Statistical Science
2004, Vol. 19, No. 4, 697–711
DOI 10.1214/088342304000000602

Multivariate Signed-Rank Tests in Vector Autoregressive Order Identification

Marc Hallin and Davy Paindaveine

Abstract. The classical theory of rank-based inference is essentially limited to univariate linear models with independent observations. The objective of this paper is to illustrate some recent extensions of this theory to time-series problems (serially dependent observations) in a multivariate setting (multivariate observations) under very mild distributional assumptions (mainly, elliptical symmetry; for some of the testing problems treated below, even second-order moments are not required). After a brief presentation of the invariance principles that underlie the concepts of ranks to be considered, we concentrate on two examples of practical relevance: (1) the multivariate Durbin–Watson problem (testing against autocorrelated noise in a linear model context) and (2) the problem of testing the order of a vector autoregressive model, testing VAR(p_0) against VAR($p_0 + 1$) dependence. These two testing procedures are the building blocks of classical autoregressive order-identification methods. Based either on pseudo-Mahalanobis (Tyler) or on hyperplane-based (Oja and Paindaveine) signs and ranks, three classes of test statistics are considered for each problem: (1) statistics of the sign-test type, (2) Spearman statistics and (3) van der Waerden (normal score) statistics. Simulations confirm theoretical results about the power of the proposed rank-based methods and establish their good robustness properties.

Key words and phrases: Ranks, signs, Durbin–Watson test, interdirections, elliptic symmetry, autoregressive processes.

1. RANKS, SIGNS AND SEMIPARAMETRIC MODELS

1.1 Rank-Based Methods: From Nonserial Univariate to Multivariate Serial

Rank-based methods for a long time have been essentially limited to statistical models that involve univariate independent observations. Save a few exceptions (such as testing against bivariate dependence, tests based on runs, tests for scale or goodness-of-fit methods that do not address any specific alternative), classical monographs mainly deal with single-response linear models with independent errors: one- and two-sample location, analysis of variance, regression and so forth.

The need for non-Gaussian, distribution-free and robust methods is certainly no less acute in problems that involve multivariate and/or serially dependent (time-series) data. Rank-based methods for multivariate observations attracted much attention in the late fifties and the sixties, leading to a fairly complete theory of hypothesis testing based on componentwise ranks. A unified account of this line of research is given in the monograph by Puri and Sen (1971). Componentwise ranks, however, are not affine-invariant and hence they crucially depend on the (often arbitrary) choice of a coordinate system; as a consequence, they cannot yield distribution-free statistics. The resulting tests are permutation tests. However, if invariance and "distribution-freeness" are lost, there is little reason to consider permutations of componentwise rank vectors rather than permutations of the observations them-

Marc Hallin and Davy Paindaveine are Professors, I.S.R.O., E.C.A.R.E.S. and Département de Mathématique, Université Libre de Bruxelles, Brussels, Belgium (e-mail: mhallin@ulb.ac.be, dpaindav@ulb.ac.be).

selves. The resulting theory, therefore, is not entirely satisfactory.

Interest in an adequate generalization of ranks and signs for multivariate observations (still in the independent case) was revived in the nineties with a series of papers by Oja, Randles, Hettmansperger and their collaborators: see Oja (1999) for a review. The signs and ranks we consider herein belong to this vein, and we refer to Section 1.3 for details.

Despite the fact that some of the earliest and most classical rank tests (such as runs tests and turning point tests) were of a genuine serial nature, no systematic and coherent theory of serial rank-based statistics was constructed until the mid-eighties. The reason for this late interest is probably the confusing idea that since ranks are intimately related with independence or, at least, exchangeability, they are inherently confined to the analysis of independent observations. This idea, however, does not resist closer examination, since ranks, whatever their definition, always should be computed from a series of residuals that reduce to white noise under some null hypothesis to be tested. Serial statistics based on the ranks of univariate observations or residuals were considered in a series of papers (Hallin, Ingenbleek and Puri, 1985; Hallin and Puri, 1988, 1991, 1994); see Hallin and Puri (1992) for a review of rank-based testing in a (univariate) autoregressive moving average (ARMA) context.

The purpose of this paper is to combine these two extensions of the classical theory: time-series in a multivariate setting. Rather than give a general exposition (for which we refer to Hallin and Paindaveine, 2004a, 2005), we concentrate on two important particular problems: (1) a multivariate version of the classical Durbin–Watson test and (2) the tests that allow for autoregressive order identification, namely, the problem of testing VAR(p_0) against VAR(p_0+1) dependence (which reduces to the Durbin–Watson problem for $p_0=0$). In both cases, we limit ourselves to constant, linear and normal rank-weighting functions (the so-called *score functions*), which yield test statistics of the sign, Spearman and van der Waerden types, respectively.

1.2 From Classical Univariate Signed Ranks to Multivariate Signs and Ranks

Denote by $Z_1^{(n)},\dots,Z_n^{(n)}$ an n-tuple of univariate i.i.d. random variables with common density f satisfying the symmetry assumption $f(-z)=f(z)$, $z\in\mathbb{Z}$, and consider the group $\mathcal{G}=\{g_g^{(n)}\}$ of transformations

$$g_g^{(n)}:(Z_1^{(n)},\dots,Z_n^{(n)})\mapsto g_g^{(n)}(Z_1^{(n)},\dots,Z_n^{(n)}) := (g(Z_1^{(n)}),\dots,g(Z_n^{(n)})), \tag{1}$$

where $g:\mathbb{R}\to\mathbb{R}$ is antisymmetric [$g(-z)=-g(z)$], continuous and order-preserving [$z_1<z_2\Rightarrow g(z_1)<g(z_2)$]. The vector of *signed ranks* $(s_1^{(n)}R_{+;1}^{(n)},\dots,s_n^{(n)}R_{+;n}^{(n)})$, where $s_t^{(n)}:=I_{[Z_t^{(n)}>0]}-I_{[Z_t^{(n)}<0]}$ stands for the sign of $Z_t^{(n)}$ and $R_{+;t}^{(n)}$ denotes the rank of $|Z_t^{(n)}|$ among $|Z_1^{(n)}|,\dots,|Z_n^{(n)}|$, constitutes (up to a factor ±1) a maximal invariant for $\mathcal{G}$. This means that, beyond the fact that the signed ranks are invariant statistics [which means they take the same value in the transformed sample $g_g^{(n)}(Z_1^{(n)},\dots,Z_n^{(n)})$ as in the original sample $(Z_1^{(n)},\dots,Z_n^{(n)})$ for all g], any invariant statistic can be expressed as a function of the signed ranks. The invariance principle, which says one should restrict to invariant test statistics, therefore naturally leads to tests based on the signed ranks. Thanks to the fact that the group $\mathcal{G}$ generates the set of all possible symmetric densities f, the resulting signed-rank tests are distribution-free.

Similarly, denote by $\mathbf{Z}_1^{(n)},\dots,\mathbf{Z}_n^{(n)}$ an n-tuple of k-dimensional i.i.d. random vectors with common density $\underline{f}$. The univariate assumption of symmetry is replaced by the assumption of *elliptical symmetry*. We say that a random vector $\mathbf{Z}$, with density $\underline{f}=\underline{f}_{\mathbf{\Sigma},f}$, is elliptically symmetric if there exist a symmetric, positive definite $k\times k$ matrix $\mathbf{\Sigma}$ and a function $f:\mathbb{R}_0^+\to\mathbb{R}_0^+$ satisfying $\int_0^\infty r^{k-1}f(r)\,dr<\infty$, with

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

where $c_{k,f}$ is a normalizing constant and

$$\|\mathbf{\Sigma}^{-1/2}\mathbf{z}\|:=(\mathbf{z}'\mathbf{\Sigma}^{-1}\mathbf{z})^{1/2}$$

denotes the norm of $\mathbf{z}$ in the metric associated with $\mathbf{\Sigma}$ [we write $\mathbf{\Sigma}^{-1/2}$ for the unique upper-triangular $k\times k$ array with positive diagonal elements satisfying $\mathbf{\Sigma}^{-1}=(\mathbf{\Sigma}^{-1/2})'\mathbf{\Sigma}^{-1/2}$]. The contours of $\underline{f}_{\mathbf{\Sigma},f}$ clearly are a family of ellipsoids centered at the origin, the shape of which is characterized by the matrix $\mathbf{\Sigma}$; the nonnegative function f will be called a *radial density*, although it does not integrate to 1. Note that $\mathbf{\Sigma}$ need not

be the covariance matrix of $\mathbf{Z}$; the rank-based Durbin–Watson tests we are describing in Section 3 do not even require finite second-order moments to exist. In practice, of course, both $\mathbf{\Sigma}$ and f remain unspecified nuisance parameters.

The multivariate generalizations of signed ranks we are now considering are based on arguments of invariance with respect to the group $\mathcal{G}_{\mathbf{\Sigma}}$ of continuous order-preserving *radial transformations*—a direct extension to the multivariate setting of the group $\mathcal{G}$ above—and the group $\mathcal{G}_a$ of affine transformations acting on $\mathbb{R}^k$.

