Econometric Theory, **24**, 2008, 948–987. Printed in the United States of America.
doi:10.1017/S0266466608080377

ROBUST OPTIMAL TESTS FOR CAUSALITY IN MULTIVARIATE TIME SERIES

ABDESSAMAD SAIDI AND ROCH ROY
Université de Montréal

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, *Asymptotic Methods in Statistical Decision Theory*) 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, *Annals of Statistics* 32, 2642–2678) 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. The level, power, and robustness (to outliers) of the resulting tests are studied by simulation and are compared to those of the Wald test. Finally, the new tests are applied to Canadian money and income data.

1. INTRODUCTION

The concept of causality introduced by Wiener (1956) and Granger (1969) is now a fundamental idea 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), Gouriéroux and Monfort (1990, Ch. X), and Lütkepohl (1991). The idea behind this concept is that if a variable X affects a variable Y, the former should help improve the predictions of the latter variable. A formal definition is presented in Section 2. The original definition of Granger (1969) refers to

The authors thank the co-editor and two anonymous referees for their helpful comments. In particular, many constructive remarks by the referees on the first draft of the paper greatly enhanced the presentation. 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). Address correspondence to Roch Roy, Département de mathématiques et de statistique, Université de Montréal, CP 6128, succursale Centre-ville, Montréal, Québec, H3C 3J7, Canada; e-mail: royr@dms.umontreal.ca.

 © 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

the predictability of a variable X, one period ahead. This is also called causality in mean. It was extended to vectors of variables; see, for example, Tjøstheim (1981), Lütkepohl (1991), and Boudjellaba, Dufour, and Roy (1992, 1994). Lütkepohl (1993) and 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. First, the characterization of noncausality in terms of the parameters of the fitted model to the observed series. Second, 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, Ch. 2) for vector autoregressive (VAR) models, and by Boudjellaba et al. (1992, 1994) for vector autoregressive moving average (VARMA) models. Characterization of noncausality and inference in possibly cointegrated autoregressions was studied by, among others, 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), and Dufour, Pelletier, and Renault (2006).

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 semiparametric (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 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. To 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 autoregressive case with linear trend by Swensen (1985) and in the ARMA case by

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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), and 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 that 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 the 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 LAN and robustness features offered by the Peters and Randles (1990) 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 established by Hallin and Puri (1994), who 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 also was 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). Although these applications are concerned with affine invariant linear problems, the general results of Hallin and Paindaveine (2005) can also be used for testing linear hypotheses that are not affine invariant. The approach we adopt in the present paper is in the same spirit. We combine robustness, invariance, and optimality concerns. The null hypothesis of interest here is not affine invariant. Indeed, the null hypothesis of no feedback in

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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, such as measurement errors or equipment failure (see, e.g., Martin and Yohai, 1985; Rousseeuw and Leroy, 1987; 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 we present some technical assumptions on the elliptical density and 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. Because this density is 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 called "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. An adaptive version of those tests is also discussed. In Section 5, we describe how the theory developed for VAR models can be adapted for testing causality in a vector moving average (VMA) context.

The particular case of testing for no feedback in the bivariate VAR(1) model is considered in Section 6, 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), this 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 that is almost as powerful as the Wald and pseudo-Gaussian tests. In the presence of observation or innovation outliers, both the Wald and the pseudo-Gaussian tests are severely affected. With innovation outliers, the

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

levels of all rank-based tests are very well controlled. However, with observation outliers, the semiparametric tests are still biased. In general, they overreject, and the bias is more important when using the empirical covariance matrix estimator.

Finally, the proposed rank-based tests are applied to a set of Canadian economic data to study the causality relations between income and money. The tests are based on a trivariate VAR(5) model of changes in nominal income and two money stocks ($M1$ and $M2$). They lead us to conclude that there is feedback between income and money.

A word on notation. Boldface throughout denotes 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 $\mathbf{P}$, $\mathbf{P}^{1/2}$ is the unique upper triangular $k \times k$ matrix with positive diagonal elements that satisfy $\mathbf{P} = (\mathbf{P}^{1/2})^T \mathbf{P}^{1/2}$. Also, $\mathbf{A} \leq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is nonnegative 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 VAR(p) process satisfying a stochastic difference equation of the form

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

where $\mathbf{A}_j$, $j = 1, \ldots, p$, are $d \times d$ real matrices and $\boldsymbol{\epsilon}_t$ is a d-variate white noise process, that is, 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 $\mathbf{A}_j$, $j = 1, \ldots, p$, into

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

Denote by

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

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

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

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

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

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 Tjøstheim (1981). Boudjellaba et al. (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

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

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 $\boldsymbol{\Sigma}(\boldsymbol{\xi}|\mathcal{H}(\mathbf{X};t))$ be the covariance matrix of the corresponding prediction error $\boldsymbol{\xi} - \text{Proj}(\boldsymbol{\xi}|\mathcal{H}(\mathbf{X};t))$.

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 et al. (1992); see also Lütkepohl (1991, Sect. 2.3).

PROPOSITION 2.1. *Suppose that the VAR(p) process* $\mathbf{X} = ((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T)^T$ *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* $\mathbf{A}_j^{(21)} = \mathbf{0}$ $\forall j = 1, \ldots, p$;

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

(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* $\mathbf{A}_j^{(12)} = \mathbf{A}_j^{(21)} = \mathbf{0}$ $\forall j = 1, \ldots, p$.

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. In our approach, linearity is crucial to construct optimal tests. On 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 et al. (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) holds 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 $\mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{(11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_j^{(22)} \end{pmatrix}$, $j = 1, \ldots, p$. Then $\mathcal{H}_0$ holds iff $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, where $\boldsymbol{\Theta}_0 := \{\boldsymbol{\theta} = (\mathrm{vec}^T \mathbf{A}_1, \ldots, \mathrm{vec}^T \mathbf{A}_p)^T \in \boldsymbol{\Theta} | (\mathbf{A}_1, \ldots, \mathbf{A}_p) \in \mathcal{A}\}$ 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

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 Fang, Kotz, and Ng (1990). 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 LAN (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)–(B3), which follow, 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 $\boldsymbol{\Sigma}$, that is, a sequence of independent, identically distributed (i.i.d.) random vectors with mean zero and probability density given by

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

Here, $\|\mathbf{z}\|_{\mathbf{\Sigma}}$ denotes the norm of $\mathbf{z}$ in the metric associated with $\mathbf{\Sigma}$, that is, $\|\mathbf{z}\|_{\mathbf{\Sigma}}^2 = \mathbf{z}^T\mathbf{\Sigma}^{-1}\mathbf{z}$. The constant $c_{d,f}$ is the normalization factor $(\kappa_d \mu_{d-1;f})^{-1}$, where κ_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 $\mathbf{\Sigma}$ and f are only identified up to an arbitrary scale factor. More precisely, for any $a > 0$, letting $\mathbf{\Sigma}_a = a^2\mathbf{\Sigma}$ and $f_a(r) = f(ar)$, we have $\underline{f}(\mathbf{z}, \mathbf{\Sigma}, f) = \underline{f}(\mathbf{z}, \mathbf{\Sigma}_a, f_a)$. This will not be a problem because we just need the multivariate scatter matrix $a\mathbf{\Sigma}$ (for some arbitrary $a > 0$), not $\mathbf{\Sigma}$ itself. We will denote by $\mathbf{\Sigma}^{-1/2}$ the unique upper triangular $d \times d$ array with positive diagonal elements that satisfies $\mathbf{\Sigma}^{-1} = (\mathbf{\Sigma}^{-1/2})^T\mathbf{\Sigma}^{-1/2}$. With this notation, $\mathbf{\Sigma}^{-1/2}\boldsymbol{\epsilon}_t / \|\mathbf{\Sigma}^{-1/2}\boldsymbol{\epsilon}_t\|$ are i.i.d. and uniformly distributed over $\mathcal{S}^{d-1}$. Similarly, $\|\mathbf{\Sigma}^{-1/2}\boldsymbol{\epsilon}_t\|$ are i.i.d. 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 necessary and sufficient condition is given by $\mu_{d+1;f} = \int_0^\infty r^{d+1} f(r)\, dr < \infty$.

(B3). Let $L^2(\mathbb{R}_0^+, \mu_l)$ denote 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 $f^{1/2}$ 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 $\underline{f}^{1/2}$, requiring the existence of a square integrable vector $\mathbf{D}\underline{f}^{1/2}$ such that, for all $\mathbf{0} \neq \mathbf{h} \to \mathbf{0}$,

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

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 $(f^{1/2})' = f'/2f^{1/2}$ is in $L^2(\mathbb{R}_0^+, \mu_{d-1})$.

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

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

Consider two arbitrary sequences of $d \times d$ matrices $\boldsymbol{\gamma}_1^{(N)}, \ldots, \boldsymbol{\gamma}_p^{(N)}$ and $\tilde{\boldsymbol{\gamma}}_1^{(N)}, \ldots, \tilde{\boldsymbol{\gamma}}_p^{(N)}$ and let $\boldsymbol{\tau}^{(N)} := (\mathrm{vec}^T \boldsymbol{\gamma}_1^{(N)}, \ldots, \mathrm{vec}^T \boldsymbol{\gamma}_p^{(N)})^T \in \mathbb{R}^K$ and $\tilde{\boldsymbol{\tau}}^{(N)} := (\mathrm{vec}^T \tilde{\boldsymbol{\gamma}}_1^{(N)}, \ldots, \mathrm{vec}^T \tilde{\boldsymbol{\gamma}}_p^{(N)})^T \in \mathbb{R}^K$. 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 $\boldsymbol{\tau} := (\mathrm{vec}^T \boldsymbol{\gamma}_1, \ldots, \mathrm{vec}^T \boldsymbol{\gamma}_p)^T$ instead of $\boldsymbol{\tau}^{(N)}$. Define the local sequences

