of freedom.

(ii) Lemma 7.1 implies that, under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f)$, the central sequence $\underline{\mathbf{\Delta}}_{\boldsymbol{\Sigma},\mathcal{N}}(\boldsymbol{\theta})$ is asymptotically normal with mean $\frac{1}{d^2}\pi_d(\tilde{F}^{-1}, f)\eta_d(\tilde{F}^{-1}, f)\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\boldsymbol{\tau}$ and covariance matrix $\left(\frac{\mu_{d+1;f}}{d\mu_{d-1;f}}\right)^2 \mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$. Note that $\left(\frac{\pi_d(\tilde{F}^{-1}, f)}{d}\right)^2 = \omega_d(f)$ where $\omega_d(f) = \frac{1}{d^2}\left(\int_0^1 \tilde{F}^{-1}(u)\varphi_f \circ \tilde{F}^{-1}(u)du\right)^2$ and $\eta_d(f) = \frac{\mu_{d+1;f}}{\mu_{d-1;f}}$. Again, substituting $\hat{\mathbf{\Sigma}}_E$ for $\mathbf{\Sigma}$ implies that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f)$, the

statistic $\underline{\mathbf{\Delta}}_{\hat{\mathbf{\Sigma}}_E,\mathcal{N}}(\boldsymbol{\theta})$ is asymptotically normal with mean $\sqrt{\omega_d(f)}\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}\boldsymbol{\tau}$ and covariance matrix $\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}$. Now, (7.7) is still valid under local alternatives, because $\mathcal{H}^{(N)}(\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau},\mathbf{\Sigma},f)$ and $\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$ are contiguous. Then, $Q_{\mathcal{N}}$ is noncentral chi-square with $r=2pd_1d_2$ degrees of freedom and with noncentrality parameter $\psi^2_{\mathcal{N},f}=\omega_d(f)\delta_{\boldsymbol{\theta},\mathbf{\Sigma},\boldsymbol{\tau}}$ under $\mathcal{H}^{(N)}(\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau},\mathbf{\Sigma},f)$.

(iii) The test $Q_{\mathcal{N}}$, is asymptotically equivalent to $\underline{Q}_{\mathbf{\Sigma},\mathcal{N}}(\boldsymbol{\theta})$ under the null hypothesis $\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},\mathcal{N})$ and under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau},\mathbf{\Sigma},\mathcal{N})$. Then, the test $\phi_{\mathcal{N}}$ is locally and asymptotically most stringent for $\mathcal{H}_0$ against $\bigcup_{\boldsymbol{\theta}\notin\mathbf{\Theta}_0}\bigcup_{\mathbf{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},\mathcal{N})$. This completes the proof of Theorem 3.1. □

Proof of Theorem 4.1. (i) To prove (i), we first need to prove the asymptotic linearity of $\hat{\mathbf{\Delta}}_J(\boldsymbol{\theta})$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$:

$$\hat{\mathbf{\Delta}}_J(\hat{\boldsymbol{\theta}}^{(N)})=\hat{\mathbf{\Delta}}_J(\boldsymbol{\theta})-\frac{1}{d^2}\pi_d(J_1,f)\eta_d(J_2,f)\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})+o_p(1). \tag{7.8}$$

Let $\mathbf{M}_i^{(s)}:=\mathbf{M}_i^{(s)}(\boldsymbol{\theta})$ be the ith block-column of the matrix $\mathbf{M}^{(s)}(\boldsymbol{\theta})$: $\mathbf{M}^{(s)}(\boldsymbol{\theta})=\left[\mathbf{M}_1^{(s)},...,\mathbf{M}_{s-1}^{(s)}\right]$. Similarly, let $\hat{\mathbf{M}}_i^{(s)}:=\mathbf{M}_i^{(s)}(\hat{\boldsymbol{\theta}}^{(N)})$ be the ith block-column of the matrix $\mathbf{M}^{(s)}(\hat{\boldsymbol{\theta}}^{(N)})$. Using this notation, for any fixed integer s $(N>s)$, we have the following decomposition

