Estimation of Non-Gaussian SVAR Using Tensor Singular Value Decomposition*

Alain Guay and Dalibor Stevanovic†

August 2025

Abstract

This paper introduces a tensor singular value decomposition (TSVD) approach for estimating non-Gaussian Structural Vector Autoregressive (SVAR) models. The proposed methodology applies to both complete and partial identification of structural shocks. The estimation procedure relies on third- and/or fourth-order cumulants. We establish the asymptotic distribution of the estimator and conduct a simulation study to evaluate its finite-sample performance. The results demonstrate that the estimator is highly competitive in small samples compared to alternative methods under complete identification. In cases of partial identification, the estimator also exhibits very good performance in small samples. To illustrate the practical relevance of the procedure under partial identification, two empirical applications are presented.

JEL classification: C12, C32, C51.
Keywords: Non-Gaussian SVAR, tensor decomposition, cumulants.

*We thank Christian Gouriéroux and Éric Ghysels for helpful discussions, as well as participants at the NBER-NSF Time Series Conference 2024, the CEA and CESG Annual Meetings 2024, the CIREQ-CMP Econometrics Conference in Honor of Éric Ghysels 2024 and the 2025 BSE Summer Forum workshop on Advances in Structural Shocks Identification for helpful comments. *Correspondence*: Alain Guay, Department of Economics, ESG-UQAM Montréal, 3120 Sainte-Catherine est, Montréal, Québec, Canada, H2X 3X2. E-mail: guay.alain@uqam.ca.

†Université du Québec à Montréal, CIREQ and Chaire en macroéconomie et prévisions ESG-UQAM.

1. Introduction

The Structural Vector Autoregressive (SVAR) model is one of the most important tools in applied macroeconomics for estimating the effects of structural shocks. A prominent branch of the SVAR literature relies on the standard assumption that structural shocks are orthogonal, using the information contained in the unconditional covariances of the reduced-form innovations to identify the structural parameters. However, this information alone is insufficient to fully identify all parameters. As a result, additional identifying assumptions are required, such as short-run restrictions (Sims, 1980), long-run restrictions (Blanchard and Quah, 1989) or sign restrictions (Uhlig, 2005).

Recently, the statistical identification of SVAR parameters through independent non-Gaussian shocks has been proposed by Lanne et al. (2017) and Gouriéroux et al. (2017). Under the assumption that at most one structural shock is Gaussian, various estimation methods are available to recover the structural shocks and their corresponding impulse response functions (IRFs). Parametric approaches include Maximum Likelihood (Lanne et al., 2017) and Pseudo Maximum Likelihood (PML) (Gouriéroux et al., 2017; Fiorentini and Sentana, 2023), utilizing a discrete scale mixture of normals in symmetric cases or an unrestricted finite mixture in general. Semiparametric approaches include the method proposed by Lanne et al. (2023) and the Method of Moments (Guay, 2021; Keweloh, 2020; Lanne and Luoto, 2021). Importantly, this identification is purely statistical, yet it can serve as a useful benchmark for evaluating and testing standard economic identification.[1]

A key point is that, in this literature, identification conditions derived from non-Gaussianity hinge exclusively on the number of non-Gaussian *structural shocks*, rather than on the number of non-Gaussian reduced-form innovations. This distinction underpins the estimation procedure developed in this paper, which accommodates both full identification of structural shocks and the identification of a subset thereof. We refer to the latter as *partial identification*.[2] Partial identification is particularly relevant when only certain series exhibit asymmetric dynamics over the business cycle or time-varying volatility. Non-Gaussianity in reduced-form innovations may also arise from only a subset of structural shocks. In such cases, the number of non-Gaussian structural shocks can be smaller than the number of non-Gaussian reduced-form innovations. This situation occurs, for example, when several reduced-form innovations share a common source of

[1] Since the procedure is statistical in nature, it is not intended to provide direct economic interpretations—for instance, it does not imply whether monetary or fiscal policy shocks are more or less skewed than aggregate demand or supply shocks.

[2] Partial identification is used here in the sense of Phillips (1989), where full identification is not assumed and only a subset of parameters is identified (see also Gouriéroux and Jasiak (2023)). This differs from interpretations in which identification is not an all-or-nothing property, and partially identified models can still yield valuable information (see, e.g., Tamer (2010)).

non-Gaussianity, such as exposure to structural business-cycle or volatility factors.[3]

Here, we develop a formal procedure that explicitly relies on the dimension of structural non-Gaussianity for identification. To the best of our knowledge, this is the first approach to systematically exploit this dimension for both identification and estimation. Moreover, the distinction between Gaussian and non-Gaussian shocks is not always clear-cut in finite samples. In such cases, partial identification is especially useful because it allows researchers to focus on those shocks that deviate most strongly from Gaussianity, thereby reducing the risk of weak identification.[4]

The proposed approach relies exclusively on unconditional moments, in particular on higher-order cumulants, thereby avoiding the need for a fully specified model and enhancing robustness to misspecification.[5] Higher-order cumulants can be naturally represented as tensors. The multilinear algebraic structure of tensors enables the identification of independent sources, analogous to the Independent Component Analysis (ICA) problem. This identification can be achieved using multilinear generalizations of the singular value decomposition (SVD). Several tensor decompositions, each with distinct properties, have been proposed as extensions of the SVD.

When only a subset of the structural shocks is identified, a low-rank approximation becomes necessary. In the case of matrices, such an approximation can be readily obtained by truncating the SVD or, equivalently, by performing the eigenvalue decomposition (EVD) of the quadratic form that yields the principal components. However, extending this approach to higher-order cumulants is not straightforward. Unlike the matrix case, the best low-rank tensor approximation cannot be obtained sequentially; instead, all factors must be estimated simultaneously.

Consequently, a specific tensor decomposition, known as Tensor SVD, is introduced to identify structural shocks and their corresponding impulse response functions (IRFs). Tensor SVD maximizes the diagonality of the tensor while ensuring that the vectors remain orthogonal. A key property exploited by this method is that the higher-order cumulant of a stochastic vector with mutually independent components forms a diagonal tensor, meaning that only the entries with identical indices are nonzero. This property remains valid under the weaker conditions of no coskewness and/or no cokurtosis.

Alternatively, a Higher-Order Singular Value Decomposition (HOSVD) can be applied for dimensionality reduction by retaining only the principal components, which are computed as the

[3] See Sentana and Fiorentini (2001) and Normandin and Phaneuf (2004) for volatility factor structures. Such features are also consistent with macro risk episodes such as the Great Moderation or the COVID-19 recession (Bekaert et al., 2025; Montiel Olea et al., 2022).

[4] Recent estimation alternatives for partial identification include the Generalized Covariance (GCov) estimator of Gouriéroux and Jasiak (2023), a Bayesian method proposed by Anttonen et al. (2023), and a kernel-based estimator by Hafner et al. (2025), though these approaches do not test the dimension of structural non-Gaussianity.

[5] See also Lewis (2021) for volatility-based identification without parametric assumptions.

leading terms of the rank-truncated HOSVD. This approach can be interpreted as a form of Tensor Principal Component Analysis (see Babii et al. (2024)). Another possible approach is the Moment Component Analysis (MCA) introduced by Jondeau et al. (2018), which employs comoments of order three and four along with an efficient technique known as *Higher-Order Orthogonal Iteration* (HOOI) for computing low-rank tensor decomposition (see De Lathauwer et al. (2000b)). However, neither the truncated HOSVD nor the HOOI-based decomposition ensures a diagonal tensor, which is essential in our context for identifying structural shocks.

The proposed approach is implemented in two steps. The first step applies an orthogonal transformation to the impact matrix of the structural shocks, reducing the problem to identifying a single orthogonal matrix. In the second step, the orthogonal matrix is identified and estimated using higher-order cumulants through a tensor singular value decomposition (Tensor SVD). Because under our identification assumptions the higher-order cumulant tensors of the structural shocks are diagonal, our estimator maximizes their diagonal elements.

The advantage of maximizing cumulants with identical indices is that it reduces the optimization problem to only n cumulants, rather than minimizing $O(n^J)$ cross-cumulants as required in a full-identification setting such as the GMM framework for $J = 3$ or $J = 4$, where n denotes the number of variables and J the order of the cumulant (see Comon, 1994). Moreover, partial identification cannot be directly implemented within the GMM framework without imposing arbitrary restrictions solely to satisfy the requirements of full model identification.

This paper also establishes the asymptotic distribution of the proposed estimator, which is based on the Tensor SVD, as previously mentioned. To our knowledge, only a few studies have examined the asymptotic properties of low-rank decompositions involving third- and fourth-order moments or cumulants. Miettinen et al. (2015) derived the asymptotic properties of traditional estimation procedures in Independent Component Analysis (ICA) that utilize fourth moments, such as fourth-order blind identification, joint approximate diagonalization of eigenmatrices (JADE) and fastICA. Furthermore, the asymptotic properties of JADE procedures have been analyzed in the context of factor analysis based on third- and fourth-order cumulants by Bonhomme and Robin (2009).[6]

A simulation study shows that the proposed estimator delivers highly competitive small-sample performance compared to alternative methods under complete identification. In particular, maximizing non-Gaussianity yields clear gains relative to minimizing cross-cumulants in the GMM framework. Under partial identification, the proposed estimator also demonstrates robust small-sample properties.

[6] Mesters and Zwiernik (2023) derived the asymptotic properties of GMM estimation using moment or cumulant tensors for non-independent component analysis, but within a fully identified context.

Finally, we illustrate our estimation procedure with two applications. First, we examine the effects of fiscal policies on economic activity by analyzing a trivariate SVAR process that includes taxes, public spending, and output for the U.S. Rank tests developed by Guay (2021) indicate that all structural shocks are symmetric, while one is non-mesokurtic. Based on these findings, the identification condition and estimation results confirm that the subsystem linking all variables to the tax shock is identified. This identification allows us to compare our proposed Tensor SVD estimator with GMM estimators, which, however, require identification of the entire system.

In the second application, we investigate the effects of credit shocks in a four-variable VAR based on Boivin et al. (2020). Rank tests reveal the presence of two structural shocks that deviate clearly from Gaussianity, displaying both asymmetry and excess kurtosis. Exploiting the information embedded in these higher-order moments yields impulse responses that correspond more closely to theoretical DSGE models with financial frictions.

This paper is organized as follows. Section 2 presents the SVAR specification and outlines the sufficient conditions for local statistical identification using higher-order cumulants. Section 3 provides a general introduction to tensors and tensor decompositions. Section 4 details the proposed approach based on Tensor SVD and includes the asymptotic distribution results. Section 5 presents the results of simulation experiments, and the final section discusses two empirical applications.

2. SVAR and Cumulants

This section presents the SVAR specification and the sufficient conditions for local statistical identification using higher-order cumulants.

2.1 Specification

We consider a structural system represented by the following p-order SVAR process:

$$\Phi x_t = \Phi_0 + \sum_{\tau=1}^{p} \Phi_\tau x_{t-\tau} + \epsilon_t, \tag{1}$$

where x_t is an $(n \times 1)$ vector of endogenous variables, and ϵ_t is an $(n \times 1)$ vector of structural shocks. The shocks are assumed to have zero mean and an identity variance–covariance matrix. The vector Φ_0 contains n unrestricted intercepts, while the non-singular matrix Φ captures the n^2 unrestricted contemporaneous relations among the variables.[7] Finally, each $(n \times n)$ matrix Φ_τ encodes the n^2 unrestricted dynamic feedback effects at lag τ.

[7] Non-singularity ensures that no redundant variables are included in the SVAR system.

The reduced form associated with system (1) corresponds to the following p-order VAR process:

$$x_t = \Gamma_0 + \sum_{\tau=1}^{p} \Gamma_\tau x_{t-\tau} + \nu_t, \tag{2}$$

where $\Gamma_0 = \Theta\Phi_0$, $\Gamma_\tau = \Theta\Phi_\tau$, and the non-singular matrix $\Theta = \Phi^{-1}$ captures the impact responses of the variables of interest to the various structural shocks, while ν_t represents the reduced-form innovations. These innovations are related to the structural shocks as follows: $\nu_t = \Theta\epsilon_t$.

Now we present the corresponding cumulants of the structural shocks and the reduced-form innovations. Consider the $n \times 1$ vector of structural shocks ϵ_t. The dth-order cumulant, denoted $\mathcal{C}^d$, is given by:

$$\mathcal{C}^d_{i_1 i_2 \ldots i_d}(\epsilon) = \text{Cum}(\epsilon_{i_1,t}, \epsilon_{i_2,t}, \ldots, \epsilon_{i_d,t}).$$

An important advantage of higher-order cumulants over higher-order moments is that, for a Gaussian random variable, all cumulants of order $d > 2$ are zero. Furthermore, for independent variables, all cross-cumulants are zero.[8] This implies that only terms with identical indices are nonzero for independent variables. The same property holds for sets of variables satisfying weaker independence conditions, such as zero coskewness or zero cokurtosis, meaning that third- and fourth-order cumulants vanish except for terms where all indices match. These cumulants are theoretically diagonal, i.e., only entries where all indices are equal can be nonzero. Consequently, for uncorrelated variables, the second-order cumulant is diagonal, and the same property extends to higher-order cumulants under conditions of zero coskewness, zero cokurtosis, or independence.

The first four cumulants of the structural shocks from system (1) are given by the following expressions:

$$\begin{aligned}
\mathcal{C}_i(\epsilon) &= \text{Cum}(\epsilon_{it}) = \mathbb{E}[\epsilon_{it}], \\
\mathcal{C}^2_{i,j}(\epsilon) &= \text{Cum}(\epsilon_{it}, \epsilon_{jt}) = \mathbb{E}[\epsilon_{it}\epsilon_{jt}], \\
\mathcal{C}^3_{i,j,k}(\epsilon) &= \text{Cum}(\epsilon_{it}, \epsilon_{jt}, \epsilon_{kt}) = \mathbb{E}[\epsilon_{it}\epsilon_{jt}\epsilon_{kt}], \qquad (3) \\
\mathcal{C}^4_{i,j,k,l}(\epsilon) &= \text{Cum}(\epsilon_{it}, \epsilon_{jt}, \epsilon_{kt}, \epsilon_{lt}) \\
&= \mathbb{E}[\epsilon_{it}\epsilon_{jt}\epsilon_{kt}\epsilon_{lt}] - \mathbb{E}[\epsilon_{it}\epsilon_{jt}]\mathbb{E}[\epsilon_{kt}\epsilon_{lt}] \\
&\quad - \mathbb{E}[\epsilon_{it}\epsilon_{kt}]\mathbb{E}[\epsilon_{jt}\epsilon_{lt}] - \mathbb{E}[\epsilon_{it}\epsilon_{lt}]\mathbb{E}[\epsilon_{jt}\epsilon_{kt}]. \qquad (4)
\end{aligned}$$

Here, $\mathbb{E}$ denotes the unconditional expectation operator. The elements of the first-order cumulant represent the unconditional means of each structural shock i. The second-order cumulant corresponds to the variance-covariance matrix of the structural shocks, given by: $\mathbb{E}(\epsilon\epsilon') = I$, which

[8] See Appendix A for additional properties of cumulants.

implies that $\mathcal{C}^2_{i,j}(\epsilon) = 0$ for $i \neq j$ and $\mathcal{C}^2_{i,i}(\epsilon) = 1$ for all i.

The n unconstrained skewnesses of the structural shocks may be nonzero, i.e., $\mathcal{C}^3_{i,j,k}(\epsilon) \neq 0$ for $i = j = k$, while all cross-cumulants (coskewnesses) are assumed to be zero, i.e., $\mathcal{C}^3_{i,j,k}(\epsilon) = 0$ for $i \neq j, k$. Similarly, the n unconstrained fourth-order cumulants may be nonzero, i.e., $\mathcal{C}^4_{i,j,k,l}(\epsilon) \neq 0$ for $i = j = k = l$, while cross-cumulants are assumed to be zero, meaning that the excess cokurtoses satisfy $\mathcal{C}^4_{i,j,k,l}(\epsilon) = 0$ for $i \neq j, k, l$. For notational simplicity, the third-order cumulant is denoted by $\mathcal{C}^3 \in \mathbb{R}^{n \times n \times n}$, where each element (i, j, k) is given by equation (3). Similarly, the fourth-order cumulant, denoted by and $\mathcal{C}^4 \in \mathbb{R}^{n \times n \times n \times n}$ has each element (i, j, k, l) defined by equation (4).

These third- and fourth-order cumulants can also be expressed in matrix form, respectively:

$$\mathbf{C}^3_\epsilon = \mathbb{E}[\epsilon_t \epsilon_t' \otimes \epsilon_t'], \tag{5}$$

$$\mathbf{C}^4_\epsilon = \mathbb{E}[\epsilon_t \epsilon_t' \otimes \epsilon_t' \otimes \epsilon_t'] - \mathbb{E}[\tilde{\epsilon}_t \tilde{\epsilon}_t' \otimes \tilde{\epsilon}_t' \otimes \tilde{\epsilon}_t'], \tag{6}$$

where $\otimes$ denotes the Kronecker product, and $\tilde{\epsilon}_t$ represents hypothetical structural shocks following a multivariate normal distribution.

The first four cumulants of the reduced-form innovations are defined similarly to those of the structural shocks ϵ_t. By the multilinearity properties of cumulants, the second-, third-, and fourth-order cumulants of the innovations are related to the cumulants of the structural shocks as follows:

$$\begin{aligned}
\mathcal{C}^2_{i,j}(\nu) &= \sum_{m=1}^{n} \Theta_{im}\Theta_{jm}, \\
\mathcal{C}^3_{i,j,k}(\nu) &= \sum_{m=1}^{n} \Theta_{im}\Theta_{jm}\Theta_{km}\mathcal{C}_{m,m,m}(\epsilon), \\
\mathcal{C}^4_{i,j,k,l}(\nu) &= \sum_{m=1}^{n} \Theta_{im}\Theta_{jm}\Theta_{km}\Theta_{lm}\mathcal{C}_{m,m,m,m}(\epsilon),
\end{aligned}$$

where $\mathcal{C}^3_{m,m,m}(\epsilon)$ and $\mathcal{C}^4_{m,m,m,m}(\epsilon)$ represent the skewness and excess (positive or negative) kurtosis of the structural shocks m.[9] The objective is to recover the mixing matrix Θ using an estimator of the cumulants $\mathcal{C}^2_{i,j}(\nu)$, $\mathcal{C}^3_{i,j,k}(\nu)$, and/or $\mathcal{C}^4_{i,j,k,l}(\nu)$.

In matrix form, this becomes:

$$\begin{aligned}
\mathbf{C}^2_\nu &= \mathbb{E}[\nu\nu'] = \Sigma_\nu = \Theta\Theta', \\
\mathbf{C}^3_\nu &= \mathbb{E}[\nu_t \nu_t' \otimes \nu_t'] = \Theta\mathbf{C}^3_\epsilon(\Theta' \otimes \Theta'), \\
\mathbf{C}^4_\nu &= \Theta\mathbf{C}^4_\epsilon(\Theta' \otimes \Theta' \otimes \Theta').
\end{aligned}$$

[9] A statistical distribution with negative excess kurtosis is called a platykurtic distribution, while a distribution with positive excess kurtosis is known as a leptokurtic distribution. A mesokurtic distribution has an excess kurtosis of zero.

