INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

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

Alessio Moneta [a]
Gianluca Pallante [a]

[a] Institute of Economics and EMbeDS Department, Scuola Superiore Sant'Anna, Pisa, Italy.

2020/24 September 2020
ISSN(ONLINE) 2284-0400

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

Alessio Moneta[1] and Gianluca Pallante[2]

[1,2]Institute of Economics and EMbeDS

Sant'Anna School of Advanced Studies

Pisa, Italy

September, 2020

Abstract

Independent Component Analysis (ICA) is a statistical method that transforms a set of random variables in least dependent linear combinations. Under the assumption that the observed data are mixtures of non-Gaussian and independent processes, ICA is able to recover the underlying components, but 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 identification, and compare their performance in recovering structural coefficients. Moreover, after suggesting an algorithm that solve the ICA indeterminacy problem, we assess the size distortions of the estimators in hypothesis testing. We conduct our analysis by focusing on distributional scenarios that get gradually close the Gaussian case, which is the case where ICA methods fail to recover the independent components. In terms of statistical properties of the ICA estimators, we find no evidence that a method outperforms all others. 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.

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

JEL classification: C14, C32, E62.

Acknowledgments: We thank Fulvio Corsi, Gabriele Fiorentini, and participants to the the WEHIA Conference 2018 and the CFE Conference 2018 for comments and suggestions.

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. The latter have been proposed as an alternative against 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 a structural model.

Specifically, in order to measure the effects of exogenous interventions on the system, one needs to distinguish correlation from causation (Stock and Watson, 2017). This is because the residuals of an estimated (reduced-form) VAR model typically display cross-correlations, which are induced by the contemporaneous causal relationships among the variables 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 observational 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 Uhlig, 2005; Mountford and Uhlig, 2009). 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). These identification strategies have contributed, at least to a certain extent, to make SVAR identification based less on restrictions guided by theory and more on statistical properties of the data. Still, the picture of SVAR model as a statistical tool capable to inferring causal effects from data is not yet well-grounded.

A recent stream of literature exploits directly certain statistical properties of the data, namely the non-Gaussianity of the reduced-form VAR residuals, which results in many estimations (Lanne and Lütkepohl, 2010; Lanne and Saikkonen, 2013; Lanne et al., 2017; Lanne and Luoto, 2019; Gouriéroux et al., 2020). In this framework, linear combinations of VAR disturbances can be recovered so to be not only uncorrelated, but also statistically independent. This is possible by means of a purely data-driven statistical technique, called Independent Component Analysis (ICA) (Comon, 1994; Hyvärinen, 1999; Eriksson and Koivunen, 2004). Under the condition that the underlying structural shocks are statistically independent and non-Gaussian, the SVAR model is globally identified up to a re-scaling and re-ordering of the shocks. Examples of SVAR models identified using ICA are increasing and delivering interesting insights (Moneta et al., 2013; Herwartz and Plödt, 2016; Capasso and Moneta, 2016; Gouriéroux et al., 2017; Guerini et al., 2018; Herwartz, 2018b; Maxand, 2018). Many algorithms for learning independent components from data have been proposed and applied, especially in the field of blind signal separation, neural networks, feature extraction (Hyvärinen and Oja, 2000), as well as in fields which are more closely related to economics like finance (Back and Weigend, 1997), causal inference and structural modeling (Shimizu et al., 2006).

In this paper we evaluate a class of ICA-based methods, selectively focusing on those that have been proposed within the field of SVAR identification, namely the *fastICA* algorithm developed by Hyvärinen (1999) and employed by Moneta et al. (2013) and Guerini et al. (2018); the minimization of the Cramer-von-Mises statistics proposed by Herwartz and Plödt (2016); the pseudo-maximum likelihood estimator derived in Gouriéroux et al. (2017); and the minimization of the distance covariance by Matteson and Tsay (2017). The latter study shows how the distance-covariance method outperforms several ICA techniques under several distributional scenarios. With a different focus, Herwartz (2018a) undertakes a similar analysis, focusing more on the discriminatory power of several identification schemes in detecting structural shocks embedded in a simple DSGE model for the Euro Area. Our paper extends and completes the analysis by focusing on the statistical performances of the four methods mentioned above. Specifically, we study how these methods perform when the distribution of the structural disturbances gets gradually closer to the full Gaussian case, which corresponds to the case in which the SVAR model cannot be identified through ICA. We examine the different distributional scenarios through a p-generalized normal distribution. The novelty of our study is that, after suggesting an algorithm that solves the scale/order indeterminacy of the ICA model, we are able to analyze the distribution of the parameter's estimates derived under the different methods. Furthermore, we study the size distortions that arise when performing statistical inference on the coefficients of the impact multiplier matrix.

The paper is organized as follows: Section 2 presents the framework of our study, introducing the ICA-based SVAR model and the simulation exercise. Section 3 presents and discusses the results of our assessment. Section 4 discusses 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

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

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 $0 \leq l \leq q$, c_t is a $k \times 1$ vector of constants, which may also include a deterministic trend. (The analysis 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 of the VAR model, 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

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

where $B_l = A_0^{-1} A_l$ for $1 \leq l \leq q$, $\quad d_t = A_0^{-1} c_t$, $\quad u_t = A_0^{-1} \varepsilon_t$. 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, so that we refer to it as the ICA 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*.

Let us denote with a_i' the rows of the matrix A_0. Any ICA procedure aims at estimating the k-length weight vectors a_i' for $1 \leq i \leq k$, 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 scale (including sign) and ordering. More precisely, the matrix A_0 in the ICA 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, whose performance in recovering the mixing/unmixing matrix of equation (3) we want to comparatively assess.Although in addition to these four, more algorithms have been proposed in the literature (see Cardoso, 1989; Hyvärinen, 2013), the approaches described below are good representative of the ICA methods already 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 $H(.)$ is the differential entropy function, 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 minimize 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. Matteson and Tsay (2017) propose to estimate the ICA 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).[1] 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. Herwartz and Plödt (2016) and Herwartz (2018a),

[1] 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.

similarly to Matteson and Tsay (2017), define an objective function to be minimized in function of θ and exploits 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. 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 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 $\log g_i(\varepsilon_t)$, but also assuming that some parts of the density functions may be misspecified.

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. The chosen set 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. In 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 ICA model can be legitimately applied for identification.

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 (platykutic) distribution. Often used for robustness studies (Subbotin, 1923; Box and Tiao, 1962; 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[2]. 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.

We split our Monte Carlo experiment into two different designs: a *General Assessment* and a *Specific Assessment*. The former, 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

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

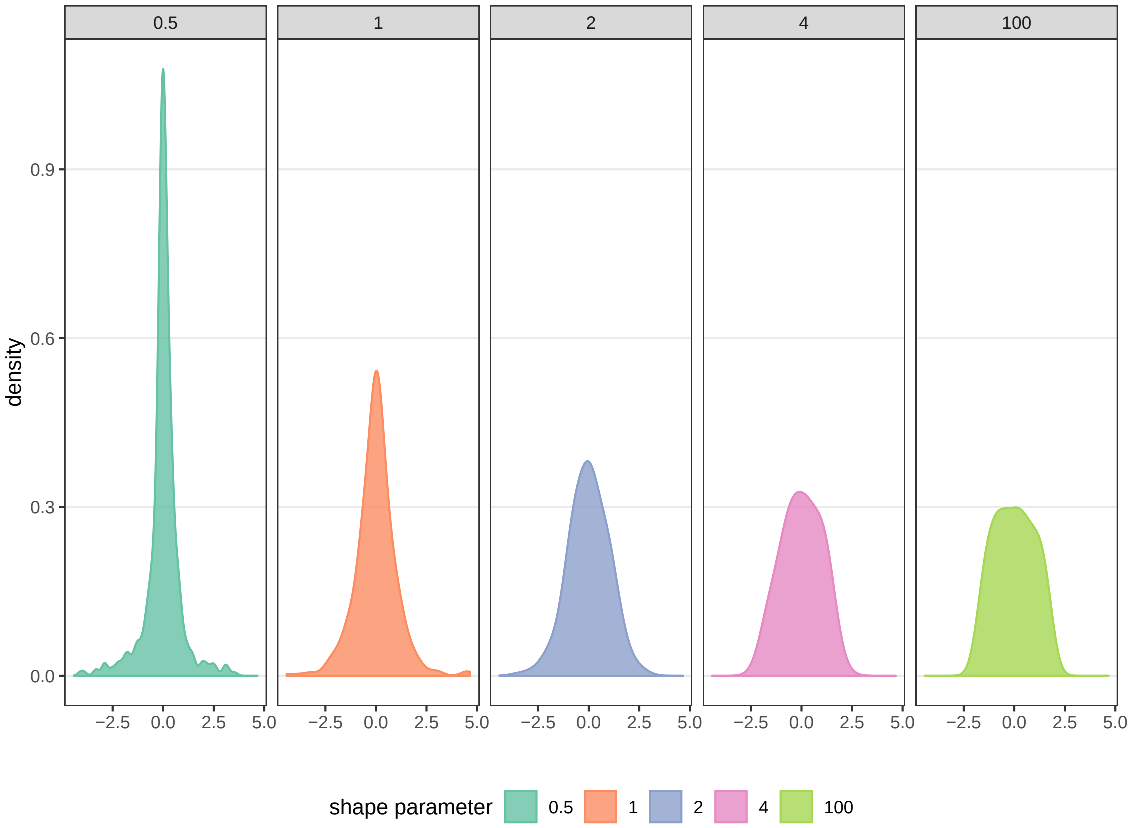

Figure 1: Kernel density estimates on data generated from the p-generalized normal distribution for different values of the shape parameter p (0.05, 1, 2, 4,100): with $p = 0.5$ super-Gaussian (Laplace), $p = 2$ Gaussian, $p = 100$ sub-Gaussian (uniform).

generates a random ICA 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 $1 \leq \mathcal{K}(B_0^{(m)}) \leq 2$, simulated with the R-package `ProDenICA` (Hastie and Tibshirani, 2010)[3]. Given the indeterminacy of the ICA model, namely its identification only up to column scale/sign and permutation of B_0 (as mentioned in the previous sub-section), 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[4]. 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 "Ilmonen index".

