Non-Gaussian shocks and recursive monetary policy SVARs for the Euro Area

Filippo Palandri

November 24, 2024

Abstract

In order to deal with the identification problem in Macroeconomics a common assumption of orthogonality and unit variance of the shock is made but this is not enough in the Gaussian framework as any orthogonal transformation will be observationally equivalent, so short run restrictions are employed so that the structural matrix is recovered from the Cholesky decomposition of the covariance matrix of the residuals. This, in practice, limits the immediate response of some variables to some shocks to zero. New Keynesian monetary policy DSGE models instead consider the structural matrix of interest to be full in principle, hence under these conditions the structural shock of interest would be misidentified if recovered through Cholesky. Interesting a non-Gaussian assumption about the nature of the shocks allow for an identification up to permutation and scaling of the columns of the structural matrix, which is enough to conduct inference about the lower triangular structure imposed by Cholesky. I hereby use both parametric and non-parametric methods borrowed from the statistical ICA literature to identify the structural shocks on a small scale monetary policy SVAR in the Euro Area finding not enough statistical evidence to reject the recursive identification scheme.

1 Introduction

The aim of this work is to test if recursive monetary SVARs holds in practice when tested against the data in the Euro Area. The need for such a question is that considering small scale DSGE models, the structural matrix in this framework is considered to be full. Hence, given that SVAR models have the ambition to recover those structural shocks from the data by imposing only a minimum of structure, using Cholesky to identify the shocks with a recursive identification scheme would bias the correct estimates as the SVAR would not be able to retrieve the original shocks.

Given that the Cholesky SVAR is only exactly identified I will exploit the non-Gaussianity of the shocks to statistically identify the structural matrix by considering both parametric and non-parametric approaches and then testing its lower triangularity.

In the first part I will go through the theory, analyzing the concept behind identification by non-Gaussianity and identification in general, in the second part I will apply those methods estiamating a trivariate monetary policy SVAR with Euro Area data and then assessing the recursive identification scheme.

Contents

2 Literature Review

Since their introduction by [Sims(1980)], SVAR models have been largely used for macroeconomic analysis and the evaluation of dynamic causal effects. SVAR models revealed to represent a good choice for researcher that intended to evaluate the effects of monetary or fiscal policy on the economy, helping to answer those fundamental question in the macroeconomics literature such as "How long does it take for inflation to respond to monetary policy?","What is the fiscal multiplier?", "Has the monetary policy real effects on the economy?", "Are we living in a New Keynesian or Neo-Classical world?". More recently, SVAR's have been used also to address the exogeneity of macroeconomic uncertainty ([Ludvigson et al.(2021)Ludvigson, Ma, and Ng]), relationship between monetary policy and the financial sector, and the effect of Financial stress shocks on real variables [Brunnermeier et al.(2021)Brunnermeier, Palia, Sastry, and Sims]. Those questions are of course high dimensional issues and have still today no trivial answer, but SVAR models had been largely employed to empirically support or object the related economic theories. Despite their popularity, SVAR models had always been object of discussion and critique because of their controversial issue of identification. The first approach to identify the structural shocks was to impose a recursive scheme on the matrix of structural parameters via the **Cholesky decomposition**. Generalizations to this approach were introduced by [Blanchard and Watson(1986)], [Bernanke(1986)], we can thus broadly refer to this approach as **short-run restrictions**. Although its simplicity this method imposes a very strong structure on the dynamics of the model, so other solutions have been proposed.
[Blanchard and Quah(1989)] introduces **long-run restrictions** by exploiting the fact that in a model with demand and supply shocks only the latter can have a permanent effect on output. More recent proposed solutions followed a softer approach, by imposing **sign restrictions** instead of specific values as the previous methods do. [Uhlig(2005)] imposes sign restrictions on the impulse responses of prices, non-borrowed reserves and the federal funds rate in response to a monetary policy shock so that its effects on output could have been freely estimated.
In the context of monetary policy shock, another possible approach is to identify shocks by using high-frequency data, focusing on narrow time windows around central bank announcements in order to isolate the exogenous component of monetary policy shocks from others confounding factors like
[Gürkaynak et al.(2005)Gürkaynak, Sack, and Swanson]; this is the so called **High-Frequency Identification** procedure. Another remarkable example is
[Nakamura and Steinsson(2018)] in which the authors perform high-frequency identification with US data to assess monetary non-neutrality. A new method is the **Proxy SVAR**, firstly developed by Stock and Watson, and then [Stock and Watson(2012)] [Mertens and Ravn(2013)]; the idea is to introduce external information into the VAR model that could be seen as a noisy measure of the true shock. If the classical *relevance* and *exogeneity* condition of the instrument are satisfied, then it could be exploited to obtain identification of the shock. In practice finding such an instrument is far from being straightforward.

A very recent literature [Lewis(2024)] focused in statistical identification techniques by exploiting higher moments without relying on economic theory based restrictions. Those methodologies can be grouped into two class, one exploit the time changing of the second moment of the structural shocks (heteroskedasticity) and so given that there is more than one covariance matrix to exploit the SVAR model could be healed from underidentification. **Heteroskedasticity-Based Identification** requires the volatility regime of the vari-

ance to change over time in order to reach indentification of the model. Examples in the literature are [Sentana and Fiorentini(2001)] and [Lanne and Ltkepohl(2008)] among the others. The second part of the literature aims at exploiting high order moments in the distribution of the structural shocks and is denoted as identification via **non-Gaussianity** as the shocks must be non-Gaussian in order to exploit higher order moments to achieve identification. This whole literature is based on the result derived in [Comon(1994)] according to, a decomposition of the form $\boldsymbol{u_t} = \mathbf{B}\boldsymbol{\varepsilon_t}$ is unique up to column order and scale provided the components $\boldsymbol{\varepsilon_t}$ are independent and at most of them is Gaussian. Those two literatures are very appealling at first sight as, they do not require particulars economic restrictions to obtain identification. Indeed both practices do need assumptions, statistical assumptions, but still requirements that need to be respected to do not incur in weak identification issues. Particularly Heteroskedasticity-Based Identification needs a strong level of variability of the variance, to avoid weak identification; similarly we need a significant departure from normality in non-Gaussian identification, along with the non-innocuous assumption of mutual independence of the components.

The non-Gaussian approach in econometrics owes its origin to the Independent Componenent Analysis (ICA) literature. The first paper that exploited those methods to identify SVAR's models was most likely [Hyvärinen et al.(2010)Hyvärinen, Zhang, Shimizu, and Hoyer]; they applied the *FastICA* algorithm with a sparsity penalty in estimation to investigate the dynamics between financial indecis.
[Moneta et al.(2013)Moneta, Entner, Hoyer, and Coad] exploit a similar algorithm and they impose a acyclic structure not known by the researcher at priori for which the structural matrix is lower triangular; they instead propose an application in monetary policy and one in firm growth. Other **non-parametric** approaches that have been applied in econometrics are the copula-based CVM distance of Genest et al. (2007) that has been used by [Herwartz(2018)]. Mattson DS, Say differently, propose a methods that consists in minimizing a distance covariance of Székely et al.(2007).

The **parametric approach** to Non-Gaussian estimation of SVAR models has been introduced by
[Lanne et al.(2017)Lanne, Meitz, and Saikkonen]. They propose to use Student-t densities to model the independent shocks which are assumed to be non-Gaussian allowing for to the utmost one Normal component, then to implement a Maximum Likelihood estimator that is consistent and asymptotically normally distributed under their identification scheme and a correct density choice but their three step estimator is efficient only if the shocks follow a symmetric distribution. Finally they made an application with US data regarding monetary policy and financial market and had been able to reject the widely employed recursive identifying restrictions.
[Gouriéroux et al.(2017)Gouriéroux, Monfort, and Renne] differently, do not pick an exact distribution but a pseudo likelihood (PML) and they estimate structural matrix by exploiting its orthogonality; their contribution differs as the the PML estimator is consistent even if the likelihood of the model is misspecified but still it negatively affects the asymptotic accuracy. A discrete mixture of Normals have been instead used to model the shocks by [Fiorentini and Sentana(2023)] who analyze the performance of their pseudo maxmium likelihood estimator under misspecification of the density, their argue that their approach is able to consistently estimates all the parameters.

There is a more recent approach in the literature that relaxes the independence assumption so moving away from the classical ICA framework and requiring instead uncorrelated shocks satisfying some zero restrictions on co-skewness and co-kurtosis. Those weaker assumptions in practice present a trade-off as the "old" stronger restrictions improve identification when valid.
A first branch of the literature focuses on using forth moments to achieve identification by excess co-kurtosis restrictions, among this we find [Guay(2021)]) who approaches the problem similarly, while [Keweloh(2023)] after imposing some restrictions in order to restrict the set of permissible matrices is able to globally identify the structural matrix following [Comon(1994)]. Another branch instead tried to exploit third cumulants and skewness combined or separately with kurtosis, in their respective works the researchers above mentioned showed conditions of under some conditions their identification scheme still holds. Interesting the usage of skewness conditions alone permit the data to have co-heteroskedasticity, which is quite common in macroeconomic time series and one of the main critique to the ICA approach to identification.

3 Macroeconomic shocks

Given the nature of macroeconomic dynamics causal inference in macroeconomics is a considerably challenging matter. Ansewering to questions such as quantitatively assessing the effect of monetary and fiscal policy on the economy is as interesting as complex especially nowdays. Central banks for example do no set randomly the interest rates, but theirs is an endogenous response to the economy conditions [Nakamura and Steinsson(2018)]; thus is impossible in practice to assign the change of a particular economic variable due to a change in another given the countless other variables that might have affected the latter in the same time. And of course central banks and governments have all the interest to have a more endogenous as possible reaction to the economy, this is a fortune for the common wellbeing but for empirical economists. In order to bypass this complication economists aim at identifying exogenous variations in macroeconomic policy and use them to assess the effects of the policy on the economic variables; we define this exogenous variation as a **macroeconomic shock**. [Bernanke(1986)] defines shocks as primitive exogenous forces that are uncorrelated with each other and they should be economically meaningful, specifically they should be:

1. Exogenous with respect to the other current and lagged endogenous variables in the model
2. Uncorrelated with other exogenous shocks
3. Unanticipated as agents information set at time t must not contain shocks that will materialize at time $t + q; q > 0$

These three assumptions if verified allow to effectively track the impact of the shocks on the system, but as a matter of facts making assumptions is easy but things are not so straightforward in practice. These objects that we call shocks are very appeling in theory but things do not come so easy in practice. In facts, one could occur in identifying as economic shocks what could not detected by the specification of the model itself and not the desired primitive shock, leading to a poor identification of the latter. The model in facts threats as exogenous everything that has not been specified, but the fact that the model is not rich enough do no prevent other variables to have a real impact in practice. We will come back on this problematic of the SVAR models afterwards. Note that so far we have settled with uncorrelation and not independence, indeed this is a different story.

Among all, the most common approach to identify economic shocks in the literature is controlling for confounding factors, namely **SVAR** models. This method consists in specifying a model containing some variables and their lags. Then a linear combination of the residuals are interpreted as primitive shocks of the included variables. The main drawback of this approach is the limited number of variables you can control for about the current economic situation but also the its expected state in the future. Monetary policy for example also reacts to the expected state of the economy which is and not tracked by the econometrician information set. This leads to an high risk in identifying shocks that are biased. The omitted variable problem is technically assessed as *non-invertibility* problem and is thought [Sims(1992)] to cause at least partially the famous "price puzzle", which imply a positive reaction of prices to a monetary policy shock in the short run. Once the shocks have been estimated, what remains to do is to assess their impact on the economy. The objects that track the dynamic response of the variables to the shocks are called **Impulse Response Function (IRF)** and could be basically constructed following two ways; the first is by iterating the response through the VAR dynamics and it is the more precise if the model is correctly specified but it carries the bias if is not; the other is called local-projections [Jordà(2005)] and only requires a linear structure as imposition, consisting basically in regressing to identified shocks along with some other control to the variable of interest but directly at the interested horizon, without any iteration. If the first is potentially more accurate the second is more robust.

