Center-Outward R-Estimation for Semiparametric VARMA Models

M. Hallin, D. La Vecchia, and H. Liu

ECARES, Université libre de Bruxelles CP 114/4
Avenue F.D. Roosevelt 50 - B-1050 Bruxelles, Belgium
Email: mhallin@ulb.ac.be

Research Center for Statistics, University of Geneva
Boulevard du Pont d'Arve 40 - CH-1211 Geneva, Switzerland
Email: davide.lavecchia@unige.ch

Department of Mathematics and Statistics, Lancaster University
LA1 4YF Lancaster, UK
Email: h.liu11@lancaster.ac.uk

Abstract

We propose a new class of estimators for semiparametric VARMA models with the innovation density playing the role of nuisance parameter. Our estimators are R-estimators based on the multivariate concepts of center-outward ranks and signs recently proposed by Hallin (2017). We show how these concepts, combined with Le Cam's asymptotic theory of statistical experiments, yield a robust yet flexible and powerful class of estimation procedures for multivariate time series. We develop the relevant asymptotic theory of our R-estimators, establishing their root-n consistency and asymptotic normality under a broad class of innovation densities including, e.g., multimodal mixtures of Gaussians or and multivariate skew-t distributions. An implementation algorithm is provided in the supplementary material, available online. A Monte Carlo study compares our R-estimators with the routinely-applied Gaussian quasi-likelihood ones; the latter appear to be quite significantly outperformed away from elliptical innovations. Numerical results also provide evidence of considerable robustness gains. Two real data examples conclude the paper.

Keywords Multivariate ranks, Distribution-freeness, Local asymptotic normality, Measure transportation, Quasi likelihood estimation, Skew innovation density.

1 Introduction

1.1 Quasi-maximum likelihood and R-estimation in time-series models

Quasi-likelihood methods, which include QMLE (quasi-maximum likelihood estimation) and correlogram-based testing, are standard daily practice in the statistical analysis of time series, univariate and multivariate, linear and non-linear. Focusing on estimation problems in linear models (autoregressive moving average, ARMA) the properties of (Gaussian) QMLE are generally considered as fully satisfactory: the estimator is root-n consistent and asymptotically normal, under finite fourth-order moment assumptions. The case of nonlinear models (with, e.g., exponential QMLE for non-linear volatility and multiplicative error models) is roughly similar.

Despite of their popularity, however, QMLE methods are not without some undesirable consequences. *(i)* While achieving efficiency under Gaussian innovations, their actual performance can be arbitrarily bad under non-Gaussian ones. More precisely, their asymptotic relative efficiency with respect to the efficient estimator, depending on the actual innovation density, can be arbitrarily close to zero. *(ii)* Due to technical reasons (the *Fisher consistency* requirement), the choice of a quasi-likelihood is the most pessimistic one: quasi-likelihoods automatically are based on the *least favorable* innovation density (e.g.,ARMA models with Gaussian innovations). *(iii)* Actual fourth-order moments might be infinite, thus the validity of the QMLE can be questionable.

In principle, the ultimate theoretical remedy to those problems is the semiparametric estimation method described in the monograph by Bickel et al. (1993), which yields uniformly locally and asymptotically semiparametrically efficient estimators (in ARMA models, which are *adaptive*, semiparametric and parametric efficiency coincide). Semiparametric estimation procedures, however, are not easily implementable, since they require kernel-based estimation of the actual innovation density. This means the choice of a kernel, the selection of a bandwidth, and the use of sample splitting techniques and needs relatively large samples. It explains why the approach is seldom considered in practice.

A more flexible and computationally less heavy alternative is R-estimation, which reaches efficiency at some chosen reference density (not necessarily Gaussian or least favorable) or class of densities. R-estimation has been proposed first in the context of location (Hodges and Lehmann 1956) and regression (Jurečková 1971, Koul 1971, van Eeden 1972, Jaeckel 1972) models with independent observations. R-estimation later on was extended to autoregressive time series models (Koul and Saleh 1993, Koul and Ossiander 1994, Terpstra et al. 2001, Hettmansperger and McKean 2008, Mukherjee and Bai 2002, Andrews 2008, 2012)—note, however, that the R-estimators considered by these authors are not genuinely rank-based, as the objective functions they are based on involve both the ranks and the observations themselves. Extensions to the estimation of non-linear time series such as AR-GARCH, discretely observed diffusions with jumps, or autoregressive conditional duration models were considered by Mukherjee (2007), Andreou and Werker (2015), and Hallin and La Vecchia (2017, 2019).

The drawbacks of quasi-likelihood methods for observations in dimension $d = 1$ only get worse as d increases. The finite-sample performance of a VAR(1) Gaussian QMLE under a mixture of Gaussians (see Figure 2), for instance, can be quite terrible—although the conditions for root-n consistency and asymptotic normality are perfectly met. Also the use of the semiparametric method of Bickel et al. becomes problematic (if not infeasible), since the higher the dimension, the larger the required sample (curse of dimensionality). Thus, a natural question is: can R-estimation palliate such drawbacks in dimension $d > 1$ the way it does in dimension one?

This question immediately comes up against another one: what are ranks and signs, hence what is R-estimation, in dimension $d > 1$? Starting with dimension $d = 2$, indeed, the space $\mathbb{R}^d$ is no longer canonically ordered. Based on measure-transportation results, a data-driven ordering yielding a concept of ranks and signs for multivariate observations recently proposed by Chernozhukov et al. (2017) has been developed by Hallin (2017) and del Barrio et al. (2018). Those *center-outward ranks and signs* (see Section 3.2 for details) enjoy all the properties that make traditional univariate ranks a natural and successful tool of inference when $d = 1$. In particular, they are distribution-free and independent, irrespective of the actual density, of the corresponding order statistic. In this paper, we make use of those new notions of ranks and signs to derive a novel class of R-estimators for VARMA models with unspecified innovation densities, we establish its asymptotics and illustrate numerically how it outperforms the routinely-applied QMLE.

Several attempts had been made previously to introduce ranks and signs in a multivariate context. Among them are the componentwise ranks (Puri and Sen 1971) and the spatial ranks (Oja 2010), which are not distribution-free, and the depth-based ranks (Liu 1992; Liu and Singh 1993) which are distribution-free, but not "maximal distribution-free": see the introduction of Hallin (2017) for details and a survey. The concept of *Mahalanobis ranks and signs*, due to Hallin and Paindaveine (2002a), has been quite successful in the time-series context (Hallin and Paindaveine 2002b, 2004); the validity of their methods, however, is limited to the case of elliptical innovations—an assumption we are not willing to make here. The center-outward ranks and signs we are considering below extend the validity of Mahalanobis ranks and signs-based methods, essentially, to arbitrary absolutely continuous distributions in $\mathbb{R}^d$.

1.2 Center-outward R-estimation at a glance

To justify our method and illustrate its performance, consider the following two continuous bivariate innovation densities. The first one is a bivariate $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$ density ($\boldsymbol{I}_2$ the 2×2 identity matrix), the second is a mixture

$$\frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \frac{3}{8}\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) + \frac{1}{4}\mathcal{N}(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3), \tag{1.1}$$

of three Gaussian densities, with

$$\boldsymbol{\mu}_1 = (-3, 0)',\ \boldsymbol{\mu}_2 = (3, 0)',\ \boldsymbol{\mu}_3 = (0, 0)',\ \boldsymbol{\Sigma}_1 = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix},\ \boldsymbol{\Sigma}_2 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix},\ \text{and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}.$$

Those two densities were used to generate independent innovations $\boldsymbol{\epsilon}_t$ in $N = 300$ replications of length $n = 1000$ of the stationary solution of the bivariate VAR(1) model

$$(\boldsymbol{I}_2 - \boldsymbol{A}L)\, \boldsymbol{X}_t \;=\; \boldsymbol{\epsilon}_t, \qquad t \in \mathbb{Z},$$

where L as usual denotes the lag operator and $\boldsymbol{A} = (a_{ij})$ is a 2×2 matrix satisfying the classical stationarity requirements. The VAR(1) parameter to be estimated is thus $\text{vec}(\boldsymbol{A}) = (a_{11}, a_{21}, a_{12}, a_{22})'$. For the simulation we set $a_{11} = 0.2$, $a_{21} = -0.6$, $a_{12} = 0.3$, and $a_{22} = 1.1$. For each replication, we computed the (Gaussian) QMLE, the R-estimator based on the signs, the Spearman, and the van der Waerden R-estimators. The results are presented as boxplots in Figures 1 and 2. In the Gaussian case (Figure 1), of course, the Gaussian QMLE achieves parametric efficiency; the mixture (1.1) generates a multimodal distribution, which nevertheless satisfies all the conditions required for a root-n consistent and asymptotically normal QMLE.

Even a rapid inspection of Figures 1 and 2 reveals the overwhelming superiority of R-estimators over the QMLE. Under the mixture innovations (Figure 2), the bias and mean squared deviations of the van der Waerden and Spearman estimators are small, while those

Figure 1: Boxplots of the QMLE and R-estimators (signs, Spearman, van der Waerden) under spherical Gaussian innovations (sample size $n = 1000$; $N = 300$ replications). The horizontal red line represents the actual parameter value.

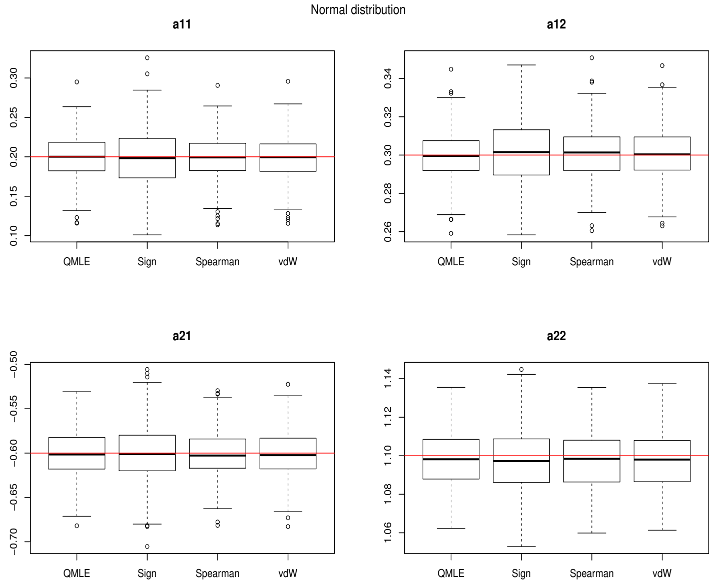
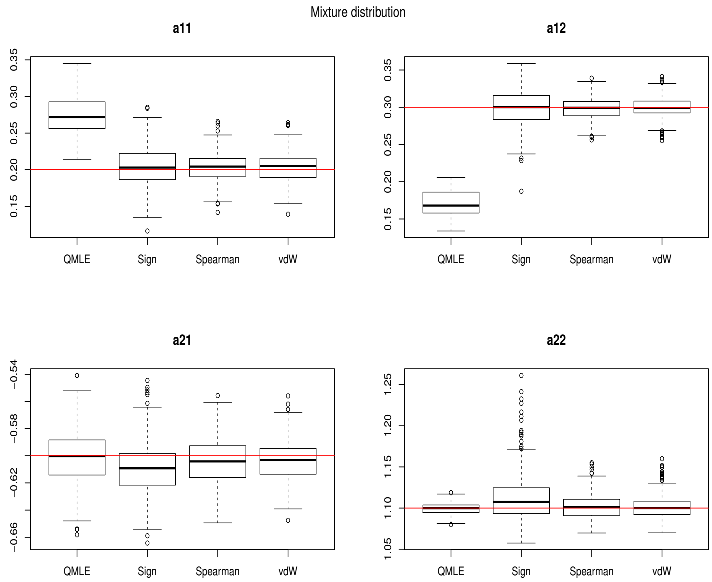


Figure 2: Boxplots of the QMLE and R-estimators (signs, Spearman, van der Waerden) under the mixture (1.1) of three multivariate normal distributions (sample size $n = 1000$; $N = 300$ replications). The horizontal red line represents the actual parameter value.



of the QMLE for a_{11} and a_{12}, for instance, both are dramatic. On the other hand, under Gaussian innovations (Figure 1), the QMLE is optimal but all R-estimators are nearly as good. The good performance of R-estimation under the multimodal mixture (1.1), thus, is not obtained at the cost of its performance under well-behaved Gaussian innovations.

2 Local asymptotic normality

2.1 Notation and basic assumptions

We throughout consider the d-dimensional VARMA(p, q) model

$$\left(\boldsymbol{I}_d - \sum_{i=1}^{p} \boldsymbol{A}_i L^i\right) \boldsymbol{X}_t = \left(\boldsymbol{I}_d + \sum_{j=1}^{q} \boldsymbol{B}_j L^j\right) \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}, \tag{2.1}$$

where $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_q$ are $d \times d$ matrices, L is the lag operator, and $\{\boldsymbol{\epsilon}_t; t \in \mathbb{Z}\}$ is an i.i.d. mean $\mathbf{0}$-innovation with probability density f. The observed series is $\{\boldsymbol{X}_1^{(n)}, \ldots, \boldsymbol{X}_n^{(n)}\}$ (superscript $^{(n)}$ omitted whenever possible), and the $(p + q)d^2$-dimensional parameter of interest is

$$\boldsymbol{\theta} := ((\text{vec}\boldsymbol{A}_1)', \ldots, (\text{vec}\boldsymbol{A}_p)', (\text{vec}\boldsymbol{B}_1)', \ldots, (\text{vec}\boldsymbol{B}_q)')'.$$

Letting $\boldsymbol{A}_0 := \boldsymbol{I}_d =: \boldsymbol{B}_0$, $\boldsymbol{A}(L) := \boldsymbol{A}_0 - \sum_{i=1}^{p} \boldsymbol{A}_i L^i$ and $\boldsymbol{B}(L) := \boldsymbol{B}_0 + \sum_{j=1}^{q} \boldsymbol{B}_j L^j$, the following conditions, which are standard in VARMA modelling, are assumed to hold (they actually define the parameter space $\boldsymbol{\Theta}$).

Assumption (A1). *(i)* All solutions of the determinantal equations

$$\det\left(\sum_{i=0}^{p} \boldsymbol{A}_i z^i\right) = 0 \text{ and } \det\left(\sum_{i=0}^{q} \boldsymbol{B}_i z^i\right) = 0, \ z \in \mathbb{C}$$

lie outside the unit ball in $\mathbb{C}$; *(ii)* $|\boldsymbol{A}_p| \neq 0 \neq |\boldsymbol{B}_q|$; *(iii)* $\boldsymbol{I}_d$ is the greatest common left divisor of $\sum_{i=0}^{p} \boldsymbol{A}_i z^i$ and $\sum_{i=0}^{q} \boldsymbol{B}_i z^i$.

To proceed with the statement of local asymptotic normality, we need some algebraic preparation, which we borrow from Garel and Hallin (1995) and Hallin and Paindaveine (2004); that preparation is needed, essentially, for the explicit expressions of the central sequence (2.3) and the Fisher information (2.6) below, and can be skipped at fist reading. The interested reader will find this technical material in Appendix A.

Throughout, we assume that f is non-vanishing with respect to the Lebesgue measure μ

on $\mathbb{R}^d$. More precisely we assume that, for all $c \in \mathbb{R}^+$, there exist $b_{c;f}$ and $a_{c;f}$ in $\mathbb{R}$ such that

$$0 < b_{c;f} \leq a_{c;f} < \infty \quad \text{and} \quad b_{c;f} \leq f(\boldsymbol{x}) \leq a_{c;f}$$

for $\|\boldsymbol{x}\| \leq c$. Denote by $\mathcal{F}_d$ this family of densities, which is the one for which the center-outward distribution functions to be defined in Section 3.1 are shown to be continuous in Hallin (2017). In order to have LAN, we moreover are making the following regularity assumptions.

Assumption (A2). The density $f \in \mathcal{F}_d$ is such that *(i)* $\int \boldsymbol{x} f(\boldsymbol{x}) d\mu = \mathbf{0}$, $\int \boldsymbol{x}\boldsymbol{x}' f(\boldsymbol{x}) d\mu = \boldsymbol{\Xi}$, with $\boldsymbol{\Xi}$ positive definite; *(ii)* there exists a square integrable random vector $\boldsymbol{D}f^{1/2}$ such that for all sequence $\boldsymbol{h} \in \mathbb{R}^d$ such that $\mathbf{0} \neq \boldsymbol{h} \to \mathbf{0}$,

$$(\boldsymbol{h}'\boldsymbol{h})^{-1} \int [f^{1/2}(\boldsymbol{x}+\boldsymbol{h}) - f^{1/2}(\boldsymbol{x}) - \boldsymbol{h}'\boldsymbol{D}f^{1/2}(\boldsymbol{x})]^2 d\mu \to 0,$$

i.e., $f^{1/2}$ is mean-square differentiable, with mean square gradient $\boldsymbol{D}f^{1/2}$; *(iii)* letting

$$\boldsymbol{\varphi}_f(\boldsymbol{x}) := (\varphi_1(\boldsymbol{x}), \ldots, \varphi_d(\boldsymbol{x}))' := -2(\boldsymbol{D}f^{1/2})/f^{1/2}, \tag{2.2}$$

$\int [\varphi_i(\boldsymbol{x})]^4 f(\boldsymbol{x}) d\mu < \infty$, $i = 1, \ldots, d$; *(iv)* the score function $\boldsymbol{\varphi}_f$ is piecewise Lipschitz, i.e., there exists a finite measurable partition of $\mathbb{R}^d$ into J non-overlapping subsets I_j, $j = 1, \ldots, J$ such that $\|\boldsymbol{\varphi}_f(\boldsymbol{x}) - \boldsymbol{\varphi}_f(\boldsymbol{y})\| \leq K\|\boldsymbol{x} - \boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y}$ in I_j, $j = 1, \ldots, J$.