As is well known, the symmetric matrix $\mathbf{C}_\nu^2 = \Sigma_\nu$ contains $\frac{n(n+1)}{2}$ distinct elements. Furthermore, the matrices $\mathbf{C}_\nu^3$ and $\mathbf{C}_\nu^4$ include $\frac{n(n+1)(n+2)}{6}$ and $\frac{n(n+1)(n+2)(n+3)}{24}$ distinct elements, respectively.

2.2 Identification Conditions with Non-gaussianity

As outlined by Gouriéroux et al. (2017) and Lanne et al. (2017), the structural shocks ϵ_t are often assumed to be i.i.d., mutually independent random variables, with at most one Gaussian component. In this paper, we relax the mutual independence assumption by requiring instead that the shocks exhibit zero cross-sectional covariances, coskewnesses, and/or excess cokurtoses, following Guay (2021) and Mesters and Zwiernik (2023), at the cost of assuming the existence of higher-order moments. Under this framework, the complete identification assumptions based on skewness and/or excess kurtosis are:

Assumption 1 *Complete Identification:* ***CI***

(i) The structural shocks ϵ_t have finite moments up to the fourth order.

(ii) The third- and fourth-order cross-cumulants of the structural shocks ϵ_t are zero.

(iii) All but at most one of the structural shocks exhibit skewness and/or excess kurtosis (i.e., are non-mesokurtic).

Under condition **CI**, the impact matrix is identifiable up to post-multiplication by $\mathcal{D}\mathcal{P}$, where $\mathcal{D}$ is a diagonal matrix with diagonal entries equal to ± 1, and $\mathcal{P}$ is a permutation matrix. Assumption 1(ii) can be relaxed by requiring that only a sufficient number of cross-cumulants vanish (see Lanne and Luoto (2021) and Mesters and Zwiernik (2023)).[10]

In the case of partial identification (**PI**), Guay (2021) provides sufficient rank conditions for the partial identification of the impact matrix Θ. These conditions ensure that the subsystem linking reduced-form innovations to the subset of structural shocks that are asymmetric and/or non-mesokurtic is statistically identified, up to sign changes and column permutations. Consequently, the columns of Θ corresponding to structural shocks characterized by skewness and/or excess kurtosis are locally identified.

[10] However, according to the following general definition of structural IRFs (Gouriéroux et al., 2020),

$$\text{IRF}_{ij}(h) = \mathbb{E}\big[x_{i,t+h} \,\big|\, \epsilon_{j,t} = 1,\ \epsilon_{k,t} = 0\ \ \forall k \neq j,\ \mathcal{F}_{t-1}\big] - \mathbb{E}\big[x_{i,t+h} \,\big|\, \epsilon_{j,t} = 0\ \ \forall k,\ \mathcal{F}_{t-1}\big]\,,$$

where $\mathcal{F}_{t-1}$ denotes the information set available up to time $t-1$, any nonzero cross-cumulant would imply joint movements of shocks, thereby contradicting this definition.

Assumption 2 *Partial Identification:* ***PI***

(i) The third- and fourth-order cross-cumulants of the structural shocks ϵ_t are zero.

(ii) Let a subset of structural shocks of dimension $r < n-1$ exhibit skewness and/or excess kurtosis. Under the rank condition in Guay (2021, Corollary 1), the subsystem corresponding to this subset is identified (up to sign and permutation).

In conclusion, the entire structural system is statistically identified up to changes in sign and column permutations, provided that all but one of the structural shocks exhibit nonzero skewness and/or excess kurtosis. Additionally, the subsystem relating all reduced-form innovations to asymmetric and/or non-mesokurtic structural shocks is statistically identified, given the information contained in the second-, third-, and/or fourth-order cumulants. Finally, we introduce the following additional assumption:

Assumption 3 *The third- and/or fourth-order cumulants of the structural shocks that exhibit skewness and/or excess kurtosis are distinct.*

This assumption is analogous to the condition for eigenvalue decomposition, where the eigenvectors are uniquely determined if and only if all eigenvalues are distinct. It also resembles the condition in Gouriéroux et al. (2017), which requires distinct and asymmetric pseudo-distributions, as well as true distributions.

3. Cumulants as Tensors

This section introduces higher-order tensors as natural generalizations of vectors and matrices to multi-dimensional spaces. Many higher-order statistics—such as cumulants—can be expressed compactly in tensor form, making tensor decompositions particularly well suited for their analysis. Since these statistics form the basis of our identification strategy, tensor methods provide a direct and efficient framework for developing the estimation procedure of structural shocks.

A tensor is a multidimensional array; more formally, an Nth-order tensor is an element of the tensor product of N vector spaces.[11] The *order* of a tensor refers to the number of dimensions (also called modes or ways). An Nth-order tensor (or N-way tensor) is defined as

$$\mathcal{A} = \{a_{i_1 i_2 \cdots i_N}\} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \quad \text{or} \quad \mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N},$$

[11] See Appendix B for a formal definition of tensors and related concepts.

where $a_{i_1 i_2 \cdots i_N}$ are the individual entries. Vectors and matrices correspond to tensors of order one and two, respectively, while tensors of order three or higher are referred to as higher-order tensors. Throughout, we adopt the notation of Kolda and Bader (2009).

A tensor is said to be *supersymmetric* (or simply *symmetric*) if its entries remain invariant under any permutation of their indices. Since the value of a cumulant does not depend on the order of the variables involved, cumulants are therefore symmetric tensors, as defined in equations (3) and (4).

The *unfolding* process reorders the elements of an Nth-order tensor into a matrix, a procedure also referred to as *matricization*. The mode-n unfolding of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is denoted by $\mathbf{A}_{(n)}$ and arranges the mode-n fibers into the columns of the resulting matrix. For a symmetric tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the mode-$n$ unfoldings coincide, i.e., $\mathbf{A}_{(1)} = \mathbf{A}_{(2)} = \cdots = \mathbf{A}_{(N)} = \mathbf{A}$. Hence, the matrix representations in (5) and (6) correspond to the mode-i unfolding of the respective symmetric tensor for $i = 1, \ldots, I$, with $I = 3$ or $I = 4$ for the third- and fourth-order cumulants, respectively.

A tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is said to be *diagonal* if $a_{i_1 i_2 \cdots i_N} \neq 0$ only when $i_1 = i_2 = \cdots = i_N$. Under the identification conditions stated in **Assumption 1**(ii) or **Assumption 2**(i), the third- and fourth-order cumulants of the structural shocks are diagonal tensors. This diagonality property is a key feature exploited in the decomposition proposed below.

Our identification and estimation strategy for structural shocks relies on explicitly imposing orthogonality constraints. Following Chen and Saad (2008)[12], this can be achieved through the *tensor singular value decomposition*, which provides an optimal low-rank orthogonal approximation of a tensor. Formally, we now define the tensor SVD for an Nth-order symmetric tensor: $\mathcal{A} = \{a_{i_1 i_2 \cdots i_N}\} \in \mathbb{R}^{n \times n \times \cdots \times n}$.

Definition 1 (Tensor SVD for Symmetric Tensors) *An Nth-order symmetric tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times \cdots \times n}$ admits a tensor singular value decomposition if it can be written as*

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i \, \mathbf{u}_i \circ \mathbf{u}_i \circ \cdots \circ \mathbf{u}_i, \tag{7}$$

where $\lambda_i \in \mathbb{R}$ are singular values and $\mathbf{u}_i \in \mathbb{R}^n$ are singular vectors.

Let

$$\mathbf{U} = \left[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\right] \in \mathbb{R}^{n \times r},$$

with the constraint that $\mathbf{U}'\mathbf{U} = I_r$, so that the columns of $\mathbf{U}$ are orthonormal. Then (7) is called the symmetric tensor singular value decomposition *(symmetric TSVD) of $\mathcal{A}$.*

[12] See also Comon (2002).

Equivalently, in matrix form this can be expressed as

$$\mathbf{A} = \mathbf{U}\,\mathrm{diag}(\lambda_i)\,(\mathbf{U} \odot \cdots \odot \mathbf{U})', \tag{8}$$

where $\odot$ *denotes the Khatri–Rao product of* $N-1$ *terms and* $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$.

In the matrix case, the decomposition reduces to the familiar singular value decomposition (SVD). The tensor SVD generalizes this concept by representing a tensor as a sum of rank-one components, each formed by the outer product of vectors $\mathbf{u}_i$, for $i = 1, \ldots, r$. The smallest integer r for which such a representation holds is called the *rank* of the tensor $\mathcal{A}$. In this setting, the associated core tensor $\mathcal{S}$ is diagonal, with entries $s_{i,i,\ldots,i} = \lambda_i$ for $i = 1, \ldots, r$, and all off-diagonal elements equal to zero.[13] The next proposition establishes an explicit expression for each singular value λ_i, linking it directly to the tensor $\mathcal{A}$ and its corresponding singular vector u_i.

Proposition 2 *Let* $\mathcal{A} \in \mathbb{R}^{n \times \cdots \times n}$ *be an* N*th–order symmetric tensor admitting the symmetric TSVD* (7)*. Then, for each* $i = 1, \ldots, r$*,*

$$\lambda_i \;=\; \langle \mathcal{A},\; u_i \circ u_i \circ \cdots \circ u_i \rangle \;=\; u_i' \mathbf{A}\,(u_i \otimes u_i \otimes \cdots \otimes u_i),$$

where the last expression involves $(N-1)$ *copies of* u_i *in the Kronecker product. Here,* $\langle \cdot, \cdot \rangle$ *denotes the Frobenius inner product, and* $\mathbf{A}$ *the mode-n unfolding of* $\mathcal{A}$*.*

The proof of the proposition is given in Appendix C. The diagonality property makes the tensor SVD particularly well suited to our framework, since under the identification assumptions the higher-order cumulants of structural shocks are diagonal tensors.

When the rank is smaller than the tensor dimension—as in the partial identification case—a best rank-r approximation must be determined. For matrices, the truncated SVD delivers the optimal low-rank approximation in the least-squares sense: according to the Schmidt–Eckart–Young theorem, this is achieved by retaining the first r terms of the singular value decomposition. This result, however, does not extend to higher-order tensors. In the tensor setting, the best rank-r approximation cannot be obtained sequentially, but requires estimating all factor matrices simultaneously (see Kolda and Bader (2009)).

The tensor SVD decomposition parallels principal component analysis (PCA), which identifies orthogonal directions that maximize variance, a second-order cumulant. When applied to higher-order cumulants, the tensor SVD extracts orthogonal directions that maximize these higher-order statistics, making PCA simply the special case where the focus is on variance.

[13] See Appendix B for formal definitions of rank-one tensors and core tensors.

The tensor SVD decomposition is also closely related to the CANDECOMP/PARAFAC (CP) decomposition introduced in Appendix B. A key distinction is that the CP decomposition does not impose orthogonality constraints on the factor matrices $\mathbf{U}$. Indeed, uniqueness of the CP decomposition can be achieved under conditions that are far weaker than orthogonality among singular vectors (see Kolda and Bader (2009) for details).[14] In general, for tensors of order greater than two, a representation based solely on orthogonal vectors cannot achieve full diagonalization.

The absence of exact diagonalization motivates the search for an alternative notion of diagonal structure. This leads to the concept of *maximal diagonality*, which formalizes the idea of approximating a tensor by a decomposition that maximizes the weight on its diagonal elements. The concept is formalized in the following proposition.

Proposition 3 (Maximal Diagonality and Tensor Decomposition) *Let $\mathcal{A} \in \mathbb{R}^{n\times\cdots\times n}$ be a symmetric tensor of order N. Since higher-order tensors cannot, in general, be diagonalized by unitary transformations, a solution is obtained by seeking the decomposition that minimizes the Frobenius distance to $\mathcal{A}$ while enforcing orthogonality:*

$$\min_{\substack{\lambda_1,\ldots,\lambda_r \\ \mathbf{u}_1,\ldots,\mathbf{u}_r}} \left\| \mathcal{A} - \sum_{i=1}^{r} \lambda_i \, \mathbf{u}_i \circ \mathbf{u}_i \circ \cdots \circ \mathbf{u}_i \right\|_F^2 \quad \textit{subject to} \quad \mathbf{U}'\mathbf{U} = I_r, \tag{9}$$

where $\mathbf{U} = \left[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\right]$ and $\|\cdot\|_F$ denotes the Frobenius norm.

This minimization problem is equivalent to

$$\max_{\mathbf{U}'\mathbf{U}=I_r} \sum_{i=1}^{r} \lambda_i^2,$$

that is, maximizing the squared norm of the diagonal of the core tensor subject to orthogonality.

Hence, the solution provides the decomposition of $\mathcal{A}$ with maximal diagonality. The equivalence between the minimization problem (9) and the maximal diagonality formulation is formally established in Chen and Saad (2008). The solution to (9) maximizes the diagonal contribution of the core tensor $\mathcal{S}$.[15] In our setting, this equivalence substantially reduces computational complexity: from $O(n^J)$ to n diagonal cumulants of order J under complete identification, and from $O(r^J)$ to r diagonal cumulants under partial identification, where r is the number of non-Gaussian structural shocks. This reduction is particularly valuable in moderate- to high-dimensional systems, where GMM estimators targeting cross-cumulants are computationally intensive.

[14] Uniqueness is defined up to scaling and permutation of the components.

[15] See Comon (1994) for a related argument in the context of Independent Component Analysis (ICA) and cumulants.

4. Tensor SVD for Non-Gaussian SVAR

In this section, we present a two-step strategy for identifying structural shocks in a non-Gaussian VAR using the tensor SVD decomposition. The first step consists in prewhitening the data so that the identification problem reduces to determining an orthogonal matrix. The second step then focuses on estimating this orthogonal matrix.

4.1 A Two-Step Procedure: Prewhitening as the First Step

The first step consists of applying an orthogonal transformation to the impact matrix Θ:

$$\nu_t = \Theta\epsilon_t = \tilde{\Theta}Q\epsilon_t = \tilde{\Theta}u_t,$$

where $u_t = Q\epsilon_t$, Q is an orthogonal matrix satisfying $QQ' = I$, and $\Sigma_\nu = \tilde{\Theta}\tilde{\Theta}'$. The matrix $\tilde{\Theta}$ provides a square-root decomposition of Σ_ν, which ensures that the transformed shocks are uncorrelated with unit variance. Any valid square-root decomposition may be used for prewhitening; common choices include the Cholesky and singular value decompositions.

Using only the covariance matrix, the orthogonal matrix Q remains unidentified. Its recovery requires exploiting higher-order cumulants of the prewhitened shocks u_t. By multilinearity of cumulants, the third- and fourth-order cumulants of u_t can be expressed in terms of those of the structural shocks as

$$\mathcal{C}^3_{i,j,k}(u) = \sum_{m=1}^{n} q_{im}q_{jm}q_{km}\,\mathcal{C}^3_{m,m,m}(\epsilon), \quad \mathcal{C}^4_{i,j,k,l}(u) = \sum_{m=1}^{n} q_{im}q_{jm}q_{km}q_{lm}\,\mathcal{C}^4_{m,m,m,m}(\epsilon),$$

where q_{ij} are the entries of Q, while $\mathcal{C}^3_{m,m,m}(\epsilon)$ and $\mathcal{C}^4_{m,m,m,m}(\epsilon)$ denote the corresponding diagonal cumulants of the structural shocks.

In compact matrix form:

$$\mathbf{C}^3_u = Q\,\mathbf{C}^3_\epsilon\,(Q\otimes Q)', \qquad \mathbf{C}^4_u = Q\,\mathbf{C}^4_\epsilon\,(Q\otimes Q\otimes Q)',$$

which, using $Q'Q = I$, yield the inverses

$$\mathbf{C}^3_\epsilon = Q'\mathbf{C}^3_u(Q\otimes Q), \qquad \mathbf{C}^4_\epsilon = Q'\mathbf{C}^4_u(Q\otimes Q\otimes Q).$$

4.2 A Two-Step Procedure: Estimation as a Second Step

The second step recovers the orthogonal matrix Q (or a subset of its columns) using the tensor SVD decomposition introduced in Section 3.

Consider the third-order cumulant tensor $\mathcal{C}^3(u)$ and $r = n$. Building on Proposition 3, the associated tensor SVD optimization problem is

$$\min_{\substack{\lambda_1,\ldots,\lambda_r \\ q_1,\ldots,q_r}} \left\| \mathcal{C}^3(u) - \sum_{i=1}^{r} \lambda_i \, q_i \circ q_i \circ q_i \right\|_F^2 \quad \text{subject to} \quad Q'Q = I_r,$$

where $Q = [q_1, \ldots, q_r]$.

Using vectorization, this problem can equivalently be written as

$$\min_{\substack{\lambda_1,\ldots,\lambda_r \\ Q'Q=I_r}} \left\| \operatorname{vec}(\mathbf{C}_u^3) - (Q \otimes Q \otimes Q) \operatorname{vec}(\mathbf{C}_\epsilon^3) \right\|_2^2,$$

where λ_i, for $i = 1, \ldots, r$, are the nonzero elements of $\mathbf{C}_\epsilon^3$.

As shown in Proposition 3, this minimization problem is equivalent to the maximization

$$\max_{Q'Q=I_r} \sum_{i=1}^{r} \lambda_i^2 = \max_{Q'Q=I_r} \sum_{i=1}^{r} \left(\mathcal{C}_{i,i,i}^3(\epsilon) \right)^2.$$

For the third-order cumulants, Proposition 2 implies that

$$\begin{aligned} q_i' \, \mathbf{C}_u^3 \, (q_i \otimes q_i) &= \lambda_i, \qquad && i = 1, \ldots, r, \\ q_i' \, \mathbf{C}_u^3 \, (q_j \otimes q_k) &= 0, && \text{for all } i \neq j, k. \end{aligned}$$

In the case of partial identification, where $r < n$, the orthogonal matrix Q can be partitioned as

$$Q = \begin{bmatrix} Q_r & Q_{n-r} \end{bmatrix},$$

with $Q_r = \begin{bmatrix} q_1 & \cdots & q_r \end{bmatrix}$ containing the first r columns of Q. The estimation task reduces to selecting the r orthogonal directions—linear combinations of the reduced-form innovations—that maximize a chosen measure of non-Gaussianity, in this case, skewness. This procedure parallels principal component analysis (PCA): whereas PCA identifies orthogonal directions that maximize variance, here we identify orthogonal directions that maximize skewness.