4 Small scale DSGE model

The effect of monetary policy is usually modeled my economists through small-scale monetary DSGE models. An example of a standard New Keyesian DSGE model is:

$$\begin{aligned}
\tilde{y}_t &= \gamma \mathbb{E}_t[\tilde{y}_{t+1}] + (1-\gamma)\tilde{y}_{t-1} - \phi(R_t - \mathbb{E}_t[\pi_{t+1}]) + \zeta_{\tilde{y},t}, \\
\pi_t &= \beta_f \mathbb{E}_t[\pi_{t+1}] + \beta_b \pi_{t-1} + \kappa \tilde{y}_t + \zeta_{\pi,t}, \\
R_t &= \rho R_{t-1} + (1-\rho)\phi_\pi(\mathbb{E}_t[\pi_{t+1}] - \pi_t^*) + (1-\rho)\phi_y \tilde{y}_t + \zeta_{R,t}
\end{aligned}$$

where the disturbances terms are defined as follows:

$$\zeta_{\tilde{y},t} = \phi_y \zeta_{\tilde{y},t-1} + \sigma_{\tilde{y}} \varepsilon_{\tilde{y},t}, \quad \varepsilon_{\tilde{y},t} \sim \text{WN}(0,1), \tag{1}$$

$$\zeta_{\pi,t} = \phi_\pi \zeta_{\pi,t-1} + \sigma_\pi \varepsilon_{\pi,t}, \quad \varepsilon_{\pi,t} \sim \text{WN}(0,1), \tag{2}$$

$$\zeta_{R,t} = \phi_R \zeta_{R,t-1} + \sigma_R \varepsilon_{R,t}, \quad \varepsilon_{R,t} \sim \text{WN}(0,1). \tag{3}$$

In compact matrix form:

$$\boldsymbol{\Gamma}_0 \mathbf{W}_t = \boldsymbol{\Gamma}_f \mathbb{E}_t \mathbf{W}_{t+1} + \boldsymbol{\Gamma}_b \mathbf{W}_{t-1} + \boldsymbol{\Theta} \boldsymbol{\zeta}_t \tag{4}$$

$$\boldsymbol{\zeta}_t = \boldsymbol{\Phi} \boldsymbol{\zeta}_{t-1} + \underbrace{\boldsymbol{\Sigma}_\varepsilon^{1/2}}_{\boldsymbol{\nu}_t} \boldsymbol{\varepsilon}_t \tag{5}$$

$$\mathbb{E}(\boldsymbol{\varepsilon}_t) = 0, \quad \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{I}_M \tag{6}$$

where $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_f$, $\boldsymbol{\Gamma}_b$, and $\boldsymbol{\Theta} = \mathbf{I}_M$, $\boldsymbol{\Phi} \equiv \text{dg}(\boldsymbol{\Phi})$, and $\boldsymbol{\Sigma}_\varepsilon \equiv \text{dg}(\boldsymbol{\Sigma}_\varepsilon)$ depend on $\boldsymbol{\theta}$.
$\boldsymbol{\theta} := (\gamma, \delta, \varpi, \beta_f, \beta_b, \kappa, \rho, \varphi_\pi, \varphi_y, \phi_{\tilde{y}}, \phi_\pi, \phi_R, \sigma_{\tilde{y}}, \sigma_\pi, \sigma_R)'$ is a vector of structural parameters which contains information about the agent preferences, policy parameters about central bank reaction, elasticity of substitution, degree of price stickiness among the others. These economic theoretical models differ in many points from their purely statistical counterparts.
Of course they have a deeper structure that is derived from economic theory, but they importantly also account for forward-looking behavior of the agents besides the backward looking inertia which is also common in SVAR models. If the DSGE admits a unique stable solution then it admits the following SVAR(2) representation:

$$\mathbf{W_t} = \mathbf{A_1} \mathbf{W_{t-1}} + \mathbf{A_2} \mathbf{W_{t-2}} + \boldsymbol{\Psi} \boldsymbol{\varepsilon_t}$$

This is basically what we define as a B-model, and its autoregressive parameters are such that to represent the the DSGE structural parameters in $\boldsymbol{\theta}$. In particular the parameters in the SVAR model depend nonlinearly and uniquely on the structural parameters $\boldsymbol{\theta}$ through the parameters in the above compact matrix form, beside the cross equation restrictions (CER). Those CER arises in order to maintain a specific structure also in reduced-form, empirical representation, as the model do contain a deep underlying economic structure and consequently parameters in different equations are linked by economic theory. From the SVAR representation

we can obtain the impulse response functions of the original DSGE under the standard VMA representation and so potentially the SVAR is alone able to retrieve the DSGE structural shocks. The key point here is that $\mathbf{\Psi}$ is a full matrix in the DSGE framework, or in other words all the variables react immediately to each shock. We know that in practice dealing with a B-model with a full structural matrix is not trivial. And so economist have put some restrictions of this matrix, namely a recursive identification scheme to bypass this problem. What could be interesting to do is testing the recursive identification scheme statistically to understand indeed if there is enough evidence in the data to reject those restrictions, that although being very comfortable are hardly digested by the theory. Unfortunately the Cholesky SVAR is exactly identified and cannot be tested, so another estimation method must be employed. In the last years a large vast of literature in macroeconometrics has focused on statistical identification taking inspiration from the statistical literature of Independent Component Analysis (ICA). This approach fits our scope has it not requires any economic assumptions and so no restrictions on $\mathbf{B}$, although this comes at the cost of statistical assumptions of non-Gaussianity and mutual independence, on the structural shocks we are striving to recover.

5 Identification problem in SVAR models

Consider a SVAR model:

$$\mathbf{y}_t = \boldsymbol{\nu} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}\boldsymbol{\varepsilon}_t, \tag{7}$$

Where $\boldsymbol{\nu}$ is a $(n \times 1)$ vector of constants, $\boldsymbol{A_i}$ is the i-th reduced form parameter matrix, $\mathbf{y}_t$ is a $(n \times 1)$ vector of time series, $\mathbf{B}$ is the $(n \times n)$ matrix of structural parameters assumed invertible, $\boldsymbol{\varepsilon_t}$ is a $(n \times 1)$ vector of structural parameters such that $\mathbf{u_t} = \mathbf{B}\boldsymbol{\varepsilon_t}$ is the $(n \times 1)$ vector of reduced form innovations. Here the structural shocks are latent and only the residuals of the innovations are observed. Besides we assume that $\mathbf{E}[\boldsymbol{\varepsilon_t}] = \mathbf{0}$ (n x 1) and $\mathbf{E}[\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t'}] = \mathbf{I}$ (n x n), hence:

$$\boldsymbol{u_t}\boldsymbol{u_t'} = \mathbf{B}\boldsymbol{\varepsilon_t}(\mathbf{B}\boldsymbol{\varepsilon_t})' = \mathbf{B}\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t'}\mathbf{B'} \tag{8}$$

By expectations on both sides and by exploiting $\boldsymbol{E}[\boldsymbol{\varepsilon_t}\boldsymbol{\varepsilon_t'}] = \mathbf{I}$ the following is obtained:

$$\boldsymbol{\Sigma_u} = \mathbf{B}\mathbf{B}' \tag{9}$$

Following [Rothenberg(1971)], identification relies on two kinds of conditions:

- **Necessary order condition:** the number of structural parameters cannot exceed the number of the reduced form parameters. So the necessary order condition for identification is:

$$p = dim(\boldsymbol{\beta}) \leq m = \frac{M(M+1)}{2} \tag{10}$$

- **Necessary and sufficient rank condition:** involves the rank of the Jacobian matrix.

$$rank[J_B(\boldsymbol{\beta})] = b \tag{11}$$

Where $\boldsymbol{\Sigma_u}$ is a symmetric matrix and thus is has only $\frac{M(M+1)}{2}$ free parameters. Hence only $\frac{M(M+1)}{2}$ ele-

ments can be recovered from it and the remaining $\frac{M(M-1)}{2}$ in $\mathbf{B}$ are not directly recoverable from $\boldsymbol{\Sigma_u}$. $J_B(\boldsymbol{\beta})$ is the Jacobian matrix computed in a neighborhood of $\boldsymbol{\beta_0}$ and b is number of free parameters. $\boldsymbol{\beta}$ is the (b x 1) vector of free parameters of $\mathbf{B}$. By applying the $vech(\cdot)$ to both sides of (6):

$$\boldsymbol{\sigma_u^+} = g(\boldsymbol{\beta}). \tag{12}$$

$g(\cdot)$ is a nonlinear, continuous and differentiable function that maps the reduced form parameters and the structural parameters. For our specific task, we want $g(\cdot)$ to map the reduced form parameters into the structural parameters by the means of its inverse. It is straightforward that to have the **necessary order condition** respected we need to impose at least $\frac{M(M-1)}{2}$ restrictions on $\mathbf{B}$:

$$l \geq \frac{M(M-1)}{2} \tag{13}$$

Where l is the number of restrictions we need to place on $\mathbf{B}$. The Jacobian matrix can be rewritten as:

$$J_B(\boldsymbol{\beta}) = 2\mathbf{D_M^+}(\mathbf{B} \otimes \mathbf{I_M})\mathbf{S_B} \tag{14}$$

$\mathbf{S_B}$ is an (M^2 x b) selection matrix, $\boldsymbol{D_M^+}$ is the (m x M^2) Moore-Penrose pseudo-inverse of the Duplication matrix, $\mathbf{D_M^+} = (\mathbf{D_M'D_M})^{-1}\mathbf{D_M'}$. And so the **necessary and sufficient rank condition for local identification is**:

$$\text{rank}\{2\mathbf{D_M^+}(\mathbf{B}\otimes\mathbf{I_M})\mathbf{S_B}\} = \text{b} \tag{15}$$

with $\mathbf{B_0} = \mathbf{B}(\beta_0)$.

The rank condition can be tested by estimating the SVAR model by Maximum Likelihood and then check the rank of the Jacobian at $\mathbf{B} = \hat{\mathbf{B}}_\mathbf{T}$. It ensures that the restrictions allow for a unique solution of the system, in facts the full rank of the matrix of restrictions guarantees that the restrictions are linearly independent and so the system can be uniquely solved.

Importantly identification can be assessed to be **local** or **global**. We say that a parameter point α^0 is *globally identified* if there is no other α in the parameter space $\mathcal{A}$ which is observationally equivalent. We refer to *local identification* instead if there exists an open neighborhood of α^0 containing no other α in which $\mathcal{A}$ is observationally equivalent. Global identification is rarely obtained in practice, as researchers must often settle with a local one.

A $(K \times K)$ positive definite matrix $\boldsymbol{\Sigma}$ is said to have a unique Cholesky decomposition if and only if there exists a lower triangular $(K \times K)$ $\mathbf{L}$ such that:

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' \tag{16}$$

where $\mathbf{L}$ is the Cholesky factor. In the B-SVAR context:

$$\mathbf{\Sigma_u} = \mathbf{BB}' \tag{17}$$

These decomposition is unique, and so the structural parameters in $\mathbf{B}$ are uniquely recovered from $\mathbf{\Sigma_u}$. Of course the trade-off is that we need $\mathbf{B}$ to be lower triangular. The drawback of this operation is that it imposes a recursive structure upon the variables. In other words as the $\mathbf{B}$ matrix corresponds to the IRF impact at horizon 0, fixing ex-ante some of its element to zero automatically imposes that same variables do not react instantly to one or more shock in the system. Those restrictions are so called **short run restrictions** and in some circumstances can be reasonably defended relying on economic theory. For example in a small scale SVAR of monetary policy with three variables, the interest rate is considered the fastest variable to move and so it is ordered last, while inflation and economic activity are put first.

$$\mathbf{B}_{\text{chol}} = \begin{pmatrix} b_{\pi\pi} & 0 & 0 \\ b_{y\pi} & b_{yy} & 0 \\ b_{r\pi} & b_{ry} & b_{rr} \end{pmatrix} \tag{18}$$

When we have more variables, or when those variables are all "fast-moving" (typical financial variables tend to present these behaviour) the defense of short-run restrictions is much harder and so the researcher must rely on more sophisticated methods to achieve identification. Unfortunately we have to settle with identification up to sign, in other words we know the magnitude and the order of the shock, but we are not able to determine their sign (i.e. the column sign of the $\mathbf{B}$ matrix), but this is not a problem in practice as one can assumed the diagonal elements to be positive. The following example should convince you on why it happens. We define a $\mathbf{B}$ matrix with a single restriction but if we change the sign of all the columns of $\mathbf{B}$ (or even of one) those models are all observationally equivalent (16).

$$\mathbf{B} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \qquad \mathbf{BB}' = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} \\ 0 & b_{22} \end{pmatrix} = \begin{pmatrix} b_{11}^2 & b_{11}b_{21} \\ b_{11}b_{21} & b_{21}^2 + b_{22}^2 \end{pmatrix} \tag{19}$$

$$\mathbf{B}^* = \begin{pmatrix} -b_{11} & 0 \\ -b_{21} & -b_{22} \end{pmatrix} \qquad \mathbf{B}^*\mathbf{B}^{*\prime} = \begin{pmatrix} -b_{11} & 0 \\ -b_{21} & -b_{22} \end{pmatrix} \begin{pmatrix} -b_{11} & -b_{21} \\ 0 & -b_{22} \end{pmatrix} = \begin{pmatrix} b_{11}^2 & b_{11}b_{21} \\ b_{11}b_{21} & b_{21}^2 + b_{22}^2 \end{pmatrix} \tag{20}$$

$$\mathbf{B}^{**} = \begin{pmatrix} -b_{11} & 0 \\ -b_{21} & b_{22} \end{pmatrix} \qquad \mathbf{B}^{**}\mathbf{B}^{**\prime} = \begin{pmatrix} -b_{11} & 0 \\ -b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} -b_{11} & -b_{21} \\ 0 & b_{22} \end{pmatrix} = \begin{pmatrix} b_{11}^2 & b_{11}b_{21} \\ b_{11}b_{21} & b_{21}^2 + b_{22}^2 \end{pmatrix} \tag{21}$$

$$\mathbf{\Sigma_u} = \mathbf{BB}' = \mathbf{B}^*\mathbf{B}^{*\prime} = \mathbf{B}^{**}\mathbf{B}^{**\prime} \tag{22}$$

The issue of identification arises when we cannot recover the parameters of the model in population only by the observations alone; in other words what we observe cannot allow us to uniquely recover the parameters of interest, but instead can be originated by a set of different models. We state that these models are **observationally equivalent**.

