T E C H N I C A L
R E P O R T

0330

MULTIVARIATE SIGNED RANK TESTS IN VECTOR AUTOREGRESSIVE ORDER IDENTIFICATION

HALLIN, M. and D. PAINDAVEINE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Multivariate Signed Rank Tests in Vector Autoregressive Order Identification

Marc Hallin* and Davy Paindaveine*

I.S.R.O., E.C.A.R.E.S., and Département de Mathématique
Université Libre de Bruxelles, Brussels, Belgium

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 underlying the concepts of ranks to be considered, we concentrate on two examples of practical relevance: (i) the multivariate Durbin-Watson problem (testing against autocorrelated noise in a linear model context), and (ii) 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: (a) statistics of the sign test type, (b) Spearman statistics, and (c) van der Waerden (normal score) ones. Simulations confirm theoretical results about the power of the proposed rank-based methods, and establish their good robustness properties.

AMS 1980 subject classification : 62G10, 62M10

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 involving univariate independent observations. Except for 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 (Hájek and Šidák, 1967 and Hájek, Šidák, and Sen, 1999; Lehmann, 1975; Randles and Wolfe, 1979; Pratt and Gibbons, 1981; Hettmansperger, 1984; Puri and Sen, 1985—to quote only a few) mainly deal with single-response linear models with independent errors: one- and two-sample location, analysis of variance, regression, etc.

The need for non Gaussian, distribution-free and robust methods is certainly no less acute in problems involving multivariate and/or serially dependent (time series) data. Rank-based methods for multivariate observations have 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

*Research supported by a P.A.I. contract of the Belgian federal Government and an Action de Recherche Concertée of the Communauté française de Belgique.

account of this line of research is given in the monograph by Puri and Sen (1971). Componentwise ranks however are not affine-invariant, hence 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 for considering permutations of componentwise rank vectors rather than permutations of the observations themselves. The resulting theory therefore is not entirely satisfactory.

The interest for 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 are considering below 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, turning point tests, etc.) 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 were 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 reducing 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 et al. 1985, Hallin and Puri 1988, 1991, 1994); see Hallin and Puri (1992) for a review of rank-based testing in a (univariate) ARMA context.

The purpose of this paper is a combination of these two extensions of the classical theory: time-series in a multivariate setting. Rather than giving a general exposition (for which we refer to Hallin and Paindaveine 2003d), we concentrate on two important particular problems: a multivariate version of the classical Durbin-Watson test, and the tests allowing for autoregressive order identification, namely, the problem of testing VAR(p_0) against VAR($p_0 + 1$) dependence (reducing to the Durbin-Watson problem for $p_0 = 0$). In both cases, we limit ourselves to statistics of the following three types: sign test statistics, Spearman statistics, and van der Waerden (normal score) ones.

1.2 Ranks, signs, and semiparametric efficiency

The reader who is not familiar with local asymptotic normality (LAN) or tangent spaces safely can skip this section, where we very briefly provide the 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, a k-dimensional one) with unspecified density $\underline{f}$ belonging to some class $\mathcal{F}$ of densities. The statistical models we are considering are thus, typically, *semiparametric* ones, of the form

$$\left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\underline{f}}, \, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \, \underline{f} \in \mathcal{F}\right\}\right). \tag{1}$$

Assume that $\boldsymbol{\theta}$ is the parameter of interest, whereas $\underline{f}$ plays the role of a nuisance. Whenever the fixed-$\underline{f}$ parametric submodels of (1) are locally asymptotically normal (LAN) 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 of reaching semiparametric efficiency (still, at $\boldsymbol{\theta}$ and $\underline{f}$) is possible when the fixed-$\boldsymbol{\theta}$ submodels of (1) are generated by some group of transformation $\mathcal{G}^{(n)}_{\boldsymbol{\theta}}$ acting over $\left(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}\right)$, with maximal invariant $\mathbf{R}^{(n)}(\boldsymbol{\theta})$. Hallin and Werker (2003) have shown that, under quite general conditions, the difference between $\boldsymbol{\Delta}^{(n)*}_{\underline{f}}(\boldsymbol{\theta})$ and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\underline{f}}(\boldsymbol{\theta}) := \mathrm{E}\left[\boldsymbol{\Delta}^{(n)}_{\underline{f}}(\boldsymbol{\theta}) \,\middle|\, \mathbf{R}^{(n)}(\boldsymbol{\theta})\right]$ 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 involving 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, etc.

