ECARES

Consistent Distribution–Free Affine–Invariant Tests for the Validity of Independent Component Models

Marc Hallin
ECARES, and Département de Mathématique, Université libre de Bruxelles
Czech Academy of Sciences, Prague, Czech Republic

Simos G. Meintanis
Department of Economics, National and Kapodistrian University of Athens
Pure and Applied Analytics, North–West University, Potchefstroom, South Africa

Klaus Nordhausen
Department of Mathematics and Statistics, University of Jyväskylä

February 2024

ECARES working paper 2024-04

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Consistent Distribution–Free Affine–Invariant Tests for the Validity of Independent Component Models

Marc Hallin*
ECARES and Department of Mathematics, Université Libre de Bruxelles
and
Czech Academy of Sciences, Prague, Czech Republic

Simos G. Meintanis
Department of Economics, National and Kapodistrian University of Athens
and
Pure and Applied Analytics, North–West University, Potchefstroom, South Africa

and

Klaus Nordhausen
Department of Mathematics and Statistics, University of Jyväskylä

Abstract

We propose a family of tests of the validity of the assumptions underlying independent component analysis methods. The tests are formulated as L2–type procedures based on characteristic functions and involve weights; a proper choice of these weights and the estimation method for the mixing matrix yields consistent and affine-invariant tests. Due to the complexity of the asymptotic null distribution of the resulting test statistics, implementation is based on permutational and resampling strategies. This leads to distribution-free procedures regardless of whether these procedures are performed on the estimated independent components themselves or the componentwise ranks of their components. A Monte Carlo study involving various estimation methods for the mixing matrix, various weights, and a competing test based on distance covariance is conducted under the null hypothesis as well as under alternatives. A real-data application demonstrates the practical utility and effectiveness of the method.

Keywords: Characteristic function; total independence; independent component model; rank test

*Marc Hallin gratefully acknowledges the support of the Czech Science Foundation grants GAČR22036365 and GA24-10078S.

1 Introduction

1.1 Testing the validity of the independent component model

Consider the *independent component model* (ICM)

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Omega}\boldsymbol{Z} \tag{1}$$

whereby the p-dimensional random vector $\boldsymbol{X} \coloneqq (X_1, ..., X_p)^\top,\ p > 1$ is linearly associated with a latent centered random vector

$$\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{\mu}, \boldsymbol{\Omega}) \coloneqq (Z_1(\boldsymbol{\mu}, \boldsymbol{\Omega}), ..., Z_p(\boldsymbol{\mu}, \boldsymbol{\Omega}))^\top = (Z_1, ..., Z_p)^\top$$

of independent components, i.e., $\boldsymbol{Z}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ is enjoying the property of *total independence*:[1] the p-dimensional distribution of $\boldsymbol{Z}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ is the product of the (marginal) distributions of its components $Z_1(\boldsymbol{\mu}, \boldsymbol{\Omega}), \ldots, Z_p(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

Denoting by $\varphi_\ell,\ \ell = 1, ..., p$ and φ the characteristic functions (CFs) of $Z_\ell(\boldsymbol{\mu}, \boldsymbol{\Omega})$ and $\boldsymbol{Z}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, respectively, the property of total independence is equivalent to

$$\varphi(\boldsymbol{t}) = \prod_{\ell=1}^{p} \varphi_\ell(t_\ell), \qquad \boldsymbol{t} = (t_1, ..., t_p)^\top \in \mathbb{R}^p. \tag{2}$$

The linear structure of (1) involves a mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and a $(p \times p)$ matrix $\boldsymbol{\Omega}$, often termed the *mixing matrix*, which belongs to the space $\mathbb{M}^p$ of full-rank $(p \times p)$ matrices.

The independent component model is underlying the broad class of methods known as *independent component analysis* (ICA) which originates in the engineering literature on signal processing, where it has countless applications, but recently also attracted much interest in finance and economics: see, for instance, Back and Weigend [1997], Comon and Jutten [2010], Garcia-Ferrer et al. [2012], Matteson and Tsay [2017], Gourieroux et al.

[1] While *total independence* (2) implies *pairwise independence*—namely, $Z_j \perp\!\!\!\perp Z_k$ for all $j \neq k = 1, \ldots, p$ where $\perp\!\!\!\perp$, as usual, denotes stochastic independence—the converse, of course, is not true.

[2017], Hai [2020], or Miettinen et al. [2020], to quote only a few. These applications rely on the hypothesis that the ICM assumptions (1)–(2) are satisfied for some observed sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of i.i.d. copies of $\boldsymbol{X}$. Our objective here is a test of the validity of these assumptions, i.e., a test of the null hypothesis

$$\mathcal{H}_0 : \text{ the random vector } \boldsymbol{X} \text{ is such that (1) and (2) hold for some } (\boldsymbol{\mu}, \boldsymbol{\Omega}) \in \mathbb{R}^p \times \mathbb{M}^p \quad (3)$$

where $\mathbb{M}^p$ denotes the space of all $(p \times p)$ square matrices of full rank. Note that, beyond its ICM structure, $\mathcal{H}_0$ does not specify anything about $\boldsymbol{X}$; in particular, the distribution of $\boldsymbol{Z}$, hence that of $\boldsymbol{X}$, remains completely unspecified under $\mathcal{H}_0$ as well as under the alternative (which includes arbitrary dependencies between the components of $\boldsymbol{X}$). In principle, that distribution even could be discrete, although we are focusing, for simplicity, on the absolutely continuous case. The problem, therefore, is of a semiparametric nature, where the nuisance parameters are unspecified distributions. Ranks, in that context, naturally come into the picture.

Our tests are based on CFs and the fact that the null hypothesis $\mathcal{H}_0$ in (3) holds if and only if there exists $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ such that (2) are satisfied. To the best of our knowledge, this problem of testing the validity of the ICM has only briefly been touched by Matteson and Tsay [2017], who are using the notion of distance covariance. Their test statistic, however, depends on the arbitrary ordering of the components of $\boldsymbol{Z}$, which is not a desirable property.

While the problem of testing bivariate independence has been treated quite extensively in the literature, the results on testing total independence are less abundant; see, however, Blum et al. [1961], Deheuvels [1981], Csörgő [1985], Kankainen [1995]. A fundamental difference with our problem, however, is the fact that independence, in that literature, is meant among the observed quantities whereas we, in the ICM context, do not observe the $\boldsymbol{Z}$'s, and hence have to rely on estimated values.

The related problem of testing vector independence—independence between vectors of finite dimension—recently attracted renewed interest: we refer the reader to the articles by Roy et al. [2020], Genest et al. [2019], Herwartz and Maxand [2020], Bilodeau and Nangue [2017], Shi et al. [2022a] for reviews of the existing literature. CF-based methods in this context date back to Csörgő and Hall [1982], Csörgő [1985], Feuerverger [1993], Kankainen and Ushakov [1998]; for more recent contributions, see Meintanis and Iliopoulos [2008], Fan et al. [2017], Pfister et al. [2017], Chakraborty and Zhang [2019], among others. In all these tests, an empirical contrast between the joint and the product of marginal CFs is considered. The advantages in independence testing of the CF-based approach over competing methods such as those based on the distribution function are demonstrated by the success of distance covariance methods and their relation [Sejdinovic et al., 2013] to RKHS statistics; the reader is referred to Edelmann et al. [2019], Székely and Rizzo [2023], and Chen et al. [2019] for reviews on distance covariance and related approaches, and to Eriksson and Koivunen [2003] and Chen and Bickel [2005] for a discussion of their advantages in the ICM context.

As already mentioned, the underlying distributions in the problem of testing $\mathcal{H}_0$ remain largely unspecified under the null as well as under the alternative; rank-based methods then are a natural solution. As shown by Shi et al. [2022b], pseudo-Gaussian methods such as Wilks' classical test for vector independence [Wilks, 1935], indeed, although asymptotically valid, very severely over-reject under skewed distributions and for moderately large dimensions. Rank-based independence testing methods have been thoroughly investigated in the bivariate context—see Chapters II.4.11 and III.6 of Hájek and Šidák [1967] for classical references. With the recent introduction of measure transportation-based concepts of multivariate ranks and signs [Chernozhukov et al., 2017, Hallin et al., 2021], this rank-based approach to bivariate independence testing has been extended to arbitrary dimensions by Shi et al. [2022a, 2023, 2022b]; see also Ghosal and Sen [2022]. The problem we are facing here is different, though, since independence, in all these references, is to be tested between observable quantities, the values of which are i.i.d. under the null, so that the distribution-freeness of rank-based statistics is straightforward. We are dealing with unobservable variables $Z_{j\ell}$, the values of which have to be estimated as $\widehat{Z}_{j\ell}$. These "estimated residuals" are no longer i.i.d. under the null, and handling them requires additional care.

Turning to the ICM context, CF-based estimation procedures have been developed by Eriksson and Koivunen [2003] and Chen and Bickel [2005]. To the best of our knowledge, however, the corresponding tests and their performance have not been studied in any detail. The present work aims to fill this gap by investigating the properties of CF-based tests for the ICM. Our tests involve weights W; when choosing these weights, we take special care of their impact on the computation of the corresponding test statistics, a feature that is particularly important in case of high-dimensional ICMs. Finally, our tests also are exact—exact size α uniformly over $\mathcal{H}_0$, irrespective of the actual (absolutely continuous) distribution of the observations, hence fully distribution-free and affine-invariant—despite the preliminary estimation of ICM residuals $\widehat{\boldsymbol{Z}}_j$'s which are not i.i.d. under the null. However, we show that, provided that $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Omega}}$ are symmetric functions of the $\boldsymbol{X}_j$'s, the $\widehat{\boldsymbol{Z}}_j$'s enjoy an exchangeability property that allows for permutational critical values. For the same reason, the rank-based versions of our tests admit fully distribution-free exact critical values.

Still in the ICM context, rank-based tests of vector independence have been proposed by Oja et al. [2016]. Their problem is quite different from ours, though: instead of the validity of the ICM assumptions, these authors are testing the independence of two given subvectors of $\boldsymbol{X}$ under maintained ICM assumptions.

1.2 The test statistics

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, with $\boldsymbol{X}_j := (X_{j1}, \ldots, X_{jp})^\top$, $j = 1, ..., n$ denote a sample of n independent copies of $\boldsymbol{X}$. To implement the test, since we only observe $\boldsymbol{X}_j$ and not $\boldsymbol{Z}_j$, the latent variables $(\boldsymbol{Z}_j,\ j = 1, ..., n)$ first need to be estimated from the data. Specifically, our test statistic is a function of the "estimated ICM residuals"

$$\widehat{\boldsymbol{Z}}_j = \widehat{\boldsymbol{Z}}_j(\widehat{\boldsymbol{\mu}}_n, \widehat{\boldsymbol{\Omega}}_n) := \widehat{\boldsymbol{\Omega}}_n^{-1}(\boldsymbol{X}_j - \widehat{\boldsymbol{\mu}}_n) =: (\widehat{Z}_{j1}, ..., \widehat{Z}_{jp})^\top,\ j = 1, ..., n \tag{4}$$

obtained by plugging estimators $\widehat{\boldsymbol{\mu}}_n$ and $\widehat{\boldsymbol{\Omega}}_n$ of $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ into (1). It is well known, however, that the matrix $\boldsymbol{\Omega}$ in the model (1) is not uniquely identified, hence that identification constraints need to be imposed which, without any loss of generality, actually define the parameter space for $\boldsymbol{\Omega}$. These identification constraints being closely related to the construction and statistical properties of $\widehat{\boldsymbol{\Omega}}_n$, their discussion is postponed to Section 2, which is dealing with the estimation of $\boldsymbol{\Omega}$.

We now rapidly describe the proposed test statistics. Denote by

$$\varphi^{(n)}(\boldsymbol{t}) := \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath \boldsymbol{t}^\top \widehat{\boldsymbol{Z}}_j}, \quad \boldsymbol{t} \in \mathbb{R}^p \tag{5}$$

and

$$\varphi_\ell^{(n)}(t_\ell) := \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath t_\ell \widehat{Z}_{j\ell}}, \quad t_\ell \in \mathbb{R}, \quad \ell = 1, \ldots, p \tag{6}$$

(where $\imath$ stands for the imaginary root of -1) the joint and marginal empirical CFs, respectively, of the estimated ICM residuals $\widehat{\boldsymbol{Z}}_j$, $j = 1, \ldots, n$.

Similarly, letting $\mathbf{R}_j^{(n)} := (R_{j1}^{(n)}, \ldots, R_{jp}^{(n)})^\top$ where $R_{j\ell}^{(n)}$, $j = 1, \ldots, n$, stands for the rank of $\widehat{Z}_{j\ell}$ among $\widehat{Z}_{1\ell}, \ldots, \widehat{Z}_{n\ell}$, $\ell = 1, \ldots, p$, define the joint and marginal rank-based statistics

$$\varphi_{\mathbf{J}}^{(n)}(\boldsymbol{t}) := \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath \boldsymbol{t}^\top \mathbf{J}(\mathbf{R}_j^{(n)}/(n+1))}, \quad \boldsymbol{t} \in \mathbb{R}^p \tag{7}$$

and

$$\underset{\sim}{\varphi}_{\mathbf{J},\ell}^{(n)}(t_\ell) := \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath t_\ell J_\ell(R_{j\ell}^{(n)}/(n+1))}, \quad t_\ell \in \mathbb{R}, \quad \ell = 1,\ldots,p \tag{8}$$

where $\mathbf{J}(\mathbf{R}_j^{(n)}/(n+1)) := (J_1(R_{j1}^{(n)}/(n+1)),\ldots,J_p(R_{jp}^{(n)}/(n+1)))^\top$, and $\mathbf{J} := (J_1,\ldots,J_p)^\top$ denotes a p-tuple of *score functions* $J_\ell : (0,1) \to \mathbb{R}$. Clearly, (7) and (8) are the joint and marginal empirical CFs, respectively, of the *scored ranks* $J_\ell(R_{j\ell}^{(n)}/(n+1))$ of the estimated ICM residuals $\widehat{\boldsymbol{Z}}_j$, $j = 1,\ldots,n$.

Our tests are rejecting the null hypothesis $\mathcal{H}_0$ in (3) for large values of the test statistics (of the Cramér-von Mises type)

$$T_{n,W} := n\int_{\mathbb{R}^p} |D_n(\boldsymbol{t})|^2\, W(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \quad \text{and} \quad \underset{\sim}{T}_{n,\mathbf{J},W} := n\int_{\mathbb{R}^p} |\underset{\sim}{D}_{n,\mathbf{J}}(\boldsymbol{t})|^2\, W(\boldsymbol{t})\mathrm{d}\boldsymbol{t}, \tag{9}$$

where

$$D_n(\boldsymbol{t}) := \varphi^{(n)}(\boldsymbol{t}) - \prod_{\ell=1}^{p} \varphi_\ell^{(n)}(t_\ell) \quad \text{and} \quad \underset{\sim}{D}_{n,\mathbf{J}}(\boldsymbol{t}) := \underset{\sim}{\varphi}_{\mathbf{J}}^{(n)}(\boldsymbol{t}) - \prod_{\ell=1}^{p} \underset{\sim}{\varphi}_{\mathbf{J},\ell}^{(n)}(t_\ell) \tag{10}$$

with $\boldsymbol{t} := (t_1,t_2,\ldots,t_p)^\top \in \mathbb{R}^p$ and some weight function $W : \boldsymbol{t} \mapsto W(\boldsymbol{t})$. Test statistics of the Kolmogorov-Smirnov type $S_n := \sqrt{n}\sup_{\boldsymbol{t}\in\mathbb{R}^p}|D_n(\boldsymbol{t})|$ and $\underset{\sim}{S}_{n,\mathbf{J}} := \sqrt{n}\sup_{\boldsymbol{t}\in\mathbb{R}^p}|\underset{\sim}{D}_{n,\mathbf{J}}(\boldsymbol{t})|$ could also be considered; see for instance Csörgő [1985]. However, computing a $\sup_{\boldsymbol{t}\in\mathbb{R}^p}$ runs into technical difficulties (already apparent in Csörgő [1985]) since the empirical CF converges weakly only over compact neighborhoods of $\mathbb{R}^p$ and this supremum has to be taken over some discrete grid. These drawbacks appear to have a direct impact on finite–sample powers, and Kolmogorov–Smirnov type tests typically are less powerful than their Cramér–von Mises counterparts (Henze et al. [2014]). We, therefore, only consider the latter.

1.3 Outline of the paper

The rest of the paper unfolds as follows. In Section 2, we consider estimation methods for the ICM parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$. Section 3 concentrates on particular conditions under which the test statistics are affine-invariant while at the same time can be expressed in convenient closed forms, and illustrates certain connections with test statistics based on distances between densities. In Section 4, we consider permutation and bootstrap resampling methods for the computation of critical values. The consistency of our tests is established in Section 5, while in Section 6 we consider an alternative method of testing based on distance covariance. Section 7 reports the results of a Monte Carlo study on the finite-sample behavior of the new tests, with various implementation choices, e.g., different estimators, weight functions, scores, and permutational or bootstrap-based critical values. Section 8 is devoted to an empirical biomedical application. Section 9 concludes. An online supplement contains some technical arguments (Supplement A), extra simulation results (Supplement B), and additional material on the real-data application of Section 8 (Supplement C).