5.1 More on the identification issue

Consider a B-SVAR model, the product $\mathbf{B}\boldsymbol{\varepsilon_t}$ can be rewritten as $\mathbf{BCC^{-1}}\boldsymbol{\varepsilon_t}$ for any nonsingular $(n \times n)$ matrix $\mathbf{C}$ and the model will not be affected.

$$\mathbf{u_t} = \mathbf{BCC^{-1}}\boldsymbol{\varepsilon_t} \tag{23}$$

In this case the true model cannot be detected from any of its linear transformation. In the case $\boldsymbol{\varepsilon_t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, meaning that the structural shocks are normally distributed with zero mean, uncorrelated and standardized to have unit variance then the matrix $\mathbf{C}$ can be restricted to $\mathbf{C} = \mathbf{DO}$. Where $\mathbf{D}$ is a diagonal matrix while $\mathbf{O}$ is orthogonal ($\mathbf{OO'} = \mathbf{I}$), those two matrices are such that they maintain the variance structure of the shocks it was previously assumed: the orthogonal matrix ensures that the every transformation of the shocks remain uncorrelated with guaussian distribution $\mathbf{O}\boldsymbol{\varepsilon_t} \sim N(\mathbf{0}, \mathbf{I})$ as the gaussian distribution is invariant under orthogonal transformation; while $\mathbf{D}$ has $+1$ or -1 on the diagonal, ensuring that the shocks maintain their unit variance.
Thus we can state that $\mathbf{B}$ is identified up to orthogonal transformation and sign while preserving the statistical structure of the shocks.

$$\mathbf{u_t} = \mathbf{BDO}(\mathbf{DO})^{-\mathbf{1}}\boldsymbol{\varepsilon_t} \tag{24}$$

Note that $\text{diag}(\mathbf{D})$ is either 1 or -1, hence $\mathbf{DO}$ is still orthogonal and so it is $(\mathbf{DO})^{-\mathbf{1}}$.
By assuming Gaussianity we were able to be more specific about $\mathbf{C}$, but the identification issue is still severe. What could help is instead assuming **non-Gaussianity** of $\boldsymbol{\varepsilon_t}$, in facts it can be shown that under some conditions, assuming non-Normality of the shocks can restrict the previously defined orthogonal matrix $\mathbf{O}$ to a permutation matrix $\mathbf{P}$, so that the structural matrix $\mathbf{B}$ is said to be identified up to permutation and scaling (sign); in other words we don't know the scale (sign) of the columns nor their original order.

$$\mathbf{u_t} = \mathbf{BDO}(\mathbf{DO})^{-\mathbf{1}}\boldsymbol{\varepsilon_t} \tag{25}$$

Proposition: *consider a random vector* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*, let* $\mathbf{Q}$ *be an* $(n \times n)$ *orthogonal matrix (i.e.,* $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$*), and define* $\mathbf{Y} = \mathbf{Q}\mathbf{X}$.
Then $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y})$.

Proof: define the density of $\mathbf{X}$ as:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{26}$$

Then because $\mathbf{Q}$ is an orthogonal transformation, $\mathbf{Y}$ is also normally distributed: $\mathbf{Y} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top)$ and so its density is defined as:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu})^\top (\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top)^{-1}(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu})\right) \tag{27}$$

Since $\mathbf{Q}$ is orthogonal then $|\mathbf{Q}| = \pm 1$ and consequently $|\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top| = |\boldsymbol{\Sigma}|$. Thus,$|\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top|^{1/2} = |\boldsymbol{\Sigma}|^{1/2}$. And so the fraction outside the exponential has been made equal.

In order to simplify the quadratic form we exploit the properties of orthogonal matrices and so by using $(\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top)^{-1} = \mathbf{Q}\boldsymbol{\Sigma}^{-1}\mathbf{Q}^\top$, we can rewrite the exponent as:

$$(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu})^\top (\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top)^{-1}(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu}) = (\mathbf{y}-\mathbf{Q}\boldsymbol{\mu})^\top \mathbf{Q}\boldsymbol{\Sigma}^{-1}\mathbf{Q}^\top(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu}). \tag{28}$$

Then we use $\mathbf{x} = \mathbf{Q}^\top \mathbf{y}$ and $\mathbf{y} = \mathbf{Q}\mathbf{x}$ to finally state:

$$(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu})^\top \mathbf{Q}\boldsymbol{\Sigma}^{-1}\mathbf{Q}^\top(\mathbf{y}-\mathbf{Q}\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}). \tag{29}$$

Hence we proved that:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{30}$$

Proposition: *consider a random vector of jointly gaussian distributions* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\boldsymbol{\Sigma}$*. Then if* $\boldsymbol{\Sigma}$ *is diagonal this implies that* $\forall i \neq j,\ X_i \perp X_j$
and so $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$.

Proof: The joint density for multivariate gaussian is defined as:

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \tag{31}$$

Exploiting the fact that $\boldsymbol{\Sigma}$ is diagonal, its determinant is just the product of the entries: $|\boldsymbol{\Sigma}| = \sigma_1^2 \cdot \sigma_2^2 \cdot \ldots \cdot \sigma_n^2$ where each σ_i^2 is the variance of X_i for the i-th diagoanal entry of $\boldsymbol{\Sigma}$.

The term at the exponent can be rewritten as:

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{\sigma_i^2}. \tag{32}$$

and so we put it all together to get:

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\prod_{i=1}^{n}\sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right). \tag{33}$$

Further simplification due the exponent's properties allow to express the joint as a product of the marginals:

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right). \tag{34}$$

And so we proved that:

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \tag{35}$$

6 Independent component analysis (ICA)

Following [Hyvärinen et al.(2004)Hyvärinen, Karhunen, and Oja], Independent Component analysis is a relatively new technique which was firstly designed to decompose a multivariate signal into independent non-Gaussian signals. The original scope of ICA was to solve problems that were closely related to the so called cocktail-party problem in the field of signal processing, but then it has been spread out in many other field of research such as Neuroscience, Biomedical Engineering, Machine Learning, Psychology and even Econometrics. In order to understand the logic of ICA, is useful to start from the original cocktail-party problem.

Imagine to be in a room where three people are talking and three microphones give the recording signals. Our goal is to clearly recover the talking from the noisy signals. We label as $x_1(t), x_2(t), x_3(t)$ the noisy signals recorded by the microphones at each time index t, while the original latent signals that we want to recover are labeled as $s_1(t), s_2(t), s_3(t)$. The problem can be mathematically formalized represented as follows where a_{ij} are some parameters that depend on the distances of the microphones from the speakers (more generally they can be though to weight how much a certain speaker enters one specific microphone)

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \qquad\qquad , \mathbf{x} = \mathbf{A}\mathbf{s} \\ x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \end{aligned}$$

In the matrix form of the system $\mathbf{A}$ is the matrix of the a_{ii} elements, $\mathbf{x}$ is the column vector of the mixed signals, while $\mathbf{s}$ is the column vector of the original sources. Retrieve $\mathbf{s}$ would have been straightforward if we $\mathbf{A}$ would be known and invertible, but this is not the case as we don't known the mixing matrix. In order to successfully recover the original sources we need three assumptions to hold:

- **The independent components are assumed statistically independent**
  The random variables $y_1, y_2, ..., y_n$ are said to be independent if and only if the joint pdf $p(y_1, y_2, ..., y_n)$ is factorizable as a product of the marginals:

  $$p(y_1, y_2, ..., y_n) = p_1(y_1)p_2(y_2)...p_n(y_n) \tag{36}$$

  In other words the value of y_i does not give any information on the value of y_j for $i \neq j$. It is important to remark that independence is a stricter requirement than uncorrelation; the latter rule out any kind of linear relationship while the first every kind of dependence, linear or non-linear. Among all the random variables, Gaussians are the only ones for which uncorrelation implies independence.

- **The independent components must have non-Gaussian distribution**
  As Normal random variables are characterized only by their first and second moment, those information are already exploited by methods like PCA. Thus if we want to rule out deeper relationships we need distributions that are characterized by higher moments, and so as a matter of fact ICA is useless for Gaussian variables.

- **We assume that the unknown mixing matrix is square**
  We want to recover $\mathbf{s}$ by inverting $\mathbf{A}$ and a square matrix implies that the number of independent

components is equal to the number of observed mixtures.

$$\mathbf{s} = \mathbf{A^{-1}x} \tag{37}$$

6.1 Identification up to permutation and scaling

The ICA procedure although being very powerful is not able to fully recover the underlying components, indeed the scale and order ambiguity remains. To see that every scaling of $\mathbf{A}$ is observationally equivalent look at:

$$\mathbf{x} = \sum_i (\frac{1}{\alpha_i})(s_i \alpha_i) \tag{38}$$

If we scale a source s_i by a constant α_i it can be canceled by multiplying the correspondent column $\mathbf{a_i}$ of $\mathbf{A}$ [1]. This problem can be easily solved in practice by assuming that the sources have unit variance, so we only need to rescale them and so the mixing matrix accordingly. Of course the issue of the sign remains and cannot be statistically solved, but needs theoretical arguments.
The another obstacle to complete identification is related to the order of the components $\mathbf{s}$ and so of the column of $\mathbf{A}$. Formally consider a permutation matrix $\mathbf{P}$, so that:

$$\mathbf{x} = \mathbf{AP^{-1}Ps} \tag{39}$$

6.2 PCA and ICA

Broadly speaking the ICA algorithm can be decomposed into three steps. What differences PCA and ICA is that the first produce components that are scaled and uncorrelated, while the second besides is able to produce compoenents that are independent. We rewrite the ICA model as by the SVD of $\mathbf{A}$

$$\mathbf{x} = \underbrace{(\mathbf{U\Sigma V})}_{\mathbf{A}}\mathbf{s} \qquad\qquad \mathbf{s} = \underbrace{(\mathbf{U^*\Sigma^{-1}V^*})}_{\mathbf{A^{-1}}}\mathbf{x} \tag{40}$$

1. **"Rotate"** multiply the data by the matrix $\mathbf{U}^*$ that contains information about the primary direction of the data and it rotates the data accordingly making the components uncorrelated.
2. **"Stretch"** scales the data by the matrix $\mathbf{\Sigma^{-1}}$ to get a diagonal covariance matrix.
3. **"Rotate"** rotates again the data by the means of $\mathbf{V}^*$ by making the components as independent as possible exploiting high-order statistics of the data.

If data are not-Gaussian the third step can be used to distinguish components that are statistically independent, differently any further rotation of the space will be useless as the standard Normal distribution is invariant under orthogonal transformation. Put differently the last rotation in ICA aims at making the components independent; but given that in the Gaussian case uncorrelation implies independence, there is no meaning in any further rotation as the data will be equally independent for any rotation of the space.

