CIRANO
Knowledge into action

A SPECTRAL FRAMEWORK FOR NON-GAUSSIAN SVARS

ALAIN GUAY
DALIBOR STEVANOVIC

CS
2025s-26
WORKING PAPER

CIRANO
Knowledge into action

The purpose of the **Working Papers** is to disseminate the results of research conducted by CIRANO research members in order to solicit exchanges and comments. These reports are written in the style of scientific publications. The ideas and opinions expressed in these documents are solely those of the authors.

Les cahiers de la série scientifique visent à rendre accessibles les résultats des recherches effectuées par des chercheurs membres du CIRANO afin de susciter échanges et commentaires. Ces cahiers sont rédigés dans le style des publications scientifiques et n'engagent que leurs auteurs.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO Partners – *Les partenaires du CIRANO*

Corporate Partners – Partenaires Corporatifs	**Governmental partners - Partenaires gouvernementaux**	**University Partners – Partenaires universitaires**
Autorité des marchés financiers	Ministère des Finances du Québec	École de technologie supérieure
Banque de développement du Canada	Ministère de l'Économie, de l'Innovation et de l'Énergie	École nationale d'administration publique de Montréal
Banque du Canada	Innovation, Sciences et Développement Économique Canada	HEC Montreal
Banque Nationale du Canada	Ville de Montréal	Institut national de la recherche scientifique
Bell Canada		Polytechnique Montréal
BMO Groupe financier		Université Concordia
Caisse de dépôt et placement du Québec		Université de Montréal
La Caisse		Université de Sherbrooke
Énergir		Université du Québec
Hydro-Québec		Université du Québec à Montréal
Intact Corporation Financière		Université Laval
Mouvement Desjardins		Université McGill
Power Corporation du Canada		
Pratt & Whitney Canada		
VIA Rail Canada		

CIRANO collaborates with many *centres* and university research chairs; list available on its website. *Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.*

The observations and viewpoints expressed in this publication are the sole responsibility of the authors; they do not represent the positions of CIRANO or its partners. *Les idées et les opinions émises dans cette publication sont sous l'unique responsabilité des auteurs et ne représentent pas les positions du CIRANO ou de ses partenaires.*

ISSN 2292-0838 (online version)

A Spectral Framework for Non-Gaussian SVARs[*]

Alain Guay[†], *Dalibor Stevanovic*[‡]

Abstract/Résumé

This paper develops a spectral framework for identification, estimation, and inference in non-Gaussian Structural Vector Autoregressive (SVAR) models using higher-order cumulants. Under independence or the absence of cross-cumulants, cumulant tensors of whitened innovations admit an orthogonal decomposition whose singular vectors recover the structural shocks. Identification is therefore governed by the spectral geometry of the population cumulant tensor. In particular, separation of tensor singular values provides a quantitative measure of identification strength through explicit perturbation bounds linking estimation error to the inverse singular-value gap. This characterization yields asymptotic normality under strong identification and nonstandard limits under local-to-weak identification sequences. We derive asymptotic distributions for tensor SVD estimators and show how statistically identified subsystems can be completed using conventional structural restrictions. Monte Carlo experiments and empirical applications illustrate the finite-sample properties and empirical relevance of the approach.

Cet article développe un cadre spectral pour l'identification, l'estimation et l'inférence dans les modèles SVAR (Structural Vector Autoregressive) non gaussiens à l'aide de cumulants d'ordre supérieur. Sous l'hypothèse d'indépendance ou d'absence de cumulants croisés, les tenseurs de cumulants des innovations blanchies admettent une décomposition orthogonale dont les vecteurs singuliers permettent de retrouver les chocs structurels. L'identification est ainsi gouvernée par la géométrie spectrale du tenseur de cumulants de la population.

En particulier, la séparation des valeurs singulières du tenseur fournit une mesure quantitative de la force de l'identification grâce à des bornes explicites de perturbation reliant l'erreur d'estimation à l'inverse de l'écart entre les valeurs singulières. Cette caractérisation conduit à une normalité asymptotique sous identification forte et à des lois limites non standard dans des séquences d'identification localement faibles.

Nous dérivons les distributions asymptotiques des estimateurs fondés sur la SVD tensorielle et montrons comment des sous-systèmes statistiquement identifiés peuvent être complétés à

---

[*] This paper is a substantially revised and expanded version of an earlier manuscript circulated under the title Estimation of Non-Gaussian SVARs Using Tensor Singular Value Decomposition. 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 CIREQCMP Econometrics Conference in Honor of Éric Ghysels 2024 and the 2025 BSE Summer Forum workshop on Advances in Structural Shocks Identification for helpful comments.

[†] Alain Guay, Department of Economics, ESG-UQAM Montreal, 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.

l'aide de restrictions structurelles conventionnelles. Des expériences de Monte Carlo et des et des applications empiriques illustrent les propriétés en échantillon fini et la pertinence empirique de l'approche.

Keywords/Mots-clés: Non-Gaussian SVAR, tensor decomposition, cumulants / SVAR non gaussien, décomposition tensorielle, cumulants

JEL Codes/Codes JEL: C12, C32, C51.

Pour citer ce document / To quote this document
Guay, A., & Stevanovic, D. (2026). A Spectral Framework for Non-Gaussian SVARs (2026s-02, Cahiers scientifiques, CIRANO.) https://doi.org/10.54932/MEID7213

1. Introduction

Structural Vector Autoregressive (SVAR) models are a central tool in applied macroeconomics for quantifying the effects of structural shocks. Traditional identification strategies exploit orthogonality of shocks and the information contained in the covariance matrix of reduced-form innovations, supplemented by economic restrictions such as short-run zero restrictions (Sims, 1980), long-run neutrality constraints (Blanchard and Quah, 1989), or sign restrictions (Uhlig, 2005).

A growing literature instead exploits statistical identification based on independent non-Gaussian structural shocks (Lanne et al., 2017; Gouriéroux et al., 2017); see Lewis (2025) for a survey. When structural innovations exhibit skewness or excess kurtosis, higher-order moments contain identifying information absent from second moments. Both likelihood-based and moment-based procedures have been proposed to recover structural shocks from this non-Gaussian information.

Identification through non-Gaussianity also provides a useful complement to theory-driven approaches. Statistical identification can be used to evaluate economic restrictions by comparing impulse responses or forecast error variance decompositions, or by formally testing structural hypotheses implied by economic models (e.g., Keweloh 2020; Anttonen et al. 2023). Hybrid strategies combining statistical and economic restrictions have therefore emerged as a promising direction for structural analysis (Bekaert et al., 2025; Drautzburg and Wright, 2023).

Despite this progress, the statistical foundations of non-Gaussian identification remain only partially understood. Existing contributions typically focus on specific estimators, while the underlying source of identification strength—and its implications for estimation stability, inference, and weak identification—remains unclear. In particular, there is no unified population measure quantifying how distinct non-Gaussian features must be for structural shocks to be reliably identified.

This paper develops a spectral framework for identification, estimation, and inference in non-Gaussian SVARs based on higher-order cumulants. After whitening reduced-form innovations, third- and fourth-order cumulants admit an orthogonal tensor decomposition whose singular vectors coincide with structural shocks (up to sign and permutation) under independence or weaker conditions excluding coskewness and cokurtosis. Structural identification therefore reduces to a spectral decomposition problem for an orthogonally decomposable cumulant tensor.

The key insight is that identification strength is governed by the separation of tensor singular values. Spectral gaps provide a population measure determining whether structural directions are sharply identified or nearly rotationally indeterminate. Sampling fluctuations in estimated cumulants induce rotations along nearby spectral directions, implying that estimation precision, finite-sample stability, and weak-identification phenomena are all governed by the same spectral quantity. Identification can thus be interpreted as a conditioning problem for the cumulant operator.

Building on this characterization, we introduce a symmetric Tensor Singular Value Decomposition (TSVD) estimator that recovers structural shocks by maximizing the diagonal mass of higher-order cumulant tensors. When third- and fourth-order cross-cumulants vanish, the cumulant tensors admit an orthogonal decomposition that restores an SVD-like geometry and delivers direct identification of the structural directions. The framework accommodates both complete system identification and *subsystem identification*, allowing non-Gaussian features to characterize only a subset of structural shocks.

A central implication is that identification is governed by the number of non-Gaussian *structural shocks*, rather than by the number of non-Gaussian reduced-form innovations. In macroeconomic environments, departures from Gaussianity often affect only a subset of shocks—for example when asymmetry or conditional heteroskedasticity is confined to specific structural disturbances, or when several reduced-form innovations load on a common non-Gaussian factor such as aggregate business-cycle or volatility shocks.[1] In such settings, the dimension of the non-Gaussian structural component may be strictly smaller than that of the reduced-form innovation vector. Identification therefore applies only to the structural subspace spanned by the non-Gaussian shocks, while the remaining directions remain unrestricted and can be completed using conventional economic restrictions.

The paper makes four main contributions.

First, we develop a spectral theory of identification for non-Gaussian SVARs. Identification strength, estimator stability, and weak-identification phenomena are shown to be governed by singular-value separation of the population cumulant tensor, providing a quantitative and estimator-independent measure of identification strength.

[1]See, for example, Sentana and Fiorentini (2001) and Normandin and Phaneuf (2004). Such mechanisms are consistent with macroeconomic risk episodes including the Great Moderation and the COVID-19 recession (Bekaert et al., 2025; Montiel Olea et al., 2022).

Second, we propose joint and sequential TSVD estimators and establish their asymptotic properties under both complete-system and subsystem identification. Under strong identification, the estimators admit linear expansions and are asymptotically normal. When spectral gaps shrink, we characterize local-to-weak identification through explicit perturbation limits in which individual structural directions become asymptotically unidentified while invariant subspaces remain consistently estimable.

Third, we derive deterministic perturbation bounds linking estimation error directly to cumulant estimation noise scaled by inverse spectral gaps. These bounds provide non-asymptotic stability guarantees and clarify the relationship between identification strength and finite-sample robustness.

Fourth, we show how a TSVD-identified subsystem can be completed into a fully identified SVAR by imposing conventional economic restrictions only on its orthogonal complement. This yields a modular identification strategy combining statistical and theory-based information.

Monte Carlo simulations demonstrate competitive finite-sample performance relative to existing methods, particularly under moderate or subsystem identification. Two empirical applications illustrate the usefulness of the framework for identifying fiscal and financial shocks in macroeconomic VARs.

Related literature The spectral perspective developed here provides a unifying lens through which existing approaches to non-Gaussian identification can be interpreted and compared.

Moment-based estimators exploiting non-Gaussianity include independent component analysis (ICA) procedures such as JADE and FastICA, as well as generalized method of moments (GMM) estimators constructed from higher-order moments. These methods rely on the same higher-order information but typically operate on matrix slices or scalar contrasts. By contrast, the spectral framework developed here exploits the full cumulant tensor and characterizes identification through singular-value separation, yielding explicit perturbation bounds governing estimation stability and weak identification.

Likelihood-based approaches, including pseudo-likelihood and flexible Bayesian non-Gaussian SVARs, identify shocks through distributional shape parameters. In these models, identification strength corresponds to curvature of the likelihood or posterior distribution. Spectral gaps provide a geometric interpretation common to

both moment-based and likelihood-based methods, clarifying why weak identification arises when higher-order features of shocks become similar.

Alternative tensor decompositions, such as truncated HOSVD or moment component analysis (Babii et al., 2025; Jondeau et al., 2018), are primarily designed for dimension reduction rather than cumulant diagonalization and therefore do not recover structural shocks in the present framework. By contrast, set-identification approaches for non-Gaussian SVARs (Anttonen et al., 2023; Lanne et al., 2023; Gourieroux and Jasiak, 2025) characterize partially identified regions, whereas TSVD isolates the point-identified structural subspace and thereby permits standard asymptotic inference without set inversion.

Overall, the analysis shows that identification in non-Gaussian SVARs is fundamentally a spectral perturbation problem, in which singular-value separation determines both what can be learned from the data and how precisely it can be estimated.

The remainder of the paper is organized as follows. Section 2 introduces the SVAR specification and identification conditions based on higher-order cumulants. Section 3 develops the tensor notation and spectral decomposition tools. Section 4 presents the TSVD estimator and establishes its asymptotic and perturbation theory. Section 5 reports simulation results, and Section 6 provides the empirical applications. All proofs are collected in the Appendix.

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. Finally, each $(n \times n)$ matrix Φ_τ contains the n^2 unrestricted autoregressive coefficients at lag τ.

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 nonsingular matrix $\Theta = \Phi^{-1}$ captures the contemporaneous impact of the structural shocks, while ν_t denotes the reduced-form innovations. These innovations are related to the structural shocks by $\nu_t = \Theta\epsilon_t$.

The corresponding structural moving-average representation is given by

$$x_t = \Gamma_0^* + \sum_{i=0}^{\infty} \Psi_i \Theta \epsilon_{t-i}, \tag{3}$$

where $\Gamma_0^* = [I_n - \Gamma(1)]^{-1}\Gamma_0$, $\Gamma(1) = \sum_{\tau=1}^{p} \Gamma_\tau$, $\Psi_0 = I_n$, and $\Psi_i = \sum_{j=1}^{i} \Psi_{i-j}\Gamma_j$ for $i = 1, 2, \ldots$ and $\Gamma_j = 0$ for $j > p$.

2.2 Cumulants of Structural Shocks and Innovations

Let $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{nt})'$ denote the vector of structural shocks. The dth-order cumulant is defined elementwise as

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

The first-order cumulant represents the unconditional means of each structural shock i and the second-order cumulant corresponds to the variance-covariance matrix. The third- and the fourth-order cumulants are given by

$$\mathcal{C}^3_{i,j,k}(\epsilon) = \mathrm{Cum}(\epsilon_{it}, \epsilon_{jt}, \epsilon_{kt}) = E[\epsilon_{it}\epsilon_{jt}\epsilon_{kt}], \tag{4}$$

$$\begin{aligned} \mathcal{C}^4_{i,j,k,l}(\epsilon) &= \mathrm{Cum}(\epsilon_{it}, \epsilon_{jt}, \epsilon_{kt}, \epsilon_{lt}) \\ &= E[\epsilon_{it}\epsilon_{jt}\epsilon_{kt}\epsilon_{lt}] - E[\epsilon_{it}\epsilon_{jt}]E[\epsilon_{kt}\epsilon_{lt}] - E[\epsilon_{it}\epsilon_{kt}]E[\epsilon_{jt}\epsilon_{lt}] \\ &\quad - E[\epsilon_{it}\epsilon_{lt}]E[\epsilon_{jt}\epsilon_{kt}]. \end{aligned} \tag{5}$$

where $E[\cdot]$ denotes the unconditional expectation operator.

Higher-order cumulants satisfy two key properties: (i) all cumulants of order $d > 2$ vanish for Gaussian variables, and (ii) all cross-cumulants vanish under independence (or under weaker conditions such as zero coskewness or zero excess cokurtosis).[2] Consequently, under ii), for $d = 3, 4$,

$$\mathcal{C}^d_{i_1 \ldots i_d}(\epsilon) = 0 \quad \text{unless } i_1 = \cdots = i_d.$$

[2]See the Supplementary Materials for additional properties of cumulants.

Under the normalization $E(\epsilon_t \epsilon_t') = I_n$, i.e. the second-order cumulant of ϵ_t. Hence, the second-, third-, and fourth-order cumulants are diagonal in the structural basis, in the sense that only entries with identical indices can be nonzero. The only potentially nonzero higher-order cumulants of the structural shocks are therefore the marginal skewnesses and excess kurtoses:

$$\lambda_{3,i} = \mathcal{C}^3_{i,i,i}(\epsilon), \qquad \lambda_{4,i} = \mathcal{C}^4_{i,i,i,i}(\epsilon).$$

The corresponding matricized representations are

$$\mathbf{C}^3_\epsilon = E[\epsilon_t \epsilon_t' \otimes \epsilon_t'], \tag{6}$$

$$\mathbf{C}^4_\epsilon = E[\epsilon_t \epsilon_t' \otimes \epsilon_t' \otimes \epsilon_t'] - E[\tilde{\epsilon}_t \tilde{\epsilon}_t' \otimes \tilde{\epsilon}_t' \otimes \tilde{\epsilon}_t'], \tag{7}$$

where $\tilde{\epsilon}_t$ denotes a Gaussian vector with the same covariance.

For the reduced-form innovations, $\nu_t = \Theta \epsilon_t$, by multilinearity of cumulants,

$$\mathcal{C}^3_{i,j,k}(\nu) = \sum_{m=1}^{n} \lambda_{3,m}\, \Theta_{im}\Theta_{jm}\Theta_{km}, \tag{8}$$

$$\mathcal{C}^4_{i,j,k,\ell}(\nu) = \sum_{m=1}^{n} \lambda_{4,m}\, \Theta_{im}\Theta_{jm}\Theta_{km}\Theta_{\ell m}. \tag{9}$$

Equivalently, in matricized form,

$$\mathbf{C}^3_\nu = \Theta \mathbf{C}^3_\epsilon (\Theta' \otimes \Theta'), \tag{10}$$

$$\mathbf{C}^4_\nu = \Theta \mathbf{C}^4_\epsilon (\Theta' \otimes \Theta' \otimes \Theta'). \tag{11}$$

The variance-covariance matrix corresponding to the second-order cumulant $\mathbf{C}^2_\nu = \Sigma_\nu$ contains $\frac{n(n+1)}{2}$ distinct elements, while $\mathbf{C}^3_\nu$ and $\mathbf{C}^4_\nu$ contain $\frac{n(n+1)(n+2)}{6}$ and $\frac{n(n+1)(n+2)(n+3)}{24}$ distinct elements, respectively. In the following, we denote a cumulant of order d for a vector x as $\mathcal{C}^d(x)$.