Let $d_t = d_{\mathbf{\Sigma};t}^{(n)} := \|\mathbf{\Sigma}^{-1/2}\mathbf{Z}_t^{(n)}\|$. Then $\mathbf{U}_{\mathbf{\Sigma};t}^{(n)} := \mathbf{\Sigma}^{-1/2}\mathbf{Z}_t^{(n)}/d_{\mathbf{\Sigma};t}^{(n)}$ is the unit vector that points in the direction of the *sphericized vector* $\mathbf{\Sigma}^{-1/2}\mathbf{Z}_t^{(n)}$. Clearly, if $\mathbf{Z}_t^{(n)}$ has density (2), then the density of $\mathbf{\Sigma}^{-1/2}\mathbf{Z}_t^{(n)}$ is constant over the spheres centered at $\mathbf{0}$ (this is why we call it sphericized), while $\mathbf{U}_{\mathbf{\Sigma};t}^{(n)}$ is uniform over the unit sphere $\mathcal{S}^{k-1}$ in $\mathbb{R}^k$, just as $s_t^{(n)}$ in the univariate setting is uniform over $\mathcal{S}^0 = \{-1, 1\}$, the unit sphere in $\mathbb{R}$. For each $\mathbf{\Sigma}$, define the *group of continuous order-preserving radial transformations* $\mathcal{G}_{\mathbf{\Sigma}}^{(n)} = \{g_g^{(n)}\}$ with [cf. with (1) above]

$$\begin{aligned} g_g^{(n)} &: (\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \\ &\mapsto g_g^{(n)}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) \\ &:= \Big(g\big(d_{\mathbf{\Sigma};1}^{(n)}\big)\mathbf{\Sigma}^{1/2}\mathbf{U}_{\mathbf{\Sigma};1}^{(n)}, \ldots, g\big(d_{\mathbf{\Sigma};n}^{(n)}\big)\mathbf{\Sigma}^{1/2}\mathbf{U}_{\mathbf{\Sigma};n}^{(n)}\Big), \end{aligned}$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a continuous, strictly increasing function such that $g(0) = 0$ and $\lim_{r\to\infty} g(r) = \infty$. The transformation $g_g^{(n)}$ is *radial* in the sense that, under the action of $g_g^{(n)}$, the residuals $\mathbf{Z}_t = d_{\mathbf{\Sigma};t}\mathbf{\Sigma}^{1/2}\mathbf{U}_{\mathbf{\Sigma};t}$ move along a half line running through the origin in $\mathbb{R}^k$. This group is a generating group for the fixed-$\mathbf{\Sigma}$ submodel and, quite analogous to the univariate case, a maximal invariant for this group is the couple $(\mathbf{U}_{\mathbf{\Sigma}}^{(n)}, \mathbf{R}_{\mathbf{\Sigma}}^{(n)})$, where the matrix $\mathbf{U}_{\mathbf{\Sigma}}^{(n)} = (\mathbf{U}_{\mathbf{\Sigma};1}^{(n)}, \ldots, \mathbf{U}_{\mathbf{\Sigma};n}^{(n)})$ collects the signs of the observations and $\mathbf{R}_{\mathbf{\Sigma}}^{(n)} = (R_{\mathbf{\Sigma};1}^{(n)}, \ldots, R_{\mathbf{\Sigma};n}^{(n)})$ is the vector of the ranks $R_{\mathbf{\Sigma};t}^{(n)}$ of $d_{\mathbf{\Sigma};t}^{(n)}$ among $d_{\mathbf{\Sigma};1}^{(n)}, \ldots, d_{\mathbf{\Sigma};n}^{(n)}$, $t = 1, \ldots, n$.

Similarly, the group $\mathcal{G}_a$ of affine transformations of $\mathbb{R}^k$ generates the fixed-f submodel. Indeed, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have elliptical densities $\underline{f}_{\mathbf{\Sigma}_1,f}$ and $\underline{f}_{\mathbf{\Sigma}_2,f}$, respectively, iff $\mathbf{Z}_2 \stackrel{d}{=} \mathbf{\Sigma}_2^{1/2}\mathbf{\Sigma}_1^{-1/2}\mathbf{Z}_1$, where $\mathbf{\Sigma}_2^{1/2}\mathbf{\Sigma}_1^{-1/2}$ clearly belongs to $\mathcal{G}_a$ ($\stackrel{d}{=}$ denotes equality in distribution).

In view of this, $\mathbf{U}_{\mathbf{\Sigma};t}^{(n)}$ and $R_{\mathbf{\Sigma};t}^{(n)}$ can be considered as multivariate generalizations of the usual signs and ranks of absolute values, respectively. We refer to Hallin and Paindaveine (2003) for a characterization of the testing problems for which this invariance approach makes sense. Of course, when $\mathbf{\Sigma}$ is unspecified, these multivariate signs and ranks cannot be computed from the observations $\mathbf{Z}_t^{(n)}$. In Sections 1.2.1 and 1.2.2, we describe two different *empirical* reconstructions of these multivariate signs and ranks.

1.2.1 *Pseudo-Mahalanobis signs and ranks*: *The Tyler signs and ranks.* The most natural way to deal with the nonspecification of $\mathbf{\Sigma}$ consists of replacing $\mathbf{U}_{\mathbf{\Sigma};t}^{(n)}$ and $R_{\mathbf{\Sigma};t}^{(n)}$ with $\mathbf{U}_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ and $R_{\widehat{\mathbf{\Sigma}};t}^{(n)}$, respectively, where $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Sigma}}^{(n)}$ is some *reasonable* estimator of $\mathbf{\Sigma}$: namely, we require $\widehat{\mathbf{\Sigma}}$ to be root-n consistent (so that this replacement asymptotically has only limited effect) and affine-equivariant (to ensure the affine invariance of the resulting test statistics). A possible choice for $\widehat{\mathbf{\Sigma}}$ is the empirical covariance matrix of the $\mathbf{Z}_t^{(n)}$'s, but this estimate is known to be highly nonrobust and its consistency requires finite moments of order 2. We therefore suggest using Tyler's (1987) estimator of shape. This estimator is defined as $\widehat{\mathbf{\Sigma}}_{\mathrm{Tyl}} := \mathbf{C}_{\mathrm{Tyl}}^{-1}\mathbf{C}_{\mathrm{Tyl}}^{\prime-1}$, where $\mathbf{C}_{\mathrm{Tyl}}$ is the unique upper-triangular $k \times k$ matrix with nonnegative diagonal and upper left element 1 such that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\mathbf{C}_{\mathrm{Tyl}}\mathbf{Z}_t}{\|\mathbf{C}_{\mathrm{Tyl}}\mathbf{Z}_t\|}\right)\left(\frac{\mathbf{C}_{\mathrm{Tyl}}\mathbf{Z}_t}{\|\mathbf{C}_{\mathrm{Tyl}}\mathbf{Z}_t\|}\right)' = \frac{1}{k}\mathbf{I}_k \tag{3}$$

($\mathbf{I}_k$ stands for the $k \times k$ identity matrix). This estimate thus is such that the empirical covariance of the resulting signs $\mathbf{U}_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ coincides with the covariance matrix $\frac{1}{k}\mathbf{I}_k$ of the uniform distribution over the unit sphere $\mathcal{S}^{k-1}$. It is affine-equivariant and, under the assumption that the $\mathbf{Z}_t^{(n)}$'s are i.i.d. with density (2), it can be shown (without making any moment assumption) that $\widehat{\mathbf{\Sigma}}$ is root-n consistent for $a\mathbf{\Sigma}$, where a is some positive constant.

The resulting *Tyler signs* $\mathbf{U}_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ are strictly equivariant under both $\mathcal{G}_{\mathbf{\Sigma}}$ and $\mathcal{G}_a$, but the *Tyler ranks* $R_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ are invariant under $\mathcal{G}_a$ only. However, it can be shown that $\mathbf{U}_{\widehat{\mathbf{\Sigma}};t}^{(n)} - \mathbf{U}_{\mathbf{\Sigma};t}^{(n)}$ and $R_{\widehat{\mathbf{\Sigma}};t}^{(n)} - R_{\mathbf{\Sigma};t}^{(n)}$ are $o_P(1)$ as $n \to \infty$, so that, although the ranks $R_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ are not invariant under $\mathcal{G}_{\mathbf{\Sigma}}$, they are at least *asymptotically invariant* under $\mathcal{G}_{\mathbf{\Sigma}}$, in the sense that they are asymptotically equivalent to the strictly invariant *exact* ranks $R_{\mathbf{\Sigma};t}^{(n)}$. When the choice of $\widehat{\mathbf{\Sigma}}$ is not imposed, we use the somewhat heavier terminology *pseudo-Mahalanobis signs* and *pseudo-Mahalanobis ranks*.

1.2.2 *Hyperplane-based signs and ranks.* Another approach to reconstructing the exact signs $\mathbf{U}^{(n)}_{\boldsymbol{\Sigma};t}$ and the exact ranks $R^{(n)}_{\boldsymbol{\Sigma};t}$ is based on counts of hyperplanes. For the signs, the idea is due to Randles (1989). For any pair $\mathbf{Z}^{(n)}_{t_1}, \mathbf{Z}^{(n)}_{t_2}$, $1 \le t_1 \ne t_2 \le n$, consider the $\binom{n-2}{k-1}$ hyperplanes going through the origin and $k-1$ out of the $n-2$ remaining $\mathbf{Z}^{(n)}_t$'s ($t_1 \ne t \ne t_2$). Define the *interdirection* $c^{(n)}_{t_1t_2}$ as the number of such hyperplanes that separate $\mathbf{Z}^{(n)}_{t_1}$ and $\mathbf{Z}^{(n)}_{t_2}$ (see Figure 1 for an illustration in the bivariate case). Interdirections are invariant under the affine group $\mathcal{G}_a$ and under the group $\mathcal{G}_{\boldsymbol{\Sigma}}$ of radial transformations, irrespective of $\boldsymbol{\Sigma}$. Due to this invariance, it is intuitively clear that $\pi p^{(n)}_{t_1t_2} := \pi c^{(n)}_{t_1t_2}/\binom{n-2}{k-1}$ is a consistent estimate of the angle $\arccos(\mathbf{U}'_{\boldsymbol{\Sigma};t_1}\mathbf{U}_{\boldsymbol{\Sigma};t_2})$ between $\mathbf{U}^{(n)}_{\boldsymbol{\Sigma};t_1}$ and $\mathbf{U}^{(n)}_{\boldsymbol{\Sigma};t_2}$. Interdirections thus allow for a reconstruction of those angles (equivalently, a reconstruction of their cosines $\mathbf{U}'_{\boldsymbol{\Sigma};t_1}\mathbf{U}_{\boldsymbol{\Sigma};t_2}$, since the $\mathbf{U}_{\boldsymbol{\Sigma};t}$'s are unit vectors): quite remarkably, they do the same job, with the same invariance properties, as the *Tyler cosines* $\mathbf{U}'_{\widehat{\boldsymbol{\Sigma}};t_1}\mathbf{U}_{\widehat{\boldsymbol{\Sigma}};t_2}$, but require no estimation of $\boldsymbol{\Sigma}$. The respective advantages of Tyler angles and Randles interdirections are discussed in Hallin and Paindaveine (2002c).