The *Specific Assessment* aims at evaluating how well the ICA methods perform in identifying the structural impulse response of a SVAR model, using different realizations of the same data generating process, for a given mixing matrix B_0. This also allows us to also 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 in the previous sub-section, an ICA algorithm, which delivers a mixing (or unmixing) matrix,

[3]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*.

[4]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.

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). For all four ICA methods, we propose here a unique criterion to address this scale-sign-ordering problem. This criterion relies inevitably on a priori assumptions, which can be justified on some general considerations that are free from any economic-theoretic argument.

The first general consideration is that in SVAR analysis it is uncontroversial to expect that a structural shock is going to affect mostly, at least in its immediate response, the variable is referring to and by which is *labelled*. A monetary policy shock, for instance, is likely to have a bigger contemporaneous impact (measured as a variable's percentage change in response to an impulse as large as the shock's standard deviation) on the interest rate — the usual monetary policy instrument — than other variables (e.g. GDP, inflation, etc.), which can be affected by the monetary policy shock only via the monetary policy instrument. Another general, but related, consideration is that, since we are interested in the impulse response function, the standard deviation of the shock can be always normalized to one by dividing each shock for its standard deviation and multiplying each column of $\widehat{B_0}$ (the estimated mixing matrix) by the same number — i.e. by right multiplying $\widehat{B_0}$ by $D = \text{diag}(\sigma_{\varepsilon_1}, \ldots, \sigma_{\varepsilon_k})$ — with no effect on the estimate of the impulse responses. Finally, the sign of the impact is a matter of labelling convention: for example, if we decide to study the effect of a *contractionary* monetary policy shock, we want the effect of the interest rate instrument to be positive, while if we study the effect of an *expansionary* monetary policy shock, we want the same effect to be negative, so that we can change the sign of the column of the mixing matrix corresponding to the monetary policy shock accordingly. In general, we are interested in the positive effects of economic shocks.

This means that we should expect that, after having normalized the mixing matrix B_0 so that the structural shocks have unit variance, the maximum entries, in absolute value, of each column of B_0 do not appear on the same row. The data generating processes specified in equation (8) satisfy this characteristic. But, of course, we cannot exclude a priori the possibility that in an empirical mixing matrix a row contains more than one column-maxima. We then suggest an identification procedure to solve the indeterminacy of the ICA model, which deals also, in its last step, with the possibility of column-maxima lying on the same row. We call this procedure *MaxFinder*, and it consists of the following steps:

1. Given the observed (estimated through a VAR model) reduced-form residuals u_t, apply an ICA algorithm and obtain an estimate of the mixing matrix $\widehat{B_0}$, of the unmixing matrix $\widehat{A_0}$, and the independent components $\hat{\varepsilon}_t$ with $\Sigma_\varepsilon = I$.

2. Find the column permutation of $\widehat{B_0}$ such that, each column, has diagonal entries which are greater than off-diagonal's. Call this matrix $B_{id} = \widehat{B_0}P$, where P is a permutation matrix.

3. If step 2 fails, the matrix B_{id} is obtained by applying step 2 of LiNGAM (Shimizu et al., 2006). Such matrix is the column-permuted version of $\widehat{B_0}$ that minimizes the quantity $1/\sum_i |\widehat{b}_{ii}|$. Be aware that, potentially step 2 and step 3 deliver the same matrix. However, if the ICA algorithm does not deliver a matrix which has a maximum for each column that enters at different rows, then step 3 is applied in order to heavily penalize those column permutations that imply low entries (in absolute value) on the main diagonal of B_{id}

The next section shows the results of our Monte Carlo experiments.

3 Results

3.1 General assessment

Before presenting our results from the Monte Carlo experiments, we discuss some practical issues related with the implementation of the different algorithms. Both *fastICA* and distance covariance (*DCov*) optimize a non-linear, locally convex objective function. Following Matteson and Tsay (2017), we use a Latin hypercube sampling of 500 parameter values, corresponding to the rotation angles $0 < \theta < 2\pi$ of the Givens rotation matrices defined in equations (4) and (5); we then select the initialization that minimizes the metric proposed in equation (7). As mentioned in section 2.1, the *CvM* approach relies on the minimization of the Cramer-von-Mises distance between the empirical copula of the shocks vector $\hat{\varepsilon}_t = G(\tilde{\theta})^{-1}\hat{u}_t$ and the implied copula under the hypothesis of mutual independence between the shocks. The term $\hat{u}_t$ refers to the VAR estimated reduced-form residuals, while $\tilde{\theta}$ in its initial formulation (Herwartz and Plödt, 2016) is taken from a uniform grid of 29 rotation angles on the interval $\{0, 2\pi\}$: in the k-dimensional case this implies testing the hypothesis of independence for 29^k vectors of rotation angles. We adopt instead a more efficient search that exploits the Differential Evolution algorithm for global optimization (Price et al., 2006)[5]. The PML approach does not require any initialization because we use pseudo-likelihood functions that satisfy the condition, derived in Gouriéroux et al. (2017) and Hyvärinen et al. (2001), under which the estimates correspond to a maximum of the asymptotic optimization problem.

Figure 2 displays the Ilmonen index, as specified in equation (7), averaged across 500 Monte Carlo replications, for different values of k, T and p. This shows the performance of the four ICA methods when the full Gaussianity ($p = 2$) of the independent structural shocks is approached, for $k = 2$ (upper part of the figure) and $k = 3$ (bottom part), and for $T = 100$ (left part of the figure) and $T = 400$ (right part). As shown in Figure 8 in appendix, the case in which $T = 200$ is not qualitatively different. As expected, when the ε_t's are not statistically different from being normally distributed (shaded areas in the four panels of Figure 2), all the methodologies score bad (the metric is high). However, in average *DCov* is consistently better than the others all over the range of the values of p, both in the two and three dimensional case and regardless the sample size. *CvM* instead is dominated in all scenarios, while *fastICA* and the PML estimator are almost equivalent, with slightly better performance of the former when the sample size is small. As the latter increases, all the methods tend to score equal, the further being the independent components distributed from Gaussianity, both in the direction of super- and sub-Gaussianity. *DCov* seems to be the most robust when normality is approached. The results are compatible with and complementary to those in 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.

3.2 Specific assessment

Having analysed 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 ICA model in equation

[5] The algorithm is adapted from the R-package `svars` (Lange et al., 2018).