2.3 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. and mutually independent, 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, identification based on skewness and/or excess kurtosis can

arise under two distinct regimes: complete system identification (**CSI**) and subsystem identification (**SSI**). These regimes differ in the number of non-Gaussian structural shocks and in the scope of the identified structural objects.

Assumption 1 *Complete System Identification (**CSI**)*

(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 nonzero skewness and/or excess kurtosis (i.e., are non-mesokurtic).

Under Assumption 1, the structural impact matrix is point identified 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) may be weakened by requiring that only a sufficient number of cross-cumulants vanish; see Lanne and Luoto (2021) and Mesters and Zwiernik (2023).

When fewer structural shocks exhibit non-Gaussian features, **CSI** fails, but identification may still be possible for a subset of structural shocks. In such cases, identification is restricted to a structural subspace.

Assumption 2 *Subsystem Identification (**SSI**)*

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

(ii) A subset of structural shocks of dimension $r < n - 1$ exhibits nonzero skewness and/or excess kurtosis.

(iii) Under the rank condition of Guay (2021, Corollary 1), the subsystem associated with this subset of shocks is point identified up to sign changes and column permutations.

Under Assumption 2, the columns of the impact matrix corresponding to the non-Gaussian structural shocks are point identified, while the orthogonal complement remains unrestricted. Thus, the mapping between reduced-form innovations and the identified subset of structural shocks is statistically identified, but the full structural system is not.

To characterize the strength of identification within the identified subsystem, we impose the following additional assumption.

Assumption 3 *The third- and/or fourth-order cumulants associated with the structural shocks exhibiting nonzero skewness and/or excess kurtosis are distinct.*

Assumption 3 plays a role analogous to the distinct-eigenvalue condition in spectral decompositions. When the relevant cumulants are distinct, the corresponding structural directions are uniquely determined (up to sign and permutation). When two cumulants become close, the associated structural directions become nearly rotationally indeterminate, leading to weak identification. This spectral separation will be shown to govern both the stability of the TSVD estimator and the precision with which structural shocks can be estimated.

3. Cumulant Tensors and Orthogonal Spectral Decomposition

This section formalizes the spectral structure of higher-order cumulants under complete system identification (**CSI**) and subsystem identification (**SSI**), and introduces the symmetric tensor singular value decomposition (TSVD) as the natural orthogonal diagonalization tool.

3.1 Whitened Innovations and Orthogonal Decomposability

Under **CSI** and **SSI**, the only nonzero higher-order cumulants of the structural shocks correspond to the marginal skewnesses and excess kurtoses:

$$\lambda_{3,i} = \mathcal{C}^3_{i,i,i}(\epsilon), \qquad \lambda_{4,i} = \mathcal{C}^4_{i,i,i,i}(\epsilon).$$

Let the reduced-form innovations satisfy $\nu_t = \Theta\epsilon_t$, where ϵ_t denotes the structural shocks and Θ is the structural impact matrix. Let $\Sigma_\nu = E(\nu_t\nu_t')$ denote the covariance matrix of the reduced-form innovations. Consider a square-root decomposition

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

Define the whitened innovations $u_t = \tilde{\Theta}^{-1}\nu_t$. Substituting $\nu_t = \Theta\epsilon_t$ yields $u_t = \tilde{\Theta}^{-1}\Theta\epsilon_t = Q\epsilon_t$, where $Q := \tilde{\Theta}^{-1}\Theta$ is orthogonal because $E(u_tu_t') = I_n$. Thus the whitened innovations are orthogonal transformations of the structural shocks. Any valid square-root decomposition of Σ_ν may be used for prewhitening; common choices include the Cholesky decomposition or the singular value decomposition. The next proposition shows that, under **CSI** and **SSI**, the third- and fourth-order cumulants of the whitened innovations admit an orthogonally decomposable structure.

Proposition 1 *Suppose Assumption 1 (**CSI**) or Assumption 2 (**SSI**) holds and let $u_t = Q\epsilon_t$, where Q is an orthogonal matrix. Then, for $d = 3$ or 4, the dth-order cumulant of u_t satisfies*

$$\mathcal{C}^d(u) = \sum_{j=1}^{r} \lambda_{d,j}\, q_j^{\circ d}, \tag{12}$$

where q_j denotes the jth column of Q and $\lambda_{d,j} = \mathcal{C}^d_{j,\ldots,j}(\epsilon)$. Here $\circ$ denotes the outer product, so that $q_j^{\circ d} = q_j \circ q_j \circ \cdots \circ q_j$ (d times).

*Under **CSI**, $r = n$ (or $n-1$ if one shock is Gaussian), whereas under **SSI**, $r < n$. Hence $\mathcal{C}^d(u)$ admits an orthogonal rank-r decomposition along the directions $\{q_j\}_{j=1}^r$.*

Thus, after whitening, identification reduces to a spectral problem: recovering the orthogonal directions $\{q_j\}_{j=1}^r$ that define the orthogonal rank-r decomposition of the cumulant tensor. The tensor formulation underlying this decomposition is introduced in the next subsection.

3.2 Tensor Representation and Symmetric TSVD

Proposition 1 shows that, under **CSI** or **SSI**, the third- and fourth-order cumulants of the whitened innovations admit an orthogonal rank-r decomposition given by (12). Such objects are naturally represented as higher-order arrays. In particular, higher-order statistics such as cumulants admit a compact representation in tensor form, making tensor decompositions a natural analytical framework. We briefly introduce the required tensor notation below.[3]

An order-N tensor is a multidimensional array $\mathcal{A} \in \mathbb{R}^{n\times\cdots\times n}$, generalizing vectors and matrices, which correspond to the cases $N = 1$ and $N = 2$, respectively. A tensor is said to be *symmetric* if its entries are invariant under any permutation of their indices. Because cumulants are permutation invariant, the third- and fourth-order cumulants $\mathcal{C}^3(u)$ and $\mathcal{C}^4(u)$ are symmetric tensors.

The orthogonal decomposition established in Proposition 1 therefore takes the form of a spectral decomposition for symmetric tensors. This motivates the use of the symmetric tensor singular value decomposition (TSVD), which provides the natural orthogonal diagonalization of higher-order cumulants.[4]

[3] See the Supplementary Material for formal definitions and technical details.

[4] See, for instance, Comon (2002) and Chen and Saad (2009).

Definition 2 (Symmetric TSVD) *A symmetric tensor $\mathcal{A} \in \mathbb{R}^{n\times\cdots\times n}$ admits a symmetric tensor singular value decomposition (TSVD) if it can be written as*

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i \, u_i^{\circ d}, \qquad U'U = I_r,$$

where $U = [u_1, \ldots, u_r]$ has orthonormal columns.

For tensors of order two, this decomposition reduces to the familiar matrix singular value decomposition. The symmetric TSVD extends this spectral representation to higher-order arrays by expressing $\mathcal{A}$ as a sum of orthogonal rank-one components formed by repeated outer products of vectors u_i. The smallest integer r for which such a representation exists is called the rank of $\mathcal{A}$. In this representation the associated core tensor is diagonal, with entries λ_i and all off-diagonal elements equal to zero.[5]

The next proposition provides an explicit characterization of the singular values λ_i as multilinear contractions of $\mathcal{A}$ along the corresponding singular directions.

Proposition 3 *Let $\mathcal{A} \in \mathbb{R}^{n\times\cdots\times n}$ be a symmetric order-N tensor admitting the symmetric TSVD*

$$\mathcal{A} = \sum_{i=1}^{r} \lambda_i u_i^{\circ N}, \qquad U'U = I_r.$$

Then, for each $i = 1, \ldots, r$,

$$\lambda_i = \langle \mathcal{A}, \, u_i^{\circ N} \rangle = u_i' \mathbf{A} \big(u_i^{\otimes (N-1)} \big), \tag{13}$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product, $\mathbf{A}$ denotes any unfolding (matricization) of $\mathcal{A}$, and $u_i^{\otimes(N-1)}$ denotes the $(N-1)$-fold Kronecker product.

The next corollary shows that, under Proposition 3, tensor singular values coincide with diagonal cumulant contractions along the corresponding structural directions.

Corollary 1 *Let $d \in \{3, 4\}$ and suppose the population dth-order cumulant of the whitened innovations admits the orthogonal decomposition*

$$\mathcal{C}^d(u) = \sum_{i=1}^{r} \lambda_{d,i} \, q_i^{\circ d}, \qquad Q_r = [q_1, \ldots, q_r], \quad Q_r' Q_r = I_r.$$

Let $\mathbf{C}_u^d$ denote any matricization (mode unfolding) of $\mathcal{C}^d(u)$. Then, for each $i = 1, \ldots, r$,

$$\lambda_{d,i} = \big\langle \mathcal{C}^d(u), \, q_i^{\circ d} \big\rangle = q_i' \mathbf{C}_u^d \big(q_i^{\otimes (d-1)} \big). \tag{14}$$

[5] See the Supplementary Material for formal definitions of rank-one tensors and core tensors.

In particular, the population diagonal cumulant functional entering the TSVD criterion coincides with the corresponding tensor singular value.

Under subsystem identification ($r < n$), the cumulant tensor therefore has orthogonal rank r. A key implication is that, unlike generic tensors, orthogonally decomposable tensors admit a best rank-r approximation obtained by truncation of the TSVD. This spectral property is exceptional: although an Eckart–Young theorem does not hold for general higher-order tensors, it remains valid within the class of orthogonally decomposable tensors (see Anandkumar et al., 2014). Consequently, **SSI** can be interpreted as a low-rank orthogonal tensor spectral problem.

The tensor SVD decomposition parallels principal component analysis (PCA), which identifies orthogonal directions that maximize variance, corresponding to the second-order cumulant. Applied to higher-order cumulants, the tensor SVD instead extracts orthogonal directions that maximize higher-order dependence measures.

In general, tensors of order greater than two cannot be fully diagonalized using orthogonal transformations alone. The absence of exact diagonalization for higher-order symmetric tensors therefore motivates an alternative notion of diagonal structure. This leads to the concept of *maximal diagonality*, which formalizes the approximation of a tensor by an orthogonal decomposition that concentrates as much mass as possible on its diagonal elements. The following result provides a precise characterization.

Proposition 4 (Maximal diagonality) *Let $\mathcal{A} \in \mathbb{R}^{n\times\cdots\times n}$ be a symmetric tensor of order N. Among all orthogonal rank-r decompositions, the best approximation to $\mathcal{A}$ in Frobenius norm solves*

$$\min_{\substack{\lambda_1,\ldots,\lambda_r \\ U'U=I_r}} \left\| \mathcal{A} - \sum_{i=1}^{r} \lambda_i \, u_i^{\circ N} \right\|_F^2, \tag{15}$$

where $U = [u_1, \ldots, u_r]$ has orthonormal columns.

This problem is equivalent to

$$\max_{U'U=I_r} \sum_{i=1}^{r} \lambda_i(U)^2, \qquad \lambda_i(U) = \langle \mathcal{A},\ u_i^{\circ N} \rangle,$$

that is, maximizing the squared diagonal mass of the associated orthogonal core tensor.

This equivalence is formally established in Proposition 6.1 of Chen and Saad (2009); see also Comon (1994) and De Lathauwer et al. (2000b). Consequently, the

symmetric TSVD identifies the orthogonal basis that maximizes the diagonal entries of the associated core tensor.

In the present setting this property leads to substantial computational gains. The complexity is reduced from $O(n^J)$ to computing only n diagonal cumulants of order J under complete identification, and from $O(n^J)$ to only r diagonal cumulants under subsystem identification, where r denotes the number of non-Gaussian shocks. This efficiency gain becomes particularly important in moderate- to high-dimensional systems, where alternative procedures—such as GMM estimators based on cross-cumulants—can become computationally burdensome.[6]

4. Estimation of the Orthogonal Structure via TSVD

4.1 TSVD Estimator

Let $\widehat{\mathcal{C}}_T^d(u)$ denote the sample cumulant tensor. By Proposition 4 and Corollary 1, estimation of the orthogonal structure can be formulated as an orthogonal rank-r approximation problem:

$$\min_{Q_r'Q_r=I_r} \left\| \widehat{\mathcal{C}}_T^d(u) - \sum_{j=1}^{r} \widehat{\lambda}_{d,j}(Q_r)\, q_j^{\circ d} \right\|_F^2,$$

where the coefficients are defined by

$$\widehat{\lambda}_{d,j}(Q_r) = \langle \widehat{\mathcal{C}}_T^d(u),\, q_j^{\circ d} \rangle.$$

This problem seeks the best orthogonal decomposition of the sample cumulant tensor within the class of orthogonally decomposable tensors.

By Proposition 4, the minimization problem is equivalent to the maximization problem

$$\widehat{Q}_{r,T} \in \arg \max_{Q_r'Q_r=I_r} \sum_{j=1}^{r} \left(\langle \widehat{\mathcal{C}}_T^d(u),\, q_j^{\circ d} \rangle \right)^2, \tag{16}$$

which defines the symmetric TSVD estimator. Hence TSVD selects orthogonal di-

[6]The GMM estimation procedure based on higher-order cumulants is described in Supplementary Material; see also Lanne and Luoto (2021) for a related approach exploiting third- and fourth-order moments. In the absence of additional identifying restrictions, this GMM implementation is valid only under complete identification. Importantly, GMM-based estimators minimize a criterion that differs from (15): rather than directly targeting the best orthogonal tensor approximation in Frobenius norm, they rely on moment conditions derived from cumulants. In the simulation section we therefore also consider weighted and unweighted minimum-distance estimators that explicitly solve the minimization problem in (15).

rections that maximize the diagonal mass of the rotated cumulant tensor.

From Corollary 1, the population singular values satisfy

$$\lambda_{d,j} = \langle \mathcal{C}^d(u),\, q_j^{\circ d} \rangle = q_j' \mathbf{C}_u^d \big(q_j^{\otimes(d-1)}\big).$$

The sample analogue replaces $\mathcal{C}^d(u)$ by $\widehat{\mathcal{C}}_T^d(u)$. The estimated directions therefore asymptotically diagonalize the cumulant operator:

$$\widehat{q}_j' \widehat{\mathbf{C}}_{u,T}^d (\widehat{q}_k^{\otimes(d-1)}) = 0, \qquad j \neq k.$$

Under **CSI**, $r = n$ (or $n-1$ if one shock is Gaussian), and TSVD recovers the orthogonal matrix Q up to sign and permutation. Under **SSI**, $r < n$ and $\mathcal{C}^d(u)$ has orthogonal rank r. The estimator then recovers the invariant subspace associated with the nonzero singular values, while the remaining directions are not identified by higher-order cumulants and may be completed using additional structural restrictions.

Thus estimation of a non-Gaussian SVAR reduces to a low-rank spectral problem for an orthogonally decomposable cumulant tensor, with identification strength governed by separation among the singular values $\{\lambda_{d,j}\}$. The following definition formalizes the joint TSVD estimator.

Definition 5 (Joint TSVD estimator) *For a given cumulant order $d \in \{3, 4\}$, the joint TSVD estimator of $Q_r = [q_1, \ldots, q_r]$ is defined as*

$$\widehat{Q}_{r,T} \;\in\; \arg \max_{Q_r' Q_r = I_r} J_T(Q_r),$$

where the sample contrast function is

$$J_T(Q_r) = \sum_{i=1}^{r} \lambda_{d,i,T}^2(Q_r), \qquad \lambda_{d,i,T}(Q_r) = q_i' \widehat{\mathbf{C}}_{u,T}^d \big(q_i^{\otimes(d-1)}\big),$$

and $\widehat{\mathbf{C}}_{u,T}^d$ denotes a sample estimator of the d-th order cumulant in matricized form.

The estimator $\widehat{Q}_{r,T}$ solves the constrained optimization problem characterized by the Lagrangian

$$\mathcal{L}_T(Q_r, \mu) = \sum_{i=1}^{r} \lambda_{d,i,T}^2(Q_r) - \sum_{j,k=1}^{r} \mu_{jk} \big(q_j' q_k - \delta_{jk}\big), \tag{17}$$

where δ_{jk} denotes the Kronecker delta and μ_{jk} are Lagrange multipliers enforcing the orthogonality constraints $Q_r' Q_r = I_r$.

4.1.1 Asymptotic distribution

In what follows, we first focus on identification driven by skewness or excess kurtosis, and subsequently consider the mixed criterion.

