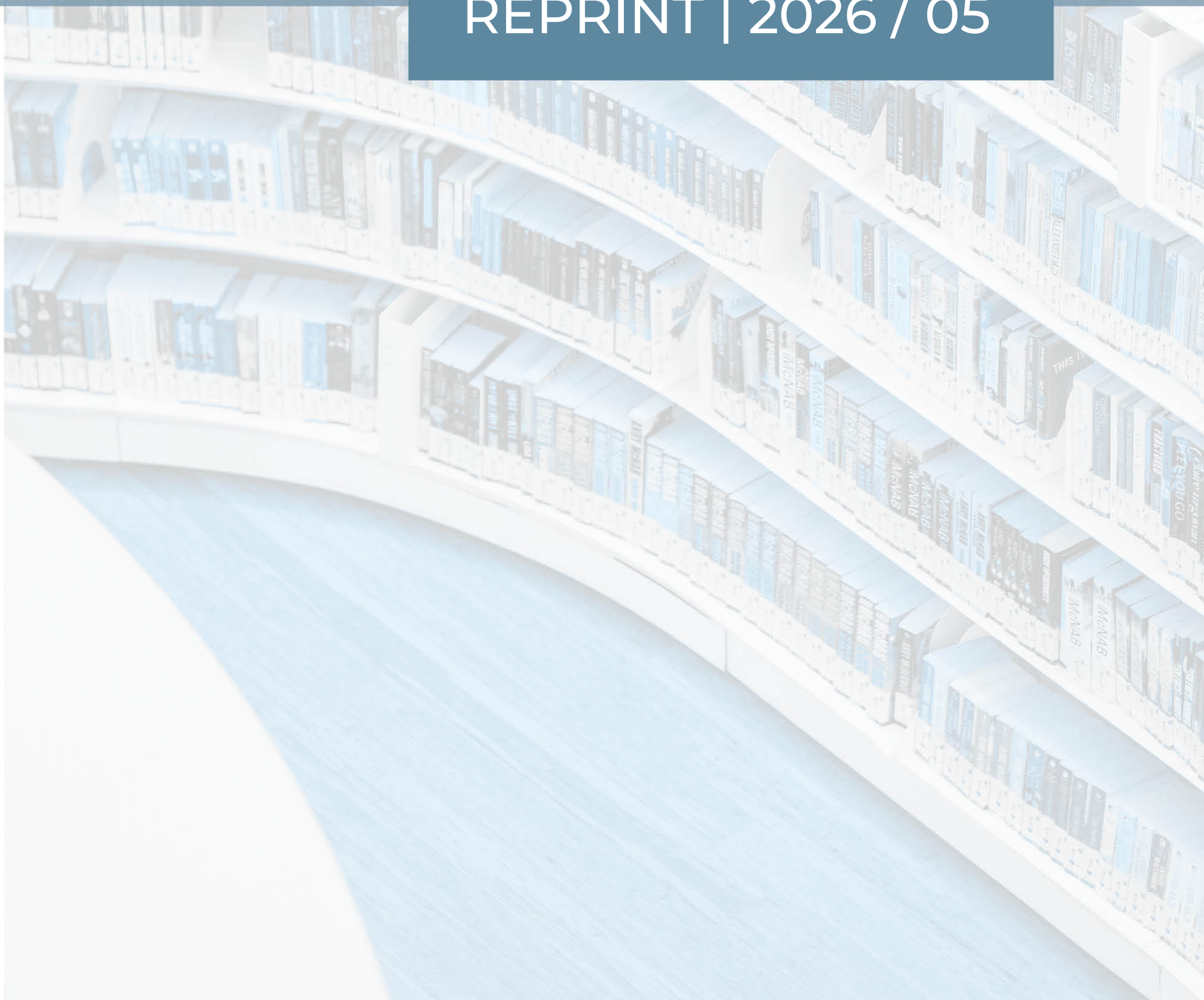
STATISTICAL IDENTIFICATION OF INDEPENDENT SHOCKS WITH KERNEL-BASED MAXIMUM LIKELIHOOD ESTIMATION AND AN APPLICATION TO THE GLOBAL CRUDE OIL MARKET

Christian M. Hafner, Helmut Herwartz, Shu Wang

REPRINT | 2026 / 05



UCLouvain

LIDAM
ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication

Statistical identification of independent shocks with kernel-based maximum likelihood estimation and an application to the global crude oil market

Christian M. Hafner* Helmut Herwartz† Shu Wang†

Abstract

Independent component analysis has emerged as a promising approach for revealing structural relationships in multivariate dynamic systems, particularly in scenarios with limited knowledge of causal patterns. This paper introduces a robust kernel-based maximum likelihood (KML) estimation method that accommodates the distributional characteristics of the structural sources of data variation. Our Monte Carlo study demonstrates the superior performance of the KML estimator compared to existing approaches for independence-based identification. Moreover, the proposed method enables partial identification and dimension reduction even in the presence of dependent shocks. We illustrate the benefits of our approach by applying it to the global oil market model of Kilian (2009), highlighting its ability to capture unmodeled higher-order dependence between oil supply and speculative oil demand shocks.

Keywords: Structural VAR, structural MGARCH, independent component analysis, kernel maximum likelihood, global crude oil market

JEL Classification: C14, C32, Q43

*UCLouvain, Louvain Institute for Data Analysis and Modelling in Economics and Statistics (LIDAM), Belgium.

†Chair of Econometrics, University of Göttingen, Humboldtallee 3, D-37073 Göttingen, Germany. Corresponding author: Shu Wang, e-mail: `shu.wang@uni-goettingen.de`.

1 Introduction

A paramount problem in causal analysis of multivariate dynamic systems, such as structural VAR (SVAR) and multivariate GARCH models, is the identification of the underlying structural representation of the data. Such a representation holds great interest for both academics and policy-makers, as it reveals the underlying forces acting upon a financial or macroeconomic system and aids in understanding the transmission mechanisms of policy actions. The identification of the structural form requires external information, which typically takes the form of instrumental variables (e.g. Mertens and Ravn 2013) and theory-motivated restrictions, either on the magnitude (e.g. Sims 1980) or the sign of the parameter (e.g. Faust 1998). However, in cases where such information is scarce or unavailable, a growing literature has emerged on statistical identification methods that leverage the stochastic characteristics of the structural shocks, particularly their heteroskedasticity or statistical independence (see Chapter 14 of Kilian and Lütkepohl 2017 for a literature review).

Several independence-based identification approaches have been explored in the literature by adapting the technique of independent component analysis (ICA), initially developed for blind source separation. While the approach followed by Moneta et al. (2013) relies on an a-priori recursive model structure, the maximum likelihood (ML) approach proposed by Lanne et al. (2017) depends on particular distributional assumptions of the latent shocks. The semi-parametric pseudo ML (PML) method, developed by Gouriéroux et al. (2017), offers consistent estimation even when the assumed non-Gaussian density does not perfectly match the actual probability density function (pdf) of the latent shocks. This approach operates under mild conditions that regulate the degree of misspecification. However, in practical settings, caution is necessary as the true source distributions are often unknown to the analyst, making prior assessments of regularity conditions tentative. If the pseudo densities are severely misspecified, PML estimators may fail to converge to the true structural parameters.

To guard against misspecification of latent distributions, this paper proposes an adaptive ML approach based on non-parametric density estimation, where kernel estimates (Rosenblatt 1971, Silverman 1978) are used to formulate the estimation equations. The consistency and asymptotic normality of the proposed estimator are proven under relatively weak conditions, and extensive simulation-based evidence is provided to demonstrate its favorable finite-sample performance. On the methodological side, we believe that this study contributes to the existing literature in two significant ways. Firstly, it presents an identification approach that is particularly advantageous when prior information on the underlying economic theory or on distributional properties of the structural innovations is lacking, particularly in small sample sizes. Secondly, we develop a less restrictive framework that incorporates partial independence, enabling the incorporation of external information into the model. These contributions make our methodology highly flexible and applicable in various empirical settings.

For a consistent estimation of structural parameters, it is crucial to properly adjust the estimation function to the distribution of latent shocks. Researchers have developed more flexible ICA routines to address this issue (see Hastie and Tibshirani 2002, Bach and Jordan 2003, Samarov and Tsybakov 2004, Boscolo et al. 2004, Chen and Bickel 2006, Robinson and Taylor 2017 and references therein). For instance, Bach and Jordan (2003) use the canonical correlation as a contrast function based on kernel methods. Chen and Bickel (2006) and Robinson and Taylor (2017) estimate the density score function using B-splines and kernel methods, respectively. Similar to our approach, Hastie and Tibshirani (2002) express the pdfs by an exponentially tilted Gaussian and estimate the density function using a cubic spline. We note that Samarov and Tsybakov (2004) and Boscolo et al. (2004) also build upon kernel density estimation, but our method is specifically developed for structural identification in SVAR models within the context of time series analysis. In this regard, our study aligns with the recent contribution of Fiorentini and Sentana (2023), who advocate for the use of discrete local scale mixture of normals (DLSMN) as pseudo densities. Their method entails a joint estimation of structural parameters and shape parameters of the mixture, demonstrating the consistency of ML estimators for key parameters relevant to structural analysis, such as impulse response functions (IRFs). A Bayesian perspective is adopted by Braun (2023), utilizing a Dirichlet process mixture to specify shock distributions. Extending the existing literature, we conduct extensive simulation experiments to study the finite-sample performance of the proposed estimator under diverse source distributions (symmetric vs. asymmetric, uni- vs. multimodal, and leptokurtic vs. platykurtic distributions). Furthermore, we assess the estimator's performance in the presence of various second-order-moment characteristics, including heteroskedasticity and co-heteroskedasticity, comparing it with prominent peers. We discuss the close connection between our method and the PML approach and the generalized method of moments (GMM), which imposes higher-order cross-moment conditions like co-skewness and co-kurtosis (Lanne and Luoto 2021, Keweloh 2021, Guay 2021). Additionally, we compare our suggested approach with fully non-parametric identification schemes, such as

ICA based on distance covariance, which has been successfully applied to macroeconometric models of small to medium sizes (Matteson and Tsay 2017, Herwartz and Wang 2023).