1.3 From classical univariate signed ranks to multivariate signs and ranks

Denote by $Z^{(n)}_1, \ldots, Z^{(n)}_n$ a n-tuple of univariate i.i.d. random variables with common density f satisfying the symmetry assumption $f(-z) = f(z)$, $z \in \mathbb{Z}$. The sign of $Z^{(n)}_t$ will be denoted by $s^{(n)}_t$, and the rank of $|Z^{(n)}_t|$ among $|Z^{(n)}_1|, \ldots, |Z^{(n)}_N|$ by $R^{(n)}_{+;t}$. The products $(s^{(n)}_1 R^{(n)}_{+;1}, \ldots, s^{(n)}_n R^{(n)}_{+;n})$ are called the *signed ranks*, and constitute (up to a factor ± 1) a maximal invariant for the group of antisymmetric, continuous order-preserving transformations acting on $Z^{(n)}_1, \ldots, Z^{(n)}_n$. When no confusion is possible, superscripts $^{(n)}$ are omitted.

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

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

where $c_{k,f}$ is a normalization constant, and $\|\boldsymbol{\Sigma}^{-1/2}\mathbf{z}\| := (\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z})^{1/2}$ denotes the norm of $\mathbf{z}$ in the metric associated with $\boldsymbol{\Sigma}$ (we write $\boldsymbol{\Sigma}^{-1/2}$ for the unique upper-triangular $k \times k$ array with positive diagonal elements satisfying $\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\Sigma}^{-1/2})'\boldsymbol{\Sigma}^{-1/2}$). The contours of $\underline{f}_{\boldsymbol{\Sigma},f}$ clearly are a family of ellipsoids centered at the origin, the shape of which is characterized by the matrix $\boldsymbol{\Sigma}$; the nonnegative function f will be called a *radial density*, though it does not integrate to one. Note that $\boldsymbol{\Sigma}$ needs 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. In practice, of course, both $\boldsymbol{\Sigma}$ 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 $\mathcal{G}_{\boldsymbol{\Sigma}}$ of continuous order-preserving radial transformations and the group $\mathcal{G}$ of affine transformations acting on $\mathbb{R}^k$.

Let $d_t = d^{(n)}_{\boldsymbol{\Sigma};t} := \|\boldsymbol{\Sigma}^{-1/2}\mathbf{Z}^{(n)}_t\|$. Then, $\mathbf{U}^{(n)}_{\boldsymbol{\Sigma};t} := \boldsymbol{\Sigma}^{-1/2}\mathbf{Z}^{(n)}_t / d^{(n)}_{\boldsymbol{\Sigma};t}$ is the unit vector pointing in the direction of the *sphericized vector* $\boldsymbol{\Sigma}^{-1/2}\mathbf{Z}^{(n)}_t$. Clearly, if $\mathbf{Z}^{(n)}_t$ has density (2), then $\mathbf{U}^{(n)}_{\boldsymbol{\Sigma};t}$ is uniform over the unit sphere $\mathcal{S}^{k-1}$ in $\mathbb{R}^k$, just as $s^{(n)}_t$, in the univariate setting, is uniform over $\mathcal{S}^0 = \{-1, 1\}$, the unit sphere in $\mathbb{R}$. For each $\boldsymbol{\Sigma}$, define the *group of continuous order-preserving* *radial transformations* $\mathcal{G}_{\mathbf{\Sigma}}^{(n)} = \{\mathcal{g}_g^{(n)}\}$, with

$$\mathcal{g}_g^{(n)}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) := (g(d_{\mathbf{\Sigma};1}^{(n)})\,\mathbf{\Sigma}^{1/2}\mathbf{U}_{\mathbf{\Sigma};1}^{(n)}, \ldots, g(d_{\mathbf{\Sigma};n}^{(n)})\,\mathbf{\Sigma}^{1/2}\mathbf{U}_{\mathbf{\Sigma};n}^{(n)}),$$

where $g\colon \mathbb{R}^+ \to \mathbb{R}^+$ is a continuous, strictly increasing function such that $g(0) = 0$ and $\lim_{r\to\infty} g(r) = \infty$. The transformation $\mathcal{g}_g^{(n)}$ is *radial*, in the sense that, under the action of $\mathcal{g}_g^{(n)}$, the residuals $\mathbf{Z}_t = d_{\mathbf{\Sigma};t}\mathbf{\Sigma}^{1/2}\mathbf{U}_{\mathbf{\Sigma};t}$ are moving along a half line running through the origin in $\mathbb{R}^k$. This group is a generating group for the fixed-$(\boldsymbol{\theta}, \mathbf{\Sigma})$ submodel, and 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 $\mathcal{G}$ of affine transformations of $\mathbb{R}^k$ generates the fixed-$(\boldsymbol{\theta}, f)$ submodel.