$$\begin{aligned}\hat{\mathbf{\Delta}}_J(\hat{\boldsymbol{\theta}}^{(N)})-\hat{\mathbf{\Delta}}_J(\boldsymbol{\theta})=\quad&\sum_{i=1}^{s-1}(N-i)^{\frac{1}{2}}\left(\hat{\mathbf{M}}_i^{(s)}-\mathbf{M}_i^{(s)}\right)\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\\&+\sum_{i=1}^{s-1}(N-i)^{\frac{1}{2}}\mathbf{M}_i^{(s)}\left(\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})-\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta})\right)\\&+\sum_{i=s}^{n-1}(N-i)^{\frac{1}{2}}\left(\hat{\mathbf{M}}_i^{(s)}\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})-\mathbf{M}_i^{(s)}\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta})\right).\end{aligned}$$

Proposition 7.2 (with $\boldsymbol{\tau}^{(N)}=\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})$), local discretness property and root-N consistency of $\hat{\boldsymbol{\theta}}^{(N)}$ (Assumptions (C1)-(ii) and (C1)-(iii)) imply that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$,

$$\begin{aligned}\sum_{i=1}^{s-1}(N-i)^{\frac{1}{2}}\mathbf{M}_i^{(s)}\left(\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})-\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta})\right)=&-\frac{\pi_d(J_1,f)\eta_d(J_2,f)}{d^2}\mathbf{M}^{(s)}(\boldsymbol{\theta})\left[\mathbf{I}_{s-1}\otimes(\mathbf{\Sigma}\otimes\mathbf{\Sigma}^{-1})\right]\\&\left(\mathbf{M}^{(s)}(\boldsymbol{\theta})\right)^T\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})+\sum_{i=1}^{s-1}\mathbf{R}_i^{(N)},\end{aligned}$$

where, $\mathbf{R}_i^{(N)}=o_p(1)$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma},f)$, as $N\to\infty$. Hence, we obtain

$$\hat{\mathbf{\Delta}}_J(\hat{\boldsymbol{\theta}}^{(N)})-\hat{\mathbf{\Delta}}_J(\boldsymbol{\theta})+\frac{1}{d^2}\pi_d(J_1,f)\eta_d(J_2,f)\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})=T_1^{N,s}+T_2^{N,s},$$

where

$$T_1^{N,s}=\sum_{i=1}^{s-1}(N-i)^{\frac{1}{2}}\left(\hat{\mathbf{M}}_i^{(s)}-\mathbf{M}_i^{(s)}\right)\text{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})+\sum_{i=1}^{s-1}\mathbf{R}_i^{(N)},$$

and

$$T_2^{N,s} = \frac{1}{d^2}\pi_d(J_1,f)\eta_d(J_2,f)\left(\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}} - \mathbf{M}^{(s)}(\boldsymbol{\theta})\left[\mathbf{I}_{s-1}\otimes\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}^{-1}\right)\right]\left(\mathbf{M}^{(s)}\right)^T\right)\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})$$
$$+\sum_{i=s}^{n-1}(N-i)^{\frac{1}{2}}\left(\hat{\mathbf{M}}_i^{(s)}\mathrm{vec}\hat{\boldsymbol{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}) - \mathbf{M}_i^{(s)}\mathrm{vec}\hat{\boldsymbol{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta})\right).$$

The continuity in $\boldsymbol{\theta}$ of the Green matrices $\mathbf{G}_u(\boldsymbol{\theta})$, the root-$N$ consistency of $\hat{\boldsymbol{\theta}}^{(N)}$, and the boundedness of $(N-i)^{\frac{1}{2}}\mathrm{vec}\hat{\boldsymbol{\Gamma}}_{i,J}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})$ under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, which follows from Proposition 7.1, Lemma 7.1 and Proposition 7.2, insure that, for any fixed s, $T_1^{N,s}=o_p(1)$, as $N\to\infty$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$. Moreover, the exponential decrease of the Green matrices and the root-N consistency of $\hat{\boldsymbol{\theta}}^{(N)}$ imply that $T_2^{N,s}=o_p(1)$ under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, as $s\to\infty$, and the convergence is uniform in N. Now, we can chose $s=S$ sufficiently large so that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, $T_1^{N,S}=T_2^{N,S}=o_p(1)$, as $N\to\infty$; then (7.8) follows.