Furthermore, we introduce a novel concept by extending the suggested approach to a less restrictive framework that accommodates systems with dependent components. Recognizing the relevance of modeling dependence among structural shocks in financial and macroeconomic variables, assuming mutual independence of all shocks can be overly restrictive in some applications (see Montiel Olea et al. 2022 for a critical assessment of ICA-based identification). A substantial dimension reduction can be achieved by relaxing the mutual independence assumption and considering partial independence or group-wise independence. For example, assuming a single independent component in an N-dimensional system ($N > 2$) enables the identification of structural quantities related to this component and reducing the number of free parameters by $N - 1$. We also illustrate that full identification can be achieved by introducing only a small number of additional restrictions. Therefore, our proposed approach combines the advantage of maximizing information content through independent shocks with the flexibility to incorporate unmodeled dependence when supported by economic theory. This opens up new possibilities for identification in models of simultaneous causality by combining limited external information with relatively weak statistical properties.

Monte Carlo experiments demonstrate that, when compared to a diverse set of competing approaches, the proposed kernel-based ML (KML) provides the most accurate structural estimates with a competitive level of estimation uncertainty. The DLSMN approach demonstrates the second-best overall performance. This holds true for small and large sample sizes, low and high-dimensional systems, and in the presence of heteroskedastic and coheteroskedastic components. In contrast, the PML approach based on fixed pdfs carries risks of misspecification and inconsistent estimation. Similar to previous findings for ICA by maximizing non-Gaussianity using high-order cumulants (see e.g. Hyvärinen et al. 2001), GMM estimation based on third- and fourth-order cross-moments is sensitive to outliers and extreme data values. It is particularly vulnerable if the source distribution exhibits heavy tails. However, due to its weaker assumptions, the GMM approach exhibits comparable performance to the proposed estimator in large samples or very large systems.

To empirically illustrate the proposed identification approach, we apply it to a stylized model of the global crude oil market (Kilian 2009, Kilian and Murphy 2012). The results demonstrate that mutually independent shocks exhibit structural characteristics consistent with those derived by Kilian (2009). Without imposing a-priori constraints on the price elasticity of oil supply as suggested by Kilian and Murphy (2012), unrestricted KML estimates reveal minimal and statistically insignificant changes in oil production following marginal fluctuations in oil prices. We further enhance the oil market model by introducing a system with one independent shock (related to aggregate flow demand for oil) and two partially dependent shocks (associated with precautionary oil demand and flow supply of oil). This combination of a weaker form of independence assumption and a restriction on the price elasticity of oil supply leads to well-identified shocks. The system with partially dependent shocks exhibits substantial improvements in log-likelihood and provides a sound interpretation of oil-specific demand shocks, primarily capturing precautionary motives. Specifically, the three-dimensional model with partially dependent shocks offers insights into the role of alternative motives for oil demand that align closely with findings in the sign-restricted four-dimensional model of Kilian and Murphy (2014).

The remainder of this paper is organized as follows. In Section 2, we present the KML estimation approach within the framework of SVAR identification. Simulation results in Section 3 provide a comprehensive comparison of the suggested approach with alternative independence-based identification schemes. Section 4 provides empirical evidence on the structural effects in the global oil market. Section 5 concludes. A proof of estimator consistency and asymptotic normality is provided in the Appendix. In the Online Appendix (OA), we provide additional theoretical discussions, simulation results, and supplementary materials to complement the empirical analysis.

2 Methodology

In this Section, we start with an overview of ICA as a tool for identifying structural data representations and its solution through ML estimation. We then introduce the KML estimator and discuss its asymptotic properties. Notably, we extend the application of this estimator to SVARs with a combination of independent and higher-order dependent orthogonal shocks. Additionally, we explore the relationship between PML estimation and the GMM approach.

2.1 Source separation and SVAR models

The identification of the structural representation of a multivariate system typically amounts to solving a system of equations,

$$\mathrm{x}_t = B\xi_t,\ \xi_t \sim (\mathbf{0}, I_N),\ t = 1, \ldots, T, \tag{1}$$

where $\mathrm{x}_t \in \mathbb{R}^N$ collects centered (sensor) variables measured at time t, which can be either directly observed or consistently estimated from the data. The vector $\xi_t \in \mathbb{R}^N$ contains latent (source) variables or shocks, which are mapped to the sensors through an unknown non-singular matrix $B \in \mathbb{R}^{N\times N}$. While observable variables in x_t are correlated with a covariance matrix Σ_{x}, elements in ξ_t represent structural information and are mutually orthogonal.

In empirical finance, for instance, one may consider x_t as returns of N speculative assets, with the (conditional) covariance following a multivariate GARCH representation. Then, the structural parameters in B unravel the contribution of each shock in ξ_t to the returns (see Hafner and Herwartz 2006, Hafner et al. 2022 and references therein). In empirical macroeconometrics, model (1) can be encountered in factor analysis or structural VARs. In the latter case, x_t is the vector of reduced form residuals retrieved from the data. The VAR model – in reduced and structural form, respectively – is

$$A(L)y_t = \mathrm{x}_t,\ \mathrm{x}_t \sim (\mathbf{0}, \Sigma_{\mathrm{x}}) \tag{2}$$

$$\Pi(L)y_t = B^{-1}\mathrm{x}_t = \xi_t, \tag{3}$$

where $y_t \in \mathbb{R}^N$ collects macroeconomic variables (after removing the deterministic terms), $A(L) = \left(I_N - \sum_{j=1}^{p} A_j L^j\right)$ is a p-th order lag-polynomial and $\Pi(L) = B^{-1}A(L)$. Under the weak stationarity condition $\det A(z) \neq 0$ for $|z| \leq 1$, y_t has a Wold moving average (MA) representation, which characterizes the effects of structural shocks on the system in the form of structural impulse response functions (IRFs)

$$y_t = \sum_{i=0}^{\infty} \Phi_i \xi_{t-i} = \Phi(L)\xi_t, \tag{4}$$

where $\Phi(L) = A(L)^{-1}B$ with $\Phi_0 = B$. Thus, the 'impact multipliers' in B carry direct information about reactions of variables in y_t to an exogenous change of shocks in ξ_t.

While the parameters in the reduced form (2), i.e. A_j's and Σ_{x}, can be consistently estimated from the data, the key structural object of interest, i.e. the matrix B, remains unidentified. In specific, the $N(N+1)/2$ equations implied by the moment condition $BB' = \Sigma_{\mathrm{x}}$ are insufficient for solving the N^2 unknowns in matrix B. The identification problem is locally resolved under the following assumption:

(A0) The random vector ξ_t contains N independent components and at most one component exhibits a Gaussian distribution.

As a consequence of the characterization theorem of Darmois–Skitovich (see Th.11 in Comon 1994), under Assumption (A0), the matrix B is uniquely identified up to a right-multiplication by ΛP with Λ being diagonal and P a permutation matrix. That is, B is identified up to column permutations and sign flips. We briefly discuss the ML estimation of the mixing matrix B next.

2.2 Maximum likelihood estimation of independent components

Define C as a matrix square root of Σ_{x} such that $CC^\top = \Sigma_{\mathrm{x}}$. For example, C can be obtained from a spectral or Cholesky decomposition. Left-multiplying (1) or (2) by C^{-1} gives

$$\mathrm{z}_t = C^{-1}\mathrm{x}_t = C^{-1}B\xi_t = Q\xi_t,$$

where Q is orthonormal such that $QQ^\top = I_N$ and $|\det Q| = 1$. It is easy to verify that $\mathrm{z}_t \sim (\mathbf{0}, I_N)$ and $B = CQ$. Prewhitening is a necessary step for the consistency of many ICA techniques (see e.g. Gouriéroux et al. 2017), since it reduces the number of free parameters from N^2 to $N(N-1)/2$. The space of the parameter of interest in Q is restricted to be the set of all N-dimensional orthonormal matrices $\mathcal{O}(N) = \{Q = (q_1, \ldots, q_N) : q_i \in \mathbb{R}^N, q_i^\top q_j = \delta_{ij}, \forall i,j = 1,\ldots,N\}$ with δ_{ij} being the Kronecker delta. Let f_i be the pdf of the i-th component of ξ_t and e_i the i-th column of an identity matrix of conformable dimension. Then, the log-likelihood and gradient are

$$T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\log f_i\left(e_i^\top Q^\top \mathrm{z}_t\right) + \log|\det Q^\top| \quad \text{and} \quad T^{-1}\sum_{t=1}^{T} S(Q^\top \mathrm{z}_t)\mathrm{z}_t^\top + Q^{-1}, \tag{5}$$

respectively, where $S : \mathbb{R}^N \mapsto \mathbb{R}^N$ is the component-wise mapping with $S(\boldsymbol{x}) = (s_1(x_1), \ldots, s_N(x_N))^\top$ and $s_i = f_i'/f_i$, where $'$ denotes the first derivative. It is convenient to use the relative gradient and right-multiply the first order condition by Q. The ML estimate of Q is obtained by solving the system $T^{-1}\sum_{t=1}^{T} S(Q^\top \mathrm{z}_t)(Q^\top \mathrm{z}_t)^\top + I_N = \mathbf{0}$ s.t. $Q \in \mathcal{O}(N)$.