2 Estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$

2.1 Estimation of $\boldsymbol{\mu}$: shift invariance

An important feature of the test statistics (9) is their invariance with respect to shifts acting on $\boldsymbol{X}$. Let us show, in particular, that $\widehat{\boldsymbol{\mu}}_n$ does not influence the test statistics (9). Let $D_n^0(\boldsymbol{t}) := n^{-1}\sum_{j=1}^n \mathrm{e}^{\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j} - \prod_{\ell=1}^p n^{-1}\sum_{j=1}^n \mathrm{e}^{\imath t_\ell (\widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j)_\ell}$: clearly, $D_n^0(\boldsymbol{t})$ is obtained by letting $\widehat{\boldsymbol{\mu}}_n = \mathbf{0}$ in $D_n(\boldsymbol{t})$. Then,

$$\begin{aligned}
D_n(\boldsymbol{t}) &= \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath \boldsymbol{t}^\top \widehat{\boldsymbol{Z}}_j} - \prod_{\ell=1}^{p} \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath t_\ell \widehat{Z}_{j\ell}} \\
&= \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}(\boldsymbol{X}_j - \widehat{\boldsymbol{\mu}}_n)} - \prod_{\ell=1}^{p} \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath t_\ell \left(\widehat{\boldsymbol{\Omega}}_n^{-1}(\boldsymbol{X}_j - \widehat{\boldsymbol{\mu}}_n)\right)_\ell} \\
&= \mathrm{e}^{-\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\widehat{\boldsymbol{\mu}}_n} \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j} - \prod_{\ell=1}^{p} \mathrm{e}^{-\imath t_\ell \left(\widehat{\boldsymbol{\Omega}}_n^{-1}\widehat{\boldsymbol{\mu}}_n\right)_\ell} \prod_{\ell=1}^{p} \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath t_\ell \left(\widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j\right)_\ell} \\
&= \mathrm{e}^{-\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\widehat{\boldsymbol{\mu}}_n} \left(\frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j} - \prod_{\ell=1}^{p} \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\imath t_\ell \left(\widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j\right)_\ell} \right) = \mathrm{e}^{-\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\widehat{\boldsymbol{\mu}}_n} D_n^0(\boldsymbol{t}),
\end{aligned}$$

where $\boldsymbol{t} = (t_1, \ldots, t_p)^\top$. Hence, $|D_n(\boldsymbol{t})| = \left|\mathrm{e}^{-\imath \boldsymbol{t}^\top \widehat{\boldsymbol{\Omega}}_n^{-1}\widehat{\boldsymbol{\mu}}_n}\right| |D_n^0(\boldsymbol{t})| = |D_n^0(\boldsymbol{t})|$ and the test statistic $T_{n,W}$, which only depends on $|D_n(\cdot)|$, safely can be computed from $D_n^0(\cdot)$ instead of $D_n(\cdot)$. This allows us to skip the estimation of $\boldsymbol{\mu}$.

The ranks of the ICM residuals $\widehat{\boldsymbol{Z}}_j$ being insensitive to location shifts, the same conclusion directly holds for the rank-based statistics $\underset{\sim}{T}_{n,\mathbf{J},W}$.

2.2 Estimation of the unmixing matrix $\boldsymbol{\Omega}^{-1}$

The ICA literature proposes a number of estimators for the *unmixing matrix* $\boldsymbol{\Omega}^{-1}$ (see, e.g., Comon and Jutten [2010], Nordhausen and Oja [2018]). Below, we are focusing on three of the most popular of them: FOBI (fourth order blind identification, [Cardoso, 1989]), JADE (joint diagonalization of eigenmatrices, [Cardoso and Souloumiac, 1993]), and symmetric FastICA [Hyvarinen and Oja, 1997]. The main reasons for these choices are that they are computationally fast and widely used.

Before estimating $\boldsymbol{\Omega}^{-1}$, however, one first needs to impose identification constraints on $\boldsymbol{\Omega}$. Let $\boldsymbol{P}$, $\boldsymbol{J}$, and $\boldsymbol{D}$ denote an arbitrary p-dimensional permutation matrix, an arbitrary p-dimensional sign-change matrix (a diagonal matrix with ± 1 on its diagonal), and an arbitrary p-dimensional scaling matrix (a diagonal matrix with strictly positive diagonal elements), respectively. Then,

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Omega}\boldsymbol{Z} = \boldsymbol{\mu} + (\boldsymbol{\Omega}\boldsymbol{P}\boldsymbol{J}\boldsymbol{D})(\boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{P}^\top\boldsymbol{Z}) = \boldsymbol{\mu} + \boldsymbol{\Omega}^*\boldsymbol{Z}^*,$$

which means that the order of the components, their signs, and their scales are not well identified. Except for these indeterminacies, it can be shown, however, that $\boldsymbol{Z}$ is identifiable if at most one of its components is Gaussian. If $\boldsymbol{Z}$ has more than one Gaussian component, then the non-Gaussian ones are identifiable and the Gaussian ones are identifiable up to an additional rotation, see e.g. Theis et al. [2011] for details. The scales of the independent components Z_ℓ are usually fixed by imposing $\mathsf{Cov}(\boldsymbol{Z}) = \boldsymbol{I}_p$ (the $(p \times p)$ identity matrix) and their order is determined by the specific ICA procedure used.

Let $\boldsymbol{X}^{st} := \mathsf{Cov}(\boldsymbol{X})^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$ denote the standardized (whitened) observed vector. Almost all ICA methods make use of the fact that $\boldsymbol{X}^{st} = \boldsymbol{U}^\top\boldsymbol{Z}$ for some $p \times p$ orthogonal matrix $\boldsymbol{U}$. The unmixing matrix then has the form $\boldsymbol{\Omega}^{-1} = \boldsymbol{U}^\top \mathsf{Cov}(\boldsymbol{X})^{-1/2}$ [see e.g. Miettinen et al., 2014] and the various ICA methods differ in the way the orthogonal matrix $\boldsymbol{U}$ is obtained.

Denote by $\overline{\boldsymbol{X}}$ and $\widehat{\mathsf{Cov}}$ the sample mean and the sample covariance matrix, respectively. Then $\boldsymbol{X}_j^{st} := \widehat{\mathsf{Cov}}^{-1/2}(\boldsymbol{X}_j - \overline{\boldsymbol{X}})$, $j = 1, \ldots, n$, and FOBI estimate the orthogonal matrix $\boldsymbol{U}$ as the $p \times p$ matrix consisting of the eigenvectors of the matrix of sample fourth moments $\widehat{\mathsf{Cov}}_4 := \frac{1}{p+2}\frac{1}{n}\sum_{j=1}^n {\boldsymbol{X}_j^{st}}^\top \boldsymbol{X}_j^{st} \boldsymbol{X}_j^{st} {\boldsymbol{X}_j^{st}}^\top$. The FOBI unmixing matrix is well defined if all independent components have distinct kurtosis values, and the components are usually ordered by decreasing kurtosis order.

To avoid the assumption of distinct kurtosis values, JADE computes the p^2 cumulant matrices $\boldsymbol{C}^{kl} := \frac{1}{n}\sum_{j=1}^n {\boldsymbol{X}_j^{st}}^\top \boldsymbol{E}^{kl} \boldsymbol{X}_j^{st} \boldsymbol{X}_j^{st} {\boldsymbol{X}_j^{st}}^\top - \boldsymbol{E}^{kl} - \boldsymbol{E}^{lk} - \delta_{kl}\boldsymbol{I}_p$, $1 \le k \le l \le p$ where $\boldsymbol{E}^{kl}$ is the $p \times p$ matrix with (k,l)th element δ_{kl} (the usual Kronecker delta). The orthogonal matrix $\boldsymbol{U}$ of JADE then is the common diagonalizer of the matrices $\boldsymbol{C}^{kl}$, $1 \le k \le l \le p$.

JADE is well defined provided that one component at most has kurtosis zero.

FOBI and JADE are called *algebraic* ICA approaches as they exploit properties of cumulant matrices. An alternative popular approach consists of estimating the rows of $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$ by maximizing some componentwise criteria of non-Gaussianity in a projection pursuit framework. The most widespread family of estimators of this type is FastICA, where one has to choose the optimization criterion and a measure G of non-Gaussianity, and decide whether the rows of $\boldsymbol{U}$ are determined sequentially or simultaneously.

The so-called *symmetric* FastICA determines the rows $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$, simultaneously and maximizes $\sum_{i=1}^{p} \mathsf{E}(|G(\boldsymbol{u}_i^\top \boldsymbol{X}^{st})|)$ under the constraint $\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{I}_p$ where the measure of non-Gaussianity G can be any twice continuously differentiable and nonquadratic function G satisfying $\mathsf{E}(G(Y)) = 0$ for $Y \sim N(0,1)$. Popular choices for G are

$$\textsf{pow3} : G(x) \coloneqq (x^4 - 3)/4 \quad \text{and} \quad \textsf{tanh} : G(x) \coloneqq \log(\cosh(x)) - c_t,$$

where $c_t \coloneqq \mathsf{E}[\log(\cosh(Y))] \approx 0.375$ for $Y \sim N(0,1)$ is a normalizing constant (the names (pow3 and tanh) of these functions are related to their derivatives, which are needed in the fixed-point algorithms used for estimation). For computational details, other variants, and other G functions, see, for example, [Miettinen et al., 2017a, 2018].

Whether symmetric FastICA is consistent depends on the choice of the function G involved. It is known [Miettinen et al., 2015] that pow3 is consistent provided that there is at most one component with kurtosis zero while, for most other G functions, consistency conditions are difficult to establish. For example, FastICA with $G =$ tanh may fail for some densities [Wei, 2014, Virta and Nordhausen, 2017]. However, it is usually argued that these cases are so artificial that tanh, in practice, remains a good choice—actually, the most popular one in FastICA [Hyvarinen, 1999, Miettinen et al., 2017a].

FOBI, JADE, and symmetric FastICA all are affine-equivariant in the sense that, under affine transformations of the data, the independent components at most change their signs and their order—which, however, can be fixed via adequate conventions. FOBI and JADE are consistent and symmetric FastICA (see the previous paragraph) is considered consistent. When consistent, all these estimators have limiting normal distributions, see Miettinen et al. [2015, 2017a] for details.

Clearly, FOBI, JADE, and symmetric FastICA all require moment assumptions. ICA techniques bypassing the existence of finite moments do exist, though: see, e.g., Nordhausen, Oja, and Ollila (2008), Ilmonen and Paindaveine (2011), as well as Hallin and Mehta (2015). Unfortunately, they also are computationally much more intensive, which makes them impractical in permutational and bootstrapping approaches.

3 Weight functions

In this section, we investigate the computational issues related to our testing procedures. In particular, we discuss the choice of the weight function W in (9) in connection with affine-invariance and tests based on contrasts between density estimators.

3.1 Computational issues and affine-invariance

The weight function in (9) in principle, may be any integrable mapping W from $\mathbb{R}^p$ to the non–negative half–line. As we shall see, however, adequate choices of W considerably simplify the computation of $T_{n,W}$ and $\underset{\sim}{T}_{n,\mathbf{J},W}$.

Denoting by $w_1, \ldots, w_p$, a collection of p symmetric (i.e., such that $w_\ell(t) = w_\ell(-t)$) probability densities over $\mathbb{R}$, set

$$W : \; \boldsymbol{t} = (t_1, \ldots, t_p)^\top \mapsto W(\boldsymbol{t}) := \prod_{\ell=1}^{p} w_\ell(t_\ell). \tag{11}$$

This particular class of weight functions is in line with Kankainen and Ushakov [1998] and Meintanis and Iliopoulos [2008]. Denoting by $\mathcal{C}_\ell(t) := \int_{-\infty}^{\infty} \cos(tx) w_\ell(x) \mathrm{d}x$ the CF of w_ℓ (since w_ℓ is symmetric, the imaginary part is zero), note that $\mathcal{C}_\ell(-t) = \mathcal{C}_\ell(t)$ for all t. Then, proceeding as in Kankainen and Ushakov [1998] or Meintanis and Iliopoulos [2008], we obtain

$$T_{n,W} = \frac{1}{n} \sum_{j,k=1}^{n} \prod_{\ell=1}^{p} \mathcal{C}_\ell(\widehat{Z}_{jk,\ell}) + \frac{1}{n^{2p-1}} \prod_{\ell=1}^{p} \sum_{j,k=1}^{n} \mathcal{C}_\ell(\widehat{Z}_{jk,\ell}) - \frac{2}{n^p} \sum_{j=1}^{n} \prod_{\ell=1}^{p} \sum_{k=1}^{n} \mathcal{C}_\ell(\widehat{Z}_{jk,\ell}) \tag{12}$$

where $\widehat{Z}_{jk,\ell} := \widehat{Z}_{j\ell} - \widehat{Z}_{k\ell} := \left(\widehat{\boldsymbol{Z}}_j - \widehat{\boldsymbol{Z}}_k\right)_\ell$, $j, k = 1, ..., n$, $\ell = 1, .., p$.

This specific choice for w_ℓ considerably simplifies computations since there exist symmetric densities with particularly simple CFs, such as the stable densities, with $\mathcal{C}(t) = \mathrm{e}^{-\gamma t^\eta}$, $\eta \in (0, 2]$, $\gamma > 0$ [Nolan, 2013], or the generalized Laplace densities, with $\mathcal{C}(t) = (1+\gamma t^2)^{-\eta}$, $\eta, \gamma > 0$ [Kozubowski et al., 2013]. For $\eta = 2$ in the stable density we get the Gaussian CF $\mathcal{C}(t) = \mathrm{e}^{-\gamma t^2}$, while $\eta = 1$ in the generalized Laplace density yields $\mathcal{C}(t) = (1 + \gamma t^2)^{-1}$, the CF of the classical Laplace density.

Moreover, the resulting tests are affine-invariant, meaning that

$$T_{n,W}(\boldsymbol{A}\boldsymbol{X}_1 + \boldsymbol{b}, ..., \boldsymbol{A}\boldsymbol{X}_n + \boldsymbol{b}) = T_{n,W}(\boldsymbol{X}_1, ..., \boldsymbol{X}_n) \tag{13}$$

for any $\boldsymbol{b} \in \mathbb{R}^p$ and any full-rank $(p \times p)$ matrix $\boldsymbol{A}$. To see this, first recall that, in light of shift invariance (Section 2.1), we can set $\boldsymbol{b} = \boldsymbol{0}$. The affine equivariance of all ICA methods considered in Section 2.2 implies $\widehat{\boldsymbol{\Omega}}_{n,A} := \widehat{\boldsymbol{\Omega}}_n(\boldsymbol{A}\boldsymbol{X}_1, ..., \boldsymbol{A}\boldsymbol{X}_n) = \boldsymbol{A}\boldsymbol{J}\boldsymbol{P}\widehat{\boldsymbol{\Omega}}_n$ for some permutation matrix $\boldsymbol{P}$ and some sign-change matrix $\boldsymbol{J}$. Therefore, from (4), we have that

$$\begin{aligned}\widehat{\boldsymbol{Z}}_j^{\boldsymbol{A}\boldsymbol{X}}(\widehat{\boldsymbol{\Omega}}_{n,\boldsymbol{A}}) &= (\widehat{\boldsymbol{\Omega}}_{n,\boldsymbol{A}})^{-1}\boldsymbol{A}\boldsymbol{X}_j = (\boldsymbol{A}\boldsymbol{J}\boldsymbol{P}\widehat{\boldsymbol{\Omega}}_n)^{-1}\boldsymbol{A}\boldsymbol{X}_j \\ &= \widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{P}^{-1}\boldsymbol{J}^{-1}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{X}_j = \widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{P}^\top\boldsymbol{J}\boldsymbol{X}_j = \boldsymbol{J}\boldsymbol{P}^\top\widehat{\boldsymbol{Z}}_j^{\boldsymbol{X}}(\widehat{\boldsymbol{\Omega}}_n),\end{aligned}$$

where we use the fact that $\boldsymbol{J}$ and $\boldsymbol{P}$ are a sign matrix and a permutation matrix, respectively, so that $\boldsymbol{J}^{-1} = \boldsymbol{J}$ and $\boldsymbol{P}^{-1} = \boldsymbol{P}^\top$. Clearly then, for the class of weights in (11), $\boldsymbol{J}$ and $\boldsymbol{P}$ have no impact on $T_{n,W}$, and affine invariance readily follows.

3.2 Connection with density contrasts

In this section, we illustrate a connection of the CF–based test with tests based on density estimators that leads to a interesting interpretation of the weight function. To this end, recall the Parseval identity

$$\int_{\mathbb{R}^p} \Big|\varphi_1(\boldsymbol{t}) - \varphi_2(\boldsymbol{t})\Big|^2 \mathrm{d}\boldsymbol{t} = (2\pi)^p \int_{\mathbb{R}^p} \Big[f_1(\boldsymbol{t}) - f_2(\boldsymbol{t})\Big]^2 \mathrm{d}\boldsymbol{t}, \tag{14}$$

whereby L2–type distances between CFs $\varphi_m,\ m = 1, 2$, translate to distances between the corresponding densities $f_m,\ m = 1, 2$. Also recall the definition of *adjoint pairs* of densities: a (univariate) density w_1 is called *adjoint* to the density w_2 with CF $\mathcal{C}_2$ if $\mathcal{C}_2(x) = w_1(x)/w_1(0)$ (this implies that $\mathcal{C}_2$ is real, hence that w_2 is symmetric). Conversely, w_2 then is adjoint to w_1 and (w_1, w_2) is called an *adjoint pair*. Typical examples of adjoint densities are the Gaussian ones (which are self–adjoint), and the pairs consisting of Laplace and Cauchy densities; for more examples, see Rossberg [1995].