Turning to the distribution under the null hypothesis of $\tilde{Q}_J$, the continuity of $\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, and the root-N consistency of $\hat{\boldsymbol{\theta}}^{(N)}$ and $\hat{\boldsymbol{\Sigma}}$, entail that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$,

$$\tilde{Q}_J := \frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\tilde{\boldsymbol{\Delta}}_J^T\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\tilde{\boldsymbol{\Delta}}_J + o_p(1).$$

Now, using (7.8) and the fact that $\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta}) = \mathbf{0}$ (it follows from Assumption (C1)-(i) and the fact that $\boldsymbol{\theta}\in\mathcal{M}(\mathbf{Q})$), we obtain,

$$\tilde{Q}_J := \frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\hat{\boldsymbol{\Delta}}_J^T(\boldsymbol{\theta})\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\hat{\boldsymbol{\Delta}}_J(\boldsymbol{\theta}) + o_p(1),$$

under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, as $N\to\infty$. Now, from Proposition 7.1 and Lemma 7.1, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, $\hat{\boldsymbol{\Delta}}_J^T(\boldsymbol{\theta})$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\frac{1}{d^2}\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$. This implies that $\tilde{Q}_J$ is asymptotically chi-square with $r=2pd_1d_2$ degrees of freedom under $\mathcal{H}_0 = \bigcup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\bigcup_{\boldsymbol{\Sigma}}\bigcup_f\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$.

(ii) Let us prove that $\tilde{Q}_J$ is asymptotically invariant with respect to the group of continuous monotone radial transformations. Let $\boldsymbol{\Delta}_{\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta}) := \mathbf{M}^{(N)}(\boldsymbol{\theta})\mathbf{T}_{\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta})$, $(\boldsymbol{\theta}\in\mathcal{M}(\mathbf{Q}))$ with

$$\mathbf{T}_{\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta}) = \left((N-1)^{\frac{1}{2}}\left(\mathrm{vec}\boldsymbol{\Gamma}_{1,\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta})\right)^T, ..., (N-i)^{\frac{1}{2}}\left(\mathrm{vec}\boldsymbol{\Gamma}_{i,\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta})\right)^T, ..., \left(\mathrm{vec}\boldsymbol{\Gamma}_{N-1,\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta})\right)^T\right)^T,$$

where

$$\boldsymbol{\Gamma}_{i,\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta}) := (N-i)^{-1}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\right)^T\left(\sum_{t=i+1}^{N}J_1\left(\frac{R_t(\boldsymbol{\theta},\boldsymbol{\Sigma})}{N+1}\right)J_2\left(\frac{R_{t-i}(\boldsymbol{\theta},\boldsymbol{\Sigma})}{N+1}\right)\mathbf{U}_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})\left(\mathbf{U}_{t-i}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})\right)^T\right)\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^T.$$

On the other hand, the continuity of $\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, implies that, under $\bigcup_f\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, as $N\to\infty$,

$$\tilde{Q}_J := \frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\tilde{\boldsymbol{\Delta}}_J^T\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\tilde{\boldsymbol{\Delta}}_J + o_p(1). \tag{7.9}$$

Now, from (7.8), under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$,

$$\tilde{\boldsymbol{\Delta}}_J=\hat{\boldsymbol{\Delta}}_J(\boldsymbol{\theta})-\frac{1}{d^2}\pi_d(J_1,f)\eta_d(J_2,f)\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\sqrt{N}(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})+o_p(1).$$