Let $\widehat{\mathbf{C}}_{u,T}^3$ denote a $T^{1/2}$-consistent and asymptotically normally distributed estimator of the coskewness matrix $\mathbf{C}_u^3$, where the dependence on the first-stage VAR parameters is omitted for simplicity. The estimator $\widehat{Q}_{r,T}$ is then defined as the maximizer of the Lagrangian

$$L_T = \sum_{i=1}^{r} \lambda_{i,T}^2 - \sum_{j,k=1}^{r} \mu_{jk} \left(q_j' q_k - \delta_{jk} \right), \tag{10}$$

where $\lambda_{i,T} = q_i' \widehat{\mathbf{C}}_{u,T}^3 (q_i \otimes q_i)$, δ_{jk} is the Kronecker delta, and μ_{jk} are Lagrange multipliers enforcing

orthogonality.

The following theorem establishes the asymptotic distribution of the estimator $\widehat{Q}_r$ obtained from (10). An analogous result applies when estimation is based on the fourth-order cumulant tensor, corresponding to the optimization problem

$$\max_{Q_r'Q_r=I_r} \sum_{i=1}^{r} \left(\mathcal{C}^4_{i,i,i,i}(\epsilon)\right)^2 = \max_{Q_r'Q_r=I_r} \sum_{i=1}^{r} \lambda_i^2.$$

For simplicity of exposition, we consider the case where the entire system or a subsystem is identified for the two distinct cases of skewness and non-mesokurticity.

Theorem 4 *Suppose that Assumption* **CI** *holds in the case of complete identification, or Assumption* **PI** *holds in the case of partial identification, as well as Assumption 3.*

- *In the case where the entire set or a subset of structural shocks of dimension r exhibits skewness, and given the additional assumption that $E(\epsilon_t^6) < \infty$, let $\widehat{\mathbf{C}}^3_{u,T}$ be a $T^{1/2}$-consistent and asymptotically normally distributed estimator of the coskewness matrix $\mathbf{C}^3_u$. The asymptotic distribution of* $\mathrm{vec}(\widehat{Q}_{r,T})$ *is given by:*

$$T^{1/2}\left(\mathrm{vec}(\widehat{Q}_{r,T}) - \mathrm{vec}(Q_r)\right) \to \mathcal{N}\left(0, Avar\left(T^{1/2}\widehat{Q}_{r,T}\right)\right), \tag{11}$$

  where

$$Avar\left(T^{1/2}\widehat{Q}_{r,T}\right) = (I_r \otimes Q_r)\Xi\left((Q_r \odot Q_r) \otimes Q_r\right)' Avar\left(\widehat{\mathbf{C}}^3_{u,T}\right)\left((Q_r \odot Q_r) \otimes Q_r\right)\Xi'(I_r \otimes Q_r)',$$

  and Avar denotes the asymptotic variance, while $\to$ indicates weak convergence.

- *In the case where the entire set or a subset of structural shocks of dimension r features non-mesokurticity, let $\widehat{\mathbf{C}}^4_{u,T}$ be a $T^{1/2}$-consistent and asymptotically normally distributed estimator of the cokurtosis matrix $\mathbf{C}^4_u$. Given the additional assumption that $E(\epsilon_t^8) < \infty$, the asymptotic distribution of* $\mathrm{vec}(\widehat{Q}_{r,T})$ *is given by expression (11) and:*

$$\begin{aligned} Avar\left(T^{1/2}\widehat{Q}_{r,T}\right) \;=\; & (I_r \otimes Q_r)\Xi\left((Q_r \odot Q_r \odot Q_r) \otimes Q_r\right)' Avar\left(\widehat{\mathbf{C}}^4_{u,T}\right) \\ & \times \left((Q_r \odot Q_r \odot Q_r) \otimes Q_r\right)\Xi'(I_r \otimes Q_r)', \end{aligned}$$

  where Avar denotes the asymptotic variance.

The proof of the theorem is given in Appendix C.

In the case of complete identification (**CI**), $r = n$ or $r = n-1$ if only one structural shock is Gaussian. For partial identification (**PI**), $r < n-1$. As becomes clear from the derivation of the

asymptotic distribution, the precision of the estimator $\widehat{Q}_r$ depends on the distance between each $\hat{\lambda}_{i,T}$, similarly to the case of eigenvalues in an eigenvalue decomposition (EVD). It is also important to note that the asymptotic distribution of $\widehat{\mathbf{C}}^3_{u,T}$ and $\widehat{\mathbf{C}}^4_{u,T}$ depends on the asymptotic estimator of the reduced-form parameters associated with the first-stage VAR:

$$\Gamma = \left(\text{vec}(\Phi_0)', \ldots, \text{vec}(\Phi_p)', \text{vech}(\Sigma_\nu)'\right)',$$

where $\Sigma_\nu = \tilde{\Theta}\tilde{\Theta}'$.

The advantage of maximizing (10) is that it involves only the n diagonal cumulants, i.e. an $O(n)$ problem. By contrast, in the GMM framework one must account for all cross-cumulants associated with these n cumulants, resulting in a computational burden of order $O(n^3)$ or $O(n^4)$. For completeness, Appendix D presents the corresponding GMM estimator with the optimal weighting matrix. The relative performance of our estimator will be examined in the next section, where simulation results indicate that it consistently outperforms the GMM approach.

In practice, several cases, or combinations thereof, are of interest. The first two cases involve non-Gaussianity arising from skewness or from non-mesokurticity, as examined above. Suppose there are r_1 skewed and r_2 non-mesokurtic structural shocks, with $r_1 + r_2 = r$. The corresponding optimization problem is:

$$\max_{Q_r'Q_r = I_r} \left(\sum_{i=1}^{r_1} \lambda^2_{3,i} + \sum_{j=1}^{r_2} \lambda^2_{4,j} \right),$$

where $\lambda_{3,i} = \mathcal{C}^3_{i,i,i}(\epsilon)$ and $\lambda_{4,j} = \mathcal{C}^4_{j,j,j,j}(\epsilon)$, with $i \neq j$. It is also possible to introduce weights on third- and fourth-order cumulants, reflecting prior information or estimated precision.

Another relevant case arises when the r structural shocks exhibit both skewness and non-mesokurticity. This situation is empirically plausible, as economic and financial shocks often display multiple forms of non-Gaussianity rather than a single deviation from normality. In this scenario, the maximization problem becomes:

$$\max_{Q_r'Q_r = I_r} \sum_{i=1}^{r} \left(\lambda^2_{3,i} + \lambda^2_{4,i}\right) = \max_{Q_r'Q_r = I_r} \sum_{i=1}^{r} \left((\mathcal{C}^3_{i,i,i}(\epsilon))^2 + (\mathcal{C}^4_{i,i,i,i}(\epsilon))^2 \right).$$

Appendix C derives the asymptotic distributions for these cases.

In empirical applications, the relevant case of interest can be determined by performing the bootstrap testing procedure proposed in Guay (2021) on the third- and fourth-order cumulants. The details of the test are provided in Appendix E.

4.3 Practical Implementation

The estimation considerations above suggest a practical sequence of steps for implementing TSVD. The procedure can be summarized as follows:

1. **Reduced-Form Estimation:** Estimate the reduced-form VAR model in (2) using standard techniques and retrieve the residuals $\widehat{\nu}_t$.

2. **Diagnostic Testing:** Apply the bootstrap rank tests of Guay (2021) to determine the number of structural shocks that exhibit non-Gaussianity (i.e., skewness and/or excess kurtosis).

3. **Model Selection:** Based on the test outcomes, select the appropriate optimization problem, which may involve maximizing third-order cumulants, fourth-order cumulants, or a weighted combination thereof. This step also determines whether the system is fully or only partially identified.

4. **Impact Matrix Estimation:** Estimate the orthogonal matrix $\widehat{Q}_r$ by solving the TSVD optimization problem (10). The structural impact matrix is then given by $\widehat{\Theta} = \bar{\Theta}\widehat{Q}_r$, where $\bar{\Theta}$ is obtained from the prewhitening step (e.g., Cholesky decomposition or singular value decomposition).

5. **Inference:** Compute structural impulse response functions (IRFs) and construct confidence intervals using a bootstrap procedure. The bootstrap replicates steps 1 and 4 conditional on the model choice in step 3, thereby incorporating estimation uncertainty.

In the final step, we propose using a bootstrap procedure instead of relying on asymptotic approximations for two reasons. First, in VAR settings, standard or corrected bootstrap methods have been shown to deliver more accurate finite-sample approximations of uncertainty than their asymptotic counterparts (Kilian, 1999). Second, when higher moments are involved—particularly in the presence of excess kurtosis—bootstrap methods tend to be more reliable than asymptotic approximations.[16]

Alternative methods have been proposed in the ICA literature to address this estimation problem in specific cases. Deflation-based FastICA and Symmetric FastICA (Hyvärinen, 1999; Hyvärinen et al., 2001) maximize fourth-order moments in the presence of excess kurtosis, but rely on a different criterion. Jacobi rotation-based methods are also commonly employed in practice. The objective is to render a symmetric tensor as diagonal as possible through successive Jacobi

[16] See Bonhomme and Robin (2009) and Keweloh (2020) for arguments in favor of bootstrap methods in this context. Additional supporting evidence is provided by simulation experiments in testing settings, as shown in Kilian and Demiroglu (2000) and Guay (2021).

rotations (see, e.g., Comon (1994); Cardoso (1989); Li et al. (2018); Bonhomme and Robin (2009)). Other approaches include Riemannian optimization techniques (see, e.g., Usevich et al. (2020)).

5. Simulation Results

This section presents the results of a Monte Carlo study designed to evaluate the finite-sample performance of the proposed estimator. We compare its properties with those of alternative estimators in scenarios where the system is either completely identified or partially identified.

5.1 Complete Identification Case

In the first set of experiments, a two-dimensional orthogonal matrix is specified, which depends on a single parameter, as described in Gouriéroux et al. (2017). The vector u_t is a function of the structural shocks via an orthogonal matrix, such that

$$u_t = Q\epsilon_t,$$

where the orthogonal matrix is given by

$$Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix},$$

with $\theta = -\pi/5$. Since the matrix Q depends on one parameter, as in Gouriéroux et al. (2017), our analysis primarily focuses on the estimation of $q_{11} = \cos(\theta) = 0.809$.

We investigate different sample sizes T, specifically $T = 200, 500$, and 5000, with 10,000 simulated samples. To assess the robustness of the estimators, we consider different distributional setups, following those used in Gouriéroux et al. (2017). In **case (1)**, both structural shocks are independently generated from Student's t-distributions with five degrees of freedom. In **case (2)**, the first structural shock is drawn from a Student's t-distribution with seven degrees of freedom, while the second is drawn from a Student's t-distribution with twelve degrees of freedom. In **case (3)**, the first structural shock is generated from a Student's t-distribution with twelve degrees of freedom, and the second from a hyperbolic secant distribution.

In the second set of experiments, we investigate the performance of alternative estimators in the presence of skewness. In the first case, which exhibits weaker skewness, $1.6808 \times \epsilon_{1,t} \sim N(1,1)$ with probability 0.5 and $1.6808 \times \epsilon_{1,t} \sim N(-1, 2.65)$ with probability 0.5, resulting in $\epsilon_{1,t}$ having a skewness of -0.5231. The second structural shock is generated as $\epsilon_{2,t} \sim N(0,1)$. In second case, which features stronger skewness, $2.1755 \times \epsilon_{1,t} \sim N(1,1)$ with probability 0.7887 and $2.1755 \times \epsilon_{1,t} \sim N(-3.7326, 1)$ with probability 0.2113, resulting in $\epsilon_{1,t}$ having a skewness of -0.9907. The second

structural shock is again generated as $\epsilon_{2,t} \sim N(0,1)$, as in the preceding case.

For each experiment, we apply different approaches to estimate the orthogonal matrix. In cases of complete identification with excess kurtosis, we consider the pseudo maximum likelihood (PML) estimator proposed by Gouriéroux et al. (2017), the higher-order singular value decomposition (HOSVD),[17] the alternating least squares (ALS) estimator, the FastICA estimator of Hyvärinen (1999), and the joint approximate diagonalization of eigenmatrices (JADE) (Cardoso, 1989).[18] We include the ALS estimator in our comparison even though it does not enforce linear independence among the structural shocks. Finally, we compare these results with the GMM estimator described in Appendix D, using both the identity matrix and an optimal data-driven weighting matrix in the second step.[19] This comparison allows us to evaluate the performance of maximizing non-Gaussianity, as done in the TSVD estimator, relative to minimizing cokurtosis and/or coskewness, as in the GMM framework.

In the case of skewness, we also consider the PML estimator with a likelihood specified under excess kurtosis in order to evaluate the robustness of inference under a misspecified likelihood. As in the first set of experiments, we compare the HOSVD, ALS, FastICA and GMM estimators, along with the tensor SVD (TSVD) estimator proposed in this paper. The JADE estimator is excluded from this comparison because it is specifically designed to capture excess kurtosis. For HOSVD, FastICA, TSVD, and GMM, the identification criteria are based on third-order cumulants.

Table 1 presents the bias and root mean squared errors (RMSE) associated with the different estimators of $q_{11} = \cos(\theta) = 0.809$. As a first validation step, our implementation of the PML estimators exactly reproduces the results reported in Gouriéroux et al. (2017). In nearly all experiments, the TSVD estimator outperforms the other estimators, including in scenarios where the likelihood function is correctly specified for the PML estimator with $T = 200$. However, as expected, in well-specified settings with larger samples ($T = 500$ and $T = 5000$), the PML estimator slightly outperforms TSVD, reflecting its status as the asymptotically efficient maximum likelihood estimator. In some experiments, the TSVD estimator achieves substantial efficiency gains, reducing the RMSE by up to 50% relative to competing estimators.

[17] See Appendix B for a full description of HOSVD.

[18] The JADE algorithm (Cardoso, 1989) is a cumulant-based method that diagonalizes a set of fourth-order cumulant matrices using Jacobi rotations. We implemented JADE using the `JadeR.m` MATLAB file provided by Jean-François Cardoso (2013). For the FastICA toolbox, we used MATLAB codes developed by H. Gävert, J. Hurri, J. Särelä, and A. Hyvärinen, version 2.5 (2005). The ALS estimator is implemented using the Tensor Toolbox by Sandia National Labs, Version 3.1. The ALS estimator does not generally guarantee orthogonal vectors, but is nevertheless included for comparison.

[19] We also considered the GMM estimator proposed by Lanne and Luoto (2021), but its small-sample performance was consistently inferior—in terms of both bias and root mean squared error—compared to the GMM estimator presented in Appendix D.

The FastICA and JADE estimators generally perform worse than TSVD, but outperform HOSVD and ALS in most configurations—specifically in cases (1) and (2)—except when the sample size increases to $T = 5000$, where their relative performance deteriorates. Regarding the GMM estimators, the version using the identity weighting matrix (GMM-I) outperforms the two-step estimator in small samples, particularly in case (1) with $T = 200$. The optimally weighted GMM estimator (GMM-opt) yields improvements over GMM-I in terms of both bias and RMSE when the sample size increases to $T = 500$ and $T = 5000$. However, it remains outperformed by the TSVD estimator with respect to RMSE in all configurations for $T = 200$ and $T = 500$.

Table 2 presents the bias and RMSE associated with the different estimators of $q_{11} = \cos(\theta) = 0.809$ in the presence of skewness. For the HOSVD, TSVD, ALS, and FastICA estimators, the ranking remains consistent with the first set of experiments, with the TSVD estimator outperforming the others. Its superior performance is especially evident in scenarios with weaker skewness and smaller sample sizes ($T = 200$ and $T = 500$). Although performance differences tend to diminish in cases of strong skewness, the TSVD estimator still maintains a clear advantage.

As expected, the performance of the PML estimator deteriorates in this setting, due to its reliance on a likelihood function based on the Student's t-distribution. Consistent with results from the excess kurtosis experiments, the first-step GMM estimator (GMM-I) performs better than the two-step version in small samples. However, its relative performance declines when the sample size increases to $T = 500$ and $T = 5000$, at which point the two-step GMM performs comparatively better. Nonetheless, both GMM estimators are clearly dominated by the TSVD estimator in all scenarios, except in large samples where their performance becomes similar.

5.1.1 Impact of Dimensionality on TSVD vs. GMM

In this set of experiments, we examine how the small-sample performance of the TSVD and GMM estimators is affected by the number of terms optimized or moment conditions employed. Specifically, we consider a trivariate system in which the orthogonal mixing matrix Q is parameterized as the product of three elementary rotations:

$$Q(\theta) = Q_1(\theta) \times Q_2(\theta) \times Q_3(\theta),$$

where

$$Q_1(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q_2(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad Q_3(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

The rotation parameter is fixed at $\theta = -\pi/5$, yielding $q_{11} = \cos(\theta) = 0.809$. Although the

dimensionality of the parameter space remains constant, the TSVD criterion involves only three terms—one per variable—whereas the GMM estimator requires 15 moment conditions based on fourth-order cumulants. Structural shocks are independently drawn from Student's t-distributions with 5, 9, and 12 degrees of freedom, respectively, thereby inducing substantial excess kurtosis.

Table 3 reports the finite-sample properties of the TSVD and GMM estimators under this setting. The results clearly indicate that the TSVD estimator consistently outperforms both GMM variants in terms of bias and RMSE across all sample sizes. For $T = 200$, the RMSE of TSVD is approximately 45% lower than that of GMM-I and more than 50% lower than GMM-opt. As the sample size increases, TSVD retains its advantage, whereas the GMM estimators suffer from significant performance degradation due to the large number of moment conditions. This results in both higher bias and larger RMSE, even with the optimal weighting matrix in GMM-opt. Notably, even at $T = 5000$, the RMSE of the GMM estimators remains considerably higher than that of the TSVD estimator.

This evidence highlights a key theoretical insight: maximizing a reduced set of cumulants, as in TSVD, yields greater small-sample efficiency than minimizing a large collection of cross-cumulants, as in GMM. The curse of dimensionality inherent in the GMM approach amplifies estimation error, particularly in finite samples, whereas TSVD remains robust by focusing only on the most informative cumulants.

5.2 Partial Identification Case

In the partial identification setting, we consider a trivariate system with only one non-Gaussian structural shock. The orthogonal matrix is parameterized as:

$$Q(\theta_1, \theta_2) = Q_1(\theta_1) \times Q_2(\theta_2),$$

where

$$Q_1(\theta_1) = \begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) & 0 \\ -\sin(\theta_1) & \cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q_2(\theta_2) = \begin{bmatrix} \cos(\theta_2) & 0 & \sin(\theta_2) \\ 0 & 1 & 0 \\ -\sin(\theta_2) & 0 & \cos(\theta_2) \end{bmatrix}.$$

In this experiment, we assume that only the first structural shock exhibits non-Gaussianity (excess kurtosis), while the remaining two shocks are Gaussian. Two cases are considered. In case (1), the non-Gaussian shock is generated from a Student's t-distribution with five degrees of freedom. In case (2), which features weaker excess kurtosis, the non-Gaussian shock follows a Student's t-distribution with twelve degrees of freedom. The two Gaussian shocks are independently generated from standard normal distributions.