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

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

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

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

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 $\mathrm{P}^{(N)}_{\boldsymbol{\Sigma}, f, \boldsymbol{\theta}}$,

$$\mathbf{e}_t^{(N)}(\boldsymbol{\theta}) := \mathbf{X}_t - \sum_{j=1}^{p} \mathbf{A}_j \mathbf{X}_{t-j}, \tag{2.5}$$

that we decompose into $\mathbf{e}_t^{(N)}(\boldsymbol{\theta}) = d_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{1/2} \mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $d_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) := \|\mathbf{e}_t^{(N)}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}$ and $\mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) := \boldsymbol{\Sigma}^{-1/2} \mathbf{e}_t^{(N)}(\boldsymbol{\theta}) / d_t^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. As in Garel and Hallin (1995), we define the residual f-cross covariance matrix at lag i as $\boldsymbol{\Lambda}^{(N)}_{i, \boldsymbol{\Sigma}, f}(\boldsymbol{\theta}) := (N - i)^{-1} \sum_{t=i+1}^{N} \varphi_f(\mathbf{e}_t^{(N)}(\boldsymbol{\theta})) (\mathbf{e}_{t-i}^{(N)}(\boldsymbol{\theta}))^T$, where $\varphi_f = -2((f^{1/2})'/f^{1/2})$. Because of the elliptical structure of f, these cross-covariance matrices take the form

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

$$\mathbf{\Lambda}^{(N)}_{i,\mathbf{\Sigma},f}(\boldsymbol{\theta}) := (N-i)^{-1}(\mathbf{\Sigma}^{-1/2})^T \times \left(\sum_{t=i+1}^{N} \varphi_f(d_t^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma}))\, d_{t-i}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma})\mathbf{U}_t^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma})(\mathbf{U}_{t-i}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma}))^T \right) \times (\mathbf{\Sigma}^{1/2})^T. \tag{2.6}$$

Denote the vector of all cross-covariance matrices by

$$\mathbf{S}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}) = ((N-1)^{1/2}(\mathrm{vec}\mathbf{\Lambda}^{(N)}_{1,\mathbf{\Sigma},f}(\boldsymbol{\theta}))^T, \ldots, (N-i)^{1/2}(\mathrm{vec}\mathbf{\Lambda}^{(N)}_{i,\mathbf{\Sigma},f}(\boldsymbol{\theta}))^T, \ldots, (\mathrm{vec}\mathbf{\Lambda}^{(N)}_{N-1,\mathbf{\Sigma},f}(\boldsymbol{\theta}))^T)^T. \tag{2.7}$$

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 $\{\mathbf{G}_u(\boldsymbol{\theta})\}$:

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

We are now ready to state the ULAN property that 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) and (B1)–(B3) are satisfied. Let* $\boldsymbol{\theta} \in \mathbf{\Theta}_0$ *and* $\boldsymbol{\theta}^{(N)}$ *and* $\tilde{\boldsymbol{\theta}}^{(N)}$ *be as defined in (2.3) and (2.4), respectively. Then,*

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

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

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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*

$$\xi_d(f) = \frac{\mathcal{I}_{d,f}\,\mu_{d+1;f}}{d^2\mu_{d-1;f}}, \tag{2.11}$$

and

$$\begin{aligned}\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}} &= \lim_{N\to+\infty} \mathbf{M}^{(N)}(\boldsymbol{\theta})[\mathbf{I}_{N-1}\otimes(\mathbf{\Sigma}\otimes\mathbf{\Sigma}^{-1})](\mathbf{M}^{(N)}(\boldsymbol{\theta}))^T \\ &= \left(\sum_{j=\max(i,i')}^{+\infty} \mathbf{G}_{j-i}\mathbf{\Sigma}\mathbf{G}_{j-i'}^T \otimes \mathbf{\Sigma}^{-1}\right)_{i,i'=1,\ldots,p}. \end{aligned} \tag{2.12}$$

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

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

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

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 $\phi(\underline{\mathbf{\Delta}}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta}))$ inherits, locally and asymptotically, the same optimality properties in the sequence of experiments $\mathcal{E}^{(N)}_{\mathbf{\Sigma},f}(\boldsymbol{\theta})$—see, for example, Le Cam (1986, Sect. 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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

$\mathcal{H}_0: \boldsymbol{\tau} \in \mathcal{M}(\mathbf{Q})$ is equivalent to $\mathcal{H}_0: \underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})\ \boldsymbol{\tau} \in \mathcal{M}(\underline{\boldsymbol{\Omega}}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})\ \mathbf{Q})$, a set of linear constraints on the location parameter of the Gaussian shift experiment $\mathcal{E}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})$. The most stringent α-level test for this problem consists in rejecting $\mathcal{H}_0$ whenever

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

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, Sect. 4.3) thus is obtained by substituting $\underline{\boldsymbol{\Delta}}^{(N)}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta})$ for $\boldsymbol{\Delta}$ in (3.1).

In view of Proposition 2.1, the null hypothesis $\mathcal{H}_0 = \bigcup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\bigcup_{\boldsymbol{\Sigma}}\bigcup_f \mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$ (here and in what follows, union on $\boldsymbol{\Sigma}$ 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)–(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 $\mathbf{Q}_\perp^T := \mathbf{I}_p \otimes \begin{pmatrix}\mathbf{L}_{d_1d_2\times d^2}\\ \mathbf{S}_{d_1d_2\times d^2}\end{pmatrix}$, 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 $\mathbf{Q} := \mathbf{I}_p \otimes (\mathbf{U}_{d^2\times d_1d_1}, \mathbf{V}_{d^2\times d_2d_2})$ 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 $\underline{\phi}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}) := I[\underline{Q}_{\boldsymbol{\Sigma},f}(\boldsymbol{\theta}) > \chi^2_{r,1-\alpha}]$, with $r = 2pd_1d_2$ and

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

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 the partially unspecified parameter value $\boldsymbol{\theta}$ under the null. 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:

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

(C1).

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

(b) (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

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

(c) $\hat{\boldsymbol{\theta}}^{(N)}$ 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 $\{\mathbf{t} \in \mathbb{R}^K : \|\sqrt{N}(\mathbf{t} - \boldsymbol{\theta})\| \leq c\}$ is bounded by J, uniformly as N tends to infinity.

The local discreteness Assumption (C1)(a) is a purely technical requirement, with little practical implication as, for fixed size, any estimate can be considered part of a locally asymptotically discrete sequence. Assumptions (C1)(a) and (C1)(b) do 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, etc.).

It is a classical result (see, e.g., Le Cam, 1986, Ch. 11) that, under ULAN (which entails the asymptotic linearity of $\underline{\boldsymbol{\Delta}}^{(N)}_{\boldsymbol{\Sigma}, f}(\boldsymbol{\theta})$) and Assumption (C1), substituting $\hat{\boldsymbol{\theta}}^{(N)}$ for $\boldsymbol{\theta}$ in (3.4) has no influence on the asymptotic behavior of $\underline{\phi}_{\boldsymbol{\Sigma}, f}(\boldsymbol{\theta})$ or hence on its local asymptotic optimality. Thus the LAN property straightforwardly allows for building locally and asymptotically optimal testing procedures, under fixed $\boldsymbol{\Sigma}$ and f. The scatter matrix $\boldsymbol{\Sigma}$ is unknown and consequently plays the role of a nuisance parameter. However, we can replace it by $\hat{\boldsymbol{\Sigma}}$, a consistent estimator of $a\boldsymbol{\Sigma}$ for some fixed $a > 0$. More precisely, we assume that $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\mathbf{e}^{(N)}_1, \ldots, \mathbf{e}^{(N)}_N)$ is such that the following condition exists.

(D1). There exists 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 $\mathbf{e}^{(N)}_t(\boldsymbol{\theta})$'s.

Note that it is sufficient to estimate $a\boldsymbol{\Sigma}$ rather than $\boldsymbol{\Sigma}$ itself because the central sequence and also the information matrix are insensitive to a variation of a. The resulting procedure $\underline{\phi}_{\hat{\boldsymbol{\Sigma}}, f}(\hat{\boldsymbol{\theta}}^{(N)})$ allows for testing the parametric null hypothesis $\mathcal{H}_0(f) = \bigcup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \bigcup_{\boldsymbol{\Sigma}} \mathcal{H}^{(N)}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, f)$.

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

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

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

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

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 $\mathbf{Q}^{(21)}:=\mathbf{I}_p\otimes(\widetilde{\mathbf{U}}_{d^2\times dd_1},\mathbf{V}_{d^2\times d_2d_2})$, where $\mathbf{V}$ is defined in (3.3) and $\widetilde{\mathbf{U}}=\left(\begin{pmatrix}\mathbf{I}_{d_1}\\ \mathbf{0}_{d_2\times d_1}\end{pmatrix}\otimes\mathbf{I}_d\right)$. 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 $\phi_f^{(21)}:=I[Q_f^{(21)}>\chi^2_{r_1,1-\alpha}]$, where

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

$$\lim_{N\to+\infty} \mathrm{E}_{\mathrm{P}^{(N)}_{\boldsymbol{\Sigma},f,\boldsymbol{\theta}+(1/\sqrt{N})\boldsymbol{\tau}}}[\phi_f^{(21)}] = 1 - F^{r_1}_{\chi^2}(\chi^2_{r_1,1-\alpha};\psi^2_{2,f}(\boldsymbol{\tau},\boldsymbol{\theta},\boldsymbol{\Sigma})),$$

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

$$\delta^{(21)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}} = \boldsymbol{\tau}^T[\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}} - \mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}^{(21)}((\mathbf{Q}^{(21)})^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}\mathbf{Q}^{(21)})^{-1}(\mathbf{Q}^{(21)})^T\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}]\boldsymbol{\tau}. \qquad \textbf{(3.10)}$$

3.3. Affine Invariance

The null hypothesis $\mathcal{H}_0$ under which $\mathbf{X}^{(2)} \not\Leftrightarrow \mathbf{X}^{(1)}$ is invariant under block-diagonal-affine transformations, in the sense that for $\forall\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, $\mathbf{I}_p \otimes ((\mathbf{B}^{-1})^T \otimes \mathbf{B})\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, for all full rank matrices $\mathbf{B}$ of the form $\mathbf{B} = \left(\begin{smallmatrix}\mathbf{B}^{(11)} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{(22)}\end{smallmatrix}\right)$, 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 $\mathbf{X}_t$, that is, $\mathbf{x} \to \mathbf{Bx}$, with $\mathbf{B} = \left(\begin{smallmatrix}\mathbf{B}^{(11)} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{(22)}\end{smallmatrix}\right)$. Because 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. 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 $\mathbf{CX}_1,\ldots,\mathbf{CX}_N$ and by $\hat{\boldsymbol{\Sigma}}(\mathbf{C})$ the value of $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\mathbf{e}_1^{(N)},\ldots,\mathbf{e}_N^{(N)})$ obtained from the transformed residuals $\mathbf{Ce}_1^{(N)},\ldots,\mathbf{Ce}_N^{(N)}$.