Note that Assumption (A2) implies the existence and finiteness of the matrix

$$\boldsymbol{\mathcal{I}}(f) = \int \boldsymbol{\varphi}_f(\boldsymbol{x}) \boldsymbol{\varphi}'_f(\boldsymbol{x}) f(\boldsymbol{x}) d\mu$$

appearing in Proposition 2.1 below.

Let $\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta})$ denote the residuals computed from the initial values $\boldsymbol{\epsilon}_{-q+1}, \ldots, \boldsymbol{\epsilon}_0$ and $\boldsymbol{X}_{-p+1}, \ldots, \boldsymbol{X}_0$, the parameter value $\boldsymbol{\theta}$, and the observations; those residuals can be computed recusively, or from (A.1). Clearly, $\boldsymbol{X}^{(n)} := \{\boldsymbol{X}_1^{(n)}, \ldots, \boldsymbol{X}_n^{(n)}\}$ is the finite realization of a solution of (2.1) with parameter value $\boldsymbol{\theta}$ iff $\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta})$ and $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ coincide. Denoting by $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ the distribution of $\boldsymbol{X}^{(n)}$ under parameter value $\boldsymbol{\theta}$ and innovation density f, the residuals $\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \ldots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta})$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ are i.i.d. with density f.

2.2 LAN

We are now ready to establish the local asymptotic normality, under very general conditions, of the VARMA model (2.1). Write $L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta};f}$ for the log-likelihood ratio of $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)};f}$ with respect to $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, where $\boldsymbol{\tau}^{(n)}$ is a bounded sequence of $\mathbb{R}^{(p+q)d^2}$. The following LAN result will be used to motivate the definition of our R-estimator and to establish the asymptotic normality of our R-estimators. Note that it does not require f to be elliptic.

Let

$$\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta}) := \boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\boldsymbol{Q}^{(n)\prime}_{\boldsymbol{\theta}}\boldsymbol{\Gamma}^{(n)}_f(\boldsymbol{\theta}), \tag{2.3}$$

where $\boldsymbol{M}_{\boldsymbol{\theta}}$, $\boldsymbol{P}_{\boldsymbol{\theta}}$, and $\boldsymbol{Q}^{(n)}_{\boldsymbol{\theta}}$ are given in (A.2) and (A.3) of Appendix A and

$$\boldsymbol{\Gamma}^{(n)}_f(\boldsymbol{\theta}) := ((n-1)^{1/2}(\mathrm{vec}\boldsymbol{\Gamma}^{(n)}_{1,f}(\boldsymbol{\theta}))', \dots, (n-i)^{1/2}(\mathrm{vec}\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta}))', \dots, (\mathrm{vec}\boldsymbol{\Gamma}^{(n)}_{n-1,f}(\boldsymbol{\theta}))')', \tag{2.4}$$

with the so-called *f-cross-covariance matrices*

$$\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta}) := (n-i)^{-1}\sum_{t=i+1}^{n}\boldsymbol{\varphi}_f(\boldsymbol{Z}^{(n)}_t(\boldsymbol{\theta}))\boldsymbol{Z}^{(n)\prime}_{t-i}(\boldsymbol{\theta}) \tag{2.5}$$

($\boldsymbol{\varphi}_f$ as in (2.2)). Recall that under Assumption (A1), the Green matrices $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$ (see Appendix A) decrease exponentially in u: we can thus safely define

$$\boldsymbol{\Lambda}_f(\boldsymbol{\theta}) := \boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\lim_{n\to\infty}\left\{\boldsymbol{Q}^{(n)\prime}_{\boldsymbol{\theta}}[\boldsymbol{I}_{n-1}\otimes(\boldsymbol{\Xi}\otimes\boldsymbol{\mathcal{I}}(f))]\boldsymbol{Q}^{(n)}_{\boldsymbol{\theta}}\right\}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}. \tag{2.6}$$

The following LAN result then is essentially the same as in Garel and Hallin (1995, (LAN 2) in their Proposition 3.1)

Proposition 2.1. *Let Assumptions (A1) and (A2) hold. Then, for any bounded sequence $\boldsymbol{\tau}^{(n)}$ in $\mathbb{R}^{(p+q)d^2}$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, as $n\to\infty$,*

$$L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta};f} = \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1), \tag{2.7}$$

and

$$\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta}) \to \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_f(\boldsymbol{\theta})).$$

This proposition follows from the LAN result in Garel and Hallin (1995, (LAN 2) in their Proposition 3.1) and, moving along the same lines as in the proof of Proposition 1 in Hallin and Paindaveine (2004) (note that there is no need there for the assumption of an elliptic f), we obtain the form (2.3) of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$. The form (2.6) and the finiteness of the asymptotic covariance matrix $\boldsymbol{\Lambda}_f(\boldsymbol{\theta})$ easily follow from applying Lemma 4.12 in Garel and Hallin (1995). Details are left to the reader.

2.3 Elliptical LAN

The class of densities $\mathcal{F}_d$, of course, contains all elliptical distributions with nonvanishing radial densities. Let $\mathbb{S}_d$ and $\mathcal{S}_{d-1}$ stand for the (open) unit ball and the unit sphere in $\mathbb{R}^d$, respectively. Define the (spherical) uniform measure U_d over $\mathbb{S}_d$ as the product of the uniform measure over $\mathcal{S}_{d-1}$ with a uniform measure over the unit interval of distances to the origin. Consider a symmetric and positive definite $d \times d$ matrix $\boldsymbol{\Sigma}^{1/2}$ and a univariate distribution function F_{rad} over $\mathbb{R}^+$. A centered d-dimensional random vector $\boldsymbol{Z}$ has elliptical distribution with *scatter matrix* $\boldsymbol{\Sigma}$ and radial distribution F_{rad} (*radial density* f_{rad}) iff $F_{\mathrm{rad}}((\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z})^{1/2})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Z}/(\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z})^{1/2} \sim \mathrm{U}_d$. In the terminology of measure transportation, the mapping $\boldsymbol{F}_{\mathrm{ell}}$ from $\mathbb{R}^d$ to $\mathbb{S}_d$ defined as

$$\boldsymbol{z} \mapsto \boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{z}) := F_{\mathrm{rad}}((\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z})^{1/2})\boldsymbol{\Sigma}^{-1/2}\boldsymbol{z}/(\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z})^{1/2} \tag{2.8}$$

is thus *pushing* the elliptical distribution of $\boldsymbol{Z}$ *forward to* the uniform U_d over the unit ball $\mathbb{S}_d$.

The form of the VARMA central sequence (2.3) considerably simplifies in the case of such elliptical innovation densities. Letting $f \in \mathcal{F}_d$ be elliptical with scatter matrix $\boldsymbol{\Sigma}$ and radial density f_{rad}, put $\boldsymbol{S}_{\boldsymbol{\Sigma},t}^{(n)} := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Z}_t^{(n)}/(\boldsymbol{Z}_t^{(n)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_t^{(n)})^{1/2}$: it follows from (Hallin and Paindaveine 2004) that the correponding f-cross-covariances are of the form

$$\boldsymbol{\Gamma}_{i,f}^{(n)}(\boldsymbol{\theta}) := (n-i)^{-1}\boldsymbol{\Sigma}'^{-1/2}\sum_{t=i+1}^{n}\varphi_1((\boldsymbol{Z}_t^{(n)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_t^{(n)})^{1/2})\varphi_2((\boldsymbol{Z}_{t-i}^{(n)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_{t-i}^{(n)})^{1/2})\boldsymbol{S}_{\boldsymbol{\Sigma},t}^{(n)}\boldsymbol{S}_{\boldsymbol{\Sigma},t-i}^{(n)\prime}\boldsymbol{\Sigma}'^{1/2} \tag{2.9}$$

where $\boldsymbol{Z}_t^{(n)} = \boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$, $\varphi_1(r) := -2Df_{\mathrm{rad}}^{1/2}(r)/f_{\mathrm{rad}}^{1/2}(r)$, and $\varphi_2(r) := r$, $r \in \mathbb{R}^+$. The existence of φ_1 is a consequence of Assumption (A2), which implies the quadratic mean differentiability, with derivative $Df_{\mathrm{rad}}^{1/2}$, of $f_{\mathrm{rad}}^{1/2}$ and the finiteness of $\int_0^\infty(\varphi_1(r))^2 f_{\mathrm{rad}}(r)\mathrm{d}r$.

3 Methodology

3.1 Center-outward ranks and signs

The main novelty of the R-estimation procedures proposed in this paper is the introduction of the measure transportation-based ranks and signs proposed by Chernozhukov et al. (2017) and developed in Hallin (2017) and del Barrio et al. (2018) under the name of *center-outward ranks and signs*.

Unlike the real line, the real space in dimension $d > 1$ is not canonically ordered, and finding an appropriate definition of ranks and signs (hence quantiles) has remained an open problem for quite some time—motivating, for instance, the development of *statistical depth*. A successful theory of rank-based inference has been proposed by Hallin and Paindaveine (2002a and b, 2004), but its validity unfortunately is restricted to models based on elliptical noise. This restriction is lifted here by considering the newly defined concepts of center-outward multivariate distribution and quantile functions, ranks, and signs. These new notions hinge on measure transportation theory; in their empirical version, they are based on the idea of an optimal (in the sense of a quadratic loss function) coupling of the sample with a regular grid over the unit ball. We refer to Villani (2009) for a book-length discussion on the mathematical detail and to Hallin (2017) for a discussion in a statistical context.

Let us briefly recall some key facts about center-outward ranks and signs. Let $\mathcal{P}_d$ denote the family of all distributions P with densities in $\mathcal{F}_d$. The *center-outward distribution function* $\boldsymbol{F}_\pm$ is the unique gradient of convex function mapping $\mathbb{R}^d$ to $\mathbb{S}_d$ and pushing P forward to the spherical uniform distribution U_d over $\mathbb{S}_d$. For $\mathrm{P} \in \mathcal{P}_d$, such mapping is (Figalli 2018) a homeomorphism between $\mathbb{S}_d \setminus \{\mathbf{0}\}$ and $\mathbb{R}^d \setminus \boldsymbol{F}_\pm^{-1}(\{\mathbf{0}\})$ and (letting, with a small abuse of notation, $\boldsymbol{Q}_\pm(\mathbf{0}) := \boldsymbol{F}_\pm^{-1}(\{\mathbf{0}\})$) we can define the corresponding *center-outward quantile function* as $\boldsymbol{Q}_\pm := \boldsymbol{F}_\pm^{-1}$. For any given distribution P, $\boldsymbol{F}_\pm$ induces a (partial) ordering of $\mathbb{R}^d$ and the *center-outward median* $\boldsymbol{Q}_\pm(\mathbf{0})$ is a uniquely defined compact set of Lebesgue measure zero.

Turning to the sample, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the residuals $\boldsymbol{Z}^{(n)}(\boldsymbol{\theta}) := (\boldsymbol{Z}_1^{(n)}(\boldsymbol{\theta}), \dots, \boldsymbol{Z}_n^{(n)}(\boldsymbol{\theta}))$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ are i.i.d. with density $f \in \mathcal{F}_d$ and center-outward distribution function $\boldsymbol{F}_\pm$. For the empirical counterpart $\boldsymbol{F}_\pm^{(n)}$ of $\boldsymbol{F}_\pm$, let n factorize into $n = n_R n_S + n_0$, for $n_R, n_S, n_0 \in \mathbb{N}$ and $0 \le n_0 < \min\{n_R, n_S\}$, where $n_R \to \infty$ and $n_S \to \infty$ as $n \to \infty$, and consider a sequence of grids, where each grid consists of the intersection between an n_S-tuple $(\boldsymbol{u}_1, \dots \boldsymbol{u}_{n_S})$ of unit vectors, and the n_R-hyperspheres centered at the origin, with radii $1/(n_R{+}1), \dots, n_R/(n_R{+}1)$, along with n_0 copies of the origin. The resulting grid is such that the discrete distribution with probability masses $1/n$ at each gridpoint and probability mass n_0/n at the origin converges weakly to the uniform U_d over the ball $\mathbb{S}_d$. Then, we define $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})$, for $t = 1, \ldots, n$ as the solution of an optimal coupling problem between the observations and the grid. Specifically, the empirical center-outward distribution function is the (random) mapping

$$\boldsymbol{F}_{\pm}^{(n)} : \boldsymbol{Z}^{(n)} := (\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}) \mapsto (\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_1^{(n)}), \ldots, \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_n^{(n)})),$$

satisfying

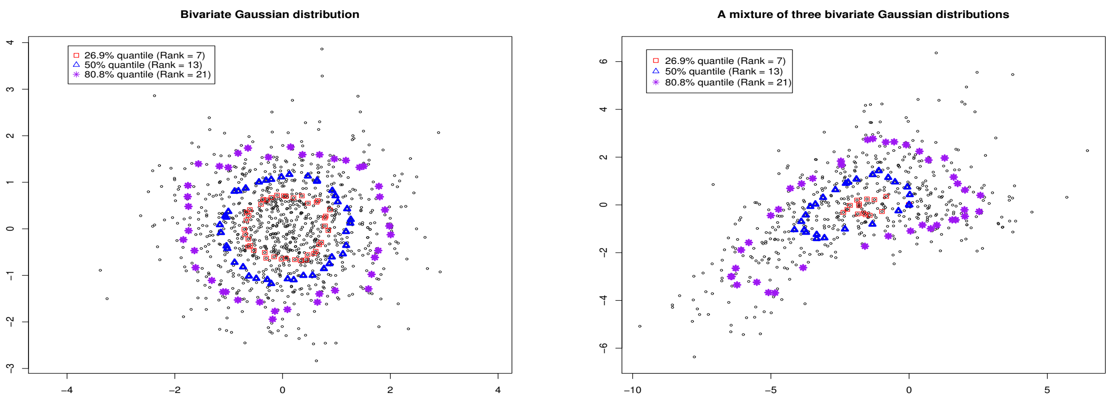
$$\sum_{t=1}^{n} \|\boldsymbol{Z}_t^{(n)} - (\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|^2 = \min_{\pi} \sum_{t=1}^{n} \|\boldsymbol{Z}_{\pi t}^{(n)} - (\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|^2, \tag{3.1}$$

(where $\boldsymbol{Z}_t^{(n)} = \boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta})$ and $\| \cdot \|$ stands for the Euclidean norm) or, equivalently,

$$\sum_{t=1}^{n} \|\boldsymbol{Z}_t^{(n)} - \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|^2 = \min_{T \in \mathcal{T}} \sum_{t=1}^{n} \|\boldsymbol{Z}_t^{(n)} - T(\boldsymbol{Z}_t^{(n)})\|^2, \tag{3.2}$$

where the set $\{\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) | t = 1, \ldots, n\}$ coincides with the n points of the grid and π in (3.1) ranges over the $n!$ possible permutations of $\{1, \ldots, n\}$, while $\mathcal{T}$ denotes the set of all possible bijective mappings between $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ and the n gridpoints.

Figure 3: Empirical center-outward quantile contours (probability contents 26.9%, 50 %, and 80%, respectively) computed from $n = 1000$ points drawn from a standard multivariate normal (left panel) and the mixture of Gaussians (1.1) (right panel).



Based on this empirical center-outward distribution function, the *center-outward ranks* are defined as

$$R_{\pm,t}^{(n)} := R_{\pm,t}^{(n)}(\boldsymbol{\theta}) := (n_R + 1)\|\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\| \tag{3.3}$$

and the *center-outward signs* as

$$\boldsymbol{S}_{\pm,t}^{(n)} := \boldsymbol{S}_{\pm,t}^{(n)}(\boldsymbol{\theta}) := \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) I[\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) \neq \boldsymbol{0}]/\|\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})\|. \tag{3.4}$$

It follows that $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)})$ factorizes into

$$\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}) = \frac{R_{\pm,t}^{(n)}}{n_R+1}\boldsymbol{S}_{\pm,t}^{(n)}, \qquad \text{hence} \qquad \boldsymbol{Z}_t^{(n)} = \boldsymbol{Q}_{\pm}^{(n)}\Big(\frac{R_{\pm,t}^{(n)}}{n_R+1}\boldsymbol{S}_{\pm,t}^{(n)}\Big). \tag{3.5}$$

Those ranks and signs are jointly distribution-free (for $f \in \mathcal{F}_d$): more precisely, under $\mathrm{P}_{\boldsymbol{\theta};f}^{(n)}$, the n-tuple $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_1^{(n)}), \ldots, \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_n^{(n)})$ is uniformly distributed over the $n!$ permutations of the n underlying gridpoints. Actually, denoting by $\mathcal{B}^{(n)}$ the σ-field generated by the sample, by $\mathcal{B}_{(.)}^{(n)}$ and $\mathcal{B}_{\pm}^{(n)}$ the sub-σ-fields generated by the order statistic and the empirical center-outward distribution function, respectively, the sample σ-field $\mathcal{B}^{(n)}$ factorizes into the product $\mathcal{B}_{(.)}^{(n)} \times \mathcal{B}_{\pm}^{(n)}$ with the remarkable property that $\mathrm{P}_{\boldsymbol{\theta};f}^{(n)}$ in turn factorizes as the product measure of its $\mathcal{B}_{(.)}^{(n)}$ and $\mathcal{B}_{\pm}^{(n)}$ marginals (see Sections 3.1 and 6 in Hallin (2017) for details).