Due to the unknown pdfs of the structural components f_i's, the MLE is often infeasible in practice. To address this, Lanne et al. (2017) assumed a Student-t density for f_i, while Gouriéroux et al. (2017) generalized the approach by resorting to pseudo density functions $\tilde{f}_i$. The PML estimator replaces the true pdfs with $\tilde{f}_i$ and uses $\tilde{S} = (\tilde{s}_1, \ldots, \tilde{s}_N)^\top$, where $\tilde{s}_i = \tilde{f}_i'/\tilde{f}_i$, to achieve estimation. Details of the PML estimator can be found in Gouriéroux et al. (2017). In OA A, we discuss the condition for local concavity of the log-likelihood and demonstrate potential violations for specific fixed pseudo densities. We also highlight the equivalence between the ML approach and mutual information minimization, and that PML can be viewed as a Kullback-Leibler (KL) matching between the distributions $\tilde{f}(Q^\top \mathrm{z}_t)$

and $f(\xi_t)$. The risk of misspecification in the pseudo densities can thus lead to inconsistency in other ICA procedures such as FastICA (Hyvärinen and Oja 1997) as well. Therefore, adjusting the estimation function to the distribution of latent shocks is crucial. Against this background, Fiorentini and Sentana (2023) have recently proposed a non-Gaussian PML approach based on discrete local scale Gaussian mixture densities

$$\tilde{f}_i(x|\varrho) = \sum_{k=1}^{K} \frac{\lambda_k(\varrho)}{\sigma_k(\varrho)} \varphi\left(\frac{x - \mu_k(\varrho)}{\sigma_k(\varrho)}\right), \ i = 1, \ldots, N,$$

where the vector ϱ collects the shape parameters of the mixture densities, $\sum_{k=1}^{K} \lambda_k(\varrho) = 1$ and $\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ (see Fiorentini and Sentana 2023 for a detailed parameterization of mixtures for centered and standardized pseudo densities). By allowing the shape parameters to adapt during estimation, Fiorentini and Sentana (2023) demonstrate consistent estimation of structural IRFs (up to a scaling factor) regardless of the true distribution f_i. In this paper, we formulate the likelihood function based on non-parametric density estimation akin to approaches by Hastie and Tibshirani (2002), Samarov and Tsybakov (2004), and Boscolo et al. (2004). In Section 3, we compare the finite-sample performance of our suggested estimator with PML based on fixed densities and DLSMN. We proceed to outline kernel-based ML estimation.

2.3 Kernel-based maximum likelihood estimation

Denote the i-th component of ξ_t by $\xi_{it} = e_i^\top Q^\top z_t = q_i^\top z_t$. The pdf of ξ_{it} is determined in the form of a kernel-density estimator

$$\hat{f}_{ih}(\xi_{it}) = \frac{1}{Th} \sum_{s=1}^{T} K\left(\frac{\xi_{it} - \xi_{is}}{h}\right), \ i = 1, \ldots, N,$$

where $K(x) : \mathbb{R} \mapsto \mathbb{R}_0^+$ is a bounded univariate kernel function and h is a bandwidth parameter. Assumptions for $K(x)$ and h will be stated later. The kernel density estimate is consistent under mild conditions (see Rosenblatt 1971, Silverman 1978) and will inherit the continuity and differentiability properties of the kernel function. For instance, if $K(x)$ is chosen to be a Gaussian kernel

$$K(x) = (2\pi)^{-1/2} \exp(-x^2/2),$$

the resulting $\hat{f}_i$ will be a smooth curve and infinitely differentiable. In this case, the r-th derivative of the density can be estimated as

$$\frac{d^r}{d\xi_{it}^r} \hat{f}_{ih} = \frac{1}{Th^{r+1}} \sum_{s=1}^{T} (-1)^r \mathcal{H}_r\left(\frac{\xi_{it} - \xi_{is}}{h}\right) K\left(\frac{\xi_{it} - \xi_{is}}{h}\right),$$

where $\mathcal{H}_r(x)$ is the r-th Hermite polynomial (see, e.g. Bhattacharya 1967). Let the orthonormal matrix Q be parameterized by a vector $\theta \in \Theta$ and $\xi_{it}(\theta) := e_i^\top Q(\theta)^\top z_t$. One particularly interesting parameterization of Q under the constraint $Q \in \mathcal{O}(N)$ is given by a sequence of Givens rotation matrices (see Gouriéroux et al. 2017). A typical choice in the literature has the following form:

$$Q(\theta) = \left(\prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \mathcal{G}_{i,j}(\theta_n) \right)^\top ,$$

where θ_n are rotation angles defined on the interval $(0, \pi]$ with $n = 1, \ldots, N(N-1)/2$ and $\mathcal{G}_{i,j}(\theta_n)$ being the Givens matrix that rotates the subspace spanned by axes i and j while holding other axes fixed. The structural orthogonal mixing matrix is then obtained by maximizing the non-parametric (pseudo) log-likelihood function, i.e.

$$\hat{Q} \equiv Q(\hat{\theta}) = \arg\max_{\theta \in \Theta} \sum_{t=1}^{T} \sum_{i=1}^{N} \log \hat{f}_{ih\theta}(\xi_{it}(\theta)). \tag{6}$$

Henceforth, we refer to this estimator as the kernel-based ML (KML) estimator.

The practical implementation of the estimator in (6) requires the analyst to decide upon a kernel function and the bandwidth parameter. Data-driven bandwidth selection is well-known to be a much more important issue than an appropriate selection of the kernel function since the resulting estimates are relatively unaffected by the choice of the kernel (see e.g. Silverman 1998). For bandwidth selection we use – for simplicity – the rule-of-thumb estimator of Silverman (1998), which assumes an underlying Gaussian density and chooses the h that minimizes the mean integrated squared error. This would imply a bandwidth choice of $1.06\hat{\sigma}T^{-1/5}$ if the analyst employs a Gaussian kernel, where the estimated standard deviation $\hat{\sigma}$ is set to unity since the estimator is applied to orthogonalized VAR residuals with unit variance. This practice becomes suboptimal, however, if the analyst has prior knowledge about the data-generating distribution and believes that the true density substantially differs from Gaussianity. For instance, if the underlying density is known to be bi-modal, the rule-of-thumb bandwidth estimator would lead to over-smoothing. A detailed discussion of more refined bandwidth selectors such as cross-validation or plug-in methods, goes beyond the scope of the paper and can be found, e.g. in the review article by Jones et al. (1996). We should therefore keep in mind that our results for the KML estimator could potentially be further improved by a more sophisticated bandwidth selection.

Next, we discuss the asymptotic properties of the estimator and describe a less restrictive framework for partial identification that allows some components to be dependent. Moreover, we also compare the ML approach with recently proposed methods that exploit selected higher-order (co-)moments for identification (Lanne and Luoto 2021, Keweloh 2021, Guay 2021, Hafner et al. 2022).

2.4 Asymptotic properties

In the following, we give results for the consistency and asymptotic normality of the proposed estimator. Let θ_0 be the true parameter vector and $\xi_{it} = \xi_{it}(\theta_0)$ the resulting vector of independent components. Denote the true density of $\xi_{it}(\theta)$ by $f_{i\theta}(\xi_{it}(\theta))$, and the true pseudo log-likelihood by

$$l_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\log f_{i\theta}(\xi_{it}(\theta))$$

and its probability limit

$$l(\theta) = \sum_{i=1}^{N}\mathbb{E}\log f_{i\theta}(\xi_{it}(\theta)).$$

For a given θ, the true log-likelihood is

$$\log L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\log f_{\theta}(\xi_t(\theta)),$$

where $f_\theta(\xi_t(\theta))$ is the joint density of the random vector $\xi_t(\theta)$. Let $\log L(\theta)$ denote the probability limit of the true log-likelihood, i.e. $\log L(\theta) = \mathbb{E}\log f_\theta(\xi_t(\theta))$. Note that at the true parameter θ_0, components in $\xi_t(\theta_0)$ are independent, so that pseudo densities and true densities coincide, which means that $\log L_T(\theta_0) = l_T(\theta_0)$, and $\log L(\theta_0) = l(\theta_0)$. We make the following additional assumptions:

(A1) The parameter space Θ is a compact subspace of $\mathbb{R}^{N(N-1)/2}$.

(A2) f_i is uniformly continuous with uniformly bounded second derivatives on its support, and $\inf_x f_i(x) \geq \epsilon > 0$.

(A3) $h \to 0$ and $Th/\log(T) \to \infty$ as $T \to \infty$.

(A4) The kernel function is such that $K(u) \geq 0$, $\int K(u)du = 1$, $\int |K(u)|du < \infty$, $|u|K(u) \to 0$ as $|u| \to \infty$, $\sup_u K(u) < \infty$, $\sup_u |dK(u)/du| < \infty$ and $K(0) \geq \delta > 0$.

Theorem 1. *Under Assumptions (A0)-(A4),*

$$\hat{\theta} \xrightarrow{p} \theta_0$$

Proof: See Appendix A.

The study of the asymptotic distribution of the estimator of θ is well-known to be challenging due to the random denominator in the estimation of the scores (see e.g. Chen and Bickel 2006, Robinson and Taylor 2017). Under certain assumptions involving, e.g. sample-splitting and trimming conditions for the score estimator, asymptotic normality can be obtained at the usual parametric rate. In the following, for notational convenience, we drop the index θ from $f_{i\theta}(\xi_{it})$. Define the density score

$$s_{it}(\theta_0) := \frac{f_i'(\xi_{it})}{f_i(\xi_{it})}$$

and its estimator

$$\hat{s}_{it}(\theta) := \begin{cases} \frac{\hat{f}_i'(\xi_{it}(\theta))}{\hat{f}_i(\xi_{it}(\theta))}, & \text{if } \xi_{it} \in \Xi,\ \hat{f}_i(\xi_{it}) \geq d_T,\ \hat{f}_i'(\xi_{it}) \leq c_T \hat{f}_i(\xi_{it}) \\ 0, & \text{otherwise,} \end{cases}$$

where we assume that $d_T \to 0$ and $c_T \to \infty$, and Ξ is defined below in (A5). To simplify notation, we will write $s_{it} = s_{it}(\theta_0)$ and $\hat{s}_{it} = \hat{s}_{it}(\theta_0)$. We further introduce notation for the infinite-dimensional nuisance parameter $\tau_i \in \mathcal{T}$, where $\mathcal{T}$ is a pseudo-metric space, $\tau_i = (\tau_{i1}, \tau_{i2})'$, with $\tau_{i1}(\xi_{it}) = f_i'(\xi_{it})$ and $\tau_{i2}(\xi_{it}) = f_i(\xi_{it})$. Let $m_i(\tau_i(\xi_{it}), \theta) = \frac{\tau_{i1}(\xi_{it})}{\tau_{i2}(\xi_{it})} \frac{\partial \xi_{it}}{\partial \theta}$, $\tau_i \in \mathcal{T}$, and $\hat{\theta}$ is a consistent estimator that solves the first order condition $T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} m_i(\hat{\tau}_i(\xi_{it}), \theta) = 0$. We also assume that the population first order condition $\sum_{i=1}^{N} \mathbb{E}[m_i(\tau_{i0}(\xi_{it}), \theta_0)] = 0$ holds. Let $\mathcal{T}$ be equipped with the pseudo-metric $\rho_{\mathcal{T}}$

$$\rho_{\mathcal{T}}(\tau_a, \tau_b) = \left[\int_{\Xi} (m(\tau_a(x)) - m(\tau_b(x)))^2 \, dx \right]^{1/2}, \tag{7}$$

where $\Xi \subset \mathbb{R}$, and the index i and the dependence of $m(\cdot)$ on θ are skipped for simplicity.