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. But, when $\mathbf{\Sigma}$ is unspecified, they cannot be computed from the $\mathbf{Z}_t^{(n)}$'s.

1.3.1 Pseudo-Mahalanobis signs and ranks: the Tyler signs and ranks

Of course, $\mathbf{\Sigma}$ in practice is not known. A natural idea therefore consists in 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 root-n consistent, affine-equivariant estimator of $\mathbf{\Sigma}$. A possible choice for $\widehat{\mathbf{\Sigma}}$ would be the empirical covariance matrix of the $\mathbf{Z}_t^{(n)}$'s. This estimate however is known to be highly non robust, and its consistency requires finite moments of order two. We therefore rather suggest using Tyler (1987)'s estimator of shape, which is defined as $\widehat{\mathbf{\Sigma}} := \mathbf{C}_{\text{Tyl}}^{-1}\mathbf{C}_{\text{Tyl}}^{\prime -1}$, where $\mathbf{C}_{\text{Tyl}}$ is the unique upper triangular $k \times k$ matrix with nonnegative diagonal and upper left element 1 such that

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

($\mathbf{I}_k$ stands for the $k \times k$ identity matrix). This estimate thus is such that the empirical covariance of the corresponding $\mathbf{U}_{\widehat{\mathbf{\Sigma}};t}^{(n)}$'s is proportional to the identity matrix. 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 $\widehat{\mathbf{\Sigma}}$ converges in probability to $a\mathbf{\Sigma}$, where a is some positive constant. The *Tyler signs* $\mathbf{U}_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ are strictly equivariant under both $\mathcal{G}_{\mathbf{\Sigma}}$ and $\mathcal{G}$, but the *Tyler ranks* $R_{\widehat{\mathbf{\Sigma}};t}^{(n)}$ are invariant under $\mathcal{G}$ 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 $o_{\mathrm{P}}(1)$ 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*, in the sense of being asymptotically equivalent to the strictly invariant genuine ranks. When the choice of $\widehat{\mathbf{\Sigma}}$ is not imposed, we use the somewhat heavier terminology *pseudo-Mahalanobis signs* and *pseudo-Mahalanobis ranks.*

1.3.2 Hyperplane-based signs and ranks

Another approach to reconstructing the genuine signs $\mathbf{U}_{\mathbf{\Sigma};t}^{(n)}$ and genuine ranks $R_{\mathbf{\Sigma};t}^{(n)}$ is based on counts of hyperplanes.

For the signs, the idea is due to Randles (1989). For any pair $\mathbf{Z}_{t_1}^{(n)}, \mathbf{Z}_{t_2}^{(n)}$, $1 \leq t_1 \neq t_2 \leq n$, consider the $\begin{pmatrix} n-2 \\ k-1 \end{pmatrix}$ hyperplanes going through the origin and $(k-1)$ out of the $(n-2)$ remaining

$\mathbf{Z}_t^{(n)}$'s ($t_1 \neq t \neq t_2$). Define the *interdirection* $c_{t_1 t_2}^{(n)}$ as the number of such hyperplanes separating $\mathbf{Z}_{t_1}^{(n)}$ and $\mathbf{Z}_{t_2}^{(n)}$. Interdirections are invariant under the affine group $\mathcal{G}$ and under the group $\mathcal{G}_{\mathbf{\Sigma}}$ of radial transformations, irrespective of $\mathbf{\Sigma}$. Due to this invariance, it is intuitively clear that $\pi p_{t_1 t_2}^{(n)} := \pi c_{t_1 t_2}^{(n)} / \binom{n-2}{k-1}$ is a consistent estimate of the angle $\arccos(\mathbf{U}'_{\mathbf{\Sigma};t_1}\mathbf{U}_{\mathbf{\Sigma};t_2})$ between $\mathbf{U}_{\mathbf{\Sigma};t_1}^{(n)}$ and $\mathbf{U}_{\mathbf{\Sigma};t_2}^{(n)}$. Interdirections thus allow for a reconstruction of those angles (equivalently, of their cosines $\mathbf{U}'_{\mathbf{\Sigma};t_1}\mathbf{U}_{\mathbf{\Sigma};t_2}$, since the $\mathbf{U}_{\mathbf{\Sigma};t}$'s are unit vectors): quite remarkably, they do the same job, with the same invariance properties, as the *Tyler cosines* $\mathbf{U}'_{\widehat{\mathbf{\Sigma}};t_1}\mathbf{U}_{\widehat{\mathbf{\Sigma}};t_2}$, but without requiring any estimation of $\mathbf{\Sigma}$.