We illustrate these definitions in Figure 3, where $n = 1000$ bivariate observations were drawn from the spherical Gaussian and from the mixture of Gaussians (1.1) considered in the introduction. Their center-outward ranks and signs were obtained via the Hungarian algorithm—see Appendix C for computational details. Figure 3 displays a few empirical center-outward quantile contours, which nicely conform to the shape of the underlying distribution.

A strong parallel exists between Mahalanobis ranks and signs and the center-outward ones: indeed, both $\boldsymbol{z} \mapsto \boldsymbol{F}_{\pm}(\boldsymbol{z})$ and $\boldsymbol{z} \mapsto \boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{z})$ are pushing P forward to the uniform U_d over the unit ball. However, $\boldsymbol{F}_{\pm}$ is the gradient of a convex function, hence, provided that P has finite moments of order two, is an optimal Monge-Kantorovich transport for quadratic transportation costs. So is $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{z} \mapsto \boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{z})$. Unless $\boldsymbol{\Sigma} = c\boldsymbol{I}_d$ for some $c > 0$, however, $\boldsymbol{z} \mapsto \boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{z})$ is not a gradient of convex function, hence is not an optimal transport. Denoting by $\widehat{\boldsymbol{\Sigma}}^{(n)}$ a consistent estimator of $\boldsymbol{\Sigma}$ measurable with respect to the order statistic of the $\boldsymbol{Z}_t$'s and by $F_{\mathrm{rad}}^{(n)}$ the empirical distribution function (with denominator $(n+1)$) of the $(\boldsymbol{Z}_t'\widehat{\boldsymbol{\Sigma}}^{(n)-1}\boldsymbol{Z}_t)^{1/2}$'s, an empirical counterpart of $\boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{Z}_t)$ is

$$\boldsymbol{F}_{\mathrm{ell}}^{(n)}(\boldsymbol{Z}_t) := F_{\mathrm{rad}}^{(n)}((\boldsymbol{Z}_t'\widehat{\boldsymbol{\Sigma}}^{(n)-1}\boldsymbol{Z}_t)^{1/2})\widehat{\boldsymbol{\Sigma}}^{(n)-1/2}\boldsymbol{Z}_t/(\boldsymbol{Z}_t'\widehat{\boldsymbol{\Sigma}}^{(n)-1}\boldsymbol{Z}_t)^{1/2} \tag{3.6}$$

with the *Mahalanobis ranks*

$$R^{(n)}_{\mathrm{ell},t} := (n+1)\|\boldsymbol{F}_{\mathrm{ell}}(\boldsymbol{Z}_t)\| = (n+1)F^{(n)}_{\mathrm{rad}}((\boldsymbol{Z}'_t\widehat{\boldsymbol{\Sigma}}^{(n)-1}\boldsymbol{Z}_t)^{1/2})$$

and *Mahalanobis signs*

$$\boldsymbol{S}^{(n)}_{\mathrm{ell},t} := \boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t)I[\boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t) \neq \mathbf{0}]/\|\boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t)\| = \widehat{\boldsymbol{\Sigma}}^{(n)-1/2}\boldsymbol{Z}_t/(\boldsymbol{Z}'_t\widehat{\boldsymbol{\Sigma}}^{(n)-1}\boldsymbol{Z}_t)^{1/2}$$

and, similar to (3.5), we have the factorization

$$\boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t) = \frac{R^{(n)}_{\mathrm{ell},t}}{n+1}\boldsymbol{S}^{(n)}_{\mathrm{ell},t}, \quad \text{hence} \quad \widehat{\boldsymbol{\Sigma}}^{(n)-1/2}\boldsymbol{Z}^{(n)}_t = F^{(n)-1}_{\mathrm{rad}}\left(\frac{R^{(n)}_{\mathrm{ell},t}}{n+1}\right)\boldsymbol{S}^{(n)}_{\mathrm{ell},t} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Z}^{(n)}_t + o_{\mathrm{P}}(1). \tag{3.7}$$

3.2 Center-outward sign- and rank-based central sequences

Intuitively, the basic idea in R-estimation consists in replacing the residuals $\boldsymbol{Z}(\boldsymbol{\theta})$ appearing in estimating equations with some adequate function of their ranks and their signs.

This, in dimension $d = 1$ and a context of asymptotic optimality, can be achieved by considering the central sequence $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$ associated with some reference density f (not necessarily the actual one, which is unknown), conditioning it on the vector of residual ranks and performing a one-step (in practice, a multistep one) Newton-Raphson iteration based on the resulting "rank-based central sequence" rather than $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$ itself. This latter strategy has been applied quite successfully in Hallin and La Vecchia (2017, 2019) in the context of nonlinear time series models; due to the classical equivalence between *exact* and *approximate score statistics* it essentially leads to substituting $F^{-1}\Big(R^{(n)}(\boldsymbol{\theta})/(n+1)\Big)$ for $\boldsymbol{Z}^{(n)}(\boldsymbol{\theta})$ in $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$, then proceeding as usual.

In dimension $d > 1$, similar ideas—namely, substituting

$$\widehat{\boldsymbol{\Sigma}}^{(n)1/2}F^{-1}_{\mathrm{rad}}(R^{(n)}_{\mathrm{ell},t}/(n+1))\boldsymbol{S}^{(n)}_{\mathrm{ell},t} = \boldsymbol{F}^{-1}_{\mathrm{ell}}((R^{(n)}_{\mathrm{ell},t}/(n+1))\boldsymbol{S}^{(n)}_{\mathrm{ell},t}) = \boldsymbol{F}^{-1}_{\mathrm{ell}}(\boldsymbol{F}^{(n)}_{\mathrm{ell}}(\boldsymbol{Z}^{(n)}_t(\boldsymbol{\theta}))) \tag{3.8}$$

for $\boldsymbol{Z}^{(n)}_t(\boldsymbol{\theta})$ in $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$—have been applied by Hallin et al. (2006) for the estimation of shape matrices based on the Mahalanobis ranks and signs.

The validity of the latter approach, however, is limited to models with elliptical noise. Here, we aim at removing that restriction by considering the center-outward counterpart of (3.8), that is, by substituting

$$\boldsymbol{F}_{\pm}^{-1}((R_{\pm,t}^{(n)}/(n_R+1))\boldsymbol{S}_{\pm,t}^{(n)}) = \boldsymbol{F}_{\pm}^{-1}(\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}))) = \boldsymbol{Q}_{\pm} \circ \boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}))$$

($\boldsymbol{F}_{\pm}$ and $\boldsymbol{Q}_{\pm}$ associated with the chosen reference density $f \in \mathcal{F}_d$—not the actual one, which remains unspecified within $\mathcal{F}_d$) for $\boldsymbol{Z}_t(\boldsymbol{\theta})$ in $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$. Except for a few particular densities such as the spherical or elliptical ones, explicit forms of center-outward distribution and quantile functions are seldom available, though, and we therefore focus, for the choice of the reference density f, on the spherical ones: let us insist again that this does not mean that the actual density, which remains unknown, has to be spherical.

In view of (2.5), writing $\boldsymbol{F}_{\pm,t}^{(n)}$, $R_{\pm,t}^{(n)}$, and $\boldsymbol{S}_{\pm,t}^{(n)}$, respectively, for $\boldsymbol{F}_{\pm}^{(n)}(\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}))$, $R_{\pm,t}^{(n)}(\boldsymbol{\theta})$, and $\boldsymbol{S}_{\pm,t}^{(n)}(\boldsymbol{\theta})$, we thus concentrate on *rank-based f-cross-covariance* statistics of the form

$$\underset{\sim}{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}) := (n-i)^{-1} \sum_{t=i+1}^{n} \varphi_1\left(\frac{R_{\pm,t}^{(n)}}{n_R+1}\right) \varphi_2\left(\frac{R_{\pm,t-i}^{(n)}}{n_R+1}\right) \boldsymbol{S}_{\pm,t}^{(n)} \boldsymbol{S}_{\pm,t-i}^{(n)\prime}, \quad i = 1, \ldots, n-1, \tag{3.9}$$

where φ_1 and $\varphi_2 : (0,1) \to \mathbb{R}$ are appropriate *score functions*; see Section 4.2 for examples. Denote by $\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ the *center-outward rank-based* central sequence resulting from substituting $\underset{\sim}{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ for $\boldsymbol{\Gamma}_{i,f}^{(n)}(\boldsymbol{\theta})$ in the elliptical central sequence $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$. Our center-outward R-estimators, which we now describe, are based on those $\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$.

4 R-estimation

4.1 One-step R-estimators

We now proceed with a precise definition of our R-estimators and establish their asymptotic properties. Throughout, φ_1 and φ_2 are assumed to satisfy the following mild assumption.

Assumption (A3). The score functions φ_1 and φ_2 are *(i)* square integrable, that is,

$$\sigma_{\varphi_l}^2 := \int_0^1 \varphi_l^2(u)\mathrm{d}u < \infty, \qquad l = 1, 2,$$

and *(ii)* continuous differences of two monotonic increasing functions (i.e., have bounded variation).

Define

$$\boldsymbol{J}_{\varphi_2,f} := \int_{\mathbb{S}_d} \varphi_2(\|\boldsymbol{u}\|)\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}\boldsymbol{F}_{\pm}^{-1\prime}(\boldsymbol{u})\mathrm{dU}_d(\boldsymbol{u}), \tag{4.1}$$

and

$$\boldsymbol{K}_{\varphi_1,\varphi_2,f} := \int_{\mathbb{S}_d} \varphi_1(\|\boldsymbol{u}\|) \left[\boldsymbol{I}_d \otimes \frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}\right] \boldsymbol{J}_{\varphi_2,f} \left[\boldsymbol{I}_d \otimes \boldsymbol{\varphi}'_f\Big(\boldsymbol{F}_{\pm}^{-1}(\boldsymbol{u})\Big)\right] \mathrm{dU}_d(\boldsymbol{u}). \tag{4.2}$$

Those two matrices under Assumptions (A2) and (A3) exist and are finite in view of the Cauchy–Schwarz inequality since $\boldsymbol{u}/\|\boldsymbol{u}\|$ is bounded.

Whether defined under one-step form (as in Hallin et al. (2006) and here) or as the solution of a minimization problem, R-estimation requires the asymptotic linearity of the rank-based objective function involved: sufficient conditions for such linearity have been established by Jurečková (1971) and van Eeden (1972) for single-output regression models, by Hallin and Puri (1994) for univariate ARMA models, by Hallin and Paindaveine (2005) for elliptical VARMA and Mahalanobis ranks and signs; see also Hallin et al. (2015) for a general criterion. The conditions we need here are regularity assumptions on φ_1 and φ_2. They are quite similar to those in Hallin and Paindaveine (2005), and so is the proof (important pieces of which actually are Lemmas B.1 and B.3 in Appendix B), which we therefore omit. We thus directly make the following assumption on the rank-based statistics $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$; the form of the linear term in the right-hand side of (4.3) follows from the form of the asymptotic shift in Lemma B.3.

Assumption (A4) For any positive integer i, as $n \to \infty$,

$$(n-i)^{1/2}\left[\mathrm{vec}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau})) - \mathrm{vec}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))\right] = -\boldsymbol{K}_{\varphi_1,\varphi_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau} + o_{\mathrm{P}}(1), \tag{4.3}$$

where $\boldsymbol{Q}_{i,\boldsymbol{\theta}}$, $\mathbf{P}_{\boldsymbol{\theta}}$ and $\mathbf{M}_{\boldsymbol{\theta}}$ are given in (B.1), (A.3), and (A.2) in Appendix.

Now, for $m \leq n-1$, let

$$\underset{\sim}{\boldsymbol{\Gamma}}^{(m,n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) := ((n-1)^{1/2}(\mathrm{vec}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{1,\varphi_1,\varphi_2}(\boldsymbol{\theta}))', \ldots, (n-m)^{1/2}(\mathrm{vec}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta}))')', \tag{4.4}$$

$$\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}} := \boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\boldsymbol{Q}^{(m+1)\prime}_{\boldsymbol{\theta}}, \quad \text{and} \quad \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta}) := \boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}\underset{\sim}{\boldsymbol{\Gamma}}^{(m,n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}); \tag{4.5}$$

clearly, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ is a version of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$ truncated after lag m. The asymptotic linearity of $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ entails, for $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$, the following result.

Proposition 4.1. *Let Assumptions (A1), (A2), (A3), and (A4) hold. Then, for any (m,n) such that $m \leq n-1$ and $m \to \infty$ (hence also $n \to \infty$),*

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta}) = -\boldsymbol{\Upsilon}^{(m+1)}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})\boldsymbol{\tau} + o_{\mathrm{P}}(1), \tag{4.6}$$

where $\boldsymbol{\Upsilon}^{(m+1)}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta}) := \boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}(\boldsymbol{I}_m \otimes \boldsymbol{K}_{\varphi_1,\varphi_2,f})\boldsymbol{T}^{(m+1)\prime}_{\boldsymbol{\theta}}$.

With the above asymptotic linearity result, we are now ready to define our R-estimators. First, let us introduce some notations. Under Assumption (A1), we can safely define the limit

$$\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta}) := \lim_{n\to\infty} \boldsymbol{\Upsilon}^{(n)}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})$$

and the *cross-information matrix*

$$\mathbf{I}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta}) := \lim_{n\to\infty} \mathrm{E}_{\boldsymbol{\theta},f}\left[\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})'\right]. \tag{4.7}$$

Let

$$\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}) := (n-i)^{-1}\sum_{t=i+1}^{n}\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i} \tag{4.8}$$

with $\boldsymbol{S}_{\pm,t} := \boldsymbol{F}_{\pm,t}/\|\boldsymbol{F}_{\pm,t}\|$ the "sign" of $\boldsymbol{F}_{\pm,t}$. Denote by $\bar{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$ the central sequence resulting from substituting $\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ for $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ in $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$. Following the proofs in Lemma B.4 and Lemma B.3, it is not difficult to see that the difference between $\bar{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ converges to zero in quadratic mean as $n\to\infty$. Therefore, in view of the proofs in Lemma B.1 and (B.19) in Appendix **??**, $\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})$ coincides with the cross-information matrix when Assumptions (A1), (A2) and (A3) hold.

Denote by $\hat{\boldsymbol{\Upsilon}}^{(n)}_{\varphi_1,\varphi_2}$ a consistent estimator of $\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})$; one way to obtain such an estimator is by using (4.6); see Appendix C for details. Also, denote by $\hat{\boldsymbol{\theta}}^{(n)}$ a root-n consistent and asymptotically discrete estimator of $\boldsymbol{\theta}$ (note that asymptotic discreteness is only a theoretical requirement since in practice $\hat{\boldsymbol{\theta}}^{(n)}$ only has a bounded number of digits; see Le Cam and Yang (2000, Chapter 6) and van der Vaart (1998, Section 5.7) for details). Then the one-step R-estimator is defined as

$$\underset{\sim}{\hat{\boldsymbol{\theta}}}_n := \hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\left(\hat{\boldsymbol{\Upsilon}}^{(n)}_{\varphi_1,\varphi_2}\right)^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\hat{\boldsymbol{\theta}}^{(n)}). \tag{4.9}$$

The following proposition then establishes the $\sqrt{n}$-consistency and asymptotic normality of $\underset{\sim}{\hat{\boldsymbol{\theta}}}_n$. We remark that asymptotic efficiency can be achieved by $\underset{\sim}{\hat{\boldsymbol{\theta}}}_n$ under spherical distributions with adequate choices of φ_1 and φ_2. More specifically, for spherical distributions, it is shown in Chernozhukov et al. (2017, Section 2.4) (see, also, Hallin (2017)) that $\boldsymbol{F}_\pm$ actually coincides with $\boldsymbol{F}_{\mathrm{ell}}$. Hence (see Section 2.3), the central sequence $\bar{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ also coincides with $\boldsymbol{\Delta}^{(n)}_f$ when

$$\varphi_1 = -2\left(Df^{1/2}_{\mathrm{rad}}/f^{1/2}_{\mathrm{rad}}\right)\circ F^{-1}_{\mathrm{rad}} \quad\text{and}\quad \varphi_2 = F^{-1}_{\mathrm{rad}}. \tag{4.10}$$

Therefore, due to the convergence of $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ to $\bar{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ in quadratic mean as mentioned above, when (4.10) holds, $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ and $\mathbf{\Delta}^{(n)}_{f}$ are asymptotically equivalent, and $\mathbf{\Upsilon}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})$ coincides with the Fisher information matrix. In this case, asymptotic efficiency can be achieved by $\underset{\sim}{\hat{\boldsymbol{\theta}}}_n$.