Moreover, we can verify that $\boldsymbol{\Delta}^{(N)}_{\boldsymbol{\Sigma},J}(\boldsymbol{\theta})-\hat{\boldsymbol{\Delta}}_J(\boldsymbol{\theta})=o_p(1)$, as $N\to+\infty$, under $\bigcup_f\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$. Then, because $\left[\mathbf{N}^{-1}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}-\mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}=0$, and $(\hat{\boldsymbol{\theta}}^{(N)}-\boldsymbol{\theta})\in\mathcal{M}(\mathbf{Q})$ under Assumption (C1), we obtain that under $\bigcup_f\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$,

$$\tilde{Q}_J=Q_{\boldsymbol{\Sigma},J}(\boldsymbol{\theta})=\frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\left(\boldsymbol{\Delta}^{(N)}_{\boldsymbol{\Sigma},J}(\boldsymbol{\theta})\right)^T\left[\mathbf{N}^{-1}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}-\mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\boldsymbol{\Delta}^{(N)}_{\boldsymbol{\Sigma},J}(\boldsymbol{\theta})+o_p(1).$$

This entails that $\tilde{Q}_J$ is asymptotically invariant with respect to the group of continuous monotone radial transformations, since $\boldsymbol{\Delta}^{(N)}_{\boldsymbol{\Sigma},J}(\boldsymbol{\theta})$ is strictly invariant with respect to that group.

Now, to prove that $\tilde{Q}_J$ is block-diagonal-affine-invariant, Let $\mathbf{B}$ be a $d\times d$ matrix of the form $\mathbf{B}=\begin{pmatrix}\mathbf{B}^{(11)} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{(22)}\end{pmatrix}$, where $\mathbf{B}^{(11)}$ and $\mathbf{B}^{(22)}$ are full rank matrices with dimension $d_1\times d_1$ and $d_2\times d_2$ respectively. Denote by $\mathbf{R}(\mathbf{B})$ the value of a statistic $\mathbf{R}$ (function of the sample $\mathbf{X}_1,...,\mathbf{X}_N$) computed from the transformed sample $\mathbf{BX}_1,...,\mathbf{BX}_N$. When the statistic is of the form $\mathbf{R}(\mathbf{H})$, we will use $\mathbf{R}(\mathbf{H}(\mathbf{B}))$ to stand for the statistic $\mathbf{R}(\mathbf{H})$ computed from the transformed sample $\mathbf{BX}_1,...,\mathbf{BX}_N$. It is clear that,

$$\mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}))=\mathbf{I}_p\otimes\left(\mathbf{B}\otimes(\mathbf{B}^T)^{-1}\right)\mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}\left[\mathbf{I}_{N-1}\otimes\left(\mathbf{B}^{-1}\otimes\mathbf{B}^T\right)\right].$$

Lemma (7.2) implies that $\tilde{\mathbf{T}}^{(N)}_J(\mathbf{B})=\mathbf{I}_{N-1}\otimes\left(\mathbf{B}\otimes(\mathbf{B}^T)^{-1}\right)\tilde{\mathbf{T}}^{(N)}_J$. Then,

$$\tilde{\boldsymbol{\Delta}}_J(\mathbf{B})=\mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}))\tilde{\mathbf{T}}^{(N)}_J(\mathbf{B})=\left[\mathbf{I}_p\otimes\left(\mathbf{B}\otimes(\mathbf{B}^T)^{-1}\right)\right]\tilde{\boldsymbol{\Delta}}_J, \tag{7.10}$$

and

$$\mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}),\hat{\boldsymbol{\Sigma}}(\mathbf{B})}:=\left[\mathbf{I}_p\otimes\left(\mathbf{B}\otimes(\mathbf{B}^T)^{-1}\right)\right]\mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}}\left[\mathbf{I}_p\otimes\left(\mathbf{B}^T\otimes\mathbf{B}^{-1}\right)\right]. \tag{7.11}$$