The hyperplane-based cosines $p_{t_1 t_2}^{(n)}$ are sufficient for the first problem we are treating below (Section 3). For the second problem (Section 4), we need the slightly more informative concept of *absolute interdirections* (Hallin and Paindaveine 2003c and d). 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 separating $\mathbf{Z}_{t_1}^{(n)}$ and $\mathbf{Z}_{t_2}^{(n)}$, we now count the number $c_{t;i}^{(n)}$ of hyperplanes separating $\mathbf{Z}_t^{(n)}$ and the transformed unit vectors $\widehat{\mathbf{\Sigma}}^{1/2}\mathbf{u}_i$, $i = 1, \ldots, k$, where $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ forms the canonical basis of $\mathbb{R}^k$. Then, for the same reasons as above, $\pi p_{t;i}^{(n)} := \pi c_{t;i}^{(n)} / \binom{n-1}{k-1}$ allows for a consistent estimation of the angles $\arccos(\mathbf{U}'_{\mathbf{\Sigma};t}\mathbf{u}_i)$, $i = 1, \ldots, k$, so that the vectors $(\cos(\pi p_{t;i}), i = 1, \ldots, k)$ are consistent estimators of the signs $\mathbf{U}_{\mathbf{\Sigma};t}$ 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 (2003). This concept relies on the so-called *lift-interdirections*.

For any $\mathbf{Z}_t^{(n)}$, consider the $\binom{n-1}{k}$ hyperplanes going through k out of the $(n-1)$ remaining $\mathbf{Z}_{t'}^{(n)}$'s ($t' \neq t$). The lift-interdirection $\ell_t^{(n)}$ associated with $\mathbf{Z}_t^{(n)}$ is defined as the number of such hyperplanes that separate $\mathbf{Z}_t^{(n)}$ and $-\mathbf{Z}_t^{(n)}$. Lift-interdirections can be shown to converge to some monotone increasing function of the distances $d_{\mathbf{\Sigma};t}^{(n)}$; as for their ranks, they converge to the genuine ranks $R_{\mathbf{\Sigma};t}^{(n)}$ (see Oja and Paindaveine 2003 for details). Again, we are able to reconstruct, as $n \to \infty$, a quantity that depends on the unspecified shape matrix $\mathbf{\Sigma}$ without estimating it.

2 The general linear model with VAR errors

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

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

of k-variate random vectors satisfying

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

where

$$\mathbf{X}^{(n)} := \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,m} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,m} \end{pmatrix} := \begin{pmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} := \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,k} \\ \vdots & \vdots & & \vdots \\ \beta_{m,1} & \beta_{m,2} & \dots & \beta_{m,k} \end{pmatrix} := \begin{pmatrix} \boldsymbol{\beta}_1' \\ \vdots \\ \boldsymbol{\beta}_m' \end{pmatrix}$$

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

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

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

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

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

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

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 $\mathbf{G}_0 = \mathbf{I}_k$, $\mathbf{G}_{-1} = \mathbf{G}_{-2} = \dots, \mathbf{G}_{-p+1} = \mathbf{0}$. The remainder term $\mathbf{r}_t$ is related to the influence of the unobserved initial values $\mathbf{V}_0, \dots, \mathbf{V}_{-p+1}$; 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 $\boldsymbol{\theta} := (\text{vec}'(\boldsymbol{\beta}'), \text{vec}'(\mathbf{A}_1), \dots, \text{vec}'(\mathbf{A}_p))' \in \mathbb{R}^{km+k^2p} =: \mathbb{R}^K$, 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

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

The Durbin-Watson testing problem deals with the null hypothesis that $\mathbf{V}_t$ is white noise, i.e., 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 square estimate of $\boldsymbol{\beta}$, and by $\mathbf{Z}_t := \mathbf{Y}_t - \hat{\boldsymbol{\beta}}_{\mathcal{N}}^{(n)\prime}\mathbf{x}_t$ 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

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

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

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 one $\hat{\boldsymbol{\beta}}_{\mathcal{N}}^{(n)}$ (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 (2003d), Section 2.1). If however the tests are to remain valid under infinite second-order moments, robust estimators, resisting heavy-tailed distributions, such as the M-estimators proposed by Davis et al. (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 mixing the two types (Tyler signs, for instance, with lift-interdirection ranks) are somewhat incoherent, and we will restrict to combining signs and ranks of the same type (either pseudo-Mahalanobis or hyperplane-based); the same notation will be used for both cases.