Proposition 4.2. *Let Assumptions (A1), (A2), (A3), and (A4) hold. Then,*

$$n^{1/2}\left(\mathbf{\Omega}^{(n)}\right)^{-1/2}(\underset{\sim}{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}) \to \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{d^2(p+q)}), \tag{4.11}$$

as both n_R and n_S tend to infinity, where

$$\mathbf{\Omega}^{(n)} := d^{-2}\sigma^2_{\varphi_1}\sigma^2_{\varphi_2}\left(\mathbf{\Upsilon}^{(n)}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{T}^{(n)}_{\boldsymbol{\theta}}\boldsymbol{T}^{(n)\prime}_{\boldsymbol{\theta}}\left(\mathbf{\Upsilon}^{(n)\prime}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})\right)^{-1}$$

and $\left(\mathbf{\Omega}^{(n)}\right)^{-1/2}$ stands for the symmetric square root of $\mathbf{\Omega}^{(n)}$.

See Appendix B for the proof. Appendices C and D provide some details on the computational aspects of the procedure and describe the algorithm we are using. Codes are available upon request.

4.2 Some examples of score functions

The rank-based central sequence $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}$ still depends on the choice, which is left to the user, of score functions φ_1 and φ_2. We provide three examples of sensible choices. The proposed scores hinge on scores widely applied in the univariate time series setting (see e.g. Hallin and La Vecchia (2019)) and in the multivariate setting under elliptical innovation density (see Hallin and Pandaveine (2004)).

Example 1 (Sign test scores). Setting $\varphi_1(u) = 1 = \varphi_2(u)$ yields the center-outward sign-based cross-covariance matrices

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\mathrm{sign}}(\boldsymbol{\theta}) = (n-i)^{-1}\sum_{t=i+1}^{n}\boldsymbol{S}^{(n)}_{\pm,t}\boldsymbol{S}^{(n)\prime}_{\pm,t-i}, \quad i = 1,\ldots,n-1. \tag{4.12}$$

The resulting central sequence $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\mathrm{sign}}(\boldsymbol{\theta})$ relies on the center-outward signs $\boldsymbol{S}^{(n)}_{\pm,t}$ and we label these scores as *sign test scores.*

Example 2 (Spearman scores). A simple choice is $\varphi_1(u) = \varphi_2(u) = u$. The corresponding rank-based cross-covariance matrices are

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\mathrm{Sp}}(\boldsymbol{\theta}) = (n-i)^{-1}\sum_{t=i+1}^{n}\boldsymbol{F}^{(n)}_{\pm,t}\boldsymbol{F}^{(n)\prime}_{\pm,t-i}, \quad i = 1,\ldots,n-1, \tag{4.13}$$

reducing, for $d = 1$, to Spearman autocorrelations, whence the terminology *Spearman rank scores*.

Example 3 (van der Waerden or normal scores). Finally, $\varphi_1(u) = \varphi_2(u) = \{\gamma_d^{-1}(u)\}^{1/2}$, where γ_d is the chi-square distribution function with d degrees of freedom, yields the *van der Waerden* (vdW) *rank scores*, with cross-covariance matrices

$$\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\mathrm{vdW}}(\boldsymbol{\theta}) = (n-i)^{-1} \sum_{t=i+1}^{n} \left[\gamma_k^{-1}\left(\frac{R^{(n)}_{\pm,t}}{n_R+1}\right)\right]^{1/2} \left[\gamma_k^{-1}\left(\frac{R^{(n)}_{\pm,t-i}}{n_R+1}\right)\right]^{1/2} \boldsymbol{S}^{(n)}_{\pm,t}\boldsymbol{S}^{(n)\prime}_{\pm,t-i}, \qquad (4.14)$$

$$i = 1, \ldots, n-1.$$

4.3 Algorithmic and computational aspects

The computational aspects of our estimation method are discussed in Appendices C and D. The corresponding codes are available on request.

5 Numerical illustration

5.1 Finite-sample performance

In this section we illustrate the finite-sample performance of our R-estimators by considering different actual innovation densities and comparing the R-estimates to the routinely-applied Gaussian QMLE. The aim of this numerical exercise is to complement and reinforce the picture glanced in section 1.2 and demonstrate the superiority of R-estimators over the QMLE under non-Gaussian densities.

To begin with, we set the bivariate ($d = 2$) VAR(1) model

$$(\boldsymbol{I}_d - \boldsymbol{A}L)\,\boldsymbol{X}_t = \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z} \qquad (5.1)$$

with parameter of interest $\boldsymbol{\theta} := \mathrm{vec}\boldsymbol{A} = (a_{11}, a_{21}, a_{12}, a_{22})'$. For $\boldsymbol{\epsilon}_t$, we consider some heavy-tailed and skew distributions: Student t_3, skew normal and skew t_3. Those densities are quite commonly used in the modelization of a number of real data in finance, economics, and biostatistics. The skew normal distribution has density ($\phi(\cdot;\boldsymbol{\Sigma})$ stands for the $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$ density, Φ for the standard normal distribution function)

$$f_{\boldsymbol{\epsilon}}(\boldsymbol{z};\boldsymbol{\xi},\boldsymbol{\Sigma},\boldsymbol{\alpha}) := 2\phi(\boldsymbol{z}-\boldsymbol{\xi};\boldsymbol{\Sigma})\Phi(\boldsymbol{\alpha}'\boldsymbol{w}^{-1}(\boldsymbol{z}-\boldsymbol{\xi})), \quad \boldsymbol{z} \in \mathbb{R}^d, \qquad (5.2)$$

where $\boldsymbol{\xi} \in \mathbb{R}^d$, $\boldsymbol{\alpha} \in \mathbb{R}^d$, and $\boldsymbol{w} = \mathrm{diag}(w_1, \ldots, w_d) > 0$ are location, shape, and scale parameters, respectively. The skew t_ν distribution has density

$$f_{\boldsymbol{\epsilon}}(\boldsymbol{z}; \boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{\alpha}, \nu) := 2\mathrm{det}(\boldsymbol{w})^{-1} t_d(\boldsymbol{x}; \boldsymbol{\Sigma}, \nu) T\left(\boldsymbol{\alpha}'\boldsymbol{x}\Big((\nu+d)/(\nu + \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})\Big)^{1/2}; \nu + d\right), \quad \boldsymbol{z} \in \mathbb{R}^d, \tag{5.3}$$

where $\boldsymbol{x} = \boldsymbol{w}^{-1}(\boldsymbol{z} - \boldsymbol{\xi})$, $T(y; \nu)$ denotes the univariate t_ν distribution function, and

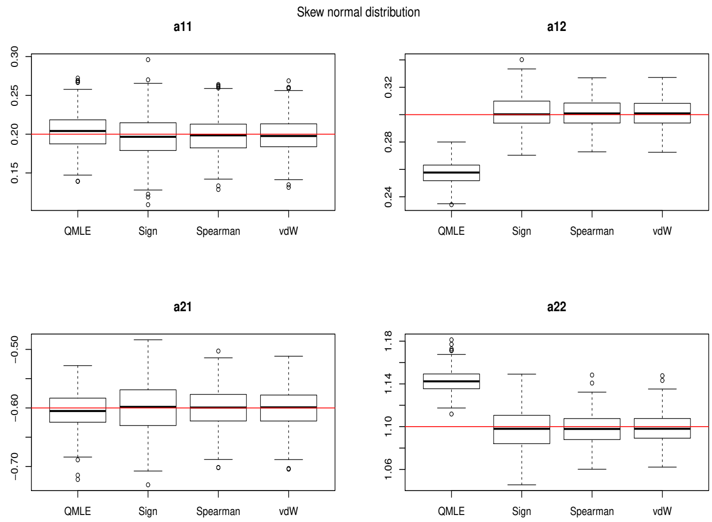
$$t_d(\boldsymbol{x}; \boldsymbol{\Sigma}, \nu) := \frac{\Gamma((\nu+d)/2)}{(\nu\pi)^{d/2}\Gamma(\nu/2)\mathrm{det}(\boldsymbol{\Sigma})^{1/2}} \left(1 + \frac{\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{\nu}\right)^{-(\nu+d)/2}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

We refer to Azzalini and Dalla Valle (1996), Azzalini and Capitanio (2003) for details.

We generate $N = 300$ Monte Carlo replications—larger values of N only show non-significant changes in the estimates—with sample size $n = 1000$ and the same parameter value $\boldsymbol{\theta} = (0.2, -0.6, 0.3, 1.1)'$ as in Section 1.2.

Throughout, the QMLEs are computed from the `MTS` package in R program, while the R-estimators are obtained using the one-step procedure as described in the online Appendix C. The simulation results are reported in Figures 4-6 under the form of boxplots. Table 1 provides a numerical summary, in terms of the first two moments of the sampling distribution of the estimators—averaged Bias ($\times 10^3$) and MSE ($\times 10^3$).

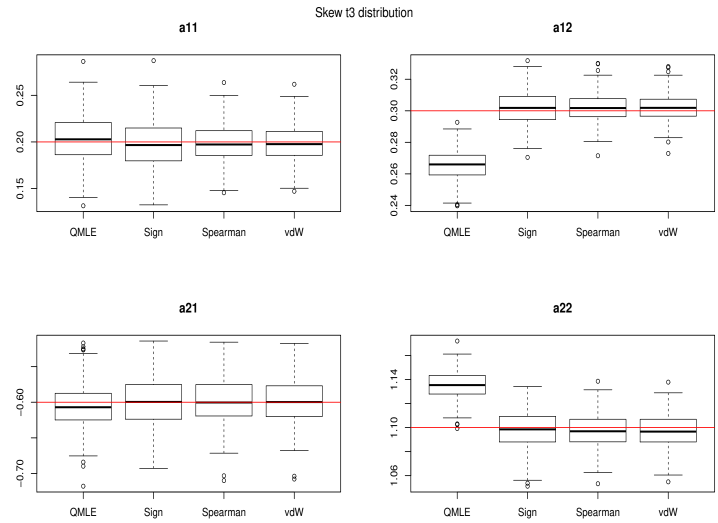
Figure 4: Boxplots of the QMLE and R-estimators (signs, Spearman, van der Waerden) under skew normal innovations (5.2); sample size $n = 1000$; $N = 300$ replications. The horizontal red line represents the actual parameter value.



First consider the case of skew innovation densities: under skew normal (Figure 4) and skew t_3 (Figure 5) distributions, the benefits of using the R-estimators rather than the

QMLE are evident. While all R-estimators unbiased, the QMLE exhibits, for a_{12} and a_{22}, quite dramatic deviations from the true parameter value, entailing large bias and large MSE.

Figure 5: Boxplots of the QMLE and R-estimators (signs, Spearman, van der Waerden) under skew t_3 innovations (5.3); sample size $n = 1000$; $N = 300$ replications. The horizontal red line represents the actual parameter value.



One may wonder, however, what happens if skewness is removed and only the heavy-tail feature is kept. Figures 1 and 6 answer this question, displaying the boxplots obtained under Gaussian and t_3 innovation densities. We see that the QMLE recovers a most reasonable performance (small bias; MSE only slightly larger than the R-estimators), even though the finite fourth moment condition is not satisfied. It does not, however, outperform our R-estimators, which keep displaying an excellent performance, with a virtually negligible bias and MSE similar to that of the QMLE. The benefits in Figures 4 and 5, thus, are not obtained at the cost of a poorer performance under Gaussian or well-behaved symmetric densities.

Finally, we consider the performance of the estimators in the presence of anomalous records—additive outliers. For univariate time series, rank-based methods have been shown resistant to outliers, while even a very small number of them can ruin the validity of correlogram-based tests and QMLEs; see e.g. Hallin and Mélard (1988) or, for a book-length presentation, Maronna et al. (2019). To investigate the robustness aspects of our R-estimators, we consider a simulation study where "clean" observations are contaminated

Figure 6: Boxplots of the QMLE and R-estimators (signs, Spearman, van der Waerden) under symmetric t_3 innovations (sample size $n = 1000$; $N = 300$ replications). The horizontal red line represents the actual parameter value.

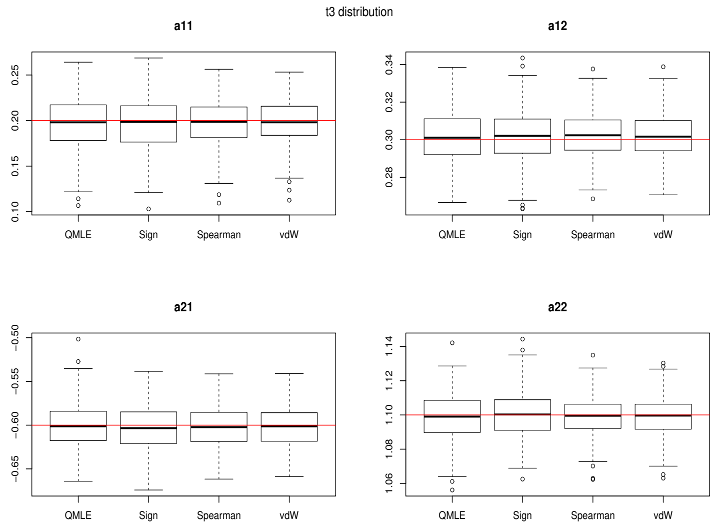
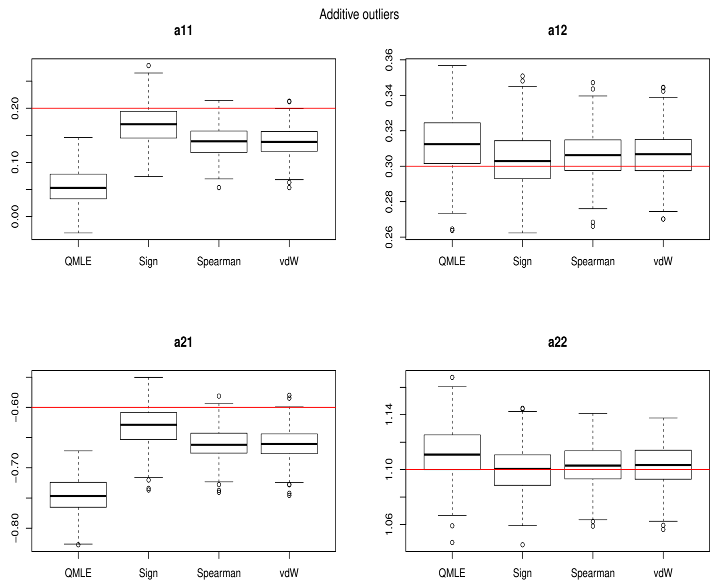


Figure 7: Boxplot of the QMLE and R-estimators (signs, van der Waerden, Spearman) under Gaussian innovations in the presence of additive outliers (sample size $n = 1000$; $N = 300$ replications). The horizontal red line represents the actual parameter value.



by some additive outliers (AO). Specifically, we first generate Gaussian VAR(1) realizations $\{\boldsymbol{X}_t\}$ of (5.1). Then, adding the outliers, we obtain a new sample, with contaminated observations $\{\boldsymbol{X}_t^* = \boldsymbol{X}_t + I(t = h)\boldsymbol{\xi}\}$ where h and $\boldsymbol{\xi}$ denote the location and size of the contamination, respectively. We set h in order to have 5% equally spaced AOs and put $\boldsymbol{\xi} = (4, 4)'$. The parameter $\boldsymbol{\theta}$ remains the same as in the previous settings and the sample size is again $n = 1000$, with $N = 300$ Monte Carlo replications. In Figure 7, we compare the three R-estimators (sign, Spearman, vdW) with the QMLE by displaying the resulting boxplots. Note that the additive outliers have a large impact on the QMLE, while their impact on the R-estimators, especially the sign-based one, is much less.

5.2 Empirical examples

We illustrate the applicability and good performance of our R-estimators in two real data problems, originating in electroencephalograms analysis and macroeconomics, respectively.

5.2.1 Electroencephalogram data

We consider two Electroencephalogram (EEG) signals from a dataset freely available in the `eegkitdata` package of the R program. That dataset contains multichannel signals from several patients; we selected the channels FP1 and FP2 of the first patient. The channels convey information on the brain reaction to the visual inputs. For each channel, the sample size is $n = 256$.