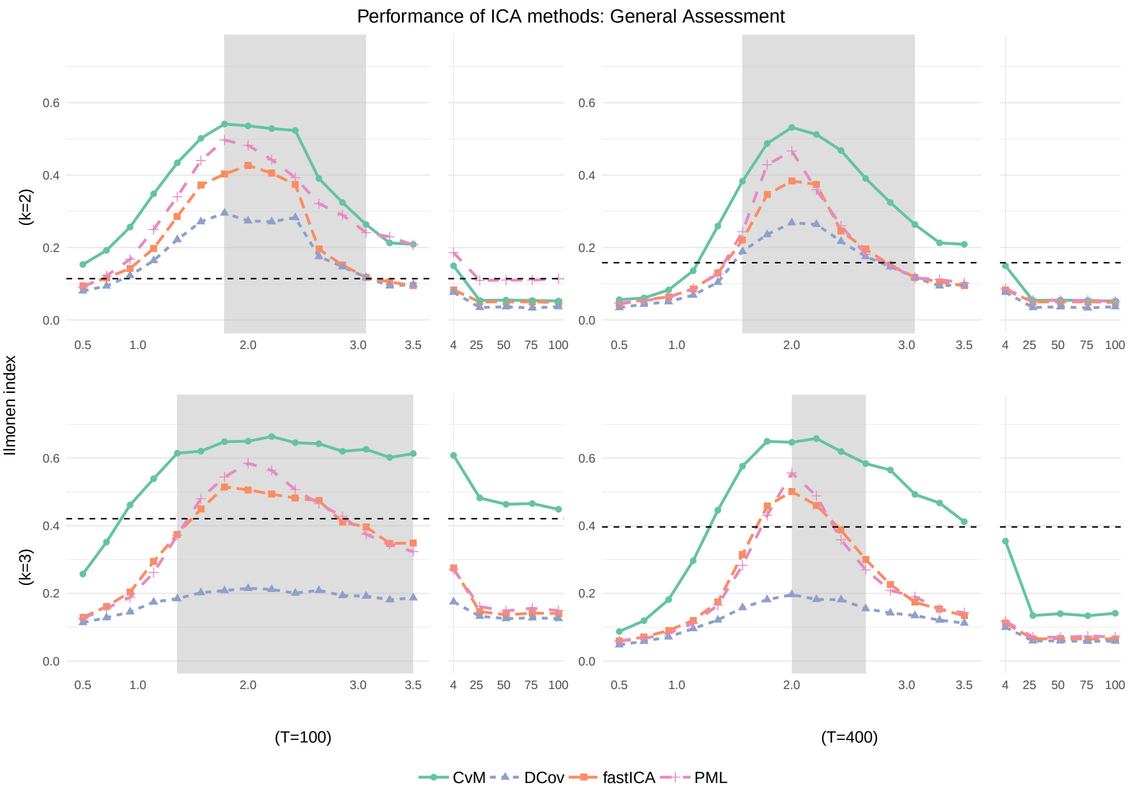

Figure 2: Ilmonen index ($y-$ axis), averaged across 500 Monte Carlo replications, for different ICA models in which ε_t follow a p-generalized normal distribution, with p ($x-$ axis) varying from the super-Gaussian ($p < 2$) case, to the Gaussian ($p = 2$), and sub-Gaussian ($p > 2$) case. Shaded area where normality is not rejected at 10% (on the basis of the Jarque-Bera test). Values below the dashed lines corresponds to estimated mixing matrices significantly different from a pure random at 5% significance level

(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. As mentioned in section 2.2, this assumption mirrors a common feature in SVAR analysis, namely the fact that each structural shock tends to be mostly (contemporaneously) correlated with the variable in the system that is referring to and is labelling with.

We apply the four ICA methods on the artificial u_t. We then apply the *MaxFinder* identification algorithm described in section 2.2, so to solve the indeterminacy that affects the ICA model.[6] We then assess the performance of the ICA model in this specific exercise

[6]The column-ordering problem, i.e. the labeling of the shocks recovered from ICA has rarely been addressed within the ICA literature on signal processing. However, in SVAR analysis, it plays a crucial role as it is necessary to identify the structural shock and its contemporaneous impact on the other variables. Lanne et al. (2017) suggest an algorithm that leads to a unique impact matrix among those belonging to the same equivalent class. We have explored other column-orderings algorithms, e.g. the column/sign permutation that minimizes the Frobenius norm between the estimated matrix and the true one. Results are

by using a variation of the Ilmonen index: differently from equation (7), the index 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 similarity metric 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 Ilmonen index, 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 independent components ε_t. Overall, the results resemble those shown for the *General Assessment*: *CvM* is dominated in all scenarios; the four methods are almost equivalent when the independent components are highly super- or sub-Gaussian; *DCov* performs better when Gaussianity is approached as well as when the sample size increases ($T = 400$). However, the latter shows a concerning variability when the distribution of the shocks is sub-Gaussian. Despite the large number of initializations (500), this result may suggest that the objective function that *DCov* solves is highly non-linear, so to imply a relatively higher uncertainty of the estimates: in fact, in some cases, the performance is not different to the case in which the full Gaussianity ($p = 2$) of ε_t's is simulated.

Although there are several measures of ICA performance proposed in the literature (see Nordhausen et al., 2011, for a review), those metrics, 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 errors 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 shows the four central moments that characterize the distribution 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

available upon request.

super-Gaussian ($p = 1.5$) or (iii) sub-Gaussian ($p = 2.5$); (iv) when the ε_t follow an almost uniform distribution ($p = 100$). For $k = 3$, only the upper left block's parameters are considered. In almost all cases, the estimates are negatively biased and their distribution is skewed to the left (a result compatible with those of Gouriéroux et al., 2017). As expected, in those scenarios where full Gaussianity is closer, the bias is 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; in the cases close to the full-Gaussianity scenario, *fastICA* has a smaller bias and variance when the independent components are super-Gaussian, whilst *DCov* seems to deliver better results when the independent components are sub-Gaussian; *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 former, derived in Gouriéroux et al. (2017) and drawn on asymptotic approximations, outperforms the bootstrap procedure, which we adopt since we do not know the asymptotic distribution of the estimate of B_0's entries under *fastICA*, *DCov* and *CvM*. This should at least happen 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 a ICA model with mixing matrix as specified in equation (8) and with shocks ε_t p-generalized-normally distributed, for different values of p and sample size $T = 400$. For each Monte Carlo replication $m = 1, \ldots, 500$ 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

$$C_m(\alpha) = [\hat{b}_{ij} - \phi_{\alpha/2}\hat{\sigma}_{ij},\ \ b_{ij} + \phi_{\alpha/2}\hat{\sigma}_{ij}], \tag{10}$$

with $\phi_{\alpha/2}$ being the $\alpha/2$-quantile of the standard normal distribution. We do this for each Monte Carlo replication m, to which it corresponds 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 $C_m(\alpha)$, 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 $C_m(\alpha)$ across Monte Carlo replications. We should expect that the frequency of observing the true value b_{ij} in the confidence interval is equal to α.

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:

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

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 $C_m(\alpha)$ 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.

Table 2 shows the results of the simulation exercise for values of $(1 - \alpha) = \{0.99, 0.95, 0.90, 0.75, 0.50\}$. The first result is that bootstrap-based confidence intervals of the main diagonal elements (b_{11}, b_{22}) built under *DCov* estimation tend to contain much more often than expected the true value of the parameter ($\hat{\alpha} < \alpha$). The small sample properties of PML and *fastICA* ensure, instead, smaller size distortions.

For what concerns other parameters' estimates, testing procedures based both on PML and on bootstrap methods remarkably display lower size distortions and *fastICA* seems to have lower size distortion than *DCov* under all the distributional scenarios. Surprisingly, the empirical coverage of *fastICA* estimator seems to be closer to the nominal coverage $(1-\alpha)$ even when compared with PML-based inference, especially in the case the independent components follow sub-Gaussian distributions. However, as stated in Gouriéroux et al. (2017), the choice of the pseudo-likelihoods does indeed matter for the asymptotic accuracy of the PML estimator. Overall, the performance of *CvM* is poorer.

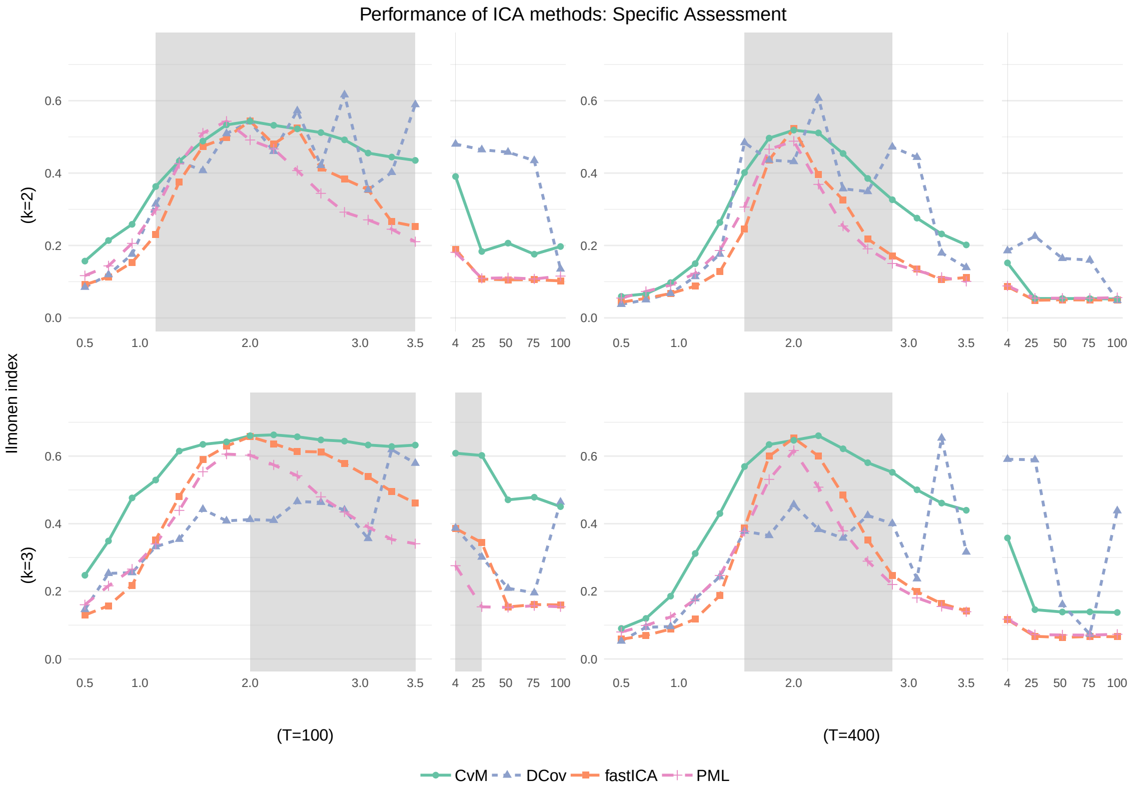

Figure 3: Ilmonen index (y- axis), as specified in equation (9) and averaged across 500 Monte Carlo replications, for different ICA-SVAR models with fixed mixing matrix (equation 8), in which ε_t follow a p-generalized normal distribution, with p (x-axis) varying from the super-Gaussian ($p < 2$), to the Gaussian ($p = 2$), and sub-Gaussian ($p > 2$) case. Shaded area where normality is not rejected at 10% (on the basis of the Jarque-Bera test).