(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 $j = 1,\ldots,p$, where $\hat{\mathbf{A}}_j(\mathbf{B})$ is the value of $\hat{\mathbf{A}}_j$ obtained from the transformed sample $\mathbf{BX}_1,\ldots,\mathbf{BX}_N$.

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

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 $\mathbf{G}_u(\hat{\boldsymbol{\theta}}^{(N)}(\mathbf{B})) = \mathbf{BG}_u(\hat{\boldsymbol{\theta}}^{(N)})\mathbf{B}^{-1}$ 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 $\mathcal{M}(\mathbf{Q})$.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

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

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, that is, the group of affine transformations $\mathbf{x} \to \mathbf{B}\mathbf{x}$, 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, that is, the group of affine transformations $\mathbf{x} \to \mathbf{B}\mathbf{x}$, 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 and also 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. $\mathcal{H}_0^{(21)}$), we are not able to construct procedures that are invariant under block-upper-triangular (resp. block-lower-triangular) affine transformations.

3.4. 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 because 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

$$\underline{\mathbf{\Delta}}_{\mathbf{\Sigma},\mathcal{N}}(\boldsymbol{\theta}) = \mathbf{M}^{(N)}(\boldsymbol{\theta})((N-1)^{1/2}(\operatorname{vec}\mathbf{\Lambda}_{1,\mathbf{\Sigma},\mathcal{N}}^{(N)}(\boldsymbol{\theta}))^T, \ldots, (\operatorname{vec}\mathbf{\Lambda}_{N-1,\mathbf{\Sigma},\mathcal{N}}^{(N)}(\boldsymbol{\theta}))^T)^T,$$

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

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

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

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

The next theorem follows from a general result established by Hallin and Paindaveine (2005). A proof is provided in Saidi and Roy (2006). It gives the asymptotic distribution-freeness, in addition to the local powers and optimality properties of the pseudo-Gaussian test. These results allow us to compute asymptotic relative efficiencies. Indeed, the Gaussian test will serve as a benchmark in Section 4.3.

THEOREM 3.1. *Assume that Assumptions (A1), (B1)–(B3), and (C1) hold. Consider the sequence of tests $\phi_{\mathcal{N}}$ (resp. $\phi^{(12)}_{\mathcal{N}}$ or $\phi^{(21)}_{\mathcal{N}}$) that reject the null hypothesis $\mathcal{H}_0$ (resp. $\mathcal{H}^{(12)}_0$ or $\mathcal{H}^{(21)}_0$) whenever $Q_{\mathcal{N}}$ (resp. $Q^{(12)}_{\mathcal{N}}$ or $Q^{(21)}_{\mathcal{N}}$) exceeds the $1-\alpha$-quantile $\chi^2_{r,1-\alpha}$ of a chi-square variable with $r = 2pd_1d_2$ (resp. $r = r_1 = pd_1d_2$) degrees of freedom. Then*

- *(i) $Q_{\mathcal{N}}$ (resp. $Q^{(12)}_{\mathcal{N}}$ or $Q^{(21)}_{\mathcal{N}}$) is asymptotically chi-square with r (resp. r_1) degrees of freedom under $\mathcal{H}_0$ (resp. $\mathcal{H}^{(12)}_0$ or $\mathcal{H}^{(21)}_0$).*
- *(ii) $Q_{\mathcal{N}}$ (resp. $Q^{(12)}_{\mathcal{N}}$ or $Q^{(21)}_{\mathcal{N}}$) is noncentral chi-square with r (resp. r_1) degrees of freedom and with noncentrality parameter $\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$, given by (3.6) (resp. $\delta^{(12)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ defined by (3.8) or $\delta^{(21)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ defined by (3.10)), under local alternatives $\mathcal{H}^{(N)}(\boldsymbol{\theta}+N^{-1/2}\boldsymbol{\tau},\boldsymbol{\Sigma},f)$.*
- *(iii) The sequence of tests $\phi_{\mathcal{N}}$ (resp. $\phi^{(12)}_{\mathcal{N}}$ or $\phi^{(21)}_{\mathcal{N}}$) is locally asymptotically most stringent for $\mathcal{H}_0$ (resp. $\mathcal{H}^{(12)}_0$ or $\mathcal{H}^{(21)}_0$) against $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}_0}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},\mathcal{N})$ (resp. $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}^{(12)}_0}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},\mathcal{N})$ or $\bigcup_{\boldsymbol{\theta}\notin\boldsymbol{\Theta}^{(21)}_0}\bigcup_{\boldsymbol{\Sigma}}\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},\mathcal{N})$).*

The pseudo-Gaussian procedure $\phi_{\mathcal{N}}$ is invariant under the group of block-diagonal transformations (in the sense that the value of the test statistic obtained from the transformed sample $\mathbf{BX}_1,\ldots,\mathbf{BX}_N$ is the same for all block-diagonal full rank matrices $\mathbf{B}$). It is of course distribution free. However, it is not even

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

asymptotically invariant under the group of *continuous monotone radial transformations* (because it is not measurable with respect to the maximal invariant); see Section 4.1 for a definition of this group.

4. OPTIMAL SEMIPARAMETRIC 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 semiparametric tests will be defined. These new tests are based on multivariate residual ranks and signs. In what follows, we focus on testing for noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ ($\mathcal{H}_0$: $\mathbf{X}^{(2)} \not\Rightarrow \mathbf{X}^{(1)}$). 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 (2.6) are measurable with respect to the spherical distances between the residuals (2.5) 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 $\mathbf{U}_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}) := \boldsymbol{\Sigma}^{-1/2}\mathbf{e}_t^{(N)}(\boldsymbol{\theta})/d_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})$. Under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, the residuals $\mathbf{U}_1^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}),\ldots,\mathbf{U}_N^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})$ 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, $d_1^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}),\ldots,d_N^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})$, under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$, are i.i.d. over the real line, with density function $\tilde{f}(r) = (\mu_{d-1;f})^{-1}r^{d-1}f(r)I[r > 0]$; their ranks, denoted $R_1(\boldsymbol{\theta},\boldsymbol{\Sigma}),\ldots,R_N(\boldsymbol{\theta},\boldsymbol{\Sigma})$, 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 $\Im_{\boldsymbol{\Sigma}}^{(N)} = \{\Im_{\boldsymbol{\Sigma},g}^{(N)}\}$ acting on $(\mathbb{R}^d)^N$ and characterized by

$$\Im_{\boldsymbol{\Sigma},g}^{(N)}(\mathbf{e}_1^{(N)}(\boldsymbol{\theta}),\ldots,\mathbf{e}_N^{(N)}(\boldsymbol{\theta})) := (g(d_1^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}))\boldsymbol{\Sigma}^{1/2}\mathbf{U}_1^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}),\ldots,$$

$$g(d_N^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}))\boldsymbol{\Sigma}^{1/2}\mathbf{U}_N^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})),$$

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 $(\mathbf{U}_1^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}),\ldots,\mathbf{U}_N^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}))$, the ranks $R_t(\boldsymbol{\theta},\boldsymbol{\Sigma})$, $t = 1,\ldots,N$, are a maximal invariant for the group $\Im_{\boldsymbol{\Sigma}}^{(N)}$ of continuous monotone radial transformations.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

Unfortunately, the multivariate ranks and signs cannot be computed from the residuals $\mathbf{e}_t^{(N)} = \mathbf{e}_t^{(N)}(\boldsymbol{\theta}^{(N)})$ because they depend on the unknown scatter matrix $\boldsymbol{\Sigma}$ that can be replaced by $\hat{\boldsymbol{\Sigma}}$ that satisfies Assumption (D1). The corresponding distances from the origin $\hat{d}_t^{(N)}(\boldsymbol{\theta}) := d_t^{(N)}(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}})$ are called *pseudo-Mahalanobis distances*, and the corresponding signs $\hat{\mathbf{U}}_t^{(N)}(\boldsymbol{\theta}) := \mathbf{U}_t^{(N)}(\boldsymbol{\theta}, \hat{\boldsymbol{\Sigma}})$ are called the *pseudo-Mahalanobis signs*, which are also known as *standardized spatial 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.

4.2. Optimal Rank-Based Tests

The semiparametric (signed rank) J-score versions of the cross-covariance matrices (2.6) are

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

where the score 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 score 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_* that is as follows.

(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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

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 $\kappa_d(u) = \Gamma_d(u)/\Gamma(d)$, with $\Gamma_d(u) = \int_0^u r^{d-1}\exp(-r)\,dr$ and $\Gamma(d) = \int_0^{+\infty} r^{d-1}\exp(-r)\,dr = (d-1)\times\cdots\times 1 = (d-1)!$, stands for the incomplete gamma function.

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

$$\hat{\mathbf{T}}_J^{(N)}(\boldsymbol{\theta}) = ((N-1)^{1/2}(\operatorname{vec}\hat{\mathbf{\Gamma}}_{1,J}^{(N)}(\boldsymbol{\theta}))^T, \ldots, (N-i)^{1/2}(\operatorname{vec}\hat{\mathbf{\Gamma}}_{i,J}^{(N)}(\boldsymbol{\theta}))^T, \ldots, (\operatorname{vec}\hat{\mathbf{\Gamma}}_{N-1,J}^{(N)}(\boldsymbol{\theta}))^T)^T$$

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 $\boldsymbol{\theta}$, let us denote

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

the resulting semiparametric (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, that is,

$$\tilde{\mathbf{T}}_J^{(N)} = \hat{\mathbf{T}}_J^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}) = ((N-1)^{1/2}(\operatorname{vec}\tilde{\mathbf{\Gamma}}_{1,J}^{(N)})^T, \ldots, (N-i)^{1/2}(\operatorname{vec}\tilde{\mathbf{\Gamma}}_{i,J}^{(N)})^T, \ldots, (\operatorname{vec}\tilde{\mathbf{\Gamma}}_{N-1,J}^{(N)})^T)^T. \tag{4.1}$$

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

$$\mathbf{\Omega}_{\mathbf{\Sigma},J}(\boldsymbol{\theta}) = \frac{1}{d^2}\,\mathrm{E}[J_1^2(U)]\,\mathrm{E}[J_2^2(U)]\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}, \tag{4.2}$$

where $\mathbf{N}_{\boldsymbol{\theta},\mathbf{\Sigma}}$ is defined in (2.12), and $\tilde{\mathbf{\Omega}}_J = \mathbf{\Omega}_{\hat{\mathbf{\Sigma}},J}(\hat{\boldsymbol{\theta}}^{(N)})$. Building on the previous notations, invariant optimal rank-based tests for $\mathcal{H}_0$ (noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$) are $\tilde{\phi}_J := I[\tilde{Q}_J > \chi^2_{r,1-\alpha}]$, where