The evaluation focuses on the recovery of the first column of the orthogonal matrix Q, i.e., $(q_{11}, q_{21}, q_{31})^\top$. We assess the performance of HOSVD, ALS, JADE, FastICA, and TSVD estimators, each designed to extract the best rank-one orthogonal approximation.

The rotation parameters are fixed at $\theta_1 = -\pi/3$ and $\theta_2 = -\pi/6$. Table 4 reports empirical bias and RMSE over 10,000 Monte Carlo replications for both excess kurtosis scenarios. As expected, performance improves under stronger non-Gaussianity. Across all configurations and for all elements of the first column of Q, the TSVD estimator delivers the best results, consistently outperforming alternative approaches. ALS and HOSVD follow in terms of accuracy, whereas FastICA and JADE perform less satisfactorily. Although differences diminish with larger sample sizes, TSVD maintains a clear and persistent advantage.

6. Applications

We consider two applications to illustrate our approach. In the first, we revisit the estimation of macroeconomic effects of fiscal policy. The second application examines the relevance of financial shocks.

6.1 Effects of Fiscal Policy

As a first application, we reexamine the effects of fiscal policies on economic activity, following the seminal paper by Blanchard and Perotti (2002). This application is particularly well suited because non-Gaussianity enables the identification of only a subsystem, implying partial identification, as shown in Guay (2021).

Consider the following trivariate SVAR process:

$$\begin{pmatrix} \nu_{\tau,t} \\ \nu_{g,t} \\ \nu_{y,t} \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}, \tag{12}$$

where $\nu_{\tau,t}$, $\nu_{g,t}$, and $\nu_{y,t}$ represent the reduced-form innovations capturing the unanticipated movements in taxes, government spending, and output, respectively, whereas $\epsilon_{1,t}$, $\epsilon_{2,t}$, and $\epsilon_{3,t}$ correspond to the structural shocks.

The dataset encompasses quarterly U.S. data spanning from the first quarter of 1980 to the third quarter of 2015.[20] The variable *Output* represents the logarithm of real GDP per capita. *Taxes* are defined as the logarithm of total real government receipts, net of transfer payments per

[20] This starting date aligns with the selections made by Perotti (2004), Favero and Giavazzi (2009), and Bouakez et al. (2014).

capita, and *Government Spending* refers to the logarithm of the combined total of real government consumption and gross government investment expenditures per capita.

These series are adjusted for inflation using the GDP deflator and for population size using total population figures. Furthermore, the measurements for taxes and government spending encompass the general government sector, which includes federal (both defense and non-defense), state, and local government levels. Data are seasonally adjusted and are sourced from the National Income and Product Accounts (NIPA). The total population data are retrieved from the Federal Reserve Bank of St. Louis's FRED database. The reduced-form model, as shown in (2), incorporates a linear deterministic trend and eight lags. This configuration was selected as the most parsimonious structure that ensures all reduced-form residuals are serially uncorrelated.

The rank tests developed by Guay (2021), described in Appendix E, reveal no significant evidence of skewness, while one of the structural shocks exhibits statistically significant excess kurtosis. This result implies that the rank of the entire system is not identified, and an additional restriction must be imposed to achieve full identification. Nonetheless, the subsystem corresponding to the non-mesokurtic structural shock is identified—specifically, the column of the impact matrix Θ associated with this shock. Without loss of generality, we assume that this shock corresponds to the first structural shock $\epsilon_{1,t}$.

The first column of Table 5 presents the results for the TSVD estimator. The estimator for θ_{11} is precisely valued at 0.471. In contrast, the estimated values for θ_{21} and θ_{31} are very close to zero, indicating that they are not significantly different from zero. The values $\theta_{21} = 0$ and $\theta_{31} = 0$ suggest that, at impact, the structural shock $\epsilon_{1,t}$ affects only taxes. This shock can thus be economically interpreted as a tax shock, denoted $\epsilon_{\tau,t} = \epsilon_{1,t}$.[21]

In this application, the statistical properties of the subsystem linking reduced-form innovations to the structural shock with excess kurtosis permit the economic identification of the tax shock. It is important to emphasize, however, that local statistical identification does not in general guarantee that structural shocks admit a meaningful economic interpretation.

Although the subsystem related to the tax shock is locally identified, the subsystem linking the reduced-form innovations to the structural shocks $\epsilon_{2,t}$ and $\epsilon_{3,t}$ remains underidentified. To achieve identification of this subsystem, and to facilitate comparison with the partial estimator, one restriction must be imposed. Blanchard and Perotti (2002) propose two sets of identifying restrictions, both of which include the condition $\theta_{23} = 0$. Imposing $\theta_{23} = 0$ assumes that the

[21] As demonstrated in Guay (2021), these findings are corroborated by applying Jarque-Bera tests to the reduced-form innovations, with finite-sample critical values approximated through the bootstrap procedure of Kilian and Demiroglu (2000). The hypothesis of symmetry is not rejected for any of the reduced-form innovations, while the hypothesis of zero excess kurtosis is exclusively rejected for the reduced-form innovation associated with taxes, denoted $\nu_{\tau,t}$.

automatic and systematic government responses of government spending to changes in output are zero. We also explore the imposition of $\theta_{32} = 0$.

Accordingly, the estimation imposes either $\theta_{23} = 0$ or $\theta_{32} = 0$. Column 2 of Table 5 reports the results for the GMM estimator of Guay (2021), using the identity matrix as the weighting matrix under the restriction $\theta_{23} = 0$. The results are very similar when $\theta_{32} = 0$ is imposed. While this procedure is consistent, it is not efficient. Estimation of the optimal weighting matrix requires computing the covariance matrix of fourth-order unconditional moments, which are typically imprecisely estimated in small samples, as illustrated in our simulation experiments (see also Bonhomme and Robin 2009; Keweloh 2020). Column 3 of Table 5 presents the results obtained with the optimal weighting matrix. Although qualitatively similar to those based on the identity matrix, these estimates are less precise. In particular, it is not possible to reject the null hypothesis that excess kurtosis equals zero.

In contrast, the estimates obtained with the partial estimator are comparable to those achieved with the identity matrix in terms of precision. However, in all three estimation methods, the estimates of θ_{11} are numerically sizable and statistically significant, while the estimates of θ_{21} and θ_{31} are negligible and insignificant.

The results have an important implication: the subsystem linking all reduced-form innovations to the tax shock is effectively identified. Within this subsystem, we observe that the effectiveness of tax policy is minimal. Specifically, the dynamic response of the output following a tax shock is small but is not statistically significant. Furthermore, the tax multiplier - defined as the dollar change in production in quarter $t+i$ resulting from a dollar reduction in the exogenous component of taxes - is modest: it is zero at impact and reaches a peak of approximately 0.61 at 14 quarters (see the bottom part of Table 5). Importantly, these results can be obtained using the partial TSVD estimator without resorting to additional restrictions that are necessary to fully identify and estimate the entire system.

6.2 Effects of Credit Shocks

In the second application, we reevaluate the macroeconomic effects of credit shocks. Inspired by Boivin et al. (2020), we specify a VAR model with four monthly U.S. variables: the personal consumption expenditure price index (PCE) inflation rate, the unemployment rate (UR), the difference between the BAA corporate bond yields and 10-year Treasury bond yields (BSPREAD), and the federal funds rate (FFR). The BSPREAD is a commonly used proxy for the external finance premium of borrowing firms. The data span from January 1959 to December 2019. We exclude observations from the COVID-19 pandemic period to avoid outliers that could affect the

higher-order moments of the data (Lenza and Primiceri, 2022). Based on the Akaike Information Criterion (AIC), we use 9 lags.

We begin by testing for the number of non-Gaussian structural shocks using the rank test. The finite-sample properties of the test, reported in Appendix F, are examined using a data-generating process (DGP) calibrated to this application. Both the Wald and likelihood-ratio tests exhibit minimal size distortion while demonstrating strong power properties. Results for the Wald version of the test are presented in Table 6. When considering only skewness, the test suggests at least two non-Gaussian shocks, while an additional shock is detected when excess kurtosis is taken into account.

However, non-Gaussianity is primarily concentrated in the first two elements, as the first two eigenvalues of the third- and fourth-order cumulant matrices are dominant. Consequently, we focus on the first two non-Gaussian shocks in the context of partial identification. This approach also mitigates potential issues of weak identification. Since statistically identified shocks do not have a clear economic interpretation ex ante, we examine whether these shocks can be associated with a common non-Gaussian factor or with identification strategies previously adopted by Boivin et al. (2020). To this end, we compare them with shocks identified through recursive ordering. The first TSVD shock is highly correlated with the Cholesky innovation in the FFR equation, while the second non-Gaussian shock comoves with the credit shock identified by innovations in the credit spread equation. Overall, the shocks identified by non-Gaussianity appear to correspond closely to structural shocks obtained under Cholesky identification, and do not appear to reflect common non-Gaussian factors.

Figure 1 displays the dynamic responses to a credit shock that increases the credit spread. The black line with a shaded area represents impulse responses along with the 90% bootstrapped confidence intervals based on Boivin et al. (2020), which employs a recursive ordering to identify the credit shock associated with the BSPREAD equation. We then add impulse responses implied by the second TSVD shock. Following a negative credit supply shock, real activity and prices decline, while monetary policy responds by lowering the interest rate—a typical scenario of an adverse demand disturbance.

When shocks are identified using information contained in higher-order cumulants, the dynamic effects on the macroeconomy are qualitatively similar, as depicted by the red, blue, and green lines. These lines represent impulse responses obtained from our TSVD procedure, where we use both the third- and fourth-order cumulants, only skewness, and only kurtosis, respectively. However, some differences merit discussion. Recursive ordering implies that neither inflation nor the unemployment rate can respond on impact. In contrast, our procedure does not impose timing restrictions, leaving

all variables' impact responses unrestricted. Consequently, both inflation and real activity respond on impact, which aligns with DSGE models incorporating financial frictions (Christiano et al., 2014; Del Negro et al., 2015). In addition, we note that information from both asymmetry and excess kurtosis is relevant for identifying credit shocks.

These findings are robust to: considering other price or real activity measures like the CPI or the industrial production, using the 1-year constant maturity rate instead of FFR to avoid the zero-lower bound issue, or adding the volume of credit to the VAR. Thus, we have shown that higher-order cumulants provide valuable information for better identifying and estimating the macroeconomic effects of credit shocks.

7. Conclusion

The paper introduces a novel methodology for estimating non-Gaussian Structural Vector Autoregressive (SVAR) models based on Tensor Singular Value Decomposition (TSVD) applied to third- and fourth-order cumulants. A central contribution of this approach is its capacity to handle both complete and partial identification of structural shocks. This flexibility is important, as it enables researchers to apply the methodology across a wider variety of settings, including those in which full identification cannot be achieved. By doing so, the proposed framework expands the empirical relevance of non-Gaussian SVAR models and provides a unified tool for exploiting higher-order statistical information.

A promising avenue for future research is to develop frameworks that combine partial identification based on a subset of non-Gaussian structural shocks with more conventional identifying restrictions, such as short-run, long-run, and sign restrictions. Such hybrid approaches could substantially broaden the applicability of non-Gaussian SVARs. Future work could also focus on deriving robust extensions of the TSVD methodology to improve performance under outliers (see Davis and Ng, 2023) and weak identification. Since skewness or excess kurtosis may be driven by only a few extreme observations, future research should explore methods that explicitly downweight influential data points or employ robust cumulant-based criteria. These extensions would not only strengthen the reliability of inference in small samples, but also ensure that non-Gaussian SVAR models remain informative and stable in empirically challenging environments.

References

Anttonen, J., Lanne, M., and Luoto, J. (2023). Bayesian Inference on Fully and Partially Identified Structural Vector Autoregressions. University of Helsinki.

Babii, A., Ghysels, E., and Pan, J. (2024). Tensor principal component analysis. UNC Chapel Hill, manuscript.

Bekaert, G., Engstrom, E., and Ermolov, A. (2025). Uncertainty and the Economy: The Evolving Distributions of Aggregate Supply and Demand Shocks. *American Economic Journal: Macroeconomics*. Forthcoming.

Blanchard, O. and Perotti, R. (2002). An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics*, 117:1329–1368.

Blanchard, O. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79:655–673.

Boivin, J., Giannoni, M., and Stevanovic, D. (2020). Dynamic Effects of Credit Shock in Data-Rich Environment. *Journal of Business and Economic Statistics*, 38(2):272–284.

Bonhomme, S. and Robin, J. (2009). Consistent Noisy Independent Component Analysis. *Journal of Econometrics*, 149:12–26.

Bouakez, H., Chihi, F., and Normandin, M. (2014). Measuring the Effects of Fiscal Policy. *Journal of Economic Dynamics and Control*, 47:123–151.

Cardoso, J. (1989). Source Separation Using Higher Order Moments. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2109–2112.

Chen, J. and Saad, Y. (2008). On the Tensor SVD and Optimal Low Rank Orthogonal Approximation of Tensors. *SIAM Journal of Matrix Analysis and Applications*, 30:1709–1734.

Christiano, L. J., Motto, R., and Rostagno, M. (2014). Risk Shocks. *American Economic Review*, 104(1):27–65.

Comon, P. (1994). Independent Component Analysis, a New Concept? *Signal Process.*, 36:287–314.

Comon, P. (2002). Tensor Decompositions: State of the Art and Applications. In McWhirter, J. G. and Proudler, I. K., editors, *Mathematics in Signal Processing V*, pages 1–24. Clarendon Press, Oxford, UK. Lecture Notes in Mathematics; survey chapter.

Davis, R. and Ng, S. (2023). Time Series Estimation of the Dynamic Effects of Disaster-type Shocks. *Journal of Econometrics*, 235(1):180–201.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000a). A Multilinear Singular Value Decomposition. *SIAM Journal Matrix Analysis and Applications*, 21:1253–1278.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000b). On the Best Rank-1 and Rank-$(R^1, R^2, \ldots, R^N)$ Approximation of Higher-order Tensors. *SIAM Journal of Matrix Analysis and Applications*, 21:1324–1342.

Del Negro, M., Giannoni, M. P., and Schorfheide, F. (2015). Inflation in the Great Recession and New Keynesian Models. *American Economic Journal: Macroeconomics*, 7(1):168–96.

Favero, C. and Giavazzi, F. (2009). How Large Are the Effects of Tax Changes. NBER Working Paper No. 15303.

Fiorentini, G. and Sentana, E. (2023). Discrete Mixtures of Normals Pseudo Maximum Likelihood Estimators of Structural Vector Autoregressions. *Journal of Econometrics*, 235(2):643–665.

Golub, G. and Van Loan, C. (2013). *Matrix Computations*. The John Hopkins University, 4th edition.

Gouriéroux, C. and Jasiak, J. (2023). Generalized Covariance-Based Inference for Models Partially Identified from Independence Restrictions. manuscript.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical Inference for Independent Component Analysis: Application to Structural VAR Models. *Journal of Econometrics*, 196:111–126.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2020). Identification and Estimation in Non-Fundamental Structural VARMA Models. *The Review of Economic Studies*, 87(4):1915–1953.

Guay, A. (2021). Identification of Structural Vector Autoregressions through Higher Unconditional Moments, journal = Journal of Econometrics. 225:27–46.

Hafner, C. M., Herwartz, H., and Wang, S. (2025). Statistical Identification of Independent Shocks with Kernel-based Maximum Likelihood Estimation and an Application to the Global Crude Oil Market. *Journal of Business & Economic Statistics*, 43:423–438.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054.

Hyvärinen, A. (1999). Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10:626–634.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley Interscience.

Jondeau, E., Jurczenko, E., and Rockinger, M. (2018). Moment Component Analysis: An Illustration with International Stock Markets. *Journal of Business & Economic Statistics*, 36(3):567–598.

Keweloh, S. (2020). A Generalized Method of Moments Estimator for Structural Autoregressions Based on Higher Moments. *Journal of Business & Economic Statistics*, 39:772–782.

Kilian, L. (1999). Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses. *The Review of Economics and Statistics*, 81(4):652–660.

Kilian, L. and Demiroglu, U. (2000). Residual-Based Tests for Normality in Autoregressions: Asymptotic Theory and Simulation Evidence. *Journal of Business & Economic Statistics*, 18:40–50.

Kolda, T. and Bader, B. (2009). Tensor Decompositions and Applications. *SIAM Review*, 51:455–500.

Lanne, M., Keyan, L., and Luoto, J. (2023). Identifying Structural Vector Autoregressions Via Non-Gaussianity of Potentially Dependent Structural Shocks. Available at SSRN: https://ssrn.com/abstract=4564713.

Lanne, M. and Luoto, J. (2021). GMM Estimation of Non-Gaussian Structural Vector Autoregression. *Journal of Business & Economic Statistics*, 39:69–81.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and Estimation of Non-Gaussian Structural Vector Autoregressions. *Journal of Econometrics*, 196:288–304.

Lenza, M. and Primiceri, G. E. (2022). How to Estimate a Vector Autoregression after March 2020. *Journal of Applied Econometrics*, 37(4):688–699.

Lewis, D. J. (2021). Identifying Shocks via Time-Varying Volatility. *The Review of Economic Studies*, 88(6):3086–3124.

Li, J., Usevich, K., and Comon, P. (2018). Globally Convergent Jacobi-Type Algorithms for Simultaneous Orthogonal Symmetric Tensor Diagonalization. *SIAM Journal of Matrix Analysis and Applications*, 39(1):1–22.

Mesters, G. and Zwiernik, P. (2023). Non-Independent Components Analysis. manuscript.

Miettinen, J., Taskinen, S., Nordhausen, K., and Oja, H. (2015). Fourth Moments and Independent Component Analysis. *Statistical Science*, 30:372–390.

Montiel Olea, J. L., Plagborg-Møller, M., and Qian, E. (2022). SVAR Identification from Higher Moments: Has the Simultaneous Causality Problem Been Solved? *AEA Papers and Proceedings*, 112:481–485.

Newey, W. (1984). A Method of Moments Interpretation of Sequential Estimators. *Economics Letters*, 14(2-3):201–206.

Normandin, M. and Phaneuf, L. (2004). Monetary Policy Shocks: Testing Identification Conditions under Time-Varying Conditional Volatility. *Journal of Monetary Economics*, 51(6):1217–1243.

Perotti, R. (2004). Estimating the Effects of Fiscal Policy in OECD Countries. Università Bocconi, Discussion Paper No. 276.

Phillips, P. (1989). Partially Identified Models. *Econometric Theory*, 5:181–240.

Robin, J.-M. and Smith, R. J. (2000). Tests of Rank. *Econometric Theory*, 16:151–175.

Sentana, E. and Fiorentini, G. (2001). Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models. *Journal of Econometrics*, 102(1):143–164.

Sims, C. (1980). Macroeconomics and Reality. *Econometrica*, 48:1–48.

Tamer, E. (2010). Partial Identification Models. *The Annual Review of Economics*, 2:167–195.