In this context, weight functions can be interpreted as (real-valued) CFs rather than densities. Let W be of the form[2] $W = \prod_{\ell=1}^p \mathcal{C}_\ell^2$ where $\mathcal{C}_1, \ldots, \mathcal{C}_p$ are real-valued CFs. The test statistic (9) then can be written as

$$T_{n,W} = n \int_{\mathbb{R}^p} \Big|\varphi^{(n)}(\boldsymbol{t}) \prod_{\ell=1}^p \mathcal{C}_\ell(t_\ell) - \prod_{\ell=1}^p \varphi_\ell^{(n)}(t_\ell)\mathcal{C}_\ell(t_\ell)\Big|^2 \mathrm{d}\boldsymbol{t}. \tag{15}$$

Let F_n and $F_{n\ell}$, $\ell = 1, ..., p$ denote the joint and marginal empirical distribution functions of $\widehat{\boldsymbol{Z}}_1, \ldots, \widehat{\boldsymbol{Z}}_n$. Then, writing $\mathcal{F}_\ell$, $\ell = 1, \ldots, p$ and $\mathcal{F}$ for the distribution functions corresponding to $\mathcal{C}_\ell$ and $\prod_{\ell=1}^p \mathcal{C}_\ell(t_\ell)$, respectively, $\varphi^{(n)}(\boldsymbol{t}) \prod_{\ell=1}^p \mathcal{C}_\ell(t_\ell)$ and $\varphi_\ell^{(n)}(t_\ell)\mathcal{C}_\ell(t_\ell)$ are the CFs associated with the convolutions $F_n * \mathcal{F}$ and $F_{n\ell} * \mathcal{F}_\ell$, $\ell = 1, ..., p$, respectively. Applying the Parseval identity (14) in (15) yields $T_{n,W} = n(2\pi)^p \int_{\mathbb{R}^p} \left[f_n(\boldsymbol{t}) - \prod_{\ell=1}^p f_{n\ell}(t_\ell)\right]^2 \mathrm{d}\boldsymbol{t}$,

[2] Recall that, if $\mathcal{C}$ is the CF of some random variable ξ, then $\mathcal{C}^2$ is the CF of $\xi_1 + \xi_2$ where ξ_1 and ξ_2 are independent copies of ξ.

where f_n is the density corresponding to $F_n * \mathcal{F}$, and $f_{n\ell}$ the density corresponding to $F_{n\ell} * \mathcal{F}_\ell$, $\ell = 1, ..., p$. Thus, the test statistic $T_{n,W}$ takes the form of an L2 distance between the (random) densities f_n and $\prod_\ell f_{n\ell}$, where $\prod_\ell \mathcal{C}_\ell$ and $\mathcal{C}_\ell$ act as convolution operators.

4 Critical values and distribution-freeness

The asymptotic null distribution of the test statistic $T_{n,W}$ is highly non–trivial and consequently cannot be used for practical implementation of the tests: see Kankainen [1995], Kankainen and Ushakov [1998], Pfister et al. [2017] for the case of specified (known) $\mathbf{\Omega}$, and Supplement A for the case of our composite hypothesis $\mathcal{H}_0$ under which $\mathbf{\Omega}$ needs to be estimated. For this reason, we recur to permutational and resampling techniques in order to compute critical values and calibrate the testing procedures.

Throughout this section, we make the assumption (satisfied by any reasonable estimator) that $\widehat{\mathbf{\Omega}}_n$ is a symmetric function of the n observations—that is, $\widehat{\mathbf{\Omega}}_n$ is invariant under permutations of the $\boldsymbol{X}_i$'s; recall (Section 2.1) that $\boldsymbol{\mu}$ can be assumed to be $\mathbf{0}$ without loss of generality, hence needs not be estimated.

4.1 Permutational critical values

Denote by $\widehat{Z}_{(j)k}$, $j = 1, \ldots, n$, $k = 1, \ldots, p$ the jth order statistic of the kth components $(\widehat{Z}_{k1}, \ldots, \widehat{Z}_{kn})$ of the estimated ICM residuals $(\widehat{\boldsymbol{Z}}_1, \ldots, \widehat{\boldsymbol{Z}}_n)$. Define the *order statistic* of $\widehat{\boldsymbol{Z}}^{(n)} := (\widehat{\boldsymbol{Z}}_1, \ldots, \widehat{\boldsymbol{Z}}_n)$ as the n-tuple $\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)} := (\widehat{\boldsymbol{Z}}_{(1)}, \ldots, \widehat{\boldsymbol{Z}}_{(n)})$ of vectors of estimated ICM residual values reordered from smallest to largest values of their first components, i.e., such that $(\widehat{\boldsymbol{Z}}_{(j)})_1 = \widehat{Z}_{(j)1}$, $j = 1, \ldots, n$. With the actual ICM residuals $\boldsymbol{Z}^{(n)} := (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ substituted for the estimated ones, the notation $\boldsymbol{Z}^{(n)}_{(\cdot)}$ will be used in an obvious similar way; contrary to $\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)}$, however, $\boldsymbol{Z}^{(n)}_{(\cdot)}$ is not observable. Note that, since $\widehat{\mathbf{Z}} = \widehat{\mathbf{\Omega}}^{-1}\mathbf{\Omega}\mathbf{Z}$ where $\widehat{\mathbf{\Omega}}$ is both $\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)}$- and $\boldsymbol{Z}^{(n)}_{(\cdot)}$-measurable, conditioning on $\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)}$ is equivalent to conditioning on $\boldsymbol{Z}^{(n)}_{(\cdot)}$.

Denoting by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)$ an arbitrary p-tuple of permutations of $\{1, \ldots, n\}$ and by $\mathbf{P}_{\pi_\ell}$, $\ell = 1, \ldots, p$ the $n \times n$ permutation matrix corresponding with π_ℓ (the matrix with entries $\left(\mathbf{P}_{\pi_\ell}\right)_{ij} = \delta(i, \pi_\ell^{-1}(j))$, with δ the Kronecker delta function), let

$$\widehat{\boldsymbol{Z}}^{\boldsymbol{\pi}} := \begin{pmatrix} (\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)})_{1\cdot}\, \mathbf{P}_{\pi_1} \\ \vdots \\ (\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)})_{p\cdot}\, \mathbf{P}_{\pi_p} \end{pmatrix}. \tag{16}$$

The ℓth row in matrix $\widehat{\boldsymbol{Z}}^{\boldsymbol{\pi}}$ results from performing the permutation π_ℓ on row ℓ of the order statistic $\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)}$. As $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)$ ranges over the $(n!\,)^p$ possible p-tuples of permutations of $\{1, \ldots, n\}$, $\widehat{\boldsymbol{Z}}^{\boldsymbol{\pi}}$ takes $(n!\,)^p$ values. Denoting by $T^{\boldsymbol{\pi}}_{n,W}$ the test statistic computed from $\widehat{\boldsymbol{Z}}^{\boldsymbol{\pi}}$, $T^{\boldsymbol{\pi}}_{n,W}$ similarly takes $(n!\,)^p$ possible values. These $(n!\,)^p$ values are not all distinct, though: in particular, since $T_{n,W}$ is invariant under column permutations (i.e., $\boldsymbol{\pi}$ of the form $\boldsymbol{\pi} = (\pi, \ldots, \pi)$), these $(n!\,)^p$ possible values, with probability one, all have the same multiplicity $n!$. We therefore restrict the range of $\boldsymbol{\pi}$ to the collection $\boldsymbol{\Pi}$ of p-tuples that are not column permutations (not of the form $\boldsymbol{\pi} = (\pi, \ldots, \pi)$); then, with probability one, $T^{\boldsymbol{\pi}}_{n,W}$ has $(n!\,)^{p-1}$ distinct possible values. Let us show that these $(n!\,)^{p-1}$ values, under the null and conditional on the order statistic $\widehat{\boldsymbol{Z}}^{(n)}_{(\cdot)}$ (equivalently, conditional on $\boldsymbol{Z}^{(n)}_{(\cdot)}$), are equiprobable.

From (16), we have

$$\widehat{\boldsymbol{Z}}^{\boldsymbol{\pi}} = \begin{pmatrix} (\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega})_{1\cdot}\, (\boldsymbol{Z}^{(n)}\, \mathbf{P}_{\pi_1}) \\ \vdots \\ (\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega})_{p\cdot}\, (\boldsymbol{Z}^{(n)}\, \mathbf{P}_{\pi_p}) \end{pmatrix} = \begin{pmatrix} (\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega})_{1\cdot}\, (\boldsymbol{Z}^{(n)}_{(\cdot)}\, \mathbf{P}_{\pi_1^*}) \\ \vdots \\ (\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega})_{p\cdot}\, (\boldsymbol{Z}^{(n)}_{(\cdot)}\, \mathbf{P}_{\pi_p^*}) \end{pmatrix} \tag{17}$$

with $\boldsymbol{\pi}^* := (\pi_1^*, \ldots, \pi_p^*)$ where $\pi_\ell^* = \pi \circ \pi_\ell$ and π is such that $\mathbf{Z}^{(n)} = \mathbf{Z}^{(n)}_{(\cdot)} \mathbf{P}_\pi$. When $\boldsymbol{\pi}$ ranges over $\boldsymbol{\Pi}$, $\boldsymbol{\pi}^*$ similarly ranges over $\boldsymbol{\Pi}$. Under the null hypothesis and conditional on $\boldsymbol{Z}_{(\cdot)}^{(n)}$, $(\boldsymbol{Z}_{(\cdot)}^{(n)}\mathbf{P}_{\pi_1^*}, \ldots, \boldsymbol{Z}_{(\cdot)}^{(n)}\mathbf{P}_{\pi_p^*})$ then takes, as $\boldsymbol{\pi}$ ranges over $\boldsymbol{\Pi}$, $(n!\,)^{p-1}$ possible values which are (conditionally) equiprobable. Now, (17) establishes an a.s. bijection between these $(n!\,)^{p-1}$ possible values and those of $\widehat{\boldsymbol{Z}}^{\boldsymbol{\pi}}$ which therefore, still under the null hypothesis and conditional on $\boldsymbol{Z}_{(\cdot)}^{(n)}$ (equivalently, on $\widehat{\boldsymbol{Z}}_{(\cdot)}^{(n)}$), are also equiprobable. The claim about the $(n!\,)^{p-1}$ possible values of $T_{n,W}^{\boldsymbol{\pi}}$ follows.

As a consequence, permutational critical values can be constructed as the quantile of order $(1-\alpha)$ of that uniform distribution. While the test statistic $T_{n,W}$ itself is not distribution-free (since the marginal order statistics of the sample are not), the resulting permutation test has exact[3] level α and therefore is strictly distribution-free. Note, however, that the permutations involved here are not the usual $n!$ permutations of the observations.

Even for moderate values of n and p, enumerating the $(n!\,)^{p-1}$ possible permutational values of our statistics, of course, is unfeasible. But these $(n!\,)^{p-1}$ values can be sampled, and the quantile of order $(1-\alpha)$ of that sample still can be used as an exact[3] α-level critical value (this, when the number of permutations is small, may induce a small loss of power but does not affect the size and distribution-freeness of the test).

4.2 Bootstrap critical values

As an alternative to the permutational approach of the previous section, one can also recur to the bootstrap to obtain critical values. An algorithm (Algorithm 1) producing bootstrap p–values for $T_{n,W}$ is described below, where the ICM assumptions are enforced by bootstrapping the p estimated independent components independently from each other.

[3]Since the permutational distribution is discrete, a randomization, in principle, might be required to match the exact size α. This, however, is highly negligible for large n; we do not implement it in practice, and no longer mention it in the sequel.

Algorithm 1 Bootstrapping algorithm for the p–value of the $T_{n,W}$-based test of the ICM assumptions.

Set the number of bootstrap samples M, select a weight function W and an ICA method, and consider centered observed data $\boldsymbol{X}_j,\ j=1,\dots,n$;

– apply the ICA method to obtain $\widehat{\boldsymbol{\Omega}}_n$ and $\widehat{\boldsymbol{Z}}_j = \widehat{\boldsymbol{\Omega}}_n^{-1}\boldsymbol{X}_j$, $j=1,\dots,n$;

– compute the test statistic $T = T_{n,W}$ based on $\boldsymbol{X}_j,\ j=1,\dots,n$:

for $m \in \{1,\dots,M\}$ **do**

- obtain bootstrap samples $\boldsymbol{Z}_j^m = (Z_{j1}^m,\dots,Z_{jp}^m)^\top$, $j=1,\dots,n$, where Z_{ji}^m is sampled from the empirical distribution of $(\widehat{Z}_{1i},\dots,\widehat{Z}_{ni})$, $i=1,\dots,p$;
- compute $\widehat{\boldsymbol{X}}_j^m = \widehat{\boldsymbol{\Omega}}_n\boldsymbol{Z}_j^m$, $j=1,\dots,n$;
- compute the test statistic $T^m = T_{n,W}$ based on $\widehat{\boldsymbol{X}}_j^m,\ j=1,\dots,n$;

– return the p–value::= $[\#\{T^m \geq T\}+1]/(M+1)$.

Simulations indicate that he resulting bootstrap critical values only slightly differ from the permutational ones. In the simulation section below, we provide guidelines for the choice of the number M of bootstrap samples and permutations required for a good approximation of critical values; see also Supplement B.4.

Critical values for the rank tests based on $\underset{\sim}{T}_{n,\mathbf{J},W}$ follow along the same lines as in Section 4.1, with the additional property that the $(n!)^{p-1}$ possible values of $\underset{\sim}{T}_{n,\mathbf{J},W}$ no longer depend on the values of the marginal order statistics. The test statistics $\underset{\sim}{T}_{n,\mathbf{J},W}$, thus, are conditionally uniform over these non-random $(n!)^{p-1}$ values—hence they are unconditionally uniform. It follows that $\underset{\sim}{T}_{n,\mathbf{J},W}$, unlike $T_{n,W}$, is genuinely distribution-free under $\mathcal{H}_0$—and so are the resulting tests.

4.3 The normal score test

A classical choice for the score function $\mathbf{J} = (J_1, \ldots, J_p)$ is the normal or van der Waerden score $J_\ell = \Phi^{-1}$, $\ell = 1, \ldots, p$ where Φ, as usual, stands for the standard normal distribution function. Then, instead of $\underset{\sim}{D}_{n,\mathbf{J}}(\boldsymbol{t})$ as defined in (10), one may like to consider

$$\underset{\sim}{D}_{n,\mathrm{vdW}}(\boldsymbol{t}) := \underset{\sim}{\varphi}^{(n)}_{\mathbf{J}}(\boldsymbol{t}) - \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{t}\|^2} \quad \text{with} \quad \|\boldsymbol{t}\|^2 := \sum_{\ell=1}^{p} t_\ell^2.$$

5 Consistency

In this section, we establish the consistency of our tests under rather mild conditions on the weight function W; namely, we make the following assumption.

Assumption 1 The weight function W is such that

(i) $W(\boldsymbol{t}) > 0$ Lebesgue–a.e. on $\mathbb{R}^p$; (ii) $\int_{\mathbb{R}^p} W(\boldsymbol{t})\mathrm{d}\boldsymbol{t} < \infty$.

The proof goes along similar lines as in Meintanis and Iliopoulos [2008], with the modifications required by the estimation of the mixing matrix $\boldsymbol{\Omega}$. The estimator $\widehat{\boldsymbol{\Omega}}_n$ is assumed to converge strongly. More precisely, we assume the following.

Assumption 2 As $n \to \infty$, $\widehat{\boldsymbol{\Omega}}_n$ converges strongly to some $\widetilde{\boldsymbol{\Omega}}$, where $\widetilde{\boldsymbol{\Omega}}$ under the null hypothesis $\mathcal{H}_0$ coincides with the actual value $\boldsymbol{\Omega}_0$ of $\boldsymbol{\Omega}$ and, under the alternative under study, denotes some *pseudo–true value.*[4]

Of course, $\boldsymbol{\Omega}$, $\widetilde{\boldsymbol{\Omega}}$, and $\widehat{\boldsymbol{\Omega}}_n$ must satisfy the identification constraints, that is, it should belong to the subset $\widetilde{\mathbb{M}}^p$ of matrices in $\mathbb{M}^p$ satisfying these constraints (see Section 2.2).

Note that Assumption 2 is not just an assumption on $\widehat{\boldsymbol{\Omega}}_n$ but on $\widehat{\boldsymbol{\Omega}}_n$ and the particular alternative under study. In the sequel, we denote by $\widehat{\boldsymbol{\Omega}}_n$ and $\mathcal{H}_1$, respectively, a sequence

[4] See, for instance, Matteson and Tsay [2017].

of estimators and the collection of alternatives under which $\widehat{\mathbf{\Omega}}_n$ satisfies Assumption 2.

Denote by $\widetilde{\varphi}$ the CF of $\widetilde{\boldsymbol{Z}} = (\widetilde{Z}_1, \ldots, \widetilde{Z}_p)^\top := \widetilde{\mathbf{\Omega}}^{-1}\boldsymbol{X}$ and by $\widetilde{\varphi}_\ell$ the CF of its ℓth component $\widetilde{Z}_\ell$, $\ell = 1, ..., p$. Then,

- under $\mathcal{H}_0$, $\widetilde{\varphi}$ and $\widetilde{\varphi}_\ell$ coincide with φ and φ_ℓ, hence $\widetilde{\varphi}(\boldsymbol{t}) = \prod_{\ell=1}^p \widetilde{\varphi}_\ell(t_\ell)$ for all $\boldsymbol{t} \in \mathbb{R}^p$;
- under the alternative $\mathcal{H}_1$, there exists some $\boldsymbol{t}_0 = (t_{01}, ..., t_{0p})^\top \in \mathbb{R}^p$ such that

$$\widetilde{\varphi}(\boldsymbol{t}_0) \neq \prod_{\ell=1}^{p} \widetilde{\varphi}_\ell(t_{0\ell}). \tag{18}$$

Proposition 5.1 *Let Assumptions 1 and 2 hold. Then, under any fixed alternative in* $\mathcal{H}_1$*,*

$$T_{n,W} \to \infty \text{ a.s. } \textit{as } n \to \infty.$$

Proof. From (9), we have

$$\frac{1}{n}T_{n,W} = \int_{\mathbb{R}^p} |D_n(\boldsymbol{t})|^2 W(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \tag{19}$$

where $D_n(\cdot)$ is defined in (10). Letting $\widehat{\boldsymbol{u}}_n := (\widehat{\mathbf{\Omega}}_n^{-1})^\top \boldsymbol{t}$, the joint empirical CF (5) takes the form $\varphi^{(n)}(\boldsymbol{t}) = \frac{1}{n}\sum_{j=1}^n \mathrm{e}^{\imath \widehat{\boldsymbol{u}}_n^\top \boldsymbol{X}_j} =: \phi_n(\widehat{\boldsymbol{u}}_n)$, where $\phi_n(\boldsymbol{t}) = n^{-1}\sum_{j=1}^n \mathrm{e}^{\imath \boldsymbol{t}^\top \boldsymbol{X}_j}$ is the empirical CF of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. By Assumption 2, as $n \to \infty$, $\widehat{\boldsymbol{u}}_n \longrightarrow (\widetilde{\mathbf{\Omega}}^{-1})^\top \boldsymbol{t}$ a.s., which, invoking the strong uniform consistency of the empirical CF over compact intervals (see Csörgő [1981], Csörgő [1985]), implies that, pointwise in $\boldsymbol{t}$,

$$\phi_n(\widehat{\boldsymbol{u}}_n) \longrightarrow \mathsf{E}(\mathrm{e}^{\imath \boldsymbol{t}^\top \widetilde{\mathbf{\Omega}}^{-1}\boldsymbol{X}}) = \widetilde{\varphi}(\boldsymbol{t}), \text{ a.s. as } n \to \infty. \tag{20}$$

Also, denoting by $\widehat{\boldsymbol{\omega}}_{n\ell}^{(-1)}$ the ℓ^{th} row, $\ell = 1, ..., p$, of $\widehat{\mathbf{\Omega}}_n^{-1}$, write the marginal empirical CFs (6) as $\varphi_\ell^{(n)}(t_\ell) = \frac{1}{n}\sum_{j=1}^n \mathrm{e}^{\imath \widehat{\boldsymbol{u}}_{n\ell}^\top \boldsymbol{X}_j} =: \phi_n(\widehat{\boldsymbol{u}}_{n\ell})$, where $\widehat{\boldsymbol{u}}_{n\ell} := (\widehat{\boldsymbol{\omega}}_{n\ell}^{(-1)})^\top t_\ell$. Then, by Assumption 2, $\widehat{\boldsymbol{u}}_{n\ell} \longrightarrow (\widetilde{\boldsymbol{\omega}}_\ell^{(-1)})^\top t_\ell$ a.s., which, for the same reasons as for (20), implies

$$\phi_n(\widehat{\boldsymbol{u}}_{n\ell}) \longrightarrow \mathsf{E}(\mathrm{e}^{\imath t_\ell \widetilde{\boldsymbol{\omega}}_\ell^{(-1)}\boldsymbol{X}}) = \widetilde{\varphi}_\ell(t_\ell), \text{ a.s. as } n \to \infty, \tag{21}$$

where $(\widetilde{\boldsymbol{\omega}}_\ell^{(-1)},\ \ell = 1, ..., p)$, stands for the ℓ^{th} row of $\widetilde{\mathbf{\Omega}}^{-1}$.

