Journal of Economic Dynamics & Control 144 (2022) 104530

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Identification of Structural VAR Models via Independent Component Analysis: A Performance Evaluation Study

Alessio Moneta[1,*], Gianluca Pallante[1]

Institute of Economics and EMbeDS, Sant'Anna School of Advanced Studies, Pisa

ARTICLE INFO

Article history:
Received 2 May 2022
Revised 18 August 2022
Accepted 18 September 2022
Available online 20 September 2022

JEL classification:
C14
C32
E62

Keywords:
Independent Component Analysis
Identification
Structural VAR
Impulse response functions
Non-Gaussianity
Generalized normal distribution

ABSTRACT

Independent Component Analysis (ICA) is a statistical method that linearly transforms a random vector. Under the assumption that the observed data are mixtures of non-Gaussian and independent processes, ICA is able to recover the underlying components, but with a scale and order indeterminacy. Its application to structural vector autoregressive (SVAR) models allows the researcher to recover the impact of independent structural shocks on the observed series from estimated residuals. We analyze different ICA estimators, recently proposed within the field of SVAR analysis, and compare their performance in recovering structural coefficients. Moreover, we assess the size distortions of the estimators in hypothesis testing. We conduct our analysis by focusing on non-Gaussian distributional scenarios that get gradually close to the Gaussian case. The latter is the case where ICA methods fail to recover the independent components. Although the ICA estimators that we analyze show similar pattern of performance, two of them — the fastICA algorithm and the pseudo-maximum likelihood estimator — tend to perform relatively better in terms of variability, stability across sub- and super-Gaussian settings, and size distortion. We finally present an empirical illustration using US data to identify the effects of government spending and tax cuts on economic activity, thus providing an example where ICA techniques can be used for hypothesis testing.

1. Introduction

The aim of this paper is to evaluate a set of methods that have been recently proposed to achieve statistical identification of structural autoregressive (SVAR) models based on non-Gaussianity. One of the most important and pursued objectives in macroeconomics is to estimate the dynamic effect of an unexpected change in one variable, usually called shock, on other variables. Since the seminal work of Sims (1980), the study of the joint dynamics of the main macroeconomic aggregates has been conducted in the framework of vector autoregressive (VAR) models. These models have been proposed as an alternative to large simultaneous equation models (Klein and Goldberger, 1955), which were criticized for their large number of identifying and arbitrary restrictions. However, while in forecasting (reduced-form) VAR models have been proven to be powerful tools, for policy analysis one needs to deal with structural models.

* Corresponding author.
E-mail addresses: a.moneta@santannapisa.it (A. Moneta), gianluca.pallante@santannapisa.it (G. Pallante).

[1] The authors acknowledge support by the project "How good is your model? Empirical evaluation and validation of quantitative models in economics" funded by PRIN grant no. 20177FX2A7. They also thank Fulvio Corsi, Gabriele Fiorentini and two anonymous referees for comments and suggestions on an earlier version of the paper.

https://doi.org/10.1016/j.jedc.2022.104530
0165-1889/

It is indeed necessary to distinguish correlation from causation (Stock and Watson, 2017) if the goal of the analysis is measuring the effects of exogenous interventions on the system. Specifically, the residuals of an estimated (reduced-form) VAR model typically display cross-correlations, induced by contemporaneous causal relationships that cannot be detected in the regression estimates. There are infinite possibilities of linearly transforming the VAR model in order to get uncorrelated error terms, corresponding to infinite observationally equivalent structural models. Researchers aim at finding the linear transformation that yields both uncorrelated (in some cases, independent) and economically meaningful disturbances, whose effects can be studied through impulse response analysis.

In the empirical macroeconomic literature, several identification criteria have been proposed. Following Sims (1980) and Sims (1986), many empirical works have exploited the Choleski decomposition of the covariance matrix of the VAR forecast errors. This procedure provides an orthogonalization of the residuals by imposing a recursive scheme on the contemporaneous causal structure, implicit in the ordering of the endogenous variables. The decision on the sequence of the variables is of crucial importance but sometimes loosely motivated. Economic theory or background knowledge may help achieve identification by imposing (typically zero) restrictions on the contemporaneous causal impact of one variable on another (Bernanke, 1986; Bernanke and Mihov, 1998; Blanchard and Perotti, 2002). The reliability of the implied causal structure, however, can be hardly justified on the basis of mere *a priori* knowledge (Stock and Watson, 2001).

Alternative identification strategies are based on long-run restrictions, use of external instruments (extraneous data in general), sign restrictions, and heteroskedasticity (see Kilian and Lütkepohl, 2017; Stock and Watson, 2016, for an overview). External instruments have been used, for example, by Gertler and Karadi (2015), who identify the unexpected change in policy interest rate taking as instrument movements of futures prices around policy announcements. Romer and Romer (2010) adopt a similar identification procedure building a narrative series based on tax change announcements. Another popular identification approach is based on sign restrictions (see Mountford and Uhlig, 2009; Uhlig, 2005). A specific feature of this approach is that the structural coefficients are set identified, rather than point identified. It is also typical to rely on Bayesian methods of inference, which in sign-identified models may introduce the problem of priors that influence the posteriors of the structural coefficients (Kilian and Lütkepohl, 2017). Identification of SVAR models by heteroskedasticity is achieved by relying on the assumption that the contemporaneous causal structure does not vary over time, but their covariances change across regimes (Rigobon, 2003). The emerging of some of these identification strategies have made SVAR identification based less on restrictions guided by theory and more on statistical properties of the data.

Another stream of literature has exploited non-Gaussianity, instead of heteroskedasticity, as a statistical property sufficient for local identification (see, among others, Fiorentini and Sentana, 2020; Gouriéroux et al., 2020; Guay, 2021; Lanne and Luoto, 2021; Lanne and Lütkepohl, 2010; Lanne et al., 2017). The idea of some of these studies is to recover the SVAR shocks as linear combinations of reduced-form VAR residuals, under the assumption that they are not just uncorrelated, but mutually statistical independent. This is possible by means of a class of data-driven statistical techniques, called Independent Component Analysis (ICA, see Comon, 1994; Eriksson and Koivunen, 2004; Hyvärinen, 1999). Under the assumptions of non-Gaussianity and independence of the shocks, the SVAR model is identified up to a re-scaling and re-ordering of the shocks. Empirical applications of SVAR analysis with ICA identification are flourishing (see, e.g., Capasso and Moneta, 2016; Gouriéroux et al., 2017; Herwartz, 2019; Herwartz and Plödt, 2016; Maxand, 2020; Moneta et al., 2013; Zema, 2022). Several different algorithms for learning independent components from data have been proposed and applied, especially in the fields of blind signal separation, neural networks, feature extraction (see, e.g., Hyvärinen and Oja, 2000), finance (see, e.g., Back and Weigend, 1997), causal inference and structural modeling (see, e.g., Shimizu et al., 2006).

In this paper we evaluate four methods of ICA. Notwithstanding the conspicuous number of ICA estimators emerged in the literature (see Acharya and Panda, 2008), we selectively focus on those that have been proposed or applied within the field of SVAR analysis, namely (1) the *fastICA* algorithm developed by Hyvärinen (1999) and employed by Moneta et al. (2013) and Guerini et al. (2020); (2) the minimization of the distance covariance proposed by Matteson and Tsay (2017); (3) the minimization of the Cramer-von-Mises statistics proposed by Herwartz and Plödt (2016); and (4) the pseudo-maximum likelihood estimator derived by Gouriéroux et al. (2017). Matteson and Tsay (2017), in fact, show how the distance-covariance method outperforms several ICA techniques under several distributional scenarios. Herwartz (2018) also undertakes a performance evaluation analysis, with a focus, however, on the discriminatory power of several identification schemes in detecting structural shocks embedded in a simple DSGE model for the Euro Area. Herwartz et al. (2022) compare different ICA methods with identification by heteroskedasticity. Our paper shares some features with these simulation studies, but it introduces the following novelties. First, in the Monte Carlo studies we vary the distributional framework using a $p-$generalized normal distribution. This allows us to study distributional settings that range from the super-Gaussian to the Gaussian and sub-Gaussian case, in a gradual manner, simply modifying the parameter p. In this manner we can precisely appreciate at which level of closeness to Gaussianity the different ICA estimators fail.

Our motivation to study gradually distinct (non-Gaussian) distributional settings hinges on the fact that many macroeconomic time series, widely analyzed by means of VAR models, tend to display fat-tailed exponential distributions. However, their departure from normality is not always statistically significant as regards, for instance, OECD economies (Fagiolo et al., 2008). Furthermore, it is often the case that different theoretical characterizations of firm dynamics and sectoral linkages move the distribution of economic activity away from the full-Gaussian case (Baqaee and Farhi, 2019; Gabaix, 2011). Finally, the log-transformation, widely implemented when estimating linearized multiplicative models or to stabilize the variance of a time series, is not enough to ensure a Gaussian approximation of the data generating process (Box and Cox, 1982; Lütkepohl and Xu, 2012; Nelson and Granger, 1979).

The second novelty of our paper is the special focus on the four ICA estimators mentioned above, including fastICA, which was considered by Matteson and Tsay (2017) only in comparison with the distance covariance estimator. Finally, we study the average performance of the four ICA methods both in estimating a randomly generated mixing matrix and in recovering a fixed mixing matrix. The latter exercise allows us to get the distribution of the parameter estimates (for each ICA estimator) and to analyze their statistical properties for the sake of hypothesis testing.

The paper is organized as follows: in Section 2 we present the framework of our study, introducing the ICA-based SVAR model and the simulation analysis. In Section 3 we present and discuss the results of our assessment. In Section 4 we discuss an empirical investigation in which the ICA-identified SVAR model is applied to study the effects of fiscal policy (government spending and tax cuts), using the data by Blanchard and Perotti (2002). Section 5 concludes.

2. The framework

2.1. SVAR and ICA

The SVAR model we study has the general form

$$A_0 y_t = c_t + \sum_{\ell=1}^{q} A_\ell y_{t-\ell} + \varepsilon_t, \tag{1}$$

in which q is the lag length, y is a $k \times 1$ vector of endogenous variables, ε_t is a $k \times 1$ vector of exogenous structural shocks $(\varepsilon_{1t}, \ldots, \varepsilon_{kt})'$, A_ℓ is a $k \times k$ matrix of parameters for $0 \le \ell \le q$, c_t is a $k \times 1$ vector of constants, which may also include a deterministic trend. (The model can also be easily extended to include exogenous variables.) We assume the $\varepsilon_{1t}, \ldots, \varepsilon_{kt}$ to be non-normally distributed (with at most one exception) and to be mutually independent, i.e. $f(\varepsilon_{1t}, \ldots, \varepsilon_{kt}) = f(\varepsilon_{1t}) \cdot \ldots \cdot f(\varepsilon_{kt})$, where $f(\cdot)$ is the probability density function. We also assume A_0 to be invertible. The model is structural because it is able to track the effect of statistically independent shocks on the endogenous variables, a crucial feature that makes the researcher able to identify, for example, the effect of a monetary or fiscal policy intervention.

The reduced-form representation implied by the structural model (1) is

$$y_t = d_t + \sum_{\ell=1}^{q} B_\ell y_{t-\ell} + u_t, \tag{2}$$

where $B_\ell = A_0^{-1} A_\ell$ for $1 \le \ell \le q$, $d_t = A_0^{-1} c_t$, $u_t = A_0^{-1} \varepsilon_t$. Thus, we have that the reduced-form residuals $u_t = (u_{1t}, \ldots, u_{kt})$ are linear mixture of the structural shocks ε_t, namely:

$$u_t = B_0 \varepsilon_t \iff \varepsilon_t = A_0 u_t, \tag{3}$$

where $B_0 = A_0^{-1}$. Equation (3) is the model commonly studied in ICA (see, e.g., Hyvärinen, 2013), so that we refer to it as the Independent-Components (IC) model. Using the ICA jargon, we call B_0 the *mixing* matrix, since it linearly mixes the statistically independent components (i.e. shocks) $\varepsilon_{1t}, \ldots, \varepsilon_{kt}$, and A_0 the *unmixing matrix*. Notice that Equation (3) corresponds to the *B*-model in Lütkepohl (2005) (*C*-model in Amisano and Giannini, 1997) but with the further assumption of independence and non-Gausssianity of the shocks.

Let us denote with $a_1', \ldots, a_k'$ the rows of the matrix A_0. Any ICA procedure aims at estimating the k-length weight vectors a_i' for $i = 1, \ldots, k$, which yield $\varepsilon_{1t}, \ldots, \varepsilon_{kt}$ as least dependent as possible. As proved by Comon (1994, Th. 11) and Eriksson and Koivunen (2004, Th. 3) (see also Gouriéroux et al., 2017), the independent components (shocks) are identifiable up to changes in scale (including sign) and ordering. More precisely, the matrix A_0 in the IC model (3) is identifiable up to the left multiplication by PD, where P is a permutation matrix and D a diagonal matrix with non-zero diagonal elements. Equivalently, B_0 is identifiable up to the right multiplication by $D^{-1}P'$ (P' is also a permutation matrix and D^{-1} a diagonal matrix).

ICA algorithms usually consist of two stages: a preliminary whitening and the actual ICA estimation. Whitening the data means to transform them so that they become uncorrelated and with unit variance. Suppose that we have estimated u_t and its non-diagonal covariance matrix Σ_u. Whitening can be obtained through the spectral (also called eigenvalue) decomposition or, as is popular in VAR analysis, via the Choleski factorization of Σ_u. The whitening transformation via the spectral decomposition consists of left multiplying u_t by $(V\Lambda^{1/2})^{-1}$, where V is the matrix containing the eigenvectors of Σ_u, and Λ is a diagonal matrix with the eigenvalues of Σ_u on the main diagonal. Whitening through the Choleski decomposition consists of left multiplying u_t by C, where C is the Cholesky factor of Σ_u (this can be done for any ordering of the variables). Without loss of generality, in this presentation of the ICA methods, we can directly assume that u_t is a vector of uncorrelated random variables (i.e., u_t has already been whitened), so that the matrix B_0 in Equation (3) is orthogonal. Thus, the second stage of ICA estimation reduces to the problem of finding the rotation (orthogonal transformation) of the data u_t that delivers least dependent components ε_t.

We briefly review here four methods for ICA estimation, which we want to comparatively assess. Although further algorithms have been proposed in the literature (see, e.g., Cardoso, 1989; Hyvärinen, 2013), the approaches described below are good representative of the ICA methods discussed and applied in the econometric literature.