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

(a) a multivariate Durbin-Watson statistic of the sign-test type

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

(b) a multivariate Durbin-Watson statistic of the Spearman type

$$\begin{aligned} W_{\mathrm{DW;\,Sp}}^{(n)} &:= \frac{9k^2}{(n-1)(n+1)^4}\sum_{s,t=2}^{n}R_s^{(n)}R_{s-1}^{(n)}R_t^{(n)}R_{t-1}^{(n)}(\mathbf{U}_s'\mathbf{U}_t)(\mathbf{U}_{s-1}'\mathbf{U}_{t-1}) \\ &= \frac{9k^2}{(n-1)(n+1)^4}\left\|\sum_{t=2}^{n}R_t^{(n)}R_{t-1}^{(n)}\mathbf{U}_t\mathbf{U}_{t-1}'\right\|^2, \end{aligned} \tag{10}$$

(c) a multivariate Durbin-Watson statistic of the van der Waerden type

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

$$= \frac{1}{(n-1)} \left\| \sum_{t=2}^{n} \Phi_k^{-1}\!\left(\frac{R_t^{(n)}}{n+1}\right) \Phi_k^{-1}\!\left(\frac{R_{t-1}^{(n)}}{n+1}\right) \mathbf{U}_t\mathbf{U}'_{t-1} \right\|^2, \tag{11}$$

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

3.3 Asymptotic relative efficiencies

The asymptotic relative efficiencies, with respect to the traditional Gaussian procedure described in Section 3.1, of the signed-rank tests of Section 3.2 have been derived in Hallin and Paindaveine (2003d), where a multivariate serial version of the classical Chernoff-Savage result is also established. This result shows that the asymptotic relative efficiency, with respect to the Gaussian procedure based on (8), of the van der Waerden tests (c) based on (11) is uniformly larger than one. 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 two degrees of freedom in order for the Gaussian procedure to be valid.

		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

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, $k = 1, 2, 4, 6, 10$.

3.4 Numerical study

3.4.1 Size and Power

In order 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

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

with initial value $\mathbf{V}_0 = \mathbf{0}$, $\boldsymbol{\beta}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\boldsymbol{\beta}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, 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 for achieving approximate stationarity), 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, and

(d2) the van der Waerden Durbin-Watson test, based on (11) with hyperplane-based signs and ranks, respectively.

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

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

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 in 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 $\hat{\boldsymbol{\Sigma}}_{\mathrm{T}}$ was computed from the algorithm of Randles (2000). 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 .014, .025, and .031, for frequencies of the order of .05 (.95), .20 (.80), and .50, respectively. It appears that none of the rejection frequencies significantly differs from the nominal 5% level. All tests thus apparently are valid and unbiased—even the Gaussian one under 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 one under Gaussian density, a slight superiority under t_8 density, and a more marked one 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 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 are doing slightly better than their Tyler counterparts.

		Autoregression matrix					Autoregression matrix			
test	innovation density	**0**	**A**	2**A**	ARE	innovation density	**0**	**A**	2**A**	ARE
$\phi_{\mathcal{N}}$		0.0460	0.2640	0.9020	1.000		0.0580	0.2380	0.9030	1.000
ϕ_{vdW}		0.0440	0.2530	0.8910	1.000		0.0410	0.3460	0.9590	1.400
ϕ_S		0.0460	0.1620	0.6800	0.617		0.0490	0.2620	0.8650	1.000
ϕ_{SP}	$\mathcal{N}$	0.0510	0.2580	0.8920	0.970	t_3	0.0420	0.3460	0.9650	1.305
ϕ^h_{vdW}		0.0420	0.2580	0.8850	1.000		0.0390	0.3380	0.9600	1.400
ϕ^h_S		0.0500	0.1590	0.6820	0.617		0.0450	0.2580	0.8600	1.000
ϕ^h_{SP}		0.0460	0.2620	0.8870	0.970		0.0450	0.3360	0.9600	1.305
$\phi_{\mathcal{N}}$		0.0440	0.2640	0.9040	1.000		0.0460	0.2640	0.9020	undefined
ϕ_{vdW}		0.0410	0.2650	0.9070	1.047		0.0440	0.2530	0.8910	undefined
ϕ_S		0.0420	0.1890	0.7410	0.711		0.0460	0.1620	0.6800	undefined
ϕ_{SP}	t_8	0.0510	0.2790	0.9030	1.022	t_1	0.0510	0.2580	0.8920	undefined
ϕ^h_{vdW}		0.0410	0.2660	0.9050	1.047		0.0420	0.2570	0.8850	undefined
ϕ^h_S		0.0430	0.1880	0.7400	0.711		0.0500	0.1590	0.6820	undefined
ϕ^h_{SP}		0.0460	0.2730	0.9050	1.022		0.0460	0.2620	0.8860	undefined

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.