Tucker, L. R. (1963). Implications of Factor Analysis of Three-Way Matrices for Measurement of Change. In Harris, C. W., editor, *Problems in Measuring Change*, pages 122–137. University of Wisconsin Press, Madison, WI.

Uhlig, H. (2005). What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. *Journal of Monetary Economics*, 52:381–419.

Usevich, K., Li, J., and Comon, P. (2020). Approximate Matrix and Tensor Diagonalization by Unitary Transformations: Convergence of Jacobi-Type Algorithms. *SIAM Journal of Optimization*, 40(4).

Yamada, I. and Ezaki, T. (2003). An Orthogonal Matrix Optimization by Dual Cayley Parametrization Technique. In *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 35–40, Nara, Japan. Tokyo Institute of Technology, Dept. of Communications and Integrated Systems.

Table 1: Finite-sample distributions in the complete identification case: Excess kurtosis

	T = 200			T = 500			T = 5000		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel (a) - Bias									
PML(1)	0.008	0.023	0.018	0.003	0.010	0.004	0.000	0.001	0.001
PML(2)	0.008	0.023	0.015	0.003	0.010	0.002	0.000	0.000	0.001
PML(3)	0.008	0.025	0.018	0.002	0.010	0.004	0.000	0.001	0.000
HOSVD	0.027	0.036	0.036	0.022	0.028	0.019	0.016	0.004	0.002
TSVD	-0.009	-0.032	-0.022	-0.001	-0.008	-0.006	0.001	0.001	0.001
ALS	0.015	0.027	0.054	0.007	0.010	0.029	0.001	0.002	0.002
FastICA	0.018	0.039	0.032	0.007	0.016	0.010	0.003	0.003	0.002
JADE	0.013	0.028	0.024	0.006	0.013	0.008	0.001	0.002	0.001
GMM-I	-0.014	0.019	0.001	-0.009	-0.008	-0.009	-0.002	0.001	0.001
GMM-opt	-0.005	0.030	0.009	-0.004	-0.002	-0.003	-0.001	0.001	0.001
Panel (b) - Root mean-squared errors									
PML(1)	0.083	0.143	0.124	0.042	0.089	0.067	0.012	0.022	0.018
PML(2)	0.083	0.145	0.122	0.043	0.089	0.070	0.012	0.021	0.020
PML(3)	0.105	0.161	0.123	0.061	0.102	0.064	0.017	0.025	0.016
HOSVD	0.156	0.182	0.168	0.140	0.156	0.130	0.117	0.057	0.037
TSVD	0.069	0.095	0.086	0.050	0.069	0.062	0.021	0.028	0.023
ALS	0.126	0.171	0.211	0.075	0.105	0.166	0.027	0.032	0.049
FastICA	0.115	0.182	0.153	0.071	0.113	0.091	0.049	0.049	0.046
JADE	0.104	0.160	0.134	0.063	0.102	0.083	0.021	0.028	0.023
GMM-I	0.109	0.193	0.153	0.075	0.110	0.092	0.030	0.034	0.029
GMM-opt	0.123	0.211	0.166	0.066	0.103	0.084	0.024	0.025	0.022

Note: The results reported are based on a Monte-Carlo exercise where we simulate $N = 10,000$ samples of i.i.d. random variables $\epsilon_{1,t}$ and $\epsilon_{2,t}$. Sample sizes are: $T = 200$, $T = 500$, and $T = 5000$. The first row of the table indicates the generating distributions of the ϵ_ts: (1) $\epsilon_{1,t} \sim t(5)$ and $\epsilon_{2,t} \sim t(5)$; (2) $\epsilon_{1,t} \sim t(7)$ and $\epsilon_{2,t} \sim t(12)$ and 3)$\epsilon_{1,t} \sim t(12)$ and $\epsilon_{2,t}$ is drawn from a hyperbolic secant distribution. Once the ϵ_ts are simulated, we compute $u_t = Q\epsilon_t$ where the entries of Q are: $q_{11} = \cos(\theta)$, $q_{21} = -\sin(\theta)$, $q_{12} = \sin(\theta)$, and $q_{22} = \cos(\theta)$ with $\theta = -\pi/5$ (so $q_{11} = 0.809$). PML(1), PML(2) and PML(3) indicate the sets of distribution used for the pseudo maximum likelihood where (1), (2) and (3) refer to the distributions specified above. Panel (a) reports the biases of the estimators (that is $E(\hat{q}_{11} - q_{11})$) and Panel (b) reports root-mean-squared errors (that is the square root of $E(\hat{q}_{11} - q_{1,1})^2$).

Table 2: Finite-sample distributions in complete identification case: Skewness

	$Skewness = -0.5231$						$Skewness = -0.9907$					
	T=200		T=500		T=5000		T=200		T=500		T=5000	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
PML	.082	.280	.089	.289	.084	.282	.067	.247	.059	.231	.017	.129
HOSVD	.028	.158	.007	.082	.000	.023	.003	.062	.001	.038	.000	.012
TSVD	.001	.112	.005	.072	.000	.021	.003	.047	.001	.029	.000	.009
ALS	.025	.144	.006	.078	.000	.023	.003	.061	.001	.038	.000	.012
FastICA	.049	.173	.019	.102	.003	.034	.016	.086	.008	.055	.001	.017
GMM-I	.042	.192	.019	.102	.000	.023	.005	.080	.001	.038	.000	.012
GMM-opt	.045	.185	.019	.102	.000	.023	.007	.076	.002	.038	.000	.012

Note: Entries are the empirical bias and root mean square errors (RMSE) for each experiment. For the weaker skewness the distributions are: $1.6808 \times \epsilon_{1,t} \sim N(1,1)$ with probability 0.5 and $1.6808 \times \epsilon_{1,t} \sim N(-1, 2.65)$ with probability 0.5 when $\epsilon_{1,t}$ exhibits a skewness of -0.5231 and $\epsilon_{2,t} \sim N(0,1)$. For the stronger skewcase $\epsilon_{2,t} \sim N(0,1)$ as well as $2.1755 \times \epsilon_{1,t} \sim N(1,1)$ with probability 0.7887 and $2.1755 \times \epsilon_{1,t} \sim N(-3.7326, 1)$ with probability 0.2113 when $\epsilon_{1,t}$ exhibits a skewness of -0.9907. For each parametrization, 10,000 simulated samples of size T are generated. Once the ϵ_ts are simulated, we compute $u_t = Q\epsilon_t$ where the entries of Q are: $q_{11} = \cos(\theta)$, $q_{21} = -\sin(\theta)$, $q_{12} = \sin(\theta)$, and $q_{22} = \cos(\theta)$ with $\theta = -\pi/5$ (so $q_{11} = 0.809$). The Table reports the biases of the estimators ($E(\hat{q}_{11} - q_{11})$) and the root-mean-squared errors (the square root of $E(\hat{q}_{11} - q_{1,1})^2$).

Table 3: Finite-sample performance in the complete identification case: Trivariate system with excess kurtosis

	$T = 200$	$T = 500$	$T = 5000$
Panel (a) — Bias			
TSVD	0.015	0.004	0.000
GMM-I	0.066	0.040	0.009
GMM-opt	0.086	0.046	0.008
Panel (b) — Root Mean Squared Error			
TSVD	0.128	0.076	0.026
GMM-I	0.235	0.186	0.079
GMM-opt	0.265	0.188	0.074

Note: This table reports results for a trivariate system with structural shocks generated from independent Student's t-distributions: $\epsilon_{1,t} \sim t(5)$, $\epsilon_{2,t} \sim t(9)$, and $\epsilon_{3,t} \sim t(12)$. The mixing matrix Q is constructed from a sequence of rotation matrices and depends only on a single parameter $\theta = -\pi/5$, yielding $q_{11} = \cos(\theta) = 0.809$. Bias is defined as $\mathbb{E}[\hat{q}_{11} - q_{11}]$, and RMSE as $\sqrt{\mathbb{E}[(\hat{q}_{11} - q_{11})^2]}$. All results are based on 10,000 Monte Carlo replications.

Table 4: Finite-sample distributions in the partial identification case: Excess kurtosis

	T = 200			T = 500			T = 5000		
	Panel (a) - $t(5)$								
	q_{11}	q_{21}	q_{31}	q_{11}	q_{21}	q_{31}	q_{11}	q_{21}	q_{31}
	Bias								
HOSVD	0.006	0.134	0.088	0.004	0.019	0.013	0.001	0.001	0.001
TSVD	0.013	0.078	0.057	0.005	0.016	0.012	0.001	0.001	0.001
ALS	0.005	0.089	0.050	0.004	0.016	0.011	0.001	0.001	0.001
FastICA	-0.015	0.286	0.196	-0.010	0.127	0.085	-0.002	0.023	0.016
JADE	-0.065	0.752	0.493	-0.064	0.755	0.503	-0.070	0.744	0.499
	Root mean-squared errors								
HOSVD	0.188	0.399	0.328	0.119	0.129	0.136	0.040	0.030	0.039
TSVD	0.173	0.295	0.262	0.114	0.112	0.120	0.039	0.029	0.037
ALS	0.179	0.307	0.264	0.117	0.117	0.125	0.040	0.030	0.039
FastICA	0.212	0.574	0.484	0.151	0.377	0.315	0.062	0.160	0.137
JADE	0.296	0.946	0.757	0.294	0.948	0.769	0.301	0.943	0.758
	Panel (b) - $t(12)$								
	q_{11}	q_{21}	q_{31}	q_{11}	q_{21}	q_{31}	q_{11}	q_{21}	q_{31}
	Bias								
HOSVD	-0.031	0.499	0.333	-0.010	0.297	0.202	0.003	0.006	0.002
TSVD	0.010	0.271	0.205	0.011	0.158	0.114	0.003	0.006	0.002
ALS	-0.035	0.346	0.182	-0.013	0.191	0.100	0.002	0.006	0.002
FastICA	-0.052	0.607	0.416	-0.035	0.436	0.293	-0.011	0.106	0.068
JADE	-0.061	0.756	0.498	-0.063	0.742	0.488	-0.064	0.745	0.503
	Root mean-squared errors								
HOSVD	0.265	0.773	0.625	0.233	0.592	0.486	0.073	0.054	0.070
TSVD	0.236	0.577	0.481	0.206	0.430	0.363	0.072	0.053	0.069
ALS	0.256	0.611	0.486	0.217	0.448	0.361	0.072	0.054	0.070
FastICA	0.278	0.846	0.696	0.248	0.714	0.586	0.120	0.345	0.290
JADE	0.292	0.950	0.761	0.294	0.940	0.756	0.296	0.943	0.767

Note: The results reported are based on a Monte-Carlo exercise where we simulate $N = 10,000$ samples. Different sample lengths T are considered: $T = 200$ (left part of the table), $T = 500$ (middle part of the table), and $T = 5000$ (right part of the table).

Table 5: Parameter Estimates and Multipliers ($\theta_{23} = 0$)

	Parameter Estimates		
Parameter	TSVD	$W = I$	Efficient W
θ_{11}	0.0471^{***}	0.0474^{***}	0.0512^{***}
θ_{12}		0.0026	0.0056
θ_{13}		0.0089^{**}	0.0087
θ_{21}	0.0001	0.0001	-0.0003
θ_{22}		0.0068^{***}	0.0069^{***}
θ_{23}		$0.0000^{\dagger}$	$0.0000^{\dagger}$
θ_{31}	0.0003	-0.0001	0.0005
θ_{32}		0.0017^{***}	0.0020
θ_{33}		0.0048^{***}	0.0048^{***}
$\mathcal{C}_{1,1,1,1}(\epsilon)$	2.8259^{***}	2.7995^{***}	3.056
Quarter	Tax Multiplier		
1	0.00	0.00	0.00
4	0.05	0.05	0.05
8	0.24	0.24	0.24
Peak	0.61	0.61	0.61
	[14]	[14]	[14]

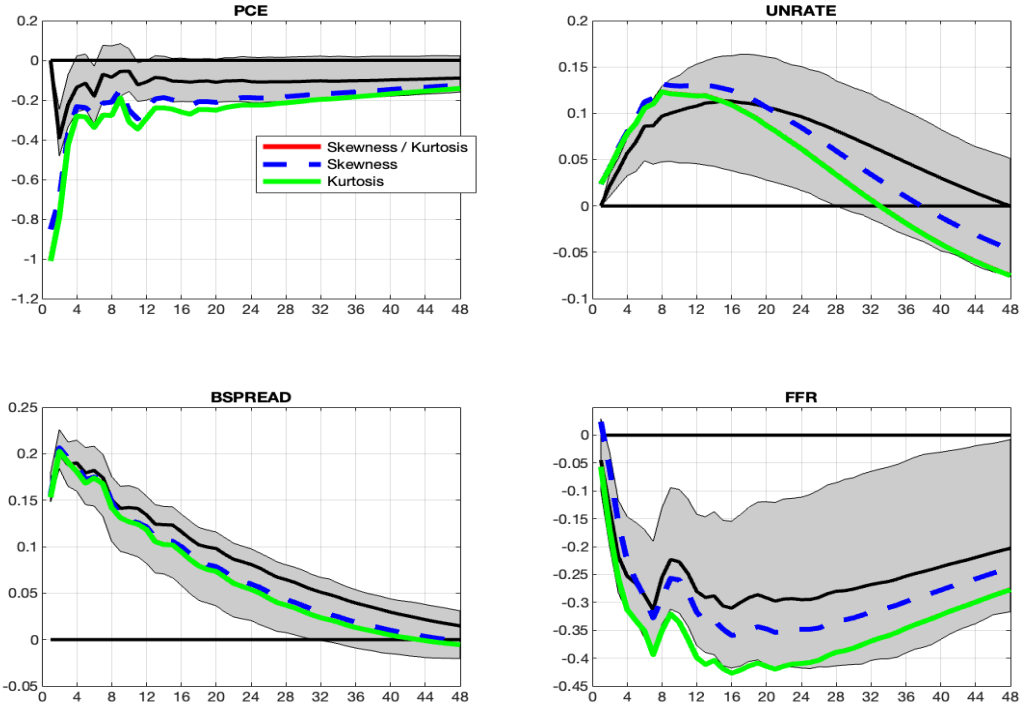
Notes. The tax multiplier measures the dollar change in output at a given horizon that results from a dollar decrease (increase) in the exogenous component of taxes. *, **, and *** indicate, respectively, that the 90, 95, and 99 percent confidence intervals do not include zero, where the confidence intervals are computed from 5,000 bootstrap samples. † indicates that the parameter is constrained. Numbers between brackets indicate the quarters in which the maximum value of the multiplier is reached.

Table 6: Testing for the number of non Gaussian structural shocks

		r=0	r=1	r=2	r=3
Skewness	Wald stat	2194.4	748.1	**88.3**	29.6
	CV (5%)	187.4	146.4	94.7	45.8
	CV (1%)	221.8	181.6	119.2	58.7
	Eigenvalues	[2.00 0.91 0.08 0.04]			
Kurtosis	Wald stat	849970	71210	6699.3	**1037.6**
	CV (5%)	1297.8	1486.5	1167.2	917.7
	CV (1%)	1665	1906.5	1512.6	1189
	Eigenvalues	[1078.6 89.4 7.8 1.4]			
Skewness /	Wald stat	852160	72073	6849.8	**1111.2**
Kurtosis	CV (5%)	1401.5	1648.6	1253.2	960.9
	CV (1%)	1787.5	2158.1	1593.6	1267.3
	Eigenvalues	[1080.5 90.3 7.9 1.5]			

Notes. This table reports rank test results for the second application. The first block allows for asymmetry only, the second for excess kurtosis and the third for both. Columns $r = 0$ to $r = 3$ indicate the number of non-Gaussian structural shocks under the null. The fourth line in each block reports the eigenvalues of the corresponding cumulant.

Figure 1: Dynamic responses to credit shocks



Note: The effects of credit shocks identified via Cholesky decomposition are displayed in black and with the corresponding 90% confidence intervals. Red, blue and green lines show impulse responses obtained from using 3th and 4th; only 3th; and only 4th moments respectively.

A Properties of the cumulants

For zero-mean real stochastic variables the cumulants up to order 4 are given by

$$\begin{aligned} Cum(X_1) &= E(X_1) \\ Cum(X_1, X_2) &= E(X_1 X_2) \\ Cum(X_1, X_2, X_3) &= E(X_1 X_2 X_3) \\ Cum(X_1, X_2, X_3, X_4) &= E(X_1 X_2 X_3 X_4) - E(X_1 X_2) E(X_3 X_4) \\ &\quad - E(X_1 X_3) E(X_2 X_4) - E(X_1 X_4) E(X_2 X_3) \end{aligned}$$

Cumulants have the following important properties:

- (Scaling) If $X_1, X_2, \ldots, X_N$ are multiplied with constants $a_1, a_2, \ldots, a_N$, then

$$Cum(a_1 X_1, a_2 X_2, \ldots, a_N X_N) = \prod_{i=1}^{N} a_i Cum(X_1, X_2, \ldots, X_N)$$

- (sum) Cumulants of a sum are the sum of the cumulants:

$$Cum(X_1 + Y_1, X_2, \ldots, X_N) = Cum(X_1, , X_2, \ldots, X_N) + Cum(Y_1, X_2, \ldots, X_N)$$

  where Y_1 is a real stochastic variable. This does not hold for moments and this explains the term cumulant.

- (Multilinearity) If a real stochastic vector X, with the components $X_1, X_2, \ldots, X_N$, is transformed into a stochastic vector Y by a real matrix multiplication $Y = AX$, with $A \in R^{J \times N}$, then we have for cumulants of order d

$$Cum_d(Y) = ACum_d(X) \left(A \otimes A \cdots \otimes A\right)'.$$

  in which $(A \otimes A \cdots \otimes A)$ is the Kronecker product of $d-1$ matrices A.

- (Symmetry) Cumulants are symmetric in their arguments, i.e.

$$Cum(X_1, X_2, \ldots, X_N) = Cum(X_{P(1)}, X_{P(2)}, \ldots, X_{P(N)}),$$

  in which P is an arbitrary permutation of $(1, 2, \ldots, N)$.

- (Independent variables) If stochastic variables $X_1, X_2, \ldots, X_N$ are independent, then we have

$$Cum(X_1, X_2, \ldots, X_N) = 0.$$

- (Gaussianity) If a stochastic variable is Gaussian, then we have

$$Cum_d(X) = 0;$$

  for $d > 2$. Higher-order cumulants of a Gaussian variable are zero.

- (Non-Gaussianity) There exists no distributions with a bound n such that $Cum_d(X) \neq 0$ for $3 \leq d \geq n$ and $Cum_d(X) = 0$ for $d > n$.