1. FastICA. A set of *fast* and *fixed-point* algorithms were proposed by Hyvärinen and Oja (1997, 2000) and Hyvärinen (1999). The *fastICA* approach is based on a fixed-point iteration scheme for finding a maximum of the non-Gaussianity of $a_i'u_t$ (for $i = 1, \ldots, k$). It is called "fast" because it finds the maximally non-Gaussian components with a cubic convergence speed. As a measure of non-Gaussianity *fastICA* adopts an approximation of negentropy $J(x)$, a notion grounded on information theory. For a continuous random variable (or vector) x with density $f(x)$, negentropy is defined as $J(x) = H(x_{gauss}) - H(x)$, where x_{gauss} is a Gaussian random variable (or vector) of the same variance (covariance matrix) of x, and $H(\cdot)$ is the differential entropy, i.e. $H(x) = -\int f(x) \log f(x) dx$. Such measure relies on the fact that a Gaussian random variable entails the largest entropy among all random variables of equal variance (Shannon, 1949). Hyvärinen and Oja (2000) also show that finding the most non-Gaussian directions $a_i'u_t$ (for $i = 1, \ldots, k$) is equivalent to minimizing the Kullback-Leibler divergence between the joint density $f(a_1'u_t, \ldots, a_k'u_t)$ and the product of the marginal densities $f(a_1'u_t) \cdot \ldots \cdot f(a_k'u_t)$, which is a measure of mutual statistical dependence among the $a_i'u_t$'s and is also referred to as *mutual information*.

2. Distance Covariance (*DCov*). Matteson and Tsay (2017) propose to estimate the IC model by finding a matrix of loadings A_0 such that the distance covariance among the $a_i'u_t$'s is minimized. Distance covariance as measure of statistical dependence between random vectors was introduced by Székely et al. (2007).[2] Matteson and Tsay (2017) define an objective function to be minimized in function of θ, which is the vector of angles defining a rotation matrix $G(\theta)$. Thus, the problem consists in finding $\hat{\theta}$ such that the dependence (measured in terms of distance covariance) among the $\varepsilon_{1t}, \ldots, \varepsilon_{kt}$ that results from $G(\hat{\theta})^{-1}u_t$ is minimized. Finally, the mixing matrix B_0 is simply set to be equal to $G(\hat{\theta})$. In this approach, it is convenient to write $G(\theta)$ as the product of $k(k-1)/2$ distinct forms of Givens rotation matrices. In the 2-dimensional case we have only one angle to estimate:

$$G(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{4}$$

In the three dimensional case we have 3 angles

$$G(\theta) = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & 0 & -\sin\theta_2 \\ 0 & 1 & 0 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_3 & -\sin\theta_3 \\ 0 & \sin\theta_3 & \cos\theta_3 \end{bmatrix} \tag{5}$$

For any $k \times k$ matrix we have then $k(k-1)/2$ rotation angles to estimate.

3. Cramer-von-Mises distance (*CvM*). Herwartz and Plödt (2016) and Herwartz (2018), similarly to Matteson and Tsay (2017), define an objective function to be minimized in function of θ and exploit the same decomposition of $G(\theta)$ in Givens matrices. But the minimization criterion is different. The selected vector of angles $\hat{\theta}$, which implies least dependent shocks, is the one that minimizes the value of the Cramer-von-Mises (*CvM*) statistics, developed by Genest et al. (2007). Specifically, this test statistics quantifies the distance between the empirical copula of the shocks vector $\varepsilon_t = G(\theta)^{-1}u_t$ and the implied copula under mutual independence.

4. Pseudo-maximum likelihood estimator (*PML*). This semi-parametric estimation method was proposed by Gouriéroux et al. (2017). It consists of a pseudo maximum likelihood (PML) estimator of the mixing matrix B_0, which maximizes the pseudo log-likelihood function, i.e. $\mathcal{L}_T(B_0) = \sum_{t=1}^{T} \sum_{i=1}^{k} \log g_i(a_i'u_t)$, where $g_i(\cdot)$'s are misspecified probability density functions, exploiting the condition that B_0 is an orthogonal matrix. Gouriéroux et al. (2017) derive the asymptotic properties of the estimator under possible specifications of $\log g_i(\varepsilon_t)$.[3]

2.2. Monte Carlo assessment

We study the performance of the just described ICA methods in estimating the model in Equation (3) with $k = \{2, 3\}$ and $T = \{100, 200, 400\}$, where k is the number of variables and T the sample size. This choice of values for k and T is due to the fact that we want to replicate a VAR model that is as close as possible to those commonly found in applied macroeconomics, where very long time series are seldom available to researchers. We have in any case to restrict the analysis of the *CvM* method to relatively small sample sizes ($T > 500$ is unfeasible, as Herwartz and Plödt (2016) point out), because the computational burden increases by an order of magnitude of $O(T^2kn)$, where n is the number of iterations implemented to generate the distribution of the statistics under the null of independence. We want to evaluate the performances of the ICA methods introduced above when the shocks/independent components in ε_t, and consequently (*a fortiori*) their linear combinations (u_t), get gradually close to be normally distributed. As a matter of fact, it is often the case in empirical applications that the reduced-form residuals of an estimated VAR model turn out to be correlated and non-normal. But it may also be the case that normality of some of the residuals is not fully and clearly rejected, so that the researcher remains doubtful whether an IC model can be legitimately applied for identification.

[2] Let $x^{(1)}$ and $x^{(2)}$ be a d_1- and a d_2-dimensional random vectors. Let $|\cdot|$ denote the Euclidean distance and let $(\dot{x}^{(1)}, \dot{x}^{(2)})$ and $(\ddot{x}^{(1)}, \ddot{x}^{(2)})$ be iid copies of $(x^{(1)}, x^{(2)})$. Székely et al. (2007) define the distance covariance between $x^{(1)}$ and $x^{(2)}$ as $\mathcal{I}(x^{(1)}, x^{(2)}) = E|x^{(1)} - \dot{x}^{(1)}||x^{(2)} - \dot{x}^{(2)}| + E|x^{(1)} - \dot{x}^{(1)}|E|x^{(2)} - \dot{x}^{(2)}| - E|x^{(1)} - \dot{x}^{(1)}||x^{(2)} - \ddot{x}^{(2)}| - E|x^{(1)} - \ddot{x}^{(1)}||x^{(2)} - \dot{x}^{(2)}|$. $\mathcal{I}(x^{(1)}, x^{(2)}) = 0$ if and only if $x^{(1)}$ and $x^{(2)}$ are independent.

[3] Following Gouriéroux et al. (2017), both in the simulation exercise and in the empirical application we assume that the misspecified density function of ε_{it} ($i = 1, \ldots, k$) is a student-t with 5 degrees of freedom if the simulated (or observed) ε_{it} is super-Gaussian; or a mixture of Gaussian distributions in the case in which it displays a sub-Gaussian behavior (Gouriéroux et al., 2017: Table 1, p. 115).

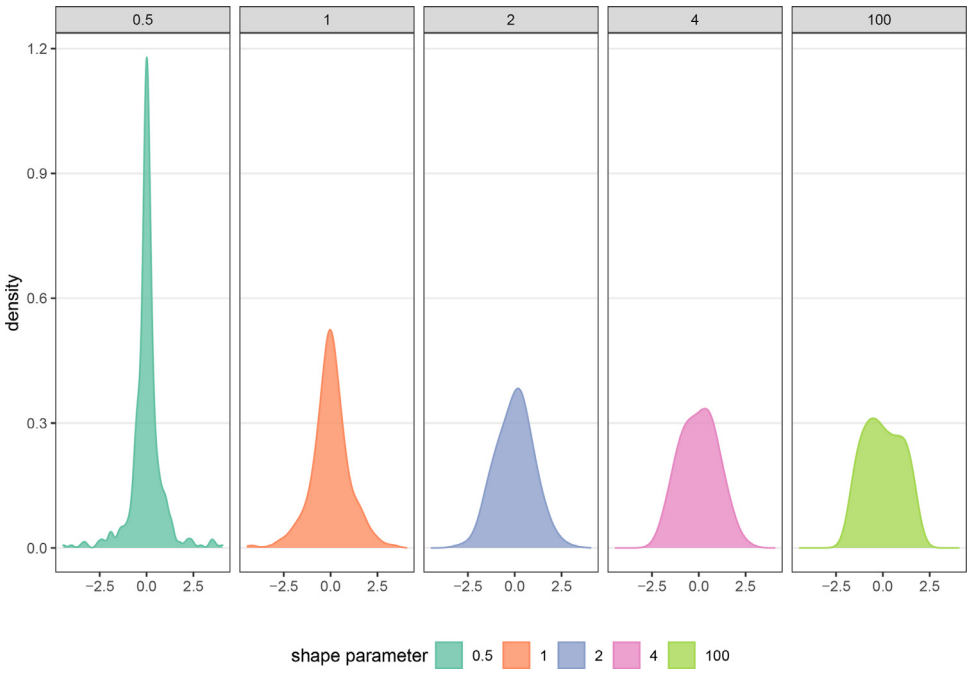

Fig. 1. Kernel density estimates on data generated from the p-generalized normal distribution for different values of the shape parameter p (0.05, 1, 2, 4,100): with $p = 0.5$ super-Gaussian (Laplace), $p = 2$ Gaussian, $p = 100$ sub-Gaussian (uniform).

We perform the analysis by exploiting the properties of a class of exponential distributions, namely the p-generalized normal distribution (Box and Tiao, 1962; Goodman and Kotz, 1973): we let the underlying processes gradually approach to or diverge from a Gaussian distribution. In this manner we can analyze how the ICA procedures behave when the independent components diverge from normality both in the direction of a super-Gaussian (leptokurtic) and a sub-Gaussian (platykurtic) distribution. Often used for robustness studies (Box and Tiao, 1962; Subbotin, 1923; Tiao and Lund, 1970), this family of distributions has also been widely adopted in studies from different fields (e.g. signal processing, audio/video encoding, face recognition, finance), in which data often display non-Gaussian behavior (see Yu et al., 2012, for a review). Following the specification of Kalke and Richter (2013), a p-generalized normal distribution has a density function f of the form:

$$\text{g.norm} = f(x, p) = \frac{p^{1-1/p}}{2\Gamma(1/p)} \exp\left[-\frac{|x|^p}{p}\right] \quad x \in \mathbb{R}, \quad p > 0, \tag{6}$$

where Γ denotes the gamma function and p is a shape parameter that is informative about the rate of decay of the density function. With $p = 2$, $f(x, p)$ is a normal density function. Given this value, as p decreases, the distribution becomes more super-Gaussian, as p increases it becomes more sub-Gaussian. Specifically, with $p = 0.5$, Equation (6) is the probability density function of a random variable with a Laplace distribution (a super-Gaussian distribution), with $p = 100$ it corresponds to the case of a sub-Gaussian distribution, i.e. the uniform distribution. The two limiting cases, $p = 0$ and $p = +\infty$ correspond to a unit impulse function and to a real line, respectively.

In our Monte Carlo experiment, we let the parameter p vary over a range of 20 values, 15 points uniformly located on the interval [0.5,3.5] and 5 on [4,100]: for each of these values we simulate k independent components.[4] As Figure 1 shows, for $0.5 < p < 3.5$ the shape of the distribution changes substantially, while for $p \geq 4$ the sub-Gaussian nature of the distribution is already pretty evident. Finally, in order to better isolate the properties of the different ICA estimators under the distributional scenarios of interest, we focus here on a simple SVAR(0) model, thus avoiding the IC model to inherit the uncertainty that stems from the estimation of VAR model with a richer lag structure. In Appendix A.3, however, we also report results stemming from a simulation exercise based on a SVAR model with one lag.

We split our Monte Carlo experiment into two different designs: a *general assessment* and a *specific assessment*. The *general assessment*, inspired by Matteson and Tsay (2017), aims at measuring the average performance of the four ICA methods in estimating the mixing matrix B_0, across random entries of the same matrix. Each Monte Carlo replication m generates a random IC model $u_t^{(m)} = B_0^{(m)} \varepsilon_t^{(m)}$, where the components $\varepsilon_t^{(m)}$ follow the p-generalized normal distribution (Equation 6) with covariance matrix equal to the identity matrix, and $B_0^{(m)}$ is a random $k \times k$ mixing matrix with condition number

[4] We use the R-package `rpgnorm` (Kalke, 2015).

$1 \leq \mathcal{K}(B_0^{(m)}) \leq 2$, simulated with the R-package `ProDenICA` (Hastie and Tibshirani, 2010).[5] Given the indeterminacy of the IC model, namely its identification only up to column scale/sign and permutation of B_0 (as mentioned in the previous subsection), Matteson and Tsay (2017) suggest to measure the performance of the ICA methods with a metric, proposed by Ilmonen et al. (2010), that is invariant to this indeterminacy. Such measure is defined as follows:

$$D(B_0^{(m)}, \widehat{B}_0^{(m)}) = \frac{1}{\sqrt{k-1}} \inf || C \widehat{B}_0^{(m)-1} B_0^{(m)} - I_k ||_F \tag{7}$$

where $B_0^{(m)}$ is the random matrix generated at replication m, $\widehat{B}_0^{(m)}$ is its estimate, $C = P_{\pm} D_+$, where $P_{\pm}$ is any $k \times k$ signed permutation matrix and D_+ is any $k \times k$ diagonal matrix with strictly positive diagonal element, and $||.||_F$ is the Frobenius norm.[6] The lower the index, the closer the estimate $\widehat{B}_0^{(m)}$ to the true value $B_0^{(m)}$. We refer to this measure as the Minimum Distance Index (MDI).