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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

$$\tilde{Q}_J^{(12)} = \tilde{\mathbf{\Delta}}_J^T[\tilde{\mathbf{\Omega}}_J^{-1} - \mathbf{Q}^{(12)}((\mathbf{Q}^{(12)})^T\tilde{\mathbf{\Omega}}_J\mathbf{Q}^{(12)})^{-1}(\mathbf{Q}^{(12)})^T]\tilde{\mathbf{\Delta}}_J. \tag{4.4}$$

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 $\tilde{\phi}_J^{(21)} := I[\tilde{Q}_J^{(21)} > \chi^2_{r_1,1-\alpha}]$, where

$$\tilde{Q}_J^{(21)} = \tilde{\mathbf{\Delta}}_J^T[\tilde{\mathbf{\Omega}}_J^{-1} - \mathbf{Q}^{(21)}((\mathbf{Q}^{(21)})^T\tilde{\mathbf{\Omega}}_J\mathbf{Q}^{(21)})^{-1}(\mathbf{Q}^{(21)})^T]\tilde{\mathbf{\Delta}}_J. \tag{4.5}$$

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_*. 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 are 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 notation. 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 $]0,1[$, and put

$$c(d,f,J_1,J_2) = \frac{1}{d^2}\,\frac{\pi_d^2(J_1,f)\eta_d^2(J_2,f)}{\mathrm{E}[J_1^2(U)]\mathrm{E}[J_2^2(U)]}. \tag{4.6}$$

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 $\pi_d(f,f)$.

The following result also follows from Hallin and Paindaveine (2005). A proof is given in Saidi and Roy (2006). It gives the optimal rank-based testing procedures for noncausality in VAR models, their invariance and distribution-freeness features, and their local powers and optimality properties.

THEOREM 4.1. *Assume that Assumptions (A1), (B1)–(B3), (C1), (D1), and (E1) hold. Consider the sequence of aligned rank tests $\tilde{\phi}_J$ (resp. $\tilde{\phi}_J^{(12)}$ or $\tilde{\phi}_J^{(21)}$)*

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

that reject 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.* $r = r_1 = pd_1d_2$*) degrees of freedom. Then*

(i) $\tilde{Q}_J$ *(resp.* $\tilde{Q}_J^{(12)}$ *or* $\tilde{Q}_J^{(21)}$*) is asymptotically chi-square with* r *(resp.* r_1*) 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 *(resp.* r_1*) degrees of freedom and with noncentrality parameter* $c(d,f,J_1,J_2)\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$*, where* $\delta_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ *is given by (3.6) and* $c(d,f,J_1,J_2)$ *by (4.6) (resp.* $c(d,f,J_1,J_2)\delta^{(12)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ *or* $c(d,f,J_1,J_2)\delta^{(21)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$*, where* $\delta^{(12)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ *and* $\delta^{(21)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},\boldsymbol{\tau}}$ *are given by (3.8) and (3.10), respectively), under the sequence of local alternatives,* $\mathcal{H}^{(N)}(\boldsymbol{\theta}+N^{-1/2}\boldsymbol{\tau},\boldsymbol{\Sigma},f)$*, 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)–(B3) and (E1′), the test* $\tilde{\phi}_{f_*}$ *(resp.* $\tilde{\phi}^{(12)}_{f_*}$ *or* $\tilde{\phi}^{(21)}_{f_*}$*) 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 exists 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).

4.3. Asymptotic Relative Efficiencies

In this section, we turn to asymptotic relative efficiencies 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 distributions of the test statistics (4.3)–(4.5) 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 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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

$\phi_{\mathcal{N}}$ (resp. $\phi_{\mathcal{N}}^{(12)}$ or $\phi_{\mathcal{N}}^{(21)}$) yields the asymptotic relative efficiency of these tests with respect to their parametric Gaussian counterparts. The next result follows from Theorems 3.1 and 4.1.

THEOREM 4.2. *Assume that Assumptions (A1), (B1)–(B3), (C1), (D1), and (E1) hold. Then the asymptotic relative efficiency 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* $\phi_{\mathcal{N}}^{(21)}$*), 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 asymptotic relative efficiency value of 1.458 is reached for four-variate Student density with 3 degrees of freedom).

Another interesting result follows from Theorem 4.2 and the multivariate Hodges–Lehmann result (see Hallin and Paindaveine, 2004a, Prop. 10). 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 $\inf_f \mathrm{ARE}_{d,f}(\tilde{\phi}_{SP}, \phi_{\mathcal{N}})$, 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 $\frac{9}{16} = 0.5625$ 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.

The asymptotic relative efficiency obtained in Theorem 4.2 may be regarded as the asymptotic relative efficiency with respect to the traditional Gaussian Wald test (see Lütkepohl, 1991, Sect. 3.6; Boudjellaba et al., 1992). Indeed, it can be proved that the pseudo-Gaussian test and the Gaussian Wald test are

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

asymptotically equivalent under the null hypothesis $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$) and also under local alternatives.

4.4. Construction of Adaptive Tests

The testing procedures $\tilde{\phi}_f$, $\tilde{\phi}_f^{(12)}$, and $\tilde{\phi}_f^{(21)}$ are semiparametric as far as their validity is concerned. However, they remain parametric from the point of view of optimality, in the sense that their optimality properties only hold for the innovation density they were constructed for (the radial density f). Recall indeed that only optimality (not validity) is affected if the innovation density f used in the construction of these tests does not coincide with the true density that governs the distribution of the innovations.

The fully semiparametric approach to the problem consists of constructing procedures that are optimal in the semiparametric model, that is, finding procedures that are (from a local and asymptotic point of view) as efficient as the optimal procedures in the parametric model with correctly specified innovation density. Semiparametric optimal tests (also called adaptive tests) could be obtained by removing the dependence of the tests $\tilde{\phi}_f$, $\tilde{\phi}_f^{(12)}$, and $\tilde{\phi}_f^{(21)}$ on the radial density f by replacing $\tilde{f}$ with an adequate estimate.

For the moment, assume that $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are known. Let $\overset{\vee}{\tilde{f}}$ be any continuous kernel density estimator of $\tilde{f}$ that is measurable with respect to the order statistic of the distances $d_1^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}),\dots,d_N^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})$ and satisfies the following consistency requirement:

$$\mathrm{E}\left[\left\{\varphi_{\overset{\vee}{\tilde{f}}}\circ\overset{\vee}{\tilde{F}}{}^{-1}\left(\frac{R_t(\boldsymbol{\theta},\boldsymbol{\Sigma})}{N+1}\right)\overset{\vee}{\tilde{F}}{}^{-1}\left(\frac{R_{t-i}(\boldsymbol{\theta},\boldsymbol{\Sigma})}{N+1}\right)\right.\right.$$
$$\left.\left.-\,\varphi_{\tilde{f}}\circ\tilde{F}^{-1}\left(\frac{R_t(\boldsymbol{\theta},\boldsymbol{\Sigma})}{N+1}\right)\tilde{F}^{-1}\left(\frac{R_{t-i}(\boldsymbol{\theta},\boldsymbol{\Sigma})}{N+1}\right)\right\}^2\Bigg|\overset{\vee}{\tilde{f}}\right]=o_{\mathrm{P}}(1) \tag{4.7}$$

under $\mathcal{H}^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma},f)$ as $N\to\infty$, where $\overset{\vee}{\tilde{F}}$ denotes the cumulative distribution function associated with $\overset{\vee}{\tilde{f}}$ and $\varphi_{\tilde{f}}=-2((\tilde{f}^{1/2})'/\tilde{f}^{1/2})$. A possible choice for $\overset{\vee}{\tilde{f}}$ is given in Hájek and Šidák (1967, Sect. 1.5 of Ch. VII). Of course, many other available density estimation methods can be considered, as long as condition (4.7) is fulfilled. An adaptive (under specified $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$) version of $\boldsymbol{\Gamma}_{i,\boldsymbol{\Sigma},f}^{(N)}(\boldsymbol{\theta})=\boldsymbol{\Gamma}_{i,\boldsymbol{\Sigma},J}^{(N)}(\boldsymbol{\theta})$ where

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

with $J_1 = \varphi_f \circ \tilde{F}^{-1}$ and $J_2 = \tilde{F}^{-1}$ is given by

$$\mathbf{\Gamma}_{i,\mathbf{\Sigma}}^{(N)}(\boldsymbol{\theta}) = (N-i)^{-1}(\mathbf{\Sigma}^{-1/2})^T \times \left(\sum_{t=i+1}^{N} \breve{\varphi}_f \circ \breve{\tilde{F}}^{-1}\left(\frac{R_t(\boldsymbol{\theta},\mathbf{\Sigma})}{N+1}\right) \breve{\tilde{F}}^{-1}\left(\frac{R_{t-i}(\boldsymbol{\theta},\mathbf{\Sigma})}{N+1}\right) \times \mathbf{U}_t^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma})(\mathbf{U}_{t-i}^{(N)}(\boldsymbol{\theta},\mathbf{\Sigma}))^T \right)(\mathbf{\Sigma}^{1/2})^T,$$

where we let $\breve{\varphi}_f(r) = \varphi_{\breve{f}}(r) + [(d-1)/r]$ (because $\varphi_f(r) = \varphi_{\tilde{f}}(r) + [(d-1)/r]$).

In practice, $\mathbf{\Sigma}$ is unknown, and the parameter value $\boldsymbol{\theta}$ is partially unspecified under the null hypothesis. The corresponding aligned version of the cross-covariance matrices is then given by

$$\tilde{\mathbf{\Gamma}}_i^{(N)} := (N-i)^{-1}(\hat{\mathbf{\Sigma}}^{-1/2})^T \times \left(\sum_{t=i+1}^{N} \tilde{\varphi}_f \circ \tilde{\tilde{F}}^{-1}\left(\frac{\tilde{R}_t}{N+1}\right) \tilde{\tilde{F}}^{-1}\left(\frac{\tilde{R}_{t-i}}{N+1}\right) \tilde{\mathbf{U}}_t^{(N)}(\boldsymbol{\theta})(\tilde{\mathbf{U}}_{t-i}^{(N)}(\boldsymbol{\theta}))^T \right)(\hat{\mathbf{\Sigma}}^{1/2})^T$$

with $\tilde{\varphi}_f(r) = \varphi_{\tilde{f}}(r) + [(d-1)/r]$ where $\tilde{\tilde{f}}$ and its cumulative distribution function $\tilde{\tilde{F}}$ are the counterparts of $\breve{f}$ and $\breve{F}$ computed from the order statistic of the aligned pseudo-Mahalanobis distances $\tilde{d}_t^{(N)}$.