- (Gram-Charlier series expansion) For a standardized distribution $p_X(x)$ of a real random variable X, with mean $m_X = 0$ and variance $\sigma_X^2 = 1$, the Gram-Charlier series expansion is given by

$$p_X(x) = \hat{p}_X(x) \left\{ 1 + \frac{1}{3!} Cum_3(X) h_3(x) + \frac{1}{4!} Cum_4(X) h_4(x) + \frac{1}{5!} Cum_5(X) h_5(x) + \frac{1}{6!} [Cum_6(X) + 10(Cum_3(X))^2] h_6(x) + \ldots \right\},$$

  where $\hat{p}_X(x)$ is the probability density function of a standardized gaussian variable and $h_i(x)$ represents the ith Hermite polynomial.

B Higher-order Tensors

Higher-order tensors are gaining importance due to developments in the field of higher-order statistics (HOS), such as higher-order moments, cumulants, spectra, and cospectra. In particular, HOS are represented as symmetric higher-order tensors. A tensor is a multidimensional array. More formally, an Nth-order tensor is an element of the tensor product of N vector spaces. A first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three or higher are called higher-order tensors. Higher-order tensors generalize vectors and matrices to dimensions of order $N > 2$. Multilinear algebra is the algebra of higher-order tensors.

The *order* of a tensor is the number of dimensions, also known as ways or modes. A Nth-order tensors (also called a N-way tensor) is defined as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ or $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ for respectively real and complex values.

Definition 5 *The* inner product $\langle \mathcal{A}, \mathcal{B} \rangle$ *of two tensors* $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is defined as*

$$\langle \mathcal{A}, \mathcal{B} \rangle \stackrel{def}{=} \sum_{i_1}^{I_1} \sum_{i_2}^{I_2} \cdots \sum_{i_N}^{I_N} a_{i_1 i_2 \cdots i_N} b_{i_1 i_2 \cdots i_N}.$$

This is the sum of the multiplication of the elements with the same indice for $\mathcal{A}$ and $\mathcal{B}$.

Definition 6 *The Frobenius norm of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is defined as*

$$\|\mathcal{A}\| \stackrel{def}{=} \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}.$$

Definition 7 *The* outer product $\mathcal{A} \circ \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N \times J_1 \times J_2 \times \ldots \times I_M}$ *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *and a tensor* $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times I_M}$ *is defined by*

$$(\mathcal{A} \circ \mathcal{B})_{i_1 i_2 \cdots i_N j_1 j_2 \cdots j_M} \stackrel{def}{=} a_{i_1 i_2 \cdots i_N} b_{j_1 j_2 \cdots j_M}$$

for all indices.

A tensor is called *cubic* if every mode is the same size, i.e, $\mathcal{A} \in \mathbb{R}^{I \times I \times \ldots \times I}$. A cubical tensor is called supersymmetric (or symmetric) if its elements remain constant under any permutation of the indices.

The *unfolding* is the process of reordering the elements of a Nth-order tensor in a matrix (also called *Matricization*) The mode-n unfolding of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is denoted by $\mathbf{A}_{(n)}$ and arranges the mode-n fibers to the the columns of the resulting matrix. For a tensor $\mathcal{A}^{2 \times 2 \times 2}$,

$$\mathbf{A}_{(1)} = \begin{bmatrix} a_{111} & a_{121} & a_{112} & a_{122} \\ a_{211} & a_{221} & a_{212} & a_{222} \end{bmatrix}, \ \mathbf{A}_{(2)} = \begin{bmatrix} a_{111} & a_{211} & a_{112} & a_{212} \\ a_{121} & a_{221} & a_{122} & a_{222} \end{bmatrix}$$

and

$$\mathbf{A}_{(3)} = \begin{bmatrix} a_{111} & a_{211} & a_{121} & a_{221} \\ a_{112} & a_{212} & a_{122} & a_{222}. \end{bmatrix}.$$

For a symmetric tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, mode-n unfoldings of the tensor are all equal, i.e., $\mathbf{A}_{(1)} = \mathbf{A}_{(2)} = \cdots = \mathbf{A}_{(N)}$. The vectorization of a tensor is denoted $vec(\mathcal{A})$.

Definition 8 *The* n-mode product *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *by a matrix* $U \in \mathbb{R}^{J_n \times I_n}$ *is denoted by* $\mathcal{A} \times_n U$ *and is of size* $I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots I_N$. *This is defined as*

$$(\mathcal{A} \times_n U)_{i_1 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 \cdots i_{n-1} i_n i_{n+1} \cdots i_N} u_{j_n i_n}$$

for all index values.

This can also be expressed in terms of unfolded tensors:

$$\mathcal{X} = \mathcal{A} \times_n U \qquad \Leftrightarrow \qquad X_{(n)} = U \mathbf{A}_{(n)}.$$

For distinct modes in a series of multiplications, the order of the multiplication is irrelevant, i.e.,

$$\mathcal{A} \times_n U \times_m V = \mathcal{A} \times_m V \times_n U.$$

If the modes are the same, then

$$\mathcal{A} \times_n U \times_n V = \mathcal{A} \times_n (VU).$$

Consider $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and $U^{(n)} \in \mathbb{R}^{J_n \times I_n}$ for all $n \in 1, 2, \ldots, N$. We have the following equivalent between the *n-mode product* and the *Kronecker product* for any $n \in 1, 2, \ldots, N$,

$$\mathcal{Y} = \mathcal{A} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)} \Leftrightarrow$$
$$Y_{(n)} = U^{(n)} A_{(n)} \left(U^{(N)} \otimes \cdots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \cdots \otimes U^{(1)} \right)'.$$

The *Khatri-Rao product* is a columnwise Kronecker product. For matrices $A \in \mathbb{R}^{I \times K}$ and $B \in \mathbb{R}^{J \times K}$, their Khatri-Rao product is denoted by $A \odot B$ and the result is a matrix of size $(IJ) \times K$ given by

$$A \odot B = \begin{bmatrix} a_1 \otimes b_1 & a_2 \otimes b_2 & \cdots & a_K \otimes b_K \end{bmatrix}.$$

If a and b are vectors, the Khatri-Rao and the Kronecker products are identical $a \otimes b = a \odot b$.

Fibers are the higher-order analogue of matrix rows and columns. A fiber is defined by fixing every index but one. A matrix column is a mode-1 fiber and a matrix row is a mode-2 fiber. For matrix A with element (i, j) the mode-1 is $A(:, j)$ with Matlab colon notation and the mode-2 fiber is $A(j, :)$. For a third-order tensor, columns (mode-1 fiber), rows (mode-2 fiber) and tube (mode-3 fiber) are respectively, $\mathcal{A}(:, j, k)$, $\mathcal{A}(j, :, k)$ and $\mathcal{A}(j, k, :)$. Slices are two-dimensional sections of a tensor, defined by fixing all but two indices. For a third order tensor the slices are $\mathcal{A}(:, i, k)$ (the horizontal slice), $\mathcal{A}(:, j, ;)$ (the lateral slice) and $\mathcal{A}(j, :, :)$ (the frontal slice).

Definition 9 Rank-1 tensors *An Nth-order tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is* rank one *if it can be written as the outer product of* N *vectors, i.e.,*

$$\mathcal{A} = \mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(N)}$$

where the symbol "$\circ$*" represents the vector outer product.*

This means that each element of the tensor is the product of the corresponding vector elements. For a Nth-order tensor $\mathcal{A}$ and N vectors $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(N)}$, this implies that $a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \cdots u^{(N)}_{i_N}$ for all values of the indices. For a matrix, this is a rank-1 matrix which can be written as an outer product of two vectors (or equivalently by a singular value decomposition of rank-1).

Definition 10 The rank of a tensor *The rank of a tensor $\mathcal{A}$ denoted rank($\mathcal{A}$) is the minimal number of* rank-1 *tensors that yield $\mathcal{A}$ in a linear combination.*

We now provide a general definition of a tensor decomposition and introduce the two most common decompositions: the Tucker decomposition and the CP decomposition.

Definition 11 Tensor Decompositions *A decomposition of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is given by*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)},$$

where $\mathcal{S} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ is called the core tensor*, and $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times R_n}$ for $n = 1, \ldots, N$ are referred to as the* side matrices*. The operator $\times_n$ denotes the* mode-n product *of tensors.*[22]

Let $\mathbf{U}^{(n)} = \left[\mathbf{u}^{(n)}_1, \mathbf{u}^{(n)}_2, \ldots, \mathbf{u}^{(n)}_{R_n}\right]$ for all n. Then, the decomposition of $\mathcal{A}$ can equivalently be expressed as a sum of outer-product tensors:

$$\mathcal{A} = \sum_{r_N=1}^{R_N} \cdots \sum_{r_1=1}^{R_1} s_{r_1 r_2 \cdots r_N} \, \mathbf{u}^{(1)}_{r_1} \circ \mathbf{u}^{(2)}_{r_2} \circ \cdots \circ \mathbf{u}^{(N)}_{r_N}, \tag{13}$$

where $\circ$ denotes the outer product.

In particular, if $\mathcal{S}$ is diagonal, i.e., $s_{r_1 r_2 \cdots r_N} = 0$ except when $r_1 = r_2 = \cdots = r_N$, then

$$\mathcal{A} = \sum_{i=1}^{r} s_{ii\cdots i} \, \mathbf{u}^{(1)}_i \circ \mathbf{u}^{(2)}_i \circ \cdots \circ \mathbf{u}^{(N)}_i, \tag{14}$$

where $r = \min\{R_1, R_2, \ldots, R_N\}$.

This decomposition can also be expressed in matrix form. For a general tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, the mode-$n$ unfolding satisfies

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \mathbf{S}_{(n)} \left(\mathbf{U}^{(N)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \cdots \otimes \mathbf{U}^{(1)}\right)',$$

where $\mathbf{A}_{(n)}$ and $\mathbf{S}_{(n)}$ denote the mode-n unfoldings of $\mathcal{A}$ and the core tensor $\mathcal{S}$, respectively.

The Tucker decomposition (Tucker, 1963) is represented by the decomposition (13), where the factor matrices $\mathbf{U}^{(n)}$ are often referred to as the principal components in the respective mode-n. In this sense, the Tucker decomposition is a form of higher-order PCA. The *core tensor* $\mathcal{S}$ expresses the interactions between the elements of the different factor matrices $\mathbf{U}^{(n)}$ for $n = 1, \ldots, N$.

[22] The *mode-n product* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $U \in \mathbb{R}^{J_n \times I_n}$ is denoted by $\mathcal{A} \times_n U$. See Appendix B for a formal definition.

Any tensor can be written in Tucker form, and the unconstrained Tucker decomposition is not unique. Imposing the orthogonality of the factor matrices $\mathbf{U}^{(n)}$, such that $\mathbf{U}^{(n)}\mathbf{U}^{(n)'} = I$, implies that the *core tensor* $\mathcal{S}$ is given by:

$$\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}^{(\mathbf{1})'} \times_2 \mathbf{U}^{(\mathbf{2})'} \cdots \times_N \mathbf{U}^{(\mathbf{N})'},$$

or equivalently:

$$\mathbf{S}_{(n)} = \mathbf{U}^{(n)'}\mathbf{A}_{(n)} \left(\mathbf{U}^{(N)} \otimes \cdots \otimes \mathbf{U}^{(n+1)} \otimes \mathbf{U}^{(n-1)} \otimes \cdots \otimes \mathbf{U}^{(1)}\right).$$

For a symmetric tensor, $\mathbf{U}^{(\mathbf{n})} = \mathbf{U}$ for all n. In this decomposition, the *core tensor* is all-orthogonal, i.e., $\mathbf{S}_{(n)}\mathbf{S}'_{(n)} = \text{diag}(\lambda_{(n)})$ for all mode-n of the *core tensor* $\mathcal{S}$, where the vector $\lambda_{(n)}$ contains the singular values of $\mathbf{A}_{(n)}$.

Formally, the CANDECOMP/PARAFAC (CP) decomposition represents a tensor as a sum of rank-one components, each expressed as the outer product of vectors $\mathbf{u}_i^{(n)}$ for $i = 1, \ldots, r$ and mode n. These vectors can be scaled arbitrarily, provided that the product of their scalings remains unchanged, reflecting the inherent indeterminacy of the decomposition. This corresponds exactly to equation (14), in which the associated core tensor $\mathcal{S}$ is diagonal.

For a symmetric tensor $\mathcal{A}$, we have $\mathbf{u}_i^{(1)} = \mathbf{u}_i^{(2)} = \cdots = \mathbf{u}_i^{(N)}$, which implies that the CP decomposition simplifies to

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i \, \mathbf{u}_i \circ \mathbf{u}_i \circ \cdots \circ \mathbf{u}_i.$$

Equivalently, in matrix form this can be expressed as

$$\mathbf{A} = \mathbf{U} \, \text{diag}(\lambda_i) \, (\mathbf{U} \odot \cdots \odot \mathbf{U})', \tag{15}$$

where $\odot$ denotes the Khatri–Rao product of $N-1$ terms and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$. The smallest integer r for which this representation holds is called the *rank* of the tensor $\mathcal{A}$. In this case, the entries of the associated core tensor $\mathcal{S}$ satisfy $s_{i,i,\ldots,i} = \lambda_i$ for $i = 1, \ldots, r$, while all other elements are zero.

In the CP decomposition, no orthogonality constraints are imposed on the factor matrices $\mathbf{U}^{(n)}$, and uniqueness is not guaranteed. Instead, uniqueness can be achieved under conditions that are generally much weaker than orthogonality (see Kolda and Bader (2009) for a detailed discussion of these conditions).

The method to compute the Tucker decomposition with orthogonal matrices is better known as the *higher-order SVD* (HOSVD), also called multilinear SVD (MLSVD), or its truncated version for model reduction. De Lathauwer et al. (2000a) show that the HOSVD is a generalization of matrix SVD and that it always exists. They demonstrate how to compute the leading left singular vectors of $\mathbf{A}_{(n)}$. The HOSVD decomposition of a symmetric tensor is given by the following SVD:

$$\mathbf{A}_{(n)} = \mathbf{A} = \mathbf{U}\Lambda V',$$

where Λ is a diagonal matrix containing the singular values and V is the matrix of right singular vectors, for all $n = 1, 2, \ldots, N$, from the invariant mode-n unfolding.

Using the resulting left singular vectors $\mathbf{U}$, the HOSVD is given by:

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U} \times_2 \mathbf{U} \cdots \times_N \mathbf{U},$$

where $\mathcal{S} = \mathcal{A} \times_1 \mathbf{U}' \times_2 \mathbf{U}' \cdots \times_N \mathbf{U}'$, or equivalently:

$$\mathbf{S} = \mathbf{U}'\mathbf{A}\left(\mathbf{U} \otimes \cdots \otimes \mathbf{U} \otimes \mathbf{U} \otimes \cdots \otimes \mathbf{U}\right),$$

where $\otimes$ represents the Kronecker product.

The link between the higher-order singular value decomposition (HOSVD) and the eigenvalue decomposition (EVD), also known as the higher-order eigenvalue decomposition (HOEVD), is straightforward. For the unfolding matrix $\mathbf{A}$, we have:

$$\mathbf{A}\mathbf{A}' = \mathbf{U}D\mathbf{U}',$$

where $D = \Lambda^2$ is a diagonal matrix with the eigenvalues on its diagonal (see De Lathauwer et al. (2000a)).

However, the truncated HOSVD is not optimal in terms of providing the best least-squares fit but it serves as a good initialization for iterative algorithms such as alternating least squares (ALS). De Lathauwer et al. (2000b) proposed an efficient method for refining the factor matrices, known as the *higher-order orthogonal iteration* (HOOI), which can be viewed as a generalization of the *power method* for matrices (see Golub and Van Loan (2013), p. 454).

However, the truncated HOSVD is not optimal in terms of giving the best fit as measured as the least square but it is a good starting point for an iterative ALS (alternating least squares) algorithm. De Lathauwer et al. (2000b) proposed an efficient techniques for calculating the factor matrices called the *higher-order orthogonal iteration* (HOOI) which is an adaptation of the *power method* for matrices (see Golub and Van Loan (2013), p. 454).

Assuming that the rank r is known, several algorithms are available to compute a CP decomposition, with the alternating least squares (ALS) method being among the most widely used. The goal of ALS is to find an approximation that minimizes the quadratic loss function

$$\min_{\widehat{\mathcal{A}}} \ \|\mathcal{A} - \widehat{\mathcal{A}}\|_F^2, \tag{16}$$

where $\widehat{\mathcal{A}} = \widehat{\mathcal{S}} \times_1 \widehat{\mathbf{U}}^{(1)} \times_2 \widehat{\mathbf{U}}^{(2)} \cdots \times_N \widehat{\mathbf{U}}^{(N)}$, $\widehat{\mathcal{S}}$ is a diagonal core tensor, and $\|\cdot\|_F$ denotes the Frobenius norm.

C Proofs

Proof of Proposition 2

For the symmetric TSVD, the mode-n unfolding satisfies

$$\mathbf{A}_{(n)} \ = \ \mathbf{A} \ = \ U\,\mathrm{diag}(\lambda)\left(U \odot \cdots \odot U\right)',$$

where the Khatri–Rao product involves $N-1$ terms. Consider

$$u_i'\,\mathbf{A}_{(n)}\,(u_i \otimes \cdots \otimes u_i).$$

Substituting the decomposition gives

$$u_i' \mathbf{A} (u_i \otimes \cdots \otimes u_i) = u_i' U \operatorname{diag}(\lambda) (U \odot \cdots \odot U)'(u_i \otimes \cdots \otimes u_i).$$

Since $u_i' U = e_i'$, where e_i is the i-th canonical basis vector in $\mathbb{R}^r$, this reduces to

$$= e_i' \operatorname{diag}(\lambda) \big((U' u_i) \odot \cdots \odot (U' u_i)\big).$$

Because $U' u_i = e_i$, we obtain

$$(U' u_i) \odot \cdots \odot (U' u_i) = e_i,$$

so that

$$u_i' \mathbf{A} (u_i \otimes \cdots \otimes u_i) = e_i' \operatorname{diag}(\lambda) e_i = \lambda_i.$$

If $r < n$, then the matrix $U \in \mathbb{R}^{n \times r}$ has orthonormal columns, and the above argument still holds with $e_i \in \mathbb{R}^r$.

Proof of Theorem 4

The proofs for consistency and the asymptotic distribution are provided for $\mathcal{C}_\epsilon^3$; the derivation is similar for $\mathcal{C}_\epsilon^4$.

Consistency: As the criterion functions

$$\sum_{i=1}^r \lambda_{i,T}^2 = \sum_{i=1}^r \left(q_i' \widehat{\mathbf{C}}_{u,T}^3 (q_i \otimes q_i) \right)^2,$$

and

$$\sum_{i=1}^r \lambda_i^2 = \sum_{i=1}^r \left(q_i' \mathbf{C}_u^3 (q_i \otimes q_i) \right)^2,$$

are continuous, it follows that $\lambda_{i,T}^2 \xrightarrow{p} \lambda_i^2$ for all q_i, given that $\widehat{\mathbf{C}}_{u,T}^3 \xrightarrow{p} \mathbf{C}_u^3$. By the compactness of the set of orthogonal matrices $\mathcal{Q}_r$, we have

$$\sup_{Q_r \in \mathcal{Q}_r} \left| \sum_{i=1}^r \lambda_{i,T}^2 - \sum_{i=1}^r \lambda_i^2 \right| \xrightarrow{p} 0.$$

The sum $\sum_{i=1}^r \lambda_i^2$ attains its maximum at any permutation or change in the signs of the columns. This implies that there exists a sequence of maximizers such that $\widehat{Q}_{r,T} \xrightarrow{p} Q_r$ for all q_i, $i = 1, \ldots, r$, and $Q_r = [q_1 \cdots q_r]$.