3.4.2 Robustness

In order 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 only consider Gaussian series.

The same Monte-Carlo scheme as in Section 3.4.1 was used to generate 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 were subjected to the following perturbations (inducing observation outliers):

($\mathcal{Y}^+$) *(observation outliers)* $\mathbf{Y}^*_t$ was replaced with $5\mathbf{Y}^*_t$ at time $t =$ 549, 550, 599, 600, 649, 650;

($\mathcal{Y}^-$) *(observation outliers)* $\mathbf{Y}^*_t$ was replaced with $5\mathbf{Y}^*_t$ at time $t =$ 549, 599, 649, and with $-5\mathbf{Y}^*_t$ at time $t =$ 550, 600, 650, respectively;

($\mathcal{E}^+$) *(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;

$N = 1000$ series of each type were generated. The last $n = 150$ observations then were subjected to the 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 one 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 one; as expected, the sign tests seem to be slightly more robust than the van der Waerden and Spearman ones.

test	type of outliers	Autoregression matrix $\mathbf{0}$	$\mathbf{A}$	$2\mathbf{A}$	type of outliers	Autoregression matrix $\mathbf{0}$	$\mathbf{A}$	$2\mathbf{A}$
$\phi_{\mathcal{N}}$	$\mathcal{Y}^+$	0.6250	0.7380	0.9020	$\mathcal{E}^+$	0.5080	0.6390	0.8820
ϕ_{vdW}		0.0730	0.3480	0.8810		0.0570	0.3140	0.8920
ϕ_S		0.0570	0.1870	0.6900		0.0460	0.1760	0.7150
ϕ_{SP}		0.0590	0.3130	0.8980		0.0560	0.3050	0.8920
ϕ^h_{vdW}		0.0670	0.3480	0.8850		0.0560	0.3190	0.8800
ϕ^h_S		0.0560	0.1800	0.6920		0.0460	0.1730	0.7140
ϕ^h_{SP}		0.0570	0.3150	0.8980		0.0520	0.3050	0.8860
$\phi_{\mathcal{N}}$	$\mathcal{Y}^-$	0.7040	0.6700	0.7130	$\mathcal{E}^-$	0.5360	0.6600	0.8770
ϕ_{vdW}		0.0850	0.1500	0.6140		0.0720	0.3300	0.9080
ϕ_S		0.0600	0.1510	0.5810		0.0510	0.1850	0.7230
ϕ_{SP}		0.0760	0.1650	0.7100		0.0690	0.3200	0.9060
ϕ^h_{vdW}		0.0840	0.1590	0.6100		0.0680	0.3230	0.9010
ϕ^h_S		0.0610	0.1490	0.5840		0.0480	0.1850	0.7210
ϕ^h_{SP}		0.0720	0.1720	0.7040		0.0630	0.3200	0.9050

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}), the Spearman (ϕ_{SP}), and the sign-test type (ϕ_S) Durbin-Watson tests, and for their hyperplane-based counterparts ϕ^h_{vdW}, ϕ^h_{SP}, and ϕ^h_S, at (asymptotic) probability level 5%; the series length throughout is $n = 150$.

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($p_0 + 1$) one; for simplicity, we assume that $\boldsymbol{\beta} = \mathbf{0}$; $\mathbf{\Sigma}$ and f of course are nuisance parameters in this problem. A sequential application of such tests can be used in the identification of 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 $\mathbf{\Theta}_{p_0}$ the set of all values of $\boldsymbol{\theta} \in \mathbb{R}^K$ such that $\mathbf{A}_{p_0+1} = \ldots = \mathbf{A}_p = \mathbf{0}$, $|\mathbf{A}_{p_0}| \neq 0$, 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, e.g., 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 $(\text{vec}'(\hat{\mathbf{A}}_1), \ldots, \text{vec}'(\hat{\mathbf{A}}_{p_0}), 0', \ldots, 0')'$. 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

$$\boldsymbol{\Gamma}_i^{(n)} := (n - p_0 - i)^{-1} \sum_{t=p_0+1+i}^{n} \mathbf{Z}_t \mathbf{Z}'_{t-i} = (n - p_0 - i)^{-1} \sum_{t=p_0+1+i}^{n} d_{\boldsymbol{\Sigma};t}\, d_{\boldsymbol{\Sigma};t-i}\, \boldsymbol{\Sigma}^{1/2} \mathbf{U}_{\boldsymbol{\Sigma};t} \mathbf{U}'_{\boldsymbol{\Sigma};t-i} \boldsymbol{\Sigma}'^{1/2};$$