For convenience we use sample splitting, i.e. let α_T be an integer sequence such that $\alpha_T/T \to \alpha \in (0,1)$ as $T \to \infty$. Then $\hat{s}_{it}(\theta)$ is estimated from the sub-sample $\xi_{i1}, \ldots, \xi_{i\alpha_T}$ if $t \in \{\alpha_{T+1}, \ldots, T\}$, and from the sub-sample $\xi_{i(\alpha_{T+1})}, \ldots, \xi_{iT}$ if $t \in \{1, \ldots, \alpha_T\}$. This effectively makes $\hat{s}_{it}(\theta)$ i.i.d. and independent of ξ_{it} which simplifies the proof. To establish asymptotic normality we add further assumptions:

(A5) $\sup_{\tau_i \in \mathcal{T}} ||m_i(\tau_i)||_{q,\Xi} < \infty$ for some integer $q \geq 1$ and $i = 1, \ldots, N$, where $||\cdot||_{q,\Xi}$ denotes the L_2-Sobolev norm of order q over the set Ξ, and $\Xi \subset \mathbb{R}$ is an open bounded subset of $\mathbb{R}$ with minimally smooth boundary.

(A6) For some constant K, $m_i(\tau_i(x)) = K \ \ \forall x \in (\mathbb{R} \setminus \Xi) \ \ \forall \tau_i \in \mathcal{T}$, $i = 1, \ldots, N$.

(A7) The bandwidth h is such that $hc_T \to 0$, $Th^3/\log(1/h) \to \infty$, $T^{1/4}h^2 \to 0$, $T^{1/2}h^3 \to \infty$.

Assumption (A5) imposes smoothness conditions on the functions $m_i(\cdot)$, while (A6) requires that functions $m_i(\cdot)$ are constant outside the open bounded set Ξ. The first condition of

Assumption (A7) gives a bound on the bandwidth as a function of the trimming sequence c_T, while the remaining conditions are needed to ensure uniform consistency of the estimator of the density derivatives, and $T^{1/4}$-convergence of the non-parametric estimator of τ_i. Note that, unlike many results on adaptive estimation such as Bickel (1982), these assumptions do not exclude asymmetric densities, at the expense of somewhat stronger smoothness conditions (A5) and (A6), and tighter bandwidth sequences (A7).

Theorem 2. *Under Assumptions (A0) to (A7),*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1})$$

where

$$\begin{aligned} I(\theta_0) &= \sum_{i=1}^{N} \mathbb{E}\left[s_{it}^2 \frac{\partial \xi_{it}(\theta_0)}{\partial \theta} \frac{\partial \xi_{it}(\theta_0)}{\partial \theta^\top} \right], \\ J(\theta_0) &= -\sum_{i=1}^{N} \mathbb{E}\left[s_{it} \frac{\partial^2 \xi_{it}(\theta_0)}{\partial \theta \partial \theta^\top} + \frac{\partial s_{it}}{\partial \xi_{it}} \frac{\partial \xi_{it}(\theta_0)}{\partial \theta} \frac{\partial \xi_{it}(\theta_0)}{\partial \theta^\top} \right]. \end{aligned}$$

Proof: See Appendix A. A consistent estimator of the asymptotic covariance matrix is obtained by replacing expectations by sample averages, and the score by its non-parametric estimate. Detailed expressions are provided in the proof.

For the main purpose of this study, the KML estimation has to be put into the context of its application in dynamic systems such as VARs, where the vector x_t is estimated and thus depends on the (reduced-form) autoregressive parameters. To jointly account for estimation uncertainty of both the reduced-form and structural parameters, we evaluate the sampling distribution of the structural parameters of interest in finite samples based on a moving-block bootstrap approach (Brüggemann et al. 2016). This will be illustrated in Section 4. Concerning the efficiency of the estimator, we demonstrate in a detailed Monte Carlo study that our proposed estimator substantially outperforms existing approaches for independence-based identification.

2.5 Systems with dependent components

The assumption that all components in ξ_t are independent is strong. In fact, structural shocks are by definition only orthogonal (see e.g. Ramey 2016). In many financial and macroeconomic models, it appears more reasonable to allow dependence among certain shocks (see, e.g. Montiel Olea et al. 2022). In the prominent three-dimensional model for the global oil market of Kilian (2009), for instance, there might indeed exist a dependence between a speculative oil demand shock and an oil supply shock. It seems plausible that the market participants may want to store and secure more oil if there appears to be a higher chance of a potential disruption or bottleneck in oil supply. However, as long as the aggregate demand shock is independent from the other shocks, it can be partially identified.

The issue of identification under partial dependence can be cast into the multidimensional ICA (MICA) framework of Cardoso (1998). Let $E_1, \ldots, E_c$ be c linear sub-spaces of $\mathbb{R}^N$ and $Q_1, \ldots, Q_c$ be c matrices of dimension $N \times n_1, \ldots, N \times n_c$ such that $Q = [Q_1, \ldots, Q_c]$ has full column rank. Then $Q\xi_t$ admits a MICA decomposition onto the spaces $E_1, \ldots, E_c$, where E_i is the n_i-dimensional linear subspace spanned by Q_i, i.e. $E_i = Span(Q_i)$ for $1 \leq i \leq c$. If the first $c-1$ subspaces allow for decompositions into individual mutually independent components, we have $n_1 = \ldots = n_{c-1} = 1$ and $n_c = N-c+1$, which corresponds to a canonical MICA decomposition in the sense that the last component Q_c cannot be further decomposed into independent non-Gaussian components. As a result, the last component with $n_c \geq 2$ is only identified up to a right-multiplication by an invertible $n_c \times n_c$ matrix. Nevertheless, the one-dimensional independent components of this decomposition are uniquely identified up to sign and permutation, as in the usual ICA case. Cardoso (1998) proposes a two-step procedure to estimate the MICA model: In the first step, estimate mono-dimensional ICA components by a standard ICA routine, and in the second step determine which estimated innovations are independent, and which other components should be grouped together due to mutual dependence. Cardoso (1998) argues that a solution to the ICA problem exists, if the estimation equations underlying the first step have the form,

$$T^{-1} \sum_{t=1}^{T} \tilde{S}(Q^\top \mathrm{z}_t)(Q^\top \mathrm{z}_t)^\top + I_N = \mathbf{0},$$

where $\tilde{S} : \mathbb{R}^N \mapsto \mathbb{R}^N$ is a properly chosen component-wise non-linear mapping. Our estimator also has this form by setting $\tilde{s}_i = f_i'/f_i$ and replacing the unknown f_i and its derivative by kernel estimates, which under our assumptions are consistent estimators of the unknown quantities. Therefore, in large samples, a solution to the ICA problem exists with probability converging to 1, if one used the KML in a two-step procedure. Note that some measures like the mutual information have to be used to group the dependent components as parts of a multidimensional component in the second step. With this being said, the approach proposed in this paper allows us to directly recover the $N-n_c$ independent components given any fixed dimension ($n_c \geq 2$) of a subspace containing mutually dependent ones in a single step.

Let $\mathcal{M} = \{C_\alpha : \alpha \in \mathcal{C}\} \equiv \binom{N}{N-n_c}$ be the collection of sets containing all combinations of $N-n_c$

components selected from $\{1, \ldots, N\}$, $\mathcal{C} = \{1 \ldots, \frac{N!}{(N-n_c)!n_c!}\}$ and $C_\alpha^- = \{1, \ldots, N\} \setminus C_\alpha$. Given the identifiability of the subsystem, the (pseudo) log-likelihood function is maximized,

$$\max_{\theta \in \Theta} \max_{C_\alpha \subset \mathcal{M}} \sum_{t=1}^{T} \left[\sum_{i \in C_\alpha} \log \hat{f}_{ih}(\xi_{it}(\theta)) + \log \hat{f}_{C_\alpha^-,h}(\xi_{C_\alpha^-,t}(\theta)) \right],$$

where $\hat{f}_{C_\alpha^-,h}$ is the n_c-dimensional joint pdf of the n_c dependent shocks in $\xi_{C_\alpha^-,t}$ estimated by

$$\hat{f}_{C_\alpha^- h}(\xi_{C_\alpha^-,t}) = \frac{1}{Th^{n_c}} \sum_{s=1}^{T} \boldsymbol{K} \left(\frac{\xi_{C_\alpha^-,t} - \xi_{C_\alpha^-,s}}{h} \right),$$

where $\boldsymbol{K}(\mathbf{x}) : \mathbb{R}^{n_c} \mapsto \mathbb{R}_0^+$ is a multivariate kernel function. For example, a standard multivariate Gaussian kernel takes the form

$$\boldsymbol{K}(\mathbf{x}) = (2\pi)^{-n_c/2} \exp\left(-\frac{\mathbf{x}^\top \mathbf{x}}{2} \right).$$

Although the parameters in the remaining n_c columns remain unidentified, the partial identification reduces the dimension of the parameter of interest from $\mathbb{R}^{N(N-1)/2}$ to $\mathbb{R}^{n_c(n_c-1)/2}$. For example, in the case of one independent component in the system as outlined earlier for the global oil market, a unique solution (up to a sign flip) for the first column in the matrix Q can be obtained. By partially identifying the independent aggregate demand shock, only one free parameter (out of three) remains unidentified. Full identification can be achieved by incorporating one additional restriction. Moreover, if it is assumed that $\xi_{C_\alpha^-,t}$ contains group-wise independent components, which is a weaker assumption compared to mutual independence, the number of structural parameters of interest can be further reduced. For instance, in a five-dimensional model representing an economic union with an area-wide shock and two country-specific shocks for each member state, imposing independence of the area-wide shock identifies this shock (along with its associated properties) and reduces the number of free parameters from 10 to 6. Further imposing independence between shocks in one country and those in the other, while allowing for mutual dependence within each country, reduces the number of free parameters from 6 to 2 (i.e. one parameter for each country). The two-dimensional subspaces can be estimated by decomposing the term $\log \hat{f}_{C_\alpha^-,h}(\xi_{C_\alpha^-,t}(\theta))$ into the sum of two joint log-likelihood functions.