The resulting adaptive aligned central sequence is $\tilde{\mathbf{\Delta}} := \mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\tilde{\mathbf{T}}^{(N)}$, with

$$\tilde{\mathbf{T}}^{(N)} = ((N-1)^{1/2}(\operatorname{vec}\tilde{\mathbf{\Gamma}}_1^{(N)})^T, \ldots, (N-i)^{1/2}(\operatorname{vec}\tilde{\mathbf{\Gamma}}_i^{(N)})^T, \ldots, (\operatorname{vec}\tilde{\mathbf{\Gamma}}_{N-1}^{(N)})^T)^T.$$

The construction of adaptive tests is then completely straightforward: simply replace the central sequence in the parametric optimal procedures by the adaptive central sequence. Note that besides the central sequence, the Fisher information matrix also depends on f, through $((\mu_{d+1;f}/\mu_{d-1;f})\mathcal{I}_{d,f})$. Therefore, we still need consistent estimates of $\mathcal{I}_{d,f} = \mathrm{E}[(\varphi_f \circ \tilde{F}^{-1}(U))^2]$ and of $\mathcal{V}_{d,f} = (\mu_{d+1;f}/\mu_{d-1;f}) = \mathrm{E}[(\tilde{F}^{-1}(U))^2]$. Such estimators are provided by

$$\tilde{\mathcal{I}}_{d,f} = \frac{1}{N}\sum_{t=1}^{N}\left(\tilde{\varphi}_f \circ \tilde{\tilde{F}}^{-1}\left(\frac{\tilde{R}_t}{N+1}\right)\right)^2, \qquad \tilde{\mathcal{V}}_{d,f} = \frac{1}{N}\sum_{t=1}^{N}\left(\tilde{\tilde{F}}^{-1}\left(\frac{\tilde{R}_t}{N+1}\right)\right)^2.$$

Building on the previous notation and working as in Section 6.2 of Hallin and Paindaveine (2004a), an adaptive test for $\mathcal{H}_0$ (noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$) is $\tilde{\phi} := I[\tilde{Q} > \chi^2_{r,1-\alpha}]$, where

$$\tilde{Q} := \frac{d^2}{\tilde{\mathcal{V}}_{d,f}\tilde{\mathcal{I}}_{d,f}}\tilde{\mathbf{\Delta}}^T[\mathbf{N}^{-1}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\mathbf{\Sigma}}} - \mathbf{Q}(\mathbf{Q}^T\mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\mathbf{\Sigma}}}\mathbf{Q})^{-1}\mathbf{Q}^T]\tilde{\mathbf{\Delta}}. \tag{4.8}$$

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

Similarly, an adaptive rank-based test 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 $\tilde{\phi}^{(12)} := I[\tilde{Q}^{(12)} > \chi^2_{r_1,1-\alpha}]$, where

$$\tilde{Q}^{(12)} = \frac{d^2}{\tilde{\mathcal{V}}_{d,f}\tilde{\mathcal{I}}_{d,f}} \tilde{\boldsymbol{\Delta}}^T [\mathbf{N}^{-1}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}} - \mathbf{Q}^{(12)}((\mathbf{Q}^{(12)})^T \mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}} \mathbf{Q}^{(12)})^{-1}(\mathbf{Q}^{(12)})^T]\tilde{\boldsymbol{\Delta}}. \quad \textbf{(4.9)}$$

Similarly, an adaptive rank-based test 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 $\tilde{\phi}^{(21)} := I[\tilde{Q}^{(21)} > \chi^2_{r_1,1-\alpha}]$, where

$$\tilde{Q}^{(21)} = \frac{d^2}{\tilde{\mathcal{V}}_{d,f}\tilde{\mathcal{I}}_{d,f}} \tilde{\boldsymbol{\Delta}}^T [\mathbf{N}^{-1}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}} - \mathbf{Q}^{(21)}((\mathbf{Q}^{(21)})^T \mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}} \mathbf{Q}^{(21)})^{-1}(\mathbf{Q}^{(21)})^T]\tilde{\boldsymbol{\Delta}}. \quad \textbf{(4.10)}$$

Note that, although the problem is multivariate, the resulting adaptive rank tests (4.8)–(4.10) do not suffer from the so-called curse of dimensionality (because only a density of a single variable needs to be estimated). The adaptive version of the parametric tests (3.5), (3.7), and (3.9) could be obtained by means of sample splitting techniques; see Hallin and Werker (1999, 2003).

5. OPTIMAL RANK-BASED TESTS IN VMA MODELS

In this section, the process $\mathbf{X}$ is assumed to be a centered VMA(q) process satisfying a stochastic difference equation of the form

$$\mathbf{X}_t = \boldsymbol{\epsilon}_t - \sum_{j=1}^{q} \mathbf{B}_j \boldsymbol{\epsilon}_{t-j}, \qquad t \in \mathbb{Z}, \quad \textbf{(5.1)}$$

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

The partition of $\mathbf{X}$ into $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ induces a partition of the coefficient matrices $\mathbf{B}_j$ into

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

Denote by

$$\boldsymbol{\theta} := (\mathrm{vec}^T \mathbf{B}_1, \ldots, \mathrm{vec}^T \mathbf{B}_q)^T \quad \textbf{(5.2)}$$

the K-dimensional vector of parameters involved in (5.1); note that $K = qd^2$. We make the following assumption.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

(A1′). The roots of the determinant of the moving average polynomial associated with (5.1) all lie outside the unit disk, that is,

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

Under Assumption (A1′), the moving average polynomial is invertible, and we write

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

The matrix coefficients $\mathbf{H}_u$ are the Green matrices associated with the moving average operator. When it is important to emphasize their dependence on $\boldsymbol{\theta}$, we will write $\mathbf{H}_u(\boldsymbol{\theta})$ instead of $\mathbf{H}_u$.

In the VMA context, the noncausality directions are characterized by linear restrictions on the moving average coefficients. A proof is given in Boudjellaba et al. (1992).

PROPOSITION 5.1. *Suppose that the VMA(q) process* $\mathbf{X} = ((\mathbf{X}_t^{(1)})^T, (\mathbf{X}_t^{(2)})^T)^T$ *in (5.1) 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* $\mathbf{B}_j^{(21)} = \mathbf{0}$ $\forall j = 1, \ldots, q$;

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

(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* $\mathbf{B}_j^{(12)} = \mathbf{B}_j^{(21)} = \mathbf{0}$ $\forall j = 1, \ldots, q$.

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 (5.2) lies in some linear subspace of $\mathbb{R}^K$. Thus, optimal rank-based tests can also be constructed using Le Cam theory. The matrices $\mathbf{Q}$, $\mathbf{Q}^{(12)}$, and $\mathbf{Q}^{(21)}$ characterizing the null hypotheses do not change (Of course, p is replaced by q.) However, the residuals, the central sequence, and the information matrix appearing in the optimal rank-based tests must be adapted to the VMA(q) context. The ULAN property in the VMA model follows from the general result established by Garel and Hallin (1995). Again, to construct optimal rank-based tests, we restrict ourselves to a class of elliptically symmetric densities, and we assume that Assumptions (B1)–(B3) are satisfied.

Let $\mathbf{e}_t^{(N)}(\boldsymbol{\theta})$ be the residuals under the parameter value $\boldsymbol{\theta}$,

$$\mathbf{e}_t^{(N)}(\boldsymbol{\theta}) := \sum_{j=0}^{t-1} \mathbf{H}_j(\boldsymbol{\theta}) \mathbf{X}_{t-j}, \tag{5.3}$$

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

that we decompose into $\mathbf{e}_t^{(N)}(\boldsymbol{\theta}) = d_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{1/2}\mathbf{U}_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})$, where $d_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}) := \|\mathbf{e}_t^{(N)}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}$ and $\mathbf{U}_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma}) := \boldsymbol{\Sigma}^{-1/2}\mathbf{e}_t^{(N)}(\boldsymbol{\theta})/d_t^{(N)}(\boldsymbol{\theta},\boldsymbol{\Sigma})$. Denote by $\mathbf{M}^{(N)}(\boldsymbol{\theta})$ the following $qd^2 \times (N-1)d^2$-dimensional matrix associated with the sequence $\{\mathbf{H}_u(\boldsymbol{\theta})\}$:

$$\mathbf{M}^{(N)}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_d \otimes \mathbf{H}_0^T(\boldsymbol{\theta}) & \mathbf{I}_d \otimes \mathbf{H}_1^T(\boldsymbol{\theta}) & \dots & \mathbf{I}_d \otimes \mathbf{H}_{N-2}^T(\boldsymbol{\theta}) \\ \mathbf{0} & \mathbf{I}_d \otimes \mathbf{H}_0^T(\boldsymbol{\theta}) & \dots & \mathbf{I}_d \otimes \mathbf{H}_{N-3}^T(\boldsymbol{\theta}) \\ & & & \vdots \\ \mathbf{0} & & & \mathbf{I}_d \otimes \mathbf{H}_{N-q-1}^T(\boldsymbol{\theta}) \end{pmatrix}.$$

The central sequence in the VMA case is

$$\underline{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma},f}^{(N)}(\boldsymbol{\theta}) := \mathbf{M}^{(N)}(\boldsymbol{\theta})\mathbf{S}_{\boldsymbol{\Sigma},f}^{(N)}(\boldsymbol{\theta}),$$

where $\mathbf{S}_{\boldsymbol{\Sigma},f}^{(N)}(\boldsymbol{\theta})$ is given by (2.7). Again, the associated information matrix is given by

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

where the constant $\xi_d(f)$ is defined by (2.11) and

$$\begin{aligned} \mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}} &= \lim_{N\to+\infty} \mathbf{M}^{(N)}(\boldsymbol{\theta})[\mathbf{I}_{N-1} \otimes (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}^{-1})](\mathbf{M}^{(N)}(\boldsymbol{\theta}))^T \\ &= \left(\sum_{j=\max(i,i')}^{+\infty} \boldsymbol{\Sigma} \otimes \mathbf{H}_{j-i}^T \boldsymbol{\Sigma}^{-1} \mathbf{H}_{j-i'} \right)_{i,i'=1,\dots q}. \end{aligned} \tag{5.4}$$

The pseudo-Gaussian tests for noncausality are $\phi_{\mathcal{N}} := I[Q_{\mathcal{N}} > \chi^2_{r,1-\alpha}]$, where

$$\begin{aligned} Q_{\mathcal{N}} &:= \underline{Q}_{\hat{\boldsymbol{\Sigma}},\mathcal{N}}(\hat{\boldsymbol{\theta}}^{(N)}) \\ &= (\underline{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\Sigma}},\mathcal{N}}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}))^T [\mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}}^{-1} - \mathbf{Q}(\mathbf{Q}^T \mathbf{N}_{\hat{\boldsymbol{\theta}}^{(N)},\hat{\boldsymbol{\Sigma}}} \mathbf{Q})^{-1} \mathbf{Q}^T] \underline{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\Sigma}},\mathcal{N}}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)}), \end{aligned}$$

where $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\theta}}^{(N)}$ are estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}$ such that Assumptions (C1) and (D1) are satisfied. Similarly we can construct pseudo-Gaussian tests ($\phi_{\mathcal{N}}^{(12)}$ and $\phi_{\mathcal{N}}^{(21)}$) for causality directions in the VMA model.