The hyperplane-based cosines $p^{(n)}_{t_1t_2}$ are sufficient for the first problem we treat (Section 3). For the second problem (Section 4), we need the slightly more informative concept of *absolute interdirections* (Hallin and Paindaveine, 2004b, 2005). The basic idea is exactly the same and the same hyperplanes are taken into account as before. However, instead of counting the number of hyperplanes that separate $\mathbf{Z}^{(n)}_{t_1}$ and $\mathbf{Z}^{(n)}_{t_2}$, we now count the number $c^{(n)}_{t;i}$ of hyperplanes that separate $\mathbf{Z}^{(n)}_t$ and the transformed unit vectors $\widehat{\boldsymbol{\Sigma}}^{1/2}\mathbf{u}_i$,

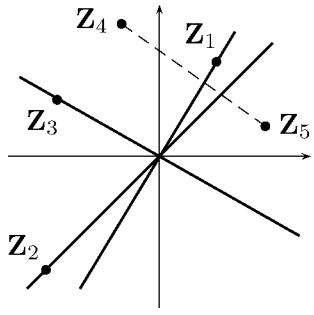

FIG. 1. *An illustration for Randles' interdirections in the bivariate case. The interdirection associated with* $\mathbf{Z}_4$ *and* $\mathbf{Z}_5$ *is* $c_{45} = 2$ *in this small sample of size* $n = 5$ [*two separating hyperplanes out of a total of* $\binom{3}{1} = 3$ *to be considered*].

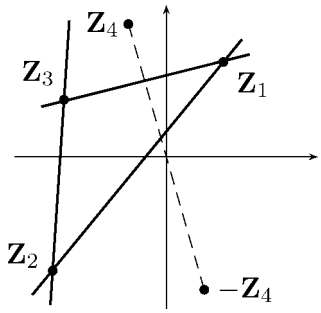

FIG. 2. *An illustration for lift interdirections in the bivariate case. The lift interdirection associated with* $\mathbf{Z}_4$ *is* $\ell_4 = 2$ *within this small sample of size* $n = 4$ [*two separating hyperplanes out of a total of* $\binom{3}{2} = 3$ *hyperplanes to be considered*].

$i = 1, \ldots, k$, where $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ forms the canonical basis of $\mathbb{R}^k$. Then, for the same reasons as above, $\pi p^{(n)}_{t;i} := \pi c^{(n)}_{t;i}/\binom{n-1}{k-1}$ allows for a consistent estimation of the angles $\arccos(\mathbf{U}'_{\boldsymbol{\Sigma};t}\mathbf{u}_i)$, $i = 1, \ldots, k$, so that the vectors $(\cos(\pi p_{t;i}), i = 1, \ldots, k)$ are consistent estimators of the signs $\mathbf{U}_{\boldsymbol{\Sigma};t}$ themselves. Absolute interdirections are invariant under the group of radial transformations; however, they are only asymptotically affine-equivariant in the sense that they converge to strictly equivariant quantities.

Along with the hyperplane-based concepts of signs just described, we propose using a hyperplane-based concept of ranks introduced by Oja and Paindaveine (2004). This concept relies on the so-called *lift interdirections*.

For any $\mathbf{Z}^{(n)}_t$, consider the $\binom{n-1}{k}$ hyperplanes going through k out of the $n-1$ remaining $\mathbf{Z}^{(n)}_{t'}$'s ($t' \ne t$). The lift interdirection $\ell^{(n)}_t$ associated with $\mathbf{Z}^{(n)}_t$ is defined as the number of such hyperplanes that separate $\mathbf{Z}^{(n)}_t$ and $-\mathbf{Z}^{(n)}_t$ (see Figure 2 for an illustration in the bivariate case). Lift interdirections can be shown to converge to some monotone increasing function of the distances $d^{(n)}_{\boldsymbol{\Sigma};t}$, so that their ranks converge to the exact ranks $R^{(n)}_{\boldsymbol{\Sigma};t}$. Again, we are able to reconstruct, as $n \to \infty$, a quantity that depends on the unspecified shape matrix $\boldsymbol{\Sigma}$ without estimating it. When used in the procedures described below, the lift interdirection ranks are those associated with a *symmetrized* version of lift interdirections (see Oja and Paindaveine, 2004, for details).

2. THE GENERAL LINEAR MODEL WITH VECTOR AUTOREGRESSIVE ERRORS

The model we are considering throughout is the k-variate general linear model with vector autoregressive (VAR) error terms [the more general case of vector

autoregressive moving average (VARMA) errors could be treated as well; we restrict to the VAR case for the sake of simplicity]. Under this model, the observation is an n-tuple

$$\mathbf{Y}^{(n)} := \begin{pmatrix} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,k} \\ \vdots & \vdots & & \vdots \\ Y_{n,1} & Y_{n,2} & \cdots & Y_{n,k} \end{pmatrix} := \begin{pmatrix} \mathbf{Y}_1' \\ \vdots \\ \mathbf{Y}_n' \end{pmatrix}$$

of k-variate random vectors that satisfies

$$\mathbf{Y}^{(n)} = \mathbf{X}^{(n)}\boldsymbol{\beta} + \mathbf{V}^{(n)}, \tag{4}$$

where

$$\mathbf{X}^{(n)} := \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{pmatrix} := \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix}$$

and

$$\boldsymbol{\beta} := \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,k} \\ \vdots & \vdots & & \vdots \\ \beta_{m,1} & \beta_{m,2} & \cdots & \beta_{m,k} \end{pmatrix} := \begin{pmatrix} c\boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_m' \end{pmatrix}$$

denote an $n \times m$ matrix of constants (the design matrix) and the $m \times k$ regression parameter, respectively. Instead of the traditional assumption that the error term

$$\mathbf{V}^{(n)} := \begin{pmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,k} \\ \vdots & \vdots & & \vdots \\ V_{n,1} & V_{n,2} & \cdots & V_{n,k} \end{pmatrix} := \begin{pmatrix} \mathbf{V}_1' \\ \vdots \\ \mathbf{V}_n' \end{pmatrix}$$

is white noise, we rather assume $(\mathbf{V}_t,\ t = 1, \ldots, n)$ to be a finite realization (of length n) of the VAR(p) process generated by

$$\mathbf{V}_t = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{V}_{t-i} + \boldsymbol{\varepsilon}_t, \quad t \in \mathbb{Z}, \tag{5}$$

where $\{\boldsymbol{\varepsilon}_t \mid t \in \mathbb{Z}\}$ is a k-dimensional white-noise process with elliptical density (2). Under (4) and (5),

$$\mathbf{Y}_t = \boldsymbol{\beta}'\mathbf{x}_t + \sum_{u=0}^{t-1} \mathbf{G}_u \boldsymbol{\varepsilon}_{t-u} + \mathbf{r}_t, \quad t = 1, \ldots, n, \tag{6}$$

with matrices $\mathbf{G}_u$ (the Green's matrices of the VAR operator) characterized by the linear recursion $\mathbf{G}_u = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{G}_{u-i}$, $u \in \mathbb{Z}$, and initial conditions $\mathbf{G}_0 = \mathbf{I}_k$, $\mathbf{G}_{-1} = \mathbf{G}_{-2} = \cdots = \mathbf{G}_{-p+1} = \mathbf{0}$. The remainder term $\mathbf{r}_t$ is related to the influence of the unobserved initial values $\mathbf{V}_0, \ldots, \mathbf{V}_{-p+1}$. It is easy to see that, under the traditional VAR stationarity assumptions, $\lim_{t\to\infty} \Lambda^t \mathbf{r}_t$ is bounded in probability, where $1 < \Lambda$ is the modulus of the smallest root of the characteristic polynomial associated with (5).

Letting $\boldsymbol{\theta} := (\mathrm{vec}'(\boldsymbol{\beta}'), \mathrm{vec}'(\mathbf{A}_1), \ldots, \mathrm{vec}'(\mathbf{A}_p))' \in \mathbb{R}^{km+k^2p} =: \mathbb{R}^K$, we write $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\boldsymbol{\Sigma},f}$ for the probability distribution of the observation $\mathbf{Y}^{(n)}$ under (6).

3. RANK-BASED DURBIN–WATSON TESTS

3.1 The Gaussian Durbin–Watson Test

Consider the first-order version ($p = 1$) of the general model described in Section 2. Writing $\mathbf{A}$ instead of $\mathbf{A}_1$, (6) takes the form

$$\mathbf{Y}_t = \boldsymbol{\beta}'\mathbf{x}_t + \sum_{u=0}^{t-1} \mathbf{A}^u \boldsymbol{\varepsilon}_{t-u} + \mathbf{A}^t \mathbf{V}_0, \quad t = 1, \ldots, n. \tag{7}$$

The Durbin–Watson testing problem deals with the null hypothesis that $\mathbf{V}_t$ is white noise, that is, that $\mathbf{A} = \mathbf{0}$. Under this hypothesis, the observations are serially independent, of the form $\mathbf{Y}_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\varepsilon}_t$. The regression parameter $\boldsymbol{\beta}$, as well, of course, as the underlying elliptic density (the shape matrix $\boldsymbol{\Sigma}$ and the radial density f of $\boldsymbol{\varepsilon}_t$), remain unspecified.