The *specific assessment* aims at evaluating how well the ICA methods perform in identifying the structural impulse responses of a SVAR model, using different realizations of the same data generating process (dgp), for a given mixing matrix B_0. This also allows us to compare, among each other, the Monte Carlo distributions of the parameter estimates derived by the four ICA methods and to analyze their statistical properties. The chosen mixing matrices for $k = \{2, 3\}$ are, respectively:

$$B_0 = \begin{bmatrix} 1.14 & -0.38 \\ 0 & 1.26 \end{bmatrix}, \quad B_0 = \begin{bmatrix} 0.9 & 0.15 & 0.65 \\ -0.75 & 1.13 & 0.22 \\ 0.21 & -0.53 & 1.5 \end{bmatrix}. \tag{8}$$

Thus, in the case of two variables, we have an essentially triangular mixing matrix, while for $k = 3$ we have a full matrix (all non-zero entries). This allows us to cover both a recursive and non-recursive mechanism of shocks' transmission. As mentioned, an ICA algorithm, which delivers a mixing (or unmixing) matrix, is not sufficient for full identification, since the mixing matrix is identified up to the right multiplication of DP (D is any diagonal matrix with all non-zero in the main diagonal and P is any permutation matrix). While the matrix D can be appropriately chosen so that the each of the shocks/components ε_{it} in ε_t have variance one and impact positively on a specific variable y_{jt} in y_t (for i and j in $1, \ldots, k$), there is not a general and unique method to choose P. Having estimated with an ICA algorithm the matrix $\widehat{B}_0$, there are different options:

1. One option is to apply the LiNGAM (Linear Non-Gaussian Acyclic Model) algorithm, as proposed by Shimizu et al. (2006), Hyvärinen et al. (2010) and Moneta et al. (2013). The permutation matrix P is found by assuming that the underlying contemporaneous causal structure is recursive (acyclic). This implies that $\widehat{B}_0$ contains at least $k(k-1)/2$ entries equal to 0 (but measurement errors) and that there exists only one P such that the main diagonal of $B_0 DP$ has all entries significantly different from 0.[7] The LiNGAM algorithm searches in the space spanned by all the possible permutation matrices the matrix $\tilde{P}$ which minimizes a cost function that penalizes small absolute values in the main diagonal of $\tilde{P} D \widehat{B}_0^{-1}$. Furthermore, it searches for a permutation matrix $\tilde{\tilde{P}}$ such that $\tilde{\tilde{P}} \tilde{P} D \widehat{B}_0^{-1} \tilde{\tilde{P}}'$ is maximally close to a lower triangular matrix and sets its upper diagonal elements equal to zero.

2. The LiNG algorithm was formulated by Lacerda et al. (2008) (see Ciarli et al., 2019, for an application) as a variant of LiNGAM algorithm which relaxes the recursiveness (acyclicity) assumption, but still exploits the possibility that some entries of the unmixing matrix are equal to zero. Having tested (e.g., by bootstrap sampling) that some entries are vanishing, LiNG finds the set of permutation matrices $\mathcal{C}$, such that for each $\tilde{P} \in \mathcal{C}$, $\tilde{P} D \widehat{B}_0^{-1}$ does not contain zeros in the main diagonal. The permutation matrices satisfying this property, however, are, in general, not unique, so that the indeterminacy problem of ICA is not uniquely solved.

3. Having estimated $\widehat{B}_0$, and having re-scaled its columns by D, such that the shocks $D^{-1} \widehat{B}_0^{-1} u_t$ have unit variance,[8] another option is to apply what we call the *Maxfinder* criterion. This corresponds to find the permutation matrix $\tilde{P}$ such that each column of $\widehat{B}_0 D \tilde{P}$ has off-diagonal elements that are smaller than the diagonal element (in absolute values, columnwise). More rigorously, the permutation matrix P is such that the entries of $C = (c_{ij}) = \widehat{B}_0 D \tilde{P}$ satisfy $|c_{ii}| > |c_{ji}|$, for all $i \neq j$ ($i, j = 1, \ldots, k$). This criterion is guaranteed to give a (unique) result only under the assumption that the shock ε_{it} affects the endogenous variable y_{it} more than other variables y_{jt} ($i \neq j$), within the sample period. This fact would justify *labelling* the shock ε_{jt} as the y_{jt}-shock (e.g. the shock to tax revenue as the tax shock). In other terms, the *Maxfinder* criterion delivers a (unique) result only if the maxima (in absolute values) of the columns of $\widehat{B}_0 D$ lie in distinct rows. If this is not the case, for instance when a row contains the maximum entry (in absolute value) of column 1 and 2, an option to solve this problem is to apply the *Maxfinder* criterion in a hierarchical manner,[9] as suggested by Bruns et al. (2021).

---

[5] The condition number of a matrix B, $\mathcal{K}(B)$, measures the "well-behavior" of B, namely the extent to which the solution x of the linear system $Bx = c$ changes with the respect to changes in c (see e.g. Horn and Johnson, 2012, ch. 5.8) If $\mathcal{K}(B) = 1$ the matrix B is said to be *perfectly conditioned*.

[6] A signed permutation matrix is like a permutation matrix, with exactly one non-zero element for each row and column, but its non-zero elements are +1 or -1.

[7] Equivalently, for any diagonal matrix D, there exists only one P such that the main diagonal of $PD\widehat{B}_0^{-1}$ has all entries significantly different from 0.

[8] This re-scaling may be irrelevant because ICA estimators usually deliver components with unit variance.

[9] This hierarchical application of *Maxfinder* delivers a unique solution: (i) If the maximum entry (in absolute value) of $\widehat{B}_0 D$ lies in position (row, column) (i, j), then the j^{th} column becomes the i^{th} column in $\widehat{B}_0 DP$, for a permutation matrix P. (ii) Repeat the same procedure starting from the output of the

4. As mentioned in Berner et al. (2022), one can label the shocks identified by ICA using a criterion derived by the estimation of the forecast error variance decomposition. Following this approach, one finds a permutation matrix P such that the shocks $\varepsilon_t = PD^{-1}\widehat{B}_0^{-1}u_t$ have the following characteristics: the prediction mean squared error (MSPE) of variable $y_{1,t+h}$ is maximally accounted for by shock ε_{1t}, the MSPE of $y_{2,t+h}$ is maximally accounted for by ε_{2t}, etc. This can be done for a specific time horizon h or summing up the contributions of the same shock across different time horizons (until a specific value of h).

5. Another criterion proposed in the literature is the sequence of transformations considered by Lanne et al. (2017), which imposes the entries of $C = (c_{ij}) = \widehat{B}_0 D_1 P$ to satisfy $|c_{ii}| > |c_{ij}|$ for all $i < j$, where D_1 is here a diagonal matrix that makes each column of $\widehat{B}_0 D_1$ have Euclidean norm one. This criterion has the advantage of providing a scheme that always deliver a unique permutation matrix P from an ICA-estimated matrix $\widehat{B}_0$.[10]

The list proposed above does not certainly exhaust the class of possible criteria for choosing the permutation matrix P. It is also conceivable to apply a mixture of different criteria: for example one can apply the *MaxFinder* criterion and check whether it is consistent with the forecast error variance decomposition criterion (see Berner et al., 2022).

In our *specific assessment* we choose P according to the *MaxFinder* criterion, avoiding the recursiveness assumption and, at the same time, exploiting the fact that the mixing matrix in the underlying dgp (see Equation 8) does not contain column maxima (in absolute values) on the same row. This allows us to focus on the identification of the single coefficients. In the empirical application (Section 4), the assumption of column maxima on the same row is not guaranteed, so that in case of its failure, which is easily verifiable, we apply the step of LiNGAM (step 2 in the original algorithm by Shimizu et al., 2006), in which it is found a permutation matrix P that minimize the quantity $1/\sum_i |w_{ii}|$ (penalizing small absolute values in the main diagonal), where w_{ii} is the (i,i) entry of the matrix $W = PD\widehat{B}_0^{-1}$. We also check for sensitivity of our results on the permutation criterion, by repeating both the specific assessment and the empirical application using the criterion discussed in the point 5 here above (these results are reported in Appendix A.2).

3. Results

3.1. General assessment

Results about the performance of the four ICA estimators in the general assessment are shown in Figure 2. For each estimator, this figure displays the MDI, as specified in Equation (7), across two different values of k ($k = 2$ and $k = 3$) and 20 values of p, fixing the sample size $T = 400$.[11] The thick lines in the plots trace the average MDI, calculated across 1000 distinct data generating processes, while the shadow areas display its (one standard deviation) variability. The performance of the four ICA methods is shown as the full Gaussianity ($p = 2$) of the independent structural shocks is approached. As expected, when the ε_t's are close to being normally distributed, all the methodologies score bad. To establish a negative benchmark, we calculated MDIs between a fixed B_0 and 1000 random matrices of the same dimension (respecting the condition number as specified above). The 90% of these MDIs are above the dashed lines in each plot in Figure 2, which can therefore be considered as negative benchmark. For values of p close to Gaussianity, MDI averages are above (i.e. score worst) this benchmark and the volatility of index increases.

On average, whatever the dimension of the system, *fastICA* and *PML* score relatively equal with the former slightly outperforming the latter when the ε's are super-Gaussian. On the other hand, the opposite is true when the independent components (herafter, ICs) are close from being normally distributed and sub-Gaussian. In such distributional scenario, *DCov* loses precision and scores worse than any other method. *CvM* is dominated in almost all scenarios, performing better only when compared to *DCov* in the highly sub-Gaussian scenario ($p > 4$). The results are compatible with and complementary to those of Matteson and Tsay (2017), where the ICs follow different families of distributions, both symmetric and asymmetric. Here instead we are interested in understanding the performances of ICA as we get closer to the full-Gaussian case. However, the exercise highlights that when the ICs are sub-Gaussian, *DCov* tends to have a lower performance.

3.2. Specific assessment

Having analyzed the average performance of the four ICA methods, we turn now to study how well they perform when they are applied to recover the impact multiplier (mixing) matrix of a SVAR model. As mentioned in Section 2.2, we focus on a specific data generating process. We artificially generate data u_t from the IC model in Equation (3) with shocks' covariance matrix $\Sigma_\varepsilon = I$, $c_t = 0_k$, and $A_0 = B_0^{-1}$, where B_0 is specified in Equation (8) for $k = 2, 3$. Both specifications of B_0 satisfy the condition that the maximum entry (in absolute value) of each column never appears on the same row. This assumption

previous step, $\widehat{B}_0 DP$, but compute the maximum entry (in absolute value) neglecting all the entries lying in the i^{th} row and column of $\widehat{B}_0 DP$, until the k^{th} column has been permuted.

[10] Note, however, that if $u_t = C\varepsilon_t$ satisfy $|c_{ii}| > |c_{ji}|$ for all $i < j$, it is not necessarily the case that the matrix $C_1 = P_1 C P_1^{-1}$ satisfy the same criterion (for some permutation matrix P_1), so that we may have: $P_1 u_t = C_1 P_1 \varepsilon_t$. This implies that if this criterion is used for shock labelling, one may get two different results for u_t and $P_1 u_t$ (same variables, but different order).

[11] Results do not change qualitatively when the sample size decreases to $T = \{100, 200\}$. As expected though, the precision decreases as the sample size gets smaller (see Figure 9 in Appendix A.1.)

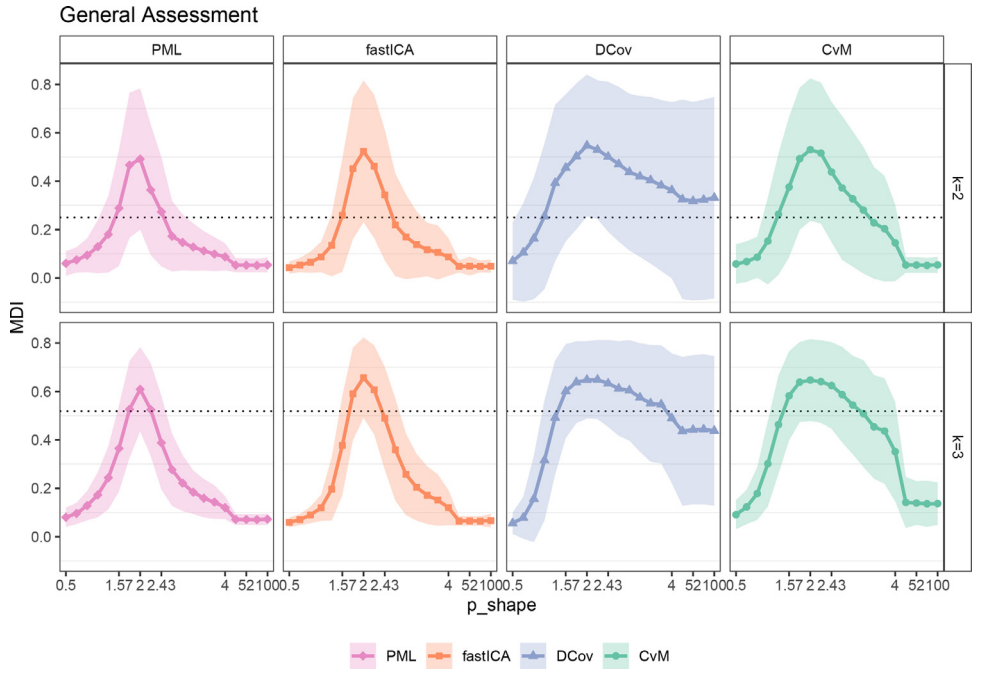

Fig. 2. Each plot shows the MDI (thick lines with marks: averages across 1000 dgps), as specified in Equation (7), for each of the four ICA estimators, across 20 values of p. The p-generalized normal distribution is super-Gaussian with $p < 2$, Gaussian with $p = 2$, sub-Gaussian with $p > 2$. Plots in the upper part of the figure correspond to the case $k = 2$, in the lower part to $k = 3$. Sample size $T = 400$. Shadow areas show one-standard deviation above and below the mean. Dashed lines constitute the negative benchmark: 90% of random estimates yield MDIs above this line.

captures a feature that is often found in SVAR analysis, namely the fact that each structural shock ε_{jt} $(j = 1, \ldots, k)$ tends to be mostly (contemporaneously) correlated with the variable (y_{jt}) that provides the label to the shock.

We apply the four ICA methods to the generated u_t. We then apply the *MaxFinder* identification algorithm described in Section 2.2 to solve the indeterminacy that affects the IC model. We then assess the performance of the IC model in this specific exercise by using a variation of MDI: differently from Equation (7), the index we use here measures the error between $\widehat{B}_0$ and the known matrix B_0 (this time fixed over Monte Carlo iterations, as specified in Equation 8). Note that $\widehat{B}_0$ is here column-permuted and scaled according the *MaxFinder* criterion. Hence, the distance index is now defined as

$$D(B_0, \widehat{B}_0) = \frac{1}{\sqrt{k-1}} ||\widehat{B}_0^{-1} B_0 - I_k||_F \tag{9}$$

Figure 3 shows the MDI, as specified in Equation (9), averaged across Monte Carlo replications, for different values of the parameter p, which determines the degree of Gaussianity of the ICs ε_t, fixing the sample size $T = 400$ (see Appendix A.1, Figure 10, for $T = 100, 200$). To some extent, the results resemble those shown for the "general assessment." The performances of *fastICA* and *PML* are almost equivalent, with the difference that *fastICA* performs slightly better when the ICs are super-Gaussian and *PML* performs slightly better when the ICs are sub-Gaussian. One striking result is that the average value of the MDI for *DCov* is remarkably lower than for other estimators, but only for the two-dimensional case. Indeed, when $k = 3$, the behavior of the performance is very similar to the *general assessment*: MDI is highly volatile and its average does not improve significantly when the distribution of the ICs becomes sub-gaussian. The estimates of the mixing matrix under *CvM* display the largest variance among the estimators under analysis in most of the distributional scenarios that we have explored.