It follows from (20) and (21) that

$$D_n(\boldsymbol{t}) \longrightarrow \widetilde{\varphi}(\boldsymbol{t}) - \prod_{\ell=1}^{p} \widetilde{\varphi}_\ell(t_\ell) =: \widetilde{D}(\boldsymbol{t}), \text{ a.s. as } n \to \infty. \tag{22}$$

Taking (22) into account, and since $|D_n(\boldsymbol{t})|^2 \leq 4$, an application to (19) of Lebesgue's Dominated Convergence Theorem yields $\frac{1}{n} T_{n,W} \longrightarrow \int_{\mathbb{R}^p} |\widetilde{D}(\boldsymbol{t})|^2 \, W(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} =: \Delta_W$ a.s. as $n \to \infty$. In view of Assumption 1, Δ_W is strictly positive unless $\widetilde{D}(\boldsymbol{t}) = 0$ $\boldsymbol{t}$-a.e., which holds if and only if $\mathcal{H}_0$ is true while, under the alternative, (18) implies $\Delta_W > 0$. The claim follows.[5]□

Turning to the rank-based tests, we have very similar results. Denoting by $\widetilde{F}_\ell$ the distribution function of $\widetilde{Z}_\ell := \left(\widetilde{\boldsymbol{\Omega}}^{-1} \boldsymbol{X}\right)_\ell$ under Assumption 2, let $\mathcal{H}_{1;\mathbf{J}}^T$ be the subset of $\mathcal{H}_1$ under which there exists some $\boldsymbol{t}_0 = (t_{01}, ..., t_{0p})^\top \in \mathcal{B}_0(T)$ such that

$$\widetilde{\varphi}_{\mathbf{J}}(\boldsymbol{t}_0) \neq \prod_{\ell=1}^{p} \widetilde{\varphi}_{\mathbf{J},\ell}(t_{0\ell}) \tag{23}$$

where $\widetilde{\varphi}_{\mathbf{J}}$ and $\widetilde{\varphi}_{\mathbf{J},\ell}$ are the CFs of $\left(J_1(\widetilde{F}_1(\widetilde{Z}_1)), \ldots, J_p(\widetilde{F}_p(\widetilde{Z}_p))\right)^\top$ and $J_\ell(\widetilde{F}_\ell(\widetilde{Z}_\ell))$, respectively. The p variables $\widetilde{Z}_\ell$, $\ell = 1, \ldots, p$ are *totally* independent iff the p variables $\widetilde{F}_\ell(\widetilde{Z}_\ell)$ (which are uniform over $[0,1]$) are. If the scores are strictly monotone, thus, (23) holds for some $\boldsymbol{t}_0 \in \mathbb{R}^p$ under any alternative in $\mathcal{H}_1$.

Proposition 5.2 *Assume that the scores J_ℓ, $\ell = 1, \ldots, p$ are strictly monotone. Let Assumptions 1 and 2 hold. Then, under any fixed alternative in $\mathcal{H}_1$, $\underset{\sim}{T}_{n,\mathbf{J},W} \to \infty$ a.s. as $n \to \infty$.*

Proof. The proof heavily relies on the continuity of the characteristic function as a mapping from the space of distribution functions (Theorem 3.6.1 in Lukacs [1970]; Theorems 2.6.8 and 2.6.9 in Cuppens [1975]; Theorem 1.6.6 in Bogachev [2018]). To be precise, the mapping from a distribution function to the corresponding CF defines a homeomorphism between

[5]Actually, due to the continuity of CFs, (18) holds over a neighborhood of $\boldsymbol{t}_0$.

the space of distribution functions equipped with the topology of weak convergence and the space of CFs equipped with the topology of pointwise convergence.

The rank-based statistics $\underset{\sim}{\varphi}^{(n)}_{\mathbf{J}}$ and $\underset{\sim}{\varphi}^{(n)}_{\mathbf{J},\ell}$ defined in (7) and (8) actually are the joint and marginal characteristic functions of the uniform distribution over the n-tuple

$$(J_1(R^{(n)}_{j1}/(n+1)), \ldots, J_p(R^{(n)}_{jp}/(n+1)))^\top, \quad j = 1, \ldots, n.$$

Denote by $\widetilde{F}_\ell$ the distribution function of $\widetilde{Z}_\ell := \left(\widetilde{\mathbf{\Omega}}^{-1}\boldsymbol{X}\right)_\ell$. Glivenko-Cantelli, combined with the continuous mapping theorem and Assumption 2, imply the weak convergence of this joint distribution function to the distribution function of $\left(J_1(\widetilde{F}_1(\widetilde{Z}_1)), \ldots, J_p(\widetilde{F}_p(\widetilde{Z}_p))\right)^\top$. The ℓth marginal distribution function similarly weakly converges to the distribution function of $J_\ell\left(\widetilde{F}_\ell\left(\widetilde{Z}_\ell\right)\right)$, $\ell = 1, \ldots, p$. Hence, in view of the continuity of CFs, $\underset{\sim}{\varphi}^{(n)}_{\mathbf{J}}$ and $\underset{\sim}{\varphi}^{(n)}_{\mathbf{J},\ell}$ pointwise converge to the CFs $\widetilde{\varphi}_{\mathbf{J}}$ and $\widetilde{\varphi}_{\mathbf{J},\ell}$ of $\left(J_1(\widetilde{F}_1(\widetilde{Z}_1)), \ldots, J_p(\widetilde{F}_p(\widetilde{Z}_p))\right)^\top$ and $J_\ell(\widetilde{F}_\ell(\widetilde{Z}_\ell))$, respectively. For monotone score functions J_ℓ, (23) holds, under any alternative in $\mathcal{H}_1$, for some $\boldsymbol{t}_0 \in \mathbb{R}^p$. The claim then follows along the same lines as in the proof of Proposition 5.1. □

An immediate corollary to Propositions 5.1 and 5.2 is the consistency of our tests.

Corollary 5.3 *The tests based on $T_{n,W}$ and $\underset{\sim}{T}_{n,\mathbf{J},W}$, under Assumptions 1 and 2 and for monotone scores in $\underset{\sim}{T}_{n,\mathbf{J},W}$, are consistent as $n \to \infty$ against any fixed alternative in $\mathcal{H}_1$.*

6 A distance covariance test

As mentioned in the Introduction, the problem of testing the validity of the ICM model seldom has been considered in the literature. To the best of our knowledge, the only competitor to our tests is the distance covariance-based method proposed by Matteson and Tsay [2017].

Distance covariance, a concept that goes back to Székely et al. [2007], is one of the most popular methods for testing independence between two random vectors. It is well known, see for instance Chen et al. [2019], that distance covariance is a special case of the general formulation implicit in (9) that suggests measuring independence by means of a weighted L2–type distance between the estimated joint CF and the product of the corresponding marginal ones. To see this, consider n independent copies $(\boldsymbol{\xi}_j^\top, \boldsymbol{\eta}_j^\top)$, $j = 1, ..., n$, of a pair $(\boldsymbol{\xi}^\top, \boldsymbol{\eta}^\top)$ of random vectors taking values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. The empirical distance covariance between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ is of the form

$$\mathrm{dc}_n(\boldsymbol{\xi}, \boldsymbol{\eta}) := \int_{\mathbb{R}^{p+q}} \left| \varphi^{(n)}_{\boldsymbol{\xi},\boldsymbol{\eta}}(\boldsymbol{t}) - \varphi^{(n)}_{\boldsymbol{\xi}}(\boldsymbol{t}_p)\varphi^{(n)}_{\boldsymbol{\eta}}(\boldsymbol{t}_q) \right|^2 W(\boldsymbol{t})\mathrm{d}\boldsymbol{t} =: T_{n,\mathrm{dc}}, \tag{24}$$

with the weight function $W(\boldsymbol{t}) := c_{p,q} \|\boldsymbol{t}_p\|^{-(1+p)} \|\boldsymbol{t}_q\|^{-(1+q)}$, where $\|\cdot\|$ denotes the usual Euclidean norm and $c_{p,q}$ is an irrelevant constant; here $\varphi^{(n)}_{\boldsymbol{\xi},\boldsymbol{\eta}}(\boldsymbol{t})$ denotes the empirical joint CF of $(\boldsymbol{\xi}_j^\top, \boldsymbol{\eta}_j^\top)$, $j = 1, \ldots, n$, computed as in (5) at $\boldsymbol{t} = (\boldsymbol{t}_p^\top, \boldsymbol{t}_q^\top)^\top$, while

$$\varphi^{(n)}_{\boldsymbol{\xi}}(\boldsymbol{t}_p) := \varphi^{(n)}_{\boldsymbol{\xi},\boldsymbol{\eta}}((\boldsymbol{t}_p^\top, 0, ..., 0)^\top) \quad \text{and} \quad \varphi^{(n)}_{\boldsymbol{\eta}}(\boldsymbol{t}_q) := \varphi^{(n)}_{\boldsymbol{\xi},\boldsymbol{\eta}}((0, ..., 0, \boldsymbol{t}_q^\top)^\top)$$

denote the corresponding empirical marginal CFs. This leads to the convenient expression

$$\mathrm{dc}_n(\boldsymbol{\xi}, \boldsymbol{\eta}) = \frac{1}{n^2} \sum_{j,k=1}^{n} \|\boldsymbol{\xi}_{jk}\| \|\boldsymbol{\eta}_{jk}\| + \frac{1}{n^2} \sum_{j,k=1}^{n} \|\boldsymbol{\xi}_{jk}\| \frac{1}{n^2} \sum_{j,k=1}^{n} \|\boldsymbol{\eta}_{jk}\| - \frac{2}{n^3} \sum_{j=1}^{n} \sum_{k,\ell=1}^{n} \|\boldsymbol{\xi}_{jk}\| \|\boldsymbol{\eta}_{j\ell}\|, \tag{25}$$

with $\boldsymbol{\xi}_{jk} := \boldsymbol{\xi}_j - \boldsymbol{\xi}_k$ and $\boldsymbol{\eta}_{jk} := \boldsymbol{\eta}_j - \boldsymbol{\eta}_k$, $j, k = 1, ..., n$. Under the general form (24), the connection between dc_n and the test statistic $T_{n,W}$ in (9) becomes quite evident.

The null distribution of dc_n depends on the distributions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, but distribution-free versions of distance covariance based on the so-called center-outward ranks and signs [Hallin et al., 2021] have been proposed recently by [Shi et al., 2022a,b].

Distance covariance and its rank-based versions, however, are not of direct use in our context since they are designed to measure dependence between pairs of observed vectors only while we need tests for total independence. This can be taken care of by adapting an idea suggested by [Matteson and Tsay, 2017] and defining

$$\mathrm{DC}_n := n \sum_{\ell=1}^{p-1} \mathrm{dc}_n(\widehat{\boldsymbol{\zeta}}_\ell, \widehat{\boldsymbol{\zeta}}_{-\ell}) =: T_{n,\mathrm{DC}},$$

where $\widehat{\boldsymbol{\zeta}}_\ell$, $\ell = 1, \ldots, p$ corresponds to the ℓth independent component and $\widehat{\boldsymbol{\zeta}}_{-\ell}$ contains the remaining $p-1$ components.

Analogously to (7)-(8), a distribution-free version of the distance–covariance test statistic DC_n can be obtained by computing it from the scored ranks of the estimated residuals ($\widehat{Z}_{j\ell},\ \ell = 1, ..., p$). And, similarly to our tests, the distance–covariance test statistic is affine-invariant as soon as the ICA method used in estimation is affine-equivariant.

7 Simulation results

We conducted a simulation study to evaluate the performance of $T_{n,W}$ in a finite-sample setting to explore the following issues: (i) how does the test perform under $\mathcal{H}_0$ and how does the performance depend on the ICA method used to estimate the sources? (ii) how does the performance depend on the score functions used, and is there a difference between using bootstrap- and permutation-based computation for the critical value? (iii) how do the tests compare to the competitor test of Matteson and Tsay [2017] described in Section 6? (iv) how is the power with respect to various alternatives?

For that purpose, we consider the following three settings:

Setting 1: a three-dimensional ICM with uniform, exponential (with parameter $\lambda = 1$), and χ^2_3 components;

Setting 2: a three-dimensional spherical t-distribution with degrees of freedom (df) in $\{1, 2, \ldots, 19, 20, 100, 500, \infty\}$, where t_∞ stands for the Gaussian distribution (except for t_∞, this setting does not yield an ICM);

Setting 3: a three-dimensional Clayton copula model with tuning parameter ω in $\{0, 0.1, \ldots, 1.4, 1.5\}$ (except for $\omega = 0$, this setting does not yield an ICM).

The ICA methods we used are the popular FOBI, JADE, and FastICA methods based on tanh, which are all affine-equivariant, and thus due to affine–invariance of the test (see Section 13) we set $\boldsymbol{X} = \boldsymbol{Z}$ without any loss of generality.

All computations were performed using R 4.2.1 [R Core Team, 2023] with the packages JADE [Miettinen et al., 2017b], fICA [Miettinen et al., 2018], FastICA [Marchini et al., 2021], steadyICA [Risk et al., 2015], mvtnorm [Genz et al., 2021], and copula [Hofert et al., 2023]. The code for the competing test of Matteson and Tsay [2017] was kindly provided by David Matteson; we used it exactly as recommended in the original paper, based on 200 permutations, with the built-in option for FastICA, which also uses tanh. All simulation results are based on 1,000 replications. For our tests, we used the warp-speed bootstrapping [Giacomini et al., 2013] (bootstrap sample size 1,000) and we applied the same principle also in the permutational case ($M = 1,000$ permutations)—call this *warp-speed permutation*. In this warp-speed framework, the number of bootstrap samples is the number of replications in the simulations. More discussion on the number of bootstrap samples is in Supplement B.4.

For our tests, we considered the following test statistics, where superscripts Boot and Perm stand for bootstrapped and permutational critical values, respectively: $T^{\text{Boot}}_{n,G}$ and $T^{\text{Perm}}_{n,G}$ with Gaussian weights with $\gamma = 1$; $T^{\text{Boot}}_{n,L}$ and $T^{\text{Perm}}_{n,L}$ with Laplace weights with $\gamma = 1$; $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$ with identity (aka Wilcoxon) scores and Gaussian weights; $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$ with identity (aka Wilcoxon) scores and Laplace weights; $\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$ for van der Waerden scores and Gaussian weights; $T_{n,\text{DC}}$ stands for the test statistic of Matteson and Tsay [2017].

7.1 Results for Setting 1

Two important preliminary questions are: Do the tests keep the nominal α level? More particularly, how does the answer to this question depend on the choice of an ICA method? Simulations under Setting 1 (which satisfies the ICM assumptions) were conducted to answer these questions; we also included in these simulations the "oracle tests" based on the true values of the $\boldsymbol{Z}$'s. Note that it is well documented that ICA methods have convergence issues if sample sizes are small: a good estimation of the latent $\boldsymbol{Z}$'s often require thousands of observations (see Miettinen et al. [2014, 2015]).

Table 1 provides the rejection frequencies for $T_{n,G}^{\text{Boot}}$, $T_{n,G}^{\text{Perm}}$, and $T_{n,\text{DC}}$ and the rank-based $\underset{\sim}{T}_{n,Id,G}^{\text{Boot}}$, and $\underset{\sim}{T}_{n,Id,G}^{\text{Perm}}$ under Setting 1 for various sample sizes and various ICA estimation methods. The table clearly shows that the choice of an ICA method has a sizeable impact on the actual size of the tests; FastICA appears to be best from that perspective, while FOBI is worst and clearly cannot be trusted even for samples as large as $n = 4,000$. The differences between permutation and bootstrap variants are minimal and, starting from $n = 2,000$ observations, the results (except for FOBI) seem satisfactory. The size of the rank-based tests is not much closer to the nominal level than that of their "parametric" counterparts, confirming the well-known fact that these tests (even the rank-based ones rely on an ICA step), just as ICA itself, are large-sample procedures. The same conclusions hold for other weight functions: see Supplement B.1 for the corresponding tables.