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

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

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

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

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

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

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

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

respectively; note that $\mathbf{w}^2$ and $\mathbf{W}^2$ only differ 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 in 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-square 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 involving 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 $\mathbf{G}_u = \hat{\mathbf{A}}^u$)

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

and

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

The order selection procedure then consists in 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 (that is, 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) and (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 of them can be computed from Tyler signs and ranks, or from hyperplane-based ones; in case interdirections are used, they should be the "absolute" ones. The three statistics are

(a) a test statistic of the sign-test type,

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

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

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

substituted for $\boldsymbol{\Gamma}_i^{(n)}$;

(b) a test statistic of the Spearman type

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

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

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

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

(c) 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*

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

(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 square 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, needs not be the empirical marginal covariance matrix anymore.

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 (2003d) thus still apply here. However, a more pertinent assesment 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 again, 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

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

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

Seven versions (Gaussian or rank-based) of those order identification procedure were performed on each series. Step one of each procedure consists in 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 in Hallin and Paindaveine (2002b). If the hypothesis of randomness cannot be rejected, the model is identified as being VAR(0), that is, white noise (under-identification). 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 over-identification of the order. Of course, it is pretty natural to use the same type of tests 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,

(c1) the van der Waerden type procedure based on Tyler's signs and ranks,

(c2) 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 for estimating $\mathbf{A}$. The Tyler estimate $\hat{\mathbf{\Sigma}}_{\mathrm{T}}$ 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 over-identification 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 .014, .025, and .031, for frequencies of the order of .05 (.95), .20 (.80), and .50, respectively. Inspection of that 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) are doing uniformly better than their Tyler-type competitors;
- although the validity of the tests used at each step of the identification procedure is not formally established 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.

test	innovation density	order identification 0	1	≥ 2	ARE	innovation density	order identification 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

Table 4: Under-identification ($p = 0$), correct identification ($p = 1$), and over-identification ($p \geq 2$) 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. All tests are 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).

4.4.2 Robustness

A robustness investigation also was conducted, on the model of Section 3.4.2, for the various order-identification procedures proposed in Sections 4.1 and 4.2. Observation $(\mathbf{Y}^*_1, \ldots, \mathbf{Y}^*_{620})$ 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, in order 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, with $-5\mathbf{Y}^*_t$ for $t = 540$, 580, and 620;

($\mathcal{E}^+$) *(innovation outliers)* $\boldsymbol{\varepsilon}_t$ was replaced with $5\boldsymbol{\varepsilon}_t$ for $t = 538$, 578, 580, 618, and 620;

($\mathcal{E}^-$) *(innovation outliers)* $\boldsymbol{\varepsilon}_t$ was replaced with $5\boldsymbol{\varepsilon}_t$ for $t = 538$, 578, and 618, with $-5\boldsymbol{\varepsilon}_t$ for $t = 540$, 580, and 620, respectively.

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 over-identification frequencies are reported in Table 5. This simulation exercise of course is somewhat limited, and only allows 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 outliers.

		Order identification				Order identification		
test	type of outliers	0	1	≥ 2	type of outliers	0	1	≥ 2
$\phi_{\mathcal{N}}$		428	293	279		88	522	390
ϕ_{vdW}		98	822	80		27	909	64
ϕ_S		189	762	49		140	819	41
ϕ_{SP}	$\mathcal{Y}^+$	81	851	68	$\mathcal{E}^+$	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
$\phi_{\mathcal{N}}$		672	180	148		77	520	403
ϕ_{vdW}		217	708	75		25	911	64
ϕ_S		290	662	48		133	817	50
ϕ_{SP}	$\mathcal{Y}^-$	179	749	72	$\mathcal{E}^-$	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

Table 5: Under-identification ($p = 0$), correct-identification ($p = 1$), and over-identification ($p \geq 2$) frequencies (out of $N = 1000$ replications) in various perturbed Gaussian VAR(1) series. 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 $n = 120$.

5 Conclusions

Rank-based methods for a long time have been confined to problems involving 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, mainly: the pseudo-Mahalanobis or Tyler ones, and the hyperplane-based or Oja-Paindaveine ones. Theoretical results establish that these methods are as efficient, locally and asymptotically, as their everyday-practice parametric competitors based on cross-correlation matrices; their van der Waerden versions even uniformly dominate them. 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.