		$p = 0.5$				$p = 1.5$				$p = 2.5$				$p = 100$			
$(k = 2)$		*mean*	*sd*	*kurt*	*skew*	*mean*	*sd*	*kurt*	*skew*	*mean*	*sd*	*kurt*	*skew*	*mean*	*sd*	*kurt*	*skew*
$\hat{b}_{11} - b_{11}$	*PML*	-0.001	0.003	64.5	-6.6	-0.055	0.095	5.7	-2.5	-0.105	0.110	1.2	-1.3	-0.001	0.002	7.6	-2.5
	fastICA	-0.001	0.001	11.5	-3.0	-0.034	0.066	15.0	-3.6	-0.119	0.113	0.2	-1.0	-0.001	0.001	3.4	-1.8
	DCov	-0.001	0.001	18.5	-3.9	-0.105	0.110	0.3	-1.1	-0.083	0.095	2.3	-1.6	-0.001	0.001	5.2	-2.1
	CvM	-0.003	0.019	93.3	-9.2	-0.081	0.105	2.1	-1.6	-0.119	0.118	0.6	-1.2	-0.001	0.002	9.2	-2.6
$\hat{b}_{22} - b_{22}$	*PML*	-0.003	0.019	2.9	-0.5	-0.039	0.100	3.1	-1.7	-0.095	0.148	-0.0	-1.0	-0.001	0.017	0.0	0.0
	fastICA	-0.000	0.013	0.3	0.2	-0.027	0.090	5.3	-2.0	-0.086	0.144	0.1	-1.0	-0.001	0.015	-0.4	0.1
	DCov	-0.000	0.012	1.5	0.2	-0.085	0.139	0.3	-1.1	-0.045	0.117	2.5	-1.6	-0.000	0.015	-0.1	0.0
	CvM	-0.003	0.036	96.8	-7.3	-0.069	0.135	1.1	-1.4	-0.082	0.142	0.1	-1.0	0.000	0.017	0.3	-0.1
$\hat{b}_{12} - b_{12}$	*PML*	0.005	0.056	2.8	0.4	-0.066	0.330	0.2	-0.5	-0.062	0.462	-1.2	-0.1	-0.001	0.052	0.0	-0.1
	fastICA	-0.002	0.040	0.4	-0.2	-0.031	0.266	1.1	-0.5	-0.136	0.475	-1.3	0.1	-0.001	0.046	-0.4	-0.2
	DCov	-0.001	0.037	1.6	-0.2	-0.095	0.456	-1.2	-0.1	-0.143	0.392	-0.9	0.1	-0.002	0.045	-0.3	-0.1
	CvM	0.001	0.087	24.9	-0.8	-0.068	0.403	-0.7	-0.2	-0.149	0.470	-1.2	0.1	-0.004	0.050	-0.0	-0.1
$\hat{b}_{21} - b_{21}$	*PML*	0.001	0.057	2.7	-0.3	0.089	0.383	0.6	0.8	0.102	0.528	-1.1	0.3	0.001	0.057	-0.1	0.1
	fastICA	0.000	0.046	2.1	-0.3	0.046	0.304	1.7	0.8	0.190	0.544	-1.3	0.1	0.001	0.049	-0.1	0.0
	DCov	-0.001	0.039	3.5	-0.6	0.140	0.525	-1.1	0.3	0.186	0.453	-1.0	0.1	0.002	0.048	0.3	-0.0
	CvM	-0.002	0.087	23.4	-0.4	0.102	0.466	-0.5	0.4	0.204	0.542	-1.2	0.0	0.004	0.053	0.9	0.0
$(k = 3)$																	
$\hat{b}_{11} - b_{11}$	*PML*	-0.190	0.036	1.7	-0.3	-0.212	0.149	0.1	-0.7	-0.209	0.180	-0.6	-0.6	-0.192	0.031	0.0	0.1
	fastICA	-0.188	0.025	0.4	-0.0	-0.213	0.149	0.3	-0.7	-0.216	0.185	-0.6	-0.6	-0.187	0.027	-0.0	-0.1
	DCov	-0.189	0.019	4.2	-0.5	-0.272	0.145	0.0	-0.4	-0.261	0.153	-0.2	-0.3	-0.162	0.076	2.9	0.5
	CvM	-0.189	0.043	8.0	0.4	-0.208	0.168	-0.5	-0.4	-0.196	0.179	-0.4	-0.7	-0.192	0.065	5.7	-1.3
$\hat{b}_{22} - b_{22}$	*PML*	-0.128	0.040	4.2	0.8	-0.155	0.143	-0.4	-0.4	-0.152	0.168	-0.6	-0.4	-0.130	0.035	-0.0	-0.0
	fastICA	-0.128	0.027	0.7	0.2	-0.142	0.136	0.1	-0.6	-0.131	0.171	-0.1	-0.6	-0.128	0.030	0.1	-0.1
	DCov	-0.128	0.023	1.2	0.1	-0.239	0.165	-0.8	-0.1	-0.213	0.155	-0.4	-0.1	-0.195	0.106	-0.6	-0.2
	CvM	-0.128	0.044	3.4	0.2	-0.145	0.164	-0.2	-0.5	-0.148	0.174	-0.6	-0.4	-0.131	0.066	2.1	-0.6
$\hat{b}_{12} - b_{12}$	*PML*	0.154	0.064	2.9	-0.4	0.082	0.312	0.7	-0.8	-0.049	0.439	-1.1	-0.1	0.159	0.055	0.3	0.2
	fastICA	0.158	0.042	0.3	-0.1	0.070	0.341	0.5	-0.8	-0.136	0.438	-1.1	0.1	0.158	0.047	-0.4	0.2
	DCov	0.159	0.034	1.7	0.3	0.127	0.402	0.2	-1.1	0.136	0.465	-0.7	-0.9	0.321	0.260	0.5	-0.6
	CvM	0.159	0.074	10.6	0.5	-0.048	0.432	-1.0	-0.2	-0.076	0.430	-1.1	-0.0	0.144	0.118	5.3	-0.4
$\hat{b}_{21} - b_{21}$	*PML*	-0.349	0.064	5.4	1.0	-0.163	0.411	2.5	1.6	0.159	0.528	-1.1	0.3	-0.357	0.050	-0.2	0.2
	fastICA	-0.351	0.041	0.3	-0.0	-0.129	0.423	1.1	1.3	0.269	0.528	-1.2	0.1	-0.349	0.048	-0.0	0.1
	DCov	-0.352	0.035	0.9	-0.3	-0.232	0.557	0.5	1.3	-0.090	0.649	-0.7	0.9	-0.256	0.218	7.3	2.5
	CvM	-0.347	0.070	5.1	0.8	0.109	0.542	-0.9	0.5	0.214	0.540	-1.1	0.1	-0.331	0.128	45.8	4.1

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 $T = 400$.