In Figure 8, we display the resulting demeaned time series. The plots indicate that the two signals tend to co-move over time. Moreover, we notice that the FP2 signal oscillates more than the FP1 one, exhibiting more extreme values, either above or below zero. The study of co-movement among EEG signals recorded at different brain locations represents an active research area. The statistical analysis of these co-movements may shed light on the joint functioning of different parts of the human brain. In this spirit, our aim is to estimate the squared coherence (related to the cross-spectrum) of those co-movements. To obtain a flexible semiparametric estimate of the cross-spectrum, we follow the widely-applied minimum entropy criterion. First, we look for the best VAR approximation to the dynamics of the bivariate (FP1, FP2) series. Selecting the order via AIC, we obtain a VAR(6). Second, we estimate the model parameters via QMLE, routinely-applied in biostatistical softwares, and via our R-estimates with the robust estimator of Croux and Joossens (2008) (R routine `varxfit` available in the package `rgarch`) as a preliminary using the algorithm described in the online Appendix C, where we refer to for details. The QMLE-based multivariate

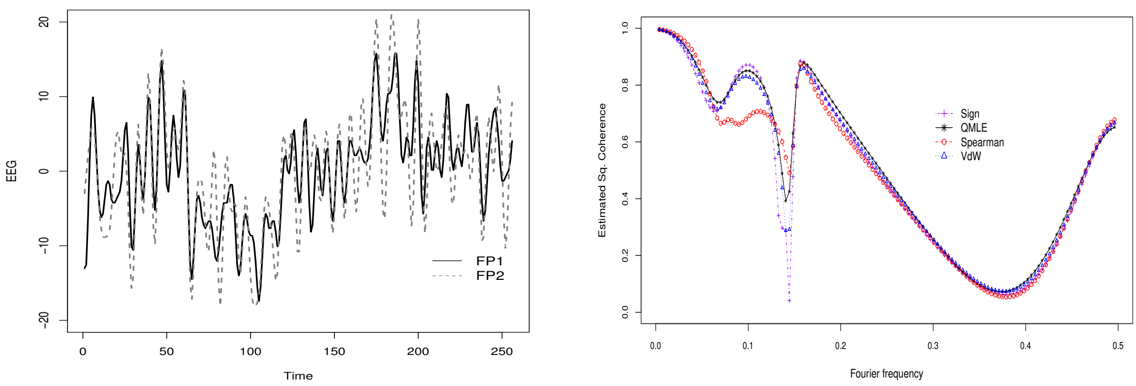
Table 1: The estimated bias ($\times 10^3$) and MSE ($\times 10^3$) for the QMLE and R-estimators (signs, van der Waerden, Spearman) under various innovation densities (normal, t_3, skew normal, skew t_3, normal mixture as in (1.1)) and with additive outliers. The sample size is $n = 1000$; $N = 300$ replications.

	Bias ($\times 10^3$)				MSE ($\times 10^3$)			
	a_{11}	a_{21}	a_{12}	a_{22}	a_{11}	a_{21}	a_{12}	a_{22}
(Normal)								
QMLE	-0.484	-0.054	0.201	-1.571	0.769	0.679	0.173	0.195
vdW	-0.662	-0.434	0.504	-1.833	0.780	0.688	0.178	0.205
Sign	-0.372	-0.600	1.545	-2.642	1.314	1.141	0.305	0.310
Spearman	-1.263	-0.979	1.274	-2.134	0.810	0.728	0.189	0.216
(t_3)								
QMLE	-3.558	-0.210	2.092	-0.967	0.844	0.671	0.205	0.185
vdW	-2.680	-1.937	2.393	-1.053	0.602	0.557	0.143	0.135
Sign	-2.204	-3.916	1.996	0.104	0.784	0.681	0.201	0.179
Spearman	-2.880	-2.014	2.663	-1.033	0.640	0.589	0.150	0.142
(Skew normal)								
QMLE	4.438	-5.728	-42.601	42.700	0.612	1.026	1.879	1.939
vdW	-1.740	-0.768	1.186	-1.384	0.590	1.146	0.103	0.219
Sign	-3.033	-0.108	1.499	-1.849	0.857	1.917	0.167	0.400
Spearman	-1.761	-0.452	1.134	-1.682	0.600	1.184	0.104	0.224
(Skew t_3)								
QMLE	4.116	-4.407	-34.612	34.898	0.634	0.965	1.280	1.349
vdW	-1.789	1.533	1.937	-2.854	0.378	0.876	0.081	0.192
Sign	-2.347	-0.039	1.858	-1.879	0.655	1.185	0.124	0.241
Spearman	-1.702	1.603	1.848	-2.968	0.388	0.907	0.084	0.198
(Mixture)								
QMLE	73.967	-1.157	-129.089	-0.472	6.137	0.399	16.940	0.054
vdW	2.600	-3.905	-0.612	1.429	0.389	0.239	0.220	0.227
Sign	4.419	-8.879	-1.815	13.927	0.713	0.437	0.622	1.168
Spearman	3.086	-4.301	-2.275	2.346	0.400	0.277	0.252	0.249
(Additive outliers)								
QMLE	-146.702	-145.957	13.106	12.255	22.584	22.221	0.443	0.509
vdW	-62.338	-61.643	6.419	3.374	4.718	4.533	0.225	0.236
Sign	-30.489	-31.571	4.390	0.172	2.169	2.262	0.297	0.295
Spearman	-62.151	-61.167	6.394	3.203	4.686	4.466	0.224	0.237

Ljung-Box test (we use the mq function in the MTS package; see Section 2.7.2 of Tsay (2014) for details) does not reject the model at nominal level 1%.

Table 2 in the online Appendix E reports the estimates along with (in parentheses) their standard errors (SEs). While some differences can be noticed (see, for instance, the first row of $\boldsymbol{A}_3$), spotting the significant ones and interpreting them is not easy. In order to do so, we plugged the various estimators into the squared coherence of the series; see Chapter 11 of Brockwell and Davis (2006) for an interpretation.

Figure 8: Left panel: EEG signals of the channels FP1 and FP2. Right panel: Squared coherence for EEG signals of the channels FP1 and FP2, as implied by different model parameter estimates.

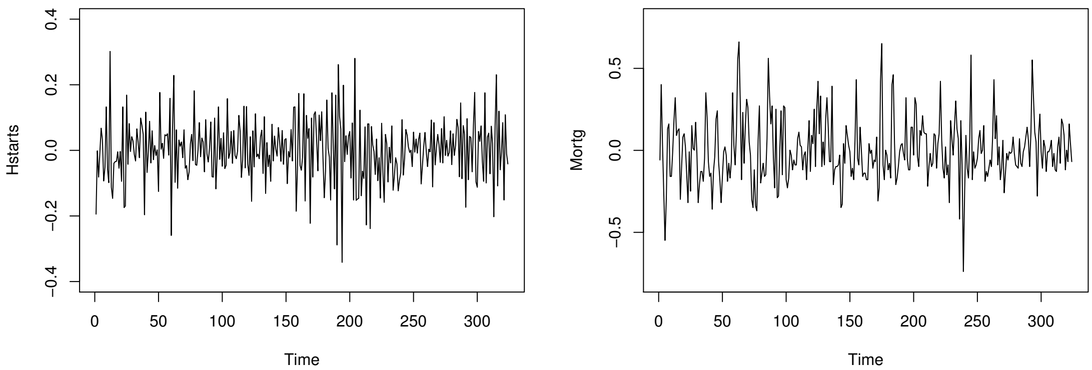


The resulting squared coherences are plotted in the right panel of Figure 8. They all reveal strong co-movements at low frequencies, with squared coherence values above 0.8. The intensity of co-movements then decreases, with a trough about 0.14 Hz where the estimated squared coherence drops below 0.3 irrespective of the estimation method. The estimated squared coherence implied by the sign R-estimator, however, exhibits a drop, followed by a sudden rise, both much sharper than the other ones. We conjecture that this difference of the sign R-estimator can be related to the presence of spiky values in FP2. Recall indeed that sign-based R-estimators are significantly more robust against outliers, hence against the extreme oscillations (see Figure 8, left plot) of FP2 which, moreover, do present some visible periodic pattern. We are thus inclined to believe that the sign-based squared coherence is more reliable than the other ones.

5.2.2 A macroeconomic application

We consider two macroeconomic time series: the seasonally adjusted monthly housing starts (Hstarts) and the 30-year conventional mortgage rate (Mortg—no need for seasonal adjustment) in the US from January 1989 to January 2016, with a sample size $n = 325$ each. Both series are freely available on the Federal Reserve Bank of Saint Louis website, where we refer to for details. Similarly to Tsay (2014, Section 3.15.2), we analyze the differenced series; Figure 9 displays plots of their demeaned differences. While the `Mortg` series seems to be driven by skew innovations (with large positive values more likely than the negative ones), the `Hstarts` series looks more symmetric about zero. Visual inspection suggests the presence of significant auto- and cross-correlations, as expected from macroeconomic theory.

Figure 9: Plots of demeaned differences of the monthly housing starts (measured in thousands of units) (left panel) and the 30-year conventional mortgage rate (in percentage) (right panel) in the US, from January 1989 through January 2016.



The AIC criterion selects a VARMA(3, 1) model, the parameters of which we estimated using the benchmark QMLE (see e.g. Tsay (2014), Chapter 3) and our R-estimators (sign, Spearman, and van der Waerden). The QMLE-based multivariate Ljung-Box test does not reject the model at nominal level 1%. We report the estimates (along with their standard error, SE, in parentheses) in Table 3 in the online Appendix E. Again, spotting the differences in Table 3 is all but simple, even though some look quite significant (see, for instance, the QMLE and R-estimates of $\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22}$) and analizing them is even more difficult.

Impulse response functions (IRFs) are easier to read, and easier to interpret. In Figures 10 and 11, we plot the estimated IRFs resulting from the QMLE and R-estimators. In accordance with macroeconometric practice, we are plotting the IRFs associated with both the original and the orthogonalized innovations—the latter in order to reduce the impact on IRFs of the off-diagonal elements of the innovation correlation matrix. Looking at the plots, we see that all IRFs decay to zero quickly; however, the QMLE-based IRFs decay uniformly faster than the R-estimator-based ones. This has interesting economic implications:

for instance, looking at the bottom-right panel of Figure 10 and Figure 11, we notice that the R-estimators, especially the sign-based ones, estimate a more persistent impact of past mortgage rates on the present one.

6 Conclusion

Starting from the LAN central sequence of VARMA models, we define a class of R-estimators based on the multivariate concept of center-outward ranks and signs recently proposed by Hallin (2017). Those R-estimators are flexible, robust, and easy to implement. They perform remarkably well both in simulations, where they significantly outperform the QMLE under non-Gaussian innovations, and in real data analysis. We conjecture that those attractive features are not limited to the VARMA case and we believe they extend to other models such as the dynamic conditional correlation model of Engle (2002), for which the QMLE is routinely applied.

Figure 10: Plots of estimated impulse response functions of the VARMA$(3, 1)$ model for the differenced `Hstarts` and `Mortg` data (QMLE and R-estimators; original innovations).

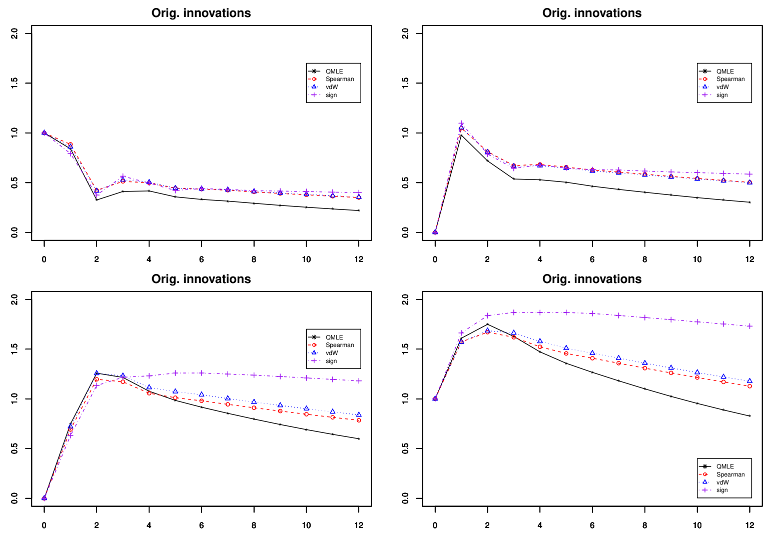
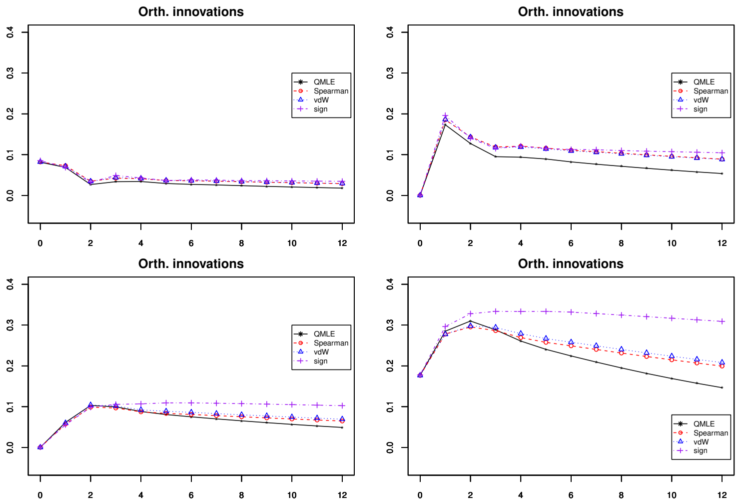


Figure 11: Plots of estimated impulse response functions of the VARMA(3, 1) model for the differenced `Hstarts` and `Mortg` data (QMLE and R-estimators; orthogonalized innovations).



References

[1] Andreou, E. and Werker, B. (2015). Residual-based rank specification tests for AR-GARCH type models. *Journal of Econometrics*, 185, 305–331.

[2] Andrews, B. (2008). Rank-based estimation for autoregressive moving average time series models. *Journal of Time Series Analysis*, 29, 51–73.

[3] Andrews, B. (2012). Rank-based estimation for GARCH processes. *Econometric Theory*, 28, 1037–1064.

[4] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B*, 65, 367–389.

[5] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726.

[6] Berndt, E., Hall, B., Hall, R., and Hausman, J. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3, 653–665.

[7] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore, MD.

[8] Brockwell, P. and Davis, R. (2006). *Time Series: Theory and Methods* (2nd edition). Springer, New York.

[9] Cassart, D., Hallin, M., and Paindaveine, D. (2010). On the estimation of cross-information quantities in R-estimation. In J. Antoch, M. Hušková and P.K. Sen, Eds: *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková*, I.M.S., 35–45.

[10] Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017). Monge- Kantorovich depth, quantiles, ranks, and signs, *Annals of Statistics*, 45, 223–256.

[11] Croux, C. and Joossens, K. (2008). Robust estimation of the vector autoregressive model by a least trimmed squares procedure. *COMPSTAT 2008*, 489-501.

[12] Dick, J. and Pillichshammer, F. (2014). Discrepancy theory and quasi-Monte Carlo integration, in W. Chen, A. Sirvastava and G. Travaglini, Eds, *A Panorama of Discrepancy Theory*, 539-620, Springer, New York.

[13] del Barrio, E., Cuesta Albertos, J., Hallin, M., and Matran, C. (2018). Center-outward distribution functions, quantiles, ranks, and signs in $\mathbb{R}^d$. Working paper, available at `https://arxiv.org/abs/1806.01238`.

[14] Engle, R. W. (2002). Dynamic Conditional Correlation. *Journal of Business & Economic Statistics*, 20:3, 339-350.

[15] Figalli, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Analysis*, 177, part B, 413-421.

[16] Garel, B. and Hallin, M. (1995). Local asymptotic normality of multivariate ARMA processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, 3, 551-579.

[17] Hájek, J. and Šidák, (1967). *Theory of Rank Tests*. New York: Academic Press.

[18] Hájek, J., Šidák, Z., and Sen, P.K. (1999). *Theory of Rank Tests*, 2nd edition. San Diego: Academic Press.

[19] Hallin, M. (1986). Non-stationary q-dependent processes and time-varying moving-average models: invertibility properties and the forecasting problem. *Advances in Applied Probability*, 18, 170-210.

[20] Hallin, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$. ECARES WP. Available at `https://ideas.repec.org/p/eca/wpaper/2013-258262.html`.

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

[22] Hallin, M. and La Vecchia, D. (2017). R-estimation in semiparametric dynamic location-scale models. *Journal of Econometrics,* 2, 233–247.

[23] Hallin, M. and La Vecchia, D. (2019). A simple R-estimation method for semiparametric duration models. *Journal of Econometrics.* Forthcoming.

[24] Hallin, M. and Mélard, G. (1988). Rank-based tests for randomness against first-order serial dependence. *Journal of the American Statistical Association,* 83, 1117–1129.

[25] Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape: II Optimal R-estimation of shape. *Annals of Statistics,* 34, 2757–2789.

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

[27] 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.

[28] Hallin, M. and Paindaveine, D. (2004). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Annals of Statistics*, 6, 2642–2678.

[29] Hallin, M. and Paindaveine, D. (2005). Asymptotic linearity of serial and nonserial multivariate signed rank statistics. *Journal of Statistical Planning and Inference*, 136, 1–32.