7.2 Results for Settings 2 and 3

Simulations under Settings 2 and 3 provide results on finite-sample powers (with sample sizes $n = 500, 1,000, 2,000$, and $4,000$). Under Setting 2, the only distribution satisfying the ICM assumptions is the spherical Gaussian one (t_∞); all others yield dependencies. Also

	$T^{\text{Boot}}_{n,G}$				$T^{\text{Perm}}_{n,G}$				$T_{n,\text{DC}}$
n	JADE	FOBI	FastICA	TRUE	JADE	FOBI	FastICA	TRUE	FastICA
500	0.0700	0.1400	0.0602	0.0640	0.0590	0.1310	0.0453	0.0580	0.058
1,000	0.0540	0.1350	0.0400	0.0640	0.0690	0.1260	0.0500	0.0470	0.063
2,000	0.0600	0.1040	0.0590	0.0750	0.0560	0.1180	0.0510	0.0750	0.049
4,000	0.0630	0.0910	0.0520	0.0570	0.0730	0.0840	0.0430	0.0530	0.053
8,000	0.0700	0.0650	0.0660	0.0680	0.0700	0.0540	0.0610	0.0380	0.039
16,000	0.0500	0.0390	0.0420	0.0550	0.0500	0.0300	0.0480	0.0670	0.047
	$\underset{\sim}{T}^{\text{Boot}}_{n,Id,G}$				$\underset{\sim}{T}^{\text{Perm}}_{n,Id,G}$				
500	0.0930	0.1680	0.0822	0.0620	0.0910	0.1870	0.0802	0.0560	
1,000	0.0680	0.1610	0.0390	0.0520	0.0710	0.1800	0.0580	0.0630	
2,000	0.0810	0.1180	0.0570	0.0420	0.0750	0.1000	0.0490	0.0460	
4,000	0.0670	0.1140	0.0460	0.0510	0.0740	0.1120	0.0550	0.0470	
8,000	0.0530	0.0670	0.0380	0.0410	0.0710	0.0580	0.0670	0.0360	
16,000	0.0670	0.0320	0.0500	0.0570	0.0380	0.0210	0.0590	0.0370	

Table 1: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $T^{\text{Boot}}_{n,G}$, $T^{\text{Perm}}_{n,G}$, $T_{n,\text{DC}}$, $\underset{\sim}{T}^{\text{Boot}}_{n,Id,G}$, and $\underset{\sim}{T}^{\text{Perm}}_{n,Id,G}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s.

note that if the degrees of freedom are smaller than df=5, the ICA methods considered below should not work, since they require finite 4th-order moments.

Given the findings of the previous section, we only display the results based on FastICA, postponing the FOBI and JADE figures to Supplement B.2. In this regard, an inspection of Figure 1 reveals that the Gaussian model (t_∞), quite correctly, is not rejected as an ICM. Although FastICA is not expected to work well, the tests correctly reject the null hypothesis under non-Gaussian $\boldsymbol{Z}$'s for small degrees of freedom, with the rank-based procedures outperforming all other procedures. This robustness to heavy tails has been also observed by Chen and Bickel [2005], Section IV. As expected, the more degrees of freedom (hence, the closer to Gaussianity), the less powerful all tests. With Gaussian or Laplace weights, in this setting, our tests are doing equally well as $T_{n,\mathrm{DC}}$ but the power of the rank-based tests decreases more sharply. Notice that Figure 1 omits results for larger degrees of freedom as all powers then tend to the nominal level $\alpha = 5\%$. Similar figures, with similar conclusions, are obtained for JADE methods, while FOBI is not working at all: see Supplement B.2.

Similarly, under Setting 3, the only independent component model is obtained for $\omega = 0$ and the dependence between the components increases with ω. Figure 2 clearly shows, for the FastICA method, that the rejection frequency matches the nominal $\alpha = 5\%$ level at the ICM (for $\omega = 0$). Then, with increasing ω, the power increases, and, with larger sample sizes, even weaker dependencies are well detected. The rank-based tests are outperforming all other ones, including the competing test based on $T_{n,\mathrm{DC}}$. Again, larger values of ω are omitted in the figure as the power there is trivially equal to one. The corresponding figures for the FOBI- and JADE-based methods are provided in Supplement B.3, indicating, as before, that FOBI should not be used while JADE seems as good as FastICA.

In all three settings, the bootstrap and permutation versions yield essentially identical performances. Which one is to be adopted, therefore, does not matter much (for a discussion of the number of bootstrap samples/permutations, see Supplement B.4).

More importantly, the ICA method involved needs to work reasonably well in the estimation of the mixing matrix $\mathbf{\Omega}$, hence the latent $\boldsymbol{Z}$'s. It is a well-known fact, in the ICA literature, that this requires quite large samples, and as a result, so do our tests. This is hardly an issue since ICA is never considered in the analysis of small or moderate samples. This also might be the main reason why FOBI, which is particularly greedy in that respect, is working poorly here. We thus recommend JADE or FastICA. Note that FastICA was

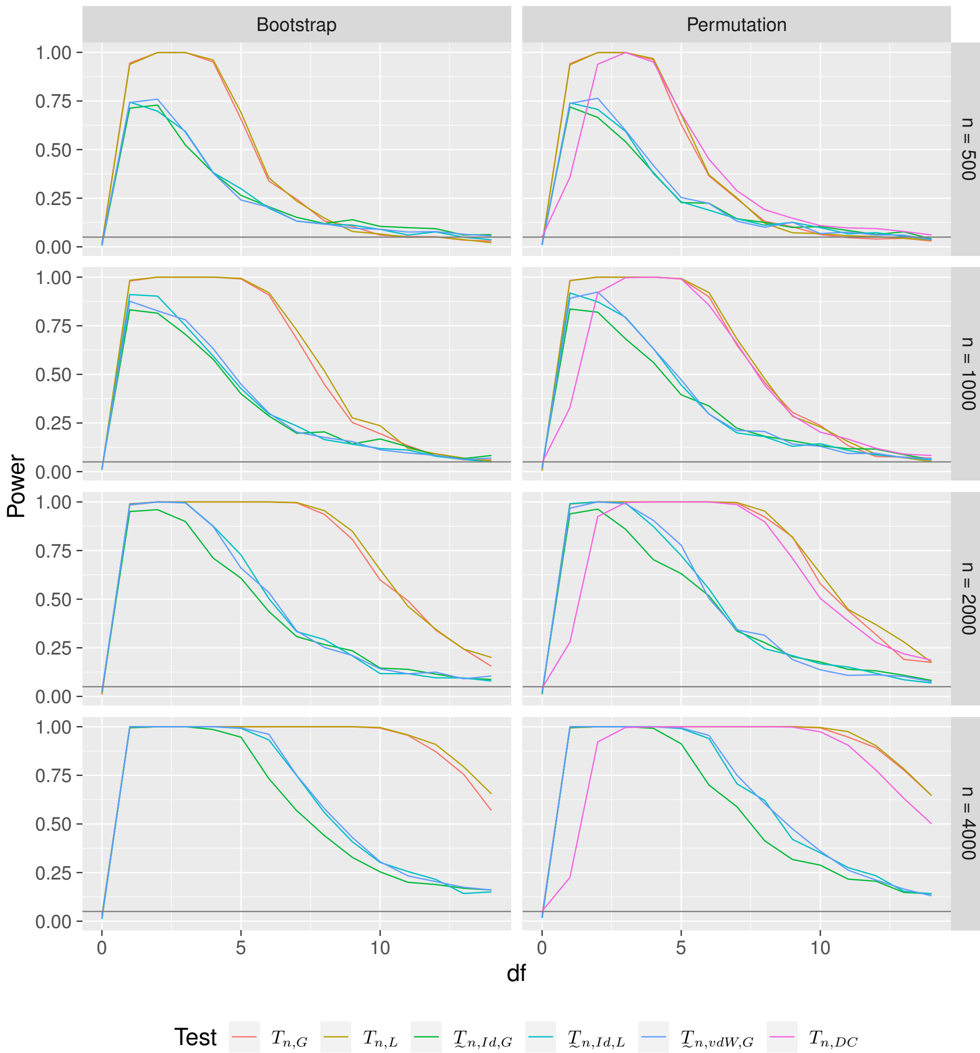

Figure 1: Empirical powers (rejection frequencies over 1,000 replications) of the tests based on $T^{\mathrm{Boot}}_{n,G}$, $T^{\mathrm{Boot}}_{n,L}$, $\underset{\sim}{T}^{\mathrm{Boot}}_{n,\mathrm{Id},G}$, $\underset{\sim}{T}^{\mathrm{Boot}}_{n,\mathrm{Id},L}$, $\underset{\sim}{T}^{\mathrm{Boot}}_{n,\mathrm{vdW},G}$, $T^{\mathrm{Perm}}_{n,G}$, $T^{\mathrm{Perm}}_{n,L}$, $\underset{\sim}{T}^{\mathrm{Perm}}_{n,\mathrm{Id},G}$, $\underset{\sim}{T}^{\mathrm{Perm}}_{n,\mathrm{Id},L}$, $\underset{\sim}{T}^{\mathrm{Perm}}_{n,\mathrm{vdW},G}$, and $T_{n,\mathrm{DC}}$ for sample sizes $n = 500, 1,000, 2,000$, and $4,000$ in Setting 2 (spherical t), as functions of the degrees of freedom when using FastICA with bootstrap and permutational critical values, respectively. The grey horizontal line represents the nominal size $\alpha = 5\%$. Note that in the figure the Gaussian distribution (t_∞) is represented under df=0.

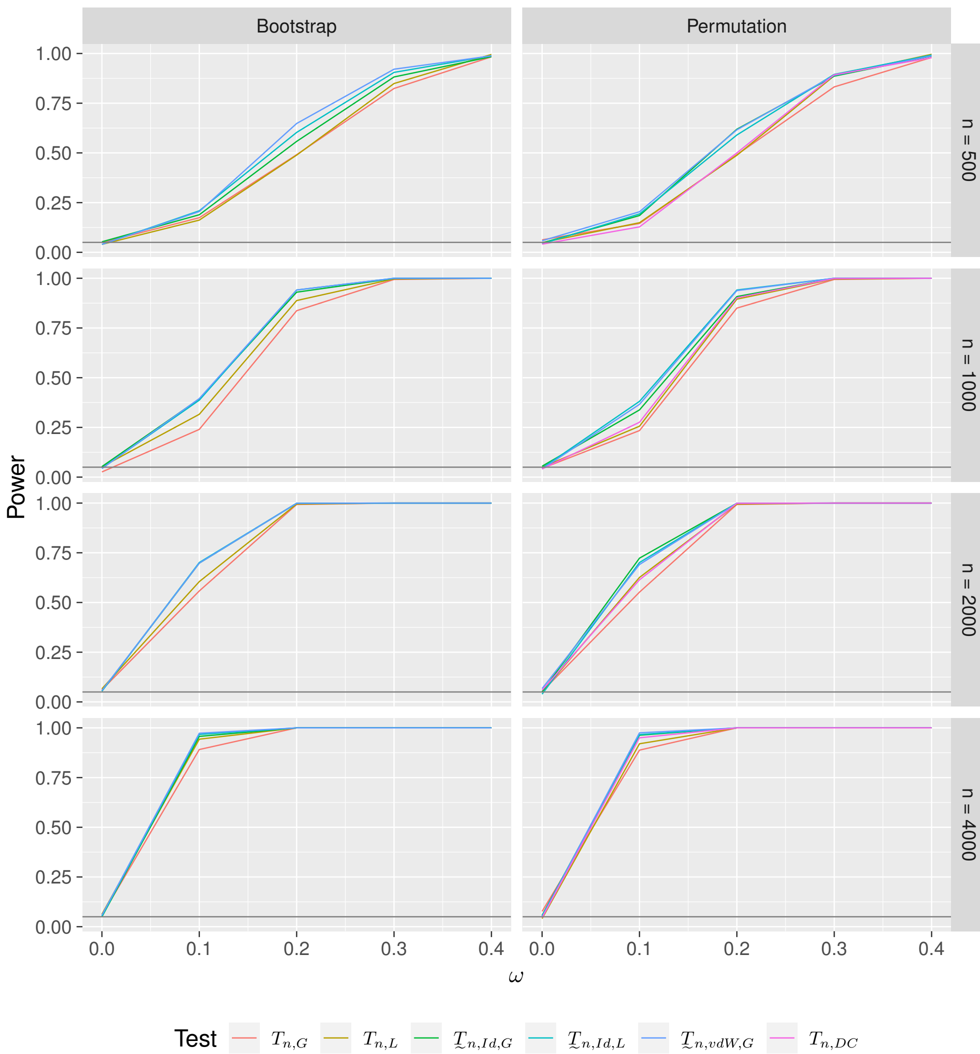

Figure 2: Empirical powers (rejection frequencies over 1,000 replications) of the tests based on $T^{\text{Boot}}_{n,G}$, $T^{\text{Boot}}_{n,L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$, $T^{\text{Perm}}_{n,G}$, $T^{\text{Perm}}_{n,L}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},L}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$, and $T_{n,\text{DC}}$ for sample sizes $n = 500$, $1,000$, $2,000$, and $4,000$ in Setting 3 (Clayton copula), as functions of the copula parameter ω when using FastICA and bootstrap or permutational critical values, respectively. The grey horizontal line represents the nominal size $\alpha = 5\%$.

also used by Matteson and Tsay [2017] in the implementation of the competing test based on $T_{n,\mathrm{DC}}$. It might be worthwhile to consider also less classical ICA methods such as, for example, the R-estimator of Hallin and Mehta [2015]. Such methods, however, usually come with a higher computational burden, which makes them impractical in the bootstrap or sampling framework used here. Different score functions $\mathbf{J}$ also are more suitable at different distributions; based on our results, it seems that a Gaussian weight function (using either the original observations or their ranks) is a good choice.

8 A real-data application

ICA is as a widely employed tool in biomedical signal processing. In this context, we will examine an electrocardiogram (ECG) dataset featuring eight sensors positioned on a pregnant woman's skin (for comprehensive details and data availability, see de Lathauwer et al. [2000] or Miettinen et al. [2017b]). ECG records the skin's electrical potential resulting from muscle activities, with the primary objective of measuring the fetal heartbeat. However, the ECG recording also captures the mother's heartbeat, as well as some other muscle activities collectively referred to as *artifacts*. ICA frequently finds application as a tool for artifact removal [He et al., 2006], with artifacts being treated as independent disturbances.

The original dataset is presented in Figure A.5 (Supplement C). Figure 3 below displays the estimated independent components based on JADE. The fourth independent component, according to experts, corresponds to the fetal heartbeat, while the last two likely are artifacts. Other components seem to be associated with the mother's heartbeat. To facilitate our tests, we initially removed second-order serial dependencies from the estimated components. We accomplished this by fitting an autoregressive (AR) model to each component, with the order selection performed via Akaike's criterion (AIC). The resulting eight residual series are depicted in Supplement C, Figure A.6.

Subsequently, we employed 500 bootstrap samples to subject the resulting vectors of residuals to hypothesis testing using the test statistics $T^{\text{Boot}}_{n,G}$ and $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ to assess if they satisfy the ICM assumptions. In both cases, we obtained a 0.002 p–value, which leads to a rejection of the null hypothesis (a result that might be due to the fact that the mother's heartbeat is present in several components). The ICA analysis of this dataset, thus, should

Figure 3: Independent components based on JADE for the ECG measurements of the pregnant women dataset.

be taken with great care.

We then restricted the analysis to the subvector identified by the experts as a vector of artifacts. Performing the same tests as above yields p–values 0.992 and 0.936, respectively, in agreement with the fact that the artifacts originate from independent sources.

9 Conclusions

Affine–invariant and globally consistent CF–based tests for the validity of the ICM are considered. Due to the complexity of the large-sample distribution of the test statistics, the tests are applied in their permutation or bootstrap versions using either the original observations or the corresponding ranks, and are implemented with a variety of estimation methods of the mixing matrix, and with suitable weight functions that lead to convenient closed-form expressions. Based on numerical results for the suggested tests we wish to register the following warning messages against casually applying the ICM: (1) large sample sizes are necessary, even in the thousands, for proper estimation and model fitting; (2) in most typical ICA applications, *deserialization* procedures for the estimated independent components are required to remove time series–type dependencies. With regards to the Monte Carlo behavior of the tests we point out the following: (1) the type of estimation does affect the empirical size of the test with FastICA being more reliable, followed by JADE; (2) rejection under alternatives reaches reasonable levels and is sample-size increasing, in line with consistency, with original and rank-based procedures being overall comparable, whether permutation- or bootstrap-based implementations are considered; (3) the new tests either seem to attain comparable powers or outperform their distance covariance–based counterpart, depending on the type of deviations from the null hypothesis.

References

A. D. Back and A. S. Weigend. A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, 8: 473–484, 1997.

M. Bilodeau and A. G. Nangue. Tests of mutual or serial independence of random vectors with applications. *Journal of Machine Learning Research*, 18:1–40, 2017.

J. R. Blum, J. Kiefer, and M. Rosenblatt. Distribution-free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, 32:485–498, 1961.

V. I. Bogachev. *Weak Convergence of Measures*. American Mathematical Society, Providence, 2018.

D. Bosq. *Linear Processes in Function Spaces: Theory and Applications*. Springer, New York, 2000.

J.-F. Cardoso. Source separation using higher order moments. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2109–2112. IEEE, 1989.

J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. In *IEE Proceedings F-Radar and Signal Processing*, volume 140, pages 362–370, 1993.

S. Chakraborty and X. Zhang. Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association*, 114: 1638–1650, 2019.

A. Chen and P. Bickel. Consistent independent component analysis and prewhitening. *IEEE Transactions on Signal Processing*, 53:3625–3632, 2005.

F. Chen, S. G. Meintanis, and L. Zhu. On some characterizations and multidimensional criteria for testing homogeneity, symmetry and independence. *Journal of Multivariate Analysis*, 173:125–144, 2019.

V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge–Kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45:223–256, 2017.

P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, Oxford, 2010.

S. Csörgő. Multivariate empirical characteristic functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55:203–229, 1981.