Theorem 6 *Let $d \in \{3, 4\}$ denote the cumulant order used for identification. Suppose Assumption 1 (**CSI**) or Assumption 2 (**SSI**) holds together with Assumption 3 (distinct diagonal cumulants), and assume the moment condition $E(\epsilon_t^{2d}) < \infty$.*

Let $\widehat{Q}_{r,T}$ denote the joint TSVD estimator (Definition 5) and suppose that

$$\sqrt{T}\Big(\mathrm{vec}(\widehat{\mathbf{C}}^d_{u,T}) - \mathrm{vec}(\mathbf{C}^d_u)\Big) \Rightarrow N(0, \Sigma_d).$$

(i) Joint asymptotic distribution

$$\sqrt{T}\Big(\mathrm{vec}(\widehat{Q}_{r,T}) - \mathrm{vec}(Q_r)\Big) \Rightarrow N(0,\ (I_r \otimes Q_r)\Xi M_d'\ \Sigma_d\ M_d \Xi'(I_r \otimes Q_r)')\,,$$

where $M_d := Q_r^{\odot(d-1)} \otimes Q_r$ and $\Xi = (I_{r^2} - P_{r,r})F$ are defined in the Appendix, and $\odot$ denotes the Khatri–Rao product.

(ii) Singular-vector perturbation expansion

For each $j \le r$,

$$\sqrt{T}\, P_j^{\perp}(\widehat{q}_{j,T} - q_j) = \sum_{\ell \neq j} \frac{\sqrt{T}\,\alpha_{\ell j,T}}{\lambda_{d,j} - \lambda_{d,\ell}}\, q_\ell + o_p(1), \tag{18}$$

where

$$\alpha_{\ell j,T} = \Big\langle \widehat{\mathcal{C}}^d_{u,T} - \mathcal{C}^d_u,\ q_\ell \otimes q_j^{\otimes(d-1)} \Big\rangle, \qquad P_j^{\perp} = I - q_j q_j'.$$

(iii) Asymptotic variance and identification strength

$$\sqrt{T}(\widehat{q}_{j,T} - q_j) \Rightarrow N(0, \Omega_j), \qquad \Omega_j = \sum_{\ell \neq j} \frac{\mathrm{Var}(\alpha_{\ell j,T})}{(\lambda_{d,j} - \lambda_{d,\ell})^2}\, q_\ell q_\ell'.$$

Hence estimation precision is governed by the inverse squared singular-value gaps δ_j^{-2}.

Theorem 6 shows that TSVD estimation is fundamentally a spectral perturbation problem. Sampling noise in estimated higher-order cumulants induces rotations of structural directions whose magnitude is governed by singular-value gaps. The estimator admits a local linear representation linking cumulant estimation error to structural directions through an explicit Jacobian mapping, while the singular-coordinate

expansion shows that estimation error is driven by projections onto neighboring spectral directions scaled by inverse gaps.

As a consequence, identification strength, asymptotic variance, and finite-sample stability are governed by the same spectral quantity. Well-separated tensor singular values yield stable identification and Gaussian inference, whereas shrinking gaps generate variance inflation and weak identification. Non-Gaussian SVAR identification can therefore be interpreted geometrically as a conditioning problem for an orthogonally decomposable cumulant tensor.

As in eigenvalue decompositions, the precision of $\widehat{Q}_{r,T}$ depends on the separation between the diagonal cumulant coefficients $\{\lambda_{d,i}\}$ for $d \in \{3,4\}$, and the strength of identification is governed by Assumption 3.[7]

The main computational advantage of TSVD is that the maximization problem (17) depends only on diagonal cumulant components. Consequently, computational complexity scales linearly with the number of identified directions. Under **CSI**, the objective involves n diagonal cumulants, yielding an $O(n)$ problem, whereas under **SSI** it involves only $r < n$ diagonal cumulants, yielding an $O(r)$ problem.

By contrast, a GMM estimator exploiting all cross-cumulants of order $J \in \{3,4\}$ must handle $O(n^J)$ moment conditions. Classical independent component analysis (ICA) methods such as FastICA and JADE also rely on higher-order moments but typically require iterative fixed-point updates (FastICA) or joint diagonalization of multiple cumulant matrices (JADE), whose computational cost grows at least quadratically in the dimension. In comparison, TSVD formulates identification as a single orthogonal spectral problem based on a symmetric cumulant object, avoiding sequential deflation or joint diagonalization steps.

In practice, empirically relevant configurations differ according to which higher-order moments drive identification. Identification may rely on skewness, excess kurtosis, or a combination of both. Suppose that r_1 shocks are identified through skewness and r_2 through non-mesokurticity, with $r_1 + r_2 = r$. The joint TSVD problem can be written compactly as

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

[7]The asymptotic distributions of $\widehat{\mathbf{C}}_{u,T}^3$ and $\widehat{\mathbf{C}}_{u,T}^4$ depend on the asymptotic distribution of the reduced-form parameters of the first-stage VAR, $\Gamma = \left(\mathrm{vec}(\Phi_0)', \ldots, \mathrm{vec}(\Phi_p)', \mathrm{vech}(\Sigma_\nu)'\right)'$, where $\Sigma_\nu = \tilde{\Theta}\tilde{\Theta}'$.

where $\lambda_{d,i}(Q_r)$ denotes the diagonal cumulant contraction of order $d \in \{3,4\}$ associated with direction q_i, and the weights $(w_{3,i}, w_{4,i})$ select or scale the relevant cumulants. Setting $w_{3,i} = 1$ (resp. $w_{4,i} = 1$) isolates skewness (resp. kurtosis) identification, while allowing both to be positive exploits joint higher-order information within a unified spectral criterion.[8] In empirical applications, the appropriate identification regime may be selected using the bootstrap rank tests for higher-order cumulants proposed by Guay (2021).

4.1.2 Sequential TSVD estimator

The orthogonal structure of higher-order cumulant tensors also allows a sequential extraction of rank-one components, analogous to the tensor power method of Anandkumar et al. (2014).

The sequential TSVD estimates directions recursively by maximizing a cumulant-based contrast subject to orthogonality with previously identified directions. Under **CSI**, the procedure applies with $r = n$ (or $r = n - 1$ if one shock is Gaussian), while under **SSI** it naturally targets the identified subspace with $r < n$.

Definition 7 (Sequential TSVD estimator) *Let $d \in \{3,4\}$ denote the cumulant order used for identification and let $\lambda_{d,T}(q)$ denote the sample cumulant contrast in direction q.*

The first sequential direction is

$$\widehat{q}_{1,T}^{\text{seq}} \in \arg\max_{\|q\|=1} \left[\lambda_{d,T}(q)\right]^2.$$

Given $\widehat{q}_{1,T}^{\text{seq}}, \ldots, \widehat{q}_{k-1,T}^{\text{seq}}$, define the orthogonal complement

$$\mathcal{S}_{k-1} = \{q \in \mathbb{R}^n : q'\widehat{q}_{j,T}^{\text{seq}} = 0,\ j = 1, \ldots, k-1\}.$$

For $k = 2, \ldots, r$,

$$\widehat{q}_{k,T}^{\text{seq}} \in \arg\max_{\substack{q \in \mathcal{S}_{k-1} \\ \|q\|=1}} \left[\lambda_{d,T}(q)\right]^2.$$

The sequential estimator is $\widehat{Q}_{r,T}^{\text{seq}} = \left[\widehat{q}_{1,T}^{\text{seq}}, \ldots, \widehat{q}_{r,T}^{\text{seq}}\right]$.

Under orthogonal decomposability and distinct diagonal cumulants (Assumption 3), the sequential procedure recovers the structural directions one at a time:

[8] The asymptotic distributions associated with these alternative specifications are derived in the Supplementary Materials.

the first step identifies the direction with the strongest non-Gaussian signal, and subsequent steps recover the remaining directions after deflation. Consequently, $\widehat{Q}_{r,T}^{\text{seq}}$ consistently estimates the population directions Q_r, up to sign and ordering.[9]

At the population level, the joint and sequential TSVD estimators recover the same identified subspace (up to permutation and sign). In finite samples, however, sequential estimation enforces orthogonality with respect to estimated directions, so its estimating equations differ from those of the joint estimator and its asymptotic variance generally differs from that in Theorem 6.

From a computational perspective, sequential TSVD is particularly attractive in high-dimensional systems. Rather than solving a joint optimization over Q_r, it reduces estimation to a sequence of low-dimensional maximization problems with orthogonality constraints, providing a scalable alternative under both **CSI** and **SSI**.

4.2 Identification Strength and Spectral Stability

Identification in non-Gaussian VARs is governed by the spectral structure of higher-order cumulants. Within the TSVD framework, structural directions correspond to singular vectors of the population cumulant tensor, and identification strength is determined by the separation of the associated tensor singular values.

Let $\mathcal{C}^d(u)$ denote the population dth-order cumulant tensor of the VAR innovations and $\widehat{\mathcal{C}}_T^d(u)$ its sample counterpart. Let $\lambda_{d,1} \geq \cdots \geq \lambda_{d,r}$ denote the nonzero tensor singular values and $q_1, \ldots, q_r$ the corresponding singular directions.

Define the spectral gap

$$\delta_j := \min_{k \neq j} |\lambda_{d,j} - \lambda_{d,k}|, \qquad d \in \{3, 4\}. \tag{19}$$

Large spectral gaps imply uniquely determined structural directions, whereas small gaps generate rotational ambiguity among nearby components. Identification strength can therefore be interpreted as a spectral conditioning property of the cumulant operator.

Definition 8 (Strong identification) *Direction q_j is strongly identified if*

$$\delta_j \geq \underline{\delta} > 0.$$

Definition 9 (Weak identification) *Direction q_j is weakly identified along a sequence of data-generating processes if $\delta_j \to 0$ as $T \to \infty$.*

[9]Formal proofs of consistency and asymptotic normality are provided in the Supplementary Materials.

Thus identification strength is a spectral property of the cumulant tensor. The perturbation expansion in Theorem 6 shows that estimation error arises from projecting sampling perturbations of the cumulant tensor onto neighboring spectral directions. These projections are scaled by inverse spectral gaps, implying that estimation variance is proportional to δ_j^{-2}.

Theorem 10 (Deterministic perturbation bound for the joint TSVD) *Let $\widehat{Q}_{r,T}$ denote the joint TSVD estimator and suppose Assumption C.1 holds. Fix direction q_j with spectral separation (19) and define*

$$E_T := \widehat{\mathcal{C}}_T^d(u) - \mathcal{C}^d(u), \qquad P_j^{\perp} := I - q_j q_j'.$$

Then there exist constants $C, c > 0$ such that whenever $\|E_T\|_F \le c\,\delta_j$,

$$\|P_j^{\perp}(\widehat{q}_{j,T} - q_j)\| \le C\,\frac{\|E_T\|_F}{\delta_j}, \qquad \|\widehat{q}_{j,T} - q_j\| \le C\,\frac{\|E_T\|_F}{\delta_j}.$$

Theorem 10 establishes a Davis–Kahan type stability principle for TSVD estimation. Estimation accuracy depends on the ratio between sampling perturbations of the cumulant tensor and the spectral gap measuring identification strength.

Under standard moment conditions,

$$\|E_T\|_F = O_p(T^{-1/2}),$$

which yields the following stochastic convergence rate.[10]

Theorem 11 (Spectral perturbation rate) *If $\delta_j > 0$, then*

$$\|\widehat{q}_{j,T} - q_j\| = O_p\left(\frac{1}{\delta_j\sqrt{T}}\right).$$

Thus estimation error equals sampling noise scaled by the inverse spectral gap. Small gaps amplify sampling fluctuations and generate weak identification.

This dependence on spectral separation is intrinsic to any method exploiting higher-order moments—including JADE, FastICA, cumulant diagonalization, or moment-based GMM—since all rely on the same population geometry. When non-Gaussian features of structural shocks are similar, the system approaches rotational indeterminacy and variance inflation becomes unavoidable for any consistent estimator.

[10]Since the Frobenius norm is invariant to matricization, the bound holds for any unfolding of the cumulant tensor.

Likelihood-based approaches exhibit the same phenomenon. Structural shocks are identified through distributional features encoding the same cumulant information. Distinct non-Gaussian characteristics produce sharply curved likelihoods and concentrated posteriors, whereas similar characteristics generate nearly flat directions and weak identification.

Hence identification in non-Gaussian SVARs can be interpreted geometrically as a spectral conditioning problem.

Local-to-weak identification

Root-T consistency requires spectral gaps bounded away from zero. When gaps shrink at rate $T^{-1/2}$, the problem enters a local-to-weak identification regime analogous to weak-instrument asymptotics.

Theorem 12 (Local-to-weak spectral degeneracy) *Let $d \in \{3, 4\}$ and suppose that all spectral gaps are bounded away from zero except between $\lambda_{d,1}$ and $\lambda_{d,2}$. Assume*

$$\lambda_{d,1} - \lambda_{d,2} = \frac{c}{\sqrt{T}}, \qquad c \neq 0.$$

Let $\widehat{q}_{1,T}$ denote the unit-norm TSVD singular vector associated with $\widehat{\lambda}_{d,1}$, with sign fixed so that $q_1'\widehat{q}_{1,T} \geq 0$. Define

$$\alpha_{21,T} = \langle \widehat{\mathcal{C}}_T^d(u) - \mathcal{C}^d(u),\, q_2 \circ q_1^{\circ(d-1)} \rangle.$$

Then:

(i) the individual direction is not root-T consistent,

$$P_1^{\perp}(\widehat{q}_{1,T} - q_1) = O_p(1), \qquad P_1^{\perp} := I - q_1 q_1';$$

(ii) the collapsing invariant subspace remains consistently identified,

$$\operatorname{span}\{\widehat{q}_{1,T}, \widehat{q}_{2,T}\} \xrightarrow{p} \operatorname{span}\{q_1, q_2\};$$

(iii) the estimator converges to a random rotation in this subspace: if $\sqrt{T}\alpha_{21,T} \Rightarrow Z$ with $Z \sim \mathcal{N}(0, \sigma_{21}^2)$, then

$$\widehat{q}_{1,T} \Rightarrow \frac{q_1 + (Z/c)\, q_2}{\sqrt{1 + (Z/c)^2}}.$$

The theorem shows that local weak identification corresponds to singular-value degeneracy. As population cumulant singular values approach each other at rate $T^{-1/2}$, individual structural directions become asymptotically unidentified while the

associated spectral subspace remains consistently estimable. Estimation uncertainty therefore takes the form of stochastic rotations within the collapsing spectral subspace rather than Gaussian fluctuations around a fixed direction. Spectral gaps summarize the intrinsic difficulty of non-Gaussian structural identification, playing a role analogous to concentration parameters in weak-instrument econometrics.

ICA methods such as FastICA identify directions by maximizing scalar contrast functions whose local curvature determines identification strength. In contrast, the TSVD framework yields a global geometric characterization in which identification strength is governed by separation of cumulant tensor singular values.

Recent statistical work on tensor singular value decomposition (e.g., Zhang and Xia, 2020) studies recovery of low-rank tensors under noise. In contrast, the present analysis focuses on structural identification: the cumulant tensor arises from economic independence restrictions, and singular-value separation governs identification strength in non-Gaussian SVARs.

4.3 Completing Identification after TSVD-Based Subsystem Recovery

Subsystem identification obtained from TSVD can be combined with standard SVAR restrictions to achieve complete system identification. The key geometric insight is that TSVD identifies a structural *subspace*, while conventional restrictions determine a rotation within its orthogonal complement.

Under subsystem identification (**SSI**), higher-order cumulants identify an r-dimensional orthogonal subspace

$$Q_r \in \mathbb{R}^{n\times r}, \qquad r < n,$$

so that the orthogonal matrix governing whitened innovations admits the decomposition

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

The columns of Q_r correspond to shocks identified through non-Gaussianity, while Q_{n-r} collects directions that remain unidentified by higher-order cumulants. The matrix Q_r is therefore point identified (up to sign and permutation), whereas Q_{n-r} spans the remaining rotational indeterminacy. Because cumulants depend only on $\text{span}(Q_r)$, additional identifying restrictions can be imposed exclusively on Q_{n-r}.

Proposition 13 (Completion of TSVD-based identification) *Let TSVD identify an r-dimensional structural subspace spanned by Q_r, and write $Q = [\, Q_r \;\; Q_{n-r} \,]$ orthogonal. Then:*

1. *complete identification is achieved by imposing restrictions on Q_{n-r} alone;*
2. *impulse responses associated with Q_r are invariant to the completion step;*
3. *under strong subsystem identification, the asymptotic distribution of estimators for Q_r is unaffected by completion.*

To construct the completion, partition the impact matrix as

$$\Theta = [\Theta_r \; \Theta_{n-r}],$$

where $\text{span}(\Theta_r) = \text{span}(Q_r)$. Let Q_{n-r} span the orthogonal complement of this space. Projecting the covariance matrix yields

$$M = Q'_{n-r}\Sigma Q_{n-r} = Q'_{n-r}\Theta_{n-r}\Theta'_{n-r}Q_{n-r},$$

which admits a factorization $M = LL'$. All admissible completions therefore satisfy

$$\Theta_{n-r} = Q_{n-r}LR, \qquad R \in \mathbb{R}^{(n-r)\times(n-r)} \text{ orthogonal},$$

so that the remaining indeterminacy reduces to a rotation within the complement space.