In the following graphical representation it can be seen visually. Imagine that the statistically independent

[1] It can also be shown by considering a diagonal matrix $\mathbf{D}$ so that $\mathbf{x = AD^{-1}DS}$

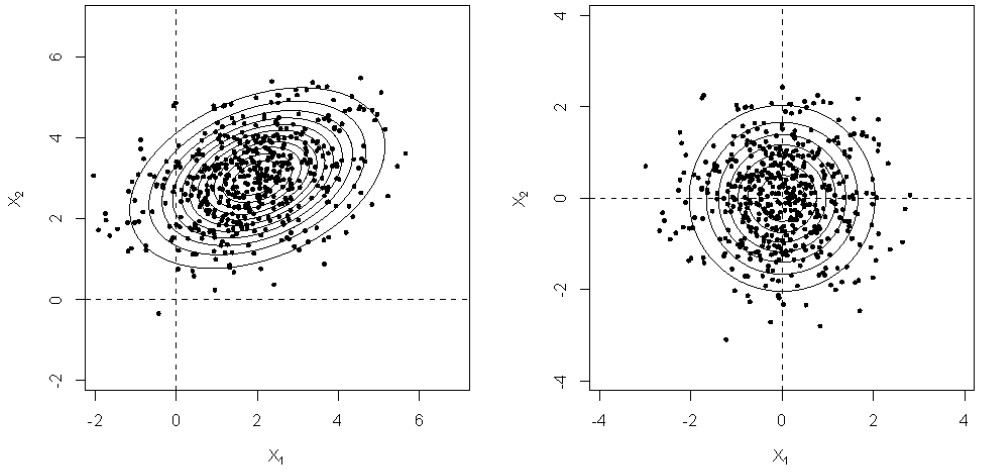


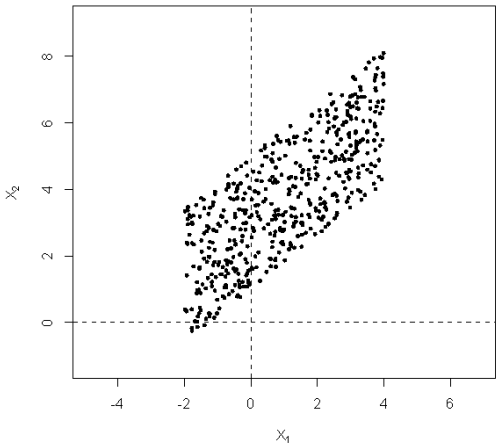
Figure 1: On the left: 500 draws from a $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (2, 3)$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{2} & \mathbf{0.5} \\ \mathbf{0.5} & \mathbf{1} \end{pmatrix}$,
On the right: same draws after being center and whitened through $\mathbf{Z} = \mathbf{E}\mathbf{D}^{-\mathbf{1/2}}\mathbf{E}'$ obtained via the spectral decomposition of the covariance matrix.

original components comes from a bivariate standardized uniform distribution (figure 2c). Then the noisy signals that you observe are obtained from multiple transformation of the original components. We can imagine that the original components have been rotate the first time by the means of $\mathbf{V}$, then stretched through $\boldsymbol{\Sigma}$ and finally rotate again to obtain linear dependence by $\mathbf{U}$ (figure 2a). What you observe is (figure 2a). In order to come back to the original components one can think to invert the process; by firstly rotate the space to obtain uncorrelation through $\mathbf{U}^*$, then rescale the space to get whitened data by $\boldsymbol{\Sigma}^{-\mathbf{1}}$ (figure 2b) and lastly to rotate again the data through $\mathbf{V}^*$ to reach independence and retrieve the original components (image 2c). The first two step can be reached by PCA, the last only by ICA as PCA cannot go further wiping out linear dependence. Figure 1a shows data from a general joint Gaussian with linear dependence, but ones you have whitened the data (figure 1b) any further rotation is useless. The whitening matrix can be defined as:

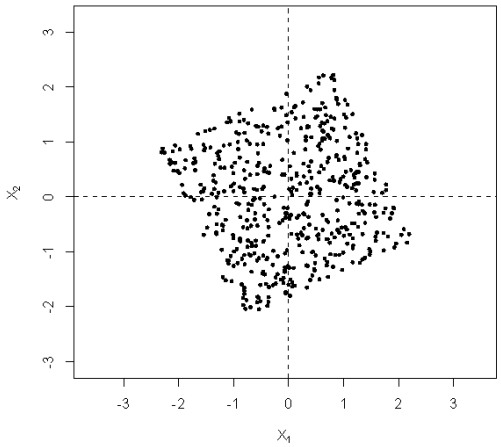
$$\mathbf{Z} = \mathbf{E}\mathbf{D}^{-\mathbf{1/2}}\mathbf{E}' \tag{41}$$

$\mathbf{Z}$ is the linear transformation that whites the data, $\mathbf{E}$ is the matrix which columns are eigenvectors of the covariance matrix and $\mathbf{D}$ is the diagonal matrix of the relative eigenvalues obtained through the spectral decomposition of the covariance matrix. Intuitively $\mathbf{E}$ rotates and $\mathbf{D}$ stretches. The eigendecomposition of the correlation matrix decompose the latter as:

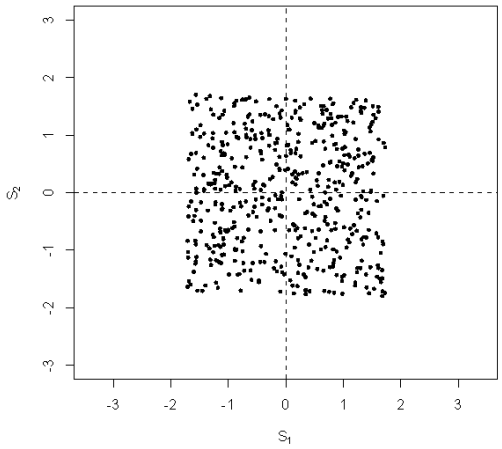
$$\Sigma = \mathbf{E}\mathbf{D}\mathbf{E}' \tag{42}$$



(a) 500 draws from a bivariate uniform distribution with linear dependence.



(b) Same draws after being centered and whitened through $\mathbf{Z} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}'$ obtained via the spectral decomposition of the covariance matrix: linear dependence has been wiped out but there is still non-linear dependence.



(c) A last rotation is performed on the whitened data by the mixing matrix obtained from a FAST ICA algorithm: the original sources are recovered and independent.

Figure 2: An application of ICA with uniform distributions.

If the data are whitened the covariance matrix is the identity matrix:

$$\mathbf{E}[\varepsilon\varepsilon'] = \mathbf{I} \tag{43}$$

Now consider that the data is whitened and the new matrix is denoted as $\tilde{\mathbf{A}}$, so:

$$\mathbf{z} = \mathbf{ZAs} = \tilde{\mathbf{A}}\mathbf{s} \tag{44}$$

Take an orthogonal transformation $\mathbf{U}$ of $\mathbf{z}$, such that $\mathbf{y} = \mathbf{Uz}$, this implies that:

$$\mathbf{E}[\mathbf{yy}'] = \mathbf{E}[\mathbf{Uyy}'\mathbf{U}] = \mathbf{UIU}' = \mathbf{I} \tag{45}$$

The takeaway is that whitening gives components only up to to an orthogonal transformation, in other words if we exploit only the first two moments, as whitening does, every orthogonal transformation of components will be *obervationally equivalent* and then indistinguible. Hence if we want to say something more about the latent variables we need a further linear transformation of the space, which this time exploit information present in higher moments. Consequently we need distributions of the shocks that are not-Gaussian.

6.3 Determine the last rotation

As we already saw the whitening process could be implemented in numerous ways; by the Choleksy decomposition, or the spectral decomposition obtained by the PCA but is inconclusive in solving the problem at a acceptable level of precision. What is more interesting is the last rotation of the space in order to retrieve the independent components performed by ICA.
There are many ways to achieve this results, both with non-parametric and parametric approaches. If you want to go parametrically Maximum Likelihood estimation (ML) is the way; this method in general tends to be more accurate under the correct specification of the likelihood but of course is at risk of miss-specification. Non-parametric approaches differently do not need a precise non-Gaussian density but a measure of non-Gaussianity you want to maximize [2]; popular choices in the literature are the *kurtosis* which is a simple measure but very sensible to outliers and the negentropy which has excellent statistical properties but computationally very demanding.
In this section I briefly summarize some popular methods in the literature which scope is finding the linear combinations which is least dependent.

6.3.1 Least dependent innovations build on Cramér-von Mises statistics (CVM)

Following [Lange et al.(2021)Lange, Dalheimer, Herwartz, and Maxand] if we consider the Cholesky decomposition C of the covariance matrix such that $\mathbf{\Sigma_u} = \mathbf{CC}'$ and $\varepsilon_t = \mathbf{C^{-1}u_t}$, then other candidates can be generated as:

$$\tilde{\varepsilon}_{\boldsymbol{t}} = \mathbf{Q}\varepsilon_t = \mathbf{QC^{-1}u_t} \tag{46}$$

$\mathbf{Q}$ is a rotation matrix and $\mathbf{Q} \neq \mathbf{I_K}, \mathbf{QQ}' = \mathbf{I_K}$. In order to recover the original shocks the residuals are rotated to obtain the least dependent rotation, so that the estimated shocks minimize the CVM rotation. This

[2] Or a measure of Gaussianity you want to minimize

rotation is obtained by the means of $\mathbf{B}_\theta = \mathbf{C}\mathbf{Q}(\theta))$
The rotation matrix could be parameterized as the product of $K(K-1)/2$ distinct forms of orthogonal givens rotation matrices. If $K = 3$, $\mathbf{Q}(\theta)$ can be defined as:

$$\mathbf{Q}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_1) & -\sin(\theta_1) \\ 0 & \sin(\theta_1) & \cos(\theta_1) \end{pmatrix} \times \begin{pmatrix} \cos(\theta_2) & 0 & -\sin(\theta_2) \\ 0 & 1 & 0 \\ \sin(\theta_2) & 0 & \cos(\theta_2) \end{pmatrix} \times \begin{pmatrix} \cos(\theta_3) & -\sin(\theta_3) & 0 \\ \sin(\theta_3) & \cos(\theta_3) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

the rotation angles must satisfy $0 \leq \theta_i \leq \pi$, $i = 1, 2, 3$. By definition, the random vector $\tilde{\varepsilon}_t$ in Equation 12 is a rotation of $\boldsymbol{\varepsilon_t}$. The set of possible structural matrices $B(\theta) = DQ(\theta)$ is defined in terms of the Choleski factor D and the vector of rotation angles (θ) of the Givens matrices $\mathbf{Q}(\theta)$. Hence the estimator is called to find such angles that minimize a certain measure of joint dependence that is captured by the Cramér-von Mises defined as follows:

$$\mathcal{B}_\theta = \int_{(0,1)^K} \left[\sqrt{T} \left(\mathbf{C}(\tilde{\boldsymbol{\varepsilon}}) - \prod_{i=1}^{K} U(\tilde{\varepsilon}_i) \right) \right]^2 d\tilde{\varepsilon}, \tag{47}$$

This approach allow to recover the original structural matrix without making any assumptions about the distribution of $\boldsymbol{\varepsilon_t}$. The minimization of this measure consists in minimizing the distance between the empirical copula of shocks vectors $\mathbf{C}$ and the copula implied under independence constructed as the product of uniformly distributed variables. In other words the estimator aims at finding the rotation of the residuals such that the joint distribution of the implied shocks is as close as possible to an independent one, formally we define the estimator of $\mathbf{B}_\theta$ as $\tilde{\mathbf{B}}_{\hat{\theta}}$, where:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta} \left\{ \mathcal{B}_\theta |\ \tilde{\boldsymbol{\varepsilon}}_\mathbf{t} = \tilde{\mathbf{B}}_\theta^{-1} \mathbf{u}_t \right\}.$$

6.3.2 Fast ICA

The Fast ICA is a fast convergence fixed point alogorithm which estimate the independent components by finding the direction (a unit vector $\mathbf{w}$) such that the projection $\mathbf{w}'\mathbf{z}$ maximizes non-gaussianity measured by $J(\mathbf{w}'\mathbf{z})$. $\mathbf{z}$ is here the matrices of the whited signals, and $J(\mathbf{w}'\mathbf{z})$ is an approximation of the negentropy defined as:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gaussian}}) - H(\mathbf{y}) \tag{48}$$

where $H(\mathbf{y})$ is the differential entropy of a random vector $\mathbf{y}$ with density $p_y(\eta)$ defined as $H(\mathbf{y}) = -\int p_y(\boldsymbol{\eta}) \log p_y(\boldsymbol{\eta})\, \boldsymbol{d\eta}$, $\mathbf{y}_{\text{gaussian}}$ is a random variable of the same correlation and covariance matrix as $\mathbf{y}$. Negentropy is always non-negative and is zero if $\mathbf{y}$ is normal and has very good statistical properties in order to represent non-gaussianity but beacuse of its computational complexity is very often approximated. In practical applications that approximations is computed as:

$$J(y) \approx \left[\mathbb{E}\{G(y)\} - \mathbb{E}\{G(v)\} \right]^2 \tag{49}$$

Here $v \sim N(0, 1)$, and $G(\cdot)$ often takes the form of $-\exp(\frac{u^2}{2})$ or $\frac{1}{\alpha}\text{logcosh}(\alpha \text{u})$

6.3.3 Maximum Likelihood Estimation

In order to derive the likelihood $p_{\mathbf{x}}(\mathbf{x})$ of the mixture $\mathbf{x} = \mathbf{A}\mathbf{s}$ we represent the density of a linear transformation as:

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{|\det \mathbf{A}|} = p_{\mathbf{s}}(\mathbf{s}), \tag{50}$$

Then, as the sources $\mathbf{s}$ are considered independent, the joint density can be factorized as the product of the marginals and by putting all together we have:

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{|\det \mathbf{A}|} \prod_i p_{s_i}(s_i), \tag{51}$$

We define $\mathbf{B} = \mathbf{A}^{-1}$ and p_i as the densities of the independent components, so given that $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$ we obtain:

$$p_{\mathbf{x}}(\mathbf{x}) = |\det \mathbf{B}| \prod_i p_i(\mathbf{b}_i'\mathbf{x}), \tag{52}$$

If we assume to have T observations in $\mathbf{x}$, then the total likelihood $L(\mathbf{B})$ will be the product of the likelihoods evaluated at each point in time:

$$L(\mathbf{B}) = \prod_{t=1}^{T} \prod_{i=1}^{n} p_i(\mathbf{b}'\mathbf{x}(t)) |\det \mathbf{B}|. \tag{53}$$

For practical purpose and thanks to the logarithmic properties we work with the log-likelihood, in this way the maximizer don't change but summation are more easily treatable objects than products:

$$\log L(\mathbf{B}) = \sum_{t=1}^{T} \sum_{i=1}^{n} \log p_i(\mathbf{b}'\mathbf{x}(t)) + T \log |\det \mathbf{B}|. \tag{54}$$

The likelihood L is a function of $\mathbf{B}$, and so the optimization process will focus on finding the $\mathbf{B}$ that maximizes the likelihood. Unfortunately there is no close form solution to this problem and so numerical methods are employed.