Optimal rank-based tests can also be obtained in a similar way to the VAR case. Define the aligned J-score version of the central sequence by

$$\tilde{\boldsymbol{\Delta}}_J := \mathbf{M}^{(N)}(\hat{\boldsymbol{\theta}}^{(N)})\tilde{\mathbf{T}}_J^{(N)},$$

where $\tilde{\mathbf{T}}_J^{(N)}$ is defined by (4.1) and the aligned signs and ranks ($\tilde{\mathbf{U}}_t$ and $\tilde{R}_t$) are now obtained from the residuals (5.3). As in the VAR case, let $\tilde{\boldsymbol{\Omega}}_J = \boldsymbol{\Omega}_{\hat{\boldsymbol{\Sigma}},J}(\hat{\boldsymbol{\theta}}^{(N)})$ where $\boldsymbol{\Omega}_{\boldsymbol{\Sigma},J}(\boldsymbol{\theta})$ is defined by (4.2) and $\mathbf{N}_{\boldsymbol{\theta},\boldsymbol{\Sigma}}$ is given by (5.4). The optimal rank-based tests for $\mathcal{H}_0$ (resp. $\mathcal{H}_0^{(12)}$ or $\mathcal{H}_0^{(21)}$) in the VMA case are given by (4.3)

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

(resp. (4.4) or (4.5)). With these definitions and notation, Theorems 3.1, 4.1, and 4.2 readily extend to this VMA case. Adaptive tests can also be constructed.

In the strict VARMA case (both orders are positive), noncausality is characterized by a set of nonlinear constraints on the parameter space; see Boudjellaba et al. (1992). Even if Le Cam theory could be used in this case, construction of the corresponding tests is more difficult than in the VAR or VMA cases. Indeed, nonlinear constraints on the parameter $\boldsymbol{\theta}$ result locally (in the neighborhood of a value of $\boldsymbol{\theta}$ satisfying these constraints) in linear constraints (on the local parameter $\boldsymbol{\tau}$) by considering only the tangent hyperplane (in $\boldsymbol{\theta}$) of the nonlinear constraints. However, these linear constraints on $\boldsymbol{\tau}$ will depend on the parameter value $\boldsymbol{\theta}$, and estimation of $\boldsymbol{\theta}$ has to be made.

6. THE BIVARIATE VAR(1) CASE AND SOME MONTE CARLO RESULTS

As an illustration, and 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, that is, $\gamma_{12} = \gamma_{21} = 0$. We consider four particular cases of the following data generating equation:

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

where the bivariate spherical density of the noise $(u_t, v_t)^T$ is a bivariate normal or Student distribution 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 Y causes X.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

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, 1,000 replications of a bivariate i.i.d. 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 1,000 replications, of length 300, of the process $(X_t, Y_t)^T$. Initial values X_0 and Y_0 were put to zero. To prevent starting values from affecting the stationarity of the generated series, only the subseries of length $N = 100$ (resp. $N = 200$) resulting from dropping the 200 (resp. 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 for 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 estimators of $\boldsymbol{\theta}$:

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

The assumption of $\sqrt{N}$-consistency on the unconstrained estimators is satisfied by the other classical estimators (least squares, maximum likelihood, etc.). 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 $\mathcal{M}(\mathbf{Q})$, that is, $\hat{\boldsymbol{\theta}}^{(N)} = \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\tilde{\boldsymbol{\theta}}$. The matrix Q in this problem is given by $Q^T = \left(\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{smallmatrix}\right)$.

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

1. The Wald test statistic Q^*
2. The pseudo-Gaussian statistics $Q_{\mathcal{N}}$ of Theorem 3.1
3. The statistics $\tilde{Q}_J^{(E)}$ that 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), $\tilde{Q}_{SP}^{(E)}$, $\tilde{Q}_L^{(E)}$, $\tilde{Q}_{vW}^{(E)}$.
4. The statistics $\tilde{Q}_J^{(T)}$ that correspond to $\tilde{Q}_J$ obtained from pseudo-Mahalanobis signs and ranks (the Tyler estimator of the scatter $\hat{\boldsymbol{\Sigma}}_T$ is used). These versions are supposed to allow for 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 $\tilde{Q}_S^{(T)}$, $\tilde{Q}_{SP}^{(T)}$, $\tilde{Q}_L^{(T)}$, $\tilde{Q}_{vW}^{(T)}$.

The asymptotic relative efficiency $\mathrm{ARE}_{d,f}(\tilde{\phi}_J, \phi_{\mathcal{N}})$ of the rank-based test $\tilde{\phi}_J$ with respect to its Gaussian counterpart $\phi_{\mathcal{N}}$, as given by Theorem 4.2, is presented in Table 1 for the four scores and the four densities considered. It is immediately seen that the sign test is less efficient and that the three other scores lead to similar efficiencies with the exception of the one based on the Laplace score, which is more powerful with the Student density $\mathcal{T}_3$. From Theorem 4.2, these asymptotic relative efficiencies do not depend on the estimator of $\boldsymbol{\Sigma}$ that is used, and they are therefore the same for the tests based on the robust estimator $\hat{\boldsymbol{\Sigma}}_T$.

TABLE 1. The asymptotic relative efficiencies ($\mathrm{ARE}_{d,f}(\tilde{\phi}_J, \phi_{\mathcal{N}})$) of the rank-based tests $\tilde{\phi}_J$ with respect to their Gaussian counterparts $\phi_{\mathcal{N}}$ for $f = \mathcal{N}, \mathcal{T}_3, \mathcal{T}_6, \mathcal{T}_9$ and $d = 2$. The constant, Spearman, Laplace, and van der Waerden scores are used.

f	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$
$\mathcal{N}$	0.617	0.970	0.765	1.000
$\mathcal{T}_3$	1.000	1.305	1.778	1.401
$\mathcal{T}_6$	0.752	1.055	1.080	1.085
$\mathcal{T}_9$	0.698	1.013	0.949	1.037

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

TABLE 2. Rejection frequencies in 1,000 replications of Experiment A under Scenario 1 for the Wald test, the Gaussian test, and the optimal rank tests at significance level $\alpha = 0.05$ for various densities f of the innovations and for series lengths $N = 100$ and 200.

N	f	Q^*	Q_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	$\mathcal{N}$	0.064	0.058	0.063	0.063	0.048	0.057	0.064	0.058	0.054	0.055
	$\mathcal{T}_3$	0.063	0.057	0.055	0.059	0.051	0.061	0.054	0.062	0.055	0.055
	$\mathcal{T}_6$	0.063	0.054	0.054	0.055	0.053	0.053	0.052	0.054	0.056	0.054
	$\mathcal{T}_9$	0.057	0.051	0.048	0.057	0.050	0.049	0.041	0.055	0.050	0.045
200	$\mathcal{N}$	0.057	0.052	0.055	0.049	0.047	0.048	0.050	0.046	0.047	0.047
	$\mathcal{T}_3$	0.055	0.056	0.048	0.042	0.050	0.044	0.054	0.043	0.048	0.048
	$\mathcal{T}_6$	0.040	0.038	0.045	0.042	0.045	0.036	0.044	0.042	0.047	0.037
	$\mathcal{T}_9$	0.063	0.057	0.048	0.057	0.048	0.052	0.048	0.053	0.049	0.049

Note: The rank tests are based either on the empirical covariance matrix or on the Tyler estimator, using constant, Spearman, Laplace, and van der Waerden scores.

In the simulation, for each replication, these statistics were compared with their asymptotic critical values. Rejection frequencies under Scenario 1 are reported in Tables 2–4, for two series lengths ($N = 100$ 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 5. 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 1,000 replications generated under Experiment A with the same scenario. The rejection frequencies based on these empirical critical values are reported in Table 6. The standard error of the empirical levels in Tables 2 and 5 is 0.0069. At the 5% significance level, we reject the hypothesis that the true level is 0.05 if the rejection frequency is outside the interval]0.0365, 0.0635[.