The multivariate version of the traditional (Gaussian) Durbin–Watson procedure relies on the following test statistic. Denote by $\hat{\boldsymbol{\beta}}^{(n)}_{\mathcal{N}} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ the usual least squares estimate of $\boldsymbol{\beta}$ and denote by $\mathbf{Z}_t := \mathbf{Y}_t - \hat{\boldsymbol{\beta}}^{(n)\prime}_{\mathcal{N}} \mathbf{x}_t$ the corresponding estimated residuals. Write $\widehat{\boldsymbol{\Sigma}}_{\mathcal{N}} := \frac{1}{n}\sum_{t=1}^{n} \mathbf{Z}_t\mathbf{Z}_t'$ for the empirical residual covariance matrix. The null hypothesis of serially independent errors is rejected (at asymptotic level α) whenever

$$\begin{aligned} W^{(n)}_{\mathrm{DW}} &:= \frac{1}{n-1}\sum_{s,t=2}^{n} (\mathbf{Z}_s'\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{N}}\mathbf{Z}_t)(\mathbf{Z}_{s-1}'\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{N}}\mathbf{Z}_{t-1}) \\ &= (n-1)\left\| \frac{1}{n-1}\sum_{t=2}^{n} \widehat{\boldsymbol{\Sigma}}^{-1/2}_{\mathcal{N}}\mathbf{Z}_t\mathbf{Z}_{t-1}'\widehat{\boldsymbol{\Sigma}}^{\prime-1/2}_{\mathcal{N}} \right\|^2 \end{aligned} \tag{8}$$

exceeds the $(1-\alpha)$ quantile $\chi^2_{k^2;1-\alpha}$ of the chi-squared distribution with k^2 degrees of freedom [$\|\mathbf{M}\| := (\sum_{i,j=1}^{k}(M_{ij})^2)^{1/2}$ stands for the Euclidean norm of the $k \times k$ matrix $\mathbf{M} = (M_{ij})$]. Being the sum of all residual squared cross-correlation coefficients at lag 1, this test statistic has a clear intuitive interpretation: in the univariate case, it reduces to the squared residual autocorrelation coefficient of order 1.

3.2 Multivariate Signed-Rank Durbin–Watson Tests

The Gaussian test just described requires finite second-order moments, whereas the signed-rank tests

we now consider remain valid under arbitrarily heavy tails: only finite radial Fisher information $(\int_0^\infty[(-f'/f)(r)]^2 r^{k-1} f(r)\,dr)/(\int_0^\infty r^{k-1} f(r)\,dr)$ is required. Any consistent sequence of estimates of $\boldsymbol{\beta}$ can be substituted for the Gaussian $\hat{\boldsymbol{\beta}}^{(n)}_{\mathcal{N}}$ [consistency here means "consistency under the null hypothesis at the appropriate (optimal) rate"; the definition of this rate depends on the asymptotic behavior of the regression constants; see Hallin and Paindaveine, 2005, Section 2.1]. If, however, the tests are to remain valid under infinite second-order moments, robust estimators that resist heavy-tailed distributions such as the M estimators proposed by Davis and Wu (1997) should be used; denote by $\hat{\boldsymbol{\beta}}^{(n)}$ such an estimator.

The residuals associated with $\hat{\boldsymbol{\beta}}^{(n)}$ are obtained as in Section 3.1. Denote by $\mathbf{U}_t^{(n)}$ and $R_t^{(n)}$ the *sign* and the *rank* (among $\mathbf{Z}_1,\ldots,\mathbf{Z}_n$), respectively, of the residual $\mathbf{Z}_t$. In principle, any combination of a pseudo-Mahalanobis or hyperplane-based sign with a pseudo-Mahalanobis or hyperplane-based rank can be considered (four possibilities, thus). However, hybrid statistics that mix the two types (Tyler signs, e.g., with lift-interdirection ranks) are somewhat incoherent, so we restrict ourselves to combining signs and ranks of the same type (either pseudo-Mahalanobis or hyperplane-based); we use the same notation for both cases.

We concentrate on three versions of signed-rank Durbin–Watson statistics:

1. A multivariate Durbin–Watson statistic of the sign-test type,

$$\underset{\sim}{W}^{(n)}_{\mathrm{DW;sign}} := \frac{k^2}{n-1}\sum_{s,t=2}^{n}(\mathbf{U}_s'\mathbf{U}_t)(\mathbf{U}_{s-1}'\mathbf{U}_{t-1}) = \frac{k^2}{n-1}\left\|\sum_{t=2}^{n}\mathbf{U}_t\mathbf{U}_{t-1}'\right\|^2. \tag{9}$$

2. A multivariate Durbin–Watson statistic of the Spearman type,

$$\underset{\sim}{W}^{(n)}_{\mathrm{DW;Sp}} := \frac{9k^2}{(n-1)(n+1)^4}\cdot\sum_{s,t=2}^{n}R_s^{(n)}R_{s-1}^{(n)}R_t^{(n)}R_{t-1}^{(n)}\cdot(\mathbf{U}_s'\mathbf{U}_t)(\mathbf{U}_{s-1}'\mathbf{U}_{t-1}) \tag{10}$$

$$= \frac{9k^2}{(n-1)(n+1)^4}\cdot\left\|\sum_{t=2}^{n}R_t^{(n)}R_{t-1}^{(n)}\mathbf{U}_t\mathbf{U}_{t-1}'\right\|^2.$$

3. A multivariate Durbin–Watson statistic of the van der Waerden type,

$$\underset{\sim}{W}^{(n)}_{\mathrm{DW;vdW}} := \frac{1}{n-1}\sum_{s,t=2}^{n}\Phi_k^{-1}\left(\frac{R_s^{(n)}}{n+1}\right)\Phi_k^{-1}\left(\frac{R_{s-1}^{(n)}}{n+1}\right)\cdot\Phi_k^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)\Phi_k^{-1}\left(\frac{R_{t-1}^{(n)}}{n+1}\right)\cdot(\mathbf{U}_s'\mathbf{U}_t)(\mathbf{U}_{s-1}'\mathbf{U}_{t-1}) \tag{11}$$

$$= \frac{1}{n-1}\left\|\sum_{t=2}^{n}\Phi_k^{-1}\left(\frac{R_t^{(n)}}{n+1}\right)\cdot\Phi_k^{-1}\left(\frac{R_{t-1}^{(n)}}{n+1}\right)\mathbf{U}_t\mathbf{U}_{t-1}'\right\|^2,$$

where, denoting by $F^{-1}_{\chi^2_k}(u)$ the quantile function of the chi-squared variable with k degrees of freedom, $\Phi_k^{-1}(u) := \sqrt{F^{-1}_{\chi^2_k}(u)}$, $u \in]0,1[$.

In all cases, the null hypothesis of serially independent errors is rejected whenever the test statistic exceeds the $(1-\alpha)$ quantile of a chi-squared distribution with k^2 degrees of freedom.

3.3 Asymptotic Relative Efficiencies

The asymptotic relative efficiencies (ARE; with respect to the traditional Gaussian procedure described in Section 3.1) of the signed-rank tests in Section 3.2 were derived by Hallin and Paindaveine (2005), who also established a multivariate serial version of the classical Chernoff–Savage result. This result shows that the asymptotic relative efficiency [with respect to the Gaussian procedure based on (8)] of the van der Waerden tests (list item 3) based on (11) is uniformly larger than 1. Some of these ARE values are reported in Table 1 for several elliptic Student distributions and several dimensions of the observation space. Note that the elliptical Student distributions considered have strictly more than 2 degrees of freedom in order for the Gaussian procedure to be valid.

TABLE 1
AREs with respect to the Gaussian procedure of the sign-type (S), Spearman-type (Sp) and van der Waerden-type (vdW) Durbin–Watson tests under various k-variate Student and normal densities, $k = 1, 2, 4, 6, 10$

		Degrees of freedom of the underlying t density								
k	Test	3	4	5	6	8	10	15	20	∞
1	S	0.657	0.563	0.519	0.494	0.467	0.453	0.435	0.427	0.405
	Sp	1.299	1.139	1.070	1.032	0.992	0.972	0.948	0.938	0.912
	vdW	1.356	1.176	1.106	1.071	1.038	1.024	1.010	1.005	1.000
2	S	1.000	0.856	0.790	0.752	0.711	0.689	0.662	0.650	0.617
	Sp	1.305	1.152	1.089	1.055	1.022	1.006	0.990	0.983	0.970
	vdW	1.400	1.204	1.125	1.085	1.047	1.030	1.013	1.007	1.000
4	S	1.266	1.084	1.000	0.952	0.900	0.872	0.838	0.823	0.781
	Sp	1.189	1.050	0.994	0.966	0.941	0.930	0.922	0.920	0.924
	vdW	1.458	1.242	1.153	1.106	1.061	1.039	1.018	1.010	1.000
6	S	1.373	1.176	1.085	1.033	0.977	0.946	0.910	0.893	0.847
	Sp	1.115	0.982	0.929	0.903	0.879	0.870	0.865	0.865	0.880
	vdW	1.493	1.267	1.172	1.122	1.071	1.047	1.022	1.013	1.000
10	S	1.467	1.256	1.159	1.104	1.043	1.011	0.972	0.954	0.905
	Sp	1.039	0.909	0.857	0.831	0.808	0.799	0.795	0.797	0.823
	vdW	1.535	1.299	1.197	1.142	1.086	1.058	1.029	1.017	1.000

3.4 Numerical Study

3.4.1 *Size and power.* To study the size and power of the Durbin–Watson tests described in Sections 3.1 and 3.2, we generated $N = 1000$ independent samples $(\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{650})$ of size $n = 650$ from various bivariate spherical densities, with mean zero and identity covariance matrix (the bivariate normal and bivariate Student distributions with 1, 3 and 8 degrees of freedom). From each of these samples, we constructed a series of 650 "observations" $\mathbf{Y}_1^*, \ldots, \mathbf{Y}_{650}^*$ characterized by the linear models

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\beta}_1 I_{[501\le t\le 575]} + \boldsymbol{\beta}_2 I_{[576\le t\le 650]} + \mathbf{V}_t, \\ \mathbf{V}_t &- (m\mathbf{A})\mathbf{V}_{t-1} = \boldsymbol{\varepsilon}_t, \quad m = 0, 1, 2, \end{aligned} \tag{12}$$

with initial value $\mathbf{V}_0 = \mathbf{0}$, $\boldsymbol{\beta}_1 = \binom{1}{0}$, $\boldsymbol{\beta}_2 = \binom{0}{1}$ and

$$\mathbf{A} = \begin{pmatrix} 0.12 & 0.06 \\ -0.04 & 0.10 \end{pmatrix}.$$

Dropping observations $\mathbf{Y}_1^*$ through $\mathbf{Y}_{500}^*$ (this *warming up* period of 500 observations allows approximate stationarity to be achieved), we performed on the remaining $n = 150$ observations $(\mathbf{Y}_1, \ldots, \mathbf{Y}_{150}) := (\mathbf{Y}_{501}^*, \ldots, \mathbf{Y}_{650}^*)$ the following seven Durbin–Watson tests (at asymptotic probability level $\alpha = 5\%$):

(a) The Gaussian Durbin–Watson test based on (8).