Although there are several measures of ICA performance proposed in the literature (see Nordhausen et al., 2011, for a review), those measures, like the one proposed in Equations (7) and (9), are not informative about the distributional properties of the parameters' estimates. As for any estimator, such properties are relevant when performing statistical inference: in our specific design, they are informative about the distribution of the entries of the mixing matrix B_0. Ultimately, these distributional properties may shed light on the contemporaneous causal relationships among the endogenous variables of the VAR system. Therefore, we study the Monte Carlo distribution of the differences between the entries $\hat{b}_{ij}$ of the estimated mixing matrix $\widehat{B}_0$ and the entries b_{ij} of the known matrix B_0 (for $i, j = 1, \ldots, k$).

Table 1 reports mean and standard deviation (across Monte Carlo runs) of $\hat{b}_{ij} - b_{ij}$, in four representative scenarios: (i) when the independent components ε_t have a strong super-Gaussian behavior $(p = 0.5)$; (ii) when Gaussianity is closer but the ε_t are still either super-Gaussian $(p = 1.5)$ or (iii) sub-Gaussian $(p = 2.5)$; (iv) when the ε_t follow an almost uniform distribution $(p = 100)$. We consider both $k = 2$ (columns 1-8) and $k = 3$ (columns 9-16). For $k = 3$, only the upper left

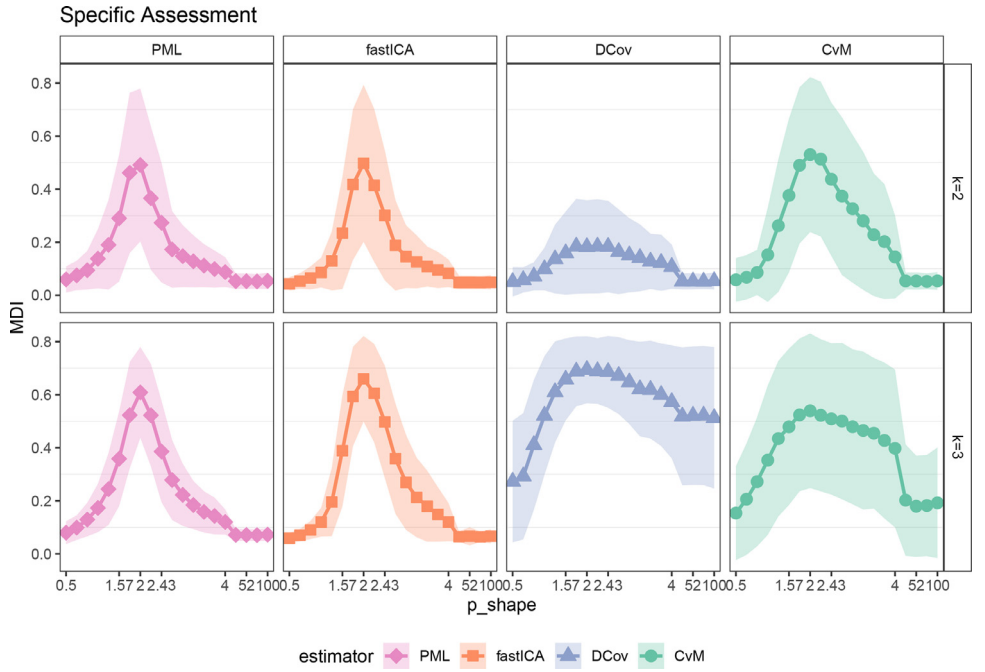

Fig. 3. Each plot shows the MDI (thick lines with marks: averages across 1000 Monte Carlo replications), as specified in Equation (9), for each of the four ICA estimators, across 20 values of p. The mixing matrix in the dgp is fixed as in Equation (8). The p-generalized normal distribution is super-Gaussian with $p < 2$, Gaussian with $p = 2$, sub-Gaussian with $p > 2$. Plots in the upper part of the figure correspond to the case $k = 2$, in the lower part to $k = 3$. Sample size $T = 400$. Shadow areas show one-standard deviation above and below the mean.

Table 1
Summary statistics of the errors between entries of $\widehat{B}_0$ and of B_0 (the latter as specified in Equation 8). Sample size $T = 400$.

		$k=2$								$k=3$								
		$p=0.5$		$p=1.57$		$p=2.47$		$p=100$		$p=0.5$		$p=1.57$		$p=2.47$		$p=100$		
		mean	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	
PML	$\hat{b_{11}} - b_{11}$	-0.002	0.007	-0.045	0.077	-0.039	0.068	-0.001	0.002	-0.005	0.042	-0.032	0.145	-0.025	0.147	-0.004	0.032	
fastICA		-0.001	0.001	-0.032	0.067	-0.049	0.081	-0.001	0.001	-0.002	0.027	-0.020	0.144	-0.007	0.165	-0.003	0.031	
DCov		-0.001	0.006	-0.015	0.034	-0.018	0.037	-0.001	0.002	0.032	0.138	0.116	0.160	0.135	0.149	0.066	0.146	
CvM		-0.003	0.018	-0.076	0.106	-0.093	0.112	-0.001	0.002	-0.001	0.044	-0.014	0.177	-0.005	0.177	-0.007	0.064	
PML	$\hat{b_{12}} - b_{12}$	-0.002	0.065	-0.029	0.308	-0.035	0.287	0.001	0.051	-0.002	0.063	-0.067	0.310	-0.102	0.330	0.000	0.052	
fastICA		0.000	0.038	-0.016	0.261	-0.044	0.317	0.001	0.046	0.000	0.041	-0.087	0.329	-0.144	0.384	0.000	0.047	
DCov		0.001	0.051	-0.009	0.179	-0.017	0.200	0.001	0.052	-0.057	0.234	-0.179	0.306	-0.169	0.282	-0.127	0.314	
CvM		0.001	0.079	-0.080	0.386	-0.112	0.421	0.001	0.052	-0.002	0.073	-0.186	0.433	-0.199	0.421	-0.013	0.108	
PML	$\hat{b_{21}} - b_{21}$	0.002	0.068	0.048	0.349	0.054	0.327	0.001	0.054	0.002	0.061	0.165	0.393	0.220	0.429	0.002	0.051	
fastICA		-0.001	0.046	0.030	0.296	0.067	0.364	0.001	0.049	0.000	0.043	0.214	0.426	0.352	0.484	0.002	0.046	
DCov		-0.002	0.077	0.015	0.205	0.028	0.231	0.002	0.054	0.164	0.329	0.458	0.371	0.458	0.344	0.388	0.361	
CvM		-0.001	0.094	0.115	0.445	0.156	0.485	0.002	0.056	0.005	0.075	0.457	0.523	0.510	0.503	0.014	0.114	
PML	$\hat{b_{22}} - b_{22}$	-0.001	0.021	-0.041	0.110	-0.032	0.097	-0.002	0.017	-0.002	0.041	-0.025	0.142	-0.022	0.148	-0.003	0.035	
fastICA		-0.001	0.013	-0.030	0.091	-0.040	0.106	-0.001	0.016	-0.002	0.026	-0.025	0.144	-0.017	0.157	-0.003	0.031	
DCov		-0.003	0.022	-0.014	0.062	-0.016	0.074	-0.001	0.017	0.011	0.090	0.105	0.134	0.115	0.130	0.039	0.103	
CvM		-0.004	0.033	-0.056	0.119	-0.065	0.131	-0.001	0.018	-0.003	0.045	-0.023	0.164	-0.011	0.174	-0.008	0.066	

block's entries are considered. In almost all the cases, the estimates are negatively biased (a result compatible with those of Gouriéroux et al., 2017). As expected, in the scenarios where Gaussianity is closer, the bias is slightly more negative and the uncertainty of the estimates gets larger; the same holds when the dimension of the system ($k = 3$) increases. For almost all parameters' estimates, in the extreme distributional scenarios ($p = 0.5, 100$), all the methods score relatively equal; *fastICA* tends to have the smallest bias and, together with *DCov*, the smallest variance. In the cases closer to Gaussianity, *DCov* performs relatively better when the dimension of the system is smaller, whilst *fastICA* seems to deliver relatively better results when dimension of the system increases; *CvM* scores better in very few cases.

Finally, we evaluate the performance of the four ICA methods when statistical inference is conducted on the basis of a bootstrap procedure. Specifically, we compare bootstrap-based inference with the results derived from the pseudo-maximum likelihood approach. We expect that the *PML* estimator, derived by Gouriéroux et al. (2017) and drawn on asymptotic approx-

imations, outperforms the bootstrap procedure, which we adopt since we do not know the asymptotic distribution of the estimates of the B_0's entries under *fastICA*, *DCov* and *CvM*. This should happen at least for large sample sizes. A small number of observations, however, may favour inference not based on asymptotic properties. Moreover, the comparison is interesting because it allows us to assess, from an another perspective, which method is more robust when the full-Gaussianity case is approached, in which, as stressed throughout the paper, ICA methods fail to recover the independent components.

The exercise is conducted in the following way. Again, we run a Monte Carlo exercise where we generate data from IC models with mixing matrices as specified in Equation (8) and with shocks distributed according a p-generalized normal distribution, for different values of p and sample size $T = 400$. For each Monte Carlo replication $m = 1, \ldots, 1000$ we estimate $\widehat{B}_0$ using one of the four ICA methods with the *MaxFinder* criterion.

For the *PML* method we derive $\hat{\sigma}_{ij}$, i.e the estimate of the standard deviation of $\hat{b}_{ij}$ (for $i, j = 1, \ldots, k$), from the asymptotic covariance matrix of the mixing matrix found in Gouriéroux et al. (2017). On the basis of this, we build the $(1-\alpha)$-confidence interval

$$\mathcal{C}_m(\alpha) = [\hat{b}_{ij} - \phi_{\alpha/2}\hat{\sigma}_{ij}, \;\; \hat{b}_{ij} + \phi_{\alpha/2}\hat{\sigma}_{ij}], \tag{10}$$

with $\phi_{\alpha/2}$ being the $\alpha/2$-quantile of the standard normal distribution. We do this for each Monte Carlo replication m, which corresponds to a specific estimate $\widehat{B}_0$, with specific entries $\hat{b}_{ij}$ (which we do not to index here with m just for simplicity) and corresponding confidence intervals $\mathcal{C}_m(\alpha)$, and for each i, j in $1, \ldots, k$. Finally, we calculate the frequency at which the true value b_{ij} (entries of the matrices specified in Equation 8) falls in $\mathcal{C}_m(\alpha)$ across Monte Carlo replications. We should expect that the frequency of observing the true value b_{ij} in the confidence interval is equal to $1-\alpha$.

As mentioned above, for the other three ICA methods we do not know the asymptotic distribution and therefore a bootstrap approach is necessary. Given the computational constraints of the exercise, we implement the *warp*-bootstrap, proposed by Giacomini et al. (2013), where it is shown that it is sufficient to have one bootstrap replication for each Monte Carlo run to obtain a reliable approximation of the statistics under analysis. The confidence interval is then built, for each i, j in $1, \ldots, k$, in the following way:

$$\mathcal{C}_m(\alpha) = [\hat{b}_{ij}^{(m)} - \hat{q}_{ij}(\alpha/2), \;\; \hat{b}_{ij}^{(m)} - \hat{q}_{ij}(1-\alpha/2)], \tag{11}$$

in which $\hat{q}_{ij}(\alpha)$ is the α-quantile of the empirical distribution (across Monte Carlo runs) of $\hat{b}_{ij}^{(m)*} - \hat{b}_{ij}^{(m)}$, where $\hat{b}_{ij}^{(m)*}$ is the estimate of the (i, j) entry of B_0 obtained at the (unique) bootstrap draw corresponding to the Monte Carlo run m, and $\hat{b}_{ij}^{(m)}$ is the estimate of the (i, j) entry of the mixing matrix B_0 at the Monte Carlo run m. We then compute the frequency at which $\mathcal{C}_m(\alpha)$ contains the true value b_{ij} across the Monte Carlo replications. If the bootstrap procedure is consistent, we should then expect that the frequency at which the bootstrap-based confidence interval contains the true value (i.e. its the empirical coverage $1-\hat{\alpha}$) to be exactly equal to $1-\alpha$, the nominal coverage.

Pooling all the entries of the mixing matrices' estimates of Equation 8, Figure 4 shows the distribution of the size distortions, measured by the ratio between the empirical coverage $1-\hat{\alpha}$ and its nominal counterpart $1-\alpha$, of each estimator for values of $(1-\alpha) = \{0.99, 0.95, 0.90, 0.75, 0.50\}$. The first result is that bootstrap-based confidence intervals built under *fastICA* display the smallest size distortion, notwithstanding the dimension k of the system. Surprisingly, this is true even when compared with *PML*-based inference, especially in the case in which the independent components follow sub-Gaussian distributions ($p = \{2.43, 100\}$). However, as stated in Gouriéroux et al. (2017), the choice of the pseudo-likelihood does indeed matter for the asymptotic accuracy of the *PML* estimator. Furthermore, all estimators but *CvM*, which also displays the worst performance, tend to contain much more often than expected the true value of the parameter ($\hat{\alpha} < \alpha$): this distortion decreases as the nominal size α increases. We repeat the same simulation exercise using the permutation criterion by Lanne et al. (2017) instead of *Maxfinder*, as well as using a SVAR(1) instead of a SVAR(0) model. These results are reported in Appendix A.2 and Appendix A.3, respectively.

4. Empirical application

In this section we discuss a macroeconomic application of the ICA approach to SVAR analysis with the aim of showing its potentials and challenges, having taken into account the results of the previous section. ICA offers the opportunity to statistically test identifying restrictions (not limited to those that are over-identifying) on the coefficients of the impact (mixing) matrix. We consider here the very influential work on fiscal policy by Blanchard and Perotti (2002), BP henceforth. Given the recent reappraisal of fiscal policy, the question on the size of multipliers continues to be highly disputed (see, e.g., Mountford and Uhlig, 2009; Ramey, 2011). BP estimate a three-variable VAR model of public spending, tax revenues and aggregate output. The SVAR model is identified through assumptions based on institutional knowledge: public spending does not respond to output in the quarter, while tax revenues do. Moreover, BP set the contemporaneous response of taxes to output on the basis of an outside estimate of the cyclical sensitivity of net taxes. Finally, they impose two alternative restrictions on the contemporaneous relationship between tax revenues and public spending, corresponding to two different models: in the first model a tax shock has an immediate effect on spending, but a spending shock does not have an immediate effect on tax revenues (except an indirect one through GDP), in the second model the other way around.