Therefore, the proposed approach introduces new possibilities for identification in models of simultaneous causality by achieving a substantial dimension reduction through weaker forms of independence assumptions. Full identification can be achieved by introducing only a small number of additional restrictions. We demonstrate this in the empirical application. Next, we briefly discuss the connection between PML approaches and the method of moments.

2.6 Method of moments

It is suggestive to write down the population first order condition using the relative gradient:

$$\mathbb{E}\left[\tilde{S}(Q^\top \mathrm{z}_t)(Q^\top \mathrm{z}_t)^\top\right] + I_N = \mathbf{0}. \tag{8}$$

The $[i, j]$-th element of the first term is $\mathbb{E}\left[\tilde{s}_i(e_i^\top Q^\top \mathrm{z}_t)e_j^\top Q^\top \mathrm{z}_t\right]$. The off-diagonal elements satisfy

$$\mathbb{E}\left[\tilde{s}_i(e_i^\top Q^\top \mathrm{z}_t)e_j^\top Q^\top \mathrm{z}_t\right] = \mathbb{E}\left[\tilde{s}_i(\xi_{it})\right]\mathbb{E}\left[\xi_{jt}\right] = 0,\ i \neq j, \tag{9}$$

where the first equality is the result of independence.

Actually, $\tilde{f}$ does not have to be a proper pdf. For instance, if one chooses $\tilde{f}^{\kappa_3}(x) = \exp\left(x^3/3\right)$ with $\tilde{s}^{\kappa_3}(x) = x^2$ or $\tilde{f}^{\kappa_4}(x) = \exp\left(x^4/4\right)$ with $\tilde{s}^{\kappa_4}(x) = x^3$, maximizing the corresponding likelihood function is equivalent to an ICA by means of maximizing third- or fourth-order moments or maximizing non-Gaussianity, if the skewness or kurtosis is considered as a measure for non-Gaussianity, respectively. Moreover, plugging $\tilde{s}^{\kappa_3}$ and $\tilde{s}^{\kappa_4}$ into equation (9) gives the moment conditions

$$\mathbb{E}[\xi_{it}^2\xi_{jt}] = 0 \quad \text{and} \quad \mathbb{E}[\xi_{it}^3\xi_{jt}] = 0,\ i \neq j,$$

respectively. These moment statistics are referred to as the 'co-skewness' and 'co-kurtosis', based on which GMM estimation has been recently developed by Lanne and Luoto (2021), Guay (2021), Keweloh (2021) and Hafner et al. (2022). In fact, these moment conditions are weaker than Assumption (A0), since independence implies zero co-moments but the converse is not true. Nevertheless, GMM requires the existence of the moments imposed. Another important merit of GMM is that it allows various forms of conditional heteroskedasticity – in particular the common stochastic volatility – if the estimation is conducted based on asymmetric moment conditions displayed above. However, an important drawback of estimation based on higher-order moments (or cumulants) is that the results are very sensitive to outliers or extreme values in the data. As we illustrate in the Monte Carlo experiment, if the source distribution has heavy tails, estimates obtained by GMM are subject to substantially larger estimation errors in comparison with KML.

3 Monte Carlo experiments

In this Section, we compare the performance of the KML approach with some prominent alternatives by means of simulation experiments. We first describe the simulation design and performance criteria. After discussing simulation results for stylized bivariate systems in detail, we sketch some Monte Carlo evidence for higher-dimensional systems in small samples, and for (actually dependent) structural shocks that are subject to non-modelled (co-)heteroskedasticity. For Monte Carlo evidence regarding partial identification in systems with dependent shocks, the reader may refer to OA B.3.

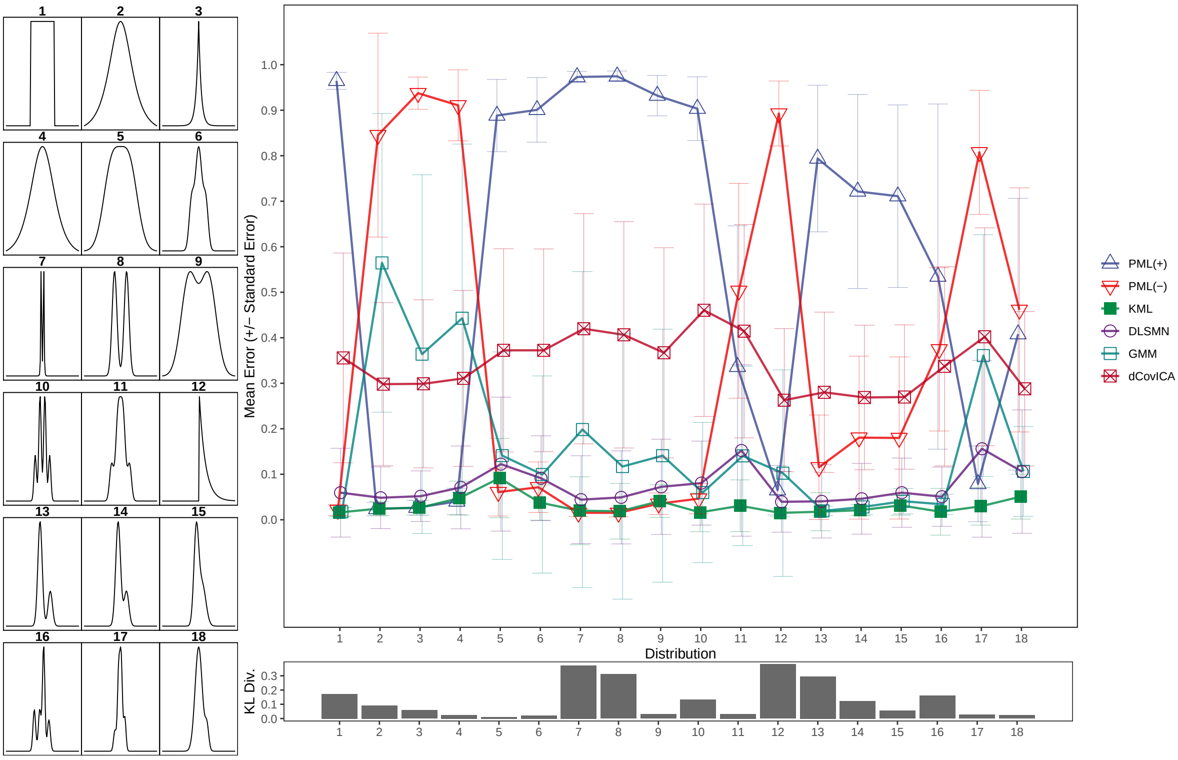


Figure 1: The left-hand-side panel shows 18 data-generating-distributions, including uniform, Student-t, exponential, symmetric and asymmetric mixtures of Gaussian, exponential distributions. Distributions 1–11 are symmetric and distributions 12–18 are asymmetric. The right-hand-side panel shows the average estimation error as given by (12) $\pm$ one estimated standard error for alternative methods and distributions based on 1,000 simulation experiments with a sample size of $T = 1,000$, as well as the KL divergence between the shock distributions and the standard Gaussian distribution.

3.1 Simulation design

For ICA by means of KML, we document only results based on a general and rather agnostic setting with a Gaussian kernel and the bandwidth selected by the rule-of-thumb. Notably, KML performance is fairly unaffected by both kernel and bandwidth choice. As alternative ICA approaches, we first consider PML estimation using a hyperbolic secant (PML(+)) and a Gaussian mixture (PML(−)) as fixed pseudo densities with the respective non-linearities

$$\tilde{s}_i^+(x) = -\frac{\pi}{2}\tanh\frac{\pi}{2}x \quad \text{and} \quad \tilde{s}_i^-(x) = 2\pi x + \frac{\pi}{2}\tanh\frac{\pi}{2}x,\ i = 1,\ldots,N. \tag{10}$$

In addition to fixed densities, we consider ML estimation based on unrestricted two-component DLSMN densities ($K = 2$), as studied in the Monte Carlo analysis of Fiorentini and Sentana (2023). Furthermore, we implement GMM estimation based on a set of third- and fourth-order moment conditions, i.e.

$$\mathbb{E}[\xi_{it}^2 \xi_{jt}] = 0, \quad \text{and} \quad \mathbb{E}[\xi_{it}^3 \xi_{jt}] = 0,\ \mathbb{E}[\xi_{it}^2 \xi_{jt}^2] - 1 = 0,\ \text{for } i \neq j \text{ and } i, j \in \{1, \ldots, N\}, \quad (11)$$

and an approach proposed by Matteson and Tsay (2017), that aims at the minimization of the distance covariance statistics of Székely et al. (2007) as a measure of dependence (dCovICA). The function for dCovICA is available in the `R` package `steadyICA` (Risk et al. 2015). Completing the set of alternative estimators, we show simulation results for some more conventional tools of ICA including the Joint Approximate Diagonalization of Eigenmatrices (JADE), and the so-called FastICA and KernelICA algorithms in OA B.1 and OA B.2, where we also provide further references regarding these established algorithms and comments on their implementation.

We start with a bivariate model in the baseline experiments and consider larger systems later in this Section. In each experiment, we simulate a pair of observable variables $\mathrm{x}_t \in \mathbb{R}^2$ as $\mathrm{x}_t = B\xi_t,\ t = 1, \ldots, T$ with sample sizes $T = 200$, 1,000 and $5,000$. Similar to previous studies on ICA (see e.g. Hastie and Tibshirani 2002), the independent components $\xi_t \in \mathbb{R}^2$ are simulated from 18 distributions that have been widely used as benchmarks to evaluate the performance of alternative ICA algorithms (see e.g. Bach and Jordan 2003, Hastie and Tibshirani 2002, Matteson and Tsay 2017). In particular, these distributions are depicted in the left-hand-side panel of Figure 1, and closely approximate those used by Matteson and Tsay (2017). The alternative distributions cover a wide range of stochastic characteristics that are relevant in empirical practice, such as symmetry vs. asymmetry, uni- vs. multimodality, sub- vs. super-Gaussian as well as nearly-Gaussian distributions. Their KL divergences from the standard Gaussian distribution are displayed in the bottom-right panel of Figure 1, ranging from approximately 0.01 to 0.40. The independent components simulated from all data-generating distributions are adjusted by their population mean and variance, such that they all have a zero mean and unit variance in population. The mixing matrix $B \in \mathbb{R}^{2\times 2}$ is formulated as

$$B(\rho, \theta) = C(\rho)Q(\theta) = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

where in each experiment the correlation of the observable variables (ρ) and the rotation angle (θ) are uniformly drawn from $(-1; 1)$ and $[-\pi; \pi]$, respectively.