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

[31] Hallin, M., van den Akker, R., and Werker, B. (2015). On quadratic expansions of log-likelihoods and a general asymptotic linearity result. In M. Hallin, D. Mason,

D. Pfeifer, and J. Steinebach Eds, *Mathematical Statistics and Limit Theorems, Festschrift in Honor of Paul Deheuvels*, Springer, 147–166.

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

[33] Hodges, J. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Annals of Mathematical Statistics*, 34, 324–335.

[34] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics*, 43, 1449–1458.

[35] Judd, K.L. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge, MA.

[36] Jurečková, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Annals of Mathematical Statistics*, 40, 1889–1900.

[37] Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *Annals of Mathematical Statistics*, 42, 1328–1338.

[38] Koul, H. (1971). Asymptotic behavior of a class of confidence regions based on ranks in regression. *Annals of Mathematical Statistics*, 42, 466–476.

[39] Koul, H. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics*, 22, 540–562.

[40] Koul, H. L. and Saleh, A. M. E. (1993). R-estimation of the parameters of autoregressive AR(p) models. *Annals of Statistics*, 21, 534–551.

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

[42] Liu, R. Y. (1992). Data depth and multivariate rank tests, in Y. Dodge, Ed., *L^1 Statistics and Related Methods*. North-Holland, Amsterdam, 279–294.

[43] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests, *J. Amer. Statist. Assoc.* **88**, 257–260.

[44] Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics : Some basic concepts (2nd ed.)*. New York: Springer.

[45] Mukherjee, K. (2007). Generalized R-estimators under conditional heteroscedasticity. *Journal of Econometrics*, 141, 383–415.

[46] Maronna, R., Martin D., Yohai V., and Salibian-Barrera M. (2019), *Robust Statistics: Theory and Methods (with R)*, Wiley.

[47] Mukherjee, K. and Bai, Z. (2002). R-estimation in autoregression with square-integrable score function. *Journal of Multivariate Analysis*, 81, 167–186.

[48] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63, SIAM, Philadelphia, PA.

[49] Oja, H. (2010). *Multivariate Nonparametric Methods with R: an approach based on spatial signs and ranks.* Springer, New York.

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

[51] Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York.

[52] Terpstra, J. T., McKean, J. W., and Naranjo, J. D. (2001). GR-estimates for an autoregressive time series. *Statistics & Probability Letters*, 51, 165–172.

[53] Tsay, R.S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications.* Hoboken, New Jersey: Wiley.

[54] van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press.

[55] van Eeden, C. (1972). An analogue, for signed rank statistics, of Jurečková's asymptotic linearity theorem for rank statistics. *Annals of Mathematical Statistics*, 43, 791–802.

[56] Villani, C. (2009). *Optimal Transport: Old and New*, Springer-Verlag, Berlin.

A Technical material: some algebraic preparation

Denote by $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$, $u \in \mathbb{Z}$ the *Green's matrices* associated with the difference operators $\boldsymbol{A}(L)$ and $\boldsymbol{B}(L)$ defined in Section 2.1: those matrices are defined as the solutions of the homogeneous linear recursions

$$\boldsymbol{A}(L)\boldsymbol{G}_u = \sum_{i=0}^{p} \boldsymbol{A}_i \boldsymbol{G}_{u-i} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{B}(L)\boldsymbol{H}_u = \sum_{i=0}^{q} \boldsymbol{B}_i \boldsymbol{H}_{u-i} = \boldsymbol{0}, \quad u \in \mathbb{Z}$$

with initial values $\boldsymbol{I}_d, \boldsymbol{0}, \ldots, \boldsymbol{0}$ at $u = 0, -1, \ldots, -p+1$ and $u = 0, -1, \ldots, -q+1$, respectively.

Then, the residual process $\{\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}); 1 \le t \le n\}$ has the representation

$$\boldsymbol{Z}_t^{(n)}(\boldsymbol{\theta}) = \sum_{i=0}^{t-1} \sum_{j=0}^{p} \boldsymbol{H}_i \boldsymbol{A}_j \boldsymbol{X}_{t-i-j}^{(n)} + \begin{bmatrix} \boldsymbol{H}_{t+q-1} & \cdots & \boldsymbol{H}_t \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_d & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{B}_1 & \boldsymbol{I}_d & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{q-1} & \boldsymbol{B}_{q-2} & \cdots & \boldsymbol{I}_d \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}_{-q+1} \\ \vdots \\ \boldsymbol{\epsilon}_0 \end{bmatrix} \tag{A.1}$$

(see Hallin (1986), Garel and Hallin (1995), or Hallin and Paindaveine (2004)).

Assumption (A1) ensures the exponential decrease of $\{\|\boldsymbol{H}_u\|, u \in \mathbb{N}\}$. Specifically, there exists some $\varepsilon > 0$ such that $\|\boldsymbol{H}_u\|(1+\varepsilon)^u$ converges to 0 as $u \to \infty$. This also holds for Green matrices $\boldsymbol{G}_u$ associated with the operator $\boldsymbol{A}(L)$. It follows that the initial values $\{\boldsymbol{\epsilon}_{-q+1}, \ldots, \boldsymbol{\epsilon}_0\}$ and $\{\boldsymbol{X}_{-p+1}, \ldots, \boldsymbol{X}_0\}$ in (A.1), which are typically unobservable, have no asymptotic influence on the residuals nor any asymptotic results. Therefore, they all can safely set to zero in the sequel. This allows us to invert the AR and MA polynomials, and to define the Green matrices $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$ as the matrix coefficients of the inverted operators $(\boldsymbol{A}(L))^{-1}$ and $(\boldsymbol{B}(L))^{-1}$:

$$\sum_{u=0}^{\infty} \boldsymbol{G}_u z^u := \left(\sum_{i=0}^{p} \boldsymbol{A}_i z^i \right)^{-1} \quad \text{and} \quad \sum_{u=0}^{\infty} \boldsymbol{H}_u z^u := \left(\sum_{i=0}^{q} \boldsymbol{B}_i z^i \right)^{-1}, \quad z \in \mathbb{C}, |z| < 1.$$

Associated with an arbitrary d-dimensional linear difference operator $\boldsymbol{C}(L) := \sum_{i=0}^{\infty} \boldsymbol{C}_i L^i$ (this of course includes operators of finite order s), define, for any integers u and v, the

$d^2u \times d^2v$ matrices

$$\boldsymbol{C}^{(l)}_{u,v} := \begin{bmatrix} \boldsymbol{C}_0 \otimes \boldsymbol{I}_d & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{C}_1 \otimes \boldsymbol{I}_d & \boldsymbol{C}_0 \otimes \boldsymbol{I}_d & \dots & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{C}_{v-1} \otimes \boldsymbol{I}_d & \boldsymbol{C}_{v-2} \otimes \boldsymbol{I}_d & \dots & \boldsymbol{C}_0 \otimes \boldsymbol{I}_d \\ \vdots & & & \vdots \\ \boldsymbol{C}_{u-1} \otimes \boldsymbol{I}_d & \boldsymbol{C}_{u-2} \otimes \boldsymbol{I}_d & \dots & \boldsymbol{C}_{u-v} \otimes \boldsymbol{I}_d \end{bmatrix}$$

and

$$\boldsymbol{C}^{(r)}_{u,v} := \begin{bmatrix} \boldsymbol{I}_d \otimes \boldsymbol{C}_0 & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{I}_d \otimes \boldsymbol{C}_1 & \boldsymbol{I}_d \otimes \boldsymbol{C}_0 & \dots & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{I}_d \otimes \boldsymbol{C}_{v-1} & \boldsymbol{I}_d \otimes \boldsymbol{C}_{v-2} & \dots & \boldsymbol{I}_d \otimes \boldsymbol{C}_0 \\ \vdots & & & \vdots \\ \boldsymbol{I}_d \otimes \boldsymbol{C}_{u-1} & \boldsymbol{I}_d \otimes \boldsymbol{C}_{u-2} & \dots & \boldsymbol{I}_d \otimes \boldsymbol{C}_{u-v} \end{bmatrix}.$$

Write $\boldsymbol{C}^{(l)}_u$ for $\boldsymbol{C}^{(l)}_{u,u}$ and $\boldsymbol{C}^{(r)}_u$ for $\boldsymbol{C}^{(r)}_{u,u}$. With this notation, note that $\boldsymbol{G}^{(l)}_u, \boldsymbol{G}^{(r)}_u, \boldsymbol{H}^{(l)}_u$, and $\boldsymbol{H}^{(r)}_u$ are the inverses of $\boldsymbol{A}^{(l)}_u, \boldsymbol{A}^{(r)}_u, \boldsymbol{B}^{(l)}_u$ and $\boldsymbol{B}^{(r)}_u$, respectively. Denoting by $\boldsymbol{C}'^{(l)}_{u,v}$ and $\boldsymbol{C}'^{(r)}_{u,v}$ the matrices associated with the transposed operator $\boldsymbol{C}'(L) := \sum_{i=0}^{\infty} \boldsymbol{C}'_i L^i$, we also have that $\boldsymbol{G}'^{(l)}_u = (\boldsymbol{A}'^{(l)}_u)^{-1}$, $\boldsymbol{H}'^{(l)}_u = (\boldsymbol{B}'^{(l)}_u)^{-1}$, and so on. Define the $d^2(p+q) \times d^2(p+q)$ matrix

$$\boldsymbol{M}_{\boldsymbol{\theta}} := (\boldsymbol{G}'^{(l)}_{p+q,p} \vdots \boldsymbol{H}'^{(l)}_{p+q,q}) : \tag{A.2}$$

under Assumption (A1), $\boldsymbol{M}_{\boldsymbol{\theta}}$ is of full rank.

Also, consider the operator $\boldsymbol{D}(L) := \boldsymbol{I}_d + \sum_{i=1}^{p+q} \boldsymbol{D}_i L^i$ (note that $\boldsymbol{D}(L)$ and most quantities defined below depends on $\boldsymbol{\theta}$; for simplicity, however, we are dropping this reference to $\boldsymbol{\theta}$), where

$$\begin{bmatrix} \boldsymbol{D}'_1 \\ \vdots \\ \boldsymbol{D}'_{p+q} \end{bmatrix} := - \begin{bmatrix} \boldsymbol{G}_q & \boldsymbol{G}_{q-1} & \dots & \boldsymbol{G}_{-p+1} \\ \boldsymbol{G}_{q+1} & \boldsymbol{G}_q & \dots & \boldsymbol{G}_{-p+2} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{G}_{p+q-1} & \boldsymbol{G}_{p+q-2} & \dots & \boldsymbol{G}_0 \\ \boldsymbol{H}_p & \boldsymbol{H}_{p-1} & \dots & \boldsymbol{H}_{-q+1} \\ \boldsymbol{H}_{p+1} & \boldsymbol{H}_p & \dots & \boldsymbol{H}_{-q+2} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{H}_{p+q-1} & \boldsymbol{H}_{p+q-2} & \dots & \boldsymbol{H}_0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{G}_{q+1} \\ \vdots \\ \boldsymbol{G}_{p+q} \\ \boldsymbol{H}_{p+1} \\ \vdots \\ \boldsymbol{H}_{p+q} \end{bmatrix}$$

(recall that $\boldsymbol{G}_{-1} = \boldsymbol{G}_{-2} = \cdots = \boldsymbol{G}_{-p+1} = \boldsymbol{0} = \boldsymbol{H}_{-1} = \boldsymbol{H}_{-2} = \cdots = \boldsymbol{H}_{-q+1}$).

Let $\{\boldsymbol{\psi}_t^{(1)}, \ldots, \boldsymbol{\psi}_t^{(p+q)}\}$ be a set of $d \times d$ matrices forming a fundamental system of solutions of the homogeneous linear difference equation associated with $\boldsymbol{D}(L)$. Such a system can be obtained from the Green matrices of the operator $\boldsymbol{D}(L)$ (see, e.g., Hallin 1986). Defining

$$\bar{\boldsymbol{\psi}}_m(\boldsymbol{\theta}) := \begin{bmatrix} \boldsymbol{\psi}_1^{(1)} & \ldots & \boldsymbol{\psi}_1^{(p+q)} \\ \boldsymbol{\psi}_2^{(1)} & \ldots & \boldsymbol{\psi}_2^{(p+q)} \\ \vdots & & \vdots \\ \boldsymbol{\psi}_m^{(1)} & \ldots & \boldsymbol{\psi}_m^{(p+q)} \end{bmatrix} \otimes \boldsymbol{I}_d,$$

the Casorati matrix $\mathbf{C}_{\boldsymbol{\psi}}$ associated with $\boldsymbol{D}(L)$ is $\bar{\boldsymbol{\psi}}_{p+q}$. Finally, let

$$\boldsymbol{P}_{\boldsymbol{\theta}} := \mathbf{C}_{\boldsymbol{\psi}}^{-1} \quad \text{and} \quad \boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)} := \boldsymbol{H}_{n-1}^{(r)} \boldsymbol{B}_{n-1}^{\prime(l)} \bar{\boldsymbol{\psi}}_{n-1}. \tag{A.3}$$

B Proofs of Propositions 4.1 and 4.2

In order to prove Propositions 4.1 and 4.2, we first need to establish the asymptotic normality, under $\mathrm{P}_{\boldsymbol{\theta};f}^{(n)}$ and $\mathrm{P}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}^{(n)}$, of the rank-based $\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$.

As in the univariate case, however, due to the fact that the ranks are not mutually independent, the asymptotic normality of a rank statistic does not follow from classical central-limit theorems. The approach we are adopting here is inspired from Hájek, and consists in establishing an asymptotic representation result for the rank-based statistic under study—namely, its asymptotic equivalence with a random variable which is no longer rank-based—then proving the asymptotic normality of the latter. This is achieved here in a series of lemmas: Lemma B.1 deals with the asymptotic normality of $(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}))$, a corollary of which is the asymptotic normality of the truncated versions $\bar{\boldsymbol{\Delta}}_{m,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ of $\bar{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$; Lemma B.3 provides the asymptotic representation of $\mathrm{vec}(\underset{\sim}{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}))$ by $\mathrm{vec}(\bar{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}))$; the asymptotic representation of $\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ by $\bar{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ and their asymptotic normality are obtained in Lemma B.4. The proofs of Propositions 4.1 and 4.2 follow.

Let us start with the asymptotic normality of $(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}))$. Considering the matrix $\boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)}$ defined in (A.3), decompose it into $d^2 \times d^2(p+q)$ blocks (note that those blocks do not depend on n), write

$$\boldsymbol{Q}_{\boldsymbol{\theta}}^{(n)} = \left(\boldsymbol{Q}_{1,\boldsymbol{\theta}}^{\prime} \cdots \boldsymbol{Q}_{n-1,\boldsymbol{\theta}}^{\prime}\right)^{\prime}. \tag{B.1}$$

Lemma B.1. *Let Assumptions (A1), (A2), and (A3) hold. Then, for any positive integer i,* *the vector* $(n-i)^{1/2}\text{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))$ *in (4.8) is asymptotically normal with mean* $\mathbf{0}$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, *mean* $\boldsymbol{K}_{\varphi_1,\varphi_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$, *and covariance* $d^{-2}\sigma^2_{\varphi_1}\sigma^2_{\varphi_2}\boldsymbol{I}_{d^2}$ *under both.*

Proof. Since $L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\theta};f} = \boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau} + o_{\mathrm{P}}(1)$, the joint asymptotic normality of $(n-i)^{1/2}\text{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))$ and $L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\theta};f}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ follows, via the classical Wold-Cramér argument, from the asymptotic normality of

$$N^{(n)}_{\boldsymbol{\alpha},\beta} := (n-i)^{1/2}\boldsymbol{\alpha}'\text{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})) + \beta\boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$$

for arbitrary $\boldsymbol{\alpha} \in \mathbb{R}^{d^2}$ and $\beta \in \mathbb{R}$.

Since $\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n$ are i.i.d. and $\boldsymbol{F}_{\pm,t} := \boldsymbol{F}_{\pm}(\boldsymbol{Z}^{(n)}_t)$ is uniform over the unit ball, $N^{(n)}_{\boldsymbol{\alpha},\beta}$ is a sum of martingale differences. If it is uniformly square-integrable, with finite variance $C^{(n)}_{\boldsymbol{\alpha},\beta}$, say, such that $\lim_{n\to\infty} C^{(n)}_{\boldsymbol{\alpha},\beta} =: C_{\boldsymbol{\alpha},\beta}$ exists and is finite, the martingale central limit theorem applies, and $N^{(n)}_{\boldsymbol{\alpha},\beta}$ is asymptotically normal with mean 0 and variance $C_{\boldsymbol{\alpha},\beta}$.