7 SVAR tools

Hereby I briefly discuss two important tools used in SVAR analysis; the first are IRFs which are used to determine the sample point estimates of the model, while the second is the bootstrap which is instead used to determine the uncertainty around those estimates.

7.1 IRF

A covariance stationary VAR model admits a VMA representation, and so the SVAR model admits a SVMA.

$$\boldsymbol{W_t} = \boldsymbol{v} + \sum_{h=0}^{\infty} \boldsymbol{C_h B \varepsilon_{t-h}} \tag{55}$$

$$\boldsymbol{C_h} = \boldsymbol{R}(\boldsymbol{\mathcal{C}})^{\boldsymbol{h}}\boldsymbol{R'}, \qquad h = 0, 1, 2, 3, ... \tag{56}$$

$$\boldsymbol{\mathcal{C}}_{Mp\ x\ Mp} = \begin{pmatrix} \boldsymbol{\Pi}_1 & \boldsymbol{\Pi}_2 & \cdots & \boldsymbol{\Pi}_{p-1} & \boldsymbol{\Pi}_p \\ \boldsymbol{I}_M & \boldsymbol{0}_{MxM} & \cdots & \boldsymbol{0}_{MxM} & \boldsymbol{0}_{MxM} \\ \boldsymbol{0}_{MxM} & \boldsymbol{I}_M & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \boldsymbol{0}_{MxM} & \boldsymbol{0}_{MxM} \\ \boldsymbol{0}_{MxM} & \boldsymbol{0}_{MxM} & \cdots & \boldsymbol{I}_M & -\boldsymbol{0}_{MxM} \end{pmatrix}, \qquad \boldsymbol{R}_{M\ x\ Mp} = (\boldsymbol{I}_M, \quad \boldsymbol{0}_{M\,x\,(Mp-m)}) \tag{57}$$

Above $\boldsymbol{\mathcal{C}}$ is the companion matrix and $\boldsymbol{R}$ is the selection matrix. This form is very convenient as it allows the researcher to quantify the dynamics effects of the structural shocks on the estimated system. This is made by the means of the so called **Impulse response functions (IRFs)**

$$\underbrace{\boldsymbol{IRF}_{\bullet,\bullet}(\boldsymbol{h})}_{MxM} = \frac{\partial \boldsymbol{W_t}}{\partial \boldsymbol{\varepsilon'_{t-h}}} \equiv \frac{\partial \boldsymbol{W_{t+h}}}{\partial \boldsymbol{\varepsilon'_t}} = \boldsymbol{C_h B} = (\boldsymbol{R}(\boldsymbol{\mathcal{C}})^{\boldsymbol{h}}\boldsymbol{R'})\boldsymbol{B} = \begin{pmatrix} c_{11}^{(h)} & \cdots & \cdots & c_{1M}^{(h)} \\ \vdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ c_{M1}^{(h)} & \cdots & \cdots & c_{MM}^{(h)} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & \cdots & b_{1M} \\ \vdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ b_{M1} & \cdots & \cdots & b_{MM} \end{pmatrix}$$

$\boldsymbol{IRF}_{\bullet,\bullet}(\boldsymbol{h})$ contains all the structural IRFs to one standard standard deviation to each shock in the system up to horizon h.

$$\underbrace{\boldsymbol{IRF}_{\bullet,j}(\boldsymbol{h})}_{Mx1} = \boldsymbol{C_h B e_j} = \boldsymbol{C_h b_j} = (\boldsymbol{R}(\boldsymbol{\mathcal{C}})^{\boldsymbol{h}}\boldsymbol{R'})\boldsymbol{b_j} = \begin{pmatrix} c_{11}^{(h)} & \cdots & \cdots & c_{1M}^{(h)} \\ \vdots & \ddots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ c_{M1}^{(h)} & \cdots & \cdots & c_{MM}^{(h)} \end{pmatrix} \begin{pmatrix} b_{11} \\ \vdots \\ \vdots \\ b_{M1} \end{pmatrix} \tag{58}$$

$\boldsymbol{e_i}$ is a $(M\ x\ 1)$ selection vector containing 1 in the position i and 0 elsewhere, so $\boldsymbol{b_j}$ is the j column of the $\boldsymbol{B}$ matrix. Hence $\boldsymbol{IRF_{\bullet,j}(h)}$ captures the impact of one-standard deviation shock to variable j to all M variables up to horizon h.

$$IRF_{i,j}(h) = \boldsymbol{e_i' C_h B e_j} = \begin{pmatrix} c_{i1}^{(h)} & \cdots & \cdots & c_{iM}^{(h)} \end{pmatrix} \begin{pmatrix} b_{11} \\ \vdots \\ \vdots \\ b_{M1} \end{pmatrix} \tag{59}$$

The latter is the structural IRF of one-standard deviation shock to variable j on variable i up to horizon h as $\boldsymbol{e_i' C_h}$ is the row i of the matrix $\boldsymbol{C_h}$.

8 Bootstrap

In the frequentist framework the uncertainty about the IRF's is mostly assessed through the method of the Bootstrap. This resampling method relies on the concept of bootstrap consistency; this implies that as the original sample size becomes large the distribution of the bootstrap estimator converges to the distribution of the estimator based on the original data. Consider that for a vector of parameters $\boldsymbol{\theta_0}$, and a vector of estimators $\hat{\boldsymbol{\theta}}_{\boldsymbol{T}}$ over a sample of length T we have the following:

$$\Gamma_T := T^{1/2} \boldsymbol{V_\theta}^{\boldsymbol{-1/2}} (\hat{\boldsymbol{\theta}}_{\boldsymbol{T}} - \boldsymbol{\theta_0}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I})$$

Where $\boldsymbol{V_\theta}^{\boldsymbol{-1/2}}$ is the original covariance matrix with element wise standard deviations. Then if we have a consistent estimator $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}} \xrightarrow{p} \hat{\boldsymbol{V}}_{\boldsymbol{\theta}}$, we can express what follows:

$$\hat{\Gamma}_T := T^{1/2} \hat{\boldsymbol{V}}_{\boldsymbol{\theta}}^{\boldsymbol{-1/2}} (\hat{\boldsymbol{\theta}}_{\boldsymbol{T}} - \boldsymbol{\theta_0}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}) \equiv \Phi_z(x)$$

Under Bootstrap consistency it can be stated that the bootstrap estimator is:

$$\hat{\Gamma}_T^* := T^{1/2} \hat{\boldsymbol{V}}_{\boldsymbol{\theta}}^{\boldsymbol{-1/2}} (\hat{\boldsymbol{\theta}}_{\boldsymbol{T}}^* - \hat{\boldsymbol{\theta}}_{\boldsymbol{T}}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I})$$

In order words the bootstrap reproduces in the bootstrap world the same asymptotic distribution as the estimator in the original probability data distribution. The bootstrap algorithm can be repeated to get N replications of $\hat{\boldsymbol{\theta}}_{\boldsymbol{T}}^*$ and so N values for $\hat{\Gamma}_T^*$ to build the empirical bootstrap cdf defined as:

$$G_{T:N}^*(x) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}(\hat{\Gamma}_{T:n}^* \leq x) \tag{60}$$

Now we can formally state that bootstrap consistency holds if $\forall \varepsilon > 0$:

$$\lim_{T\to\infty} \Pr\left(\sup_{x\in\mathbb{R}} |G_T^*(x) - \Phi_z(x)| > \epsilon\right) = 0$$

or,

$$\sup_{x\in\mathbb{R}} |G_T^*(x) - \Phi_z(x)| \xrightarrow{p} 0 \quad \text{as } T \to \infty$$

The Glivenko-Cantelli theorem guarantees that the empirical cumulative distribution function (CDF) of the bootstrap replications will converge uniformly to the true bootstrap distribution as the number of bootstrap samples, N, grows to infinity:

$$\sup_{x\in\mathbb{R}} |G_{T:N}^*(x) - G_T^*(x)| \xrightarrow{\text{a.s.}} 0$$

8.1 I.I.D. residual bootstrap

A very simple bootstrap used to compute confidence intervals for the IRF's is the i.i.d. residual bootstrap; here we will focus on its non-parametric version but the differences only lie in assuming a certain distribution for the residuals or not. The algorithm goes as it follows:

1. The VAR residuals are firstly centered and then resampled via i.i.d. resampling to obtain $u_1^*, ..., u_T^*$.
2. The bootstrap sample is generated by iterating the VAR model estimated on the original sample for $t = 1, ..., T$ with initial conditions fixed at the original sample value and then adding the randomly extracted sample residuals:

$$\mathbf{W}_t = \hat{\boldsymbol{\nu}} + \hat{\mathbf{A}}_1 \mathbf{W}_{t-1} + \hat{\mathbf{A}}_2 \mathbf{W}_{t-2} + \cdots + \hat{A}_p \mathbf{W}_{t-p} + \mathbf{u}_\mathbf{t}^*. \tag{61}$$

3. Generate the SVAR model on the bootstrap sample and then compute the IRFs $\hat{\phi}_{i,j,h}^{*\cdot}$.
4. Repet the following steps N times, in order to store a sequence of bootstrap estimates $\hat{\phi}_{i,j,h}^{*1}, \hat{\phi}_{i,j,h}^{*2}, ..., \hat{\phi}_{i,j,h}^{*N}$.
5. Compute the percentile bootstrap confidence interval $(q_{\alpha/2}^{\phi*}, q_{1-\alpha/2}^{\phi*})$ from the $\alpha/2$ and $1-\alpha/2$ quantiles of the empirical distribution of $\hat{\phi}_{i,j,h}^{*1}, \hat{\phi}_{i,j,h}^{*2}, ..., \hat{\phi}_{i,j,h}^{*N}$.

9 The model

Consider the following SVAR model:

$$\mathbf{y}_t = \boldsymbol{\nu} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}\boldsymbol{\varepsilon}_t, \tag{62}$$

$\mathbf{y}_t$ is an n-dimensional time series vector of interest, $\boldsymbol{\nu}$ is an intercept vector of size $(n \times 1)$, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are parameter matrices of the VAR model each of size $(n \times n)$, $\mathbf{B}$ is the nonsingular $(n \times n)$ matrix of structural parameters, and $\boldsymbol{\varepsilon}_t$ is the $(n \times 1)$ vector of shocks. Stationarity of $\mathbf{y}_t$, is implied by the polynomial:

$$\det(\mathbf{A}(z)) \overset{\text{def}}{=} \det(\mathbf{I}_n - \mathbf{A}_1 z - \cdots - \mathbf{A}_p z^p) \neq 0, \quad \text{for } |z| \leq 1, (z \in \mathbb{C}) \tag{63}$$

guaranteeing the stability of the time series, hence the moving average representation is:

$$\mathbf{y}_t = \mu + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{B}\varepsilon_{t-j}, \quad \boldsymbol{\Psi}_0 = \mathbf{I}_n, \tag{64}$$

where $\mu = \mathbf{A}(1)^{-1}\nu$ and the matrices $\boldsymbol{\Psi}_j$ (for $j = 0, 1, \ldots$) are determined by the power series expansion $\boldsymbol{\Psi}(z) = \mathbf{A}(z)^{-1} = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j z^j$.. We assume that the n structural shocks $\boldsymbol{\varepsilon_t} = (\boldsymbol{\varepsilon_{1,t}, \ldots, \varepsilon_{n,t}})$ are $i.i.d.$ random vector and at least $n - 1$ components are Student-t distributed with zero mean and unit variance: $\varepsilon_{it} \sim t(0, 1)$[3], besides the elements in $\boldsymbol{\varepsilon}$ are **mutually** independent. Those assumption rule out every kind of dependence both between than across each $\boldsymbol{\varepsilon_t}$, thus consisting in a stronger assumption that mere uncorrelatedness. In facts every kind of common driver of volatility between shock is not allowed. It is true that independence is implicitly imposed in guassian SVAR models as well, but it only comes from a modelling choice and the normal distribution represent the exception and not the rule.
Because of identification up to permutation and sign, the following two moving average representations in hold true:

$$\mathbf{y_t} = \mu + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{B}\varepsilon_{t-j} = \mu^* + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j^* \mathbf{B}^* \varepsilon_{t-j}^*, \tag{65}$$

Then, for some diagonal matrix $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_n)$ with nonzero diagonal elements, and for some permutation matrix $\mathbf{P}$ $(n \times n)$, each of these equivalence classes defines a set of observationally equivalent SVAR processes. An identification problem is still present, but thanks to non-gaussianity it is reduced to permutation and sign of the columns of $\mathbf{B}$ and so the following can be stated.

$$\mathbf{B}^* = \mathbf{BDP}, \quad \varepsilon_t^* = \mathbf{P}'\mathbf{D}^{-1}\boldsymbol{\varepsilon}_t, \quad \mu^* = \mu, \quad \text{and} \quad \boldsymbol{\Psi}_j^* = \boldsymbol{\Psi}_j \quad (j = 0, 1, \ldots).$$

[3] this assumption is made only for the ML estimation the model, non-parametric approaches only needs non-Gaussianity

Here we will assume that the structural shocks have unit variance, i.e. the covariance matrix is diagonal $\mathbf{\Sigma} = \mathbf{I}$. This allow to relax the scaling problem to sign indeterminacy as $\mathbf{D}$ contains only -1 or 1.