Standard SVAR restrictions—recursive, sign, or long-run—may then be imposed on Q_{n-r} (or the associated impulse responses) without affecting the shocks identified by TSVD. Identification is therefore naturally separated into two orthogonal steps: higher-order moments identify the statistically determined shocks, while economic restrictions select the rotation of the remaining ones. The resulting impact matrix satisfies $\Sigma = \Theta\Theta'$ by construction and integrates cumulant-based and conventional SVAR identification without requiring joint nonlinear optimization.

5. Simulation Evidence

This section evaluates the finite-sample implications of the spectral identification framework developed in the paper. The simulation design targets the central theoretical prediction: estimation accuracy is governed by the separation of tensor singular values of the population cumulant tensor.

Across experiments, identification strength is varied through controlled changes in higher-order cumulants while second-moment dynamics are held fixed. This design isolates the role of spectral gaps in determining estimation stability, convergence behavior, and recovery of invariant subsystems. We compare joint and sequential TSVD estimators with alternative non-Gaussian identification methods. All results are based on $N = 10{,}000$ Monte Carlo replications.

5.1 Spectral-gap experiment

To quantify the role of identification strength predicted by the spectral theory in Section 4.2, we conduct a Monte Carlo experiment in which the separation between population cumulant singular values is controlled directly.

Innovations are generated as

$$\nu_t = Q(\alpha)\epsilon_t, \qquad Q(\alpha) = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix}, \qquad \alpha = -\pi/5,$$

following Gouriéroux et al. (2017), where ϵ_t are the bivariate structural shocks. The target parameter is $q_{11} = \cos(\alpha) = 0.809$, and sample sizes are $T \in \{200, 500, 5000\}$.

The first structural shock has fixed degrees of freedom $\nu_1 = 12$, while the second shock is calibrated so that its excess kurtosis satisfies

$$\kappa_2 = \kappa_1 + \Delta.$$

Because cumulant singular values are proportional to excess kurtosis, Δ determines the spectral gap governing identification strength. We consider $\Delta \in \{1, 0.5, 0.2, 0.05, 0.01\}$, spanning regimes from strong to weak identification.

We compare the proposed TSVD estimator with ICA-type procedures: the higher-order singular value decomposition (HOSVD), the FastICA estimator of Hyvärinen (1999), and the joint approximate diagonalization of eigenmatrices (JADE) proposed by Cardoso (1989).[11] Restricting attention to this class isolates the role of spectral separation, abstracting from differences in likelihood weighting or distributional assumptions present in GMM or likelihood-based approaches.

[11]The JADE algorithm (Cardoso, 1989) diagonalizes a collection of fourth-order cumulant matrices using Jacobi rotations. We implemented JADE using the `JadeR.m` MATLAB file provided by Jean-François Cardoso (2013). For FastICA, we used the MATLAB toolbox developed by H. Gävert, J. Hurri, J. Särelä, and A. Hyvärinen, version 2.5 (2005). See the Supplementary Materials for a detailed description of HOSVD.

Table 1 reports bias and the angle-based root mean-squared error

$$\text{Angle-RMSE} = \sqrt{\mathbb{E}\big[\|P_1^{\perp}(\hat{q}_{1,T} - q_1)\|^2\big]},$$

which measures rotational distance between estimated and true directions and is invariant to sign and permutation. Bias remains small across all designs and declines with sample size, indicating that weaker identification primarily affects precision rather than systematic distortion.

In contrast, Angle-RMSE increases markedly as Δ decreases, reflecting the approach to the local-to-weak identification regime characterized in Theorem 12, where structural directions become difficult to distinguish as cumulant singular values converge.

This behavior follows from Theorems 10 and 11, which imply that estimation error scales with sampling noise and the inverse spectral gap. Consequently, Angle-RMSE increases with $1/\Delta$ and declines with $\sqrt{T}$, highlighting the joint role of identification strength and sample size.

Across estimators, TSVD exhibits uniformly smaller Angle-RMSE, particularly when the spectral gap is small, while maintaining negligible bias. FastICA and JADE deteriorate more rapidly as singular values become nearly degenerate, illustrating that weak identification manifests as rotational instability rather than optimization failure.

Overall, the experiment confirms that weak identification in non-Gaussian SVARs is fundamentally a spectral conditioning problem: estimation error equals sampling noise amplified by the inverse singular-value gap.

5.2 Complete System Identification

We next consider settings in which all structural shocks are non-Gaussian and the orthogonal mixing matrix is point identified (up to sign and permutation). Innovations are generated as in the previous experiment,

$$\nu_t = Q(\alpha)\epsilon_t,$$

where $Q(\alpha)$ is the same Givens rotation with $\alpha = -\pi/5$. Structural shocks are independent and standardized, and identification arises exclusively through differences in excess kurtosis.

We consider three data-generating specifications that preserve complete identification while varying the degree and nature of non-Gaussianity. The first specification combines Student-$t(5)$ and Student-$t(12)$ shocks, producing a pronounced contrast

in tail thickness. The second pairs Student-$t(7)$ and Student-$t(20)$ shocks, yielding a more moderate separation in excess kurtosis. The third combines a Student-$t(12)$ shock with a hyperbolic secant shock, introducing non-Gaussianity arising from distributions with qualitatively different tail behavior.

Across these designs, independence and standardization are maintained, while variation in higher-order cumulants provides the sole source of identification strength. Because singular values of the population cumulant tensor are proportional to excess kurtosis in this setting, differences in tail thickness directly translate into spectral separation and therefore determine identification strength.

For each experiment, we compare the TSVD estimator with alternative approaches for estimating the orthogonal mixing matrix. In settings with complete identification driven by excess kurtosis, we consider the pseudo–maximum likelihood (PML) estimator of Gouriéroux et al. (2017), HOSVD, FastICA, and JADE. We additionally include a GMM estimator using an identity weighting matrix as well as an optimally weighted two-step GMM estimator.[12]

This comparison contrasts estimators based on maximizing non-Gaussianity, as in the TSVD framework, with approaches relying on minimization of higher-order cross-moments, such as GMM and joint diagonalization methods.

Table 2 reports bias and root mean squared error (RMSE). Across nearly all configurations, TSVD delivers the smallest RMSE, particularly in small and moderate samples. When the likelihood is correctly specified and the sample size is large, PML slightly outperforms TSVD, consistent with its asymptotic efficiency. However, TSVD exhibits substantial robustness gains in finite samples, reducing RMSE by up to 50% relative to competing estimators. FastICA and JADE typically outperform HOSVD but remain dominated by TSVD. GMM estimators improve under optimal weighting but remain less accurate overall, reflecting the higher dimensionality of their moment conditions. [13]

[12]The implemented GMM estimator is fully described in the Supplementary Materials. We also considered the GMM estimator of Lanne and Luoto (2021), whose small-sample performance was generally comparable or inferior in terms of both bias and RMSE. Their moment conditions are formulated in terms of structural shocks, reducing the number of moments to estimate but requiring inversion of the impact matrix Θ, which may behave poorly in small samples; see Lewis (2025).

[13]Additional simulations with skewness are reported in the Supplementary Materials.

5.2.1 Role of Objective Dimensionality

To examine how estimator performance varies with the dimensionality of the optimization problem, we consider a trivariate system in which the orthogonal mixing matrix is parameterized by three Givens rotations,

$$Q(\alpha) = Q_1(\alpha)Q_2(\alpha)Q_3(\alpha),$$

with $\alpha = -\pi/5$. Structural shocks are independent Student-t with degrees of freedom $(5, 9, 12)$, ensuring identification through fourth-order cumulants.

In addition to TSVD and conventional GMM, we consider two minimum-distance estimators based on the orthogonally decomposable representation of the fourth-order cumulant tensor, denoted OD–GMM–I and OD–GMM–opt. Both estimators match the sample fourth-order cumulants to the orthogonally decomposable structure implied by $Q(\alpha)$ and a vector of diagonal kurtosis parameters (Proposition 4 and equation (15)), using the 15 distinct reduced-form fourth-order cumulant elements as moment conditions. OD–GMM–I employs the identity weighting matrix, whereas OD–GMM–opt uses a data-driven Newey–West estimate of the optimal weighting matrix.

All minimum-distance estimators impose orthonormality of Q through the same low-dimensional parameterization and estimate Q by minimizing a quadratic form in the 15 cumulant moments. They differ only in how the orthogonal decomposition restrictions enter the moment conditions and in the choice of weighting matrix. By contrast, TSVD does not minimize a quadratic loss over the full cumulant vector. Instead, it maximizes a low-dimensional spectral criterion based on diagonal projections of the cumulant tensor. In the trivariate case, this concentrates identifying information into three directional kurtosis components—one for each column of Q—rather than treating all 15 cumulant entries symmetrically.

Table 3 reports finite-sample bias and RMSE for TSVD, conventional GMM, and OD–GMM under this design. TSVD consistently achieves the lowest bias and RMSE across all sample sizes, reflecting the efficiency gains from optimizing a low-dimensional objective with strong curvature. Conventional GMM performs poorly in small samples despite imposing orthonormality of Q, with bias and RMSE inflated by estimation and weighting of a high-dimensional vector of cumulant moments. OD–GMM exhibits intermediate performance: relative to conventional GMM, it benefits from explicitly enforcing the orthogonal tensor structure, yet remains less efficient

than TSVD because it still minimizes a quadratic form over all 15 fourth-order moments.

Overall, these results indicate that the primary source of small-sample inefficiency in cumulant-based estimation is not the absence of orthogonality restrictions—which are imposed throughout—but the dimensionality of the objective function itself. Estimators that concentrate identifying information into a small number of informative spectral directions, as in TSVD, achieve substantial efficiency gains relative to minimum-distance approaches, even when those approaches exploit the same orthogonal structure and employ optimal weighting matrices.

5.3 Subsystem Identification

We now consider subsystem identification, in which only a subset of structural shocks is non-Gaussian and therefore only a structural subspace—rather than the full orthogonal matrix—is point identified.

The data are generated from the VAR(1)

$$x_t = \Gamma x_{t-1} + v_t, \qquad v_t = \Theta \epsilon_t, \tag{20}$$

where Γ is a stable $n \times n$ matrix, $\Theta \in \mathbb{R}^{n \times n}$ is the structural impact matrix, and ϵ_t contains mutually independent structural shocks with $\mathbb{E}[\epsilon_t] = 0$ and $\mathrm{Var}(\epsilon_t) = I_n$. A burn-in of 300 observations is used.

Identification relies on fourth-order cumulants (excess kurtosis). The first two shocks are heavy-tailed while the remaining $n - 2$ shocks are Gaussian:

$$\begin{aligned} &\epsilon_{it} \sim \text{standardized Student-}t(\nu_i), \qquad i = 1, 2, \\ &\epsilon_{jt} \sim \mathcal{N}(0, 1), \qquad j = 3, \ldots, n. \end{aligned}$$

independently across time and components. The Student-t shocks are rescaled to ensure $\mathrm{Var}(\epsilon_t) = I_n$.

We set

$$\Gamma = 0.35 I_n + \frac{0.10}{n-1}(\mathbf{1}\mathbf{1}' - I_n),$$

where $\mathbf{1}$ is the $n \times 1$ vector of ones, implying diagonal elements 0.35 and off-diagonal elements $0.10/(n-1)$. Unless otherwise stated, the baseline simulations use $n = 4$[14].

[14] Additional experiments were conducted for larger systems ($n = 5, \ldots, 8$), alternative numbers of non-Gaussian shocks (r), and different persistence levels of the VAR(1) matrix Γ. Across all configurations, the qualitative ranking of estimators and the main conclusions of the simulations remain unchanged.

We consider two identification regimes: stronger non-Gaussianity $(\nu_1, \nu_2) = (5, 12)$ and weaker $(7, 20)$.

The impact matrix Θ is constructed so that its first two columns,

$$\Theta_{12} \equiv [\theta_1 \; \theta_2] \in \mathbb{R}^{n \times 2},$$

span the non-Gaussian structural subspace. In practice, Θ is generated as an orthogonal matrix obtained from a sequence of Givens rotations, allowing controlled geometry of the identification problem while preserving orthogonality.

Because non-Gaussian structural directions are identified only up to sign and permutation within the non-Gaussian block, each estimate is aligned prior to computing performance measures. For every Monte Carlo replication, admissible permutations of the $r = 2$ columns are considered and sign flips are applied so that each estimated column has a nonnegative inner product with its assigned population counterpart. Among all admissible transformations, we retain the aligned estimate minimizing the Frobenius distance $\|\widehat{\Theta}_{12} - \Theta_{12}\|_F$.

For each sample size T and estimator, accuracy is summarized using column-wise bias and RMSE. To estimate the non-Gaussian structural subspace, we compare tensor-based methods (HOSVD and TSVD), ICA-type benchmarks (FastICA and JADE), and two orthogonally decomposable GMM procedures (OD–GMM with identity weighting and OD–GMM with an estimated optimal weighting matrix). To ensure comparability and reduce sensitivity to local optima, all iterative estimators are initialized using the HOSVD solution. In the subsystem setting, standard GMM cannot be implemented without additional identifying restrictions.

Table 4 reports finite-sample results. Tensor-based estimators dominate alternatives across all configurations. TSVD consistently improves upon HOSVD, particularly for smaller samples ($T = 200$ and $T = 500$). Performance deteriorates as excess kurtosis weakens, in line with the spectral-gap theory developed in Section 4.2. JADE performs poorly and exhibits limited convergence as sample size increases, while OD–GMM estimators display higher sampling variability, especially under optimal weighting.

Overall, the simulation evidence supports the spectral identification framework developed in this paper. Concentrating higher-order information through TSVD yields stable and accurate recovery of the non-Gaussian structural subspace across identification regimes, sample sizes, and degrees of non-Gaussianity.

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 Fiscal policy shocks under subsystem identification

As a first application, we revisit the effects of fiscal policy on economic activity in the trivariate SVAR framework of Blanchard and Perotti (2002). This setting is particularly relevant because identification from higher-order cumulants typically yields *subsystem* identification rather than full system identification. In our data, rank tests (see Guay, 2021) reject Gaussianity only through excess kurtosis for a single structural shock, implying that only one structural direction is identified by non-Gaussianity. We therefore use TSVD to estimate directly the impact vector associated with the non-Gaussian shock, and then complete identification of the remaining shocks by imposing a conventional short-run restriction.

Consider the trivariate SVAR

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

where $\nu_{\tau,t}$, $\nu_{g,t}$, and $\nu_{y,t}$ are reduced-form innovations to taxes, government spending, and output, and $\epsilon_{1,t}$, $\epsilon_{2,t}$, $\epsilon_{3,t}$ are structural shocks. The sample uses quarterly U.S. data from 1980Q1–2015Q3; the reduced form is a VAR with a deterministic trend and eight lags (see Guay (2021)).

Table 5 reports TSVD estimates. The identified impact vector loads almost entirely on the tax innovation: $\widehat{\theta}_{11} = 0.0471$ is precisely estimated, whereas $\widehat{\theta}_{21}$ and $\widehat{\theta}_{31}$ are small and insignificant. Thus, the non-Gaussian shock primarily affects taxes at impact and is naturally interpreted as an exogenous tax shock.

To obtain a fully identified system, we fix the remaining 2×2 rotation in the orthogonal complement of the TSVD-identified direction using the short-run restriction $\theta_{23} = 0$ (with similar results under $\theta_{32} = 0$) as investigated by Blanchard and Perotti (2002). Columns 2–3 of Table 5 report the corresponding conventional GMM estimates using either $W = I$ or an estimated optimal weighting matrix. Both approaches deliver impact estimates and tax multipliers that are very similar to TSVD, although

the efficient-weight version is less precise, consistent with the weak finite-sample performance of optimal weighting matrices documented in Section 5.3. This loss of precision is also visible in the estimation of the excess kurtosis: the TSVD-based 95% confidence interval for $\mathcal{C}_{1,1,1,1}(\epsilon)$ is $(0.952, 3.754)$, compared to $(0.155, 10.031)$ under GMM with $W = I$ and $(0.389, 4.724)$ with an estimated optimal weighting matrix.

Across all identification strategies, the implied tax multipliers are modest: the multiplier is essentially zero at impact and peaks around 0.6 after 14 quarters (Table 5). Importantly, these conclusions follow directly from the TSVD-identified subsystem, without requiring additional restrictions to identify the tax shock itself.

Finally, TSVD exhibits superior numerical stability. Estimates are insensitive to initialization, whereas GMM procedures frequently converge to alternative local minima and produce materially different impact matrices. Tensor-based identification thus delivers greater computational robustness in empirical applications.