S. Csörgő. Testing for independence by the empirical characteristic function. *Journal of Multivariate Analysis*, 16:290–299, 1985.

S. Csörgő and P. Hall. Estimable versions of Griffiths' measure of association. *Australian Journal of Statistics*, 24:296–308, 1982.

R. Cuppens. *Decomposition of Multivariate Probabilities*. Academic Press, New York, 1975.

L. de Lathauwer, B. de Moor, and J. Vandewalle. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*, 47: 567–572, 2000.

P. Deheuvels. An asymptotic decomposition for multivariate distribution-free tests of independence. *Journal of Multivariate Analysis*, 11:102–113, 1981.

D. Edelmann, K. Fokianos, and M. Pitsillou. An updated literature review of distance correlation and its applications to time series. *International Statistical Review*, 87:237–262, 2019.

J. Eriksson and V. Koivunen. Characteristic-function-based independent component analysis. *Signal Processing*, 83:2195–2208, 2003.

Y. Fan, P. Lafaye de Micheaux, S. Penev, and D. Salopek. Multivariate nonparametric test of independence. *Journal of Multivariate Analysis*, 153:189–210, 2017.

A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61:419–433, 1993.

A. Garcia-Ferrer, E. Gonzalez-Prieto, and D. Pena. A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting*, 28:70–93, 2012.

C. Genest, J. G. Nešlehová, B. Rémillard, and O. A. Murphy. Testing for independence in arbitrary distributions. *Biometrika*, 106:47–68, 2019.

A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, and T. Hothorn. *mvtnorm: Multivariate Normal and t Distributions*, 2021. URL `https://CRAN.R-project.org/package=mvtnorm`. R package version 1.1-3.

P. Ghosal and B. Sen. Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *Annals of Statistics*, 50:1012–1037, 2022.

R. Giacomini, D. Politis, and H. White. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, 29:567–589, 2013.

C. Gourieroux, A. Monfort, and J.-P. Renne. Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics*, 196: 111–126, 2017.

T. Hai. Estimation of volatility causality in structural autoregressions with heteroskedasticity using independent component analysis. *Statistical Papers*, 61:1–16, 2020.

J. Hájek and Z. Šidák. *Theory of Rank Tests*. Academic Press, New York, 1967.

M. Hallin and C. Mehta. R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, 110:218–232, 2015.

M. Hallin, E. del Barrio, J. Cuesta-Albertos, and C. Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Annals of Statistics*, 49:1139–1165, 2021.

T. He, G. Clifford, and L. Tarassenko. Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing & Applications*, 15: 105–116, 2006.

N. Henze, Z. Hlávka, and S. Meintanis. Testing for spherical symmetry via the empirical characteristic function. *Statistics*, 48:1282–1296, 2014.

H. Herwartz and S. Maxand. Nonparametric tests for independence: a review and comparative simulation study with an application to malnutrition in India. *Statistical Papers*, 61:2175–2201, 2020.

M. Hofert, I. Kojadinovic, M. Maechler, and J. Yan. *copula: Multivariate Dependence with Copulas*, 2023. URL `https://CRAN.R-project.org/package=copula`. R package version 1.1-2.

A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10:626–634, 1999.

A. Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

P. Ilmonen and D. Paindaveine. Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Annals of Statistics*, 39:2448–2476, 2011.

A. Kankainen. *Consistent testing of total independence based on the empirical characteristic function*. PhD thesis, University of Jyväskylä, 1995.

A. Kankainen and N. Ushakov. A consistent modification of a test for independence based on the empirical characteristic function. *Journal of Mathematical Sciences*, 89:1486–1494, 1998.

T. J. Kozubowski, K. Podgórski, and I. Rychlik. Multivariate generalized Laplace distribution and related random fields. *Journal of Multivariate Analysis*, 113:59–72, 2013.

E. Lukacs. *Characteristic Functions*. Hafner Publishing Company, London, 1970.

J. L. Marchini, C. Heaton, and B. D. Ripley. *fastICA: FastICA Algorithms to Perform ICA and Projection Pursuit*, 2021. URL `https://CRAN.R-project.org/package=fastICA`. R package version 1.2-3.

D. S. Matteson and R. S. Tsay. Independent component analysis via distance covariance. *Journal of the American Statistical Association*, 112:623–637, 2017.

S. G. Meintanis and G. Iliopoulos. Fourier methods for testing multivariate independence. *Computational Statistics & Data Analysis*, 52:1884–1895, 2008.

J. Miettinen, K. Nordhausen, H. Oja, and S. Taskinen. Deflation-based FastICA with adaptive choices of nonlinearities. *IEEE Transactions on Signal Processing*, 62:5716–5724, 2014.

J. Miettinen, S. Taskinen, K. Nordhausen, and H. Oja. Fourth moments and independent component analysis. *Statistical Science*, 30:372–390, 2015.

J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, and J. Virta. The squared symmetric FastICA estimator. *Signal Processing*, 131:402–411, 2017a.

J. Miettinen, K. Nordhausen, and S. Taskinen. Blind source separation based on joint diagonalization in R: The packages `JADE` and `BSSasymp`. *Journal of Statistical Software*, 76:1–31, 2017b.

J. Miettinen, K. Nordhausen, and S. Taskinen. fICA: FastICA algorithms and their improved variants. *The R Journal*, 10:148–158, 2018.

J. Miettinen, M. Matilainen, K. Nordhausen, and S. Taskinen. Extracting conditionally heteroskedastic components using independent component analysis. *Journal of Time Series Analysis*, 41:293–311, 2020.

J. Nolan. Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*, 28:2067–2089, 2013.

K. Nordhausen and H. Oja. Independent component analysis: A statistical perspective. *WIREs: Computational Statistics*, 10:e1440, 2018.

H. Oja, D. Paindaveine, and S. Taskinen. Affine-invariant rank tests for multivariate independence in independent component models. *Electronic Journal of Statistics*, 10:2372–2419, 2016.

N. Pfister, P. Bühlmann, B. Schölkopf, and J. Peters. Kernel-based tests for joint independence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80: 5–31, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL `https://www.R-project.org/`.

B. B. Risk, N. A. James, and D. S. Matteson. *steadyICA: ICA and Tests of Independence via Multivariate Distance Covariance*, 2015. URL `https://CRAN.R-project.org/package=steadyICA`. R package version 1.0.

H. Rossberg. Positive definite probability densities and probability distributions. *Journal of Mathematical Sciences*, 76:2181–2197, 1995.

A. Roy, S. Sarkar, A. K. Ghosh, and A. Goswami. On some consistent tests of mutual independence among several random vectors of arbitrary dimensions. *Statistics and Computing*, 30:1707–1723, 2020.

D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS–based statistics in hypothesis testing. *Annals of Statistics*, 41:2263–2291, 2013.

H. Shi, M. Drton, and F. Han. Distribution-free consistent independence tests via center-outward ranks and signs. *Journal of the American Statistical Association*, 117:395–410, 2022a.

H. Shi, M. Hallin, M. Drton, and F. Han. On universally consistent and fully distribution-free rank tests of vector independence. *Annals of Statistics*, 50:1933–1959, 2022b.

H. Shi, M. Drton, M. Hallin, and F. Han. Distribution-free tests of multivariate independence based on center-outward quadrant, Spearman, and Kendall statistics. *Bernoulli,* to appear, 2023.

G. J. Székely and M. L. Rizzo. *The Energy of Data and Distance Correlation.* CRC Press, Florida, USA, 2023.

G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35:2769–2794, 2007.

F. J. Theis, M. Kawanabe, and K.-R. Muller. Uniqueness of non-Gaussianity-based dimension reduction. *IEEE Transactions on Signal Processing*, 59:4478–4482, 2011.

J. Virta and K. Nordhausen. On the optimal non-linearities for Gaussian mixtures in FastICA. In P. Tichavský, M. Babaie-Zadeh, O. J. Michel, and N. Thirion-Moreau, editors, *Latent Variable Analysis and Signal Separation*, pages 427–437. Springer, 2017.

T. Wei. On the spurious solutions of the FastICA algorithm. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, pages 161–164, 2014.

S. S. Wilks. On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3:309–326, 1935.

Appendix

A On the asymptotic behavior of the empirical independence process

Letting $\delta^{(n)}(\boldsymbol{t}) \coloneqq \sqrt{n}D_n(\boldsymbol{t}) \coloneqq \delta^{(n)}_{\Re}(\boldsymbol{t}) + \imath\, \delta^{(n)}_{\Im}(\boldsymbol{t})$, with $D_n(\cdot)$ defined in (10), call *empirical independence process* the collection $\{\delta^{(n)}(\boldsymbol{t}) | \boldsymbol{t} \in \mathbb{R}^p\}$. In this appendix, we study the behavior as $n \to \infty$ of $\{\delta^n(\boldsymbol{t})|\boldsymbol{t}\}$ under the null hypothesis. In doing so we illustrate the fact that, since $T_{n,W}$ is a functional of the empirical independent process, the respective limit null distribution, as we shall see, depends on the unspecified distributions of the independent components of the vector $\boldsymbol{Z} = (Z_1, ..., Z_p)^\top$, and the estimation method employed for the mixing matrix $\boldsymbol{\Omega}$. Due to this dependence on unspecified nuisances, this asymptotic distribution is hardly tractable and thus inappropriate for computing critical values. This is why we privilege bootstrapped or permutational critical values or distribution-free rank-based test statistics in the practical implementation of our tests.

Let $\phi_n(\boldsymbol{t}) \coloneqq n^{-1}\sum_{j=1}^n \mathrm{e}^{\imath \boldsymbol{t}^\top \mathbf{X}_j}$ denote the empirical CF computed from the n i.i.d. copies $(\mathbf{X}_j,\ j = 1, ..., n)$ of the arbitrary random vector $\mathbf{X}$, and write F and ϕ for its distribution and CFs, respectively. It is well known, see Csörgő [1985], that, under a rather mild tail condition, the empirical CF process $\Xi_n(\boldsymbol{t}) \coloneqq \{\sqrt{n}(\phi_n(\boldsymbol{t}) - \phi(\boldsymbol{t}))|\,\boldsymbol{t}\}$ converges weakly to a limit Gaussian process $\{\Xi(\boldsymbol{t})|\,\boldsymbol{t}\}$, over compact subsets C of $\mathbb{R}^p$. We will also use the fact that $\sup_{\boldsymbol{t}\in C}|\phi_n(\boldsymbol{t}) - \phi(\boldsymbol{t})| \to 0$, almost surely.

Now consider the empirical independence process. Recall from (2) the notation φ and φ_ℓ, respectively, for the characteristic functions of $\boldsymbol{Z}$ and Z_ℓ, $\ell = 1, ..., p$, and note that, under the null hypothesis $\mathcal{H}_0$ in (3), we have

$$\begin{aligned}\delta^{(n)}(\boldsymbol{t}) &= \sqrt{n}(\varphi^{(n)}(\boldsymbol{t}) - \varphi(\boldsymbol{t})) - \sqrt{n}\left(\prod_{\ell=1}^{p}\varphi_\ell^{(n)}(t_\ell) - \prod_{\ell=1}^{p}\varphi_\ell(t_\ell)\right)\\ &= \sqrt{n}(\varphi^{(n)}(\boldsymbol{t}) - \varphi(\boldsymbol{t})) \qquad (26)\\ &\quad -\sum_{\ell=1}^{p}\sqrt{n}(\varphi_\ell^{(n)}(t_\ell) - \varphi_\ell(t_\ell))\prod_{m<\ell}\varphi_m(t_m)\prod_{m>\ell}\varphi_m^{(n)}(t_m),\end{aligned}$$

where the last equality follows from the algebraic identity

$$\prod_{\ell=1}^{p} x_\ell - \prod_{\ell=1}^{p} y_\ell = \sum_{\ell=1}^{p}(x_\ell - y_\ell)\prod_{m=1}^{\ell-1} y_m \prod_{m=\ell+1}^{p} x_m.$$

Recall also from (5)–(6) the notation $\varphi^{(n)}$ and $\varphi_\ell^{(n)}$ for the joint and marginal empirical CF of $\widehat{\boldsymbol{Z}}$ and $\widehat{Z}_\ell$, respectively, and write

$$\varphi^{(n)}(\boldsymbol{t}) := \varphi_{\Re}^{(n)}(\boldsymbol{t}) + \imath\,\varphi_{\Im}^{(n)}(\boldsymbol{t}), \qquad \varphi(\boldsymbol{t}) := \varphi_{\Re}(\boldsymbol{t}) + \imath\,\varphi_{\Im}(\boldsymbol{t}),$$

$$\varphi_\ell(t) := \varphi_{\ell\Re}(t) + \imath\,\varphi_{\ell\Im}(t), \qquad \text{and} \qquad \varphi_\ell^{(n)}(t) := \varphi_{\ell\Re}^{(n)}(t) + \imath\,\varphi_{\ell\Im}^{(n)}(t), \quad \ell = 1, ..., p,$$

where the subscript $\Re$ (resp. $\Im$) is used for the real part (resp. imaginary part) of CFs and empirical CFs. Then notice that, under $\mathcal{H}_0$, the real part $\varphi_{\Re}^{(n)}$ of $\varphi^{(n)}$ admits the expansion

$$\begin{aligned}\sqrt{n}\varphi_{\Re}^{(n)}(\boldsymbol{t}) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\cos(\widehat{\boldsymbol{u}}_n^\top \boldsymbol{X}_j)\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\cos(\boldsymbol{u}_0^\top \boldsymbol{X}_j) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sin(\boldsymbol{u}_0^\top \boldsymbol{X}_j)\boldsymbol{X}_j^\top(\widehat{\boldsymbol{u}}_n - \boldsymbol{u}_0) + o_{\mathbb{P}}(1), \quad (27)\end{aligned}$$

where $\widehat{\boldsymbol{u}}_n = (\widehat{\boldsymbol{\Omega}}_n^{-1})^\top \boldsymbol{t}$ and $\boldsymbol{u}_0 = (\boldsymbol{\Omega}_0^{-1})^\top \boldsymbol{t}$, with $\boldsymbol{\Omega}_0$ denoting the actual mixing matrix under the null hypothesis. Hereafter we assume $\boldsymbol{\Omega}_0 = \boldsymbol{I}_p$, where $\boldsymbol{I}_p$ stands for the identity matrix of order $(p \times p)$. This does not imply any loss of generality in light of the affine-invariance property of Section 3.1. As argued in Section 3.1, permutation and sign matrices will likewise be suppressed below.

In this connection, assume that $\widehat{\boldsymbol{\Omega}}_n^{-1}$ admits the Bahadur asymptotic representation

$$\sqrt{n}\left(\widehat{\boldsymbol{\Omega}}_n^{-1} - \boldsymbol{I}_p\right)^\top = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\boldsymbol{V}_j + o_{\mathbb{P}}(1), \qquad (28)$$

where the columns $\boldsymbol{v}_{j\ell} = \boldsymbol{v}_\ell(\boldsymbol{Z}_j)$ of $(\boldsymbol{V}_j = \boldsymbol{V}(\boldsymbol{Z}_j),\ j = 1, ..., n)$ are such that $\mathsf{E}[\boldsymbol{v}_\ell] = \boldsymbol{0}$ and $\mathsf{E}[\boldsymbol{v}_\ell \boldsymbol{v}_\ell^\top] < \infty$, $\ell = 1, ..., p$. Such expansions may be found for the CF–based estimators of Chen and Bickel [2005], as well as the R-estimators proposed by Ilmonen and Paindaveine [2011] and Hallin and Mehta [2015].

Plugging (28) into (27), and using the consistency of empirical CFs, we have that

$$\begin{aligned}\sqrt{n}\left(\varphi_{\Re}^{(n)}(\boldsymbol{t}) - \varphi_{\Re}(\boldsymbol{t})\right) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\cos(\boldsymbol{t}^\top \boldsymbol{Z}_j) - \varphi_{\Re}(\boldsymbol{t})\right) \\ &+ \frac{\partial \varphi_{\Re}(\boldsymbol{t})}{\partial \boldsymbol{t}^\top}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\boldsymbol{V}_j\right)\boldsymbol{t} + \mathrm{o}_{\mathbb{P}}(1) \quad \text{as } n \to \infty.\end{aligned} \tag{29}$$

An analogous expansion also holds for the imaginary part $\varphi_{\Im}^{(n)}$ of $\varphi^{(n)}(\cdot)$, and this completes the asymptotic treatment of the first term in the right–hand side of (26).

For the second term in the the right–hand side of the decomposition (26) of $\delta^{(n)}$, first note that

$$\begin{aligned}&\sum_{\ell=1}^{p}\sqrt{n}(\varphi_\ell^{(n)}(t_\ell) - \varphi_\ell(t_\ell))\prod_{m<\ell}\varphi_m(t_m)\prod_{m>\ell}\varphi_{nm}(t_m) \\ &\qquad = \sum_{\ell=1}^{p}\sqrt{n}(\varphi_\ell^{(n)}(t_\ell) - \varphi_\ell(t_\ell))\prod_{m\neq\ell}\varphi_m(t_m) + \mathrm{o}_{\mathbb{P}}(1) \\ &\qquad = \sum_{\ell=1}^{p}\sqrt{n}(\varphi_\ell^{(n)}(t_\ell) - \varphi_\ell(t_\ell))\,\varphi_{(\ell)}(\boldsymbol{t}_{(\ell)}) + \mathrm{o}_{\mathbb{P}}(1),\end{aligned} \tag{30}$$

where $\varphi_{(\ell)}(\boldsymbol{t}_{(\ell)}) := \varphi_{(\ell)\Re}(\boldsymbol{t}_{(\ell)}) + \imath\varphi_{(\ell)\Im}(\boldsymbol{t}_{(\ell)})$ denotes the joint CF of $(Z_m,\ m \neq \ell)$, computed at $\boldsymbol{t}_{(\ell)} := (t_1, ..., t_{\ell-1}, t_{\ell+1}, ..., t_p)^\top \in \mathbb{R}^{p-1}$, $\ell = 1, ..., p$.