9.1 Maximum likelihood

We assume that for each $i = 1, ..., n$, $\varepsilon_{i,t}$ has a Lebesgue density (is a continuous random variable) $f_{i,\sigma}(x; \lambda_i) = \sigma^{-1} f_i(\sigma^{-1}x; \lambda_i)$ where the scale σ is assumed to be one.
λ is a parameter that further describes the distribution, we assume student the density to be univariate Student-t, hence in our context it describes the heaviness of the tails by the means of the **degrees of freedom**. The model is quite general, the only assumption that we make on each $\varepsilon_{i,t}$ is to belong to the same family of univariate Student-t, then each shock distribution is allowed to be characterized by its own λ_i. Each density is not dependent on t, hence we do no let the distribution of each shock to vary depending on time. The standardized log-likelihood function can be written as:

$$L_T(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}), \tag{66}$$

$$\ell_t(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f_i \left(\sigma_i^{-1} \iota_i' \mathbf{B}(\boldsymbol{\beta})^{-1} u_t(\boldsymbol{\pi}); \lambda_i \right) - \log |\det(\mathbf{B}(\boldsymbol{\beta}))| - \sum_{i=1}^{n} \log \sigma_i$$

All the parameters of the model are contained in the vector $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$. $\boldsymbol{\pi} = (\pi_1, \boldsymbol{\pi_2})$ contains the autoregressive parameters $\pi_1 = \nu$ and $\boldsymbol{\pi_2} = vec([\boldsymbol{A_1} : \cdots : \boldsymbol{A_p}])$. $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_n)$ is 1 x n vector containing the scale of the n errors, and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$ contains the degrees of freedom. Finally $\boldsymbol{\beta} = vec(\mathbf{B})$ is a 1 x n vector containing the structural parameters. Given the computational issue that could arise with the ML estimator above defined when T is small and n is quite small, a three-step estimation is defined as follows: $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\gamma})$ where $\boldsymbol{\pi}$ contains the autoregressive parameters as before while the others are stacked into $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$

1. the autoregressive parameters are estimated through LS estimator and $\tilde{\boldsymbol{\pi}}_{LS,T}$ is obtained.
2. $\boldsymbol{\pi}$ is replaced with $\tilde{\boldsymbol{\pi}}_{LS,T}$ and so $u_t(\boldsymbol{\pi})$ with $u_t(\tilde{\boldsymbol{\pi}}_{\boldsymbol{LS,T}})$, hence the following is is maximized with respect to $\boldsymbol{\gamma}$ to obtain estimates of $\mathbf{B}$ and of each parameter of the shock's distributions.

$$\tilde{L}_T(\boldsymbol{\gamma}) = T^{-1} \sum_{t=1}^{T} \ell_t(\tilde{\boldsymbol{\pi}}_{\boldsymbol{LS,T}}, \boldsymbol{\gamma}) \tag{67}$$

3. finally what follows is maximized with respct to $\boldsymbol{\pi}$ instead to get a final maximum likelihood estimates of the autoregressive parameters.

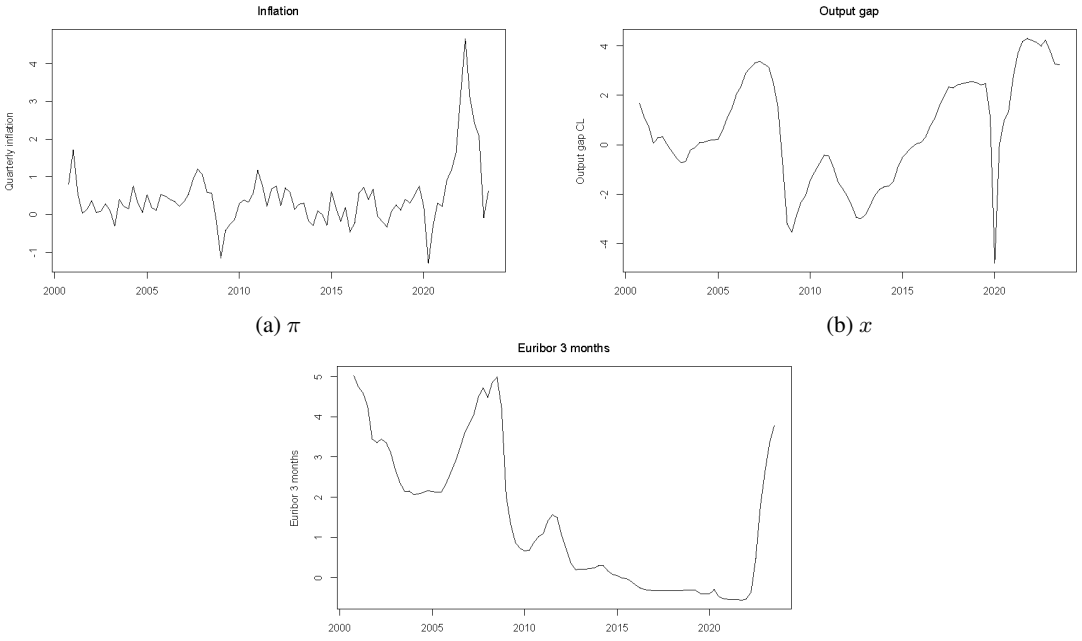
$$\widetilde{\tilde{L}}_T(\boldsymbol{\pi}) = T^{-1} \sum_{t=1}^{T} \ell_t(\boldsymbol{\pi}, \tilde{\boldsymbol{\gamma}}_{\boldsymbol{T}}) \tag{68}$$

Still identification holds up to permutation, scaling and sign indeterminacy for the parameter vector $\boldsymbol{\beta}(\boldsymbol{B})$.

10 Empirical analysis

10.1 Data

The sample data contains 92 observations data from three variables: quarterly inflation (π_t), the 3 months Euribor rate (R_t) and a quarterly measure of output gap (x_t). The observations are referred to the EA from December 2000 to September 2023, thus the dataset contains also data from the financial crisis, the sovereign debt crisis and the Covid 19 pandemic. During the estimation the variables are kept in levels.



(a) π (b) x

(c) R

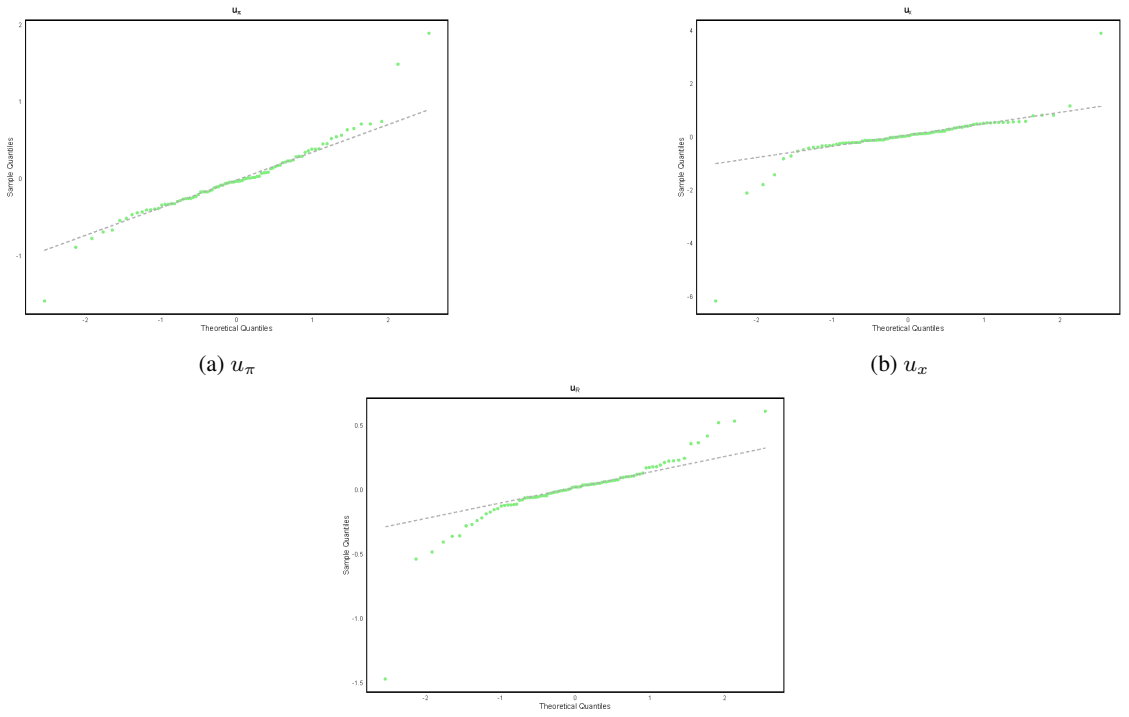
Figure 3: Time series plots of the variables

10.2 Empirical analysis

The reduced form VAR is estimated in levels with a constant by OLS, the AIC, HQ, FPE and the SC all suggest 2 lags and moreover the system is reported to be stable as all the roots are inside the unit circle [4] All the residuals have zero mean and do not show any significant level of cross correlation and their squares.

The application of ICA related method needs two assumptions; mutual independence of the shocks and non-Guassinity of at least $n-1$ shocks where n is the number of variables. Non-Gaussianity is assessed through the Shapiro-Wilk test of the residi Independence cannot be assessed directly ex-ante as you do not directly observe the structural shocks. As we are interested in assessing recursive identification we firstly estimate a Cholesky SVAR; this is straightforward as the structural matrix is directly obtained by the decomposition

[4] The roots are 0.9178894, 0.9178894, 0.6444214, 0.6444214, 0.1125162, 0.1125162



(a) u_π

(b) u_x

(c) u_R

Figure 4: Q-Q plot of the three residuals

of the covariance matrix, and then the columns are rescaled to express the reaction to one standard deviation shock.

$$\hat{\mathbf{B}}_{\text{chol}} = \begin{pmatrix} 0.4596 & 0 & 0 \\ 0.0161 & 0.906 & 0 \\ -0.0029 & 0.05601 & 0.2497 \end{pmatrix} \tag{69}$$

Cholesky SVARs needs the structural matrix to be lower triangular hence limiting the immediate response of some variables to some shocks depending on the way you ordered them. Those assumptions about the immediate response of some variables in the system are necessary to achieve identification.

The confidence intervals at 90% are computed via an i.i.d. non-parametric recursive bootstrap constructed as showed above with starting points of each bootstrap sample set to the sample mean. The IRF's are showed (figure 5) with the median bootstrap value in green and the sample results in blue dashed line. The model seems to behave accordingly to the theory in general, importantly there is no price puzzle and a contractionary monetary policy shock is reported to lower inflation and output gap, while a positive demand shock tends to higher inflation and the short term interest rate; finally the supply shock tends to lower inflation and higher interest rate.

	$\boldsymbol{\nu}$	$\mathbf{A_1}$			$\mathbf{A_2}$		
π	0.1173	**0.7261**	**0.2635**	-0.2954	-0.01145	**-0.1789**	0.2545
x	**0.3120**	0.2099	**0.9032**	-0.0433	0.0524	0.0345	-0.0995
R	-0.0288	0.0883	**0.0828**	**1.4251**	0.0544	**-0.0646**	**-0.4601**

Table 1: Coefficients of the reduced form VAR(2) model, significant parameters are in bold.