6.2 Effects of Credit Shocks

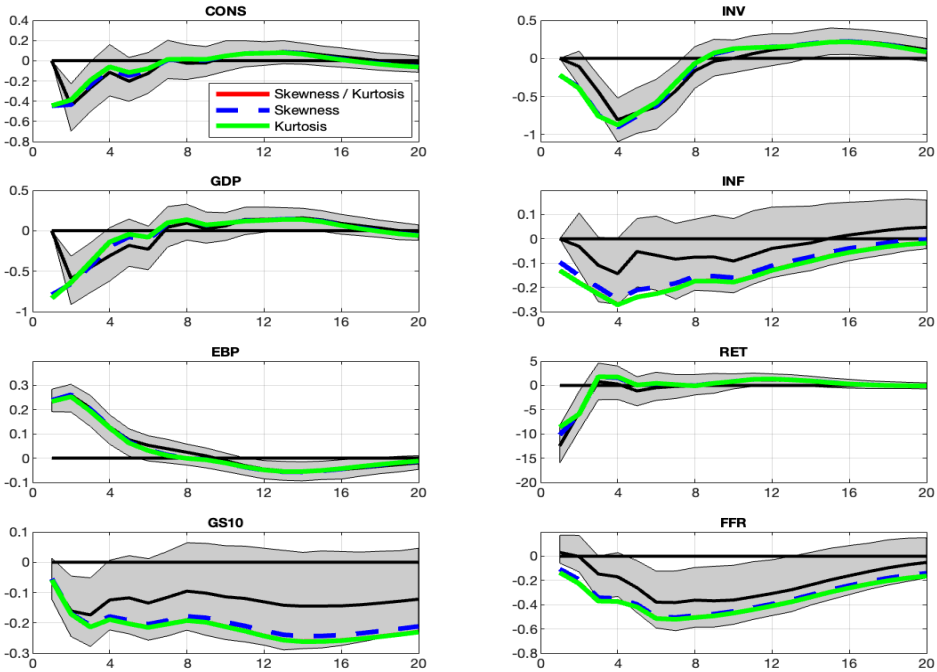
We reevaluate the macroeconomic effects of credit shocks using the quarterly eight-variable U.S. VAR of Gilchrist and Zakrajšek (2012), including real activity, inflation, financial variables, and interest rates. The excess bond premium (EBP) serves as a proxy for financial conditions. The sample spans 1973Q1–2019Q4, and the VAR is estimated with four lags.

This setting provides a demanding environment for higher-order identification. In an $n = 8$ system, the fourth-order cumulant tensor contains 330 distinct elements, rendering moment-based GMM computationally burdensome and potentially unstable. By contrast, TSVD targets the identified non-Gaussian subspace directly and remains tractable. Macro-financial variables are particularly suitable for this approach, as credit spreads exhibit pronounced skewness and fat tails.

Rank tests, reported in Table 6, indicate that non-Gaussianity is concentrated in a limited number of directions. We therefore implement subsystem identification with $r = 5$ components. Consistent with the spectral identification framework of Section 4.2, the first singular value is much larger than the remaining ones, implying a large spectral gap and therefore strong identification of the corresponding structural direction.

Figure 1 compares impulse responses to a credit shock under recursive identification (Gilchrist and Zakrajšek, 2012) and TSVD-based identification. Both approaches

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.

imply that tighter credit conditions reduce real activity and inflation and trigger an accommodative monetary policy response. The leading TSVD shock is correlated (approximately 65%) with the Cholesky-identified EBP innovation, indicating that the dominant non-Gaussian component captures a credit-related disturbance.

Quantitative differences arise at short horizons. Recursive identification imposes zero-impact restrictions on real activity and inflation, whereas TSVD imposes no timing assumptions, allowing immediate responses. These impact effects are consistent with DSGE models featuring financial frictions, in which credit conditions affect demand and costs contemporaneously (Christiano et al., 2014; Del Negro et al., 2015). Both skewness and excess kurtosis contribute materially to identification.

Overall, higher-order cumulants contain informative variation for credit-shock identification. The TSVD approach recovers an economically interpretable disturbance closely related to the credit shock in Gilchrist and Zakrajšek (2012), while delivering short-run dynamics free of recursive timing restrictions.

7. Conclusion

This paper develops a spectral framework for identification and estimation in non-Gaussian SVARs. Structural directions correspond to singular vectors of orthogonally decomposable cumulant tensors, and identification strength is governed by singular-value separation. This perspective unifies complete and subsystem identification within a common geometric structure and links weak identification and finite-sample instability to spectral degeneracy.

Within this framework, tensor singular value decomposition (TSVD) provides a transparent and computationally tractable estimator. We establish asymptotic properties for joint and sequential TSVD estimators and show that estimation accuracy is determined by spectral gaps, so that inference reduces to a spectral perturbation problem. Monte Carlo and empirical evidence confirm the theoretical predictions and illustrate stable recovery of structural directions across identification regimes.

Extensions include robust cumulant estimation, inference robust to spectral degeneracy, and applications to high-dimensional, time-varying, or nonlinear macroeconomic environments.

References

Anandkumar, A., R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky (2014). Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research 15*, 2772–2932.

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

Babii, A., E. Ghysels, and J. Pan (2025). Tensor PCA for factors. *Journal of Econometrics In Press*.

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

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

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

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

Chen, J. and Y. Saad (2009). 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., R. Motto, and M. Rostagno (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 J. G. McWhirter and I. K. Proudler (Eds.), *Mathematics in Signal Processing V*, pp. 1–24. Oxford, UK: Clarendon Press. Lecture Notes in Mathematics; survey chapter.

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

1278.

De Lathauwer, L., B. De Moor, and J. Vandewalle (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., M. P. Giannoni, and F. Schorfheide (2015). Inflation in the Great Recession and New Keynesian Models. *American Economic Journal: Macroeconomics 7*(1), 168–96.

Drautzburg, T. and J. H. Wright (2023). Refining set-identification in VARs through independence. *Journal of Econometrics 235*(2), 1827–1847. .

Gilchrist, S. and E. Zakrajšek (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review 102*(4), 1692–1720.

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

Gourieroux, C. and J. Jasiak (2025). Generalized Covariance-Based Inference for Models Set-Identified from Independence Restrictions. *Journal of Time Series Analysis 46*(2), 300–324. Published online 24 September 2024.

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

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

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.

Jondeau, E., E. Jurczenko, and M. Rockinger (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.

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

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

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

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

Lewis, D. J. (2025). Identification Based on Higher Moments in Macroeconometrics. *Annual Review of Economics 17*, 665–693. Review Article.

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

Montiel Olea, J. L., M. Plagborg-Møller, and E. Qian (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 L. Phaneuf (2004). Monetary Policy Shocks: Testing Identification Conditions under Time-Varying Conditional Volatility. *Journal of Monetary Economics 51*(6), 1217–1243.

Sentana, E. and G. Fiorentini (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.

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

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.

Table 1: Finite-sample distributions in the complete identification case: excess kurtosis with identification strength controlled by Δ.

	Identification strength Δ				
Method	1	0.5	0.2	0.05	0.01
	Panel (A) – $T = 200$				
	Bias: $\mathbb{E}(\hat{q}_{11} - q_{11})$				
HOSVD	-0.038	-0.043	-0.056	-0.056	-0.060
TSVD	0.035	0.042	0.047	0.052	0.055
FastICA	-0.043	-0.045	-0.049	-0.057	-0.054
JADE	-0.032	-0.040	-0.050	-0.049	-0.050
	Angle-RMSE: $\sqrt{\mathbb{E}\left[\|P_1^{\perp}(\hat{q}_{1,T} - q_1)\|^2\right]}$				
HOSVD	0.306	0.317	0.336	0.336	0.340
TSVD	0.221	0.236	0.246	0.254	0.258
FastICA	0.297	0.319	0.327	0.344	0.348
JADE	0.268	0.289	0.304	0.316	0.318
	Panel (B) – $T = 500$				
	Bias: $\mathbb{E}(\hat{q}_{11} - q_{11})$				
HOSVD	-0.030	-0.040	-0.028	-0.036	-0.038
TSVD	0.012	0.018	0.020	0.022	0.023
FastICA	-0.011	-0.016	-0.024	-0.029	-0.028
JADE	-0.012	-0.013	-0.016	-0.023	-0.026
	Angle-RMSE: $\sqrt{\mathbb{E}\left[\|P_1^{\perp}(\hat{q}_{1,T} - q_1)\|^2\right]}$				
HOSVD	0.263	0.284	0.287	0.287	0.290
TSVD	0.147	0.165	0.178	0.183	0.184
FastICA	0.194	0.213	0.238	0.245	0.255
JADE	0.176	0.191	0.211	0.218	0.223
	Panel (C) – $T = 5000$				
	Bias: $\mathbb{E}(\hat{q}_{11} - q_{11})$				
HOSVD	-0.004	-0.013	-0.026	-0.020	-0.028
TSVD	0.000	-0.001	-0.000	0.001	0.000
FastICA	-0.001	-0.003	-0.001	-0.001	-0.005
JADE	0.000	-0.001	0.000	0.001	0.000
	Angle-RMSE: $\sqrt{\mathbb{E}\left[\|P_1^{\perp}(\hat{q}_{1,T} - q_1)\|^2\right]}$				
HOSVD	0.115	0.178	0.245	0.268	0.282
TSVD	0.050	0.052	0.053	0.053	0.055
FastICA	0.085	0.074	0.082	0.090	0.093
JADE	0.050	0.052	0.053	0.053	0.055

Notes: $\epsilon_{1,t} \sim t(\nu_1)$ with $\nu_1 = 12$. $\epsilon_{2,t} \sim t(\nu_2)$ where $\nu_2 = 4+6/\kappa_2$ and $\kappa_2 = \kappa_1+\Delta$, $\kappa(\nu) = 6/(\nu-4)$. The mixing matrix is a rotation with $\alpha = -0.628319$ so $q_{11} = \cos(\alpha) = 0.809$. Angle-RMSE measures the subspace error $\|P_1^{\perp}(\hat{q}_{1,T} - q_1)\|$.

Table 2: Finite-sample distributions in the CSI case: Excess kurtosis

	T = 200			T = 500			T = 5000		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Panel (a) - Bias									
PML(1)	-0.013	-0.024	-0.013	-0.004	-0.012	-0.002	-0.000	-0.001	-0.001
PML(2)	-0.013	-0.023	-0.014	-0.004	-0.011	-0.003	-0.000	-0.001	-0.000
PML(3)	-0.010	-0.026	-0.016	-0.001	-0.008	-0.005	0.000	0.000	-0.000
HOSVD	-0.026	-0.035	-0.033	-0.015	-0.020	-0.020	-0.001	-0.002	-0.002
TSVD	0.020	0.040	0.024	0.003	0.011	0.004	-0.001	-0.001	-0.001
FastICA	-0.021	-0.036	-0.026	-0.007	-0.015	-0.010	-0.003	-0.003	-0.002
JADE	-0.015	-0.032	-0.020	-0.006	-0.012	-0.008	-0.001	-0.001	-0.001
GMM-I	0.004	-0.008	0.006	0.003	0.002	0.006	-0.001	-0.001	-0.001
GMM-opt	0.003	-0.003	0.004	0.001	-0.001	0.002	-0.000	-0.001	-0.001
Panel (b) - Root mean-squared errors									
PML(1)	0.113	0.157	0.119	0.060	0.101	0.072	0.016	0.024	0.022
PML(2)	0.113	0.155	0.121	0.060	0.098	0.073	0.016	0.023	0.022
PML(3)	0.118	0.167	0.120	0.070	0.108	0.068	0.021	0.031	0.018
HOSVD	0.153	0.181	0.169	0.109	0.134	0.132	0.028	0.033	0.038
TSVD	0.081	0.100	0.086	0.058	0.073	0.063	0.024	0.029	0.024
FastICA	0.141	0.184	0.155	0.084	0.121	0.094	0.050	0.051	0.047
JADE	0.124	0.172	0.138	0.075	0.106	0.084	0.025	0.030	0.024
GMM-I	0.129	0.176	0.137	0.081	0.112	0.090	0.029	0.034	0.028
GMM-opt	0.128	0.176	0.136	0.073	0.108	0.079	0.023	0.028	0.022

Notes: The first row of the table indicates the generating distributions of the ϵ_ts: (1) $\epsilon_{1,t} \sim t(5)$ and $\epsilon_{2,t} \sim t(12)$; (2) $\epsilon_{1,t} \sim t(7)$ and $\epsilon_{2,t} \sim t(20)$; (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 $q_{11} = \cos(\alpha)$, $q_{21} = -\sin(\alpha)$, $q_{12} = \sin(\alpha)$, and $q_{22} = \cos(\alpha)$ with $\alpha = -\pi/5$ (so $q_{11} = 0.809$). PML(1), PML(2) and PML(3) indicate the sets of distributions used for the pseudo maximum likelihood. Panel (a) reports the biases $\mathbb{E}(\hat{q}_{11,T} - q_{11})$ and Panel (b) reports root-mean-squared errors $\sqrt{\mathbb{E}(\hat{q}_{11,T} - q_{11})^2}$.

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
OD-GMM-I	0.026	0.016	0.009
OD-GMM-opt	0.016	0.007	0.000
GMM-I	0.072	0.046	0.008
GMM-opt	0.094	0.051	0.007
Panel (b) — Root Mean Squared Error			
TSVD	0.129	0.075	0.026
OD-GMM-I	0.163	0.106	0.041
OD-GMM-opt	0.181	0.108	0.027
GMM-I	0.241	0.199	0.077
GMM-opt	0.271	0.201	0.073

Notes. 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 $\alpha = -\pi/5$, yielding $q_{11} = \cos(\alpha) = 0.809$. Bias is defined as $\mathbb{E}[\hat{q}_{11} - q_{11}]$, and RMSE as $\sqrt{\mathbb{E}[(\hat{q}_{11} - q_{11})^2]}$.

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

	T = 200		T = 500		T = 5000	
Panel (a) – $\epsilon_{1t} \sim t(5)$, $\epsilon_{2t} \sim t(7)$						
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
Bias						
HOSVD	.125	.313	.044	.171	.004	.007
TSVD	.115	.305	.0315	.158	.003	.006
FastICA	.134	.330	.0349	.177	.003	.006
JADE	.440	.421	.424	.415	.407	.406
OD-GMM-I	.131	.339	.401	.204	.003	.019
OD-GMM-opt	.305	.408	.212	.342	.003	.045
Root mean-squared errors						
HOSVD	.500	.792	.297	.584	.085	.114
TSVD	.479	.781	.250	.562	.077	.105
FastICA	.517	.812	.264	.595	.075	.104
JADE	.938	.918	.921	.911	.902	.901
OD-GMM-I	.511	.823	.283	.639	.080	.191
OD-GMM-opt	.781	.903	.651	.827	.245	.299

Table 4 continued

	T = 200		T = 500		T = 5000	
Panel (b) – $\epsilon_{1t} \sim t(7)$, $\epsilon_{2t} \sim t(20)$						
	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
Bias						
HOSVD	.196	.418	.080	.358	.005	.038
TSVD	.187	.413	.071	.352	.004	.032
FastICA	.216	.433	.083	.379	.005	.034
JADE	.428	.434	.409	.426	.390	.415
OD-GMM-I	.206	.429	.0843	.383	.004	.061
OD-GMM-opt	.342	.436	.256	.430	.022	.109
Root mean-squared errors						
HOSVD	.626	.915	.401	.846	.096	.277
TSVD	.612	.909	.377	.839	.092	.238
FastICA	.657	.931	.408	.870	.099	.262
JADE	.925	.931	.905	.923	.883	.911
OD-GMM-I	.641	.926	.410	.875	.093	.349
OD-GMM-opt	.826	.933	.715	.927	.206	.466

Notes. In Panel (a), the two non-Gaussian shocks follow standardized Student-t distributions with degrees of freedom $(5, 7)$, whereas in Panel (b) they follow $(7, 20)$. All estimators are reported after aligning $(\hat{\theta}_1, \hat{\theta}_2)$ to (θ_1, θ_2) up to sign and permutation within the non-Gaussian block.

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

	Parameter estimates		
Parameter	TSVD	$W = I$	Efficient W
θ_{11}	0.0471^{***}	0.0470^{***}	0.0476^{***}
θ_{12}	0.0045	0.0029	0.0016
θ_{13}	0.0098	0.0110	0.0080
θ_{21}	−0.0003	0.0000	0.0002
θ_{22}	0.0068^{***}	0.0068^{***}	0.0069^{***}
θ_{23}	$0.0000^{\dagger}$	$0.0000^{\dagger}$	$0.0000^{\dagger}$
θ_{31}	−0.0002	−0.0002	0.0001
θ_{32}	0.0017^{**}	0.0017^{***}	0.0017^{**}
θ_{33}	0.0048^{***}	0.0048^{***}	0.0048^{***}
$\mathcal{C}_{1,1,1,1}(\epsilon)$	2.7919^{**}	2.9118^{***}	2.7525^{**}

	Parameter estimates (continued)		
	Tax multiplier		
Quarter	TSVD	$W = I$	Efficient W
1	0.03	0.04	−0.02
4	0.10	0.13	0.02
8	0.28	0.31	0.21
Peak	0.63	0.66	0.59
	[14]	[14]	[14]

Notes. The tax multiplier measures the dollar change in output at a given horizon that results from a one-dollar decrease (increase) in the exogenous component of taxes. *, **, and *** indicate, respectively, that the 90, 95, and 99 percent confidence intervals exclude zero, based on 5,000 bootstrap replications. † indicates a constrained parameter. Bracketed values indicate the quarter at which the peak multiplier is attained.