(b1) The sign-test-type Durbin–Watson test based on (9) with Tyler signs.

(b2) The sign-test-type Durbin–Watson test based on (9) with hyperplane-based signs (interdirections).

(c1) The Spearman Durbin–Watson test based on (10) with Tyler signs and ranks.

(c2) The Spearman Durbin–Watson test based on (10) with hyperplane-based signs and ranks.

(d1) The van der Waerden Durbin–Watson test based on (11) with Tyler signs and ranks.

(d2) The van der Waerden Durbin–Watson test based on (11) with hyperplane-based signs and ranks.

In the Gaussian test (a), the least squares estimator

$$\hat{\boldsymbol{\beta}}'_{\mathcal{N}} := \left(\tfrac{1}{75} \sum_{t=1}^{75} \mathbf{Y}_t \vdots \tfrac{1}{75} \sum_{t=76}^{150} \mathbf{Y}_t \right)$$

was used for $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1 \vdots \boldsymbol{\beta}_2)$, while in the rank-based procedures, the location center of each group was estimated by the multivariate affine-equivariant median introduced by Hettmansperger and Randles (2002); the latter is root-n consistent—and, consequently, the resulting rank-based procedures are valid—*without any assumptions on the tails of the underlying densities* (so that, unlike the Gaussian test, the rank-based tests are valid under the t_1 distribution). The Tyler estimate $\widehat{\boldsymbol{\Sigma}}_{\mathrm{Tyl}}$ was computed from the algorithm of Randles (2000).

TABLE 2

Rejection frequencies (out of $N = 1000$ replications) under various values $m\mathbf{A}$, $m = 0, 1, 2$ [cf. (12)], of the autoregression matrix and various innovation densities of the Gaussian parametric Durbin–Watson test $\phi_{\mathcal{N}}$, the Tyler signed-rank van der Waerden ϕ_{vdW}, Spearman ϕ_{Sp} and sign ϕ_{S} Durbin–Watson tests, and their hyperplane-based counterparts ϕ^h_{vdW}, ϕ^h_{Sp} and ϕ^h_{S}; the series length is 150

Test	Innovation density	Autoregression matrix 0	A	2A	ARE	Innovation density	Autoregression matrix 0	A	2A	ARE
$\phi_{\mathcal{N}}$	$\mathcal{N}$	0.046	0.264	0.902	1.000	t_3	0.058	0.238	0.903	1.000
ϕ_{vdW}		0.044	0.253	0.891	1.000		0.041	0.346	0.959	1.400
ϕ_{S}		0.046	0.162	0.680	0.617		0.049	0.262	0.865	1.000
ϕ_{Sp}		0.051	0.258	0.892	0.970		0.042	0.346	0.965	1.305
ϕ^h_{vdW}		0.042	0.258	0.885	1.000		0.039	0.338	0.960	1.400
ϕ^h_{S}		0.050	0.159	0.682	0.617		0.045	0.258	0.860	1.000
ϕ^h_{Sp}		0.046	0.262	0.887	0.970		0.045	0.336	0.960	1.305
$\phi_{\mathcal{N}}$	t_8	0.044	0.264	0.904	1.000	t_1	0.046	0.264	0.902	Undefined
ϕ_{vdW}		0.041	0.265	0.907	1.047		0.044	0.253	0.891	Undefined
ϕ_{S}		0.042	0.189	0.741	0.711		0.046	0.162	0.680	Undefined
ϕ_{Sp}		0.051	0.279	0.903	1.022		0.051	0.258	0.892	Undefined
ϕ^h_{vdW}		0.041	0.266	0.905	1.047		0.042	0.257	0.885	Undefined
ϕ^h_{S}		0.043	0.188	0.740	0.711		0.050	0.159	0.682	Undefined
ϕ^h_{Sp}		0.046	0.273	0.905	1.022		0.046	0.262	0.886	Undefined

Iterations were stopped as soon as the Frobenius norm of the difference between the two members of (3) fell below 10^{-6}.

Rejection frequencies are reported in Table 2. The corresponding individual confidence intervals (for $N = 1000$ replications), at confidence level 0.95, have half-widths 0.014, 0.025 and 0.031 for frequencies of the order of 0.05 (0.95), 0.20 (0.80) and 0.50, respectively. It appears that, under the null hypothesis, none of the rejection frequencies significantly differs from the nominal 5% level. All tests thus apparently are valid and unbiased—even the Gaussian test under the Cauchy density, although in principle it is not valid. Except for the sign test, the rank-based procedures yield the same overall performance as the Gaussian test under the Gaussian density, a slight superiority under t_8 density and a more marked superiority under t_3 density. This confirms the ARE values (which we also report in the table). Somewhat disappointingly, all methods (except again for the sign tests) have more or less the same power under the Cauchy density, a fact that is not explained by any ARE value, since the latter is not defined. As a rule, the hyperplane versions of all rank-based tests do slightly better than their Tyler counterparts.

3.4.2 *Robustness.* To investigate the robustness properties of the various Durbin–Watson procedures proposed in Section 3.3, we studied their resistance to *innovation* and *observation outliers*, respectively. For simplicity, in this section, we consider only Gaussian series.

The same Monte Carlo scheme as in Section 3.4.1 was used to generate a bivariate series of length $n = 650$ from model (12), with i.i.d. Gaussian innovations $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{650}$. The resulting series $\mathbf{Y}^*_1, \ldots, \mathbf{Y}^*_{650}$ then was subjected to the following perturbations (inducing observation outliers):

- $\mathcal{Y}^+$ (*observation outliers*): Observations $\mathbf{Y}^*_t$ were replaced with $5\mathbf{Y}^*_t$ at time $t = 549, 550, 599, 600, 649, 650$.
- $\mathcal{Y}^-$ (*observation outliers*): Observations $\mathbf{Y}^*_t$ were replaced with $5\mathbf{Y}^*_t$ at time $t = 549, 599, 649$ and with $-5\mathbf{Y}^*_t$ at time $t = 550, 600, 650$.
- $\mathcal{E}^+$ (*innovation outliers*): The Gaussian innovations $\boldsymbol{\varepsilon}_t$ were replaced with $5\boldsymbol{\varepsilon}_t$ at time $t = 549, 550, 599, 600, 649$ and 650.
- $\mathcal{E}^-$ (*innovation outliers*): The Gaussian innovations $\boldsymbol{\varepsilon}_t$ were replaced with $5\boldsymbol{\varepsilon}_t$ and $-5\boldsymbol{\varepsilon}_t$ at time $t = 549, 599, 649$ and $t = 550, 600, 650$, respectively.

We generated $N = 1000$ series of each type. The last $n = 150$ observations then were subjected to the

TABLE 3

Rejection frequencies (out of $N = 1000$ replications) for various perturbed Gaussian VAR(1) *processes of the Gaussian parametric ($\phi_{\mathcal{N}}$), the Tyler signed-rank van der Waerden (ϕ_{vdW}), Spearman (ϕ_{Sp}) and sign-test-type (ϕ_{S}) Durbin–Watson tests, and their hyperplane-based counterparts ϕ^h_{vdW}, ϕ^h_{Sp} and ϕ^h_{S} at (asymptotic) probability level 5%; the series length throughout is $n = 150$*

Test	Type of outliers	Autoregression matrix 0	A	2A	Type of outliers	Autoregression matrix 0	A	2A
$\phi_{\mathcal{N}}$	$\mathcal{Y}^+$	0.625	0.738	0.902	$\mathcal{E}^+$	0.508	0.639	0.882
ϕ_{vdW}		0.073	0.348	0.881		0.057	0.314	0.892
ϕ_{S}		0.057	0.187	0.690		0.046	0.176	0.715
ϕ_{Sp}		0.059	0.313	0.898		0.056	0.305	0.892
ϕ^h_{vdW}		0.067	0.348	0.885		0.056	0.319	0.880
ϕ^h_{S}		0.056	0.180	0.692		0.046	0.173	0.714
ϕ^h_{Sp}		0.057	0.315	0.898		0.052	0.305	0.886
$\phi_{\mathcal{N}}$	$\mathcal{Y}^-$	0.704	0.670	0.713	$\mathcal{E}^-$	0.536	0.660	0.877
ϕ_{vdW}		0.085	0.150	0.614		0.072	0.330	0.908
ϕ_{S}		0.060	0.151	0.581		0.051	0.185	0.723
ϕ_{Sp}		0.076	0.165	0.710		0.069	0.320	0.906
ϕ^h_{vdW}		0.084	0.159	0.610		0.068	0.323	0.901
ϕ^h_{S}		0.061	0.149	0.584		0.048	0.185	0.721
ϕ^h_{Sp}		0.072	0.172	0.704		0.063	0.320	0.905

various Durbin–Watson procedures described in Section 3.4.1. The resulting rejection frequencies are reported in Table 3, which thus consists of four parts (one for each type of outlier), each of which is to be compared with the left upper part of Table 2. Inspection of the table reveals that, quite significantly, the type I risk of the Gaussian test is exploding (up to a 70% rejection rate under $\mathcal{Y}^-$). The Gaussian procedure thus is totally unreliable in the presence of outliers, whatever their type; the corresponding rejection frequencies under the alternative thus are meaningless. The rank-based tests also are affected, but considerably less so, with a rejection rate under the null that, in general, does not significantly differ from the nominal. As expected, the sign tests seem to be slightly more robust than the van der Waerden and Spearman tests.

4. RANK-BASED SELECTION OF THE ORDER OF A VAR PROCESS

4.1 Gaussian Parametric VAR Order Selection

Going back to the general model described in Section 2, we now turn to the problem of testing a VAR(p_0) dependence in (5) against a VAR($p_0 + 1$) dependence. For simplicity, we assume that $\boldsymbol{\beta} = \mathbf{0}$; $\boldsymbol{\Sigma}$ and f of course are nuisance parameters in this problem. A sequential application of such tests can be used to identify the actual order of the unobserved autoregressive errors; see Pötscher (1983) or Garel and Hallin (1999) for the univariate counterpart of the problem.