10.2.1 Non-Guassian identification

Variable	Sample Skewness	Sample Kurtosis	W statistic	p-value (W)
$u_{\pi t}$	0.6710	6.9909	0.9297	0.0001
$u_{\mathrm{x}t}$	-2.7857	27.6489	0.6176	0.0000
$u_{\mathrm{R}t}$	-2.1135	15.3957	0.8188	0.0000
Joint	·	·	0.6431	0.0000

Table 2: *Shapiro-Wilk normality test on the reduced form residuals along with sample kurtosis and skewness*

We essentially need two more to hold, the first regarding the independence of the underlying components (shocks) the other about the Gaussianity of at most one of this components. The first imply independence in all the moments and cannot be directly tested ex-ante, while the second can be more easily assessed by analyzing the residuals. In practice the Normality assumption can be relaxed to "the residuals are linear combinations of underlying components that are jointly non-Gaussian" which is a weaker assumption. In table(2) are reported the *Shapiro-Wilk* results on the residuals; joint normality is strongly rejected for all residuals at any level. The independence assumption is instead not directly testable obviously as the original components are not observed, and so we must rely on theoretical arguments. As already mentioned, identification through Non-Gaussianity can be implemented through both parametric and non-parametric methods. My parametric log-likelihood estimator relies on **Student-t** density of the shocks and asymptotically efficient under their symmetry [Lanne et al.(2017)Lanne, Meitz, and Saikkonen]; thus is very prone to misspecification and lacking of efficiency. The *Student-t* distribution is common pick both in macro and finance [Brunnermeier et al.(2021)Brunnermeier, Palia, Sastry, and Sims] but still assuming a particular density should be carefully considered. A non-parametric approach could instead be preferred when little is known about the true densities of the shocks. Maximum likelihood estimation is of course very convenient given its known asymptotic distributions, while the other parametric methods asymptotics is not known. So inference is made by relying on a bootstrap procedure, which also allows joint hypothesis testing. The identification problem could also be assessed also via non-parametric estimators; the latters do not need any distributional assumptions but are of course overperformed by a well specified maximum likelihood estimator. Two possible approaches are the *CVM* and *Fast ICA* estimator, here they are implemented by the means of the *svars* and *fast ica* R-packages. [Moneta and Pallante(2022)]perform a simulation study to test the performance of those estimators in a SVAR time series context with scenarios that approach a Gaussian case, claiming that in a general situation no method outperforms the others, thus is worth to estimate the model by the means of all methods, analyzing then the results.
Of course the estimated structural matrix is not enough as we don't have yet any measure of uncertainty

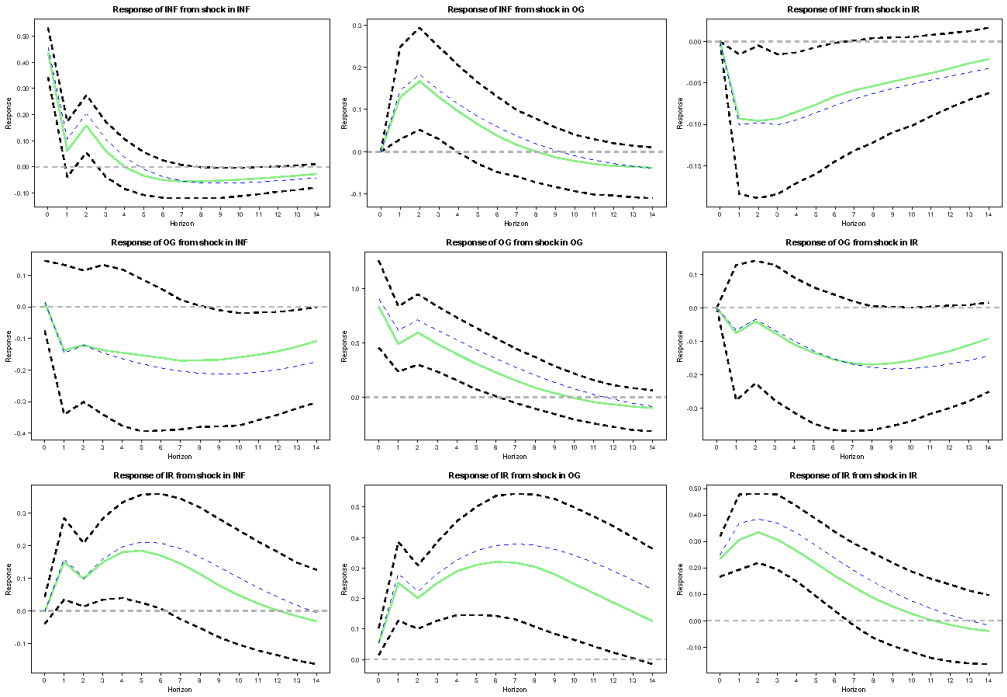


Figure 5: Response to one standard-deviation shock (Cholesky SVAR), 1000 i.i.d. bootstrap median in green, sample IRF in blue

around the estimates and so we are not ready to make inference. In order to obtain such measure we will rely on the bootstrap, that under some conditions and adjustments because of the sign and permutation issue is a suitable method.

In order to assess sign and order of the columns of $\hat{\mathbf{B}}$ I will rely economic theory and I interpret shocks from positive effect on their own variable response. So a monetary policy shock is considered contractionary as reflected by an increase in the interest rate; similarly a supply shock is considered to be negative as reflected by an increase in inflation and lastly the demand shock is positive as represented by an increase in the output gap. If the response of a variable to its own shock is not positive, this column will be changed in sign.

Given the permutation and sign issue the bootstrap procedure is not straightforward and must be assessed accordingly following [Herwartz(2018)]. In fact the initial estimates $\hat{\mathbf{B}}_*$ is not garanteed to follow the shape of $\hat{\mathbf{B}}$. $\mathbf{B_r}$ represent the all possible combination of sign column manipulations and column permutation of a matrix with K columns so that $(2^K)K$ combinations are possible. The bootstrapped confidence intervals are computed via a i.i.d. non-parametric bootstrap in which, at each iteration the $\hat{\mathbf{B}}_*$ matrix estimated on the bootstrap sample is permuted and change in sign in the columns in order to minimize a measure of similarity (distance) with respect to the original $\hat{\mathbf{B}}$ matrix estimated on the sample data. Ideally this procedure aims at preserving the desired shape of the estimated matrix, contrasting the sign and permutation indeterminacy. In this way I pursue the objective of getting a version as close as possible to the structure of $\hat{\mathbf{B}}$. The algorithm is

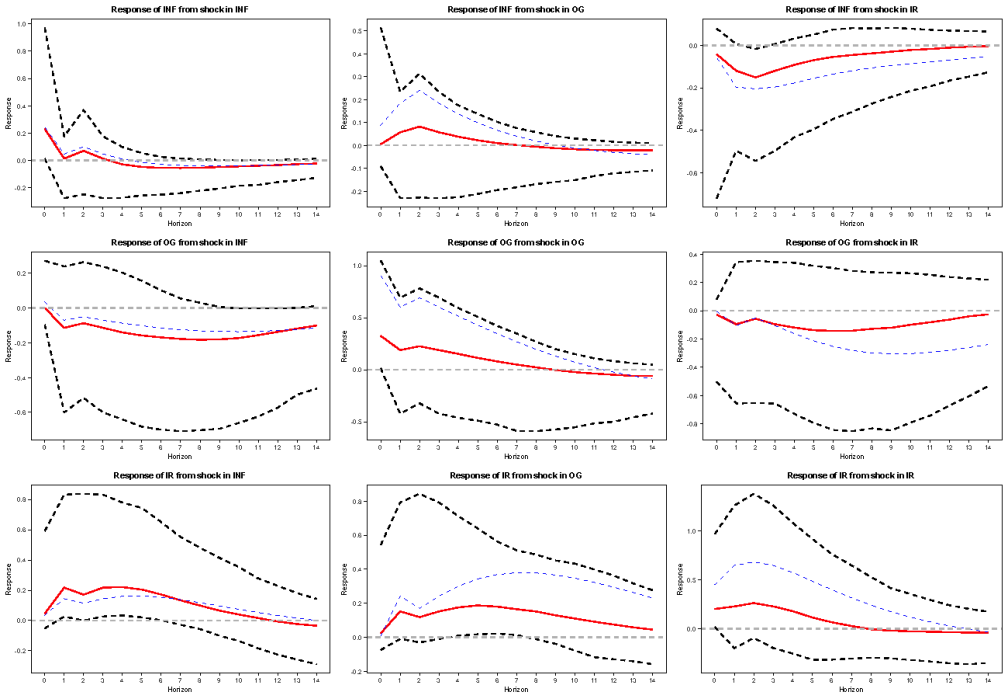


Figure 6: Response to one standard-deviation shock (Fast ICA) with 90% confidence intervals, 1000 residual bootstrap median in red, sample IRF in blue

illustrated in detail above. In order to address the problem of scaling that affects ICA, the bootstrap estimates is rescaled in order to get a comparable scale; as each component is assumed to have unit variance the structural matrix is rescaled accordingly. The similarity criteria I was previously referring to, is defined as follows:

$$\widehat{\mathbf{B}}_{**} = \mathbf{B}_{\hat{r}} = \text{argmin}_r \, \mathcal{D} = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{1}{\widehat{\omega}_{ii}} \left(\widehat{b}_{ij} - \widetilde{b}_{ij}^{(r)} \right)^2 I(\widehat{b}_{ij} \widetilde{b}_{ij}^{(r)} < 0),$$

In the above expression $I(\cdot)$ is the indicator function, $\widehat{\omega}_{ii}$ is i-th diagonal element of $\widehat{\mathbf{\Omega}}$, $\widehat{b}_{ij}$ is an element of $\widehat{\mathbf{B}}$ and $\widetilde{b}_{ij}^{(r)}$ is an element of $\widetilde{\mathbf{B}}_r$, the latter is one of all the possible $(2^K)K! = 16, \; K = 3$ combinations caused by the sign and permutation indeterminacy of the columns. This criteria aims at finding the specific manipulation of the structural matrix among all the possible combinations of the bootstrap estimation that most resembles the "well behaved" $\widehat{\mathbf{B}}$. Dividing by $\widehat{\omega}_{ii}$ consent to take in consideration the distinct conditional variances while the indicator function gives more weight to combination that differ in sign from the well behaved counterpart and so punishing them more. A similar approach to address this matter is using the *Frobenius norm* as a measure of distance to minimize, the results are similar.

Now that the bootstrap estimates are designed to reflect the sample structure of the $\mathbf{B}$ and recalling that every permutation and change of sign of the columns is observationally equivalent, we can reshape the structure of $\hat{\mathbf{B}}$ such it actually hopefully reflects the original shape of $\mathbf{B}$. To do so we rely on economic theory and com-

Algorithm 1 Bootstrap Algorithm for ICA SVAR model

1: **for** $i = 1$ to B **do**
2: Generate bootstrap sample via i.i.d. non-paramatric bootstrap
3: Refit the VAR model on the bootstrap sample
4: Compute the eigen-decompositions of the covariance matrix $\hat{\mathbf{\Omega}}_*$
5: Use any ICA estimator to get $\hat{\mathbf{B}}_*$
6: Normalize $\hat{\mathbf{B}}_*$ by multiplying by diag$(\sigma_{\boldsymbol{\varepsilon}1t}, \sigma_{\boldsymbol{\varepsilon}2t}, \sigma_{\boldsymbol{\varepsilon}3t})$ i.e. multiply each column by the correspondent shock standard deviation
7: Select the permutation and sign variant from the set of all possible combination $(2^K)K! = 16,\ K = 3$ which minimizes a similarity criterion with respect to the original $\hat{\mathbf{B}}$ matrix and call it $\hat{\mathbf{B}}_{**}$.
8: Compute and store the IRF's
9: Store $\hat{\mathbf{B}}_{**}$
10: **end for**

bining the fact that we want to evaluate positive shocks effect with the plausible assumption (also reflected by the Cholesky model) that each variable is likely to respond more heavily to an own shock, we get to the following IRFs order and sign (figure 7):

Once the bootstrap procedure has been peformed, the recursive identification scheme can be tested in the form of this test:

$$\lambda_{JS} = \left(R\,\text{vec}(\hat{\mathbf{B}}) - \mathbf{r}\right)' \left(\text{Cov}\left(\text{vec}(\hat{\mathbf{B}}^{**})\right)\right)^{-1} \left(R\,\text{vec}(\hat{\mathbf{B}}) - \mathbf{r}\right) \approx \chi^2_{(J)}$$

where $\mathbf{R}$ is a known $J \times K^2$ dimensional selection matrix of rank J,$\mathbf{r}$ is a known $J \times 1$ vector representing the considered restrictions, and that the composite null hypothesis is

$$H_0 : \mathbf{R}\,\text{vec}(\mathbf{B}) = \mathbf{r}.$$

$\hat{\mathbf{B}}_{**}$ is the bootstrap version of the covariance decomposition matrix ad is constructed as follows:FORMULA
As we are interested in testing the lower-triangularity of $\hat{\mathbf{B}}$, $\mathbf{r} = (0, 0, 0)'$ and $\mathbf{R}$ is such that to select the elements above the diagonal in the structural matrix.
This test do not reject the recursive identification scheme of the structural matrix at any level of significance.