In spite of the soundness of the identification strategy, it is important to evaluate the plausibility of such restrictions, at least for two reasons: (i) as Caldara and Kamps (2017) show, the use of different estimated elasticities may lead to

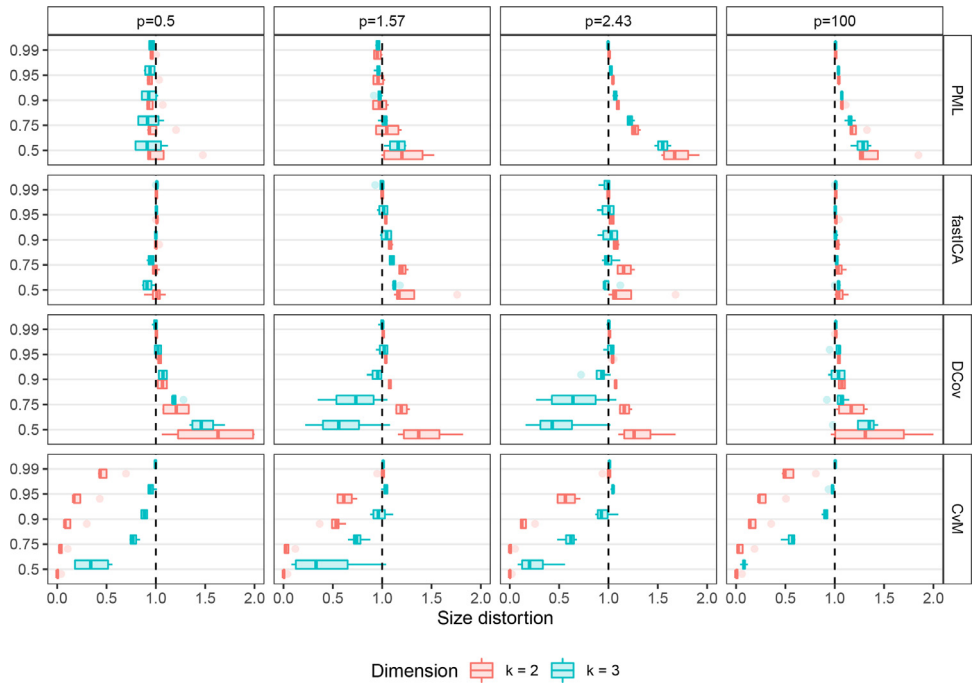

Fig. 4. Distribution of size distortions (measured on the $x-$ axis by the ratio between the empirical coverage $1-\hat{\alpha}$ and its nominal counterpart $1-\alpha$) in drawing confidence intervals for the estimates of the mixing matrix entries, for different estimators (row-wise ordered panels), different distributional scenarios (column-wise ordered panels), when the dimension of the system k changes. Confidence intervals are constructed at nominal α significance level. For *PML*, we use the asymptotic approximation derived by Gouriéroux et al. (2017), while for *fastICA*, *DCov* and *CvM* we implement the *warp*-bootstrap procedure described in text.

Table 2
Impact coefficients in Blanchard and Perotti (2002).

G ordered first				Tax ordered first		
	ε^{g}	ε^{tax}	ε^{gdp}	ε_t^{g}	ε_t^{tax}	ε^{gdp}
G	1.00	0.00	0.00	1	-0.05	0.00
Tax	0.16	1.00	2.18	0.33	1.00	2.18
GDP	0.18	-0.15	1.00	0.15	-0.16	1

dynamic responses and fiscal shocks that significantly differ in size and persistence; (ii) the authors are not able to identify the contemporaneous relationship between government spending and tax revenues and, consequently, to check whether a tax shock has an immediate effect on spending. Our empirical exercise is similar to the recent study by Karamysheva and Skrobotov (2022), who exploit the non-Gaussianity of the underlying reduced-form residuals to estimate and identify the same VAR as BP, using the GMM estimator proposed by Lanne and Luoto (2021). Differently from this study, however, our exercise tests whether the identifying restrictions proposed by BP are rejected through our inference procedure. Moreover, we study to what extent results change when we use different ICA estimators, whose performance has been evaluated in the previous sections.

Table 2 shows the impact coefficients, derived in BP, under the unit normalization of the direct contemporaneous effects, i.e. the impact (mixing) matrix has been normalized so that it displays only ones on the main diagonal. Since, as mentioned, BP use two different models in function of the different contemporaneous impacts between spending and taxes, we display both of them in the table. Note that, in terms of zero-entries, the only difference between the two models is the (1,2) entry. In the second model (right panel) the tax shock has a non-zero effect on spending but the spending shock has still a non-zero effect on tax because of the causal chain from *G* to *Tax* via *GDP*.

We estimate a three-variables VAR model using the same data as in BP (quarterly US data 1960-1997). The objective of the exercise is twofold: (i) using our *Maxfinder* algorithm, to globally identify via the IC model the independent shocks on government spending and tax increase and estimate their dynamic responses on economic activity; (ii) to assess the validity of the restrictions proposed by BP, by tests of hypothesis on the single coefficients, relying on the statistical inference procedure studied in Section 3.2.

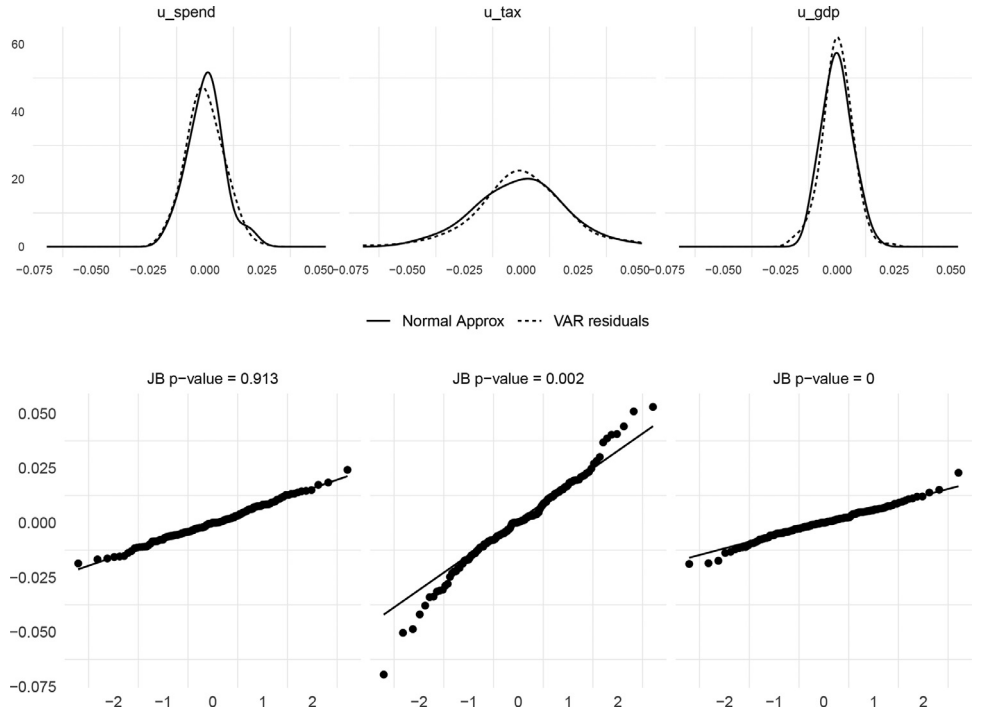

Fig. 5. Distribution (upper panels) and q-q plots (bottom panels) of reduced-form residuals, where also p-values of the Jarque-Bera (JB) test are reported.

For the analysis, we estimate a reduced-form VAR model analogous to Equation (2), namely

$$y_t = C_t\Gamma + \sum_{l=1}^{q} B_l y_{t-l} + B_l^Q (Q1 + Q2 + Q3) y_{t-l} + u_t, \tag{12}$$

where y_t is the $k \times 1$ vector of endogenous variables, C_t is $k \times p$ vector of deterministic terms with coefficients Γ ($p \times 1$), Q1, Q2, Q3 are quarter dummies, B_l and B_l^Q are $k \times k$ coefficient matrices, and u_t is k-dimensional vector of reduced-form residuals. In our case $k = 3$ and $q = 4$, as in BP.[12] In line with the setting of the IC model, we assume that the reduced-form residuals are a linear combination of statistically independent components, as specified in Equation (3). As in BP, we have

$$y_t = (G_t, TAX_t, GDP_t)' \tag{13}$$

where y_t is a vector that contains the logarithm of real per capita values of government spending, taxes and GDP, observed in U.S. from 1960:Q1 to 1997:Q4. The estimated VAR process results to be stable with no-serial autocorrelation. Moreover, estimated reduced-form residuals show departures from normality (see Figure 5), with the exception of the government-spending residual, u_t^G, for which the Jarque-Bera test does not reject the null hypothesis of normality. Non-Gaussianity of the residuals allows application of ICA, but are not *per se* a test of the requirements of the IC model, namely non-Gaussianity and independence of the *structural shocks* (with at most one exception).[13]

Using the estimated reduced-form residuals, we estimate mixing matrices and structural shocks by applying the four ICA methods studied in this paper. We perform, on the estimated shocks, both Jarque-Bera tests of normality and kernel test of independence, using the Hilbert-Schmidt independence criterion (Gretton et al., 2007). By rejecting normality (with the exception of one shock as regards *CvM*) and not rejecting independence (significance level: 0.05), the requirements of the IC model turn out to be fulfilled as regards *fastICA*, *DCov*, and *CvM*. However, this is not be the case for *PML*, where the estimated shocks turn out to be Gaussian (with a possible exception at 0.1 level of significance) and independent. The results on *PML* should be therefore taken with caution, but we report them for the sake of comparison.

Table 3 reports the estimates $\tilde{b}_{ij}$ of the entries b_{ij} of the mixing matrix B_0, for each ICA method, which has been transformed by applying the *MaxFinder* scheme and further normalised such that the entries of the main diagonal are set equal to

[12] The series of p deterministic terms include a constant, linear and quadratic time trend, quarter dummies plus a current and four-period lagged dummy for 1974:Q2 (when a large tax cut has been observed). When estimating the IRFs instead, closely following BP, we drop from the specification the non-linear terms $Q_\bullet\, y_{t-l}$. The purpose of this exercise is to drop any serial correlation from the residuals, upon which the structural analysis is conducted.

[13] Indeed, it is in principle possible that a k−length vector of shocks with a number of non-Gaussian shocks between 1 and $k-2$ produces a k−length vector of (reduced-form) residuals, among which $k-1$ or all of them are non-Gaussian. We thank an anonymous referee for pointing this out.

Table 3
Estimates of the contemporaneous impacts coefficients from IC models. The entries of the mixing matrices are calculated using the bootstrap-median with the corresponding confidence intervals and are normalized so that each structural innovation has a unit contemporaneous impact on the log of the variable that refers to.

	PML			*fastICA*			*DCov*			*CvM*		
	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
G_t	1	-0.036***	-0.197***	1	0.006	-0.061	1	0.011	0.05	1	0.013	-0.745
Tax_t	0.955***	1	3.908***	0.642*	1	1.122***	0.442	1	1.146**	0.839*	1	1.393*
GDP_t	0.287***	-0.165***	1	0.31*	-0.018	1	0.227*	-0.019	1	0.574*	-0.01	1

Notes: asterisks denote the level of significance α at which you can reject the null hypothesis of the impact matrix entry being equal to zero. The confidence interval drawn at 68% (*), 90% (**), 95% (***) level does not contain the value 0.

one. The estimates $\tilde{b}_{ij}$ are the median of 500 estimates obtained by 500 bootstrap replications of the Equation (12) model (through replications of u_t), with confidence intervals derived as in Hall (1992):

$$CI(\alpha) = \left[\tilde{b}_{ij} - q^*_{(1-\alpha/2)}, \tilde{b}_{ij} - q^*_{\alpha/2}\right] \tag{14}$$

where $q^*_{(1-\alpha/2)}$ and $q^*_{\alpha/2}$ are the $(1-\alpha/2)$ and $\alpha/2$-quantiles of empirical distribution of the root $(\hat{b}^*_{ij} - \tilde{b}_{ij})$, with $(\hat{b}^*_{ij}$ being the estimate of b_{ij} from the bootstrap replication).[14] For sake of comparison, we normalize all the coefficients so that each structural innovation has a unit contemporaneous impact on the log of the variable that refers to.

For all estimators, the entries of the resulting mixing matrix do not satisfy the property that the greatest value of each column lies in a distinct row. This is case in which the *Maxfinder* criterion is not able to deliver a permuted mixing matrix such that the off-diagonal elements are greater than main diagonal entries. As mentioned at the end of Section 2.2, we pick here the permutation matrix that penalizes low absolute values on the main diagonal according to the LiNGAM criterion.[15] On the base of this criterion, we identify two independent shocks that increase mostly taxes and government spending, and a third shock that increases substantially output and taxes. Moreover, this configuration is compatible with Table 2, which reports BP's findings and their shocks' labeling. Therefore, we can label the first (ε^1) and the second (ε^2) shock as a spending and tax shock, respectively. All the methods deliver a positive and significant impact coefficients of spending on GDP. Moreover, we get an estimate of a positive and statistically significant contemporaneous response of the third shock on tax revenues, but which is lower than BP's estimate of the GDP shock on tax revenues (except for *PML* that delivers a higher value). Finally, with the exception of *PML*, we get a non-significant impact of ε^3 on *G*, which is consistent with one of the BP's zero restriction if ε^3 is interpreted as the GDP shock. Overall, the ICA-estimated tax shock (ε^2) seems not to have a significant impact on both spending and GDP (except for the *PML* estimate). This result about spending is in tune with the zero-restriction of the model 1 in BP (zero impact from tax shock to *G*). Given the results of our assessments in the previous section, we tend to rely more on the higher precision and the better empirical coverage of *PML* and *fastICA* estimators. In Appendix A.2 (see Table 5) we report the permuted mixing matrices using the identification scheme by Lanne et al. (2017).

Having estimated the structural shocks and having identified the impact coefficients via ICA, we can now compute the impulse response functions and compare them with those implied by the BP model, shown in Figure 6. We first comment the impulse responses of a positive public spending shock, shown in Figure 7. The response of output is clearly positive and statistically significant, both at impact and within the first year. Figure 8 instead, shows the responses to an independent positive shock in tax revenues. All methods show that the effects of independent tax shocks are negative in the long run, which is a finding consistent with BP. However, all ICA-methods (except *PML*) show non-significant effect on the impact and in the short run, which is at odds with the finding by BP.