Asymptotic distribution: The partial derivatives of the Lagrangian in finite samples are given by:

$$\frac{\partial L_T}{\partial q_i} : 2\widehat{\lambda}_{i,T} \widehat{v}_{i,T} - 2\widehat{\mu}_{ii} \widehat{q}_i - \sum_{j \neq i} \widehat{\mu}_{ij} \widehat{q}_j = 0, \tag{17}$$

for $i = 1, \ldots, r$, where

$$\widehat{v}_{i,T} := \frac{\partial \widehat{q}_i' \widehat{\mathbf{C}}^3_{u,T}(\widehat{q}_i \otimes \widehat{q}_i)}{\partial q_i} = 3\widehat{\mathbf{C}}^3_{u,T}(\widehat{q}_i \otimes \widehat{q}_i),$$

with $\widehat{q}_i'\widehat{v}_{i,T} = 3\widehat{\lambda}_{i,T}$, $\widehat{\lambda}_{i,T} = \widehat{q}_i'\widehat{\mathbf{C}}^3_{u,T}(\widehat{q}_i \otimes \widehat{q}_i)$, and

$$\begin{aligned} \widehat{q}_i'\widehat{q}_j &= 0, \quad \text{for } i < j, \\ \widehat{q}_i'\widehat{q}_i &= 1, \quad \text{for } i = 1, \ldots, r. \end{aligned}$$

Multiplying the first-order conditions (17) by $\widehat{q}_i'$ yields:

$$2\widehat{\lambda}_{i,T}\widehat{q}_i'\widehat{v}_{i,T} - 2\widehat{\mu}_{ii} = 0,$$

which implies:

$$6\widehat{\lambda}^2_{i,T} - 2\widehat{\mu}_{ii} = 0,$$

or $\widehat{\mu}_{ii} = 3\widehat{\lambda}^2_{i,T}$.

Therefore, the finite sample first-order conditions (FOC)

$$\widehat{\lambda}_{i,T}\widehat{q}_j'\widehat{v}_{i,T} = \widehat{\lambda}_{j,T}\widehat{q}_i'\widehat{v}_{j,T}$$

with $\widehat{q}_i'\widehat{q}_j = 0$ for $i < j$ and $\widehat{q}_i'\widehat{q}_i = i$ for $i = 1, \ldots, r$, provide the estimate of Q_r.

A first-order Taylor expansion around the true value Q_r yields

$$\begin{aligned} \widehat{\lambda}_{i,T}\widehat{q}_j'\widehat{v}_{i,T} - \widehat{\lambda}_{j,T}\widehat{q}_i'\widehat{v}_{j,T} &= \lambda_{i,T}q_j'\widehat{\mathbf{C}}^3_{u,T}(q_i \otimes q_i) - \lambda_{j,T}q_i'\widehat{\mathbf{C}}^3_{u,T}(q_j \otimes q_j) \\ &+ q_j'\widehat{\mathbf{C}}^3_{u,T}(q_i \otimes q_i)(\widehat{\lambda}_{i,T} - \lambda_{i,T}) - q_i'\widehat{\mathbf{C}}^3_{u,T}(q_j \otimes q_j)(\widehat{\lambda}_{j,T} - \lambda_{j,T}) \\ &+ \lambda_{i,T}(q_i \otimes q_i)'\widehat{\mathbf{C}}^{3'}_{u,T}(\widehat{q}_j - q_j) - \lambda_{j,T}(q_j \otimes q_j)'\widehat{\mathbf{C}}^{3'}_{u,T}(\widehat{q}_i - q_i) \\ &+ \lambda_{i,T}q_j'(\widehat{v}_{i,T} - v_{i,T}) - \lambda_{j,T}q_i'(\widehat{v}_{j,T} - v_{j,T}) + o_p(1). \end{aligned}$$

Using $\widehat{\mathbf{C}}^3_{u,T} = \mathbf{C}^3_u + (\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)$ and $(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u) = O_p(T^{-1/2})$, and noting that $\lambda_{i,T} = \lambda_i + (\lambda_{i,T} - \lambda_i) = \lambda_i + q_i'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_i \otimes q_i) = \lambda_i + O_p(T^{-1/2})$, we find that the first line is equal to

$$\lambda_i q_j'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_i \otimes q_i) - \lambda_j q_i'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_j \otimes q_j) + o_p(T^{-1/2})$$

by $q_j'\mathbf{C}^3_u(q_i \otimes q_i) = 0$ and $q_i'\mathbf{C}^3_u(q_j \otimes q_j) = 0$.

For the second line, since $\widehat{\lambda}_{k,T} - \lambda_{k,T} = 3(q_k \otimes q_k)'\widehat{\mathbf{C}}^{3'}_{u,T}(\widehat{q}_k - q_k) + o_p(1)$ for $k = i, j$ and $\widehat{q}_k = q_k + o_p(1)$, we obtain that $\widehat{\lambda}_{k,T} = \lambda_{k,T} + o_p(1)$ for $k = i, j$. By the same arguments as above for $\mathbf{C}^3_u$, the second line becomes $o_p(T^{-1/2})$.

For the third line,

$$\lambda_i(q_i \otimes q_i)'\mathbf{C}^{3'}_u(\widehat{q}_j - q_j) - \lambda_j(q_j \otimes q_j)'\mathbf{C}^{3'}_u(\widehat{q}_i - q_i) + o_p(T^{-1/2}),$$

and using $\mathbf{C}^3_u(q_i \otimes q_i) = \lambda_i q_i$ (similarly for j), we obtain

$$\lambda_i^2 q_i'(\widehat{q}_j - q_j) - \lambda_j^2 q_j'(\widehat{q}_i - q_i) + o_p(T^{-1/2}).$$

Finally, using $\widehat{v}_{i,T} = 3\widehat{\mathbf{C}}^3_{u,T}(\widehat{q}_i \otimes \widehat{q}_i)$, the last line is equal to

$$3\lambda_{i,T} q_j' \frac{\partial \widehat{\mathbf{C}}^3_{u,T}(q_i \otimes q_i)}{\partial q_i'}(\widehat{q}_i - q_i) - 3\lambda_{j,T} q_i' \frac{\partial \widehat{\mathbf{C}}^3_{u,T}(q_j \otimes q_j)}{\partial q_j'}(\widehat{q}_j - q_j).$$

Using $\widehat{\mathbf{C}}^3_{u,T} = \mathbf{C}^3_u + (\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)$ and $\widehat{q}_k = q_k + o_p(1)$, we get

$$3\lambda_i \frac{\partial q_j' \mathbf{C}^3_u(q_i \otimes q_i)}{\partial q_i'}(\widehat{q}_i - q_i) - 3\lambda_j \frac{\partial q_i' \mathbf{C}^3_u(q_j \otimes q_j)}{\partial q_j'}(\widehat{q}_j - q_j) + o_p(T^{-1/2}),$$

and given that $q_j' \mathbf{C}^3_u(q_i \otimes q_i) = 0$ for $i \neq j$, the left-hand side of the last line is then $o_p(T^{-1/2})$.

The first-order Taylor expansion around the true value Q gives

$$\lambda_i q_j'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_i \otimes q_i) - \lambda_j q_i'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_j \otimes q_j) + \lambda_i^2 q_i'(\widehat{q}_j - q_j) - \lambda_j^2 q_j'(\widehat{q}_i - q_i) + o_p(T^{-1/2}).$$

Since $\widehat{q}_j' \widehat{q}_i = 0$ for $i \neq j$, this implies

$$q_j'(\widehat{q}_i - q_i) + q_i'(\widehat{q}_j - q_j) + o_p(1) = 0.$$

We then have

$$q_i'(\widehat{q}_j - q_j) = \frac{\lambda_j q_i'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_j \otimes q_j) - \lambda_i q_j'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_i \otimes q_i)}{\lambda_i^2 + \lambda_j^2} + o_p(T^{-1/2}),$$

and similarly,

$$q_j'(\widehat{q}_i - q_i) = \frac{\lambda_i q_j'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_i \otimes q_i) - \lambda_j q_i'(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)(q_j \otimes q_j)}{\lambda_i^2 + \lambda_j^2} + o_p(T^{-1/2})$$

for all $i \neq j$.

The above equations can be represented in matrix form as

$$Q_r' \left(\widehat{Q}_{r,T} - Q_r\right) = -\Xi \left((Q_r \odot Q_r) \otimes Q_r\right)' \operatorname{vec}(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u) + o_p(T^{-1/2}),$$

and using $\operatorname{vec}\left(Q_r' \left(\widehat{Q}_{r,T} - Q_r\right)\right) = (I_r \otimes Q_r') \left(\operatorname{vec}(\widehat{Q}_{r,T}) - \operatorname{vec}(Q_r)\right)$, this implies

$$T^{1/2} \left(\operatorname{vec}(\widehat{Q}_{r,T}) - \operatorname{vec}(Q_r)\right) = (I_r \otimes Q_r)\, \Xi\, \left((Q_r \odot Q_r) \otimes Q_r\right)' T^{1/2} \left(\operatorname{vec}(\widehat{\mathbf{C}}^3_{u,T} - \mathbf{C}^3_u)\right) + o_p(1).$$

The matrix Ξ is defined as:

$$\Xi = (I_{r^2} - \mathcal{P}_{r,r})F,$$

where $\mathcal{P}_{r,r} = I_{r^2}(v_i, :)$ is a *mod-r perfect shuffle permutation* matrix. The vector $v_i = [(1 : r : n), (2 : r : n), \ldots, (r : r : n)]$ specifies the column where the "1" occurs in row i, with zeroes elsewhere, and

$n = r^2$ (see Golub and Van Loan, 2013, pp. 18–20). For example, with $r = 2$,

$$v_i = [1\ 3\ 2\ 4] \quad \text{and} \quad \mathcal{P}_{2,2} = I_4(v_i, :) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

For $r = 3$, we have $v_i = [1\ 4\ 7\ 2\ 5\ 8\ 3\ 6\ 9]$, and so on.

Next, we define the $r \times r$ matrix W:

$$W = \left[\frac{\lambda_i}{\lambda_i^2 + \lambda_j^2} \right]_{i,j=1,\ldots,r},$$

where each element (i, j) of the matrix W is defined as above. Let $w = \text{vec}(W')$, and matrix F is then the following $r^2 \times r^2$ diagonal matrix:

$$F = \text{diag}(w).$$

For $r = 2$, this matrix F is given by

$$F = \begin{bmatrix} \frac{\lambda_1}{\lambda_1^2+\lambda_1^2} & 0 & 0 & 0 \\ 0 & \frac{\lambda_1}{\lambda_1^2+\lambda_2^2} & 0 & 0 \\ 0 & 0 & \frac{\lambda_2}{\lambda_2^2+\lambda_1^2} & 0 \\ 0 & 0 & 0 & \frac{\lambda_2}{\lambda_2^2+\lambda_2^2} \end{bmatrix}.$$

For the case with non-mesokurtic structural shocks, the finite sample partial derivatives of the Lagrangian are:

$$\frac{\partial L_T}{\partial q_i} : 2\widehat{\lambda}_{i,T}\widehat{v}_{i,T} - 2\widehat{\mu}_{ii}\widehat{q}_i - \sum_{j \neq i} \widehat{\mu}_{ij}\widehat{q}_j = 0, \tag{18}$$

for $i = 1, \ldots, r$, where

$$\widehat{v}_{i,T} := \frac{\partial \widehat{q}_i' \widehat{\mathbf{C}}_{u,T}^4 (\widehat{q}_i \otimes \widehat{q}_i \otimes \widehat{q}_i)}{\partial q_i} = 4\widehat{\mathbf{C}}_{u,T}^3 (\widehat{q}_i \otimes \widehat{q}_i \otimes \widehat{q}_i),$$

with $\widehat{q}_i'\widehat{v}_{i,T} = 4\widehat{\lambda}_{i,T}$, where $\widehat{\lambda}_{i,T} = \widehat{q}_i'\widehat{\mathbf{C}}_u^4(\widehat{q}_i \otimes \widehat{q}_i \otimes \widehat{q}_i)$, and

$$\begin{aligned} \widehat{q}_i'\widehat{q}_j &= 0, \quad \text{for } i < j, \\ \widehat{q}_i'\widehat{q}_i &= 1, \quad \text{for } i = 1, \ldots, r. \end{aligned}$$

The remaining derivation follows similarly to the case with skewness.

Now, consider the case where there are r_1 structural shocks with skewness and r_2 that are

non-mesokurtic, with $r_1 + r_2 = r$. The optimization problem is then given by:

$$\max_{Q_r'Q_r = I_r} \left(\sum_{i=1}^{r_1} \lambda_{3,i}^2 + \sum_{j=1}^{r_2} \lambda_{4,j}^2 \right),$$

where $\lambda_{3,i} = \mathcal{C}_{i,i,i}^3(\epsilon)$ and $\lambda_{4,j} = \mathcal{C}_{j,j,j,j}^4(\epsilon)$, with $i \neq j$.

Let $Q_r = [Q_{r1} \quad Q_{r2}]$, where Q_{r1} consists of the first r_1 columns of the orthogonal matrix Q_r, and Q_{r2} contains the remaining columns. Define $\Psi(u) = \left[\mathbf{C}(u)^3 \quad \mathbf{C}(u)^4\right]$.

The finite sample partial derivatives of the respective Lagrangian are the same as in (17) for the first term of the objective function, i.e., $\lambda_{3,i}^2$, and as in (18) for the second term. A similar derivation as in the two preceding cases straightforwardly gives the following asymptotic distribution:

$$T^{1/2} \begin{bmatrix} \text{vec}(\widehat{Q}_{r1,T}) - \text{vec}(Q_{r1}) \\ \text{vec}(\widehat{Q}_{r2,T}) - \text{vec}(Q_{r2}) \end{bmatrix} = \begin{bmatrix} (I_{r1} \otimes Q_{r1})\Xi_1 \\ (I_{r2} \otimes Q_{r2})\Xi_2 \end{bmatrix} \begin{bmatrix} ((Q_{r1} \odot Q_{r1}) \otimes Q_{r1})' & 0 \\ 0 & ((Q_{r2} \odot Q_{r2} \odot Q_{r2}) \otimes Q_{r2})' \end{bmatrix} \times T^{1/2}(\text{vec}(\widehat{\Psi}_{u,T}) - \text{vec}(\Psi_u)) + o_p(1).$$

where Ξ_1 and Ξ_2 correspond to the matrix Ξ defined above, containing the terms related to the columns r_1 and r_2 of the orthogonal matrix Q_r. This asymptotic distribution therefore depends on the variance-covariance matrix of $\widehat{\Psi}_{u,T}$, which is also a function of the first-stage VAR reduced-form estimation.

Finally, for the case where the r structural shocks exhibit both skewness and non-mesokurticity, the maximization problem is given by:

$$\max_{Q_r'Q_r = I_r} \sum_{i=1}^{r} \left(\lambda_{3,i}^2 + \lambda_{4,i}^2\right) = \max_{Q_r'Q_r = I_r} \sum_{i=1}^{r} \left(\mathcal{C}_{i,i,i}^2(\epsilon) + \mathcal{C}_{i,i,i,i}^2(\epsilon)\right).$$

The finite sample partial derivatives of the Lagrangian are:

$$\frac{\partial L_T}{\partial q_i} : 2(\widehat{\lambda}_{3,i,T}\widehat{v}_{3,i,T} + \widehat{\lambda}_{4,i,T}\widehat{v}_{4,i,T}) - 2\widehat{\mu}_{ii}\widehat{q}_i - \sum_{j \neq i} \widehat{\mu}_{ij}\widehat{q}_j = 0,$$

for $i = 1, \ldots, r$, where

$$\begin{aligned} \widehat{v}_{3,i,T} &:= \frac{\partial \widehat{q}_i' \widehat{\mathbf{C}}_{u,T}^3 (\widehat{q}_i \otimes \widehat{q}_i)}{\partial q_i} = 3\widehat{\mathbf{C}}_{u,T}^3 (\widehat{q}_i \otimes \widehat{q}_i), \\ \widehat{v}_{4,i,T} &:= \frac{\partial \widehat{q}_i' \widehat{\mathbf{C}}_{u,T}^4 (\widehat{q}_i \otimes \widehat{q}_i \otimes \widehat{q}_i)}{\partial q_i} = 4\widehat{\mathbf{C}}_{u,T}^4 (\widehat{q}_i \otimes \widehat{q}_i \otimes \widehat{q}_i), \end{aligned}$$

with $\widehat{q}_i'\widehat{v}_{3,i,T} = 3\widehat{\lambda}_{3,i,T}$, $\widehat{q}_i'\widehat{v}_{4,i,T} = 4\widehat{\lambda}_{4,i,T}$, where $\widehat{\lambda}_{3,i,T} = \widehat{q}_i'\widehat{\mathbf{C}}_u^3(\widehat{q}_i \otimes \widehat{q}_i)$ and $\widehat{\lambda}_{4,i,T} = \widehat{q}_i'\widehat{\mathbf{C}}_u^4(\widehat{q}_i \otimes \widehat{q}_i \otimes \widehat{q}_i)$. We also have

$$\begin{aligned} \widehat{q}_i'\widehat{q}_j &= 0, \quad \text{for } i < j, \\ \widehat{q}_i'\widehat{q}_i &= 1, \quad i = 1, \ldots, r. \end{aligned}$$

Following similar derivations as above, we obtain the asymptotic distribution:

$$T^{1/2}\left(\text{vec}(\widehat{Q}_{r,T}) - \text{vec}(Q_r)\right) = (I \otimes Q_r)\Xi \begin{bmatrix} ((Q_r \odot Q_r) \otimes Q_r)' \\ ((Q_r \odot Q_r \odot Q_r) \otimes Q_r)' \end{bmatrix} \times T^{1/2}(\text{vec}(\widehat{\Psi}_{u,T}) - \text{vec}(\Psi_u)) + o_p(1).$$

D GMM Framework

The estimator of the structural parameters is based on GMM (Hansen (1982)). We need to embed the estimation procedure in a sequential procedure to derive the optimal GMM estimator of the structural parameters (see Newey (1984)). The estimation procedure depends first on the estimation of the reduced form (2). This estimation is performed by exploiting the following orthogonality conditions between the lagged values of the variables of interest and the statistical innovations,

$$Eg_1(Z_t, \Gamma) = \begin{bmatrix} E\left(X_{t-1} \otimes \nu_t\right) \\ E\left(vech(\Sigma_v) - vech(\nu_t \nu_t')\right) \end{bmatrix} = 0, \tag{19}$$

where $Z_t = \left(x_t', \ldots, x_{t-p}'\right)'$, $X_{t-1} = \left(x_{t-1}', \ldots, x_{t-p}'\right)'$, $\Gamma = (vec(\Phi_0)', \ldots, vec(\Phi_p)', vech(\Sigma_\nu)')'$ and $\Sigma_\nu = \tilde{\Theta}\tilde{\Theta}'$. The GMM estimator $\widehat{\Gamma}_T$ based on moment conditions (19) corresponds to the estimator of the reduced-form parameters associated with the VAR.