More formally, denote by $\boldsymbol{\Theta}_{p_0}$ the set of all values of $\boldsymbol{\theta} \in \mathbb{R}^K$ such that $\mathbf{A}_{p_0+1} = \cdots = \mathbf{A}_p = \mathbf{0}$, $|\mathbf{A}_{p_0}| \neq 0$, and for which the VAR($p_0$) model with parameters $\mathbf{A}_1, \ldots, \mathbf{A}_{p_0}$ is stationary and invertible. The null hypothesis then is of the form $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{p_0}$. Gaussian parametric optimal tests for this problem can be obtained, for example, by the Lagrange multiplier method; they require finite second-order moments.

Denote by $\hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_{p_0}$ the estimators obtained under the assumption that the VAR model in (5) is of order p_0. Write $\hat{\boldsymbol{\theta}}$ for $(\mathrm{vec}'(\hat{\mathbf{A}}_1), \ldots, \mathrm{vec}'(\hat{\mathbf{A}}_{p_0}), \mathbf{0}', \ldots, \mathbf{0}')'$. Defining the residuals

$$\mathbf{Z}_t = \mathbf{Z}_t(\hat{\boldsymbol{\theta}}) = \mathbf{Y}_t - \sum_{i=1}^{p_0} \hat{\mathbf{A}}_i \mathbf{Y}_{t-i}, \quad t = p_0 + 1, \ldots, n,$$

the residual cross-covariance matrix at lag i takes the form

$$\begin{aligned}\boldsymbol{\Gamma}_i^{(n)} &:= (n - p_0 - i)^{-1} \sum_{t=p_0+1+i}^{n} \mathbf{Z}_t \mathbf{Z}'_{t-i} \\ &= (n - p_0 - i)^{-1} \\ &\quad \cdot \sum_{t=p_0+1+i}^{n} d_{\boldsymbol{\Sigma};t} d_{\boldsymbol{\Sigma};t-i} \boldsymbol{\Sigma}^{1/2} \mathbf{U}_{\boldsymbol{\Sigma};t} \mathbf{U}'_{\boldsymbol{\Sigma};t-i} \boldsymbol{\Sigma}'^{1/2}.\end{aligned}$$

Write $\widehat{\mathbf{\Sigma}}_{\mathcal{N}}$ for $\mathbf{\Gamma}_0^{(n)}$. The Gaussian test statistic for this problem then is

$$W_{p_0}^{(n)} := n\mathbf{T}'_{p_0;\widehat{\mathbf{\Sigma}}_{\mathcal{N}}}\mathbf{Q}_{\widehat{\mathbf{\Sigma}}_{\mathcal{N}}}\mathbf{T}_{p_0;\widehat{\mathbf{\Sigma}}_{\mathcal{N}}}, \tag{13}$$

where, writing $\mathbf{G}_u = \mathbf{G}_u(\hat{\boldsymbol{\theta}})$ for the Green's matrices associated with $(\hat{\mathbf{A}}_1, \ldots, \hat{\mathbf{A}}_{p_0})$,

$$\begin{aligned} n^{1/2}&\mathbf{T}_{p_0;\widehat{\mathbf{\Sigma}}_{\mathcal{N}}} \\ &:= \begin{pmatrix} (n-1)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_1^{(n)}) \\ \sum_{u=2}^{n-p_0-1}(n-u)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_u^{(n)}\mathbf{G}'_{u-1}) \\ \sum_{u=2}^{n-p_0-1}(n-u)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_u^{(n)}\mathbf{G}'_{u-2}) \\ \vdots \\ \sum_{u=i}^{n-p_0-1}(n-u)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_u^{(n)}\mathbf{G}'_{u-i}) \\ \vdots \\ \sum_{u=p_0}^{n-p_0-1}(n-u)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_u^{(n)}\mathbf{G}'_{u-p_0}) \end{pmatrix} \end{aligned} \tag{14}$$

and (for $p_0 = 1$, $\mathbf{0}_{k^2\times k^2(p_0-1)}$ is void)

$$\begin{aligned} \mathbf{Q}_{\widehat{\mathbf{\Sigma}}_{\mathcal{N}}} &:= \begin{pmatrix} \widehat{\mathbf{\Sigma}}_{\mathcal{N}}\otimes\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1} & \mathbf{0}_{k^2\times k^2p_0} \\ \mathbf{0}_{k^2p_0\times k^2} & \mathbf{w}^2 \end{pmatrix}^{-1} \\ &\quad - \begin{pmatrix} \mathbf{I}_{k^2} & \mathbf{0}_{k^2\times k^2(p_0-1)} \\ & \mathbf{I}_{k^2p_0} \end{pmatrix}(\mathbf{W}^2)^{-1} \\ &\quad \cdot \begin{pmatrix} \mathbf{I}_{k^2} & \mathbf{0}_{k^2\times k^2(p_0-1)} \\ & \mathbf{I}_{k^2p_0} \end{pmatrix}', \end{aligned}$$

with the $(k^2p_0 \times k^2p_0)$ matrices $\mathbf{w}^2$ and $\mathbf{W}^2$ having (i, j) blocks [of dimension $(k^2 \times k^2)$]

$$(\mathbf{w}^2)_{ij} := \sum_{u=\max(2,i,j)}^{n-p_0-1}(\mathbf{G}_{u-i}\widehat{\mathbf{\Sigma}}_{\mathcal{N}}\mathbf{G}'_{u-j})\otimes\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1} \quad \text{and}$$

$$(\mathbf{W}^2)_{ij} := \sum_{u=\max(i,j)}^{n-p_0-1}(\mathbf{G}_{u-i}\widehat{\mathbf{\Sigma}}_{\mathcal{N}}\mathbf{G}'_{u-j})\otimes\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}, \qquad i, j = 1, \ldots, p_0,$$

respectively. Note that $\mathbf{w}^2$ and $\mathbf{W}^2$ differ only by their upper left $(k^2 \times k^2)$ block. The structure of this test statistic is the same as that of the univariate Gaussian Lagrange multiplier test statistic described by Garel and Hallin (1999).

The null hypothesis of AR(p_0) dependence is rejected whenever $W_{p_0}^{(n)}$ exceeds the $(1-\alpha)$ quantile of a chi-squared distribution with k^2 degrees of freedom. The intuition behind the test statistic (13) is a little bit less straightforward than in the Durbin–Watson case. Actually, $W_{p_0}^{(n)}$ is a quadratic form that involves all estimated residual cross-correlation matrices, with weights that neutralize the effect of parameter estimation on the residuals and optimize the power. For instance, $p_0 = 1$ yields (writing $\hat{\mathbf{A}}$ instead of $\hat{\mathbf{A}}_1$, we have $\mathbf{G}_u = \hat{\mathbf{A}}^u$)

$$n^{1/2}\mathbf{T}_{1;\widehat{\mathbf{\Sigma}}_{\mathcal{N}}} := \begin{pmatrix} (n-1)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_1^{(n)}) \\ \sum_{u=2}^{n-1}(n-u)^{1/2}\operatorname{vec}(\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\mathbf{\Gamma}_u^{(n)}(\hat{\mathbf{A}}^{u-1})') \end{pmatrix}$$

and

$$\begin{aligned} \mathbf{Q}_{\widehat{\mathbf{\Sigma}}_{\mathcal{N}}} &:= \begin{pmatrix} \widehat{\mathbf{\Sigma}}_{\mathcal{N}}\otimes\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1} & \mathbf{0} \\ \mathbf{0} & \sum_{u=2}^{n-1}(\mathbf{A}^{u-1}\widehat{\mathbf{\Sigma}}_{\mathcal{N}}(\mathbf{A}^{u-1})')\otimes\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1} \end{pmatrix}^{-1} \\ &\quad - \begin{pmatrix} \mathbf{I}_{k^2} \\ \mathbf{I}_{k^2} \end{pmatrix}\left(\sum_{u=1}^{n-1}(\mathbf{A}^{u-1}\widehat{\mathbf{\Sigma}}_{\mathcal{N}}(\mathbf{A}^{u-1})')\otimes\widehat{\mathbf{\Sigma}}_{\mathcal{N}}^{-1}\right)^{-1} \\ &\quad \cdot (\mathbf{I}_{k^2}\ \mathbf{I}_{k^2}). \end{aligned}$$

The order selection procedure then consists of first running a Durbin–Watson test (reducing to a simple test for randomness when $\boldsymbol{\beta} = \mathbf{0}$). In case this is inconclusive, a VAR of order zero (i.e., white noise) is selected and a traditional regression model is considered for the analysis. If Durbin–Watson is significant, then turn to testing VAR(1) against VAR(2) (i.e., the particular case just discussed) and so on. This procedure as a whole is of a heuristic nature and no precise risk can be evaluated for the final output. However, consistency results have been obtained, possibly with α values varying from step to step; see Pötscher (1983, 1985).

4.2 Signed-Rank VAR Order Selection

The procedure runs exactly as in the Gaussian parametric case, but is based on multivariate signed-rank statistics. Here again, we propose three particular test statistics. Each test can be computed from Tyler signs and ranks or from hyperplane-based signs and ranks. In case interdirections are used, they should be "absolute." The three statistics are the following:

1. A test statistic of the sign-test type,

$$\underset{\sim}{W}^{(n)}_{p_0;\mathrm{sign}} := k^2 n \underset{\sim}{\mathbf{T}}'_{p_0;\widehat{\mathbf{\Sigma}};\mathrm{sign}} \mathbf{Q}_{\widehat{\mathbf{\Sigma}}} \underset{\sim}{\mathbf{T}}_{p_0;\widehat{\mathbf{\Sigma}};\mathrm{sign}},$$

with $n^{1/2}\underset{\sim}{\mathbf{T}}_{p_0;\mathrm{sign}}$ as in (14), but with the "*sign-test*" *type cross-covariance matrices*

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i;\widehat{\mathbf{\Sigma}};\mathrm{sign}} := \widehat{\mathbf{\Sigma}}^{1/2}\left(\frac{1}{n-p_0-i}\sum_{t=p_0+i+1}^{n}\mathbf{U}_t\mathbf{U}'_{t-i}\right)\widehat{\mathbf{\Sigma}}^{\prime 1/2}$$

substituted for $\mathbf{\Gamma}^{(n)}_i$.