Now, the variance of $N^{(n)}_{\boldsymbol{\alpha},\beta}$ takes the form

$$\begin{aligned} C^{(n)}_{\boldsymbol{\alpha},\beta} = {} & (n-i)\boldsymbol{\alpha}'\text{Var}\Big(\text{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))\Big)\boldsymbol{\alpha} \\ & + 2\beta\boldsymbol{\alpha}'(n-i)^{1/2}\text{Cov}\Big(\text{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})), \boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})\Big) \\ & + \beta^2\boldsymbol{\tau}'\text{Var}\Big(\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})\Big)\boldsymbol{\tau}. \end{aligned}$$

The entries of each $\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ are uniformly square-integrable. As for $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$, it follows from Lemma 2.2 in Hallin and Werker (2003) that, for any LAN family, a uniformly pth-order integrable version of the central sequence exists: without loss of generality, let us assume that $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$, for $p = 2$, is one of them. The sequence $N^{(n)}_{\boldsymbol{\alpha},\beta}$ thus has a limiting $\mathcal{N}(0, C_{\boldsymbol{\alpha},\beta})$ distribution provided that $\lim_{n\to\infty} C^{(n)}_{\boldsymbol{\alpha},\beta} =: C_{\boldsymbol{\alpha},\beta}$ exists and is finite.

Due to the independence between the signs $\boldsymbol{S}_{\pm,t} := \boldsymbol{F}_{\pm,t}/\|\boldsymbol{F}_{\pm,t}\|$ and the moduli $\|\boldsymbol{F}_{\pm,t}\|$ (which follows from the fact that $\boldsymbol{F}_{\pm,t} \sim \mathrm{U}_d$), and due to the fact that $\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n$ are i.i.d.,

$$\begin{aligned} \lim_{n\to\infty} \text{Var}\Big(\text{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))\Big) &= \lim_{n\to\infty} \mathrm{E}\left\{(n-i)\text{vec}\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})(\text{vec}\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))'\right\} \\ &= \lim_{n\to\infty}(n-i)^{-1}\mathrm{E}\left\{\left[\sum_{t=i+1}^{n}\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\text{vec}(\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i})\right]\right. \\ &\qquad \left. \times \left[\sum_{t=i+1}^{n}\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\text{vec}(\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i})\right]'\right\} \\ &= \frac{1}{d^2}\sigma^2_{\varphi_1}\sigma^2_{\varphi_2}\boldsymbol{I}_{d^2}, \end{aligned} \qquad \text{(B.2)}$$

where the last equation follows from the uniform distribution of $\boldsymbol{S}_{\pm,t}$ over $\mathcal{S}_{d-1}$. Next, the uniform square-integrability of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ and its asymptotic normality in Proposition 2.1 yield

$$\begin{aligned}&\lim_{n\to\infty}(n-i)^{1/2}\mathrm{Cov}\Big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})),\boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})\Big)\\&=\lim_{n\to\infty}\mathrm{E}\left[(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))\boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})\right]\\&=\lim_{n\to\infty}\mathrm{E}\left[(n-i)^{1/2}\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))\boldsymbol{\Gamma}_f^{(n)\prime}(\boldsymbol{\theta})\right]\boldsymbol{Q}^{(n)}_{\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau},\end{aligned}\tag{B.3}$$

where the last equality follows from (2.3). Due to the independence of $\boldsymbol{Z}_i^{(n)}$ and $\boldsymbol{Z}_j^{(n)}$ for $i\neq j$, only $\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta})$ in $\boldsymbol{\Gamma}^{(n)}_{f}(\boldsymbol{\theta})$ is contributing to (B.3). Therefore, using the block matrix form of $\boldsymbol{Q}^{(n)}_{\boldsymbol{\theta}}$ in (B.1), (B.3) reduces to

$$\lim_{n\to\infty}(n-i)\mathrm{E}\left[\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))(\mathrm{vec}(\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta})))'\right]\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}.\tag{B.4}$$

From (2.5), we have

$$\begin{aligned}&(n-i)\mathrm{E}\left[\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))(\mathrm{vec}(\boldsymbol{\Gamma}^{(n)}_{i,f}(\boldsymbol{\theta})))'\right]\\&\quad=(n-i)^{-1}\mathrm{E}\left\{\left[\sum_{t=i+1}^{n}\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\mathrm{vec}(\boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i})\right]\left[\sum_{t=i+1}^{n}\mathrm{vec}(\boldsymbol{\varphi}_f(\boldsymbol{Z}_t^{(n)}))\boldsymbol{Z}'_{t-i})\right]'\right\}\\&\quad=\mathrm{E}\Big[\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)(\boldsymbol{I}_d\otimes\boldsymbol{S}_{\pm,t})\boldsymbol{S}_{\pm,t-i}\boldsymbol{Z}'_{t-i}(\boldsymbol{I}_d\otimes\boldsymbol{\varphi}'_f(\boldsymbol{Z}_t^{(n)}))\Big]\end{aligned}\tag{B.5}$$

where the last two equalities follow from the independence of $\boldsymbol{Z}_1^{(n)},\ldots,\boldsymbol{Z}_n^{(n)}$ and the uniform distribution of $\boldsymbol{F}_{\pm,t}\sim\mathrm{U}_d$. In view of (2.7), (B.3), (B.4) and (B.5), we thus obtain

$$\lim_{n\to\infty}(n-i)^{1/2}\mathrm{Cov}\Big(\mathrm{vec}(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})),\boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})\Big)=\boldsymbol{K}_{\varphi_1,\varphi_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}.\tag{B.6}$$

Combining (B.2), (B.6) and the asymptotic normality of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ in Proposition 2.1 yields, for arbitrary $\boldsymbol{\alpha}$ and β,

$$\lim_{n\to\infty}C^{(n)}_{\boldsymbol{\alpha},\beta}=\boldsymbol{\alpha}'\boldsymbol{\alpha}d^{-2}\sigma^2_{\varphi_1}\sigma^2_{\varphi_2}+2\beta\boldsymbol{\alpha}'\boldsymbol{K}_{\varphi_1,\varphi_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau}+\beta^2\boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau}.\tag{B.7}$$

It follows that $\Big((n-i)^{1/2}\mathrm{vec}'(\bar{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})),\ L^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\theta};f}\Big)'$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, is asymptotically jointly normal, with mean $\left(\mathbf{0}', -\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau}\right)'$ and covariance

$$\begin{bmatrix} d^{-2}\sigma^2_{\varphi_1}\sigma^2_{\varphi_2}\boldsymbol{I}_{d^2} & \boldsymbol{K}_{\varphi_1,\varphi_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau} \\ (\boldsymbol{K}_{\varphi_1,\varphi_2,f}\boldsymbol{Q}_{i,\boldsymbol{\theta}}\boldsymbol{P}_{\boldsymbol{\theta}}\boldsymbol{M}_{\boldsymbol{\theta}}\boldsymbol{\tau})' & \boldsymbol{\tau}'\boldsymbol{\Lambda}_f(\boldsymbol{\theta})\boldsymbol{\tau} \end{bmatrix}. \tag{B.8}$$

The desired result then readily follows from applying Le Cam's third Lemma. □

Recall that $\boldsymbol{T}^{(n)}_{\boldsymbol{\theta}} = \boldsymbol{M}'_{\boldsymbol{\theta}}\boldsymbol{P}'_{\boldsymbol{\theta}}\boldsymbol{Q}^{(n)\prime}_{\boldsymbol{\theta}}$. For any positive integer $m \le n-1$, let

$$\bar{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta}) := \boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}\bar{\boldsymbol{\Gamma}}^{(m,n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}), \tag{B.9}$$

where

$$\bar{\boldsymbol{\Gamma}}^{(m,n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) := \left((n-1)^{1/2}(\mathrm{vec}\bar{\boldsymbol{\Gamma}}^{(n)}_{1,\varphi_1,\varphi_2}(\boldsymbol{\theta}))', \ldots, (n-m)^{1/2}(\mathrm{vec}\bar{\boldsymbol{\Gamma}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta})\right)')' :$$

clearly, $\bar{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta})$, it is the truncated version of $\bar{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$ defined in Section 4.1. The asymptotic normality of $\bar{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ follows from Lemma B.1 as a corollary.

Corollary B.1. *Let Assumptions (A1), (A2), and (A3) hold. Then, for any positive integer m, the vector $\bar{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ in (B.9) is asymptotically normal, with mean $\mathbf{0}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$, mean*

$$\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}(\boldsymbol{I}_m \otimes \boldsymbol{K}_{\varphi_1,\varphi_2,f})\boldsymbol{T}^{(m+1)\prime}_{\boldsymbol{\theta}}\boldsymbol{\tau} \tag{B.10}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$, and covariance $d^{-2}\sigma^2_{\varphi_1}\sigma^2_{\varphi_2}\boldsymbol{T}^{(m+1)}_{\boldsymbol{\theta}}\boldsymbol{T}^{(m+1)\prime}_{\boldsymbol{\theta}}$ under both.

The following auxiliary lemma, which follows along the same lines as Lemma 4 in Hallin and Paindaveine (2002) and Lemma 5 in Hallin and Paindaveine (2004), will be useful in subsequent proofs.

Lemma B.2. *Let $i \in \{1, \ldots, n-1\}$ and $t, t' \in \{i+1, \ldots, n\}$ be such that $t \neq t'$. Assume that $g : \mathbb{R}^{nd} = \mathbb{R}^d \times \cdots \times \mathbb{R}^d \to \mathbb{R}$ is even in all its arguments, and such that the expectation below exists. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$,*

$$\mathrm{E}\Big[g(\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n)(\boldsymbol{P}'_t\boldsymbol{Q}_t)(\boldsymbol{R}'_{t-i}\boldsymbol{S}_{t'-i})\Big] = 0, \tag{B.11}$$

where $\boldsymbol{P}_t, \boldsymbol{Q}_t, \boldsymbol{R}_t$ and $\boldsymbol{S}_t$ are any four random vectors among $\boldsymbol{S}^{(n)}_{\pm,t}$ and $\boldsymbol{S}^{(n)}_{\pm,t} - \boldsymbol{S}_{\pm,t}$.

The next lemma establishes an asymptotic representation result for the rank-based cross-covariance matrices $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ defined in (3.9) by showing their asymptotic equivalence with $\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ defined in (4.8). LAN implies that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$ and $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ are mutually contiguous; (B.12) therefore holds under both. This asymptotic representation in the Hájek style of a center-outward serial rank statistic extends to a multivariate setting a univariate result first established by Hallin et al. (1985).

Lemma B.3. *Let Assumptions (A1), (A2), and (A3) hold. Then, for any positive integer i,*

$$\mathrm{vec}\left(\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}) - \bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})\right) = o_{\mathrm{P}}(n^{-1/2}) \tag{B.12}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$ *and* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}$*, as* $n \to \infty$*.*

Proof. Note that $(n-i)^{1/2}(\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}) - \bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})) = (n-i)^{-1/2}(\boldsymbol{\delta}^{(n)}_1 + \boldsymbol{\delta}^{(n)}_2)$ where

$$\boldsymbol{\delta}^{(n)}_1 := (n-i)^{-1/2}\sum_{t=i+1}^{n}\left(\varphi_1(\frac{R^{(n)}_{\pm,t}}{n_R+1})\varphi_2(\frac{R^{(n)}_{\pm,t-i}}{n_R+1}) - \varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\right)\boldsymbol{S}^{(n)}_{\pm,t}\boldsymbol{S}^{(n)\prime}_{\pm,t-i}$$

and

$$\boldsymbol{\delta}^{(n)}_2 := (n-i)^{-1/2}\sum_{t=i+1}^{n}\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\left(\boldsymbol{S}^{(n)}_{\pm,t}\boldsymbol{S}^{(n)\prime}_{\pm,t-i} - \boldsymbol{S}_{\pm,t}\boldsymbol{S}'_{\pm,t-i}\right).$$

It suffices to show that $\mathrm{vec}(\boldsymbol{\delta}^{(n)}_1)$ and $\mathrm{vec}(\boldsymbol{\delta}^{(n)}_2)$ both converge in quadratic mean to zero as $n \to \infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta};f}$.

Let $\|\cdot\|_{L^2}$ denote the l_2-norm. For $\boldsymbol{\delta}^{(n)}_1$, we make use of Lemma B.2, and we exploit the independence of the ranks $\{R^{(n)}_{\pm,t}; t = 1,\dots,n\}$ and the signs $\{\boldsymbol{S}^{(n)}_{\pm,t}; t = 1,\dots,n\}$ (see Hallin (2017)). Given that $(\mathrm{vec}\boldsymbol{A})'(\mathrm{vec}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}'\boldsymbol{B})$, we have

$$\left\|\mathrm{vec}(\boldsymbol{\delta}^{(n)}_1)\right\|^2_{L^2} = (n-i)^{-1}\sum_{t=i+1}^{n}\mathrm{E}\left[\left(\varphi_1(\frac{R^{(n)}_{\pm,t}}{n_R+1})\varphi_2(\frac{R^{(n)}_{\pm,t-i}}{n_R+1}) - \varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|)\right)^2\right].$$

The Glivenko-Cantelli result in Hallin (2017, Proposition 5.1) entails

$$\max_{1\le t\le n}\left|R^{(n)}_{\pm,t}/(n_R+1) - \|\boldsymbol{F}_{\pm,t}\|\right| \to 0 \quad a.s. \quad \text{as} \quad n \to \infty. \tag{B.13}$$

In view of the assumptions made on φ_1 and φ_2, Lemma 6.1.6.1 of Hájek et al. (1999) yields

$$\|\mathrm{vec}(\boldsymbol{\delta}^{(n)}_1)\|^2_{L^2} \to 0 \quad \text{as } n \to \infty. \tag{B.14}$$

For $\boldsymbol{\delta}_2^{(n)}$, we have

$$\boldsymbol{\delta}_2^{(n)} = (n-i)^{-1/2} \sum_{t=i+1}^{n} \varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|) \left[\left(\boldsymbol{S}_{\pm,t}^{(n)} - \boldsymbol{S}_{\pm,t}\right) \boldsymbol{S}_{\pm,t-i}^{(n)\prime} + \boldsymbol{S}_{\pm,t} \left(\boldsymbol{S}_{\pm,t-i}^{(n)\prime} - \boldsymbol{S}'_{\pm,t-i}\right)\right].$$

Similar to the arguments used for $\boldsymbol{\delta}_1^{(n)}$, Lemma B.2 and the fact that $(\text{vec}\boldsymbol{A})'(\text{vec}\boldsymbol{B}) = \text{tr}(\boldsymbol{A}'\boldsymbol{B})$ imply

$$\|\text{vec}(\boldsymbol{\delta}_2^{(n)})\|_{L^2}^2 \leq 2(n-i)^{-1} \sum_{t=i+1}^{n} \text{E}\left[(\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|))^2 \|\boldsymbol{S}_{\pm,t}^{(n)} - \boldsymbol{S}_{\pm,t}\|^2\right] \tag{B.15}$$

$$+ 2(n-i)^{-1} \sum_{t=i+1}^{n} \text{E}\left[(\varphi_1(\|\boldsymbol{F}_{\pm,t}\|)\varphi_2(\|\boldsymbol{F}_{\pm,t-i}\|))^2 \|\boldsymbol{S}_{\pm,t-i}^{(n)} - \boldsymbol{S}_{\pm,t-i}\|^2\right]. \tag{B.16}$$

Still in view of Proposition 5.1 in Hallin (2017), $\max_{1\leq t\leq n}\|\boldsymbol{S}_{\pm,t}^{(n)} - \boldsymbol{S}_{\pm,t}\| \to 0$ a.s. as $n \to \infty$. Since φ_1, φ_2 are square integrable and $\boldsymbol{Z}_1^{(n)}, \ldots, \boldsymbol{Z}_n^{(n)}$ are independent, both (B.15) and (B.16) converge to 0. The result follows. □

We now can extend the above asymptotic representation and asymptotic normality results from the rank-based cross-covariance matrices $\underset{\sim}{\boldsymbol{\Gamma}}_{i,\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ to the rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$.

Lemma B.4. *Let Assumptions (A1), (A2), and (A3) hold. Then,*

$$\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}) = o_{\text{P}}(1) \quad \text{as } n \to \infty \tag{B.17}$$

both under $\text{P}_{\boldsymbol{\theta};f}^{(n)}$ *and* $\text{P}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}^{(n)}$. *Moreover,* $\underset{\sim}{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})$ *is asymptotically normal, with mean* $\mathbf{0}$ *under* $\text{P}_{\boldsymbol{\theta};f}^{(n)}$, *mean*

$$\lim_{n\to\infty} \left\{\boldsymbol{T}_{\boldsymbol{\theta}}^{(n)}(\boldsymbol{I}_{n-1} \otimes \boldsymbol{K}_{\varphi_1,\varphi_2,f})\boldsymbol{T}_{\boldsymbol{\theta}}^{(n)\prime}\right\}\boldsymbol{\tau} \tag{B.18}$$

under $\text{P}_{\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau};f}^{(n)}$, *and covariance* $d^{-2}\sigma_{\varphi_1}^2\sigma_{\varphi_2}^2 \lim_{n\to\infty}\left\{\boldsymbol{T}_{\boldsymbol{\theta}}^{(n)}\boldsymbol{T}_{\boldsymbol{\theta}}^{(n)\prime}\right\}$ *under both.*

Note that the limits appearing in the above asymptotic means and covariances exist due to Assumption (A1) on the characteristic roots of the VARMA operators involved.