		$p=0.5$					$p=1.5$					$p=2.5$					$p=100$				
$(k=2)$	$(1-\alpha)$	0.99	0.95	0.90	0.75	0.50	0.99	0.95	0.90	0.75	0.50	0.99	0.95	0.90	0.75	0.50	0.99	0.95	0.90	0.75	0.50
b_{11}	*PML*	0.98	0.96	0.95	0.93	0.87	0.74	0.70	0.66	0.59	0.49	0.60	0.52	0.48	0.44	0.36	0.97	0.94	0.92	0.86	0.69
	fastICA	1.00	0.98	0.94	0.82	0.51	1.00	1.00	0.97	0.94	0.84	1.00	1.00	0.99	0.96	0.86	1.00	0.99	0.97	0.79	0.55
	DCov	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.93	0.86	1.00	1.00	0.99	0.95	0.86	1.00	0.89	0.74	0.59	0.37
	CvM	0.77	0.58	0.38	0.11	0.01	0.96	0.84	0.57	0.10	0.01	0.89	0.67	0.31	0.05	0.00	1.00	1.00	1.00	1.00	0.14
b_{22}	*PML*	0.96	0.93	0.88	0.78	0.55	0.94	0.89	0.85	0.78	0.59	0.99	0.99	0.98	0.92	0.77	1.00	0.99	0.97	0.88	0.60
	fastICA	0.99	0.96	0.88	0.73	0.52	1.00	1.00	1.00	0.88	0.56	1.00	1.00	1.00	0.92	0.54	0.99	0.95	0.89	0.74	0.48
	DCov	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.96	0.89	0.77	1.00	0.99	0.98	0.92	0.68	1.00	1.00	1.00	0.80	0.57
	CvM	0.60	0.24	0.10	0.00	0.00	0.93	0.64	0.59	0.02	0.00	1.00	0.82	0.61	0.02	0.00	0.82	0.71	0.63	0.55	0.00
b_{12}	*PML*	0.96	0.91	0.87	0.76	0.55	0.92	0.86	0.81	0.68	0.50	1.00	0.99	0.98	0.95	0.76	1.00	0.99	0.97	0.88	0.60
	fastICA	0.99	0.95	0.89	0.72	0.51	0.97	0.96	0.94	0.84	0.55	0.97	0.96	0.95	0.83	0.56	0.99	0.94	0.87	0.74	0.48
	DCov	1.00	0.97	0.93	0.76	0.52	1.00	0.98	0.94	0.84	0.61	1.00	1.00	0.96	0.84	0.56	0.78	0.61	0.58	0.37	0.34
	CvM	0.61	0.25	0.11	0.01	0.00	0.98	0.52	0.48	0.00	0.00	1.00	0.49	0.42	0.02	0.00	0.71	0.59	0.54	0.48	0.01
b_{21}	*PML*	0.94	0.87	0.81	0.62	0.41	0.91	0.86	0.80	0.67	0.49	1.00	0.99	0.98	0.96	0.79	1.00	0.99	0.97	0.87	0.68
	fastICA	1.00	0.95	0.91	0.73	0.49	0.97	0.96	0.93	0.84	0.56	0.97	0.96	0.94	0.83	0.59	0.99	0.97	0.91	0.79	0.52
	DCov	1.00	0.98	0.96	0.88	0.62	1.00	0.98	0.94	0.84	0.58	1.00	0.99	0.96	0.83	0.57	0.82	0.61	0.49	0.37	0.35
	CvM	0.60	0.27	0.08	0.01	0.00	0.96	0.52	0.49	0.01	0.00	1.00	0.48	0.43	0.02	0.00	0.72	0.63	0.56	0.48	0.01
$(k=3)$																					
b_{11}	*PML*	0.98	0.95	0.92	0.86	0.66	0.99	0.98	0.92	0.82	0.61	1.00	0.99	0.98	0.95	0.81	1.00	0.98	0.96	0.63	0.63
	fastICA	1.00	0.96	0.89	0.73	0.46	1.00	1.00	0.98	0.83	0.60	1.00	1.00	0.97	0.82	0.55	1.00	0.96	0.91	0.72	0.52
	DCov	1.00	1.00	0.99	0.95	0.88	1.00	1.00	1.00	0.95	0.72	1.00	1.00	1.00	0.85	0.50	1.00	1.00	1.00	0.98	0.76
	CvM	1.00	0.90	0.75	0.28	0.02	1.00	0.98	0.83	0.56	0.36	1.00	0.93	0.69	0.46	0.43	0.99	0.89	0.70	0.25	0.04
b_{22}	*PML*	0.84	0.74	0.65	0.49	0.34	0.98	0.93	0.89	0.75	0.53	1.00	0.99	0.99	0.93	0.80	0.99	0.99	0.95	0.82	0.55
	fastICA	1.00	0.98	0.92	0.73	0.46	1.00	1.00	0.98	0.91	0.62	1.00	1.00	0.97	0.81	0.58	1.00	0.96	0.95	0.84	0.59
	DCov	1.00	1.00	0.99	0.94	0.86	1.00	1.00	0.99	0.93	0.70	1.00	0.99	0.97	0.79	0.35	1.00	1.00	1.00	0.98	0.85
	CvM	0.99	0.92	0.64	0.23	0.02	1.00	0.99	0.86	0.59	0.50	1.00	1.00	0.91	0.55	0.45	0.99	0.93	0.70	0.33	0.02
b_{12}	*PML*	0.87	0.80	0.72	0.57	0.36	0.97	0.92	0.88	0.76	0.54	0.99	0.98	0.97	0.92	0.79	1.00	0.99	0.97	0.87	0.64
	fastICA	1.00	0.98	0.94	0.79	0.58	0.97	0.93	0.90	0.82	0.59	0.98	0.91	0.86	0.76	0.54	1.00	0.97	0.92	0.81	0.56
	DCov	0.99	0.97	0.95	0.88	0.69	0.93	0.85	0.78	0.56	0.25	0.98	0.54	0.45	0.36	0.15	1.00	1.00	0.99	0.96	0.83
	CvM	1.00	0.90	0.68	0.25	0.02	1.00	1.00	1.00	0.85	0.17	1.00	1.00	1.00	0.99	0.30	1.00	0.94	0.82	0.32	0.03
b_{21}	*PML*	0.98	0.95	0.91	0.81	0.60	0.95	0.91	0.88	0.80	0.62	0.97	0.95	0.93	0.89	0.77	1.00	1.00	0.99	0.89	0.67
	fastICA	0.99	0.95	0.89	0.73	0.48	0.93	0.91	0.88	0.83	0.53	0.89	0.84	0.83	0.75	0.56	0.98	0.96	0.91	0.76	0.51
	DCov	0.98	0.98	0.97	0.92	0.75	0.86	0.81	0.80	0.74	0.56	0.54	0.48	0.45	0.32	0.15	1.00	0.99	0.99	0.94	0.57
	CvM	0.98	0.91	0.69	0.22	0.01	1.00	1.00	0.90	0.49	0.21	1.00	1.00	1.00	0.88	0.27	1.00	0.88	0.72	0.32	0.03

Table 2: Empirical coverage $(1-\hat{\alpha})$ of confidence intervals for the estimates of the mixing matrix entries. 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 a *warp*-bootstrap procedure.

4 Empirical application

In this section we discuss a macroeconomic application of the ICA approach to SVAR analysis with the aim of showing its practical potentials and challenges, having taken into account the results of the previous section. A conspicuous stream of literature has attempted to estimate, quantify and identify the effects of monetary and fiscal policy interventions on main macroeconomic aggregates in many developed countries. After the 2007 financial crisis triggered what economists and policy makers referred to as the Great Recession, a fresher interest in fiscal policy emerged. At the same time, the efficacy of a pure monetary policy was questioned. In spite of unconventional measures, monetary policy has struggled to push the economy out of a low-growth and stagnant-prices regime, in a situation of a prolonged zero-lower-bound scenario. Given the reappraisal of fiscal policy, the question on the size of multipliers has become again highly disputed.

Focusing on the empirical literature, Barro (1981) and, more recently, Ramey (2011) attempt to overcome the identification problem by relying on narrative series of military buildups, so to identify an anticipated and exogenous increase in government spending. Blanchard and Perotti (2002) for the US and Ilzetzki et al. (2013) for 44 countries, estimate elasticities of tax revenues to output in a quarter so to purge out the contemporaneous effect of output on government revenues. They also use institutional knowledge to insert zero contemporaneous impact restrictions. Recently, Auerbach and Gorodnichenko (2012) have identified exogenous spending shocks as those not foreseen by US professional forecasters. Batini et al. (2012) analize the effects of government expenditure and revenues increase for the main developed economies by relying on the Choleski decomposition and so imposing a recursive causal order on the variables included in the VAR. Mountford and Uhlig (2009), instead, in a Bayesian VAR framework with sign restrictions, identify spending and tax shocks as those that increase the fiscal variables while being uncorrelated to economic activity and monetary policy.

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 of on fiscal policy by Blanchard and Perotti (2002), BP henceforth. 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.

Despite the plausibility of the identification strategy, it is important to test such restrictions, at least for two reasons: (i) as Caldara and Kamps (2017) show, the use of plausible range of estimated elasticities may lead to dynamical responses and fiscal shocks that significantly differ in size and persistence; (ii) the authors are not able to distinguish the contemporaneous relationship between government spending and tax revenues and, consequently, whether a tax shock has an immediate effect on spending.

Table 3 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.

	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

Table 3: Impact coefficients in Blanchard and Perotti (2002)

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 two-fold: (i) using our *Maxfinder* algorithm, to globally identify via the ICA model an independent shock in government spending and tax increase and estimate their dynamic responses on economic activity; (ii) to test the validity of restrictions proposed by BP, relying on the results on statistical inference obtained in Section 3.2.

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

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

where C is $k \times p$ matrix of deterministic terms, β is a trend coefficient, y_t is the vector of endogenous variables, $Q1, Q2, Q3$ are quarter dummies, B_l and B_l^Q are coefficient matrices, and u_t is k-dimensional vector of reduced-form residuals. In our case $k = 3$ and $q = 4$, as in BP. Also following BP, in C we formalize the constant, a series of deterministic terms, as well as current and lagged values of the endogenous variables interacted with quarter dummies[7]. In line with the setting of the ICA 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. After estimation, the VAR is stable with no-serial autocorrelation. Moreover, as Figure 4 suggests, only for the spending's reduced-form residual, u_t^G, the Jarque-Bera test does not reject the null hypothesis of normality, so that ICA can be applied, since a single exception to non-Gaussianity is allowed in the ICA model.

Table 4 reports the estimates $\tilde{b}_{ij}$ of the entries b_{ij} of the mixing matrix B_0, estimated with the four ICA methods and identified by applying our *MaxFinder* scheme. 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

[7]The series of deterministic terms include a 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. Our analysis departs from the original specification only because it excludes a quadratic time trend. Since the variables are logs of per capital value, it seems a reasonable choice. However, this choice does not change the nature of results in both settings.

root $(\hat{b}^*_{ij} - \tilde{b}_{ij})$, with ($\hat{b}^*_{ij}$ being the estimate of b_{ij} from the bootstrap replication). 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 PML and *fastICA*, the *Maxfinder* algorithm does not deliver a maximum for each column on different rows. However, the column permutation that penalizes low absolute values on the main diagonal (step 3) delivers an interesting result: we have identified two independent shocks that increase taxes and government spending the most, and a third shock that increases substantially output and taxes. Moreover, this configuration is compatible with Table 3 derived in BP. Given the results, 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 (with exception of *CvM* where none of the estimated coefficients seems to be statistically different from zero). From PML and *fastICA* we get an estimate of a positive and statistically significant contemporaneous response of the third shock on tax revenues, but which is substantially lower than BP's estimate of the GDP shock on tax revenues. At the same time from the same method 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 *DCov* 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). Finally, coefficients estimated under *CvM* do not appear to be significant and *DCov*'s seem more difficult to interpret. 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.