Now, from (7.10), (7.11) and the fact that $\mathcal{M}(\mathbf{Q})=\mathcal{M}(\mathbf{I}_p\otimes\left(\mathbf{B}\otimes(\mathbf{B}^T)^{-1}\right)\mathbf{Q})$.

$$\begin{aligned}\tilde{Q}_J(\mathbf{B}) &:= \frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\tilde{\boldsymbol{\Delta}}_J^T(\mathbf{B})\left[\mathbf{N}^{-1}_{\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}),\hat{\boldsymbol{\Sigma}}(\mathbf{B})}-\mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B}),\hat{\boldsymbol{\Sigma}}(\mathbf{B})}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\tilde{\boldsymbol{\Delta}}_J(\mathbf{B})\\ &= \frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\tilde{\boldsymbol{\Delta}}_J^T\left[\mathbf{N}^{-1}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}}-\mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\tilde{\boldsymbol{\Delta}}_J\end{aligned}$$

This implies that, $\tilde{Q}_J(\mathbf{B})=\tilde{Q}_J$. Consequently, block-diagonal-affine-invariance is achieved.

(iii) The LAN property implies that equation (7.9) is also valid under local (contiguous) alternatives. Then, for any $\boldsymbol{\theta}\in\mathcal{M}(\mathbf{Q})$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}+N^{-\frac{1}{2}}\boldsymbol{\tau},\boldsymbol{\Sigma},f)$,

$$\tilde{Q}_J:=\frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\tilde{\boldsymbol{\Delta}}_J^T\left[\mathbf{N}^{-1}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}-\mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\tilde{\boldsymbol{\Delta}}_J+o_p(1).$$

Again, using (7.8) and the fact that $\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}^{(N)} - \boldsymbol{\theta}) = 0$ (which follows from Assumption (C1)-(i)) we obtain that $\tilde{Q}_J$ has the same asymptotic distribution (under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f)$) as

$$\frac{d^2}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\hat{\boldsymbol{\Delta}}_J^T(\boldsymbol{\theta})\left[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\hat{\boldsymbol{\Delta}}_J(\boldsymbol{\theta}).$$

Using Proposition 7.1 and Lemma 7.1 (which are still valid under local alternatives) we obtain that $\hat{\boldsymbol{\Delta}}_J^T(\boldsymbol{\theta})$, is asymptotically normal, with mean $\left(\frac{1}{d^2}\pi_d(J_1, f)\eta_d(J_2, f)\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\boldsymbol{\tau}\right)$ and covariance matrix $\frac{1}{d^2}\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ under $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f)$. This implies that $\tilde{Q}_J$ is asymptotically noncentral chi-square with $r = 2pd_1d_2$ degrees of freedom and with noncentrality parameter $\tilde{\psi}_{J,f}^2 = \tilde{\psi}_{J,f}^2(\boldsymbol{\tau}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) = \frac{1}{d^2}\frac{\pi_d^2(J_1,f)\eta_d^2(J_2,f)}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$, under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f)$.

(iv) From Section 3.2, under radial density f_*, the test $\underline{\phi}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) := I\left[\underline{Q}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) > \chi^2_{r,1-\alpha}\right]$, where

$$\underline{Q}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) := \left(\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta})\right)^T\left[\left(\underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta})\right)^{-1} - \mathbf{Q}\left(\mathbf{Q}^T\underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta})\mathbf{Q}\right)^{-1}\mathbf{Q}^T\right]\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}),$$

with

$$\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) = \mathbf{M}^{(N)}(\boldsymbol{\theta})\mathbf{S}_{\boldsymbol{\Sigma},f_*}^{(N)}(\boldsymbol{\theta}), \quad \text{and} \quad \underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) := \xi_d(f_*)\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}},$$