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

	Skewness			Kurtosis			Skewness / Kurtosis		
r	Wald stat	CV (5%)	CV (1%)	Wald stat	CV (5%)	CV (1%)	Wald stat	CV (5%)	CV (1%)
0	2291.9	858.3	950.3	77519.3	9701.4	11390.5	79811.2	10368.9	12221.9
1	1196.0	655.7	720.6	23632.3	6877.3	7610.4	24843.9	7498.2	8374.3
2	868.1	529.2	578.6	15267.2	5881.3	6556.4	16168.2	6408.0	7084.0
3	602.9	435.5	478.9	9061.8	4630.7	5035.8	9729.6	5065.7	5623.8
4	410.5	334.9	367.7	4985.4	3845.6	4216.6	5456.4	4245.6	4648.4
5	261.0	248.3	276.4	3059.4	2843.0	3117.6	3368.0	3098.1	3383.7
6	128.3	165.2	183.1	1754.6	1958.2	2162.8	1949.9	2182.5	2434.7
7	59.6	78.8	88.7	659.4	912.1	1001.5	756.6	1045.2	1125.6
λ	[5.9, 1.8, 1.4, 1.0, 0.8, 0.7, 0.4, 0.3]			[294.5, 45.7, 33.9, 22.3, 10.5, 7.1, 5.9, 3.6]			[300.4, 47.4, 35.2, 23.3, 11.4, 7.7, 6.5, 4.1]		

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. Rows $r = 0$ to $r = 7$ indicate the number of non-Gaussian structural shocks under the null. The fourth line in each block reports the singular values (λ) of the corresponding cumulant.

Appendix

Proof of Proposition 1. Let $u_t = Q\epsilon_t$ with Q orthogonal. By multilinearity of cumulants, for any indices $i_1, \ldots, i_d$,

$$\mathcal{C}^d_{i_1,\ldots,i_d}(u) = \mathrm{Cum}(u_{i_1 t}, \ldots, u_{i_d t}) = \sum_{j_1,\ldots,j_d=1}^{n} q_{i_1 j_1} \cdots q_{i_d j_d}\, \mathcal{C}^d_{j_1,\ldots,j_d}(\epsilon).$$

Under Assumption 1(ii) or Assumption 2(i), all cross-cumulants of order $d \in \{3, 4\}$ vanish, so $\mathcal{C}^d_{j_1,\ldots,j_d}(\epsilon) = 0$ unless $j_1 = \cdots = j_d = j$. Hence

$$\mathcal{C}^d_{i_1,\ldots,i_d}(u) = \sum_{j=1}^{r} \mathcal{C}^d_{j,\ldots,j}(\epsilon)\, q_{i_1 j} \cdots q_{i_d j} = \sum_{j=1}^{r} \lambda_{d,j}\, (q_j^{\circ d})_{i_1,\ldots,i_d},$$

where $\lambda_{d,j} := \mathcal{C}^d_{j,\dots,j}(\epsilon)$ and q_j is the jth column of Q. Since this holds for all $(i_1, \dots, i_d)$, we obtain

$$\mathcal{C}^d(u) = \sum_{j=1}^{r} \lambda_{d,j}\, q_j^{\circ d}.$$

Proof of Proposition 3. Let $\mathbf{A}$ denote a matricization (mode unfolding) of the symmetric order-N array $\mathcal{A}$. Under the symmetric TSVD,

$$\mathcal{A} = \sum_{j=1}^{r} \lambda_j\, u_j^{\circ N}, \qquad U = [u_1, \dots, u_r], \qquad U'U = I_r,$$

the unfolding admits the representation

$$\mathbf{A} = U\,\mathrm{diag}(\lambda)\,\big(U^{\odot(N-1)}\big)', \qquad U^{\odot(N-1)} := \underbrace{U \odot \cdots \odot U}_{N-1 \text{ times}},$$

where $\odot$ denotes the Khatri–Rao product.

Consider the multilinear contraction $u_i'\mathbf{A}\big(u_i^{\otimes(N-1)}\big)$. Substituting the above decomposition gives

$$u_i'\mathbf{A}\big(u_i^{\otimes(N-1)}\big) = u_i'U\,\mathrm{diag}(\lambda)\,\big(U^{\odot(N-1)}\big)'\big(u_i^{\otimes(N-1)}\big).$$

Since $U'u_i = e_i$ (the ith canonical basis vector in $\mathbb{R}^r$), we have $u_i'U = e_i'$. Moreover, using the defining property of the Khatri–Rao product,

$$\big(U^{\odot(N-1)}\big)'\big(u_i^{\otimes(N-1)}\big) = \big(U'u_i\big)^{\otimes(N-1)} = e_i^{\otimes(N-1)}.$$

Therefore,

$$u_i'\mathbf{A}\big(u_i^{\otimes(N-1)}\big) = e_i'\,\mathrm{diag}(\lambda)\,e_i = \lambda_i,$$

which establishes the second equality in (13).

Finally, by the equivalence between Frobenius inner products and multilinear contractions of a matricization,

$$\langle \mathcal{A},\, u_i^{\circ N}\rangle = u_i'\mathbf{A}\big(u_i^{\otimes(N-1)}\big),$$

and since $\langle u_j^{\circ N}, u_i^{\circ N}\rangle = (u_j'u_i)^N = \delta_{ij}$ under $U'U = I_r$, it follows that

$$\langle \mathcal{A},\, u_i^{\circ N}\rangle = \sum_{j=1}^{r} \lambda_j \langle u_j^{\circ N}, u_i^{\circ N}\rangle = \lambda_i.$$

Proof of Theorem 6

Assumptions for the joint TSVD. Let $d \in \{3, 4\}$ denote the cumulant order used for identification.

Assumption C.1 (Orthogonal decomposability). There exist $r \le n$ and or-

thonormal vectors $q_1, \ldots, q_r$ such that

$$\mathcal{C}^d(u) = \sum_{i=1}^{r} \lambda_{d,i} q_i^{\circ d}, \qquad |\lambda_{d,1}| > \cdots > |\lambda_{d,r}| > 0.$$

Assumption C.2 (Uniform consistency).

$$\sup_{\|q\|=1} \left| q' \widehat{\mathbf{C}}^d_{u,T}(q^{\otimes(d-1)}) - q' \mathbf{C}^d_u(q^{\otimes(d-1)}) \right| \xrightarrow{p} 0.$$

Assumption C.3 (Cumulant CLT).

$$\sqrt{T}\,\mathrm{vec}\Big(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u\Big) \Rightarrow \mathcal{N}(0, \Sigma_d).$$

Let $\mathcal{Q}_r = \{Q_r \in \mathbb{R}^{n\times r} : Q_r' Q_r = I_r\}$ and

$$\lambda_{d,i,T}(Q_r) = q_i' \widehat{\mathbf{C}}^d_{u,T}(q_i^{\otimes(d-1)}), \qquad J_T(Q_r) = \sum_{i=1}^{r} \lambda_{d,i,T}(Q_r)^2.$$

The joint TSVD estimator satisfies

$$\widehat{Q}_{r,T} \in \arg\max_{Q_r \in \mathcal{Q}_r} J_T(Q_r).$$

Consistency of the joint TSVD estimator.

By Assumption C.2,

$$\sup_{Q_r \in \mathcal{Q}_r} |J_T(Q_r) - J(Q_r)| \xrightarrow{p} 0.$$

Under orthogonal decomposability (C.1) the population criterion $J(Q_r)$ is uniquely maximized (up to permutations and sign changes) by $Q_r = [q_1, \ldots, q_r]$. Since $\mathcal{Q}_r$ is compact, the argmax theorem (Newey and McFadden, 1994) implies

$$\widehat{Q}_{r,T} P_T S_T \xrightarrow{p} Q_r,$$

where P_T is a permutation matrix and S_T is diagonal with ± 1 entries. Hence each column $\widehat{q}_{j,T}$ converges in probability to q_j up to sign and permutation.

Asymptotic distribution for $d \in \{3, 4\}$. Part (i) of the theorem:

Let $\mathcal{L}_T(Q_r, \mu)$ denote the sample Lagrangian associated with the joint TSVD problem. The finite-sample first-order conditions for $i = 1, \ldots, r$ are

$$2\widehat{\lambda}_{d,i,T} \widehat{v}_{i,T} - 2\widehat{\mu}_{ii} \widehat{q}_i - \sum_{j \neq i} \widehat{\mu}_{ij} \widehat{q}_j = 0, \tag{22}$$

where

$$\widehat{v}_{i,T} = d\,\widehat{\mathbf{C}}^d_{u,T}(\widehat{q}_i^{\otimes(d-1)}), \qquad \widehat{\lambda}_{d,i,T} = \widehat{q}_i' \widehat{\mathbf{C}}^d_{u,T}(\widehat{q}_i^{\otimes(d-1)}).$$

Premultiplying (22) by $\widehat{q}_i'$ yields

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

Hence the identifying equations reduce to

$$\widehat{\lambda}_{d,i,T}\,\widehat{q}_j'\widehat{v}_{i,T} = \widehat{\lambda}_{d,j,T}\,\widehat{q}_i'\widehat{v}_{j,T}, \qquad i \neq j. \tag{23}$$

Step 1: Linearization around the population solution.

Let $Q_r = [q_1, \ldots, q_r]$ denote the population maximizer. A first-order Taylor expansion of (23) around Q_r and $\mathbf{C}_u^d$ gives

$$\begin{aligned} 0 = {} & \lambda_{d,i}\, q_j'(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u)(q_i^{\otimes(d-1)}) - \lambda_{d,j}\, q_i'(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u)(q_j^{\otimes(d-1)}) \\ & + \lambda^2_{d,i} q_i'(\widehat{q}_j - q_j) - \lambda^2_{d,j} q_j'(\widehat{q}_i - q_i) + o_p(T^{-1/2}). \end{aligned}$$

Using the orthogonality constraint gives

$$q_i'(\widehat{q}_j - q_j) = \frac{\lambda_{d,j}\, q_i'(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u)(q_j^{\otimes(d-1)}) - \lambda_{d,i}\, q_j'(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u)(q_i^{\otimes(d-1)})}{\lambda^2_{d,i} + \lambda^2_{d,j}} + o_p(T^{-1/2}).$$

Step 2: Matrix representation. Stacking the relations yields

$$\sqrt{T}\big(\mathrm{vec}(\widehat{Q}_{r,T}) - \mathrm{vec}(Q_r)\big) = (I_r \otimes Q_r)\,\Xi M_d' \sqrt{T}\,\mathrm{vec}(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u) + o_p(1),$$

where $M_d := Q_r^{\odot(d-1)} \otimes Q_r$. and 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). Next, we define the $r \times r$ matrix W:

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

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

Step 3: Asymptotic distribution. Applying Assumption C.3 and Slutsky's theorem gives

$$\sqrt{T}\big(\mathrm{vec}(\widehat{Q}_{r,T}) - \mathrm{vec}(Q_r)\big) \Rightarrow \mathcal{N}(0,\ (I_r \otimes Q_r)\Xi M_d' \Sigma_d M_d \Xi'(I_r \otimes Q_r)')\,.$$

Part (ii) of the theorem: Singular-vector perturbation.

Extracting column j from the expansion above yields

$$\sqrt{T}(\widehat{q}_{j,T} - q_j) = \sum_{\ell=1}^{r} \omega_{\ell j,T} q_\ell + o_p(1).$$

Orthogonality of Q_r implies $\omega_{jj,T} = 0$ and therefore

$$P_j^{\perp}(\widehat{q}_{j,T} - q_j) = \frac{1}{\sqrt{T}} \sum_{\ell \neq j} \omega_{\ell j,T} q_\ell + o_p(T^{-1/2}), \qquad P_j^{\perp} = I - q_j q_j'.$$

Using the linearization in Part (i) gives

$$\omega_{\ell j,T} = \frac{\lambda_{d,\ell}\, \alpha_{\ell j,T} - \lambda_{d,j}\, \alpha_{j\ell,T}}{\lambda_{d,\ell}^2 + \lambda_{d,j}^2},$$

where

$$\alpha_{\ell j,T} = \sqrt{T} \left\langle \widehat{\mathcal{C}}_{u,T}^d - \mathcal{C}_u^d, q_\ell \otimes q_j^{\otimes(d-1)} \right\rangle.$$

Rearranging yields the singular-vector perturbation expansion

$$\sqrt{T}\, P_j^{\perp}(\widehat{q}_{j,T} - q_j) = \sum_{\ell \neq j} \frac{\sqrt{T}\, \alpha_{\ell j,T}}{\lambda_{d,j} - \lambda_{d,\ell}}\, q_\ell + o_p(1).$$

Part (iii) of the theorem: Asymptotic variance and identification strength.

Under the cumulant CLT each $\sqrt{T}\alpha_{\ell j,T}$ is asymptotically normal. Therefore

$$\sqrt{T}(\widehat{q}_{j,T} - q_j) \Rightarrow N(0, \Omega_j), \qquad \Omega_j = \sum_{\ell \neq j} \frac{(\alpha_{\ell j,T})}{(\lambda_{d,j} - \lambda_{d,\ell})^2}\, q_\ell q_\ell'.$$

Thus estimation precision is governed by the inverse squared singular-value gaps.

Proof of Theorem 10.

Let $E_T := \widehat{\mathcal{C}}_T^d(u) - \mathcal{C}^d(u)$ denote the perturbation of the population cumulant tensor. From the deterministic perturbation expansion for the joint TSVD maximizer (Theorem 6(iii)), whenever $\|E_T\|_F \leq c\, \delta_j$ for sufficiently small $c > 0$,

$$P_j^{\perp}(\widehat{q}_{j,T} - q_j) = \sum_{\ell \neq j} \frac{\alpha_{\ell j,T}}{\lambda_{d,j} - \lambda_{d,\ell}}\, q_\ell + R_T, \tag{24}$$

where $\alpha_{\ell j,T} = \left\langle E_T,\ q_\ell \circ q_j^{\circ(d-1)} \right\rangle$ and the remainder satisfies

$$\|R_T\| \leq C_0\, \frac{\|E_T\|_F^2}{\delta_j^2}$$

for some constant $C_0 > 0$. By Cauchy–Schwarz,

$$|\alpha_{\ell j,T}| \leq \|E_T\|_F\, \|q_\ell \circ q_j^{\circ(d-1)}\|_F.$$

Since $\|q_j\| = \|q_\ell\| = 1$, the rank-one tensor $q_\ell \circ q_j^{\circ(d-1)}$ has unit Frobenius norm, so

$$|\alpha_{\ell j,T}| \leq \|E_T\|_F.$$

Because the population singular values are separated,

$$|\lambda_{d,j} - \lambda_{d,\ell}| \geq \delta_j, \qquad \ell \neq j.$$

Taking norms in (24) yields

$$\|P_j^{\perp}(\widehat{q}_{j,T} - q_j)\| \leq \sum_{\ell \neq j} \frac{|\alpha_{\ell j,T}|}{|\lambda_{d,j} - \lambda_{d,\ell}|} + \|R_T\| \leq \frac{r-1}{\delta_j}\|E_T\|_F + C_0 \frac{\|E_T\|_F^2}{\delta_j^2}.$$

When $\|E_T\|_F \leq c\,\delta_j$ with c sufficiently small, the quadratic term is dominated by the linear term, implying for some constant $C > 0$ that

$$\|P_j^{\perp}(\widehat{q}_{j,T} - q_j)\| \leq C\,\frac{\|E_T\|_F}{\delta_j}.$$

Finally, since $\|\widehat{q}_{j,T}\| = \|q_j\| = 1$, we have

$$q_j'(\widehat{q}_{j,T} - q_j) = -\frac{1}{2}\|\widehat{q}_{j,T} - q_j\|^2,$$

so the component along q_j is of second order. Hence, for $\|\widehat{q}_{j,T} - q_j\|$ sufficiently small,

$$\|\widehat{q}_{j,T} - q_j\| \leq 2\,\|P_j^{\perp}(\widehat{q}_{j,T} - q_j)\|.$$

Combining the bounds possibly with a different constant $C > 0$ yields

$$\|\widehat{q}_{j,T} - q_j\| \leq C\,\frac{\|E_T\|_F}{\delta_j}.$$

Proof of Theorem 12.

Throughout assume the local-to-weak parameterization

$$\lambda_{d,1} - \lambda_{d,2} = \frac{c}{\sqrt{T}}, \qquad c \neq 0,$$

while all remaining spectral gaps remain separated:

$$\min_{k \geq 3} |\lambda_{d,1} - \lambda_{d,k}| \geq \underline{\delta} > 0, \qquad \min_{k \geq 3} |\lambda_{d,2} - \lambda_{d,k}| \geq \underline{\delta} > 0.$$

Let $\widehat{q}_{1,T}$ denote the unit-norm TSVD singular vector associated with $\widehat{\lambda}_{d,1}$, with sign fixed so that $q_1'\widehat{q}_{1,T} \geq 0$. Define

$$P_1^{\perp} := I - q_1 q_1', \qquad P_{\{1,2\}}^{\perp} := I - q_1 q_1' - q_2 q_2',$$

and $\alpha_{21,T} := \left\langle E_T,\ q_2 \circ q_1^{\circ(d-1)} \right\rangle$.

Step 1: perturbation expansion.