After estimating structural shocks and 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 5. We first comment the impulse responses to a positive public spending shock, shown in Figure 6. The response of output is clear and statistically significant, both at impact and within the first year. Figure 7 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 *DCov*) show non-significant effect on the impact and in the short run, which is at odds with the finding by BP.

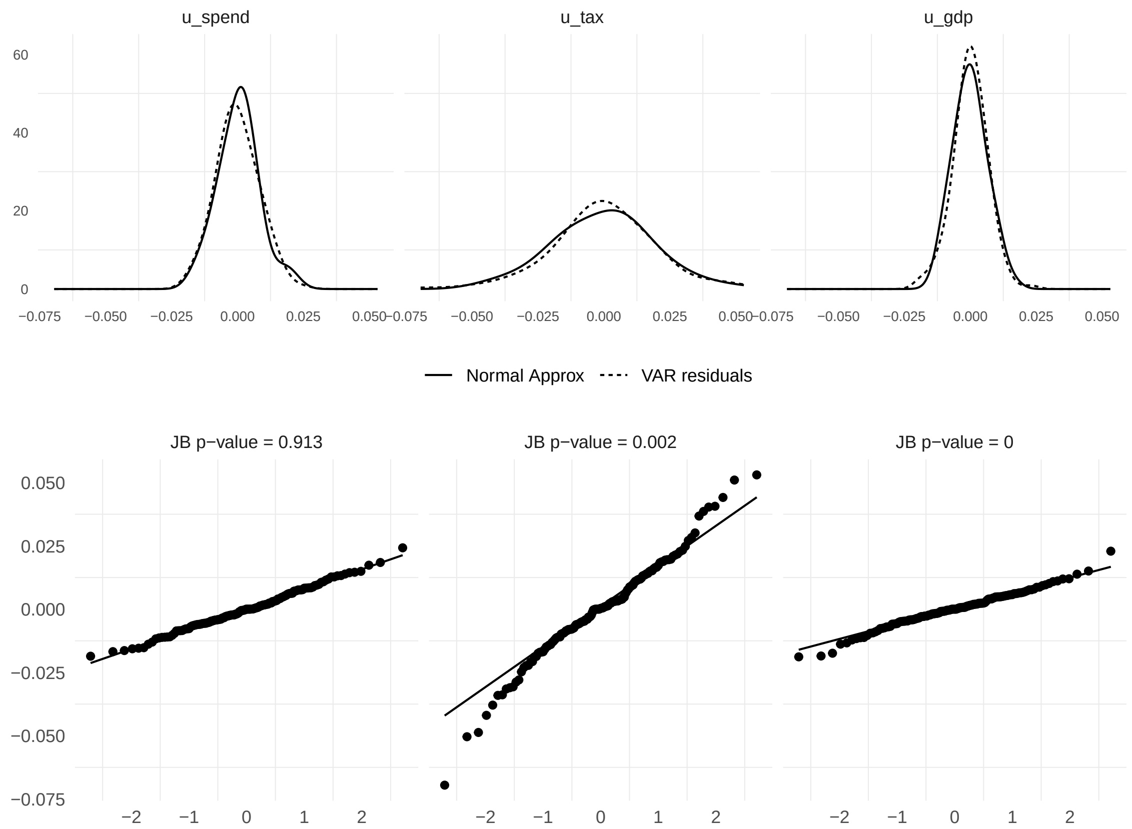

Figure 4: Distribution (upper panels) and q-q plots (bottom panels) of reduced-form residuals, where also p-values of the Jarque-Bera (JB) test are reported.

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. The BP identifying restrictions are plausible not only because of institutional knowledge and insights from economic theory: they are also present in the data and the ICA 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 it is public spending 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.

	PML			*fastICA*		
	ε^1	ε^2	ε^3	ε^1	ε^2	ε^3
G_t	1	−0.06	−0.12	1	−0.09	−0.10
TAX_t	0.47	1	1.07***	0.62	1	1.14***
GDP_t	0.24***	0.002	1	0.25*	−0.01	1
	DCov			*CvM*		
G_t	1	0.08	−0.33*	1	−0.05	−0.001
TAX_t	−0.27	1	0.21	0.39	1	0.79
GDP_t	0.34**	0.13*	1	0.19	0.04	1

Estimate falls in 64% (*), 90% (**), 95% (***) confidence interval

Table 4: Estimates of the contemporaneous impacts coefficients from ICA models. The entries of the mixing matrices are calculated using the bootstrap-median with the corresponding confidence intervals

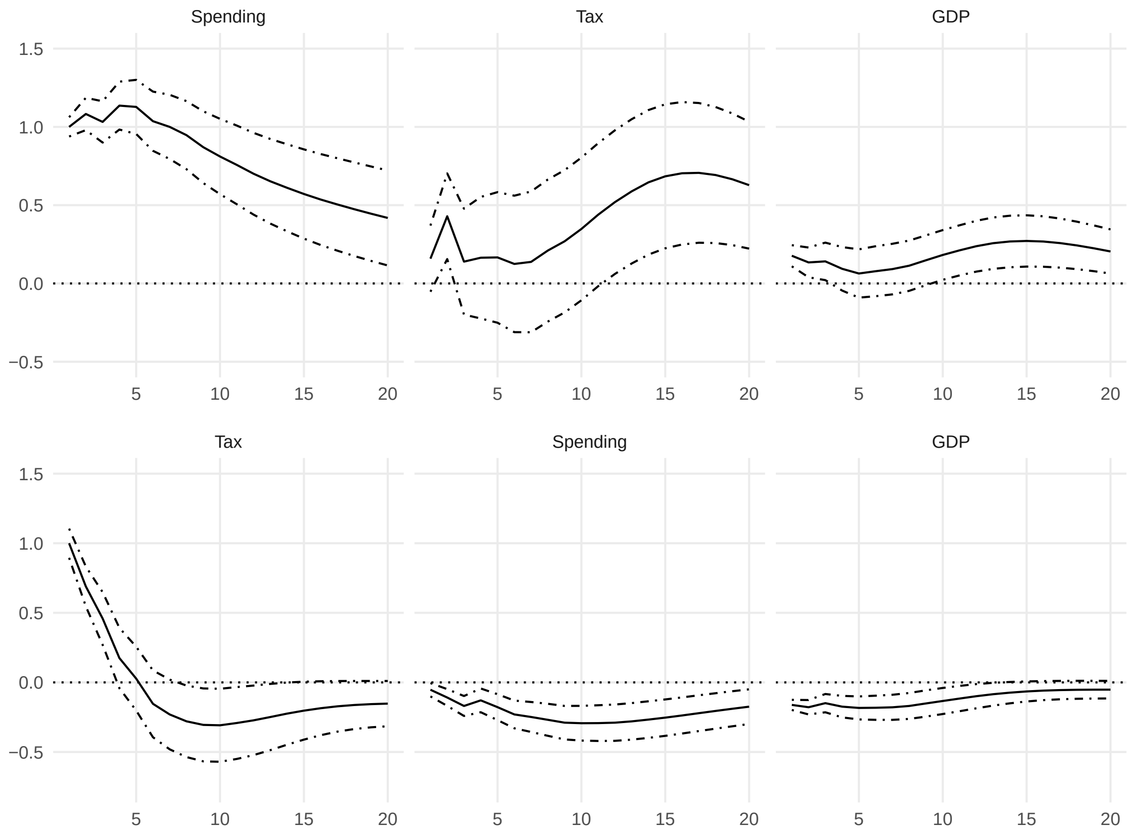

Figure 5: Impulse response functions estimated by Blanchard and Perotti (2002). The upper panel shows the responses of the three variables to a spending shock (on G). The bottom panel shows the responses to a tax-revenue shock (on Tax). Dashed lines denote an equal-tailed 68% confidence interval