6.1. Discussion of the Level and Power under Scenario 1

Rejection frequencies for Experiment A under Scenario 1 are reported in Table 2. 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 3 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 proce-

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

TABLE 3. Rejection frequencies in 1,000 replications of Experiments B and C under Scenario 1 for the Wald test, the Gaussian test, and the optimal rank tests 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^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
B	$\mathcal{N}$	100	1	0.181	0.171	0.119	0.170	0.123	0.164	0.121	0.163	0.128	0.153
			2	0.516	0.478	0.322	0.484	0.355	0.466	0.322	0.467	0.354	0.463
			3	0.881	0.869	0.643	0.854	0.742	0.855	0.642	0.849	0.726	0.844
		200	1	0.296	0.294	0.195	0.286	0.226	0.283	0.192	0.283	0.232	0.280
			2	0.833	0.820	0.595	0.818	0.690	0.816	0.593	0.815	0.684	0.811
			3	0.994	0.994	0.928	0.996	0.959	0.994	0.926	0.996	0.960	0.994
	$\mathcal{T}_3$	100	1	0.207	0.188	0.140	0.190	0.186	0.189	0.145	0.190	0.197	0.195
			2	0.528	0.501	0.389	0.546	0.545	0.541	0.393	0.555	0.568	0.545
			3	0.815	0.796	0.716	0.854	0.858	0.852	0.727	0.859	0.866	0.851
		200	1	0.318	0.307	0.233	0.336	0.352	0.333	0.230	0.333	0.358	0.335
			2	0.790	0.786	0.711	0.863	0.883	0.867	0.714	0.864	0.883	0.873
			3	0.965	0.966	0.948	0.992	0.991	0.992	0.954	0.993	0.991	0.991
	$\mathcal{T}_6$	100	1	0.171	0.152	0.128	0.161	0.144	0.157	0.128	0.160	0.141	0.148
			2	0.538	0.520	0.376	0.527	0.459	0.511	0.383	0.516	0.474	0.510
			3	0.844	0.830	0.675	0.843	0.794	0.841	0.673	0.831	0.787	0.833
		200	1	0.305	0.304	0.224	0.302	0.261	0.302	0.224	0.296	0.260	0.301
			2	0.844	0.833	0.683	0.833	0.810	0.844	0.689	0.835	0.807	0.841
			3	0.987	0.985	0.944	0.990	0.986	0.991	0.943	0.989	0.985	0.991
	$\mathcal{T}_9$	100	1	0.199	0.174	0.139	0.177	0.147	0.171	0.139	0.174	0.148	0.167
			2	0.529	0.495	0.366	0.513	0.465	0.486	0.370	0.502	0.459	0.482
			3	0.871	0.851	0.660	0.838	0.768	0.845	0.672	0.839	0.763	0.841
		200	1	0.311	0.296	0.216	0.306	0.275	0.299	0.221	0.306	0.268	0.298
			2	0.816	0.813	0.653	0.812	0.768	0.817	0.651	0.808	0.762	0.817
			3	0.990	0.990	0.937	0.987	0.979	0.991	0.930	0.987	0.977	0.989
C	$\mathcal{N}$	100	1	0.176	0.158	0.122	0.151	0.129	0.151	0.120	0.145	0.131	0.139
			2	0.534	0.509	0.351	0.517	0.387	0.495	0.349	0.507	0.391	0.486
			3	0.864	0.853	0.642	0.851	0.711	0.843	0.649	0.842	0.700	0.835
		200	1	0.305	0.296	0.201	0.292	0.232	0.285	0.204	0.287	0.237	0.286
			2	0.565	0.533	0.415	0.571	0.573	0.577	0.414	0.580	0.581	0.581
			3	0.994	0.993	0.927	0.993	0.960	0.992	0.921	0.993	0.954	0.988
	$\mathcal{T}_3$	100	1	0.202	0.187	0.135	0.191	0.180	0.185	0.141	0.199	0.199	0.195
			2	0.565	0.533	0.415	0.571	0.573	0.577	0.414	0.580	0.581	0.581
			3	0.837	0.813	0.723	0.871	0.871	0.867	0.728	0.877	0.871	0.868
		200	1	0.310	0.310	0.223	0.332	0.360	0.336	0.235	0.336	0.368	0.336
			2	0.811	0.800	0.755	0.873	0.897	0.871	0.760	0.880	0.904	0.877
			3	0.964	0.967	0.958	0.993	0.997	0.993	0.958	0.994	0.998	0.994
	$\mathcal{T}_6$	100	1	0.202	0.185	0.138	0.187	0.172	0.182	0.139	0.185	0.173	0.176
			2	0.531	0.505	0.390	0.532	0.483	0.522	0.398	0.536	0.484	0.518
			3	0.874	0.857	0.684	0.862	0.823	0.863	0.676	0.854	0.814	0.850
		200	1	0.301	0.282	0.221	0.301	0.282	0.299	0.223	0.305	0.288	0.301
			2	0.833	0.828	0.684	0.841	0.797	0.837	0.678	0.841	0.806	0.835
			3	0.992	0.991	0.948	0.994	0.990	0.994	0.943	0.994	0.989	0.994
	$\mathcal{T}_9$	100	1	0.180	0.157	0.128	0.155	0.123	0.149	0.125	0.147	0.128	0.147
			2	0.548	0.521	0.378	0.525	0.460	0.512	0.378	0.517	0.454	0.507
			3	0.864	0.852	0.657	0.843	0.753	0.835	0.649	0.839	0.744	0.830
		200	1	0.316	0.302	0.211	0.302	0.254	0.298	0.213	0.301	0.257	0.299
			2	0.840	0.829	0.648	0.833	0.782	0.831	0.651	0.830	0.771	0.829
			3	0.988	0.986	0.932	0.989	0.980	0.989	0.934	0.991	0.980	0.990

Note: The rank tests are based either on the empirical covariance matrix or on the Tyler estimator, using constant, Spearman, Laplace, and van der Waerden scores.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

TABLE 4. Rejection frequencies in 1,000 replications of Experiment D under Scenario 1 for the Wald test, the Gaussian test, and the optimal rank tests 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	0.305	0.290	0.178	0.285	0.223	0.282	0.180	0.286	0.229	0.277
			2	0.840	0.834	0.648	0.826	0.722	0.825	0.647	0.816	0.717	0.817
			3	0.995	0.993	0.960	0.993	0.983	0.994	0.956	0.991	0.983	0.992
		200	1	0.524	0.525	0.364	0.512	0.411	0.510	0.361	0.509	0.412	0.515
			2	0.987	0.988	0.936	0.985	0.957	0.987	0.933	0.984	0.954	0.983
			3	1.00	1.00	0.999	1.00	1.00	1.00	0.999	1.00	1.00	1.00
	$\mathcal{T}_3$	100	1	0.347	0.340	0.276	0.357	0.359	0.357	0.279	0.363	0.372	0.361
			2	0.867	0.863	0.791	0.890	0.894	0.889	0.786	0.887	0.896	0.883
			3	0.997	0.997	0.989	0.999	0.997	0.998	0.988	0.999	0.997	0.998
		200	1	0.545	0.547	0.457	0.617	0.670	0.626	0.456	0.627	0.685	0.635
			2	0.990	0.987	0.977	0.995	0.997	0.995	0.977	0.994	0.996	0.995
			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	0.336	0.317	0.243	0.314	0.293	0.314	0.245	0.322	0.285	0.312
			2	0.878	0.878	0.735	0.878	0.834	0.870	0.750	0.866	0.830	0.865
			3	0.999	0.999	0.980	0.999	0.991	0.998	0.982	0.998	0.990	0.998
		200	1	0.552	0.548	0.420	0.557	0.547	0.554	0.425	0.553	0.547	0.549
			2	0.992	0.991	0.955	0.993	0.987	0.993	0.956	0.993	0.989	0.993
			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	0.306	0.294	0.240	0.300	0.253	0.289	0.233	0.299	0.255	0.287
			2	0.866	0.861	0.740	0.854	0.806	0.856	0.734	0.854	0.803	0.854
			3	0.996	0.995	0.979	0.996	0.991	0.996	0.977	0.996	0.990	0.995
		200	1	0.528	0.526	0.421	0.537	0.478	0.536	0.417	0.536	0.476	0.537
			2	0.991	0.988	0.944	0.989	0.983	0.990	0.945	0.989	0.985	0.990
			3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note: The rank tests are based either on the empirical covariance matrix or on the Tyler estimator, using constant, Spearman, Laplace, and van der Waerden scores.

TABLE 5. Rejection frequencies in 1,000 replications of Experiment A under Scenario 2 for the Wald test, the Gaussian test, and the optimal rank tests at significance level $\alpha = 0.05$, with the Gaussian density for the innovations and $N = 100$.

Type	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
$\mathcal{O}_1$	0.920	0.931	0.180	0.558	0.350	0.538	0.082	0.148	0.086	0.174
$\mathcal{O}_2$	0.000	0.000	0.121	0.253	0.146	0.213	0.063	0.074	0.041	0.068
$\mathcal{O}_3$	0.878	0.881	0.196	0.533	0.359	0.531	0.096	0.114	0.093	0.121
$\mathcal{I}_1$	0.006	0.007	0.050	0.048	0.039	0.038	0.057	0.061	0.053	0.048
$\mathcal{I}_2$	0.016	0.013	0.051	0.044	0.038	0.036	0.063	0.057	0.054	0.053
$\mathcal{I}_3$	0.002	0.002	0.050	0.039	0.041	0.039	0.063	0.052	0.048	0.048

Note: The rank tests are based either on the empirical covariance matrix or on the Tyler estimator, using constant, Spearman, Laplace, and van der Waerden scores.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

TABLE 6. Rejection frequencies (based on the empirical critical values) in 1,000 replications of Experiment D under Scenario 2 for the Wald test, the Gaussian test, and the optimal rank tests 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	$\mathcal{O}_1$	0.004	0.014	0.009	0.008	0.009	0.007	0.123	0.065	0.075	0.050
	$\mathcal{O}_2$	0.042	0.041	0.041	0.009	0.015	0.007	0.127	0.111	0.106	0.102
	$\mathcal{O}_3$	0.008	0.007	0.008	0.004	0.006	0.004	0.136	0.124	0.083	0.091
	$\mathcal{I}_1$	0.519	0.525	0.305	0.443	0.496	0.466	0.294	0.429	0.486	0.466
	$\mathcal{I}_2$	0.467	0.479	0.273	0.465	0.503	0.482	0.255	0.429	0.450	0.452
	$\mathcal{I}_3$	0.777	0.786	0.334	0.584	0.616	0.627	0.354	0.639	0.659	0.679
2	$\mathcal{O}_1$	0.000	0.008	0.028	0.000	0.000	0.000	0.515	0.400	0.352	0.286
	$\mathcal{O}_2$	0.076	0.076	0.292	0.152	0.177	0.140	0.467	0.451	0.424	0.409
	$\mathcal{O}_3$	0.000	0.000	0.036	0.000	0.000	0.000	0.589	0.590	0.486	0.501
	$\mathcal{I}_1$	0.985	0.986	0.854	0.966	0.978	0.977	0.855	0.960	0.972	0.975
	$\mathcal{I}_2$	0.972	0.972	0.855	0.966	0.976	0.973	0.842	0.960	0.971	0.963
	$\mathcal{I}_3$	0.999	0.999	0.917	0.994	0.988	0.995	0.924	0.996	0.991	0.998
3	$\mathcal{O}_1$	0.000	0.003	0.203	0.006	0.009	0.005	0.920	0.862	0.826	0.771
	$\mathcal{O}_2$	0.254	0.249	0.764	0.642	0.628	0.609	0.866	0.868	0.829	0.835
	$\mathcal{O}_3$	0.000	0.000	0.351	0.013	0.023	0.008	0.921	0.957	0.907	0.928
	$\mathcal{I}_1$	1.00	1.00	0.997	1.00	1.00	1.00	0.995	1.00	1.00	1.00
	$\mathcal{I}_2$	1.00	1.00	1.00	0.999	0.999	1.00	0.998	0.999	0.999	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