Next, the estimation procedure is based on estimating the structural form (1). Consider the vector of moment conditions that depend only on the third-order cross-cumulants of the statistical innovations. The moment conditions are given by

$$E[g_2(Z_t, \widehat{\Gamma}_T, \beta)] = E[g_2(u_t(\hat{\Gamma}), \beta)] = E[\text{vecht}(\mathbf{C}^3(u_t(\widehat{\Gamma}_T))) - \text{vecht}(Q\mathbf{C}_\epsilon^3(Q \otimes Q)')] = 0, \tag{20}$$

where $\beta = (\text{vec}(Q_r)', \mathcal{C}^3_{1,1,1}, \ldots, \mathcal{C}^3_{r,r,r})'$, subject to the constraint $Q_r'Q_r = I$, and $u_t(\Gamma) = \tilde{\theta}^{-1}(x_t - \Phi_0 - \sum_{\tau=1}^p \Phi_\tau x_{t-\tau})$. The operator vecht$(\cdot)$ represents the vectorization of the distinct elements of the corresponding matrix. These moment conditions depend on the parameter vector Γ through $u_t(\Gamma)$.

The constrained GMM estimator of the structural parameter vector β is given by solving the following problem:

$$\widehat{\beta}_T = \arg\min \bar{g}_{2T}(Z_t, \widehat{\Gamma}_T, \beta)' W_{2T} \bar{g}_{2T}(Z_t, \widehat{\Gamma}_T, \beta),$$

where $\bar{g}_{2T}(Z_t, \widehat{\Gamma}_T, \beta) = \frac{1}{T}\sum_{t=1}^T g_2(Z_t, \widehat{\Gamma}_T, \beta)$, subject to the constraint $Q_r'Q_r = I$, and where W_{2T} may depend on the data.

This problem can be transformed into an unconstrained minimization problem using the Cayley transform of an orthogonal matrix (see Golub and Van Loan (2013), Yamada and Ezaki (2003) and Gouriéroux et al. (2017)). For Q_r an orthogonal matrix that does not have -1 as an eigenvalue, we define:

$$A = (Q_r - I)(Q_r + I)^{-1},$$

where A is a skew-symmetric matrix, i.e., $A' = -A$. This gives a one-to-one mapping:

$$Q_r = (I - A)^{-1}(I + A).$$

Therefore, the unconstrained GMM minimizes with respect to the parameter vector $\tilde{\beta} = (\text{vec}(A)', \mathcal{C}^3_{1,1,1}, \ldots, \mathcal{C}^3_{r,r,r})'$.

An expansion around the true value π of the second set of moments gives:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_2(u_t(\Gamma), \beta) + \sum_{t=1}^{T} \frac{\partial g_2}{\partial \Gamma'}(u_t(\Gamma), \beta)\sqrt{T}(\widehat{\Gamma}_T - \Gamma) + o_p(1).$$

An estimator of the optimal weighting matrix is given by the inverse of a consistent estimator of the variance-covariance matrix for the following expression:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_2(u_t(\Gamma), \beta) + \sum_{t=1}^{T} \frac{\partial g_2}{\partial \Gamma'}(u_t(\Gamma), \beta)\sqrt{T}(\widehat{\Gamma}_T - \Gamma).$$

The optimal two-step GMM estimator is obtained using W_{2T} as the inverse of an optimal estimator of the long-run covariance matrix of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_2(Z_t, \widehat{\Gamma}_T, \beta)$. This long-run covariance matrix depends on the estimation from the first set of moment conditions (19). Define the optimal long-run covariance matrix of the entire set of moment conditions $\bar{g}_T(Z_T, \Gamma, \beta) = (\bar{g}_{1T}(Z_t, \Gamma)', \bar{g}_{2T}(Z_t, \Gamma, \beta)')'$ as

$$\lim_{T\to\infty} TE\left[\bar{g}_T(Z_t, \Gamma, \beta)\bar{g}_T(Z_t, \Gamma, \beta)'\right] = \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

Due to the presence of skewness, the covariance matrix between the moment conditions (19) and (20) is non-zero, i.e., $\Omega_{12} \neq 0$.

Define $G_{2,\Gamma} = E\frac{\partial g_2}{\partial \Gamma'}(u_t(\Gamma), \beta)$, and by the first-stage VAR estimation:

$$\sqrt{T}(\widehat{\Gamma}_T - \Gamma) \xrightarrow{d} \mathcal{N}(0, V_\Gamma),$$

where V_Γ is the asymptotic variance-covariance matrix of $\sqrt{T}(\widehat{\Gamma}_T - \Gamma)$.

The optimal weighting matrix is then given by an estimator of the inverse of the following matrix:

$$S_2 = \begin{bmatrix} G_{2,\Gamma}V_\Gamma & I_d \end{bmatrix} \Omega \begin{bmatrix} G_{2,\Gamma}V_\Gamma & I_d \end{bmatrix}'.$$

E Rank Test of Identification

The rank condition for identification can be verified based on the number of asymmetric and/or non-mesokurtic structural shocks. Testing the number of asymmetric and/or non-mesokurtic structural shocks relies on the reduced-form innovations, which can be evaluated from the reduced form (2) prior to estimating the structural form (1), as the structural shocks are not directly observable. Specifically, the number of skewed structural shocks corresponds to the rank of the third-order cross-cumulant in matrix form $\mathbf{C}^3(\nu)$, denoted $\text{rk}(\mathbf{C}^3(\nu)) = r_s$, of the reduced-form innovations, while the number of non-mesokurtic structural shocks is given by the rank of the fourth-order cross-cumulant in matrix form $\mathbf{C}^4(\nu)$, denoted $\text{rk}(\mathbf{C}^4(\nu)) = r_k$, of the reduced-form innovations. Finally, the number of structural shocks that are both skewed and non-mesokurtic is given by the rank of the matrix $\Psi(\nu) = [\mathbf{C}^3(\nu)\ \mathbf{C}^4(\nu)]$.

The ranks of $\mathbf{C}^3(\nu)$, $\mathbf{C}^4(\nu)$, and Ψ_ν allow us to determine the numbers of structural shocks

displaying exclusively non-zero skewness (denoted r_{ss}), excess kurtosis (denoted r_{kk}), and both (denoted r_{sk}). By noting that $\text{rk}(\Psi_\nu) = r_{ss} + r_{kk} + r_{sk}$, the number of structural shocks displaying exclusively non-zero skewness, r_{ss}, excess kurtosis, r_{kk}, and both, r_{sk}, can be readily deduced — given that $r_s = r_{ss} + r_{sk}$ and $r_k = r_{kk} + r_{sk}$.

Guay (2021) proposed using the rank test of Robin and Smith (2000). Let us define the estimate of the normalized reduced-form innovations as $\hat{u}_t = \hat{\Omega}^{-1}\hat{\nu}_t$, where $\hat{\nu}_t$ represents the OLS residuals of the reduced form (2), and $\hat{\Omega}$ is a lower triangular matrix obtained from the Cholesky decomposition of the estimated covariance matrix of the OLS residuals; i.e., $\hat{\Sigma}_\nu = \hat{\Omega}\hat{\Omega}'$. The rank test uses the following likelihood-ratio (LR) and Wald (W) statistics:

$$\widehat{CRT}_{r^*}^{LR} = (T-p)\sum_{i=r^*+1}^{n} \ln(1+\hat{\lambda}_i^2), \tag{21}$$

$$\widehat{CRT}_{r^*}^{W} = (T-p)\sum_{i=r^*+1}^{n} \hat{\lambda}_i^2, \tag{22}$$

where $\hat{\lambda}_i$ are the estimates of the singular values of the matrix $\mathbf{C}^3(\nu)$, $\mathbf{C}^4(\nu)$, or Ψ_u (with $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_n \geq 0$), and r^* is the rank of this matrix under the null hypothesis.

Robin and Smith (2000) show that, under certain regularity conditions, the statistics (21) and (22) have limiting distributions that are weighted sums of independent chi-squared variables. The main drawback of such tests is that the statistics (21) and (22) are not pivotal. An estimator of the weights of the sum of the independent chi-squared distribution can be obtained, allowing for the estimation of asymptotic critical values for the statistics $CRT_{r^*}^{LR}$ and $CRT_{r^*}^{W}$ under the null hypothesis that the rank is r^*. However, the finite-sample critical values for testing the null hypothesis converge very slowly to their asymptotic counterparts, especially in the case of excess kurtosis (see simulation results in Guay (2021)).

To circumvent this problem, Guay (2021) propose a bootstrap procedure to compute the finite-sample critical values for the statistics $CRT_{r^*}^{LR}$ and $CRT_{r^*}^{W}$ associated with the ranks of $\mathbf{C}^3(u)$, $\mathbf{C}^4(u)$, and Ψ_u.

We illustrate the steps of the procedure for determining the rank of $\mathbf{C}^4(u)$.

Step 1. Under the null hypothesis that $\text{rk}[\mathbf{C}^4(u)] = r^*$ (i.e., r^* is the assumed number of non-mesokurtic structural shocks), the vector $u_t^b = \left(u_{r^*,t}^{b\prime} \quad u_{n-r^*,t}^{b\prime}\right)'$ is generated as follows. The elements in the $(r^* \times 1)$ subvector $u_{r^*,t}^b$ are obtained by bootstrapping the elements in the vector $w_{r^*,t} = \widehat{C}_{r^*}'\hat{u}_t$ for $t = (p+1), \ldots, T$, where $\widehat{C}_{r^*}$ is an $(n \times r^*)$ matrix stacking the left singular vectors associated with the r^* largest singular values of $\widehat{\mathbf{C}}^4(u)$, and $\hat{u}_t$ is the $(n \times 1)$ vector of estimated normalized reduced-form innovations. This implies that the elements in $w_{r^*,t}$ correspond to linear combinations of the normalized reduced-form innovations that display the largest excess kurtoses. The elements in the $[(n-r^*) \times 1]$ subvector $u_{n-r^*,t}^b$ are drawn from a symmetric and mesokurtic distribution, i.e., $u_{n-r^*,t}^b \sim N(0, I)$ for $t = (p+1), \ldots, T$.

Step 2. The bootstrap sample is generated recursively from the VAR process (2) as:

$$x_t^b = \hat{\Gamma}_0 + \sum_{\tau=1}^{p} \hat{\Gamma}_\tau x_{t-\tau}^b + \hat{\Omega} u_t^b,$$

for $t = (p+1), \ldots, T$. The starting values of x_t^b for $t = 1, \ldots, p$ are generated by randomly drawing a block of the actual data of length p, while $\hat{\Gamma}_0$, $\hat{\Gamma}_\tau$, and $\hat{\Omega}$ are the estimates of the reduced-form parameters obtained via OLS on the actual sample. These estimates are treated as the population values of the reduced-form parameters.

Step 3. The VAR process is then estimated, yielding:

$$x_t^b = \hat{\Gamma}_0^b + \sum_{\tau=1}^{p} \hat{\Gamma}_\tau^b x_{t-\tau}^b + \hat{\Omega}^b \hat{u}_t^b,$$

where $\hat{\Gamma}_0^b$, $\hat{\Gamma}_\tau^b$, and $\hat{\Omega}^b$ are the estimates obtained by performing OLS on the bootstrap sample, and $\hat{u}_t^b$ corresponds to the normalized residuals.

Step 4. The normalized residuals $\hat{u}_t^b$ are used to compute the bootstrap analogues of the statistics (21) and (22).

Step 5. Steps 1 to 4 are repeated for $b = 1, \ldots, B$, where $B = 2000$, to compute the empirical distributions of the statistics (21) and (22). Selecting the appropriate quantiles of these empirical distributions yields the finite-sample critical values to test the null hypothesis that the rank is r^* against the alternative that the rank is larger than r^*.

Step 6. Steps 1 to 5 are repeated for $r^* = 0, 1, \ldots, n-1$. If the null hypothesis $\text{rk}[\mathbf{C}^4(u)] = r^*$ is rejected for $r^* = 0, 1, \ldots, m-1$ but not rejected for $r^* = m$ with $m < n$, then the number of non-mesokurtic structural shocks is $r_k = m$. However, if the null hypothesis $\text{rk}[\mathbf{C}^4(u)] = r^*$ is rejected for $r^* = 0, 1, \ldots, n-1$, then $r_k = n$.

F Simulations of the Rank Test Using the Second Application as the DGP

We investigate the finite sample properties of the identification rank test using the data-generating process (DGP) specified in the second application. The DGP is simulated based on parameter estimates from the reduced-form VAR model, with a sample size of 716 observations and nine lags in the VAR. We then apply the bootstrap procedure as previously described. The estimated variance-covariance matrix is decomposed using the Cholesky method, i.e., $\hat{\Sigma}_\nu = \hat{\Omega}\hat{\Omega}'$, and the normalized errors u_t^b are generated according to the distributions specified below. The number of simulations is set to 10,000 draws, with the number of bootstrap replications fixed at $B = 2000$.

First, we examine the properties of the test under the null hypothesis of Gaussianity. For this, the 4×1 vector u_t is drawn from a multivariate normal distribution, $\mathcal{N}(0, I_4)$.

Second, as a deviation from Gaussianity, we consider a skewed-t distribution. The skewed-t distribution is a generalization of the Student's t-distribution that incorporates skewness, allowing it to model asymmetric data with heavier tails, as observed in the application. We examine three cases based on the degree of skewness and excess kurtosis:

- **Case 1**: The most substantial deviation from Gaussianity, with a skewness of -1 and an excess kurtosis of 6.
- **Case 2**: An intermediate deviation, with a skewness of -0.75 and an excess kurtosis of 2.
- **Case 3**: A mild deviation, with a skewness of -0.5 and an excess kurtosis of 0.75.

Under the null hypothesis, when $r^* = 0$, the 4×1 vector u_t is drawn entirely from a multivariate normal distribution, $\mathcal{N}(0, I_4)$.

When $r^* \neq 0$, still under the null hypothesis, an r^*-dimensional subvector of u_t is drawn from one of the three skewed-t cases specified above, while the remaining $(4 - r^*)$ elements are drawn from a $\mathcal{N}(0, I_{4-r^*})$ distribution.

Under the alternative hypothesis, an $(r^* + 1)$-dimensional subvector of u_t is drawn from one of the three skewed-t cases specified above, while the remaining $(4 - (r^* + 1))$ elements are drawn from a $\mathcal{N}(0, I_{4-(r^*+1)})$ distribution.

Table F.1 reports the empirical sizes for finite-sample distributions, where the critical values are derived from the bootstrap procedure described above. First, under the null hypothesis, when $r^* = 0$, the entries are identical across all three cases, as the vector u_t is drawn entirely from a multivariate normal distribution, $\mathcal{N}(0, I_4)$, in each scenario.

Importantly, both the Wald and likelihood-ratio tests exhibit minimal size distortions: the empirical sizes closely align with the nominal sizes for all values of r^*, with only a slight tendency toward overrejection, primarily at higher values of r^* and, unexpectedly, in the case with low skewness and excess kurtosis.

As anticipated, the power of the tests increases substantially with higher levels of skewness and excess kurtosis. In the first two scenarios, the rejection rate exceeds 80% for most alternatives. However, in cases with low skewness and low excess kurtosis, the analysis tends to be less conservative. This implies that an analyst is more likely to mistakenly conclude that the entire system is under-identified (even when it is actually identified) or to underestimate the size of the identified subsystem (even when the system is only partially under-identified).

Overall, our bootstrap procedure for rank tests effectively mitigates size distortions and demonstrates good power properties for the sample size used in the second application.

Table F.1. Empirical Sizes and Power of Rank Tests: Kurtosis

Finite-Sample Distributions

Case 1

Under the Null Hypothesis							Under the Alternative Hypothesis						
	Wald			LR				Wald			LR		
r^*	10%	5%	1%	10%	5%	1%	r^*+1	10%	5%	1%	10%	5%	1%
0	10.62	5.34	1.16	10.54	5.30	1.19	1	96.12	93.74	87.67	95.30	92.53	84.98
1	10.77	5.75	1.34	10.70	5.74	1.34	2	94.99	91.95	83.25	94.14	90.86	80.39
2	10.62	5.50	1.32	10.65	5.45	1.29	3	95.40	92.33	82.46	94.93	91.62	80.98
3	10.59	5.41	1.53	10.59	5.41	1.53	4	97.03	94.73	85.28	97.03	94.73	85.28

Case 2

Under the Null Hypothesis							Under the Alternative Hypothesis						
	Wald			LR				Wald			LR		
r^*	10%	5%	1%	10%	5%	1%	r^*+1	10%	5%	1%	10%	5%	1%
0	10.62	5.34	1.16	10.54	5.30	1.19	1	86.84	81.14	68.14	85.22	78.70	64.08
1	10.37	5.25	1.15	10.32	5.26	1.18	2	83.67	76.32	58.80	82.49	74.10	54.94
2	10.58	5.72	1.58	10.60	5.70	1.51	3	83.46	75.21	55.58	82.60	73.89	53.33
3	10.96	5.98	1.70	10.96	5.98	1.70	4	88.30	81.16	62.14	88.30	81.16	62.14

Case 3

Under the Null Hypothesis							Under the Alternative Hypothesis						
	Wald			LR				Wald			LR		
r^*	10%	5%	1%	10%	5%	1%	r^*+1	10%	5%	1%	10%	5%	1%
0	10.62	5.34	1.16	10.54	5.30	1.19	1	57.95	46.92	29.23	55.86	44.84	26.65
1	11.12	5.75	1.24	11.07	5.77	1.30	2	51.81	38.87	19.05	50.49	36.87	17.33
2	11.66	6.66	1.78	11.57	6.66	1.74	3	49.49	36.53	17.14	48.88	35.67	16.21
3	12.25	6.94	2.16	12.25	6.94	2.16	4	53.46	40.57	20.19	53.46	40.57	20.19

Notes: The entries represent the empirical sizes and power (in percentage) of the rank tests under the null hypothesis that $\text{rk}[\mathbf{C}_u^4] = r^*$ and the alternative that $\text{rk}[\mathbf{C}_u^4] = r^* + 1$. For each parametrization, 10,000 simulated samples of size T are generated to compute the proportion of times that the Wald statistic, $\widehat{CRT}_{r^*}^{W}$, and the likelihood-ratio (LR) statistic, $\widehat{CRT}_{r^*}^{LR}$, exceed the bootstrapped critical values, following the procedure described above.