References

[1] Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1993). *Efficient and Adaptive Statistical Inference for Semiparametric Models*, John Hopkins University Press, Baltimore.

[2] Davis, R.A. and W. Wu. (1997). M-estimation for linear regression with infinite variance, *Probability and Mathematical Statistics* **17**, 1-20.

[3] Garel, B. and M. Hallin (1999). Rank-based AR order identification, *Journal of the American Statistical Association* **94**, 1357-1371.

[4] Hájek, I. and Z. Šidák (1967). *Theory of Rank Tests*, Academic Press, New York.

[5] Hájek, I., Z. Šidák, and P.K. Sen (1999). *Theory of Rank Tests*, 2nd edition, Academic Press, New York.

[6] Hallin, M., J.-Fr. Ingenbleek, and M. L. Puri (1985). Linear serial rank tests for randomness against ARMA alternatives, *Annals of Statistics* **13**, 1156-1181.

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

[8] Hallin, M. and D. Paindaveine (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence, *Bernoulli* **8**, 787-816.

[9] Hallin, M. and D. Paindaveine (2002c). Multivariate signed ranks : Randles' interdirections or Tyler's angles? In Y. Dodge, Ed., *Statistical Data Analysis Based on the L_1 Norm and Related Procedures*, Birkhäuser, Basel, 271-282.

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

[11] Hallin, M. and D. Paindaveine (2003b). Rank-based optimal tests of the adequacy of an elliptic VARMA model, submitted.

[12] Hallin, M. and D. Paindaveine (2003c). Asymptotic linearity of serial and nonserial multivariate signed rank statistics, submitted.

[13] Hallin, M. and D. Paindaveine (2003d). Affine invariant aligned rank tests for the multivariate general linear model with ARMA errors, submitted.

[14] Hallin, M. and M.L. Puri (1988). Optimal rank-based procedures for time-series analysis : testing an ARMA model against other ARMA models, *Annals of Statistics* **16**, 402-432.

[15] Hallin, M. and M.L. Puri (1991). Time-series analysis via rank-order theory : signed-rank tests for *ARMA* models, *Journal of Multivariate Analysis* **39**, 1-29.

[16] Hallin, M. and M.L. Puri (1992).Rank tests for time-series analysis : a survey. In D. Brillinger, E. Parzen, and M. Rosenblatt Eds, *New Directions in Time Series Analysis*, Springer-Verlag, New York, 111-154.

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

[18] Hallin, M. and B. J. M. Werker (2003). Semiparametric efficiency, distribution-freeness, and invariance, *Bernoulli* **9**, 55-65.

[19] Hettmansperger, T.P. (1984). *Statistical Inference Based on Ranks*, John Wiley, New York.

[20] Hettmansperger, T.P. and R.H. Randles (2002). A practical affine equivariant multivariate median, *Biometrika* **89**, 851-860.

[21] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2nd edition, J. Wiley, New York.

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

[23] Oja, H. and D. Paindaveine (2003). Optimal testing procedures based on hyperplanes, submitted.

[24] Pötscher, B.M. (1983). Order estimation in ARMA models by Lagrangian multiplier tests, *Annals of Statistics* **11**, 872-885.

[25] Pötscher, B.M. (1985). The behaviour of the Lagrange multiplier test in testing the orders of an ARMA model, *Metrika* **32**, 129-150.

[26] Pratt, J.W. and J.D. Gibbons (1981). *Concepts of Nonparametric Theory*, Springer Verlag, New York.

[27] Puri, M.L. and P.K. Sen (1971). *Nonparametric Methods in Multivariate Analysis*, John Wiley, New York.

[28] Puri, M.L. and P.K. Sen (1985). *Nonparametric Methods in General Linear Models*, John Wiley, New York.

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

[30] Randles, R.H. and D.A. Wolfe (1979). *Introduction to the Theory of Nonparametric Statistics*, John Wiley, New York.

[31] Randles, R.H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test, *Journal of the American Statististical Association* **95**, 1263-1268.

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

Marc Hallin
Département de Mathématique,
I.S.R.O., and E.C.A.R.E.S.
Université Libre de Bruxelles
Campus de la Plaine CP 210
B-1050 Bruxelles Belgium
mhallin@ulb.ac.be

Davy Paindaveine
Département de Mathématique,
I.S.R.O., and E.C.A.R.E.S.
Université Libre de Bruxelles
Campus de la Plaine CP 210
B-1050 Bruxelles Belgium
dpaindav@ulb.ac.be