is locally asymptotically most stringent. We can easily check that, under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_*)$, $\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) = \hat{\boldsymbol{\Delta}}_J(\boldsymbol{\theta}) + o_p(1)$, with $J_1 = \varphi_{f_*} \circ \tilde{F}_*^{-1}$ and $J_2 = \tilde{F}_*^{-1}$. Moreover, we can also verify that $\boldsymbol{\Omega}_{f_*} = \tilde{\boldsymbol{\Omega}}_J$. Then, $\underline{Q}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta}) = \tilde{Q}_J + o_p(1)$ under $\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_*)$ as well as under contiguous alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta} + N^{-\frac{1}{2}}\boldsymbol{\tau}, \boldsymbol{\Sigma}, f_*)$. Finally, the test $\tilde{\phi}_J$, with $J_1 = \varphi_{f_*} \circ \tilde{F}_*^{-1}$ and $J_2 = \tilde{F}_*^{-1}$ is asymptotically equivalent to $\underline{\phi}_{\boldsymbol{\Sigma},f_*}(\boldsymbol{\theta})$ under the null and under local alternatives. Therefore, $\tilde{\phi}_J$ is also locally asymptotically most stringent for $\mathcal{H}_0(f_*)$ against $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f_*)$. This completes the proof of Theorem 4.1. □

References

Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics*. Springer: New York.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992). Testing causality between two vectors in multivariate autoregressive moving average models. *Journal of the American Statistical Association* **87**, 1082-1090.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1994). Simplified conditions for noncausality between vectors in multivariate ARMA models. *Journal of Econometrics* **63**, 271-287.

Dolado, J. J., and Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* **15**, 369-386.

Dufour, J.-M., and Renault, E. (1998). Short run and long run causality in time series: Theory.

Econometrica **66**, 1099-1125.

Dufour, J.-M., Pelletier, D., and Renault, E. (2006). Short run and long run causality in time series: Inference. *Journal of Econometrics* **133**, 337-362.

Drost, F. C., Klaassen, C. A. J., and Werker, B. J. M. (1997). Adaptive estimation in time-series models, *Annals of Statistics* **25**, 786-818.

Garel, B., and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics* **47**, 551-579.

Geweke, J. (1984). Inference and causality in economic time series. In *Handbook of Econometrics* (Z. Griliches, and M. D. Intrilligator, Eds.), 1102-1144. North-Holland: Amsterdam.

Gouriéroux, C., et Monfort, A. (1990). *Séries temporelles et modèles dynamiques*. Economica: Paris.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424-459.

Hallin, M., and Paindaveine, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Annals of Statistics* **30**, 1103-1133.

Hallin, M., and Paindaveine, D. (2002b). Multivariate signed ranks : Randles' interdirections or Tyler's angles?. In *Statistical data analysis based on the L1-norm and related methods* (Y. Dodge, Ed.), 271-282. Birkhuser, Basel.

Hallin, M., and Paindaveine, D. (2002c). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8**, 787-815.

Hallin, M., and Paindaveine, D. (2003). Affine invariant linear hypotheses for the multivariate general linear model with ARMA error terms. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden, I.M.S. Lecture Notes-Monograph Series, I.M.S.* (M. Moore, S. Froda, and C. Léger, Eds.), 417-434. Hayward: California.

Hallin, M., and Paindaveine, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Annals of Statistics* **32**, 2642-2678.

Hallin, M., and Paindaveine, D. (2004b). Multivariate signed rank tests in vector autoregressive order identification. *Statistical Science* **19**, 697-711.

Hallin, M., and Paindaveine, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *Journal of Multivariate Analysis* **93**, 122-163.

Hallin, M., and Paindaveine, D. (2006). Asymptotic linearity of serial and nonserial multivariate signed rank statistics. *Journal of Statistical Planning and Inference* **136**, 1-32.