ICA estimates are evaluated in terms Amari distances (Bach and Jordan 2003), i.e.

$$d_{Amari}(B,\hat{B}) = \sum_{i=1}^{N}\left(\frac{\sum_{j=1}^{N}|r_{ij}|}{\max_j |r_{ij}|}-1\right)+\sum_{j=1}^{N}\left(\frac{\sum_{i=1}^{N}|r_{ij}|}{\max_i |r_{ij}|}-1\right),\ r_{ij}=(B^{-1}\hat{B})_{ij}. \tag{12}$$

Notice that the Amari distance is invariant to sign slips and column permutations.

3.2 Simulation results

Figure 1 shows simulation results for alternative distributions with a sample size $T = 1,000$, which is – to some extent – informative for the asymptotic properties of the alternative estimators under specific distributions of the structural components (similar figures for cases $T = 200$ and $5,000$ can be found in OA B.1). Overall, the suggested KML and the DLSMN estimator are unique in the sense that these are among the best-performing estimators irrespective of the underlying distributional model. On average, the KML approach shows a stable performance lead over DLSMN. Unsurprisingly, the performance of PML depends on the underlying pseudo pdf assumed ($\tilde{f}$) and the shape of the true source distribution (f). PML estimation could be subject to a sizeable estimation bias once $\tilde{f}$ fails to capture the higher-order characteristics of f. Notably, PML(+) (PML(−)) suffers from inconsistency if the data-generating distribution is platykurtic (leptokurtic) or in general multimodal. Moreover, for certain Gaussian mixtures (11 and 16), neither of the assumed pseudo pdfs in (10) perform well. As mentioned previously in Section 2.6, extreme realizations can lead to a distortion of GMM estimation based on higher-order moments in (11). If the underlying shocks exhibit a super-Gaussian distribution with heavy tails (2–4), the performance of the GMM approach is weak. ICA estimation through the minimization of the distance covariance as measure of dependence (dcovICA) performs quite robustly over the set of underlying distributions. On average, however, KML and DLSMN clearly outperform this approach.

The simulation results in Table 1 provide insights for scenarios where analysts have limited information about the distributional origin of the structural components. Structural shocks are randomly sampled from the 18 alternative distributions, while analysts use specific ICA approaches based on ad-hoc selection. The mean estimation errors ($\bar{d}_{Amari}$) in Table 1 summarize the average performance of each estimation procedure across all alternative distributional models and alternative sample sizes of $T = 200, 1000$ and $T = 5000$. Throughout, we find that the suggested KML approach results in Amari distances that are smallest on average, and also exhibit smallest standard deviations. Interestingly, 95% intervals of KML-based performance measures do not overlap with corresponding intervals provided for DLSMN being the second-best estimation device. Both KML and DLSMN estimation are consistent,

		PML(+)	PML(−)	KML	DLSMN	GMM	dCovICA
$T = 200$	$\bar{d}_{Amari}$	.537	.383	.098	.153	.244	.362
	$s.e.(d_{Amari})$	.373	.352	.172	.189	.299	.214
	$CI^{95\%}(d_{Amari})$	[.517, .556]	[.365, .402]	[.089, .107]	[.143, .162]	[.229, .260]	[.351, .373]
$T = 1,000$	$\bar{d}_{Amari}$	.574	.363	.031	.071	.163	.350
	$s.e.(d_{Amari})$	.402	.377	.049	.102	.276	.218
	$CI^{95\%}(d_{Amari})$	[.553, .595]	[.343, .382]	[.028, .033]	[.066, .077]	[.149, .178]	[.338, .361]
$T = 5,000$	$\bar{d}_{Amari}$	.585	.338	.012	.044	.137	.195
	$s.e.(d_{Amari})$	.440	.394	.011	.082	.283	.360
	$CI^{95\%}(d_{Amari})$	[.561, .609]	[.316, .359]	[.011, .013]	[.040, .049]	[.121, .152]	[.176, .215]

Table 1: Mean ($\bar{d}_{Amari}$) and standard deviation ($s.e.(d_{Amari})$) of the estimation error. Numbers in brackets show confidence intervals for the mean estimation error with 95% coverage (determined as $\bar{d}_{Amari} \pm 1.96 s.e.(d_{Amari})/\sqrt{1,000}$).

as evidenced by the convergence of estimation errors and standard deviations to zero. In contrast, alternative estimators, such as PML(+) and PML(−), suffer from inconsistency, leading to larger average estimation errors and standard deviations.

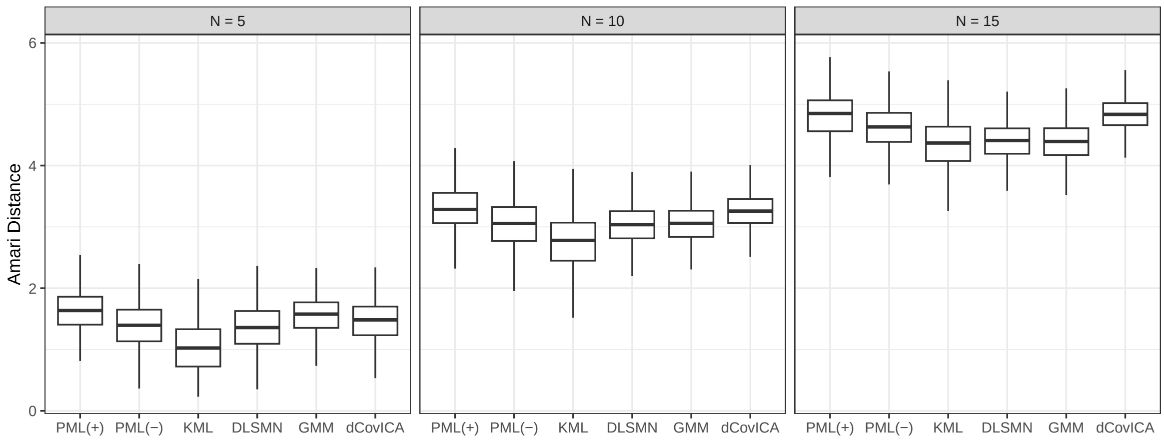


Figure 2: Distribution of the Amari distance illustrated through boxplots for alternative methods, system dimensions N, and shocks from distinct distributions with $T = 50$.

3.3 Small samples and higher dimensions

Allowing the estimation equation to adjust to the source distribution is crucial for accurately estimating the causal scheme, especially when data is limited. We compare the performance of different identification approaches in small-sample settings ($T = 50$) and systems of higher dimensions ($N = 5$, 10, and 15). The distributions of the shocks are uniformly drawn (with replacement) from the 18 distributions displayed in Figure 1. To simulate the structural mixing matrix B in an N-dimensional system, we first draw N^2 standard Gaussian variables and put them in a $N \times N$ matrix denoted by M in each experiment. Then, a covariance matrix is generated as $\widetilde{\Sigma} = M^\top \mathrm{diag}(N, \ldots, 1)M$ and standardized to obtain unit variance $\Sigma = \widetilde{\Sigma}\ \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_N)$, where σ_i^2 is the i-th main diagonal element of $\widetilde{\Sigma}$. Finally, $B = C(\Sigma)Q(\theta)$ where $C(\Sigma)$ is the lower-triangular Cholesky factor of Σ and entries in the $N(N-1)/2$-dimensional vector θ are uniformly drawn from $[-\pi; \pi]$. The results in Figure 2 demonstrate that as the system dimension increases, the estimation error, measured by the Amari distance, increases for all methods. The KML estimator outperforms other methods in terms of median estimation error, regardless of system size. In the larger systems ($N = 10, 15$), DLSMN and GMM estimators also show competitive performance.

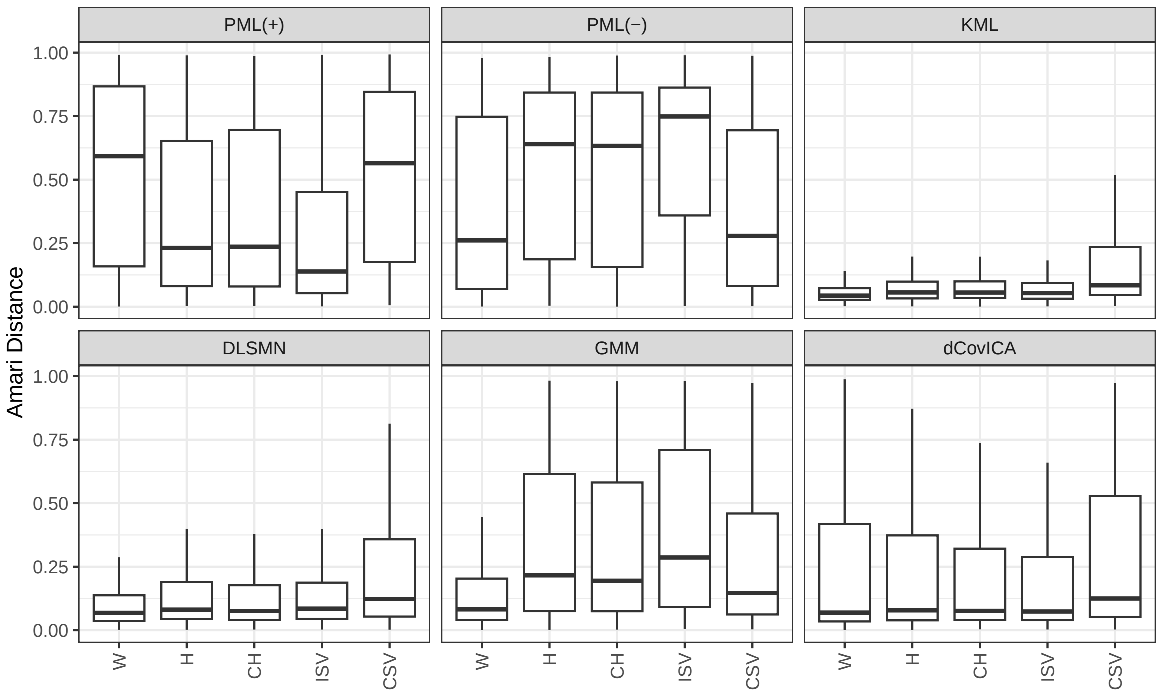


Figure 3: Distribution of the Amari distance illustrated through boxplots for various methods and DGPs with $T = 200$, which incorporate shocks with distinct distributions and second-order moment features (W: white noise, H: heteroskedasticity, CH: co-heteroskedasticity, ISV: independent stochastic volatility, CSV: common stochastic volatility).