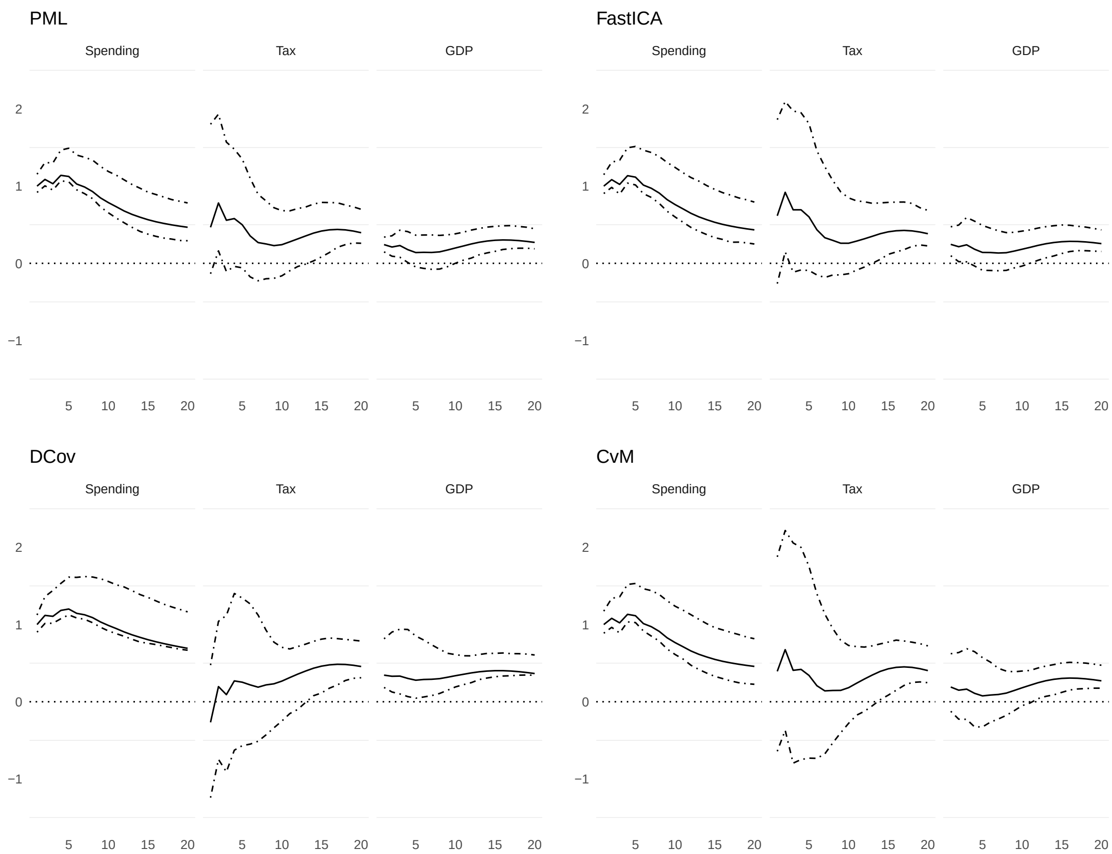

Figure 6: Impulse response functions of a positive public spending shock from different ICA estimation methods. The upper and the lower dashed lines represent respectively the 84% quantile and the 16% quantile of the bootstrap estimates.

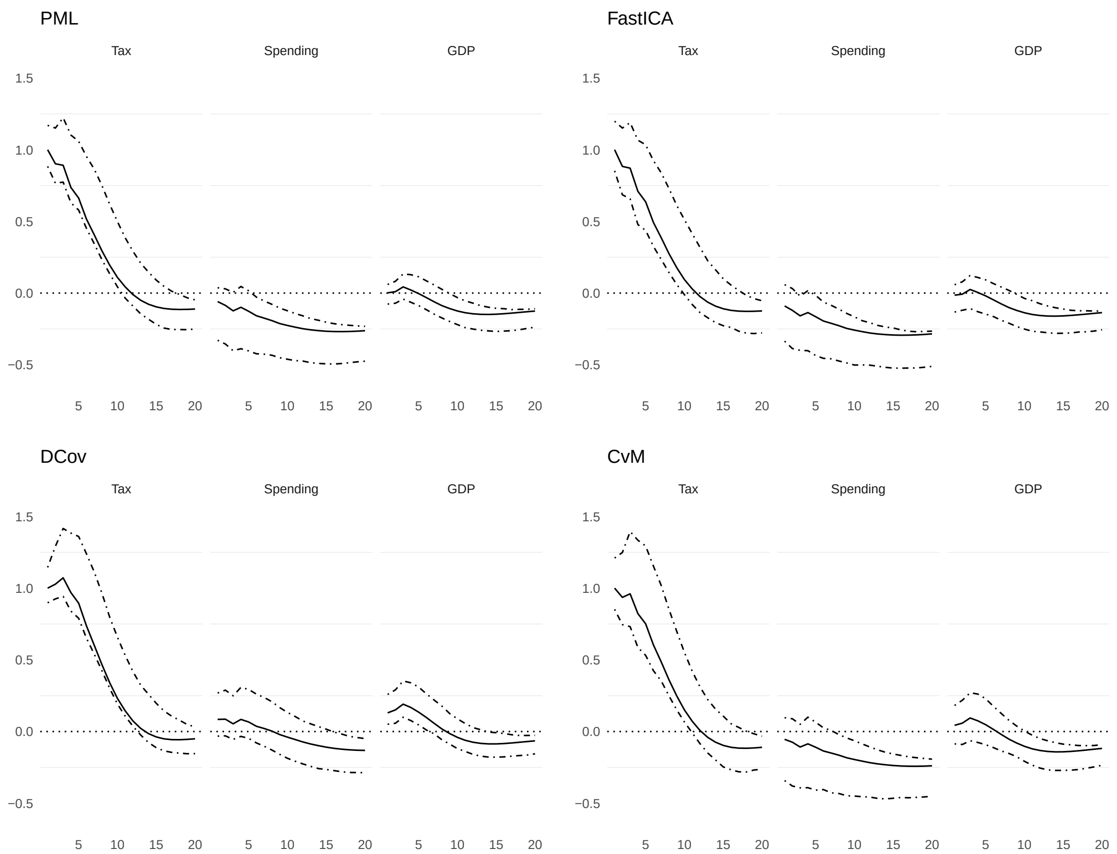

Figure 7: Impulse response functions of a positive tax shock from different ICA estimation methods. The upper and the lower dashed lines represent respectively the 84% quantile and the 16% quantile of the bootstrap estimates.

5 Conclusions

In this paper we assess, through Monte Carlo experiments, the performance of four ICA techniques (*fastICA*, *DCov*, *CvM* , PML) that have been recently used in SVAR analysis. We specifically study the cases of structural disturbances following distributions that are approaching normality, where the ICA model by construction cannot recover the independent components. The method based on distance covariance seems to be, on average, the most performing when the shocks' distributions are relatively close to be Gaussian.

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. In this context, we suggest the implementation of an identification scheme that searches for those column permutations of the mixing matrix such that its main diagonal elements reflect, as close as possible, the condition according to which the structural shock hits with greatest magnitude the variable is referring to. 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.

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 ICA model cannot reject the identification scheme implemented in Blanchard and Perotti (2002).

References

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Back, A. D. and A. S. Weigend (1997). A first application of independent component analysis to extracting structure from stock returns. *International journal of neural systems 8*(04), 473–484.

Barro, R. J. (1981). Output effects of government purchases. *Journal of political Economy 89*(6), 1086–1121.

Batini, N., G. Callegari, and G. Melina (2012). Successful austerity in the United States, Europe and Japan. IMF Working Papers 12/190, International Monetary Fund.

Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy 25*, 49–99.

Bernanke, B. S. and I. Mihov (1998). Measuring monetary policy. *The quarterly journal of economics 113*(3), 869–902.

Blanchard, O. and R. Perotti (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. E. P. and G. C. Tiao (1962). A further look at robustness via Bayes's theorem. *Biometrika 49*(3/4), 419–432.

Caldara, D. and C. Kamps (2017). The analytics of SVARs: a unified framework to measure fiscal multipliers. *The Review of Economic Studies 84*(3), 1015–1040.

Capasso, M. and A. Moneta (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,*, pp. 2109–2112. IEEE.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing 36*(3), 287–314.

Eriksson, J. and V. Koivunen (2004). Identifiability, separability, and uniqueness of linear ICA models. *IEEE signal processing letters 11*(7), 601–604.

Genest, C., J.-F. Quessy, B. Rémillard, et al. (2007). Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *The Annals of Statistics 35*(1), 166–191.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Giacomini, R., D. N. Politis, and H. White (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory 29*(3), 567–589.

Goodman, I. R. and S. Kotz (1973). Multivariate θ-generalized normal distributions. *Journal of Multivariate Analysis 3*(2), 204–219.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics 196*(1), 111–126.

Gouriéroux, C., A. Monfort, and J. P. Renne (2020). Identification and estimation in non-fundamental structural VARMA models. *The Review of Economic Studies 87*(4), 1915–1953.

Guerini, M., A. Moneta, M. Napoletano, and A. Roventini (2018). The Janus-faced nature of debt: Results from a data-driven cointegrated SVAR approach. *Macroeconomic Dynamics*, 1–31.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, 675–694.

Hastie, T. and R. Tibshirani (2010). *ProDenICA: Product Density Estimation for ICA using tilted Gaussian density estimates*. R package version 1.0.

Herwartz, H. (2018a). Hodges–Lehmann detection of structural shocks–an analysis of macroeconomic dynamics in the Euro area. *Oxford Bulletin of Economics and Statistics*.

Herwartz, H. (2018b). Long-run neutrality of demand shocks: Revisiting Blanchard and Quah (1989) with independent structural shocks. *Journal of Applied Econometrics*.

Herwartz, H. and M. Plödt (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. and C. R. Johnson (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., J. Karhunen, and E. Oja (2001). *Independent component analysis*. John Wiley & Sons.

Hyvärinen, A. and E. Oja (1997). A fast fixed-point algorithm for independent component analysis. *Neural computation 9*(7), 1483–1492.

Hyvärinen, A. and E. Oja (2000). Independent component analysis: algorithms and applications. *Neural networks 13*(4-5), 411–430.