Proof. For (B.17), due to Lemma B.3 and contiguity, it is sufficient to prove that, under $\text{P}_{\boldsymbol{\theta};f}^{(n)}$, for $m = m(n) \leq n-1$ and provided that $m(n) \to \infty$ as $n \to \infty$,

$$\limsup_{n\to\infty}\|\bar{\boldsymbol{\Delta}}_{\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta}) - \bar{\boldsymbol{\Delta}}_{m(n),\varphi_1,\varphi_2}^{(n)}(\boldsymbol{\theta})\| = o_{\text{P}}(1) \tag{B.19}$$

and

$$\limsup_{n\to\infty} \|\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) - \underset{\sim}{\mathbf{\Delta}}^{(n)}_{m(n),\varphi_1,\varphi_2}(\boldsymbol{\theta})\| = o_{\mathrm{P}}(1). \tag{B.20}$$

For $m = n-1$, the left-hand sides in (B.19) and (B.20) are exactly zero. Therefore, we only need to consider $m \leq n-2$.

It follows from Proposition 3.1 (LAN2) in Garel and Hallin (1995) that

$$\bar{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) - \bar{\mathbf{\Delta}}^{(n)}_{m(n),\varphi_1,\varphi_2}(\boldsymbol{\theta})$$
$$= \begin{bmatrix} \sum_{i=m+1}^{n-1} \sum_{j=0}^{i-1} \sum_{k=0}^{\min(q,i-j-1)} [(\boldsymbol{G}_{i-j-k-1}\boldsymbol{B}_k) \otimes \boldsymbol{H}'_j](n-i)^{1/2}(\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))) \\ \vdots \\ \sum_{i=m+1}^{n-1} \sum_{j=0}^{i-p} \sum_{k=0}^{\min(q,i-j-p)} [(\boldsymbol{G}_{i-j-k-p}\boldsymbol{B}_k) \otimes \boldsymbol{H}'_j](n-i)^{1/2}(\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))) \\ \sum_{i=m+1}^{n-1}(\boldsymbol{I}_d \otimes \boldsymbol{H}'_{i-1}(n-i)^{1/2}(\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})) \\ \vdots \\ \sum_{i=m+1}^{n-1}(\boldsymbol{I}_d \otimes \boldsymbol{H}'_{i-q}(n-i)^{1/2}(\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})) \end{bmatrix}$$

for any $p \leq m \leq n-2$,. Due to the square-integrability of φ_1, φ_2 and the fact that $\boldsymbol{Z}^{(n)}_1, \ldots, \boldsymbol{Z}^{(n)}_n$ are i.i.d., it follows from $(\mathrm{vec}\boldsymbol{A})'(\mathrm{vec}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}'\boldsymbol{B})$ that

$$\|(n-i)^{1/2}(\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})))\|^2_{L^2} = (n-i)^{-1}\sum_{t=i+1}^{n} \mathrm{E}\left[\varphi_1^2(\|\boldsymbol{F}_{\pm,t}\|)\right]\mathrm{E}\left[\varphi_2^2(\|\boldsymbol{F}_{\pm,t-i}\|)\right] = \sigma^2_{\varphi_1}\sigma^2_{\varphi_2} < \infty.$$

Recall that, under Assumption (A1), the Green matrices $\boldsymbol{G}_u$ and $\boldsymbol{H}_u$ decrease exponentially fast (see Appendix A). Using the fact that $\|\mathbf{Ax}\|_{L^2} \leq \|\mathbf{A}\|\,\|\mathbf{x}\|_{L^2}$ (where $\|\mathbf{A}\|$ denotes the operator norm of $\mathbf{A}$) and the triangular inequality, we thus obtain

$$\limsup_{n\to\infty} \|\bar{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) - \bar{\mathbf{\Delta}}^{(n)}_{m(n),\varphi_1,\varphi_2}(\boldsymbol{\theta})\|_{L^2} = 0.$$

The result (B.19) follows. Turning to (B.20), we have, in view of (B.14), (B.15) and (B.16),

$$\max_{1\leq i\leq n-1} \|(n-i)^{1/2}[\mathrm{vec}(\bar{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta})) - \mathrm{vec}(\underset{\sim}{\mathbf{\Gamma}}^{(n)}_{i,\varphi_1,\varphi_2}(\boldsymbol{\theta}))]\|^2_{L^2} = o(1)$$

as $n \to \infty$. Hence, (B.20) follows along the same lines as (B.19).

The asymptotic normality of $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$ then follows from (B.17) and the asymptotic normality of $\bar{\mathbf{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$, itself implied by (B.19) and Lemma B.1. The asymptotic mean and variance are the limits as $m = m(n)$ and n tend to infinity, of the asymptotic mean and variance of $\bar{\mathbf{\Delta}}^{(n)}_{m(n),\varphi_1,\varphi_2}(\boldsymbol{\theta})$ and do not depend on the way m grows with n. □

Proof of Proposition 4.1.

Proof. Proposition 4.1 readily follows from (B.20) and the asymptotic linearity of the truncated $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{m,\varphi_1,\varphi_2}(\boldsymbol{\theta})$ implied by Assumption (A4). □

Proof of Proposition 4.2.

Proof. From the definition of $\underset{\sim}{\hat{\boldsymbol{\theta}}}_n$ in (4.9), the asymptotic linearity in Proposition 4.1, the consistency of $\hat{\boldsymbol{\Upsilon}}^{(n)}_{\varphi_1,\varphi_2}$, the convergence of $\boldsymbol{\Upsilon}^{(n)}_{\varphi_1,\varphi_2,f}$ to $\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}$, and the asymptotic discreteness of $\hat{\boldsymbol{\theta}}^{(n)}$ (which allows us to treat $n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})$ as if it were a bounded constant: see Lemma 6.1 in Kreiss (1987)), we have

$$\begin{aligned}
n^{1/2}(\underset{\sim}{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}) &= n^{1/2}\left\{\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\left[\left(\hat{\boldsymbol{\Upsilon}}^{(n)}_{\varphi_1,\varphi_2}\right)^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\hat{\boldsymbol{\theta}}^{(n)})\right] - \boldsymbol{\theta}\right\}\\
&= n^{1/2}\left\{\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\left[\boldsymbol{\Upsilon}^{-1}_{\varphi_1,\varphi_2,f}\left(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}^{(n)}_{\varphi_1,\varphi_2,f} n^{1/2}(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})\right)\right] - \boldsymbol{\theta}\right\} + o_{\mathrm{P}}(1)\\
&= \boldsymbol{\Upsilon}^{-1}_{\varphi_1,\varphi_2,f}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1).
\end{aligned}$$

This, in view of the asymptotic normality of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\boldsymbol{\theta})$ in Lemma B.4, completes the proof of Proposition 4.2. □

C Computational aspects

In this section, we briefly discuss some computational issues related, mainly, to the measure transportation aspects of the estimation method and the one-step procedure.

Consistency requires that both n_R and n_S tend to infinity. In practice, we factorize n into $n_R n_S + n_0$ in such a way that both n_R and n_S are large—typically, n_R of order $n^{1/d}$ and n_S of order $n^{(d-1)/d}$. Generating "regular grids" of n_S points over the unit sphere $\mathcal{S}_{d-1}$ as described in Section 3.1 is easy for $d = 2$, where perfect regularity can be achieved by dividing the unit circle into n_S arcs of equal length $2\pi/n_S$. For $d \geq 3$, however, this typically becomes impossible. A random array of n_S independent and uniformly distributed unit vectors does satisfy (almost surely) the weak convergence (to U_d) requirement. More regular deterministic arrays (with faster convergence) can be constructed, though, such as the *low-discrepancy sequences* (see, e.g., Niederreiter (1992), Judd (1998), Dick and Pillichshammer (2014), or Santner et al. (2003)) considered in numerical integration and the design of computer experiments. To compute $\{\boldsymbol{F}^{(n)}_{\pm,t}; t = 1, \ldots, n\}$ via (3.2), we first create a $n \times n$ matrix, with (i, j) entry the squared Euclidean distance between $\mathbf{Z}^{(n)}_i$ and the j-th gridpoint;

the optimal coupling then can be obtained following del Barrio et al. (2018), using, for instance, a Hungarian algorithm (already included in the `clue` package of R program).

The computation of the one-step R-estimator in (4.9) involves two basic ingredients: a preliminary root n-consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ and an estimator of the cross-information matrix $\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}$. For the preliminary $\hat{\boldsymbol{\theta}}^{(n)}$, robust M-estimators such as the Multivariate Least Trimmed Square Estimator (MLTS) derived by Croux and Joossens (2008) for VAR models are obvious candidates; provided that fourth-order moments can be assumed to be finite, the QMLE still constitutes a reasonable choice, though. Different preliminary estimators may lead to different one-step R-estimators. Differences, however, gradually wane on iterating (for fixed n) the one-step procedure and the asymptotic impact (as $n \to \infty$) of the choice of $\hat{\boldsymbol{\theta}}^{(n)}$ is nil. Turning to the estimation of $\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}$, the issue is that this matrix depends on the unknown actual density f. A simple consistent estimator is obtained by letting $\boldsymbol{\tau} = \mathbf{e}_i$ in (4.6) where $\mathbf{e}_i$, $i = 1, \dots, (p+q)d^2$ denotes the ith vector of the canonical basis in the parameter space $\mathbb{R}^{(p+q)d^2}$: $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\mathbf{e}_i) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\hat{\boldsymbol{\theta}}^{(n)})$ then provides a consistent estimator of the i-th column of $-\boldsymbol{\Upsilon}_{\varphi_1,\varphi_2,f}(\boldsymbol{\theta})$. More sophisticated constructions also are possible: see Hallin et al. (2006) or Cassart et al. (2010).

D Algorithm.

We provide here a detailed description of the estimation algorithm; codes are available on request.

Step 1. Given a sample of n (demeaned) observations of the VARMA process, compute a preliminary root-n consistent estimator $\hat{\boldsymbol{\theta}}^{(n)}$ and, setting the initial values $\boldsymbol{\epsilon}_{-q+1}, \dots, \boldsymbol{\epsilon}_0$ and $\boldsymbol{X}_{-p+1}, \dots, \boldsymbol{X}_0$ all equal to zero, recursively compute residuals $\boldsymbol{Z}^{(n)}_1(\hat{\boldsymbol{\theta}}^{(n)}), \dots, \boldsymbol{Z}^{(n)}_n(\hat{\boldsymbol{\theta}}^{(n)})$ as in (A.1).

Step 2. Factorize n into $n_R n_S + n_0$ and generate, as explained in *(i)* above, a "regular grid" of $n_R n_S$ points over the unit ball $\mathbb{S}_d$.

Step 3. Create a $n \times n$ matrix $\mathbf{D}$ with (i, j) entry the squared Euclidean distance between $\mathbf{Z}^{(n)}_i$ and the j-th gridpoint. Based on that matrix, compute $\{\boldsymbol{F}^{(n)}_\pm(\boldsymbol{Z}^{(n)}_t); t = 1, \dots, n\}$ solving the optimal pairing problem in (3.2) using, e.g., the Hungarian algorithm as coded in the `clue` package of R program.

Step 4. From $\boldsymbol{F}^{(n)}_\pm$, compute the center-outward ranks (3.3), signs (3.4), and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\hat{\boldsymbol{\theta}}^{(n)})$; for some chosen $\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_{(p+q)d^2}$, compute $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\varphi_1,\varphi_2}(\hat{\boldsymbol{\theta}}^{(n)} + n^{-1/2}\boldsymbol{\tau})$, then, via (4.6), $\hat{\boldsymbol{\Upsilon}}^{(n)}_{\varphi_1,\varphi_2}$.

Step 5. Using (4.9), compute $\hat{\underset{\sim}{\boldsymbol{\theta}}}_n$ and perform the one-step iteration to update it; iterate this until numerical stabilization; in practice, four or five iterations yield a stable result.

E Tables for Section 5.2

We provide here the tables of estimated coefficients for the empirical examples of Section 5.2 that did not fit into the main paper.

Table 2: The QMLE and R-estimates of $\boldsymbol{\theta}$ in the VAR(6) fitting of the EEG data in Section 5.2.1; the standard errors are shown in parentheses.

	$\boldsymbol{A}_1$		$\boldsymbol{A}_2$		$\boldsymbol{A}_3$		$\boldsymbol{A}_4$		$\boldsymbol{A}_5$		$\boldsymbol{A}_6$	
QMLE	2.333	0.242	-2.227	-0.447	0.443	0.260	0.999	0.190	-0.899	-0.330	0.250	0.162
	(0.070)	(0.062)	(0.164)	(0.146)	(0.210)	(0.187)	(0.210)	(0.187)	(0.163)	(0.145)	(0.068)	(0.061)
	0.413	2.351	-0.713	-2.237	0.276	0.353	0.689	0.980	-0.915	-0.784	0.405	0.165
	(0.076)	(0.067)	(0.178)	(0.158)	(0.228)	(0.203)	(0.228)	(0.203)	(0.177)	(0.158)	(0.074)	(0.066)
vdW	2.303	0.250	-2.146	-0.450	0.363	0.221	1.010	0.270	-0.859	-0.397	0.235	0.173
	(0.043)	(0.032)	(0.106)	(0.052)	(0.082)	(0.073)	(0.158)	(0.105)	(0.103)	(0.245)	(0.127)	(0.147)
	0.374	2.379	-0.648	-2.285	0.258	0.362	0.634	1.036	-0.858	-0.843	0.395	0.179
	(0.048)	(0.045)	(0.122)	(0.065)	(0.143)	(0.204)	(0.213)	(0.025)	(0.104)	(0.087)	(0.107)	(0.027)
Sign	2.327	0.237	-2.173	-0.447	0.360	0.229	1.043	0.270	-0.865	-0.413	0.230	0.181
	(0.033)	(0.070)	(0.219)	(0.154)	(0.021)	(0.034)	(0.147)	(0.150)	(0.213)	(0.208)	(0.469)	(0.033)
	0.374	2.393	-0.638	-2.310	0.254	0.369	0.634	1.035	-0.873	-0.846	0.402	0.203
	(0.085)	(0.208)	(0.257)	(0.144)	(0.345)	(0.104)	(0.523)	(0.037)	(0.366)	(0.049)	(0.115)	(0.116)
Spearman	2.279	0.274	-2.139	-0.435	0.358	0.186	0.989	0.329	-0.853	-0.400	0.207	0.175
	(0.083)	(0.114)	(0.291)	(0.050)	(0.329)	(0.116)	(0.117)	(0.252)	(0.461)	(0.266)	(0.183)	(0.145)
	0.369	2.385	-0.656	-2.289	0.256	0.368	0.652	1.036	-0.887	-0.856	0.419	0.183
	(0.118)	(0.116)	(0.040)	(0.076)	(0.118)	(0.084)	(0.074)	(0.142)	(0.201)	(0.251)	(0.139)	(0.095)

Table 3: The QMLE and R-estimates of $\boldsymbol{\theta}$ in the VARMA$(3,1)$ fitting of the econometric data (demeaned differenced Hstarts and Mortg series) in Section 5.2.2; the standard errors are shown in parentheses. The datasets are demeaned changes in Hstarts and Mortg.

	$\boldsymbol{A}_1$		$\boldsymbol{A}_2$		$\boldsymbol{A}_3$		$\boldsymbol{B}_1$	
QMLE	0.137	0.487	-0.154	-0.199	0.032	0.056	-0.703	-0.490
	(0.265)	(0.353)	(0.284)	(0.130)	(0.171)	(0.072)	(0.258)	(0.350)
	0.596	0.974	0.030	-0.400	0.070	0.110	-0.152	-0.636
	(0.327)	(0.537)	(0.436)	(0.189)	(0.285)	(0.077)	(0.282)	(0.533)
vdW	0.155	0.526	-0.096	-0.181	0.017	0.038	-0.705	-0.527
	(0.141)	(0.088)	(0.122)	(0.079)	(0.133)	(0.062)	(0.088)	(0.071)
	0.561	0.943	0.094	-0.386	0.011	0.128	-0.161	-0.627
	(0.148)	(0.079)	(0.133)	(0.100)	(0.098)	(0.040)	(0.081)	(0.015)
Sign	0.087	0.536	-0.032	-0.198	0.075	-0.044	-0.705	-0.562
	(0.148)	(0.079)	(0.133)	(0.100)	(0.098)	(0.040)	(0.081)	(0.015)
	0.471	1.036	0.107	-0.403	0.035	0.148	-0.161	-0.627
	(0.178)	(0.084)	(0.165)	(0.073)	(0.138)	(0.061)	$(< 10^{-3})$	$(< 10^{-3})$
Spearman	0.180	0.511	-0.090	-0.180	0.030	0.049	-0.705	-0.537
	(0.066)	(0.033)	(0.092)	(0.046)	(0.113)	(0.049)	$(< 10^{-3})$	(0.014)
	0.531	0.946	0.072	-0.374	0.011	0.121	-0.161	-0.627
	(0.124)	(0.054)	(0.115)	(0.075)	(0.112)	(0.042)	$(< 10^{-3})$	$(< 10^{-3})$