Under orthogonal decomposability, the singular-vector perturbation expansion (Theorem 6) yields

$$P_1^{\perp}(\widehat{q}_{1,T} - q_1) = \sum_{k \neq 1} \frac{\alpha_{k1,T}}{\lambda_{d,1} - \lambda_{d,k}}\, q_k + r_{1,T}, \tag{25}$$

where $\alpha_{k1,T} = \left\langle E_T,\ q_k \circ q_1^{\circ(d-1)} \right\rangle$ and $\|r_{1,T}\| = o_p(\|E_T\|)$ provided the corresponding

gaps remain bounded away from zero. For $k \geq 3$ the denominators satisfy $|\lambda_{d,1} - \lambda_{d,k}| \geq \underline{\delta}$, so

$$\sum_{k\geq 3} \frac{\alpha_{k1,T}}{\lambda_{d,1} - \lambda_{d,k}} q_k = O_p(\|E_T\|) = O_p(T^{-1/2}), \qquad r_{1,T} = o_p(T^{-1/2}).$$

Thus the dominant term in (25) arises from $k = 2$:

$$P_1^{\perp}(\widehat{q}_{1,T} - q_1) = \frac{\alpha_{21,T}}{\lambda_{d,1} - \lambda_{d,2}} q_2 + O_p(T^{-1/2}). \tag{26}$$

Step 2: failure of root-T consistency (part (i)).

Since $\alpha_{21,T}$ is a linear functional of E_T, we have $\alpha_{21,T} = O_p(T^{-1/2})$. Using $\lambda_{d,1} - \lambda_{d,2} = c/\sqrt{T}$ in (26) gives

$$P_1^{\perp}(\widehat{q}_{1,T} - q_1) = \frac{\sqrt{T}\,\alpha_{21,T}}{c} q_2 + O_p(T^{-1/2}) = O_p(1).$$

Hence $\widehat{q}_{1,T}$ is not root-T consistent.

Step 3: consistency of the invariant subspace (part (ii)).

From (26),

$$P_{\{1,2\}}^{\perp}(\widehat{q}_{1,T} - q_1) = O_p(T^{-1/2}) = o_p(1),$$

so asymptotically $\widehat{q}_{1,T}$ lies in $\operatorname{span}\{q_1, q_2\}$. Applying the same argument to $\widehat{q}_{2,T}$ yields

$$P_{\{1,2\}}^{\perp}(\widehat{q}_{2,T} - q_2) = o_p(1),$$

and therefore $\operatorname{span}\{\widehat{q}_{1,T}, \widehat{q}_{2,T}\} \xrightarrow{p} \operatorname{span}\{q_1, q_2\}$.

Step 4: random rotation limit (part (iii)).

Let $\beta_T := q_2'(\widehat{q}_{1,T} - q_1)$. Taking the q_2-component in (26) gives

$$\beta_T = \frac{\alpha_{21,T}}{\lambda_{d,1} - \lambda_{d,2}} + o_p(1) = \frac{\sqrt{T}\,\alpha_{21,T}}{c} + o_p(1).$$

By the assumed CLT for $\sqrt{T}\operatorname{vec}(E_T)$ and linearity of the map $E_T \mapsto \langle E_T, q_2 \circ q_1^{\circ(d-1)} \rangle$,

$$\sqrt{T}\,\alpha_{21,T} \Rightarrow Z, \qquad Z \sim \mathcal{N}(0, \sigma_{21}^2),$$

so $\beta_T \Rightarrow Z/c$. Since $\widehat{q}_{1,T}$ is unit norm and $P_{\{1,2\}}^{\perp}(\widehat{q}_{1,T} - q_1) = o_p(1)$, we can write

$$\widehat{q}_{1,T} = \frac{q_1 + \beta_T q_2 + o_p(1)}{\sqrt{1 + \beta_T^2 + o_p(1)}}.$$

Applying the continuous mapping theorem yields

$$\widehat{q}_{1,T} \Rightarrow \frac{q_1 + (Z/c)\, q_2}{\sqrt{1 + (Z/c)^2}},$$

which proves the random-rotation limit and completes the proof.

A Spectral Framework for Non-Gaussian SVARs

SUPPLEMENTARY MATERIAL

Alain Guay Dalibor Stevanovic

Université du Québec à Montréal, CIREQ and Chaire en macroéconomie et prévisions (ESG–UQAM)

March 6, 2026

Abstract

This supplementary material contains six sections. The first summarizes important properties of the cumulants. The second defines higher-order tensors. The third and fourth sections present the joint TSVD estimator with mixed third- and fourth-orther identification, and the sequential TSVD estimator respectively. Additional simulations results are reported in the fifth section while the sixth presents the GMM framework.

Section S.1 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
$$Cum_d(Y) = A Cum_d(X) \left(A \otimes A \cdots \otimes A\right)'.$$
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.

Section S.2 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 S.1 *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 S.2 *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 S.3 *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 S.4 *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 S.5 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 S.6 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 S.7 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.*[15]

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

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

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

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

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.

[15] 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 $\mathcal{A} \times_n U$.

The Tucker decomposition (Tucker, 1963) is represented by the decomposition (S.1), 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 $n = 1, \ldots, N$.

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 $\mathbf{U}^{(n)}\mathbf{U}^{(n)\prime} = I$, implies that the *core tensor* $\mathcal{S}$ is given by:

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

or equivalently:

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

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 (S.2), 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

$$\mathbf{A} = \mathbf{U} \, \text{diag}(\lambda_i) \, (\mathbf{U} \odot \cdots \odot \mathbf{U})', \tag{S.3}$$

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.

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).

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

$$\min_{\widehat{\mathcal{A}}} \ \|\mathcal{A} - \widehat{\mathcal{A}}\|_F^2, \tag{S.4}$$

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.

Section S.3 Joint TSVD with Mixed Third- and Fourth-Order Identification

We first consider the case where structural shocks are identified by different higher-order cumulants. Suppose that r_1 shocks exhibit nonzero third-order cumulants (skewness) and r_2 shocks are non-mesokurtic, with

$$r_1 + r_2 = r.$$

Population problem. The population optimization problem is

$$\max_{Q_r'Q_r = I_r} \left(\sum_{i=1}^{r_1} \lambda_{3,i}^2(Q_{r1}) + \sum_{j=1}^{r_2} \lambda_{4,j}^2(Q_{r2}) \right), \tag{S.5}$$

where

$$Q_r = \begin{bmatrix} Q_{r1} & Q_{r2} \end{bmatrix}, \qquad Q_{r1} \in \mathbb{R}^{n \times r_1}, \quad Q_{r2} \in \mathbb{R}^{n \times r_2},$$

collect the directions identified respectively by skewness and kurtosis.

Let the stacked cumulant operator be

$$\Psi_u = \begin{bmatrix} \mathbf{C}_u^3 \\ \mathbf{C}_u^4 \end{bmatrix}, \qquad \widehat{\Psi}_{u,T} = \begin{bmatrix} \widehat{\mathbf{C}}_{u,T}^3 \\ \widehat{\mathbf{C}}_{u,T}^4 \end{bmatrix}.$$

Sample criterion and Lagrangian

For $i = 1, \ldots, r_1$, define

$$\lambda_{3,i,T} = q_i' \widehat{\mathbf{C}}_{u,T}^3 (q_i \otimes q_i),$$

and for $j = 1, \ldots, r_2$,

$$\lambda_{4,j,T} = q_{r_1+j}' \widehat{\mathbf{C}}_{u,T}^4 (q_{r_1+j} \otimes q_{r_1+j} \otimes q_{r_1+j}).$$

The sample Lagrangian associated with (S.5) is

$$\mathcal{L}_T(Q_r, \mu) = \sum_{i=1}^{r_1} \lambda_{3,i,T}^2 + \sum_{j=1}^{r_2} \lambda_{4,j,T}^2 - \sum_{k,\ell=1}^{r} \mu_{k\ell} \big(q_k' q_\ell - \delta_{k\ell} \big), \tag{S.6}$$

where the orthogonality constraints satisfy

$$Q_r'Q_r = I_r \quad \Longleftrightarrow \quad q_k' q_\ell = \delta_{k\ell}.$$

The first-order conditions therefore combine the skewness-based TSVD FOCs for the first r_1 columns, and the kurtosis-based TSVD FOCs for the remaining r_2 columns, under a common orthogonality restriction. Hence the mixed estimator is obtained by stacking the two systems of equations.

Asymptotic distribution in the mixed (r_1, r_2) case

Let Ξ_1 and Ξ_2 denote the weight matrices arising from the skewness and kurtosis blocks, respectively. A first-order expansion of the stacked FOCs around the population solution Q_r yields

$$\begin{bmatrix} \sqrt{T}\big(\mathrm{vec}(\widehat{Q}_{r1,T}) - \mathrm{vec}(Q_{r1})\big) \\ \sqrt{T}\big(\mathrm{vec}(\widehat{Q}_{r2,T}) - \mathrm{vec}(Q_{r2})\big) \end{bmatrix} = \begin{bmatrix} (I_{r_1} \otimes Q_{r1})\Xi_1 & 0 \\ 0 & (I_{r_2} \otimes Q_{r2})\Xi_2 \end{bmatrix} \begin{bmatrix} ((Q_{r1}^{\odot 2}) \otimes Q_{r1})' \\ ((Q_{r2}^{\odot 3}) \otimes Q_{r2})' \end{bmatrix} \times \sqrt{T}\big(\mathrm{vec}(\widehat{\Psi}_{u,T}) - \mathrm{vec}(\Psi_u)\big) + o_p(1).$$

Hence the asymptotic covariance depends on the joint covariance matrix of the stacked cumulant estimator $\widehat{\Psi}_{u,T}$, which reflects first-stage VAR estimation uncertainty.

Case where all shocks exhibit both skewness and excess kurtosis.

We now consider the case where each of the r structural shocks displays both nonzero third- and fourth-order cumulants. The population problem becomes

$$\max_{Q_r'Q_r = I_r} \sum_{i=1}^{r} \big(\lambda_{3,i}^2(Q_r) + \lambda_{4,i}^2(Q_r)\big). \tag{S.7}$$

The corresponding sample Lagrangian is

$$\mathcal{L}_T(Q_r, \mu) = \sum_{i=1}^{r} \big(\lambda_{3,i,T}^2 + \lambda_{4,i,T}^2\big) - \sum_{j,k=1}^{r} \mu_{jk}(q_j'q_k - \delta_{jk}).$$

The first-order conditions satisfy

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

where

$$\widehat{v}_{3,i,T} = 3\,\widehat{\mathbf{C}}_{u,T}^{3}(\widehat{q}_i \otimes \widehat{q}_i), \qquad \widehat{v}_{4,i,T} = 4\,\widehat{\mathbf{C}}_{u,T}^{4}(\widehat{q}_i^{\otimes 3}).$$

Linear representation. A first-order Taylor expansion around (Q_r, Ψ_u) gives

$$\sqrt{T}\big(\mathrm{vec}(\widehat{Q}_{r,T}) - \mathrm{vec}(Q_r)\big) = (I_r \otimes Q_r)\,\Xi \begin{bmatrix} ((Q_r^{\odot 2}) \otimes Q_r)' \\ ((Q_r^{\odot 3}) \otimes Q_r)' \end{bmatrix} \sqrt{T}\big(\mathrm{vec}(\widehat{\Psi}_{u,T}) - \mathrm{vec}(\Psi_u)\big) + o_p(1).$$

Let

$$W_3 = \left[\tfrac{\lambda_{3,i}}{\lambda_{3,i}^2 + \lambda_{3,j}^2}\right], \qquad W_4 = \left[\tfrac{\lambda_{4,i}}{\lambda_{4,i}^2 + \lambda_{4,j}^2}\right],$$

and define $F_3 = \mathrm{diag}(\mathrm{vec}(W_3'))$, $F_4 = \mathrm{diag}(\mathrm{vec}(W_4'))$. With $\mathcal{P}_{r,r}$ denoting the mod-r perfect shuffle matrix,

$$\Xi = (I_{r^2} - \mathcal{P}_{r,r}) \begin{bmatrix} F_3 & F_4 \end{bmatrix}.$$

This representation nests the pure third-order and pure fourth-order TSVD esti-

mators as special cases and characterizes the asymptotic covariance of the joint TSVD estimator when structural shocks are identified jointly by skewness and kurtosis.

Section S.4 Sequential TSVD Estimator: Consistency and Asymptotic Distribution

This section establishes consistency and asymptotic normality of the sequential TSVD estimator.

Sequential TSVD: criterion and first-order conditions

For any unit vector $q \in \mathbb{S}^{n-1} = \{q \in \mathbb{R}^n : \|q\| = 1\}$, define the sample cumulant contrast

$$\lambda_{d,T}(q) = q'\widehat{\mathbf{C}}^d_{u,T}\big(q^{\otimes(d-1)}\big), \qquad J_T(q) = \lambda_{d,T}(q)^2, \tag{S.8}$$

and the population analogues

$$\lambda_d(q) = q'\mathbf{C}^d_u\big(q^{\otimes(d-1)}\big), \qquad J(q) = \lambda_d(q)^2.$$

Under the orthogonal decomposition

$$\mathcal{C}^d(u) = \sum_{i=1}^{r} \lambda_{d,i}\, q_i^{\circ d}, \qquad Q_r = [q_1, \ldots, q_r], \quad Q_r'Q_r = I_r, \qquad \lambda_{d,1}^2 > \cdots > \lambda_{d,r}^2 > 0. \tag{S.9}$$

this implies

$$\mathbf{C}^d_u\big(q_i^{\otimes(d-1)}\big) = \lambda_{d,i}\, q_i, \qquad i = 1, \ldots, r, \tag{S.10}$$

and $\mathbf{C}^d_u\big(q_i^{\otimes(d-1)}\big) = 0$ for $i > r$ if $r < n$.

The sequential estimator is defined recursively. For $k = 1$,

$$\widehat{q}^{\,\mathrm{seq}}_{1,T} \in \arg\max_{\|q\|=1} J_T(q).$$

For $k \geq 2$, given $\widehat{q}^{\,\mathrm{seq}}_{1,T}, \ldots, \widehat{q}^{\,\mathrm{seq}}_{k-1,T}$, let

$$\widehat{V}_{k-1} = \mathrm{span}\{\widehat{q}^{\,\mathrm{seq}}_{1,T}, \ldots, \widehat{q}^{\,\mathrm{seq}}_{k-1,T}\}, \qquad P^{\perp}_{k-1,T} = I_n - \sum_{j=1}^{k-1} \widehat{q}^{\,\mathrm{seq}}_{j,T}\widehat{q}^{\,\mathrm{seq}\prime}_{j,T},$$

and define

$$\widehat{q}^{\,\mathrm{seq}}_{k,T} \in \arg\max_{\substack{\|q\|=1 \\ q \perp \widehat{V}_{k-1}}} J_T(q).$$

Collecting columns yields $\widehat{Q}^{\mathrm{seq}}_{r,T} = [\widehat{q}^{\,\mathrm{seq}}_{1,T}, \ldots, \widehat{q}^{\,\mathrm{seq}}_{r,T}]$.

Gradients and Lagrangians. Differentiating (S.8) gives

$$\partial_q \lambda_{d,T}(q) = d\,\widehat{\mathbf{C}}^d_{u,T}\big(q^{\otimes(d-1)}\big), \qquad \partial_q J_T(q) = 2\lambda_{d,T}(q)\,\partial_q \lambda_{d,T}(q) = 2d\,\lambda_{d,T}(q)\,\widehat{\mathbf{C}}^d_{u,T}\big(q^{\otimes(d-1)}\big).$$

Optimization takes place on the unit sphere; equivalently, we enforce the constraint through Lagrange multipliers.

For $k = 1$, the Lagrangian is

$$\mathcal{L}_{1,T}(q, \alpha_1) = J_T(q) - \alpha_1(q'q - 1),$$

and the first-order conditions are

$$2d\,\lambda_{d,T}(q)\,\widehat{\mathbf{C}}^d_{u,T}\big(q^{\otimes(d-1)}\big) - 2\alpha_1 q = 0, \tag{S.11}$$

$$q'q = 1. \tag{S.12}$$

For $k \geq 2$, the Lagrangian is

$$\mathcal{L}_{k,T}(q, \alpha_k, \beta_k) = J_T(q) - \alpha_k(q'q - 1) - \sum_{j=1}^{k-1} \beta_{kj}\, q'\widehat{q}^{\text{seq}}_{j,T},$$

and the first-order conditions are

$$2d\,\lambda_{d,T}(q)\,\widehat{\mathbf{C}}^d_{u,T}\big(q^{\otimes(d-1)}\big) - 2\alpha_k q - \sum_{j=1}^{k-1} \beta_{kj}\,\widehat{q}^{\text{seq}}_{j,T} = 0, \tag{S.13}$$

$$q'q = 1, \tag{S.14}$$

$$q'\widehat{q}^{\text{seq}}_{j,T} = 0, \qquad j = 1, \ldots, k-1. \tag{S.15}$$

Let $P^{\perp}_{k-1}$ be the orthogonal projector onto the complement of $\widehat{V}_{k-1}$:

$$P^{\perp}_{k-1} = I_n - \sum_{j=1}^{k-1} \widehat{q}^{\text{seq}}_{j,T}\widehat{q}^{\text{seq}\prime}_{j,T}.$$

where we define the population subspace $V_{k-1} = \text{span}\{q_1, \ldots, q_{k-1}\}$ and its empirical counterpart $\widehat{V}_{k-1} = \text{span}\{\widehat{q}^{\text{seq}}_{1,T}, \ldots, \widehat{q}^{\text{seq}}_{k-1,T}\}$. Premultiplying (S.13) by $P^{\perp}_{k-1,T}$ and using $P^{\perp}_{k-1,T}\widehat{q}^{\text{seq}}_{j,T} = 0$ yields the projected equation

$$P^{\perp}_{k-1,T}\,\widehat{\mathbf{C}}^d_{u,T}\big(\widehat{q}^{\text{seq}\,\otimes(d-1)}_{k,T}\big) = \gamma_{k,T}\,\widehat{q}^{\text{seq}}_{k,T}, \qquad \gamma_{k,T} = \frac{\widehat{\alpha}_{k,T}}{d\,\lambda_{d,T}(\widehat{q}^{\text{seq}}_{k,T})}.$$

Consistency of the sequential TSVD

We now show that the sequential TSVD estimator is consistent for Q_r, up to signs and permutations, under Assumptions C.1 and C.2.