This illustrative exercise has shown that a purely data-driven identification procedure of VAR models is possible and, with careful modeling decisions, can lead to convincing conclusions based exclusively on statistical properties of the data. BP's identifying restrictions are plausible not only because of institutional knowledge and insights from economic theory: they are also present in the data and the IC model supports them, at least as regards the zero restrictions of the BP's first model (*G* ordered first). On the contemporaneous relation between tax revenues and public spending, our results suggest that public spending is the first mover and its immediate impact on taxes is positive. In partial contrast to BP's results, our estimated impulse response functions suggest that a fiscal policy guided by public spending has a clearer effect on economic activity than a fiscal policy guided by tax revenues.

[14] We adopt the moving block bootstrap with a recursive procedure, where the u_t's are resampled. A well-specified SVAR model is then generated using the estimated coefficient in order to obtain the bootstrapped time series y^*_t. Then a VAR model is estimated and identified via ICA and by applying *Maxfinder*. The bootstrap distribution of the entries of the mixing matrices are then used to draw the confidence interval defined in Equation (14). It follows that, for all estimators, the confidence intervals for the impulse response functions take into account the two-step estimation VAR-ICA.

[15] Notice that application of the mere criterion of penalisation of small values (second step of the LiNGAM algorithm by Shimizu et al. (2006)) on the main diagonal does not involve the causal recursiveness assumption on which LiNGAM is based.

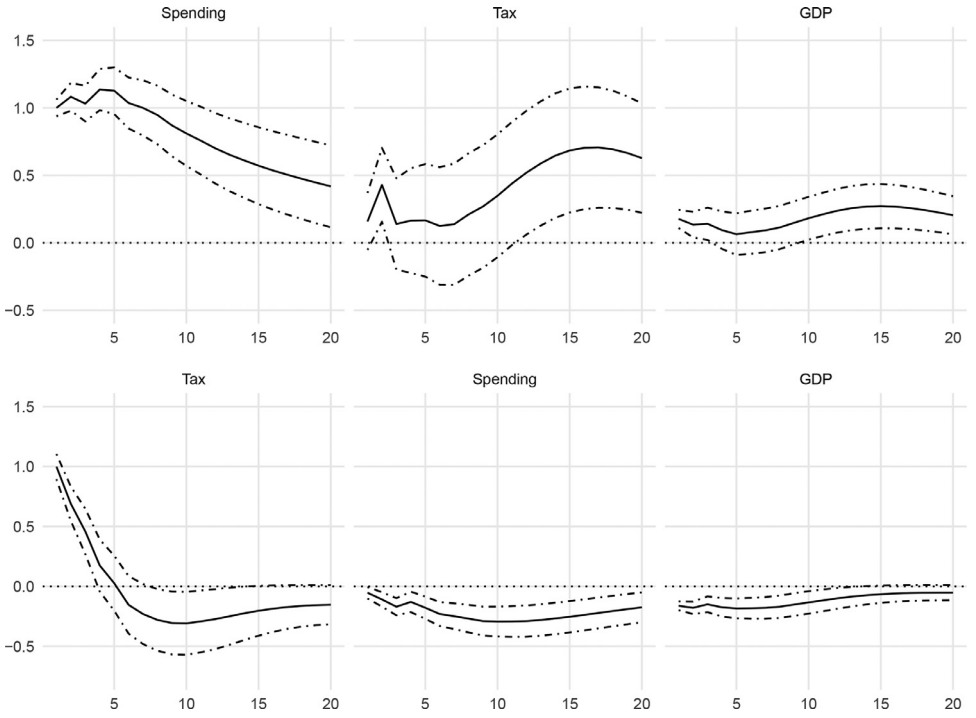

Fig. 6. Impulse response functions estimated by Blanchard and Perotti (2002). The upper panel shows the responses of the three variables to a spending shock (on *G*). The bottom panel shows the responses to a tax-revenue shock (on *Tax*). Dashed lines denote an equal-tailed 68% confidence interval.

5. Conclusions

In this paper we assess, through Monte Carlo experiments, the performance of four ICA techniques (*fastICA*, *DCov*, *CvM*, *PML*), which have been recently used in SVAR analysis. We study the cases of structural disturbances that follow distributions which are both sub- and super-Gaussian but also approaching normality, that is the case in which the IC model cannot recover the shocks by construction. The *fastICA* and the *PML* estimators result to be the ones showing a relatively lower variability and a more stable performance across both sub- and super-Gaussian settings, in comparison to the other estimators. In specific cases (e.g., number of variables $k = 2$), the *DCov* estimator performs on average better than other estimators when the shocks' distributions are in the neighbourhood of Gaussianity. The variability of the *DCov* estimates, however, is relatively high and it is shared with the *CvM* estimator.

We also consider the distributions of the mixing matrix coefficients, which is the matrix that identifies a SVAR model and contains the simultaneous interactions of the variable of the system. Our Monte Carlo studies show that, as the dimensionality of the system increases, uncertainty in the estimates increases, as well as their negative bias. We also analyze the ICA methods' performance in statistical inference. Specifically, we have considered size distortions when testing the significance of the coefficients of the mixing matrix, comparing the performances of maximum likelihood versus bootstrap based inference. The *DCov* method, despite being relatively accurate on average, shows concerning variability. In the statistical inference exercise, on the other hand, the method based on the *PML* and *fastICA* estimators show lower size distortions and a better empirical coverage in almost all distributional scenarios. The *fastICA* estimator, more specifically, is the one which displays the lowest size distortion. Finally, an empirical application on fiscal policy highlights that a purely data-driven procedure such as ICA may help the researcher to test the significance of identifying restrictions or to suggest where to insert the latter. In particular, our exercise shows that the IC model cannot reject the identification scheme implemented in Blanchard and Perotti (2002).

Appendix A. Further results

A1. Lower sample size

In this section we repeat the general and specific assessment (as regards the MDI calculation) with a sample size of $T = 100$ and $T = 200$. Results are displayed in Figures 9 (general assessment) and 10 (specific assessment). These Figures should be compared with Figures 2 and 3, respectively. In both cases, results do not change qualitatively when the sample size decreases to $T = \{100, 200\}$. In general, however, the average MDI tends to increase with the decreasing of the sample size.

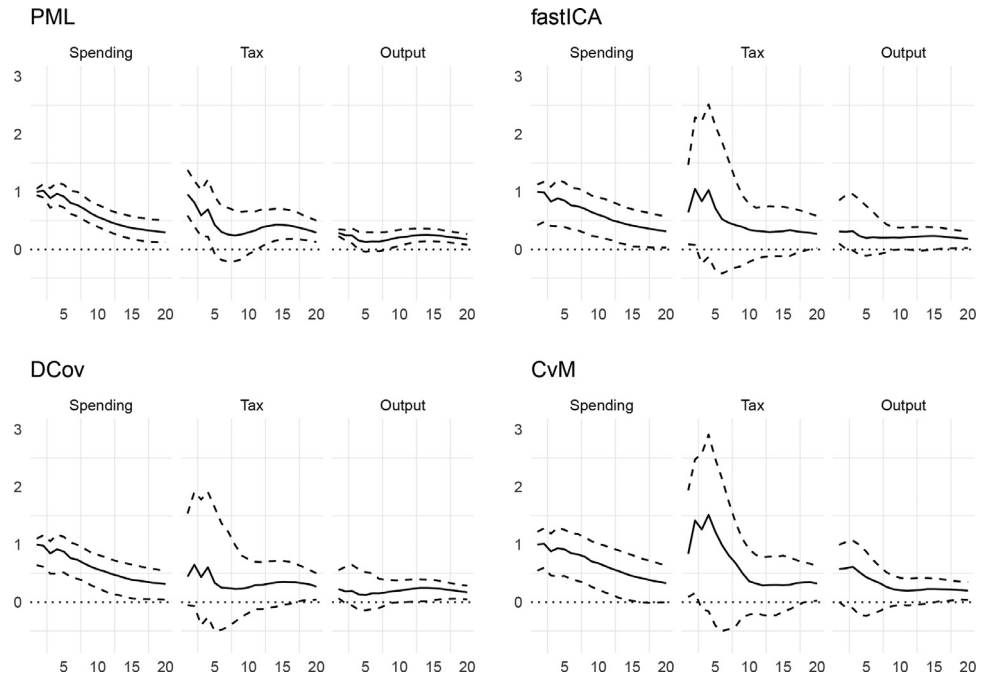

Fig. 7. Impulse response functions of a positive public spending shock from different ICA estimation methods. The upper and the lower dashed lines represent respectively the 84% quantile and the 16% quantile of the bootstrap estimates.

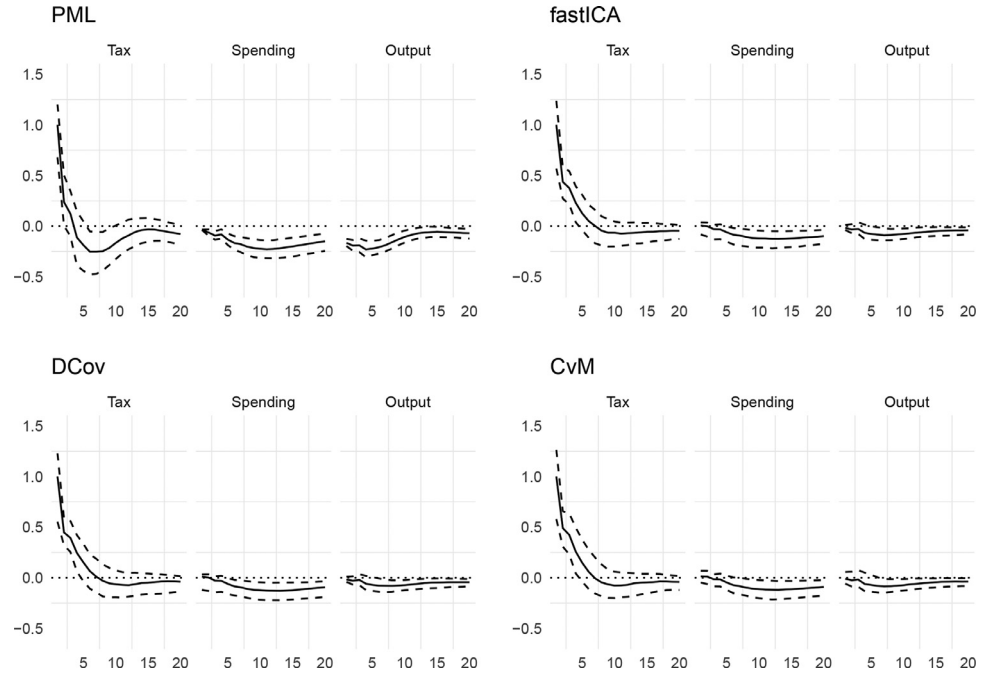

Fig. 8. Impulse response functions of a positive tax shock from different ICA estimation methods. The upper and the lower dashed lines represent respectively the 84% quantile and the 16% quantile of the bootstrap estimates.

A2. Results from Lanne's et al. identification scheme

In this section, we further check the validity of the results from both the simulation and empirical analysis, by changing the criterion of permutation of the mixing matrix. Instead of *Maxfinder* (or second step of LiNGAM, as we did in the empirical application), we apply here the criterion by Lanne et al. (2017), discussed Section 2.2 (point 5). Results are reported in Tables 4 and Figure 11, as regards the simulation study (specific assessment), and Table 5, as regards the empirical application.

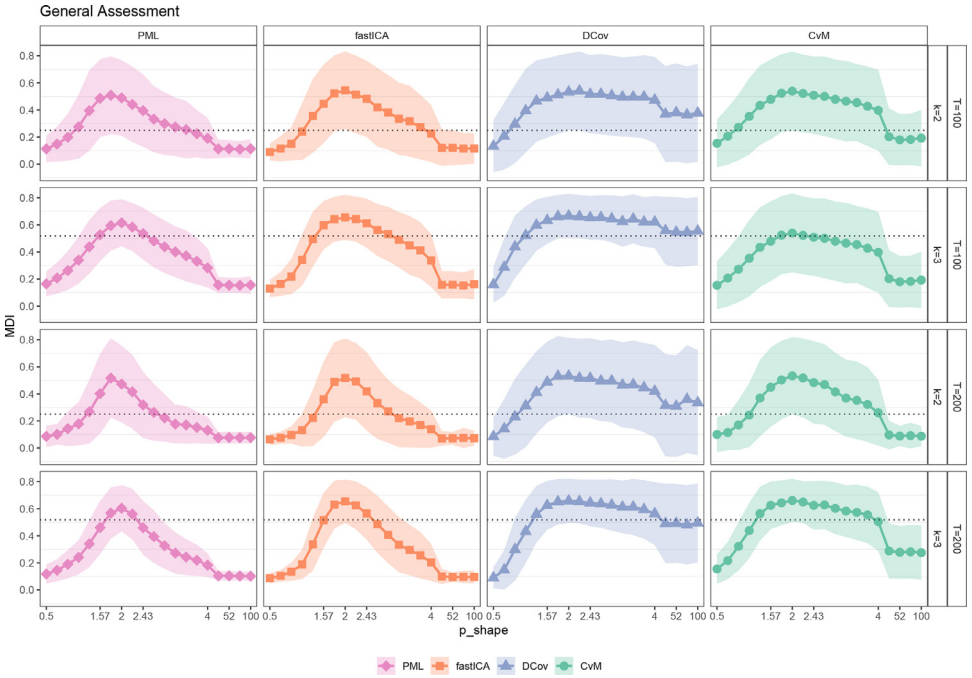

Fig. 9. Each plot shows the MDI (thick lines with marks: averages across 1000 dgps), as specified in Equation (7), for each of the four ICA estimators, across 20 values of p (*general assessment*). The p-generalized normal distribution is super-Gaussian with $p < 2$, Gaussian with $p = 2$, sub-Gaussian with $p > 2$. Plots in the upper part of the figure correspond to the case in which the sample size is $T = 100$ (for $k = 2$ and $k = 3$, respectively), in the lower part to $T = 200$ (for $k = 2$ and $k = 3$, respectively). Shadow areas show one-standard deviation above and below the mean. Dashed lines constitute the negative benchmark: 90% of random estimates yield MDIs above this line.