2. A test statistic of the Spearman type,

$$\underset{\sim}{W}^{(n)}_{p_0;\mathrm{Sp}} := 9k^2 n \underset{\sim}{\mathbf{T}}'_{p_0;\widehat{\mathbf{\Sigma}};\mathrm{Sp}} \mathbf{Q}_{\widehat{\mathbf{\Sigma}}} \underset{\sim}{\mathbf{T}}_{p_0;\widehat{\mathbf{\Sigma}};\mathrm{Sp}},$$

with $n^{1/2}\underset{\sim}{\mathbf{T}}_{p_0;\mathrm{Sp}}$ as in (14), but with the *Spearman cross-covariance matrices*

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i;\widehat{\mathbf{\Sigma}};\mathrm{Sp}} := \widehat{\mathbf{\Sigma}}^{1/2}\left(\frac{1}{(n-p_0-i)(n-p_0+1)^2}\cdot\sum_{t=p_0+i+1}^{n}R_tR_{t-i}\mathbf{U}_t\mathbf{U}'_{t-i}\right)\widehat{\mathbf{\Sigma}}^{\prime 1/2}$$

substituted for $\mathbf{\Gamma}^{(n)}_i$.

3. A test statistic of the van der Waerden type,

$$\underset{\sim}{W}^{(n)}_{p_0;\mathrm{vdW}} := n \underset{\sim}{\mathbf{T}}'_{p_0;\widehat{\mathbf{\Sigma}};\mathrm{vdW}} \mathbf{Q}_{\widehat{\mathbf{\Sigma}}} \underset{\sim}{\mathbf{T}}_{p_0;\widehat{\mathbf{\Sigma}};\mathrm{vdW}},$$

with $n^{1/2}\underset{\sim}{\mathbf{T}}_{p_0;\mathrm{vdW}}$ as in (14), but with the *van der Waerden cross-covariance matrices*

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i;\widehat{\mathbf{\Sigma}};\mathrm{vdW}} := \widehat{\mathbf{\Sigma}}^{1/2}\left(\frac{1}{n-p_0-i}\cdot\sum_{t=p_0+i+1}^{n}\Phi_k^{-1}\left(\frac{R_t}{n-p_0+1}\right)\cdot\Phi_k^{-1}\left(\frac{R_{t-i}}{n-p_0+1}\right)\cdot\mathbf{U}_t\mathbf{U}'_{t-i}\right)\widehat{\mathbf{\Sigma}}^{\prime 1/2}$$

[Φ_k is as in (11)] substituted for $\mathbf{\Gamma}^{(n)}_i$.

The null hypothesis of AR(p_0) dependence is rejected whenever the test statistic exceeds the $(1-\alpha)$ quantile of a chi-squared distribution with k^2 degrees of freedom.

We insist upon the fact that $\widehat{\mathbf{\Sigma}}$, contrary to the estimate $\widehat{\mathbf{\Sigma}}_{\mathcal{N}}$ appearing in the Gaussian statistic, need no longer be the empirical marginal covariance matrix.

4.3 Asymptotic Relative Efficiencies

The asymptotic relative efficiencies, with respect to their Gaussian counterparts, of the rank-based tests used at each step of the order selection procedure are the same as in the Durbin–Watson case. The figures in Table 1 as well as the generalized Chernoff–Savage result of Hallin and Paindaveine (2005) thus still apply here. However, a more pertinent assessment of the respective relative efficiencies of order selection procedures considered as a whole would be provided by ratios of correct identification probabilities. Deriving exact values for such ratios is probably infeasible. Monte Carlo evaluations, however, are possible; some numerical values are given in the simulation study below.

4.4 Numerical Study

4.4.1 *Efficiency.* Here we generated $N=1000$ independent samples $(\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_{620})$ of size $n=620$ from various bivariate spherical densities, with mean zero and identity covariance matrix: the bivariate normal and bivariate Student distributions with 1 (in this case, the *shape*, not the covariance matrix, is identity), 3 and 8 degrees of freedom. These samples were used in the VAR(1) model

$$\mathbf{Y}_t - \mathbf{A}\mathbf{Y}_{t-1} = \boldsymbol{\varepsilon}_t \quad \text{with } \mathbf{A} = \begin{pmatrix} 0.30 & 0.12 \\ -0.06 & 0.24 \end{pmatrix} \tag{15}$$

and initial value $\mathbf{Y}_0 = \mathbf{0}$, yielding VAR(1) series $(\mathbf{Y}^*_1,\ldots,\mathbf{Y}^*_{620})$ of length 620. Of these observations, the last 120, denoted as $(\mathbf{Y}_1,\ldots,\mathbf{Y}_{120})$, were subjected to various sequential order-identification procedures.

Seven versions (Gaussian or rank based) of the order-identification procedure were performed on each series. Step one of each procedure consisted of testing for white noise against VAR(1) dependence using a (degenerate—since no trend has to be estimated) Durbin–Watson test which coincides with the tests for randomness developed by Hallin and Paindaveine (2002b). If the hypothesis of randomness cannot be rejected, the model is identified as being VAR(0), that is, white noise (order underidentification). If randomness is rejected, the tests developed in Sections 4.1 and 4.2 are performed for testing VAR(1) against VAR(2) dependence. If VAR(1) is not rejected, the order $(p=1)$

is correctly identified; if not, the procedure is pursued further, but we simply record overidentification of the order. Of course, it is pretty natural to use the same type of test throughout the procedure. The following seven types of identification procedures were considered:

(a) The parametric Gaussian procedure.
(b1) The sign-test-type procedure based on Tyler's signs and ranks.
(b2) The hyperplane-based sign-test-type procedure.
(c1) The Spearman-type procedure based on Tyler's signs and ranks.
(c2) The hyperplane-based Spearman-type procedure.
(d1) The van der Waerden-type procedure based on Tyler's signs and ranks.
(d2) The hyperplane-based van der Waerden-type procedure.

All *individual* tests were performed at nominal (asymptotic) level $\alpha = 5\%$. In each case, the Yule–Walker estimator

$$\hat{\mathbf{A}} := \left(\frac{1}{119}\sum_{t=2}^{120}\mathbf{Y}_t\mathbf{Y}_{t-1}'\right)\left(\frac{1}{119}\sum_{t=2}^{120}\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}'\right)^{-1}$$

was used to estimate $\mathbf{A}$. The Tyler estimate $\widehat{\mathbf{\Sigma}}_{\mathrm{Tyl}}$ was computed from the algorithm of Randles (2000) [again, iterations were stopped as soon as the Frobenius norm of the difference between the two members of (3) fell below 10^{-6}].

Under-, correct and overidentification frequencies are reported in Table 4, along with the corresponding ARE figures. The corresponding individual confidence intervals (for $N = 1000$ replications), at confidence level 0.95, have half-widths 0.014, 0.025 and 0.031 for frequencies on the order of 0.05 (0.95), 0.20 (0.80) and 0.50, respectively. Inspection of the table reveals the excellent overall performance of all rank-based procedures considered:

- Hyperplane-based van der Waerden procedures uniformly outperform the Tyler-type van der Waerden ones, which in turn perform at least as well as their parametric Gaussian counterpart, even under Gaussian innovations.
- More generally, hyperplane-based procedures (van der Waerden, signs, Spearman) do uniformly better than their Tyler-type competitors.
- Although the validity of the tests used at each step of the identification procedure is not formally estab-

TABLE 4

Underidentification ($p = 0$), *correct identification* ($p = 1$) *and overidentification* ($p \geq 2$) *frequencies* (*out of* $N = 1000$ *replications*) *for the* VAR(1) *model* (15) *under various Gaussian and Student innovation densities. The seven procedures considered are based on the Gaussian parametric tests* $\phi_{\mathcal{N}}$, *the Tyler signed-rank van der Waerden and Spearman tests* ϕ_{vdW} *and* ϕ_{Sp}, *the Tyler sign test* ϕ_{S}, *and their hyperplane-based counterparts* ϕ^h_{vdW}, ϕ^h_{Sp} *and* ϕ^h_{S}

		Order identification					Order identification			
Test	**Innovation density**	**0**	**1**	**≥ 2**	**ARE**	**Innovation density**	**0**	**1**	**≥ 2**	**ARE**
$\phi_{\mathcal{N}}$	$\mathcal{N}$	42	898	60	1.000	t_3	41	915	44	1.000
ϕ_{vdW}		54	898	48	1.000		35	914	51	1.400
ϕ_{S}		186	764	50	0.617		139	809	52	1.000
ϕ_{Sp}		55	891	54	0.970		37	903	60	1.305
ϕ^h_{vdW}		55	906	39	1.000		40	925	35	1.400
ϕ^h_{S}		186	771	43	0.617		139	812	49	1.000
ϕ^h_{Sp}		51	910	39	0.970		42	915	43	1.305
$\phi_{\mathcal{N}}$	t_8	37	903	60	1.000	t_1	31	919	50	Undefined
ϕ_{vdW}		50	898	52	1.047		9	930	61	Undefined
ϕ_{S}		161	791	48	0.711		84	864	52	Undefined
ϕ_{Sp}		48	903	49	1.022		12	925	63	Undefined
ϕ^h_{vdW}		48	911	41	1.047		10	956	34	Undefined
ϕ^h_{S}		159	794	47	0.711		88	873	39	Undefined
ϕ^h_{Sp}		45	904	51	1.022		12	952	36	Undefined

NOTE. All tests were performed at probability level 5%; the series length throughout is $n = 120$ (AREs refer to individual tests, not to the order identification procedure as a whole).