Hallin, M., and Puri, M. L. (1992). Rank tests for time-series analysis: a survey. In *New Directions in Time Series Analysis*, Vol. 45 (D. Brillinger, E. Parzen, and M. Rosenblatt, Eds.), 111-153. New York: Springer-Verlag.

Hallin, M., and Puri, M. L. (1994). Aligned rank tests for linear models with autocorrelated errors. *Journal of Multivariate Analysis* **50**, 175-237.

Hallin, M., and Werker, B. J. M. (1999). Optimal testing for semi-parametric autoregressive models: from Gaussian Lagrange multipliers to regression rank scores and adaptive tests. In *Asymptotics, Nonparametrics, and Time Series* (S. Ghosh, Ed.), 295-358. New York: M. Dekker.

Hwang, S. Y., and Basawa, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series. *Stochastic Processes and their Applications* **46**, 91-114.

Koul, H. L., and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Annals of Statistics* **24**, 1025-1052.

Koul, H. L., and Schick, A. (1997). Efficient estimation in nonlinear autoregressive time series models. *Bernoulli* **3**, 247-277.

Kreiss, J.-P. (1987). On adaptative estimation in stationary ARMA processes. *Annals of Statistics* **15**, 112-133.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag: Berlin.

Lütkepohl H. (1993). Testing for causation between two variables in higher dimensional VAR models. In *Studies in Applied Econometrics* (H. Schneeweis and K. Zimmerman Eds.), 75-91. Heidelberg: Springer-Verlag.

Martin, R. D. and Yohai, V. J. (1985). Robustness in time series and estimating ARMA models. In *Handbook of Statistics*, **5**, Time Series in the Time Domain (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, Eds.), 119-155. North-Holland: Amsterdam.

Newbold, P. (1982). Causality testing in economics. In *Time Series Analysis: Theory and Practice 1* (O. D. Anderson, Ed.). North-Holland: Amsterdam.

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates. *Scandinavian Journal of Statistics* **26**, 319-343.

Oja, H., and Paindaveine, D. (2005). Optimal signed-rank tests based on hyperplanes. *Journal of Statistical Planning and Inference* **135**, 300-323.

Peters, D., and R. H. Randles (1990). A multivariate signed-rank test for the one-sample location problem. *Journal of the American Statistical Association* **85**, 552-557.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.

Pierce, D.A., and Haugh, L. D. (1977). Causality in temporal systems: characterization and survey. *Journal of Econometrics* **5**, 265-293.

Puri, M. L., and P. K. Sen (1971). *Nonparametric Methods in Multivariate Analysis*. J. Wiley: New York.

Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *Journal of the American Statistical Association* **84**, 1045-1050.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley: New York.

Sims, C. A., Stock, J. H., Watson, M. W. (1990). Interference in linear time series models with some unit roots. *Econometrica* **58**, 113-144.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis* **16**, 54-70.

Taniguchi, M., and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer.

Taylor, S. A. (1989). A comparison of classical tests of criterion when determining Granger causality with a bivariate ARMA models. *Empirical Economics* **14**, 257-271.

Tjostheim, D. (1981). Granger causality in multiple time series. *Journal of Econometrics* **17**, 157-176.

Toda, H. Y., and Phillips, P. C. B., (1993). Vector autoregressions and causality. *Econometrica* **61**, 1367-1393.

Toda, H. Y., and Phillips, P. C. B., (1994). Vector autoregressions and causality: a theoretical overview and simulation study. *Econometric Reviews* **13**, 259-285.

Tsay, R. S., Peña, D. and Pankratz, A. E. (2000). Outliers in multivariate time series. *Biometrika* **87**, 789-804.

Tyler, D. E. (1987). A distribution-free M-estimate of multivariate scatter. *Annals of Statistics* **15**, 234-251.

Wiener, N. (1956). The theory of prediction. *In Modern Mathematics for Engineers* (E.F. Beckenback, Ed.), 165-190. McGraw-Hill: New-York.