3.4 Heteroskedastic components

In addition to the baseline settings of homoskedastic components, we also assess the performance of alternative ICA procedures in the presence of heteroskedastic and co-heteroskedastic shocks. In specific, after generating the shocks as in the baseline bivariate experiments, we sample a Gaussian random variable $w_t \sim \mathcal{N}(-g^2/2, g^2)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian law with mean μ and variance σ^2 – either independently for each shock, or commonly for both shocks – and multiply the original shocks by $\exp w_t$. In addition, we consider cases where components follow stochastic volatility (log-variance) processes, which are either specific to each shock or common to all shocks. Results depicted in Figure 3 underline that the performance of KML and DLSMN is markedly stable under (co-)heteroskedasticity, while KML estimation has a considerable performance lead. Detailed descriptions and simulation results for alternative heteroskedastic DGPs and sample sizes are documented in OA B.2.

3.5 Summary

Summarizing simulation-based evidence from this Section (as well as from OA B), we conclude that KML has outstandingly accurate and stable performance within a family of alternative ICA-based methods for identifying structural relationships. We identify DLSMN estimation as a robust second-best estimation approach. In particular, the suggested KML approach is not inferior to pseudo ML estimation under the favorable condition that the a-priori presumed pseudo densities match the actual distributions of the independent components. Hence, in the case that an analyst has little or no information about the distribution or higher-order-moment features of the underlying structural components, the KML estimation might be considered as the first choice for structural estimation and inference.

4 Structural signals underlying the global crude oil market

Oil price changes play a crucial role in determining the business cycle and consumer prices in developed economies. To analyse the effects of oil price variations on major macroeconomic aggregates, it has become a convention to resort to the structural representation of such variations. In our empirical application, we apply the KML approach to reexamine the SVAR model of the global oil market introduced by Kilian (2009). The model comprises three variables $y_t = (\Delta prod_t, rea_t, rpoil_t)$, where $\Delta prod_t$ is the percentage change in world crude oil production, rea_t is an index of global real economic activity constructed by Kilian (2009) using dry cargo bulk freight rates, and $rpoil_t$ is the log real price of oil. Monthly time series data have been updated by Kilian and Murphy (2014) and cover the period 1973M2 – 2009M8. Following the benchmark study, we include 24 lags and an intercept term in the reduced-form specification. After estimating the reduced-form model, KML estimation is applied to the orthogonalized residuals. While two-step estimation procedures are prevalent in the SVAR literature, it is noteworthy that joint estimation of reduced-form and structural parameters can enhance efficiency (see, e.g. Fiorentini and Sentana 2023). Nevertheless, one-step estimation is computationally demanding, and the two-step method offers the advantage of isolating the estimation of structural parameters, so that it enables comparisons of alternative identification schemes. Furthermore, it is imperative to test the key identifying assumption (A0) (non-Gaussianity and independence) for ICA-based approaches (see Montiel Olea et al. 2022, Amengual et al. 2022). Results from specification tests provided in OA C.1 confirm that assumption (A0) holds with strong significance, supporting the use of ICA-based approaches for identifying oil price shocks.

Much of the related literature considers the following structural shocks behind observable dynamics: (i) an 'oil supply shock' moving the global production of oil, (ii) an 'aggregate demand shock' that shifts the demand for oil as a reflection of global economic activity, and (iii) an 'oil specific demand shock'. While the second shock is necessarily a shock to the 'flow' of oil, the third shock might also stem from 'precautionary' motives to stabilize the oil 'stocks'. Identification of structural shocks to the global oil market is one of the most viable issues in contemporary empirical macroeconomics (see Kilian and Zhou 2023 for a recent review). To point identify the shocks of interest, Kilian (2009) suggests a recursive causal structure, with the implication of a vertical impact supply curve of crude oil.

In this Section, we first discuss the structural implications (IRFs) of a model evaluated under the mutual independence assumption using the KML approach. We also examine alternative estimates of the price elasticity of oil supply. Additionally, we explore the capability of the KML approach to evaluate joint densities under the assumption of partial dependence. We compare the KML-based models and the benchmark models of Kilian (2009) and Kilian and Murphy (2014) by examining the implied historical decompositions of real price of oil.

4.1 Impulse responses and shock labelling

Figure 4 presents IRFs as implied by the estimated structural parameter matrix using the KML approach under the assumption of mutual independence. To account for estimation uncertainty, we employ bootstrap techniques commonly used in empirical SVAR studies. In selecting the bootstrap scheme, we rely on the results of Brüggemann et al. (2016) who have shown that stylized bootstrap schemes (e.g. wild bootstrap) fail to strictly approximate the asymptotic distribution of statistics that depend on both autoregressive and covariance parameters. As structural IRFs fall into this category, we employ the residual-based moving block bootstrap (MBB) advocated by Brüggemann et al. (2016) and follow their suggestion by

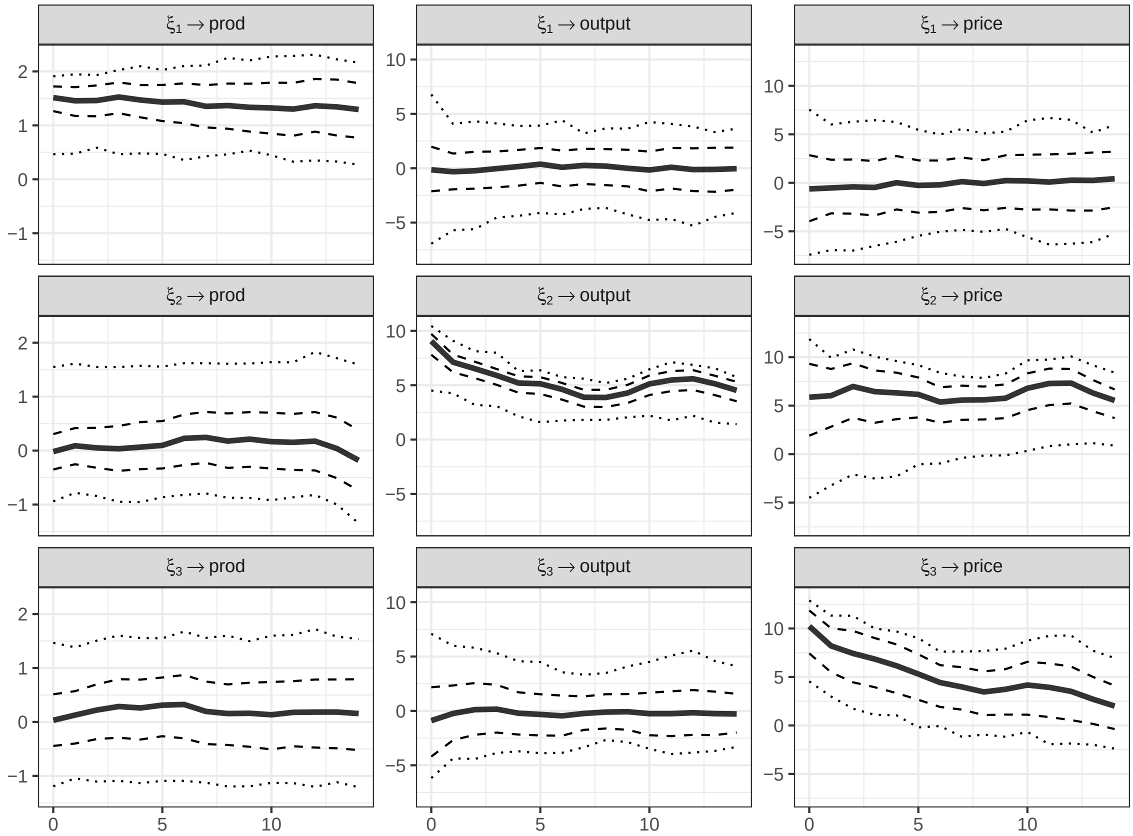


Figure 4: Impulse responses to one-standard-deviation shocks obtained from imposing mutual independence. The structural shocks are oil supply shock (ξ_1), aggregate demand shock (ξ_2) and speculative oil demand shock (ξ_3). The solid, dashed and dotted lines indicate the median, point-wise and joint Wald confidence bands with a coverage of 90% obtained by a moving block bootstrap with 1,000 replications, respectively.

setting the block length to approximately $5.03T^{1/4} \approx 23$. For detailed analysis of asymptotic coverage probabilities and simulation-based evidence, we refer the reader to Brüggemann et al. (2016). Moreover, since point-wise confidence intervals do not account for the dependence of IRFs across horizons and different response functions, these intervals have been criticized for insufficient coverage accuracy (see e.g. Inoue and Kilian 2016). In order to account for estimation uncertainty jointly for all IRFs and at all horizons, we use the Wald approach developed by Lütkepohl et al. (2015) and combine it with the MBB procedure (see OA C.2 for details). For convenience of the reader, we also document so-called shotgun trajectories in OA C.6. As the structural identification purely relies on the statistical properties of the shocks, it is unclear if the detected independent components can be reasonably considered as structural shocks in an economic sense. However, the directional patterns of the KML structural IRFs align closely with the findings of Kilian (2009), particularly regarding the insignificance of oil supply shocks on the real price of oil across all horizons. Furthermore, the oil price response to oil-specific demand shocks exhibits signs of overshooting, confirming the benchmark results. Moreover, the price responses to shocks in the flow demand for oil are largely stable over the displayed horizons and maintain long-run significance. In sum, we can conclude that the independent components retrieved by means of KML estimation allow for a sound economic labelling as structural shocks with similar features as those reported by Kilian (2009).