TABLE 5
Underidentification ($p = 0$), *correct identification* ($p = 1$) *and overidentification* ($p \geq 2$) *frequencies* (*out of* $N = 1000$ *replications*) *in various perturbed Gaussian* VAR(1) *series*

Test	Type of outliers	Order identification			Type of outliers	Order identification		
		0	1	≥ 2		0	1	≥ 2
$\phi_{\mathcal{N}}$	$\mathcal{Y}^+$	428	293	279	$\mathcal{E}^+$	88	522	390
ϕ_{vdW}		98	822	80		27	909	64
ϕ_{S}		189	762	49		140	819	41
ϕ_{Sp}		81	851	68		26	916	58
ϕ^h_{vdW}		95	831	74		29	926	45
ϕ^h_{S}		197	757	46		141	824	35
ϕ^h_{Sp}		83	857	60		27	933	40
$\phi_{\mathcal{N}}$	$\mathcal{Y}^-$	672	180	148	$\mathcal{E}^-$	77	520	403
ϕ_{vdW}		217	708	75		25	911	64
ϕ_{S}		290	662	48		133	817	50
ϕ_{Sp}		179	749	72		25	916	59
ϕ^h_{vdW}		222	715	63		22	917	61
ϕ^h_{S}		294	663	43		134	827	39
ϕ^h_{Sp}		185	754	61		26	919	55

NOTE. The various order identification procedures are based on the Gaussian parametric tests $\phi_{\mathcal{N}}$, the Tyler signed-rank van der Waerden (ϕ_{vdW}), Spearman (ϕ_{Sp}) and sign-test-type (ϕ_{S}) tests, and their hyperplane-based counterparts ϕ^h_{vdW}, ϕ^h_{Sp} and ϕ^h_{S} at (asymptotic) probability level 5%; the series length throughout is $n = 120$.

lished under multivariate Cauchy (t_1) innovations, the final result under such densities remains excellent, with a remarkable 95% frequency of correct identification for the hyperplane-based van der Waerden and Spearman versions.

4.4.2 *Robustness.* A robustness investigation also was conducted on the model in Section 3.4.2 for the various order identification procedures proposed in Sections 4.1 and 4.2. Observations, $(\mathbf{Y}^*_1, \ldots, \mathbf{Y}^*_{620})$ were generated in the same way as in the previous section from model (15) with Gaussian $\boldsymbol{\varepsilon}_t$'s. These observations then were perturbed, as in Section 3.4.2, to produce observation outliers and innovation outliers respectively:

- $\mathcal{Y}^+$ (*observation outliers*): Observations $\mathbf{Y}^*_t$ were replaced with $5\mathbf{Y}^*_t$ for $t = 538$, 540, 578, 580, 618 and 620.
- $\mathcal{Y}^-$ (*observation outliers*): Observations $\mathbf{Y}^*_t$ were replaced with $5\mathbf{Y}^*_t$ for $t = 538$, 578 and 618, and with $-5\mathbf{Y}^*_t$ for $t = 540$, 580 and 620.
- $\mathcal{E}^+$ (*innovation outliers*): Gaussian innovations $\boldsymbol{\varepsilon}_t$ were replaced with $5\boldsymbol{\varepsilon}_t$ for $t = 538$, 578, 580, 618 and 620.
- $\mathcal{E}^-$ (*innovation outliers*): Gaussian innovations $\boldsymbol{\varepsilon}_t$ were replaced with $5\boldsymbol{\varepsilon}_t$ for $t = 538$, 578 and 618, and with $-5\boldsymbol{\varepsilon}_t$ for $t = 540$, 580 and 620.

The last $n = 120$ observations then were subjected to the seven order identification procedures described in Section 4.4.1. The resulting under-, correct and overidentification frequencies are reported in Table 5. This simulation exercise of course is somewhat limited and allows only for very general conclusions. The frequencies reported in Table 5, however, very clearly show how fragile the traditional parametric method can be in the presence of a small number of outliers: the observed proportion of correct identification (based on the parametric tests) drops from 0.898 in the unperturbed case to 0.180 under the observation outlier scheme $\mathcal{Y}^-$. Quite on the contrary, the rank-based methods apparently resist quite well, irrespective of the type of outlier.

5. CONCLUSIONS

Rank-based methods have been confined for a long time to problems that involve univariate independent observations. We show, on the basis of two particular examples (the Durbin–Watson and the autoregressive order selection problems), that rank methods also apply

to serial (i.e., time-series) multivariate problems. Two concepts of signs and ranks are considered: pseudo-Mahalanobis or Tyler and the hyperplane-based or Oja–Paindaveine. Theoretical results establish that these methods are as efficient, locally and asymptotically, as their everyday-practice parametric competitors based on cross-correlation matrices; the van der Waerden versions even uniformly dominate the competition. Simulations moreover show that the rank-based procedures successfully resist the presence of observation as well as innovation outliers, whereas traditional parametric methods literally collapse under such perturbations.

APPENDIX: RANKS, SIGNS AND SEMIPARAMETRIC EFFICIENCY

For the reader who is familiar with local asymptotic normality or tangent spaces, we conclude this paper with a brief theoretical justification for considering rank-based methods in the analysis of a broad class of semiparametric models. Details can be found in Hallin and Werker (2003).

Rank-based methods apply whenever the data are generated, through some model involving a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^K$, by some unobserved white noise (here k-dimensional) with unspecified density $\underline{f}$ belonging to some class $\mathcal{F}$ of densities. The statistical models we are considering are thus, typically, *semiparametric*, of the form

$$(16)\quad \left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\underline{f}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \underline{f} \in \mathcal{F}\right\}\right).$$

Assume that $\boldsymbol{\theta}$ is the parameter of interest, whereas $\underline{f}$ plays the role of a nuisance parameter. Whenever the fixed-$\underline{f}$ parametric submodels of (16) are locally asymptotically normal with *central sequence* $\boldsymbol{\Delta}^{(n)}_{\underline{f}}(\boldsymbol{\theta})$ and provided that some other regularity assumptions are met, the theory of semiparametric efficiency (see Bickel, Klaassen, Ritov and Wellner, 1993) stipulates that semiparametrically efficient (at $\boldsymbol{\theta}$ and $\underline{f}$) inference can be based on the projection $\boldsymbol{\Delta}^{(n)*}_{\underline{f}}(\boldsymbol{\theta})$ of $\boldsymbol{\Delta}^{(n)}_{\underline{f}}(\boldsymbol{\theta})$ along the so-called *tangent spaces*.

Another way to reach semiparametric efficiency (still at $\boldsymbol{\theta}$ and $\underline{f}$) is possible when the fixed-$\boldsymbol{\theta}$ submodels of (16) are generated by some group of transformations $\mathcal{G}^{(n)}_{\boldsymbol{\theta}}$ acting over $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)})$, with maximal invariant $\mathbf{R}^{(n)}(\boldsymbol{\theta})$. Hallin and Werker (2003) showed that, under quite general conditions, the difference between $\boldsymbol{\Delta}^{(n)*}_{\underline{f}}(\boldsymbol{\theta})$ and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\underline{f}}(\boldsymbol{\theta}) := \mathrm{E}[\boldsymbol{\Delta}^{(n)}_{\underline{f}}(\boldsymbol{\theta})|\mathbf{R}^{(n)}(\boldsymbol{\theta})]$ tends to zero as $n \to \infty$, in probability, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\underline{f}}$. Conditioning on the maximal invariant thus does the same job as projecting along tangent spaces. Now, in most models that involve unobserved white noise with unspecified density $\underline{f}$, residual *ranks* and/or *signs* (their definitions depend on the class of densities $\mathcal{F}$) provide a maximal invariant $\mathbf{R}^{(n)}(\boldsymbol{\theta})$.

Rank-based methods thus, in a sense, allow for bypassing tangent space calculations in the construction of semiparametrically efficient inference procedures. Besides these semiparametric efficiency features, of course, they also enjoy their usual properties of distribution-freeness (a consequence of invariance), robustness and so forth.

ACKNOWLEDGMENT

This research was supported by a P.A.I. contract with the Belgian Federal Government and an Action de Recherche Concertée of the Communauté Française de Belgique.

REFERENCES

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.

DAVIS, R. A. and WU, W. (1997). M-estimation for linear regression with infinite variance. *Probab. Math. Statist.* **17** 1–20.

GAREL, B. and HALLIN, M. (1999). Rank-based autoregressive order identification. *J. Amer. Statist. Assoc.* **94** 1357–1371.

HALLIN, M., INGENBLEEK, J.-F. and PURI, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.* **13** 1156–1181.

HALLIN, M. and PAINDAVEINE, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133.

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

HALLIN, M. and PAINDAVEINE, D. (2002c). Multivariate signed ranks: Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (Y. Dodge, ed.) 271–282. Birkhäuser, Basel.

HALLIN, M. and PAINDAVEINE, D. (2003). Affine invariant linear hypotheses for the multivariate general linear model with VARMA error terms. In *Mathematical Statistics and Applications*: *Festschrift for Constance van Eeden* (M. Moore, S. Froda and C. Léger, eds.) 417–434. IMS, Beachwood, OH.

HALLIN, M. and PAINDAVEINE, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32** 2642–2678.

HALLIN, M. and PAINDAVEINE, D. (2004b). Asymptotic linearity of serial and nonserial multivariate signed rank statistics. *J. Statist. Plann. Inference*. To appear.

HALLIN, M. and PAINDAVEINE, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *J. Multivariate Anal.* **93** 122–163.

HALLIN, M. and PURI, M. L. (1988). Optimal rank-based procedures for time-series analysis: Testing an ARMA model against other ARMA models. *Ann. Statist.* **16** 402–432.

HALLIN, M. and PURI, M. L. (1991). Time-series analysis via rank-order theory: Signed-rank tests for ARMA models. *J. Multivariate Anal.* **39** 1–29.

HALLIN, M. and PURI, M. L. (1992). Rank tests for time-series analysis: A survey. In *New Directions in Time Series Analysis I*. (D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu, eds.) 111–154. Springer, New York.

HALLIN, M. and PURI, M. L. (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.* **50** 175–237.

HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165.

HETTMANSPERGER, T. P. and RANDLES, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860.

OJA, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Statist.* **26** 319–343.

OJA, H. and PAINDAVEINE, D. (2004). Optimal signed-rank tests based on hyperplanes. *J. Statist. Plann. Inference*. To appear.

PÖTSCHER, B. M. (1983). Order estimation in ARMA models by Lagrangian multiplier tests. *Ann. Statist.* **11** 872–885.

PÖTSCHER, B. M. (1985). The behaviour of the Lagrange multiplier test in testing the orders of an ARMA model. *Metrika* **32** 129–150.

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

RANDLES, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050.

RANDLES, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.* **95** 1263–1268.

TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251.