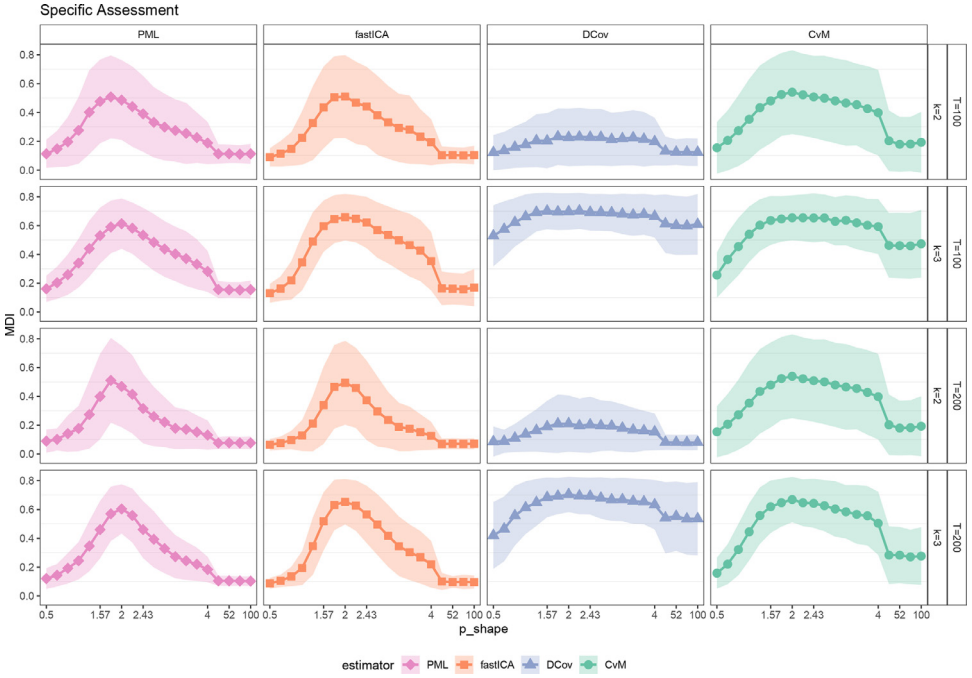

Fig. 10. Each plot shows the MDI (thick lines with marks: averages across 1000 Monte Carlo replications), as specified in Equation (9) (*specific assessment*), for each of the four ICA estimators, across 20 values of p. The mixing matrix in the dgp is fixed as in Equation (8). The p-generalized normal distribution is super-Gaussian with $p < 2$, Gaussian with $p = 2$, sub-Gaussian with $p > 2$. Plots in the upper part of the figure correspond to the case in which the sample size is $T = 100$ (for $k = 2$ and $k = 3$, respectively), in the lower part to $T = 200$ (for $k = 2$ and $k = 3$, respectively). Shadow areas show one-standard deviation above and below the mean.

Table 4
Summary statistics of the errors between entries of $\widehat{B}_0$ and of B_0 (the latter as specified in Equation 8). Sample size $T = 400$. The columns of the mixing matrix $\widehat{B}_0$ are ordered following the identification scheme of Lanne et al. (2017).

		$k=2$							$k=3$							
		$p=0.5$		$p=1.57$		$p=2.47$		$p=100$		$p=0.5$		$p=1.57$		$p=2.47$		$p=100$	
		mean	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
PML	$\hat{b_{11}} - b_{11}$	-0.002	0.007	-0.046	0.081	-0.039	0.068	-0.001	0.002	-0.005	0.042	-0.026	0.140	-0.025	0.148	-0.004	0.032
fastICA		-0.001	0.001	-0.032	0.067	-0.049	0.081	-0.001	0.001	-0.002	0.027	-0.019	0.143	-0.009	0.167	-0.003	0.031
DCov		-0.001	0.006	-0.015	0.034	-0.018	0.037	-0.001	0.002	0.034	0.137	0.115	0.163	0.129	0.161	0.066	0.146
CvM		-0.003	0.018	-0.076	0.107	-0.093	0.113	-0.001	0.002	-0.001	0.044	-0.015	0.177	-0.012	0.184	-0.007	0.064
PML	$\hat{b_{12}} - b_{12}$	-0.002	0.065	-0.035	0.310	-0.035	0.287	0.001	0.051	-0.003	0.069	-0.073	0.310	-0.103	0.324	0.000	0.052
fastICA		0.000	0.038	-0.016	0.261	-0.044	0.317	0.001	0.046	0.000	0.041	-0.088	0.325	-0.141	0.377	0.000	0.047
DCov		0.001	0.051	-0.009	0.179	-0.017	0.200	0.001	0.052	-0.063	0.240	-0.172	0.302	-0.178	0.282	-0.126	0.313
CvM		0.001	0.079	-0.085	0.387	-0.116	0.421	0.001	0.052	-0.002	0.073	-0.189	0.423	-0.205	0.410	-0.013	0.108
PML	$\hat{b_{21}} - b_{21}$	0.002	0.068	0.055	0.353	0.054	0.327	0.001	0.054	0.004	0.076	0.177	0.397	0.214	0.421	0.002	0.051
fastICA		-0.001	0.046	0.030	0.296	0.067	0.364	0.001	0.049	0.000	0.043	0.215	0.422	0.342	0.477	0.002	0.046
DCov		-0.002	0.077	0.015	0.205	0.028	0.231	0.002	0.054	0.175	0.335	0.449	0.367	0.477	0.352	0.388	0.361
CvM		-0.001	0.094	0.120	0.446	0.162	0.486	0.002	0.056	0.005	0.075	0.464	0.526	0.523	0.507	0.014	0.114
PML	$\hat{b_{22}} - b_{22}$	-0.001	0.021	-0.040	0.107	-0.032	0.097	-0.002	0.017	-0.002	0.039	-0.023	0.143	-0.022	0.149	-0.003	0.035
fastICA		-0.001	0.013	-0.030	0.091	-0.040	0.106	-0.001	0.016	-0.002	0.026	-0.024	0.143	-0.018	0.159	-0.003	0.031
DCov		-0.003	0.022	-0.014	0.062	-0.016	0.074	-0.001	0.017	0.013	0.085	0.104	0.137	0.113	0.136	0.038	0.105
CvM		-0.004	0.033	-0.055	0.117	-0.064	0.129	-0.001	0.018	-0.003	0.045	-0.028	0.172	-0.013	0.178	-0.008	0.066

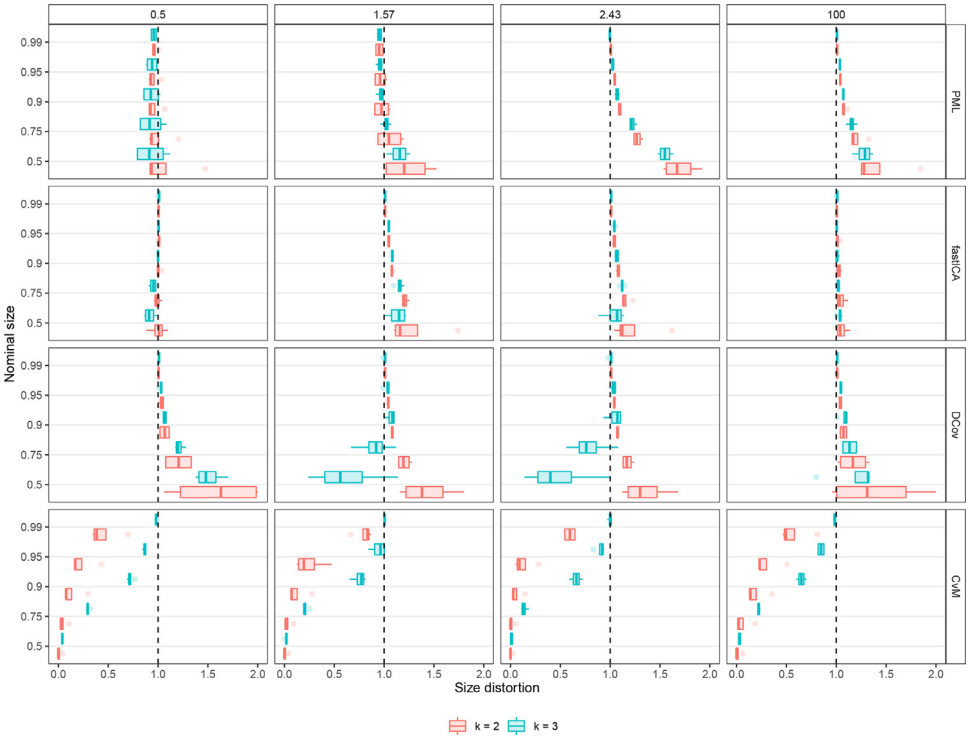

Fig. 11. Distribution of size distortions (measured on the $x-$ axis by the ratio between the empirical coverage $1-\hat{\alpha}$ and its nominal counterpart $1-\alpha$) in drawing confidence intervals for the estimates of the mixing matrix entries, for different estimators (row-wise ordered panels), different distributional scenarios (column-wise ordered panels), when the dimension of the system k changes. Confidence intervals are constructed at nominal α significance level. For *PML*, we use the asymptotic approximation derived by Gouriéroux et al. (2017), while for *fastICA*, *DCov* and *CvM* we implement the *warp*-bootstrap procedure described in text. The columns of the mixing matrix $\widehat{B}_0$ are ordered following the identification scheme of Lanne et al. (2017).

Table 4 shows mean and standard deviation (across Monte Carlo runs) of $\hat{b}_{ij} - b_{ij}$. These results approximately mirror the numbers shown in Table 1): there is not substantial difference in the distributional properties of the mixing matrix's entries. Clearly, both identification schemes in our specific assessment deliver the same permutation. Figure 11 complements by showing that the distribution of size distortions when making inference is essentially invariant to the identification scheme implemented. Finally, Table 5 shows the results for the empirical application, suggesting that, except for *CvM*, the

Table 5
Estimates of the contemporaneous impacts coefficients from IC models. The entries of the mixing matrices are calculated using the bootstrap-median with the corresponding confidence intervals. The columns of the mixing matrix $\widehat{B}_0$ are ordered following the identification scheme of Lanne et al. (2017).

	PML			*fastICA*			*DCov*			*CvM*		
	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
G_t	1	-0.036***	-0.197***	1	0.016	0.145	1	0.013	0.117	1	0.132	11.153*
Tax_t	0.947***	1	3.947***	0.107	1	1.191**	0.113	1	1.285***	8.052	1	0.92
GDP_t	0.293***	-0.165***	1	0.164	-0.012	1	0.158*	-0.017	1	-1.36	0.279	1

Notes: asterisks denote the level of significance α at which you can reject the null hypothesis of the impact matrix entry being equal to zero. The confidence interval drawn at 68% (*), 90% (**), 95% (***) level does not contain the value 0.

Table 6
Summary statistics of the errors between entries of $\widehat{B}_0$ and of B_0 (the latter as specified in Equation 8). Sample size $T = 400$. The IC model is estimated on the residuals of an estimated SVAR(1) model.

		$k=2$							$k=3$							
		$p=0.5$		$p=1.57$		$p=2.47$		$p=100$		$p=0.5$		$p=1.57$		$p=2.47$		$p=100$	
		mean	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*	*mean*	*sd*
PML	$\hat{b_{11}} - b_{11}$	-0.002	0.007	-0.045	0.077	-0.039	0.068	-0.001	0.002	-0.007	0.041	-0.033	0.142	-0.025	0.145	-0.007	0.032
fastICA		-0.003	0.006	-0.034	0.063	-0.053	0.081	-0.004	0.003	-0.006	0.028	-0.025	0.145	-0.012	0.163	-0.006	0.029
DCov		-0.004	0.015	-0.019	0.038	-0.022	0.044	-0.004	0.004	0.036	0.126	0.116	0.156	0.129	0.149	0.061	0.146
CvM		-0.005	0.015	-0.082	0.107	-0.098	0.111	-0.004	0.004	-0.003	0.045	-0.017	0.178	-0.004	0.180	-0.012	0.064
PML	$\hat{b_{12}} - b_{12}$	-0.002	0.065	-0.029	0.308	-0.035	0.287	0.001	0.051	0.001	0.059	-0.067	0.311	-0.106	0.332	-0.001	0.053
fastICA		0.000	0.039	-0.015	0.257	-0.051	0.321	0.001	0.046	-0.001	0.041	-0.084	0.327	-0.150	0.383	-0.001	0.048
DCov		0.001	0.056	-0.010	0.189	-0.022	0.203	0.001	0.053	-0.052	0.215	-0.173	0.303	-0.177	0.288	-0.143	0.324
CvM		0.001	0.075	-0.092	0.393	-0.115	0.428	0.001	0.054	-0.001	0.076	-0.186	0.426	-0.231	0.422	-0.005	0.111
PML	$\hat{b_{21}} - b_{21}$	0.002	0.068	0.048	0.349	0.054	0.327	0.001	0.054	0.002	0.057	0.165	0.391	0.230	0.421	0.004	0.052
fastICA		0.000	0.047	0.029	0.290	0.076	0.369	0.002	0.050	0.002	0.044	0.211	0.414	0.357	0.485	0.004	0.047
DCov		0.000	0.083	0.018	0.217	0.034	0.236	0.003	0.055	0.157	0.300	0.454	0.361	0.476	0.344	0.411	0.366
CvM		0.000	0.088	0.129	0.454	0.162	0.490	0.002	0.057	0.007	0.079	0.466	0.516	0.541	0.511	0.012	0.114
PML	$\hat{b_{22}} - b_{22}$	-0.001	0.021	-0.041	0.110	-0.032	0.097	-0.002	0.017	-0.008	0.039	-0.035	0.144	-0.022	0.148	-0.008	0.035
fastICA		-0.004	0.015	-0.032	0.089	-0.043	0.107	-0.005	0.016	-0.006	0.027	-0.030	0.142	-0.023	0.160	-0.007	0.032
DCov		-0.006	0.023	-0.019	0.063	-0.019	0.074	-0.005	0.018	0.008	0.086	0.104	0.132	0.106	0.133	0.039	0.105
CvM		-0.007	0.030	-0.060	0.117	-0.070	0.136	-0.005	0.019	-0.009	0.049	-0.031	0.163	-0.014	0.165	-0.016	0.071

identification scheme by Lanne et al. (2017) tends to deliver, across bootstrap iterations, the same permutation we get using the criterion of penalizing small (absolute) values on the main diagonal (see Table 3).

A3. Results from a SVAR(1) model

In this section we show how the ICA estimators perform in recovering the impact multiplier (mixing) matrix of a SVAR(1) model, rather than of a SVAR(0) model (our baseline set up). Does the two step procedure (estimation and identification via ICA) alter the distributional properties of the entries of the mixing matrix, once we allow for a richer lag structure? To see whether the uncertainty of the VAR estimates plays a role in the overall performance of the ICA estimators, we have replicated the simulation analysis, as regards the specific assessment, for a SVAR(1) model. We follow this procedure:

- Simulate a stable SVAR(1) model as written in Equations (2) and (3) (see Section 2.1), with $T = 400$ and structural shocks generated from the p-generalized normal distribution.
- Estimate a VAR(1) and obtain the reduced-form residuals.
- Estimate the mixing matrix B that identifies the model using the ICA techniques and applying *Maxfinder*.
- Evaluate the performance as in Section 3.2.