Note: The rank tests are based either on the empirical covariance matrix or on the Tyler estimator, using constant, Spearman, Laplace, and van der Waerden scores-

dures considered. The figures in that table also indicate that the performances of the rank tests based either on the empirical covariance matrix or on a robustified version given by the Tyler estimator are similar. The sign test seems to be the weakest among the semiparametric tests. These conclusions are consistent with the asymptotic relative efficiencies of Table 1. Under the Gaussian density, the Wald test does slightly better than the others. However, as N increases, we observe that the rejection frequencies of the 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 0.994. Under the Student $\mathcal{T}_3$ density (except for $N = 100$ and $m = 1$), the van der Waerden, Laplace, and Spearman tests slightly dominate the Wald test. However, when the degrees of freedom ν increase, the 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 4, 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.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

6.2. Discussion of the Level and Power under Scenario 2

The rejection frequencies for Experiment A under Scenario 2 are reported in Table 5. The rejection frequencies very clearly show that the 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 the 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 the Tyler estimator better resist, but we cannot say that the level is satisfactorily controlled because all rejection frequencies except two are outside the 5% significance limits. There is a tendancy to overreject (four frequencies out of 12 are greater than 0.10). The use of the empirical covariance matrix is clearly inappropriate in that situation because 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 6. It is immediately seen that with observation outliers, the 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 the Tyler estimator. On the other hand, with innovation outliers, there is at least one rank test whose power is similar to those of the 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 whereas the power of the more powerful rank test $\tilde{Q}_{vW}^{(T)}$ is 0.679.

7. TESTING CAUSALITY BETWEEN CANADIAN MONEY STOCK AND INCOME

As an illustration of the causality tests presented in Section 4, we investigate causality relations between the Canadian money stock and nominal national income. Causality in the relations between money stock and income has been much debated in the economic literature. On the one side, some monetarists view the money stock as an independent source of economic disturbance. Some others strongly disagree with that point of view, arguing that the money stock is passively adapting to business conditions and has little independent influence. Causality between Canadian money stock and income has been studied by, among others, Hsiao (1979) and Boudjellaba et al. (1992). The data set consists of a seasonally adjusted quarterly series of nominal gross national product (GNP) and two seasonally adjusted alternative measures of the money stock ($M1$ and $M2$), for Canada, over the period 1955–1977.

As in Boudjellaba et al. (1992), the natural logarithm and a regular difference of all three variables were taken before modeling. Denote

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

$$\mathbf{X}_t^{(1)} = \begin{bmatrix} (1-B)\ln(M1(t)) \\ (1-B)\ln(M2(t)) \end{bmatrix} \qquad \mathbf{X}_t^{(2)} = (1-B)\ln(GNP(t)).$$

Using the Tiao and Box (1981) methodology, a VAR(5) model has been adjusted to the global process $((\mathbf{X}_t^{(1)})^T, \mathbf{X}_t^{(2)})^T$. The autoregressive order was obtained by minimizing the Bayesian information criterion. The stationarity condition and also the diagnostic checks suggested by Tiao and Box (1981) are satisfied. The empirical significance levels associated with the Wald test, the pseudo-Gaussian test, and the rank-based tests are reported in Table 7.

We first test the null hypothesis of noncausality between $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. From Table 7, we observe that all considered tests very significantly reject the null hypothesis at the usual significance level $\alpha = 0.05$. This rejection is in agreement with economic intuition and with Boudjellaba et al. (1992).

Turning to causality directions, we now test $\mathcal{H}_0^{(12)}$: money stock does not Granger cause GNP and $\mathcal{H}_0^{(21)}$: GNP does not Granger cause money stock. From Table 7, it is clear that the Wald test and also our tests very significantly reject the null hypothesis $\mathcal{H}_0^{(12)}$ at the level $\alpha = 0.05$. This conclusion is also compatible with Boudjellaba et al. (1992). Similarly, the Wald test and our tests reject $\mathcal{H}_0^{(21)}$ at the 5% significance level except $\tilde{Q}_S^{(E)}$, $\tilde{Q}_S^{(T)}$, and $\tilde{Q}_L^{(T)}$ whose p-values are, respectively, 0.0514, 0.0730, and 0.0616. Because there are no outliers in these three series, there is no need to use a robust estimator of $\mathbf{\Sigma}$, and among those based on the empirical covariance matrix $\hat{\mathbf{\Sigma}}_E$, only $\tilde{Q}_S^{(E)}$ rejects with a p-value of 0.0514. Therefore, our tests lead us to conclude that income also causes money. This is in disagreement with the results of Boudjellaba et al., in which the Wald test leads (at probability level $\alpha = 0.05$) to the monetarist conclusion that causality is unidirectional, from the money stock to the national income—a view that most economists would criticize. This contradiction may be explained by the difference between the models used. Indeed, even if the transformed series $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are the same, a VARMA(2,5) model was adjusted to the global process in Boudjellaba et al.

TABLE 7. The marginal significance levels of the Wald test, the Gaussian test, and the optimal rank tests in the problem of testing causality between Canadian money stock and income.

	Q^*	$Q_{\mathcal{N}}$	$\tilde{Q}_S^{(E)}$	$\tilde{Q}_{SP}^{(E)}$	$\tilde{Q}_L^{(E)}$	$\tilde{Q}_{vW}^{(E)}$	$\tilde{Q}_S^{(T)}$	$\tilde{Q}_{SP}^{(T)}$	$\tilde{Q}_L^{(T)}$	$\tilde{Q}_{vW}^{(T)}$
$\mathcal{H}_0$	0.0000	0.0000	0.0000	0.0001	0.0012	0.0000	0.0000	0.0000	0.0004	0.0000
$\mathcal{H}_0^{(12)}$	0.0001	0.0005	0.0000	0.0032	0.0052	0.0005	0.0000	0.0003	0.0004	0.0000
$\mathcal{H}_0^{(21)}$	0.0137	0.0129	0.0514	0.0082	0.0411	0.0133	0.0730	0.0120	0.0616	0.0204

Note: The rank tests are based either on the empirical covariance matrix or on the Tyler estimator, using constant, Spearman, Laplace, and van der Waerden scores.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

8. CONCLUSION

In this paper, we have introduced a new parametric (with respect to the density of the noise) test and a class of semiparametric 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 semiparametric 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 semiparametric tests 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 the Tyler (1987) 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 that is almost as powerful as the Wald and pseudo-Gaussian tests. In the presence of observation or innovation outliers, both the 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 semiparametric tests are still biased. In general, they overreject, and the bias is more important when using the empirical covariance matrix estimator.

REFERENCES

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

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

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

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

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

Dufour, J.-M. & E. Renault (1998) Short run and long run causality in time series: Theory. *Econometrica* 66, 1099–1125.

Fang, K.T., S. Kotz, & K.W. Ng (1990) *Symmetric Multivariate and Related Distributions*. Chapman and Hall.

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

Geweke, J. (1984) Inference and causality in economic time series. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, pp. 1102–1144. North-Holland.

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

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

Hájek, J. & Z. Šidák (1967) *Theory of Rank Tests*. Academic Press.

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

Hallin, M. & D. Paindaveine (2002b) Multivariate signed ranks: Randles' interdirections or Tyler's angles? In Y. Dodge (ed.), *Statistical Data Analysis Based on the L1-norm and Related Methods*, pp. 271–282. Basel.

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

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

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

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

Hallin, M. & M.L. Puri (1992) Rank tests for time-series analysis: A survey. In D. Brillinger, E. Parzen, & M. Rosenblatt (eds.), *New Directions in Time Series Analysis*, vol. 45, pp. 111–153. Springer-Verlag.

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

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

Hallin, M. & B.J.M. Werker (2003) Semiparametric efficiency, distribution-freeness and invariance. *Bernoulli* 9, 137–165.

Hsiao, C. (1979) Autoregressive modeling of Canadian money and income data. *Journal of the American Statistical Association* 74, 553–560.

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

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

Koul, H.L. & A. Schick (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*. Springer-Verlag.

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

Lütkepohl, H. (1993) Testing for causation between two variables in higher dimensional VAR models. In H. Schneeweis & K. Zimmerman (eds.), *Studies in Applied Econometrics*, pp. 75–91. Springer-Verlag.

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

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press

Newbold, P. (1982) Causality testing in economics. In O.D. Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, pp. 701–716. North-Holland.
Oja, H. (1999) Affine invariant multivariate sign and rank tests and corresponding estimates. *Scandinavian Journal of Statistics* 26, 319–343.
Oja, H. & D. Paindaveine (2005) Optimal signed-rank tests based on hyperplanes. *Journal of Statistical Planning and Inference* 135, 300–323.
Peters, D. & 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. & L.D. Haugh (1977) Causality in temporal systems: Characterization and survey. *Journal of Econometrics* 5, 265–293.
Puri, M.L. & P.K. Sen (1971) *Nonparametric Methods in Multivariate Analysis*. Wiley.
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. & A.M. Leroy (1987) *Robust Regression and Outlier Detection*. Wiley.
Saidi, A. & R. Roy (2006) Robust Optimal Tests for Causality in Multivariate Time Series. Technical report CRM-3228, Centre de recherches mathématiques, Université de Montréal.
Sims, C.A., J.H. Stock, & M.W. Watson (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. & Y. Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag.
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.
Tiao, G.C. & G.E.P. Box (1981) Modeling multiple time series with applications. *Journal of the American Statistical Association* 76, 802–816.
Tjøstheim, D. (1981) Granger causality in multiple time series. *Journal of Econometrics* 17, 157–176.
Toda, H.Y. & P.C.B. Phillips (1993) Vector autoregressions and causality. *Econometrica* 61, 1367–1393.
Toda, H.Y. & P.C.B. Phillips (1994) Vector autoregressions and causality: A theoretical overview and simulation study. *Econometric Reviews* 13, 259–285.
Tsay, R.S., D. Peña, & A.E. Pankratz (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 E.F. Beckenback (ed.), *Modern Mathematics for Engineers*, pp. 165–190. McGraw-Hill.

https://doi.org/10.1017/S0266466608080377 Published online by Cambridge University Press