10.2.2 Fast ICA

In order to estimate the model through Fast ICA I rely on the R package *fastICA*. The algorithm runs as follows: first the data are centered by subtracting the mean of each column of the residuals, then the data are whitened and lastly the program search for the structural matrix that maximizes the neg-entropy approximation subject to constrained that the desired matrix is orthonormal.

$$\hat{\mathbf{B}}_{\text{FastICA}} = \begin{pmatrix} 0.2437 & 0.08676 & -0.0587 \\ \text{(0.2299)} & \text{(0.0053)} & \text{(-0.0422)} \\ 0.0329 & 0.9081 & -0.0065 \\ \text{(-0.0001)} & \text{(0.3249)} & \text{(-0.0275)} \\ 0.0350 & 0.0082 & 0.4532 \\ \text{(0.0465)} & \text{(0.0236)} & \text{(0.2034)} \end{pmatrix} \tag{70}$$

The sample estimates are reported above, with the related bootstrap median in parenthesis.

11 Conclusions

I estimated a SVAR(2) model and under the assumptions of non-Normality and independence of the shocks, I performed both parametric and non-parametric techniques aim at finding the structural matrix. These estimates are subject to an high level of uncertainty even after controlling for permutation and sign of the columns and so I am not able to reject the recursive identification scheme in the Euro Area sample. Given the nature of the sample a more sophisticated model could have helped to achieve identification more accurately.

References

[Bernanke(1986)] Ben Bernanke. Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy*, 25(1):49–99, 1986. URL https://EconPapers.repec.org/RePEc:eee:crcspp:v:25:y:1986:i::p:49-99.

[Blanchard and Watson(1986)] Olivier Blanchard and Mark Watson. Are business cycles all alike? In *The American Business Cycle: Continuity and Change*, pages 123–180. National Bureau of Economic Research, Inc, 1986. URL https://EconPapers.repec.org/RePEc:nbr:nberch:10021.

[Blanchard and Quah(1989)] Olivier Jean Blanchard and Danny Quah. The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79(4):655–673, September 1989. URL https://ideas.repec.org/a/aea/aecrev/v79y1989i4p655-73.html.

[Brunnermeier et al.(2021)Brunnermeier, Palia, Sastry, and Sims] Markus Brunnermeier, Darius Palia, Karthik A. Sastry, and Christopher A. Sims. Feedbacks: Financial markets and economic activity. *American Economic Review*, 111(6):1845–79, June 2021. doi: 10.1257/aer.20180733. URL https://www.aeaweb.org/articles?id=10.1257/aer.20180733.

[Comon(1994)] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3): 287–314, 1994. ISSN 0165-1684. doi: https://doi.org/10.1016/0165-1684(94)90029-9. URL https://www.sciencedirect.com/science/article/pii/0165168494900299. Higher Order Statistics.

[Fiorentini and Sentana(2023)] Gabriele Fiorentini and Enrique Sentana. Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. *Journal of Econometrics*, 235(2):643–665, 2023. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2022.02.010. URL https://www.sciencedirect.com/science/article/pii/S0304407622001269.

[Gouriéroux et al.(2017)Gouriéroux, Monfort, and Renne] Christian Gouriéroux, Alain Monfort, and Jean-Paul Renne. Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics*, 196(1):111–126, 2017. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2016.09.007. URL https://www.sciencedirect.com/science/article/pii/S0304407616301749.

[Guay(2021)] Alain Guay. Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics*, 225(1):27–46, 2021. doi: 10.1016/j.jeconom.2020.10. URL https://ideas.repec.org/a/eee/econom/v225y2021i1p27-46.html.

[Gürkaynak et al.(2005)Gürkaynak, Sack, and Swanson] Refet S Gürkaynak, Brian Sack, and Eric Swanson. Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements. *International Journal of Central Banking*, 1(1), May 2005. URL https://ideas.repec.org/a/ijc/ijcjou/y2005q2a2.html.

[Herwartz(2018)] Helmut Herwartz. Hodges–Lehmann Detection of Structural Shocks – An Analysis of Macroeconomic Dynamics in the Euro Area. *Oxford Bulletin of Economics and Statistics*, 80(4):

736–754, August 2018. doi: 10.1111/obes.12234. URL https://ideas.repec.org/a/bla/obuest/v80y2018i4p736-754.html.

[Hyvärinen et al.(2004)Hyvärinen, Karhunen, and Oja] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2004. ISBN 9780471464198. URL https://books.google.it/books?id=96D0ypDwAkkC.

[Hyvärinen et al.(2010)Hyvärinen, Zhang, Shimizu, and Hoyer] Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O. Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(56):1709–1731, 2010. URL http://jmlr.org/papers/v11/hyvarinen10a.html.

[Jordà(2005)] Òscar Jordà. Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182, March 2005. doi: 10.1257/0002828053828518. URL https://www.aeaweb.org/articles?id=10.1257/0002828053828518.

[Keweloh(2023)] Sascha A. Keweloh. Structural Vector Autoregressions and Higher Moments: Challenges and Solutions in Small Samples. Papers 2310.08173, arXiv.org, October 2023. URL https://ideas.repec.org/p/arx/papers/2310.08173.html.

[Lange et al.(2021)Lange, Dalheimer, Herwartz, and Maxand] Alexander Lange, Bernhard Dalheimer, Helmut Herwartz, and Simone Maxand. svars: An r package for data-driven identification in multivariate time series analysis. *Journal of Statistical Software*, 97(5):1–34, 2021. doi: 10.18637/jss.v097.i05. URL https://www.jstatsoft.org/index.php/jss/article/view/v097i05.

[Lanne and Ltkepohl(2008)] Markku Lanne and Helmut Ltkepohl. Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking*, 40(6):1131–1149, 2008. URL https://EconPapers.repec.org/RePEc:mcb:jmoncb:v:40:y:2008:i:6:p:1131-1149.

[Lanne et al.(2017)Lanne, Meitz, and Saikkonen] Markku Lanne, Mika Meitz, and Pentti Saikkonen. Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304, 2017. URL https://EconPapers.repec.org/RePEc:eee:econom:v:196:y:2017:i:2:p:288-304.

[Lewis(2024)] Daniel Lewis. Identification based on higher moments. CeMMAP working papers 03/24, Institute for Fiscal Studies, February 2024. URL https://ideas.repec.org/p/azt/cemmap/03-24.html.

[Ludvigson et al.(2021)Ludvigson, Ma, and Ng] Sydney C. Ludvigson, Sai Ma, and Serena Ng. Uncertainty and business cycles: Exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics*, 13(4):369–410, 2021. URL https://EconPapers.repec.org/RePEc:aea:aejmac:v:13:y:2021:i:4:p:369-410.

[Mertens and Ravn(2013)] Karel Mertens and Morten O. Ravn. The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, 103(4):1212–47, June

2013. doi: 10.1257/aer.103.4.1212. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.4.1212.

[Moneta and Pallante(2022)] Alessio Moneta and Gianluca Pallante. Identification of Structural VAR Models via Independent Component Analysis: A Performance Evaluation Study. *Journal of Economic Dynamics and Control*, 144(C), 2022. doi: 10.1016/j.jedc.2022.10453. URL https://ideas.repec.org/a/eee/dyncon/v144y2022ics0165188922002342.html.

[Moneta et al.(2013)Moneta, Entner, Hoyer, and Coad] Alessio Moneta, Doris Entner, Patrik O. Hoyer, and Alex Coad. Causal Inference by Independent Component Analysis: Theory and Applications. *Oxford Bulletin of Economics and Statistics*, 75(5):705–730, October 2013. URL https://ideas.repec.org/a/bla/obuest/v75y2013i5p705-730.html.

[Nakamura and Steinsson(2018)] Emi Nakamura and Jón Steinsson. Identification in macroeconomics. *Journal of Economic Perspectives*, 32(3):59–86, August 2018. doi: 10.1257/jep.32.3.59. URL https://www.aeaweb.org/articles?id=10.1257/jep.32.3.59.

[Rothenberg(1971)] Thomas J. Rothenberg. Identification in parametric models. *Econometrica*, 39(3):577–591, 1971. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913267.

[Sentana and Fiorentini(2001)] Enrique Sentana and Gabriele Fiorentini. Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics*, 102(2):143–164, June 2001. URL https://ideas.repec.org/a/eee/econom/v102y2001i2p143-164.html.

[Sims(1980)] Christopher Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980. URL https://EconPapers.repec.org/RePEc:ecm:emetrp:v:48:y:1980:i:1:p:1-48.

[Sims(1992)] Christopher A. Sims. Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5):975–1000, 1992. ISSN 0014-2921. doi: https://doi.org/10.1016/0014-2921(92)90041-T. URL https://www.sciencedirect.com/science/article/pii/001429219290041T.

[Stock and Watson(2012)] James H. Stock and Mark Watson. Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity*, 43(1 (Spring)):81–156, 2012. URL https://EconPapers.repec.org/RePEc:bin:bpeajo:v:43:y:2012:i:2012-01:p:81-156.

[Uhlig(2005)] Harald Uhlig. What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419, March 2005. URL https://ideas.repec.org/a/eee/moneco/v52y2005i2p381-419.html.

12 Appendix

$$\hat{\mathbf{B}}_{\text{ML}} = \begin{pmatrix} 0.4593 & 0.0511 & -0.024 \\ -0.0087 & 0.9073 & 0.0105 \\ 0.0359 & 0.112 & 0.2463 \end{pmatrix}, \hat{\boldsymbol{\lambda}}_{ML} = (3.612989, 1.853487, 2.225749) \tag{71}$$

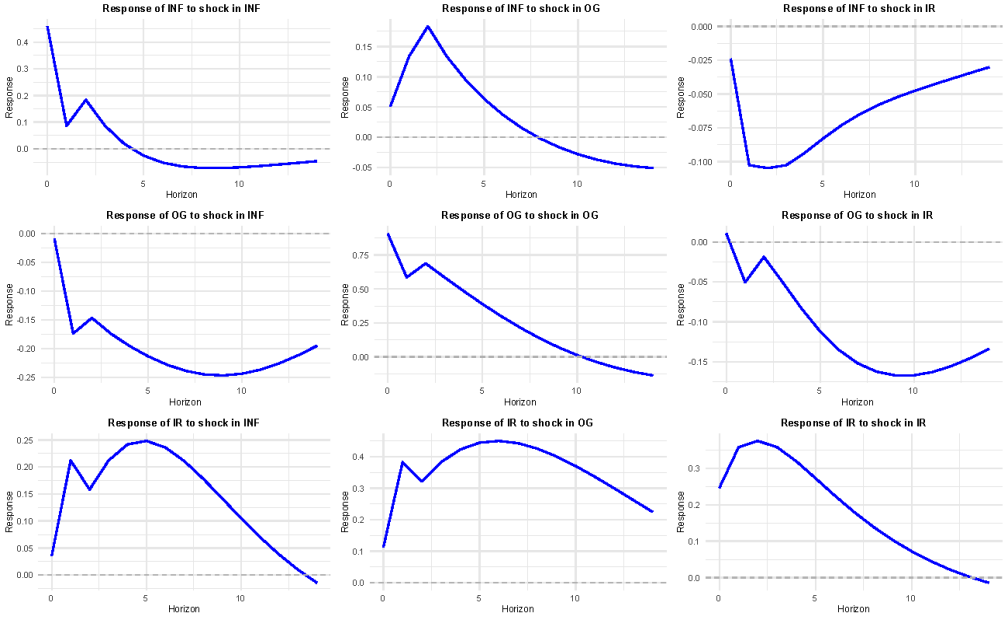


Figure 7: Response to one standard-deviation shock, Student-t ML point estimates

The parametric model instead allow for a *likelihood ratio test*, comparing the likelihood of the unrestricted model denoted as $\tilde{\mathbf{B}}$ and restricted model $\tilde{\mathbf{B}}_r$ in which the elements above the diagonal are set to zero. The likelihood ratio statistics is computed as follows with $N = 3$:

$$\lambda_{LR} = 2\left[\log \mathcal{L}\left(\text{vec}(\tilde{\mathbf{B}})\right) - \log \mathcal{L}\left(\text{vec}(\tilde{\mathbf{B}}_{\mathbf{r}})\right)\right] \sim \chi^2(N)$$

The test to not reject the recursive identification scheme.