We now treat the real part of the ℓ^{th} summand $\varphi_{\ell\Re}^{(n)}$ in (30). Specifically, by taking a two-term expansion and using similar arguments as in (27), we have

$$\begin{aligned}\sqrt{n}\left(\varphi_{\ell\Re}^{(n)}(t_\ell) - \varphi_{\ell\Re}(t_\ell)\right) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\cos(t_\ell Z_{j\ell}) - \varphi_{\ell\Re}(t_\ell)\right) \\ &+ \frac{\partial \varphi_{\Re}(\boldsymbol{t}_\ell)}{\partial \boldsymbol{t}^\top}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\boldsymbol{v}_{j\ell}\right)t_\ell + \mathrm{o}_{\mathbb{P}}(1),\end{aligned} \tag{31}$$

where $\boldsymbol{t}_\ell := (0,...,0,t_\ell,0,...,0)^\top$, $\ell = 1,...,p$. Using (27), (29) and (31) and analogous expansions for the imaginary parts, we obtain $\delta_m^{(n)} = Z_m^{(n)*} + o_{\mathbb{P}}(1)$ with $m = \Re$ or $\Im$, where

$$Z_m^{(n)*}(\boldsymbol{t}) = \frac{1}{\sqrt{n}}\sum_{j=1}^n Y_{jm}(\boldsymbol{t}) \quad \text{and} \quad Y_{jm}(\boldsymbol{t}) := Y_m(\boldsymbol{Z}_j;\boldsymbol{t})$$

with

$$\begin{aligned} Y_\Re(\boldsymbol{Z};\boldsymbol{t}) &= \big(\cos(\boldsymbol{t}^\top\boldsymbol{Z}) - \varphi_\Re(\boldsymbol{t})\big) + \frac{\partial\varphi_\Re(\boldsymbol{t})}{\partial\boldsymbol{t}^\top}\boldsymbol{V}(\boldsymbol{Z})\boldsymbol{t} \\ &+ \sum_{\ell=1}^p\left\{(\sin(t_\ell Z_\ell) - \varphi_{\ell\Im}(t_\ell)) + \frac{\partial\varphi_\Im(\boldsymbol{t}_\ell)}{\partial\boldsymbol{t}^\top}\boldsymbol{v}_\ell(\boldsymbol{Z})t_\ell\right\}\varphi_{(\ell)\Im}(\boldsymbol{t}_{(\ell)}) \\ &- \left\{(\cos(t_\ell Z_\ell) - \varphi_{\ell\Re}(t_\ell)) + \frac{\partial\varphi_\Re(\boldsymbol{t}_\ell)}{\partial\boldsymbol{t}^\top}\boldsymbol{v}_\ell(\boldsymbol{Z})t_\ell\right\}\varphi_{(\ell)\Re}(\boldsymbol{t}_{(\ell)}) \end{aligned}$$

and

$$\begin{aligned} Y_\Im(\boldsymbol{Z};\boldsymbol{t}) &= \big(\sin(\boldsymbol{t}^\top\boldsymbol{Z}) - \varphi_\Im(\boldsymbol{t})\big) + \frac{\partial\varphi_\Im(\boldsymbol{t})}{\partial\boldsymbol{t}^\top}\boldsymbol{V}(\boldsymbol{Z})\boldsymbol{t} \\ &- \sum_{\ell=1}^p\left\{(\cos(t_\ell Z_\ell) - \varphi_{\ell\Re}(t_\ell))\frac{\partial\varphi_\Re(\boldsymbol{t}_\ell)}{\partial\boldsymbol{t}^\top}\boldsymbol{v}_\ell(\boldsymbol{Z})t_\ell\right\}\varphi_{(\ell)\Im}(\boldsymbol{t}_{(\ell)}) \\ &+ \left\{(\sin(t_\ell Z_\ell) - \varphi_{\ell\Im}(t_\ell)) + \frac{\partial\varphi_\Im(\boldsymbol{t}_\ell)}{\partial\boldsymbol{t}^\top}\boldsymbol{v}_\ell(\boldsymbol{Z})t_\ell\right\}\varphi_{(\ell)\Re}(\boldsymbol{t}_{(\ell)}). \end{aligned}$$

Hence, under the null hypothesis $\mathcal{H}_0$, we have, by the Central Limit Theorem (see, e.g., Theorem 2.7 in Bosq [2000]), that $\delta_m^{(n)}(\boldsymbol{t}) \xrightarrow{\mathcal{D}} \delta_m(\boldsymbol{t})$, $m = \Re$ or $\Im$, pointwise in $\boldsymbol{t} \in \mathbb{R}^p$, where $\{\delta_m\}$ is a zero–mean Gaussian process with covariance kernel

$$K_m(\boldsymbol{s},\boldsymbol{t}) = \mathsf{E}[Y_m(\boldsymbol{s})Y_m(\boldsymbol{t})], \quad m = \Re \text{ or } \Im,$$

and consequently

$$\delta^{(n)}(\boldsymbol{t}) \xrightarrow{\mathcal{D}} \delta(\boldsymbol{t}) := \delta_\Re(\boldsymbol{t}) + \imath\delta_\Im(\boldsymbol{t}).$$

Remark A.1 Clearly, each of the aforementioned covariance kernels $K_\Re(\boldsymbol{s},\boldsymbol{t})$ and $K_\Im(\boldsymbol{s},\boldsymbol{t})$ depends on the (joint) characteristic function φ of $\boldsymbol{Z}$, as well as the corresponding marginal CFs (φ_ℓ, $\ell = 1,...,p$). Moreover, the type of estimator used for $\boldsymbol{\Omega}$

also enters these kernels via the Bahadur expansion (28). These facts make the limiting null distribution intractable for actual test implementation. Even if the joint and marginal characteristic functions in $\boldsymbol{Z}$ were known, the asymptotic null distribution of the test statistic $T_{n,W}$ would still remain highly non–standard—the distribution of a weighted infinite sum of i.i.d. chi–squared random variables with one degree of freedom, with weights requiring the computation of the eigenvalues of a certain integral equation; see for instance Pfister et al. [2017], Section 3.1.

B Additional Simulation results

In this appendix, we provide the simulation results for Setting 1 that could not be included in Section 7 of the main text.

B.1 Additional results for Setting 1

	$T_{n,L}^{\mathrm{Boot}}$				$T_{n,L}^{\mathrm{Perm}}$			
n	JADE	FOBI	FastICA	TRUE	JADE	FOBI	FastICA	TRUE
500	0.0530	0.1170	0.0471	0.0550	0.0780	0.0980	0.0481	0.0660
1,000	0.0730	0.1090	0.0520	0.0540	0.0420	0.1160	0.0450	0.0540
2,000	0.0610	0.1060	0.0320	0.0500	0.0520	0.1020	0.0460	0.0650
4,000	0.0610	0.0840	0.0490	0.0660	0.0530	0.0930	0.0480	0.0530
8,000	0.0420	0.0530	0.0690	0.0480	0.0840	0.0590	0.0490	0.0560
16,000	0.0560	0.0230	0.0440	0.0410	0.0610	0.0200	0.0380	0.0440

Table A.1: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $T_{n,L}^{\mathrm{Boot}}$ and $T_{n,L}^{\mathrm{Perm}}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s.

	$\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},L}$				$\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},L}$			
n	JADE	FOBI	FastICA	TRUE	JADE	FOBI	FastICA	TRUE
500	0.1090	0.1470	0.0603	0.0720	0.0870	0.1470	0.0672	0.0550
1,000	0.0600	0.1480	0.0410	0.0470	0.0680	0.1590	0.0380	0.0420
2,000	0.0520	0.1260	0.0430	0.0590	0.0750	0.1200	0.0480	0.0430
4,000	0.0710	0.0880	0.0410	0.0450	0.0580	0.1070	0.0440	0.0570
8000	0.0860	0.0650	0.0540	0.0410	0.0500	0.0680	0.0440	0.0310
16,000	0.0440	0.0220	0.0670	0.0470	0.0550	0.0290	0.0510	0.0600

Table A.2: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},L}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},L}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s.

	$\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$				$\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$			
n	JADE	FOBI	FastICA	TRUE	JADE	FOBI	FastICA	TRUE
500	0.0970	0.1830	0.0724	0.0670	0.0900	0.1650	0.0653	0.0730
1,000	0.0640	0.1510	0.0500	0.0600	0.0630	0.1620	0.0520	0.0570
2,000	0.0570	0.1100	0.0420	0.0630	0.0710	0.1150	0.0500	0.0660
4,000	0.0770	0.1130	0.0460	0.0550	0.0630	0.0870	0.0420	0.0610
8,000	0.0510	0.0520	0.0570	0.0310	0.0540	0.0580	0.0620	0.0410
16,000	0.0560	0.0300	0.0660	0.0610	0.0660	0.0300	0.0470	0.0630

Table A.3: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s.

B.2 Additional results for Setting 2

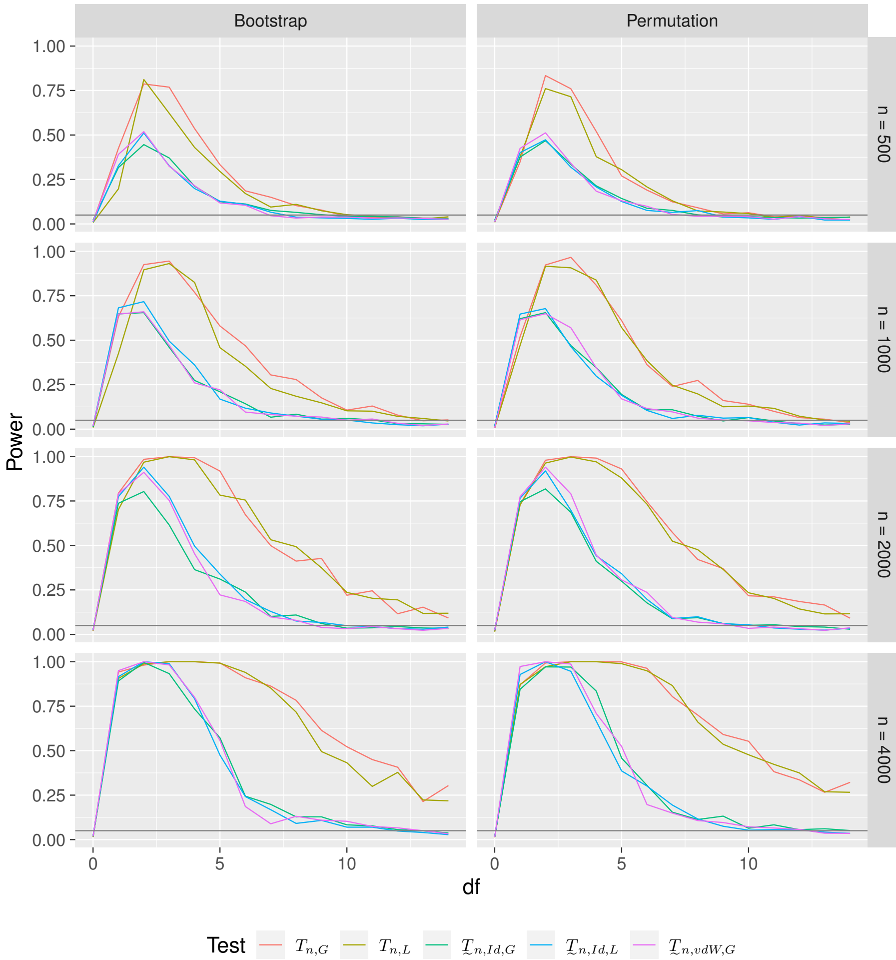

Figure A.1: Empirical powers (rejection frequencies over 1,000 replications) of the tests based on $T^{\text{Boot}}_{n,G}$, $T^{\text{Boot}}_{n,L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$, $T^{\text{Perm}}_{n,G}$, $T^{\text{Perm}}_{n,L}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},L}$, and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$, for sample sizes $n = 500$, $1,000$, $2,000$, and $4,000$ in Setting 2 (spherical t), as functions of the degrees of freedom when using FOBI, with bootstrap and permutational critical values, respectively. The grey horizontal line represents the nominal size $\alpha = 5\%$. Note that in the figure the Gaussian distribution (t_∞) is represented under df=0.

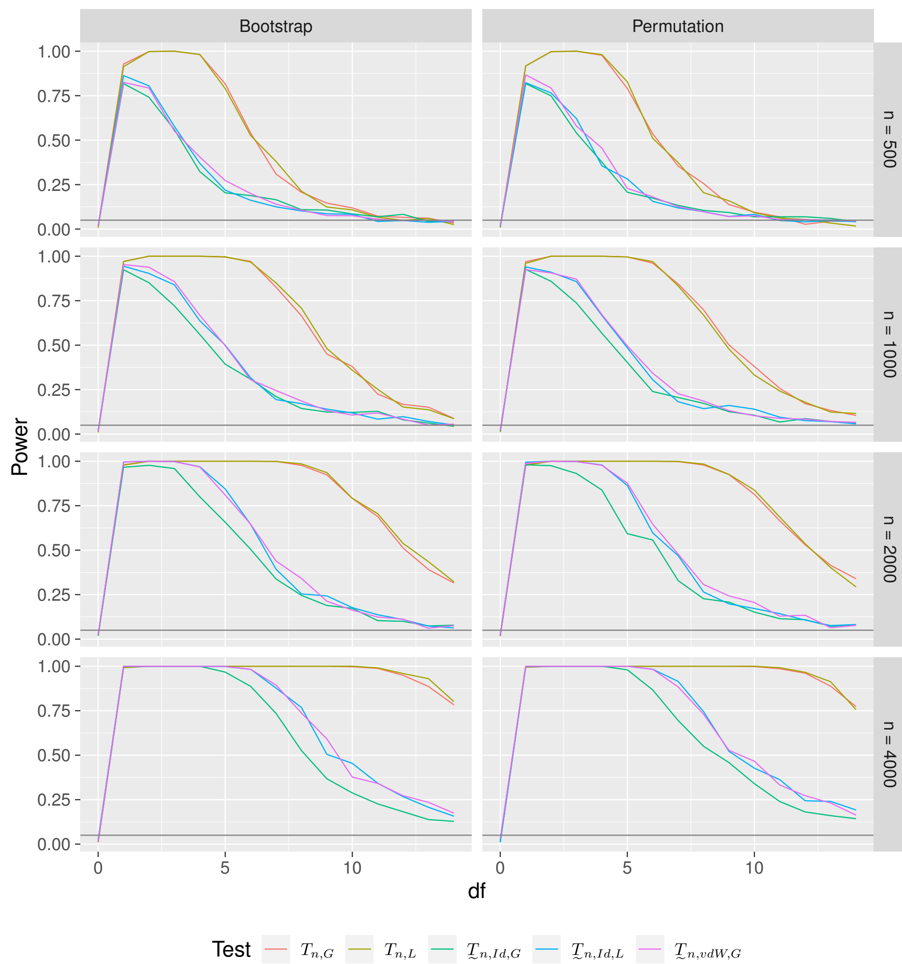

Figure A.2: Empirical powers (rejection frequencies over 1,000 replications) of the tests based on $T^{\text{Boot}}_{n,G}$, $T^{\text{Boot}}_{n,L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$, $T^{\text{Perm}}_{n,G}$, $T^{\text{Perm}}_{n,L}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},L}$, and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$ for sample sizes $n = 500$, $1,000$, $2,000$, and $4,000$ in Setting 2 (spherical t), as functions of the degrees of freedom when using JADE, with bootstrap and permutational critical values, respectively. The grey horizontal line represents the nominal size $\alpha = 5\%$. Note that in the figure the Gaussian distribution (t_∞) is represented under df=0.

B.3 Additional results for Setting 3

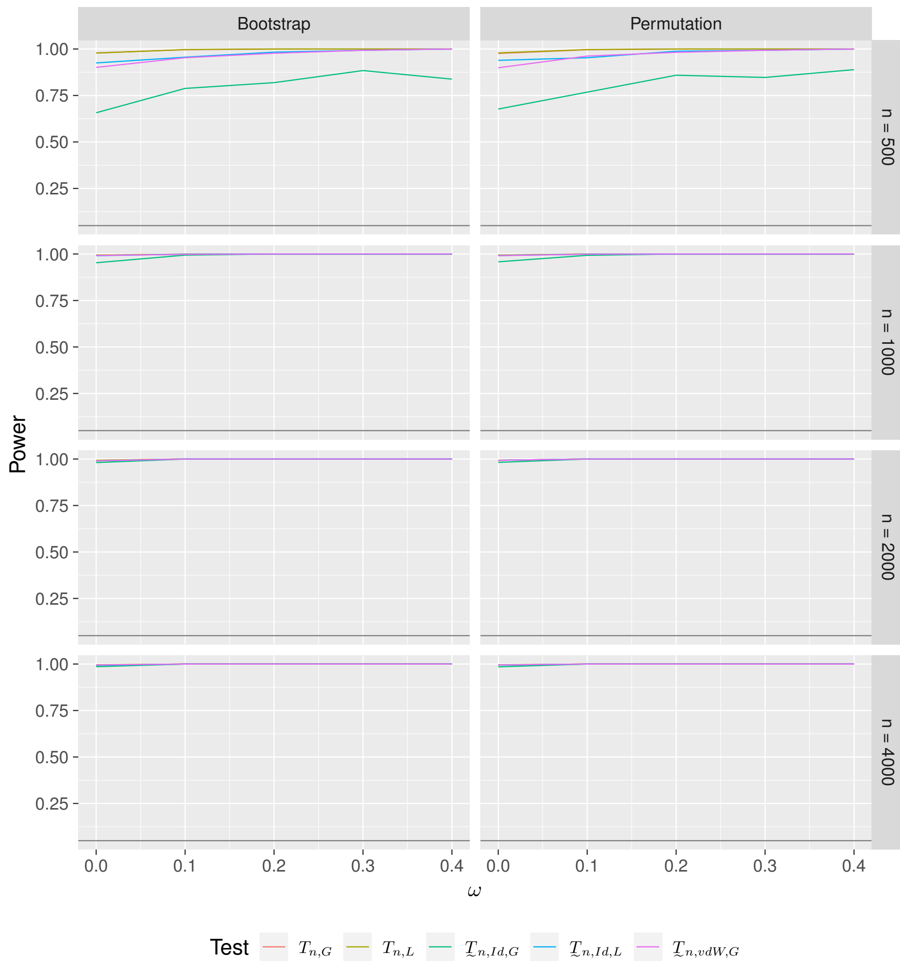

Figure A.3: Empirical powers (rejection frequencies over 1,000 replications) of the tests based on $T^{\text{Boot}}_{n,G}$, $T^{\text{Boot}}_{n,L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},L}$, $\underset{\sim}{T}^{\text{Boot}}_{n,\text{vdW},G}$, $T^{\text{Perm}}_{n,G}$, $T^{\text{Perm}}_{n,L}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$, $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},L}$, and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{vdW},G}$ for sample sizes $n = 500$, $1,000$, $2,000$, and $4,000$ in Setting 3 (Clayton copula), as functions of the copula parameter ω when using FOBI and bootstrap or permutational critical values, respectively. The grey horizontal line represents the nominal size $\alpha = 5\%$.