Ilmonen, P., K. Nordhausen, H. Oja, and E. Ollila (2010). A new performance index for ICA: properties, computation and asymptotic analysis. In *International Conference on Latent Variable Analysis and Signal Separation*, pp. 229–236. Springer.

Ilzetzki, E., E. G. Mendoza, and C. A. Végh (2013). How big (small?) are fiscal multipliers? *Journal of Monetary Economics 60*(2), 239–254.

Kalke, S. (2015). *pgnorm: The p-Generalized Normal Distribution*. R package version 2.0.

Kalke, S. and W.-D. Richter (2013). Simulation of the p-generalized Gaussian distribution. *Journal of Statistical Computation and Simulation 83*(4), 641–667.

Kilian, L. and H. Lütkepohl (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Klein, L. R. and A. S. Goldberger (1955). An econometric model of the United States.

Lange, A., B. Dalheimer, H. Herwartz, and S. Maxand (2018). *svars: Data-Driven Identification of SVAR Models*. R package version 1.2.0.

Lanne, M. and J. Luoto (2019). GMM estimation of non-Gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, 1–13.

Lanne, M. and H. Lütkepohl (2010). Structural vector autoregressions with nonnormal residuals. *Journal of Business & Economic Statistics 28*(1), 159–168.

Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics 196*(2), 288–304.

Lanne, M. and P. Saikkonen (2013). Noncausal vector autoregression. *Econometric Theory 29*(3), 447–481.

Matteson, D. S. and R. S. Tsay (2017). Independent component analysis via distance covariance. *Journal of the American Statistical Association 112*(518), 623–637.

Maxand, S. (2018). Identification of independent structural shocks in the presence of multiple Gaussian components. *Econometrics and Statistics*.

Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics 75*(5), 705–730.

Mountford, A. and H. Uhlig (2009). What are the effects of fiscal policy shocks? *Journal of applied econometrics 24*(6), 960–992.

Nordhausen, K., E. Ollila, and H. Oja (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*, pp. 486–490. IEEE.

Price, K. V., R. M. Storn, and J. A. Lampinen (2006, January). *Differential Evolution - A Practical Approach to Global Optimization*. Natural Computing. Springer-Verlag. ISBN 540209506.

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. and D. H. Romer (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., P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research 7*(Oct), 2003–2030.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 1–48.

Sims, C. A. (1986). Are forecasting models usable for policy analysis? *Quarterly Review, Federal Reserve Bank of Minneapolis* (Winter), 2–16.

Stock, J. H. and M. W. Watson (2001). Vector autoregressions. *Journal of Economic perspectives 15*(4), 101–115.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, Volume 2, pp. 415–525. Elsevier.

Stock, J. H. and M. W. Watson (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., M. L. Rizzo, N. K. Bakirov, et al. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics 35*(6), 2769–2794.

Tiao, G. and D. R. Lund (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., A. Zhang, and H. Li (2012). A review of estimating the shape parameter of generalized Gaussian distribution. *J. Comput. Inf. Syst 8*(21), 9055–9064.

Appendix

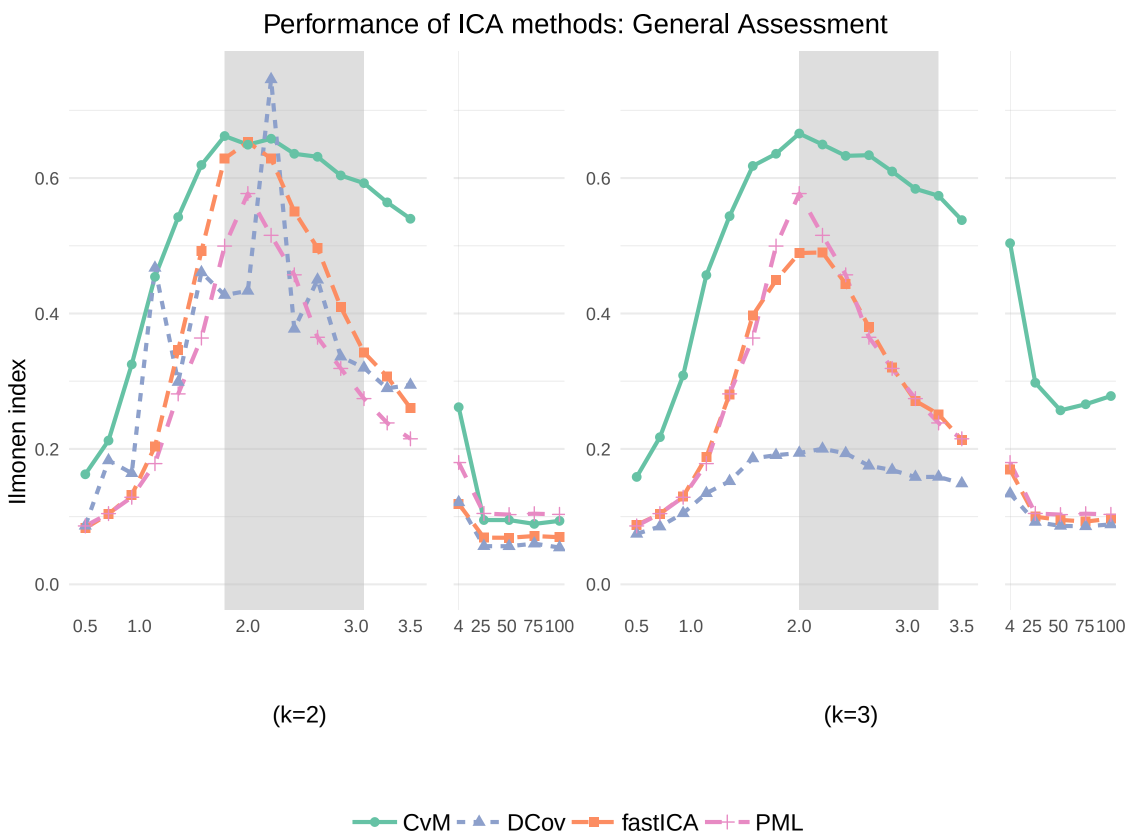

Figure 8: General Assessment of ICA methods when $n = 200$, $k = 2$ (left panel) and $k = 3$ (right panel)

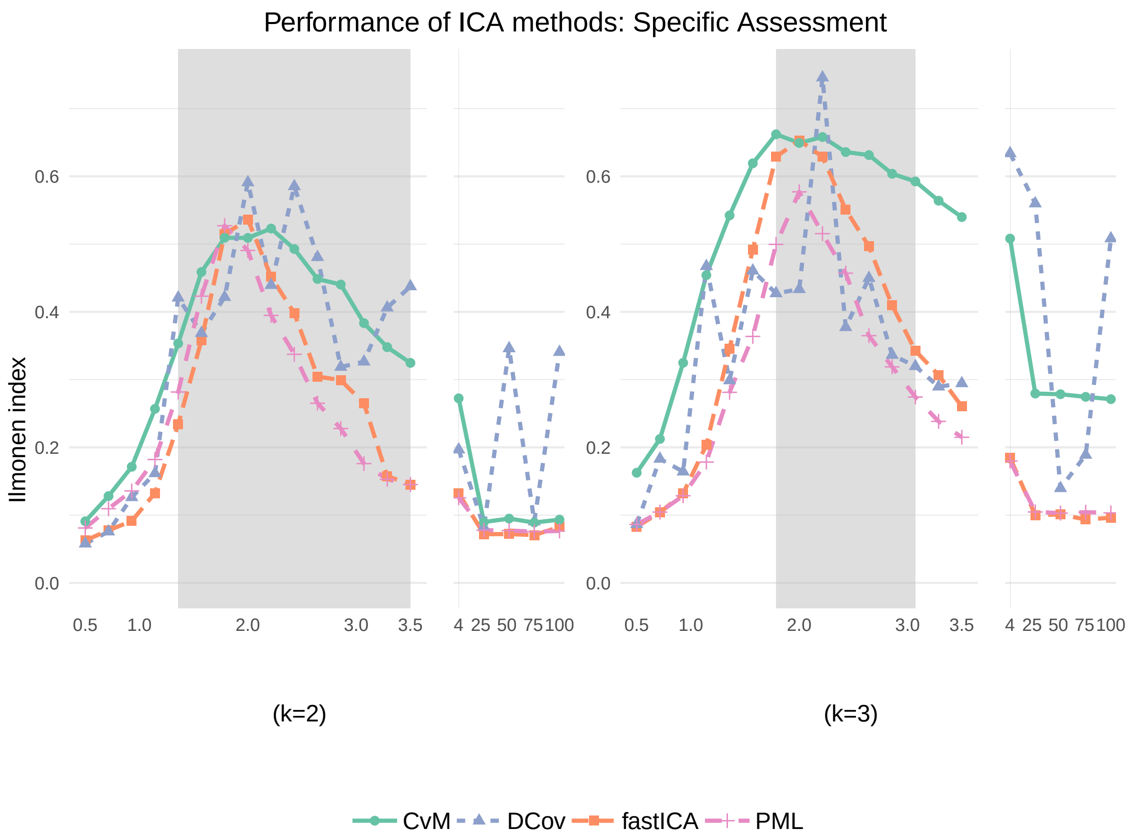

Figure 9: Specific Assessment of ICA methods when $n = 200$, $k = 2$ (left panel) and $k = 3$ (right panel)