Theorem S.8 (Consistency) *Suppose Assumptions C.1 and C.2 hold. Then there exists a sequence of permutation matrices P_r and diagonal sign matrices S_r (with*

entries ± 1*) such that*

$$\widehat{Q}_{r,T}^{\text{seq}} \xrightarrow{p} Q_r P_r S_r, \qquad T \to \infty.$$

The proof proceeds by induction on k.

Step 1 ($k = 1$*).* Let $J(q) = \lambda(q)^2 = \left[q' \mathbf{C}_u^3 (q \otimes q]^2\right.$ be the population contrast. Under Assumption C.1, $|\lambda_1| > \cdots > |\lambda_r|$. It is standard (and shown, e.g., in Anandkumar et al. (2014) that $J(q)$ is uniquely maximized on the unit sphere by q_1 and $-q_1$. Let

$$q_1^\star \in \arg\max_{\|q\|=1} J(q) = \{\pm q_1\}.$$

Assumption C.2 implies

$$\sup_{\|q\|=1} \left| J_T(q) - J(q) \right| \xrightarrow{p} 0.$$

The feasible set $\{q : \|q\| = 1\}$ is compact, so the argmax theorem (e.g., Newey and McFadden, 1994) yields

$$\widehat{q}_{1,T}^{\text{seq}} \xrightarrow{p} q_1^\star \in \{\pm q_1\}.$$

Induction step. Assume that for some $k \in \{2, \ldots, r\}$ we have

$$\widehat{q}_{j,T}^{\text{seq}} \xrightarrow{p} q_j^\star \in \{\pm q_j\}, \qquad j = 1, \ldots, k-1.$$

By the induction hypothesis and continuity of the projection operator, we have

$$\widehat{V}_{k-1} \xrightarrow{p} V_{k-1}.$$

Consider the population problem

$$q_k^\star \in \arg\max_{q \in \mathbb{S}^{n-1}(V_{k-1})} J(q),$$

where $\mathbb{S}^{n-1}(V_{k-1}) = \{q : \|q\| = 1,\, q'v = 0 \;\forall v \in V_{k-1}\}$. Under Assumption C.1, the maximizer is unique up to sign and equal to $\pm q_k$:

$$q_k^\star \in \{\pm q_k\}.$$

At the sample level, the k-th sequential estimator solves

$$\widehat{q}_{k,T}^{\text{seq}} \in \arg\max_{q \in \mathbb{S}^{n-1}(\widehat{V}_{k-1})} J_T(q).$$

Assumption C.2 again implies uniform convergence of J_T to J on neighborhoods of $\mathbb{S}^{n-1}(V_{k-1})$, and the convergence $\widehat{V}_{k-1} \to V_{k-1}$ ensures that the feasible sets converge in the sense required by the argmax theorem. Hence,

$$\widehat{q}_{k,T}^{\text{seq}} \xrightarrow{p} q_k^\star \in \{\pm q_k\}.$$

By induction, each column converges (up to sign) to the corresponding population

direction, and collecting them yields

$$\widehat{Q}_{r,T}^{\text{seq}} \xrightarrow{p} Q_r P_r S_r$$

for suitable permutation and sign matrices (P_r, S_r). This completes the proof.

Asymptotic distribution

Theorem S.9 (Asymptotic normality of the sequential TSVD) *Suppose Assumptions C.1, C.2 and C.3. Assume additionally that the population first-order conditions at each step are locally regular so that the Jacobians defined below are nonsingular. Then there exist matrices J_{seq} and D_{seq} such that*

$$\sqrt{T}\Big(\text{vec}(\widehat{Q}_{r,T}^{\text{seq}}) - \text{vec}(Q_r)\Big) = -\, J_{\text{seq}}^{-1} D_{\text{seq}}\, \sqrt{T}\,\text{vec}\big(\widehat{\mathbf{C}}_{u,T}^{d} - \mathbf{C}_u^{d}\big) + o_p(1), \tag{S.16}$$

and

$$\sqrt{T}\Big(\text{vec}(\widehat{Q}_{r,T}^{\text{seq}}) - \text{vec}(Q_r)\Big) \Rightarrow N\Big(0,\ \Omega_Q^{\text{seq}}\Big), \qquad \Omega_Q^{\text{seq}} = J_{\text{seq}}^{-1} D_{\text{seq}} \Sigma_d D_{\text{seq}}' J_{\text{seq}}^{-1\prime}.$$

Moreover, J_{seq} is block lower-triangular, reflecting the recursive dependence of step k on the previously estimated directions.

Proof:

Step 1 ($k = 1$). Let $\theta_1 = (q_1', \alpha_1)'$ and define the estimating equations corresponding to (S.11)–(S.12):

$$\psi_{1,T}(\theta_1, \widehat{\mathbf{C}}_{u,T}^{d}) = \begin{bmatrix} 2d\,\lambda_{d,T}(q)\,\widehat{\mathbf{C}}_{u,T}^{d}\big(q^{\otimes(d-1)}\big) - 2\alpha_1 q \\ q'q - 1 \end{bmatrix}, \qquad \psi_{1,T}(\widehat{\theta}_{1,T}, \widehat{\mathbf{C}}_{u,T}^{d}) = 0.$$

Let $\psi_1(\theta_1, \mathbf{C}_u^d)$ denote the population counterpart and $\theta_{1,0} = (q_1', \alpha_1)'$ the population solution. A first-order Taylor expansion around $(\theta_{1,0}, \mathbf{C}_u^d)$ yields

$$0 = \psi_{1,T}(\widehat{\theta}_{1,T}, \widehat{\mathbf{C}}_{u,T}^{d}) = \psi_1(\theta_{1,0}, \mathbf{C}_u^d) + J_1(\widehat{\theta}_{1,T} - \theta_{1,0}) + D_1\,\text{vec}(\widehat{\mathbf{C}}_{u,T}^{d} - \mathbf{C}_u^{d}) + o_p(T^{-1/2}),$$

where

$$J_1 = \partial_{\theta_1}\psi_1(\theta_{1,0}, \mathbf{C}_u^d), \qquad D_1 = \partial_C \psi_1(\theta_{1,0}, \mathbf{C}_u^d).$$

Since $\psi_1(\theta_{1,0}, \mathbf{C}_u^d) = 0$ and J_1 is nonsingular by local regularity, rearranging and multiplying by $\sqrt{T}$ gives

$$\sqrt{T}(\widehat{\theta}_{1,T} - \theta_{1,0}) = -\, J_1^{-1} D_1\, \sqrt{T}\,\text{vec}(\widehat{\mathbf{C}}_{u,T}^{d} - \mathbf{C}_u^{d}) + o_p(1).$$

Step $k \geq 2$. Let $\theta_k = (q_k', \alpha_k, \beta_k')'$ where $\beta_k \in \mathbb{R}^{k-1}$ collects the multipliers enforcing orthogonality to the previously estimated directions. Let $\psi_{k,T}(\theta_k, \widehat{\mathbf{C}}_{u,T}^{d}, \widehat{Q}_{k-1,T}^{\text{seq}}) = 0$ denote the system corresponding to (S.13)–(S.15). Let $\psi_k(\theta_k, \mathbf{C}_u^d, Q_{k-1})$ be its population counterpart and $\theta_{k,0}$ the population solution.

A first-order Taylor expansion yields

$$0 = \psi_{k,T}(\widehat{\theta}_{k,T}, \widehat{\mathbf{C}}^d_{u,T}, \widehat{Q}^{\text{seq}}_{k-1,T})$$
$$= \psi_k(\theta_{k,0}, \mathbf{C}^d_u, Q_{k-1}) + J_k(\widehat{\theta}_{k,T} - \theta_{k,0}) + D_k \operatorname{vec}(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u) + E_k \operatorname{vec}(\widehat{Q}^{\text{seq}}_{k-1,T} - Q_{k-1}) + o_p(T^{-1/2}),$$

where

$$J_k = \partial_{\theta_k}\psi_k(\theta_{k,0}, \mathbf{C}^d_u, Q_{k-1}), \quad D_k = \partial_C \psi_k(\theta_{k,0}, \mathbf{C}^d_u, Q_{k-1}), \quad E_k = \partial_Q \psi_k(\theta_{k,0}, \mathbf{C}^d_u, Q_{k-1}).$$

Since $\psi_k(\theta_{k,0}, \mathbf{C}^d_u, Q_{k-1}) = 0$ and J_k is nonsingular, we obtain

$$\sqrt{T}(\widehat{\theta}_{k,T} - \theta_{k,0}) = -\, J_k^{-1}\Big[D_k\sqrt{T}\operatorname{vec}(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u) + E_k\sqrt{T}\operatorname{vec}(\widehat{Q}^{\text{seq}}_{k-1,T} - Q_{k-1})\Big] + o_p(1).$$

This displays the sequential propagation of uncertainty through the dependence on $\widehat{Q}^{\text{seq}}_{k-1,T}$.

Stacking. Collect all parameters and multipliers across $k = 1, \ldots, r$ in a vector ϑ_T (with population value ϑ_0). Stacking the above linearizations yields

$$\sqrt{T}(\vartheta_T - \vartheta_0) = -\, J^{-1}_{\text{seq}} D_{\text{seq}}\, \sqrt{T}\operatorname{vec}(\widehat{\mathbf{C}}^d_{u,T} - \mathbf{C}^d_u) + o_p(1),$$

where J_{seq} is block lower-triangular. Selecting the coordinates corresponding to $\operatorname{vec}(\widehat{Q}^{\text{seq}}_{r,T})$ gives (S.16). Finally, Assumption C.3 and Slutsky's theorem imply the stated asymptotic normality with covariance $\Omega^{\text{seq}}_Q = J^{-1}_{\text{seq}} D_{\text{seq}} \Sigma_d D'_{\text{seq}} J^{-1\prime}_{\text{seq}}$.

The sequential TSVD enforces orthogonality relative to estimated directions (deflation), which generates the additional E_k term in the linearization and a block lower-triangular Jacobian J_{seq}. Consequently, Ω^{seq}_Q generally differs from the asymptotic variance of the joint TSVD estimator, even though both procedures recover the same population directions up to permutation and sign.

Specialization to $d = 3$ and $d = 4$

For $d = 3$,

$$\lambda_{3,T}(q) = q'\widehat{\mathbf{C}}^3_{u,T}(q \otimes q), \qquad \partial_q \lambda_{3,T}(q) = 3\,\widehat{\mathbf{C}}^3_{u,T}(q \otimes q),$$

and all expressions above apply with $\Sigma_d = \Sigma_3$.

For $d = 4$,

$$\lambda_{4,T}(q) = q'\widehat{\mathbf{C}}^4_{u,T}(q^{\otimes 3}), \qquad \partial_q \lambda_{4,T}(q) = 4\,\widehat{\mathbf{C}}^4_{u,T}(q^{\otimes 3}),$$

and the same arguments apply under finite eighth moments, with $\Sigma_d = \Sigma_4$.

Section S.5 Additional Simulations

In a second set of experiments, we investigate performance in the presence of skewness, considering two non-Gaussian structural shocks that differ in asymmetry while

remaining independent and standardized.

The first shock is generated to exhibit stronger skewness:

$$2.1755\,\epsilon_{1,t} \sim \begin{cases} N(1,1), & \text{with probability } 0.7887, \\ N(-3.7326,1), & \text{with probability } 0.2113, \end{cases}$$

implying $\epsilon_{1,t}$ has skewness -0.9907.

The second shock exhibits weaker skewness:

$$1.6808\,\epsilon_{2,t} \sim \begin{cases} N(1,1), & \text{with probability } 0.5, \\ N(-1,2.65), & \text{with probability } 0.5, \end{cases}$$

yielding skewness -0.5231 for $\epsilon_{2,t}$.

In this design, identification is driven by differences in third-order cumulants rather than excess kurtosis. Consequently, singular-value separation of the third-order cumulant tensor governs identification strength, providing a complementary setting in which asymmetry rather than tail thickness generates spectral separation.

We additionally consider the PML estimator under a likelihood specified for excess kurtosis in order to assess robustness to likelihood misspecification. As in the kurtosis-based experiments, we compare HOSVD, FastICA, GMM, and the TSVD estimator proposed in this paper. The JADE estimator is excluded because it is specifically designed to exploit fourth-order cumulants. For HOSVD, FastICA, TSVD, and GMM, identification criteria are therefore based on third-order cumulants.

Under skewness-based identification, TSVD again achieves uniformly lower RMSE, see Table S.1. The similarity between TSVD and GMM performance in this case reflects the low dimensionality of the $n = 2$ setting: TSVD effectively concentrates information into directional skewness projections, whereas GMM optimizes a quadratic criterion over the full set of cumulant moments. As expected, the performance of the PML estimator deteriorates under skewness identification, owing to reliance on a likelihood based on the Student-t distribution.

Table S.1: Finite-sample distributions in complete identification case: Skewness

	T=200		T=500		T=5000	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
PML	.052	.243	.039	.229	-0165	.144
HOSVD	.008	.100	.002	.054	.000	.015
TSVD	.001	.043	.001	.026	.000	.008
FastICA	.002	.050	.001	.031	.000	.010
GMM-I	-.002	.046	.000	.027	.000	.008
GMM-opt	.001	.044	.000	.027	.000	.008

Notes. Entries are the empirical bias and root mean square errors (RMSE) for each experiment. The experiment involves two structural shocks with different degrees of skewness. The strongly skewed shock is generated 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, implying that $\epsilon_{1,t}$ has skewness -0.9907. The weakly skewed shock is generated as $1.6808 \times \epsilon_{2,t} \sim N(1,1)$ with probability0.5 and $1.6808 \times \epsilon_{2,t} \sim N(-1,2.65)$ with probability 0.5, implying that $\epsilon_{2,t}$ has skewness -0.5231. 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: $q_{11} = \cos(\alpha)$, $q_{21} = -\sin(\alpha)$, $q_{12} = \sin(\alpha)$, and $q_{22} = \cos(\alpha)$ with $\alpha = -\pi/5$ (so $q_{11} = 0.809$). The table reports the biases ($E(\hat{q}_{11} - q_{11})$) and the root-mean-squared errors.

Section S.6 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,

$$Eg_1(Z_t,\Gamma) = \begin{bmatrix} E\left(X_{t-1}\otimes\nu_t\right) \\ E\left(vech(\Sigma_v) - vech(\nu_t\nu_t')\right) \end{bmatrix} = 0, \tag{S.17}$$

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 (S.17) 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

$$E[g_2(Z_t,\widehat{\Gamma}_T,\beta)] = E[g_2(u_t(\hat{\Gamma}),\beta)] = E[\text{vecht}(\mathbf{C}^3(u_t(\widehat{\Gamma}_T))) - \text{vecht}(Q\mathbf{C}_\epsilon^3(Q\otimes Q)')] = 0, \tag{S.18}$$

where $\beta = (\text{vec}(Q_r)', \mathcal{C}_{1,1,1}^3, \ldots, \mathcal{C}_{r,r,r}^3)'$, 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.

The estimator depends on the orthonormal matrix Q_r, which satisfies the Stiefel constraint $Q_r' Q_r = I_r$. To convert the constrained GMM problem into an unconstrained optimization, we reparameterize Q_r using a mapping from unconstrained Euclidean parameters. The Cayley transform provides one convenient option by mapping a skew-symmetric matrix A into an orthogonal matrix via $Q(A) = (I - A)^{-1}(I + A)$ (defined whenever -1 is not an eigenvalue of Q), alternative parameterizations are also available. In particular, one may use the matrix exponential parameterization $Q(S) = \exp(S)$ with $S' = -S$, which generates an orthogonal matrix, or parameterize Q_r directly by orthonormalizing an unconstrained matrix $B \in \mathbb{R}^{n \times r}$, for example through the polar factor

$$Q_r(B) = B(B'B)^{-1/2},$$

which automatically enforces $Q_r(B)' Q_r(B) = I_r$.

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 (S.17). 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 (S.17) and (S.18) 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:

$$S_2 = \left[G_{2,\Gamma}V_\Gamma \quad I_d\right]\Omega\left[G_{2,\Gamma}V_\Gamma \quad I_d\right]'.$$