As shown in Table 6, given the two-step estimation, when the IC model is applied to the residuals of an estimated VAR, the estimates of the mixing matrices' entries are slightly more biased and volatile. As for the size distortion that may arise when conducting statistical inference, the performance of the four ICA methods is qualitatively similar to our baseline results shown in Figure 12: *fastICA* displays the smallest size distortion; all estimators but *CvM*, which still displays the worst performance, tend to contain much more often than expected the true value of the parameter ($\hat{\alpha} < \alpha$); this distortion decreases as the nominal size α increases. The only notable difference, however, is that the distribution of the size distortion is slightly wider than in the baseline scenario, thus confirming that the two-step procedure does not change qualitatively the performance of the four ICA estimators. The introduction of lags can increase the uncertainty of the estimates and the coverage of the confidence intervals, especially if the sample size remains fixed.

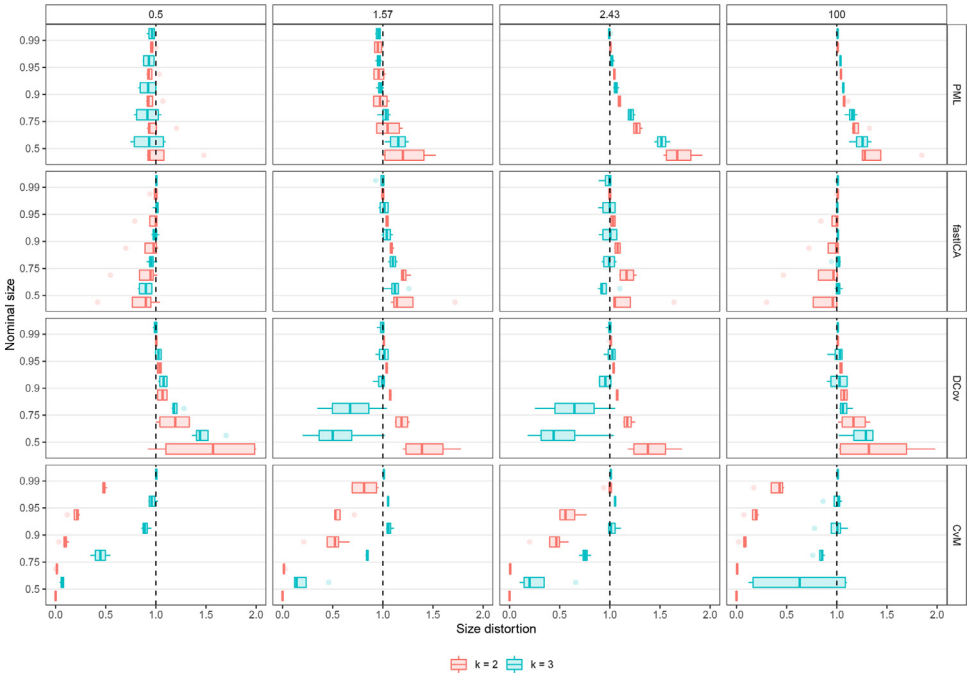

Fig. 12. Distribution of size distortions (measured on the $x-$ axis by the ratio between the empirical coverage $1-\hat{\alpha}$ and its nominal counterpart $1-\alpha$) in drawing confidence intervals for the estimates of the mixing matrix entries, for different estimators (row-wise ordered panels), different distributional scenarios (column-wise ordered panels), when the dimension of the system k changes. Confidence intervals are constructed at nominal α significance level. For *PML*, we use the asymptotic approximation derived by Gouriéroux et al. (2017), while for *fastICA*, *DCov* and *CvM* we implement the *warp*-bootstrap procedure described in text. The IC model is estimated on the residuals of an estimated SVAR(1) model.

Appendix B. R Codes and Data

The R codes and data to replicate the results reported in this article are available at the following GitHub repository: https://github.com/gianlucapallante/ica-svars-comparative-analysis.

References

Acharya, D., Panda, G., 2008. A review of independent component analysis techniques and their applications. IETE Technical Review 25 (6), 320–332.
Amisano, G., Giannini, C., 1997. From var models to structural var models. In Topics in structural VAR econometrics 1–28. Springer
Back, A.D., Weigend, A.S., 1997. A first application of independent component analysis to extracting structure from stock returns. International Journal of Neural Systems 8 (04), 473–484.
Baqaee, D.R., Farhi, E., 2019. The macroeconomic impact of microeconomic shocks: beyond Hulten's theorem. Econometrica 87 (4), 1155–1203.
Bernanke, B.S., 1986. Alternative explanations of the money-income correlation. Carnegie-Rochester Conference Series on Public Policy 25, 49–99.
Bernanke, B.S., Mihov, I., 1998. Measuring monetary policy. The Quarterly Journal of Economics 113 (3), 869–902.
Berner, A., Bruns, S., Moneta, A., Stern, D.I., 2022. Do energy efficiency improvements reduce energy use? empirical evidence on the economy-wide rebound effect in europe and the united states. Energy Economics 110, 105939.
Blanchard, O., Perotti, R., 2002. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. Quarterly Journal of economics 117 (4), 1329–1368.
Box, G., Cox, D., 1982. An analysis of transformations revisited, rebutted. Journal of the American Statistical Association 77 (377), 209–210.
Box, G.E.P., Tiao, G.C., 1962. A further look at robustness via Bayes's theorem. Biometrika 49 (3/4), 419–432.
Bruns, S.B., Moneta, A., Stern, D.I., 2021. Estimating the economy-wide rebound effect using empirically identified structural vector autoregressions. Energy Economics 97, 105158.
Caldara, D., Kamps, C., 2017. The analytics of SVARs: a unified framework to measure fiscal multipliers. The Review of Economic Studies 84 (3), 1015–1040.
Capasso, M., Moneta, A., 2016. Macroeconomic responses to an independent monetary policy shock: A (more) agnostic identification procedure. Technical report, LEM Working Paper Series..
Cardoso, J.F., 1989. Source separation using higher order moments. In International Conference on Acoustics, Speech, and Signal Processing, 2109–2112. IEEE
Ciarli, T., Coad, A., Moneta, A., 2019. Exporting and productivity as part of the growth process: Causal evidence from a data-driven structural VAR. Technical report, LEM Working Paper Series.
Comon, P., 1994. Independent component analysis, a new concept? Signal Processing 36 (3), 287–314.
Eriksson, J., Koivunen, V., 2004. Identifiability, separability, and uniqueness of linear ICA models. IEEE Signal Processing Letters 11 (7), 601–604.
Fagiolo, G., Napoletano, M., Roventini, A., 2008. Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. Journal of Applied Econometrics 23 (5), 639–669.
Fiorentini, G., Sentana, E., 2020. Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. Working Papers wp2020-2023, CEMFI..
Gabaix, X., 2011. The granular origins of aggregate fluctuations. Econometrica 79 (3), 733–772.
Genest, C., Quessy, J.-F., Rémillard, B., et al., 2007. Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. The Annals of Statistics 35 (1), 166–191.
Gertler, M., Karadi, P., 2015. Monetary policy surprises, credit costs, and economic activity. American Economic Journal: Macroeconomics 7 (1), 44–76.

Giacomini, R., Politis, D.N., White, H., 2013. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Econometric Theory 29 (3), 567589.
Goodman, I.R., Kotz, S., 1973. Multivariate θ-generalized normal distributions. Journal of Multivariate Analysis 3 (2), 204–219.
Gouriéroux, C., Monfort, A., Renne, J.P., 2017. Statistical inference for independent component analysis: Application to structural VAR models. Journal of Econometrics 196 (1), 111–126.
Gouriéroux, C., Monfort, A., Renne, J.P., 2020. Identification and estimation in non-fundamental structural VARMA models. The Review of Economic Studies 87 (4), 1915–1953.
Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., Smola, A., 2007. A kernel statistical test of independence. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.. Volume 20
Guay, A., 2021. Identification of structural vector autoregressions through higher unconditional moments. Journal of Econometrics 225 (1), 27–46.
Guerini, M., Moneta, A., Napoletano, M., Roventini, A., 2020. The janus-faced nature of debt: Results from a data-driven cointegrated SVAR approach. Macroeconomic Dynamics 24 (1), 2454.
Hall, P., 1992. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. The Annals of Statistics 20 (2), 675–694.
Hastie, T., Tibshirani, R., 2010. ProdenICA: Product density estimation for ICA using tilted Gaussian density estimates. R Foundation for Statistical Computing. R package version 1.0..
Herwartz, H., 2018. HodgesLehmann detection of structural shocks An analysis of macroeconomic dynamics in the Euro area. Oxford Bulletin of Economics and Statistics 80 (4), 736–754.
Herwartz, H., 2019. Long-run neutrality of demand shocks: Revisiting Blanchard and Quah 1989) with independent structural shocks. Journal of Applied Econometrics 34 (5), 811–819.
Herwartz, H., Lange, A., Maxand, S., 2022. Data-driven identification in svarswhen and how can statistical characteristics be used to unravel causal relationships? Economic Inquiry 60 (2), 668–693.
Herwartz, H., Plödt, M., 2016. The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. Journal of International Money and Finance 61, 30–44.
Horn, R.A., Johnson, C.R., 2012. Matrix analysis. Cambridge University Press..
Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE transactions on Neural Networks 10 (3), 626–634.
Hyvärinen, A., 2013. Independent component analysis: recent advances. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371 (1984), 20110534.
Hyvärinen, A., Oja, E., 1997. A fast fixed-point algorithm for independent component analysis. Neural computation 9 (7), 1483–1492.
Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural networks 13 (4-5), 411–430.
Hyvärinen, A., Zhang, K., Shimizu, S., Hoyer, P.O., 2010. Estimation of a structural vector autoregression model using non-Gaussianity. Journal of Machine Learning Research 11 (56), 1709–1731.
Ilmonen, P., Nordhausen, K., Oja, H., Ollila, E., 2010. A new performance index for ICA: properties, computation and asymptotic analysis. International Conference on Latent Variable Analysis and Signal Separation 229–236. Springer.
Kalke, S., 2015. pgnorm: The p-generalized normal distribution. R Foundation for Statistical Computing. R package version 2.0.
Kalke, S., Richter, W.D., 2013. Simulation of the p-generalized Gaussian distribution. Journal of Statistical Computation and Simulation 83 (4), 641–667.
Karamysheva, M., Skrobotov, A., 2022. Do we reject restrictions identifying fiscal shocks? Identification based on non-Gaussian innovations. Journal of Economic Dynamics and Control 138, 104358.
Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive analysis. Cambridge University Press.
Klein, L.R., Goldberger, A.S., 1955. An Econometric Model of the United States. North Holland. 1929-1952
Lacerda, G., Spirtes, P., Ramsey, J., Hoyer, P.O., 2008. Discovering cyclic causal models by independent components analysis. AUAI Press, Virginia, USA. In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 08, Arlington 366-374
Lanne, M., Luoto, J., 2021. GMM estimation of non-Gaussian structural vector autoregression. Journal of Business & Economic Statistics 39 (1), 69–81.
Lanne, M., Lütkepohl, H., 2010. Structural vector autoregressions with nonnormal residuals. Journal of Business & Economic Statistics 28 (1), 159–168.
Lanne, M., Meitz, M., Saikkonen, P., 2017. Identification and estimation of non-Gaussian structural vector autoregressions. Journal of Econometrics 196 (2), 288–304.
Lütkepohl, H., 2005. New introduction to multiple time series analysis. Springer Science & Business Media..
Lütkepohl, H., Xu, F., 2012. The role of the log transformation in forecasting economic variables. Empirical Economics 42 (3), 619–638.
Matteson, D.S., Tsay, R.S., 2017. Independent component analysis via distance covariance. Journal of the American Statistical Association 112 (518), 623–637.
Maxand, S., 2020. Identification of independent structural shocks in the presence of multiple gaussian components. Econometrics and Statistics 16, 55–68.
Moneta, A., Entner, D., Hoyer, P.O., Coad, A., 2013. Causal inference by independent component analysis: Theory and applications. Oxford Bulletin of Economics and Statistics 75 (5), 705–730.
Mountford, A., Uhlig, H., 2009. What are the effects of fiscal policy shocks? Journal of Applied Econometrics 24 (6), 960–992.
Nelson, H.L., Granger, C., 1979. Experience with using the box-cox transformation when forecasting economic time series. Journal of Econometrics 10 (1), 57–69.
Nordhausen, K., Ollila, E., Oja, H., 2011. On the performance indices of ICA and blind source separation. In Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on 486–490. IEEE
Ramey, V.A., 2011. Identifying government spending shocks: It's all in the timing. Quarterly Journal of Economics 126 (1), 1–50.
Rigobon, R., 2003. Identification through heteroskedasticity. Review of Economics and Statistics 85 (4), 777–792.
Romer, C.D., Romer, D.H., 2010. The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. American Economic Review 100 (3), 763–801.
Shannon, C. E., 1949. The mathematical theory of communication. by CE shannon and warren weaver. urbana.
Shimizu, S., Hoyer, P.O., Hyvärinen, A., Kerminen, A., 2006. A linear non-Gaussian acyclic model for causal discovery. Journal of Machine Learning Research 7, 2003–2030.
Sims, C.A., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.
Sims, C.A., 1986. Are forecasting models usable for policy analysis? Quarterly Review 10, 2–16.
Stock, J.H., Watson, M.W., 2001. Vector autoregressions. Journal of Economic Perspectives 15 (4), 101–115.
Stock, J.H., Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and. Handbook of macroeconomics 2, 415–525. Elsevier.and structural vector autoregressions in macroeconomics.
Stock, J.H., Watson, M.W., 2017. Twenty years of time series econometrics in ten pictures. Journal of Economic Perspectives 31 (2), 59–86.
Subbotin, M.T., 1923. On the law of frequency of error. Math. USSR-Sb 31 (2), 296–301.
Székely, G.J., Rizzo, M.L., Bakirov, N.K., et al., 2007. Measuring and testing dependence by correlation of distances. Annals of Statistics 35 (6), 2769–2794.
Tiao, G., Lund, D.R., 1970. The use of OLUMV estimators in inference robustness studies of the location parameter of a class of symmetric distributions. Journal of the American Statistical Association 65 (329), 370–386.
Uhlig, H., 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure. Journal of Monetary Economics 52 (2), 381–419.
Yu, S., Zhang, A., Li, H., 2012. A review of estimating the shape parameter of generalized Gaussian distribution. J. Comput. Inf. Syst 8 (21), 9055–9064.
Zema, S.M., 2022. Directed acyclic graph based information shares for price discovery. Journal of Economic Dynamics and Control 139, 104434.