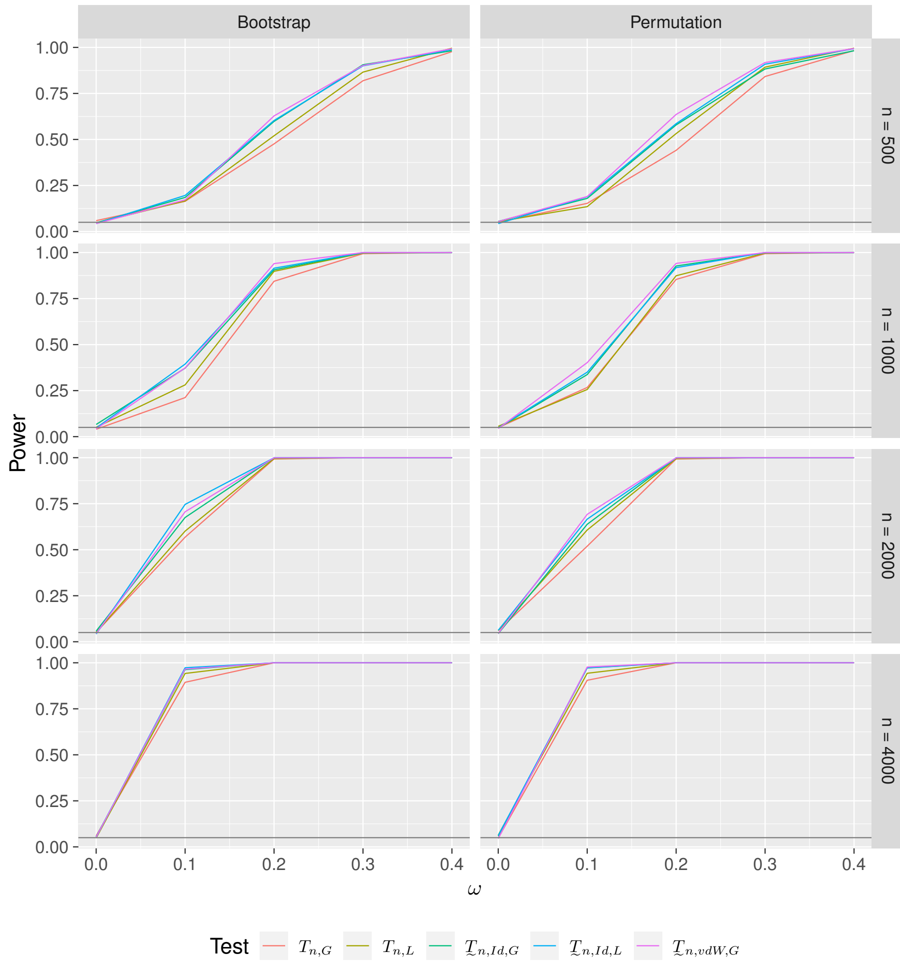

Figure A.4: Empirical powers (rejection frequencies over 1,000 replications) of the tests based on $T^{\mathrm{Boot}}_{n,G}$, $T^{\mathrm{Boot}}_{n,L}$, $\underset{\sim}{T}^{\mathrm{Boot}}_{n,\mathrm{Id},G}$, $\underset{\sim}{T}^{\mathrm{Boot}}_{n,\mathrm{Id},L}$, $\underset{\sim}{T}^{\mathrm{Boot}}_{n,\mathrm{vdW},G}$, $T^{\mathrm{Perm}}_{n,G}$, $T^{\mathrm{Perm}}_{n,L}$, $\underset{\sim}{T}^{\mathrm{Perm}}_{n,\mathrm{Id},G}$, $\underset{\sim}{T}^{\mathrm{Perm}}_{n,\mathrm{Id},L}$, and $\underset{\sim}{T}^{\mathrm{Perm}}_{n,\mathrm{vdW},G}$ for sample sizes $n = 500$, $1,000$, $2,000$, and $4,000$ in Setting 3 (Clayton copula), as functions of the copula parameter ω when using JADE and bootstrap or permutational critical values, respectively. The grey horizontal line represents the nominal size $\alpha = 5\%$.

B.4 Comments about warp-speed methods

To make simulations feasible we used the *warp-speed bootstrap* approach of Giacomini et al. [2013] in which in each simulation iteration only one bootstrap sample is created. Then the bootstrap samples from all M iterations are combined and can be used to compute the critical value or the p-value of interest. As in our simulation settings $M = 1,000$, the simulation results in the previous sections are based on a number $M = 1,000$ of bootstrap samples. To evaluate the impact of the number of bootstrap samples, we also checked how in Setting 1 the size changes when using only a random sample out of these $M = 1,000$ warp-speed bootstrap samples. The results are shown in Tables A.4-A.6 and reveal that 200 samples might not be sufficient for the approximation while 500 seems to be a good value. For other score functions, the results, which are not shown here, are quite comparable.

	$\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$				$\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$			
n	JADE	FOBI	FastICA	TRUE	JADE	FOBI	FastICA	TRUE
500	0.0700	0.1400	0.0602	0.0640	0.0590	0.1310	0.0453	0.0580
1,000	0.0540	0.1350	0.0400	0.0640	0.0690	0.1260	0.0500	0.0470
2,000	0.0600	0.1040	0.0590	0.0750	0.0560	0.1180	0.0510	0.0750
4,000	0.0630	0.0910	0.0520	0.0570	0.0730	0.0840	0.0430	0.0530
8,000	0.0700	0.0650	0.0660	0.0680	0.0700	0.0540	0.0610	0.0380
16,000	0.0500	0.0390	0.0420	0.0550	0.0500	0.0300	0.0480	0.0670

Table A.4: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s. Critical values are based on 1,000 bootstrap samples and 1,000 permutations, respectively.

While warp-speed bootstrap has a sound theoretical foundation [Giacomini et al., 2013], it seems that warp-speed permutations have not been considered so far in the literature.

,

n	$T_{rG,B}$ JADE	FOBI	FastICA	TRUE	$T_{rG,P}$ Jade	Fobi	FastICA	TRUE
500	0.0700	0.1560	0.0562	0.0750	0.0760	0.1310	0.0453	0.0510
1000	0.0550	0.1400	0.0320	0.0560	0.0600	0.1220	0.0430	0.0460
2000	0.0560	0.1120	0.0620	0.0840	0.0490	0.1260	0.0460	0.0780
4000	0.0490	0.0850	0.0600	0.0580	0.0730	0.0860	0.0380	0.0540
8000	0.0630	0.0710	0.0720	0.0680	0.0600	0.0590	0.0630	0.0350
16000	0.0480	0.0390	0.0570	0.0580	0.0370	0.0250	0.0510	0.0670

Table A.5: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s. Critical values are based on 500 bootstrap samples and 500 permutations, respectively.

We leave the theoretic justification for further research and only show below some empirical evidence that supports such an approach. In all simulations shown above, the warp-speed permutation p–values are computed such that, at each iteration $m = 1, \ldots, M$, only one permutation is performed, and all these permutations are combined for the individual p–values. While in the traditional, full permutational, approach at each iteration P permutation should be executed, and the p–value at that iteration based on these permutations only. As in our simulations study $M = 1,000$ above all warp-speed permutations are therefore based on a comparison to 1000 reference values. Assuming that like in the bootstrap case, as investigated in Tables A.4-A.6, also a smaller number P of permutations might be of interest, we considered smaller values for P and restricted the warp-speed p–value computation to a random subsample of the M permuation values of size P. We did this comparison in Setting 1 using the tests $T_{n,G}$ and $\underset{\sim}{T}_{n,\text{Id},G}$ using $n = 2,000$ based on $M = 1,000$ replications. The FastICA method was used here. The results are presented

n	$\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ JADE	FOBI	FastICA	TRUE	$\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$ Jade	Fobi	FastICA	TRUE
500	0.0970	0.1490	0.0602	0.0660	0.0760	0.0840	0.0735	0.0580
1000	0.0310	0.0920	0.0420	0.0610	0.0640	0.1240	0.0600	0.0410
2000	0.0420	0.1100	0.0560	0.0750	0.0740	0.1500	0.0630	0.0750
4000	0.0260	0.0970	0.0690	0.0160	0.0830	0.0920	0.0320	0.0580
8000	0.0700	0.0600	0.0720	0.0530	0.0540	0.0540	0.0630	0.0540
16000	0.0540	0.0190	0.0310	0.0560	0.0500	0.0190	0.0380	0.0670

Table A.6: Empirical sizes (rejection frequencies over 1,000 replications) of the tests based on $\underset{\sim}{T}^{\text{Boot}}_{n,\text{Id},G}$ and $\underset{\sim}{T}^{\text{Perm}}_{n,\text{Id},G}$ (nominal size $\alpha = 5\%$) under Setting 1 and various ICA methods. Column TRUE corresponds to the oracle test based on the actual values of the $\boldsymbol{Z}$'s. Critical values are based on 200 bootstrap samples and 200 permutations, respectively.

P	Power $T^{\text{wsPerm}}_{n,G}$	Power $T^{\text{fPerm}}_{n,G}$	Disagree	Power $\underset{\sim}{T}^{\text{wsPerm}}_{n,\text{Id},G}$	Power $\underset{\sim}{T}^{\text{fPerm}}_{n,\text{Id},G}$	Disagree
1,000	0.054	0.057	1.5 %	0.068	0.072	2.4%
500	0.053	0.053	2.0 %	0.072	0.072	2.4%
200	0.057	0.052	2.3 %	0.069	0.059	2.8%

Table A.7: Comparison of warp-speed (wsPerm) permutation tests versus full (fPerm) permutation tests for different numbers of permutations P used in Setting 1 with $n = 2,000$ based on 1,000 replications. The column "Disagree" gives the percentage of cases where the test decisions in the two variants differed at the level $\alpha = 0.05$.

in Table A.7. The table reports also the percentage of cases where the full permutation case and the warp-speed case lead to different decisions at probability level $\alpha = 0.05$. As the table shows, the differences in overall power between the full permutational approach and the warp-speed approach are minimal and lead in almost all cases to the same decision, thereby fully justifying the usage of the warp-speed approach in our simulation study. The number of permutations seems to have little impact here but we nevertheless recommend to use rather more than 200.

C Additional results for the application in Section 8

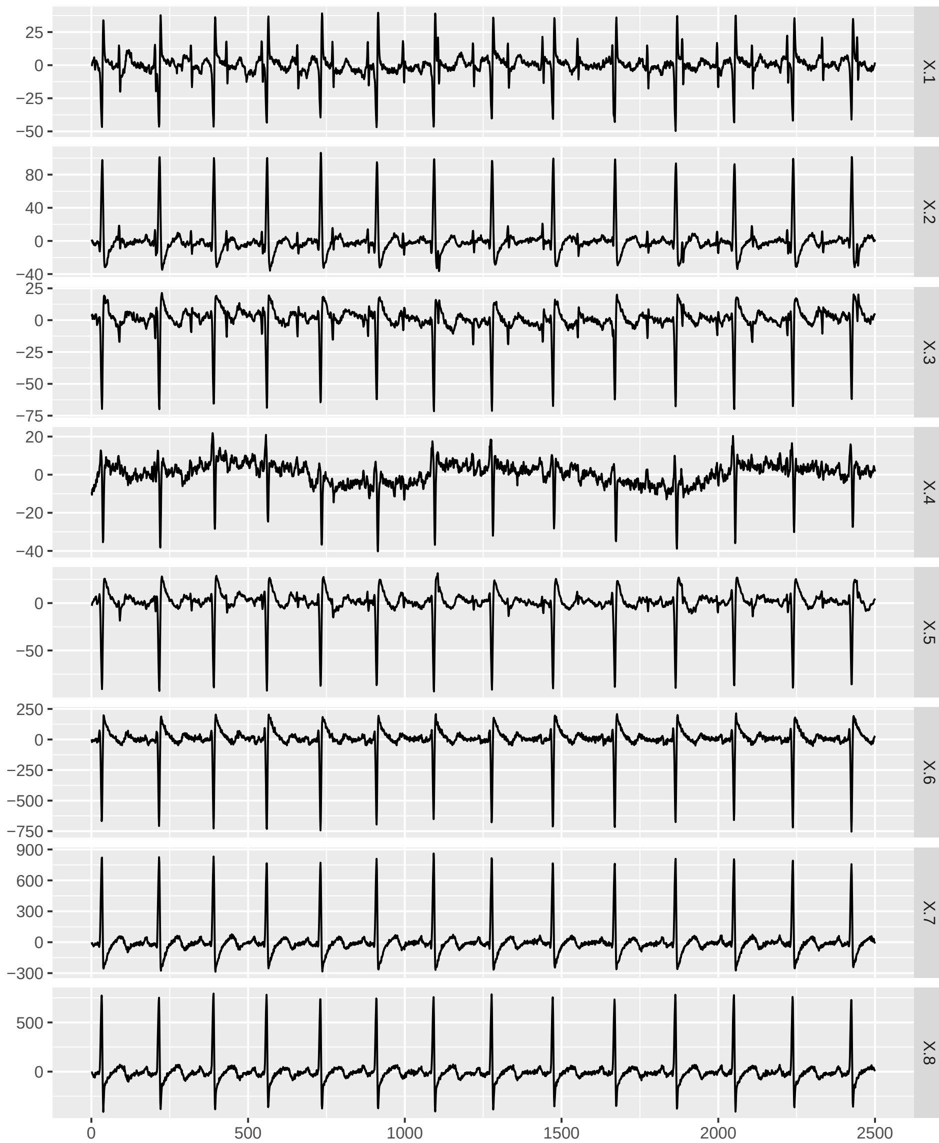

Figure A.5: ECG measurements for the pregnant women.

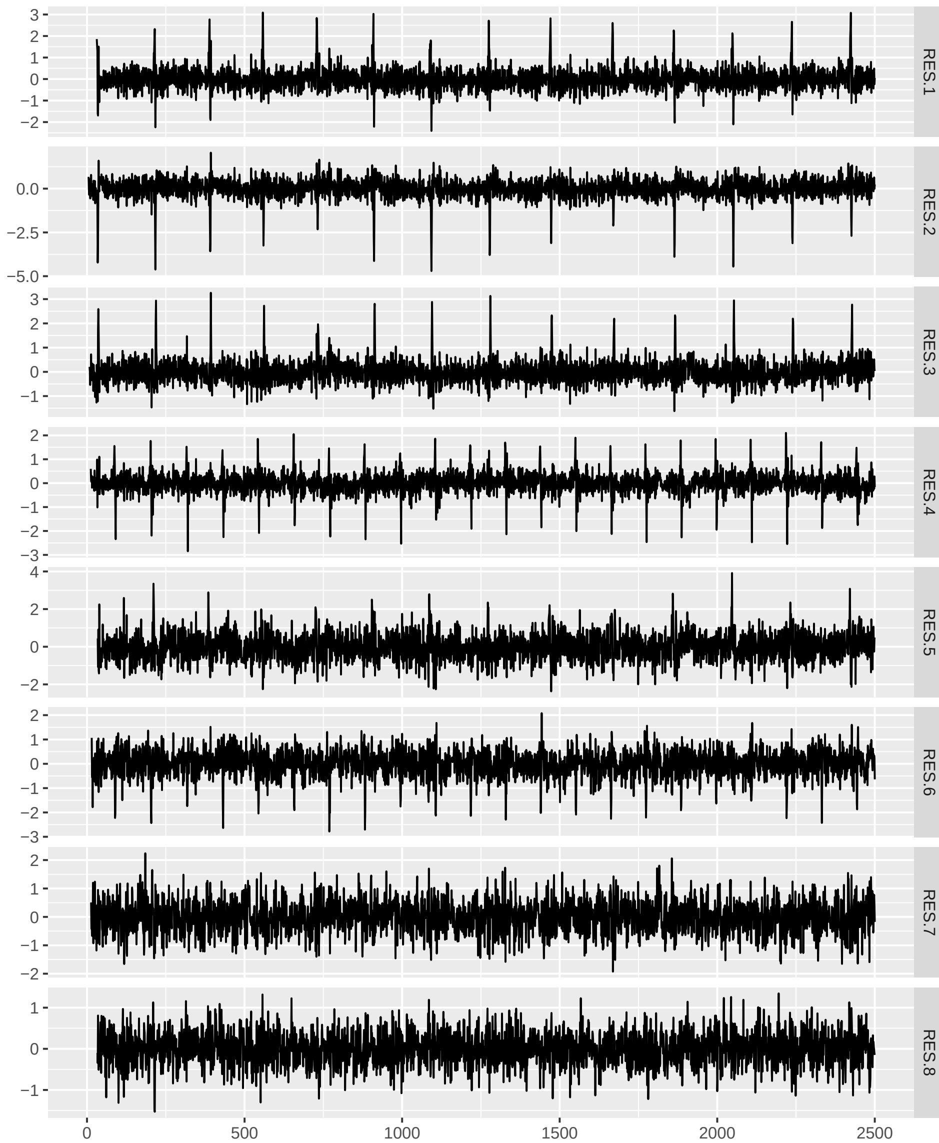

Figure A.6: Residuals after fitting AR processes to all independent components based on JADE for the ECG measurements of the pregnant women.