4.2 Estimating the short-run price elasticity of oil supply

Kilian and Murphy (2012) strongly advocate for the imposition of an upper bound of 0.0258 on the short-run price elasticity of oil supply. Specifically, the authors underpin that 'the prevailing view in the literature that the short-run oil supply elasticity is very low and perhaps zero' (p. 1175). A similar reasoning also applies to the price elasticity of oil supply given a change in the aggregate demand. Conditional on the model representation, Kilian and Murphy (2012) suggest to determine impact elasticities in the form of the ratios of parameters $\varepsilon_s = b_{13}/b_{33}$ (oil-specific demand or demand for stocks of oil) and $\varepsilon_f = b_{12}/b_{32}$ (flow demand for oil). The construction of impact elasticities directly from the structural parameter matrix has been recently critically discussed by Baumeister and Hamilton (2022) who notice that such parameters of interest are defined for the observable model variables. Accordingly, these authors argue for determining elasticities of interest from the inverse of the structural parameter matrix, i.e. B^{-1}. With $\tilde{b}_{ij}$ denoting typical elements of B^{-1}, the short-run price elasticity of oil supply in the vein of Baumeister and Hamilton (2022) is $\tilde{\varepsilon}_s := -\tilde{b}_{13}/\tilde{b}_{11}$. KML estimates support the notion that oil supply exhibits highly inelastic responses to price changes. The estimated elasticities $\varepsilon_s, \varepsilon_f$ and $\tilde{\varepsilon}_s$ are 0.003, -0.003 and 0.003, respectively, and 90% confidence intervals include zero throughout. In OA C.4, we explore alternative approaches to point identification using ICA methods (discussed in Section 3) and examine whether they yield similar magnitude estimates for the elasticities.

4.3 A model with partially dependent shocks

Applying the KML principle to a system of three mutually independent shocks yields a log likelihood statistic of -1569.193. Alternatively, we adopt the KML approach to maximize the joint density by allowing dependence between a univariate independent shock and a bivariate vector of orthogonal shocks, where the oil supply and precautionary demand shocks can exhibit higher-order dependence. Allowing for such dependence could be meaningful, since agents may worry about potential disruptions in supply chains following announcements of reduced oil production by major suppliers like OPEC, leading to additional precautionary demand due to heightened uncertainty. This is supported by the structural IRFs in Figure 4, which indicate overshooting price effects resulting from precautionary demand shocks. The KML estimation yields an improved log-likelihood of -1549.105, indicating that the data align well with the proposed refinement of the distribution of structural shocks.

While the independent aggregate demand shock (ξ_{2t}) is identified by the KML approach, two dependent shocks are left unidentified. To identify the precautionary demand shock, we impose a short-run restriction on the price elasticity of oil supply, i.e. $b_{13} = 0$. Then, the estimate for θ^* is obtained by solving the equation

$$e_1^\top \hat{C}\hat{Q} \begin{bmatrix} \cos\theta^* & 0 & -\sin\theta^* \\ 0 & 1 & 0 \\ \sin\theta^* & 0 & \cos\theta^* \end{bmatrix} e_3 = 0,$$

where e_1 and e_3 are the first and third column of I_3. The resulting structural IRFs are displayed in Figure C.2 in OA C.3. It emerges that the shape and magnitude of the IRFs are largely similar between models with mutual independence and partial dependence of structural shocks, except for the restricted response. Additionally, due to the impact restriction, the dynamic effect of flow demand shocks on oil production exhibits a similar level of precision as in Kilian (2009).

4.4 Historical decompositions

As a natural complement of average structural effects, historical decompositions shown in Figure 5 enable a more specific view at the time-varying sources of oil price in terms of the underlying identified shocks. Figure 5 is in full analogy to Figure 2 of Kilian and Murphy (2014), except for the fact that we provide a comparison among alternative structural models, i.e. (i) the recursive model of Kilian (2009), (ii) the model implied by the evaluation of mutual independence, and (iii) the model of partial dependence.

The historical decompositions are generally very similar to those of Kilian and Murphy (2014). Regarding the contributions of the oil price shock, however, we notice that the results displayed in Figure 5 are closer to zero impacts in comparison with benchmark outcomes. To explain this heterogeneity, it is worth noticing that the three-dimensional model might be subject to an aggregation of oil-supply shock effects with those of the unidentified shock in

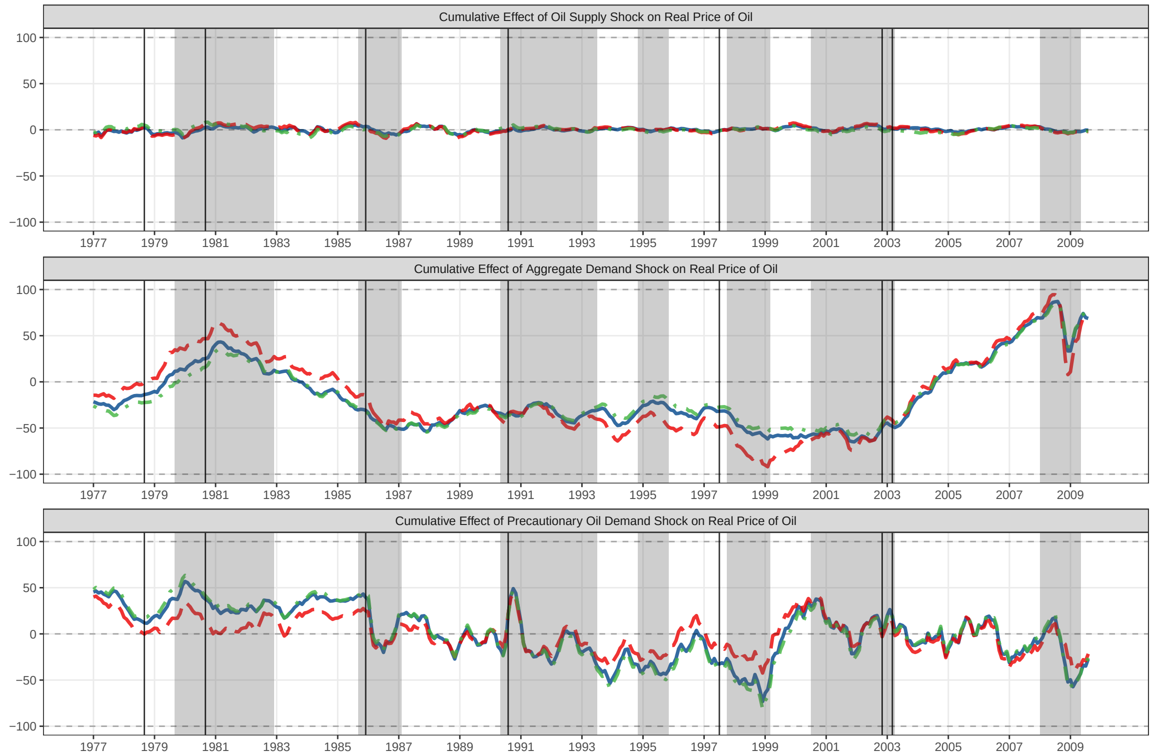


Figure 5: Historical decompositions of the real price of oil for the period 1977M1 – 2009M8. Solid (blue), dashed (red) and dot-dashed (green) lines indicate results from an unrestricted model imposing mutual independence, a model with partial dependence and one elasticity restriction, and the Cholesky scheme, respectively. The initial 22 observations are discarded in order to mitigate truncation biases. Shaded areas indicate episodes of economic recession given by the OECD based Recession Indicators. Vertical lines represent major exogenous events in oil markets: The outbreak of the Iranian Revolution in 1978M9, the Iran–Iraq War in 1980M9, the OPEC collapse in 1985M12, the outbreak of the Persian Gulf War in 1990M8, the Asian Financial Crisis of 1997M7, the Venezuelan crisis in 2002M11 and the Iraq War in 2003M3. The Figure is in full analogy to Figure 2 of Kilian and Murphy (2014).

the four-dimensional model of Kilian and Murphy (2014). As a general impression, it is interesting to observe that all structural models confirm a core insight of Kilian and Murphy (2014) noticing that the surge of oil prices between 2003 and mid-2008 is well attributed to shocks originating in the aggregate demand for oil, and not to 'excessive speculation' (in the three-dimensional model captured by oil-specific demand shocks). While the latter result is typical for all three alternatives, the model with partially dependent shocks offers a specific view of the V-shaped dip in the real price of oil occurring in 2008/09. Both models, KML with mutually independent shocks and the recursive model attribute this dip to both demand shocks with comparable importance. The model with partially dependent shocks, however, attributes the lion's share of the dip to aggregate demand shocks and less to speculative demand shocks. These findings are fully analogous to those of Kilian and Murphy (2014).

It is further worth comparing outcomes related to the oil price surge following the Iranian revolution, which is hardly the result of shortages in oil production and more a reflection of increased flow and precautionary demands for oil. While the model of Kilian (2009) and the specification featuring mutually independent shocks attribute stronger contributions to speculative demand, the model with partially dependent shocks assigns a larger weight to aggregate demand shocks explaining the price surge. While Kilian and Murphy (2014) do not provide historical decompositions for the recursive three-dimensional model, it is compelling to observe that in numerical terms the contribution of the speculative demand implied by the model with partially dependent shocks shows very similar magnitude as results displayed in the third panel of Figure 2 of Kilian and Murphy (2014).

In summary, it is worth highlighting and very interesting to observe how close the features of both types of demand shocks obtained from the partially dependent model are to their counterparts in the four-dimensional sign-restricted model of Kilian and Murphy (2014).

5 Conclusion

To support the detection of structural relations in multivariate dynamic systems (such as SVARs and multivariate GARCH models), this work suggests a kernel-based ML estimation as a particular tool of independent component analysis (ICA). Unlike alternative ICA-based approaches that have been suggested for the identification of SVARs in the literature (see e.g. Moneta et al. 2013, Lanne et al. 2017, Gouriéroux et al. 2017), the KML approach is fully agnostic with regard to the underlying structural form or distributional assumptions. Simulation-based evidence highlights that both the first- and second-order performance features of the KML estimator are throughout (among the) best within a family of alternative independence-based identification approaches. We derive the asymptotic properties of the estimator under relatively weaker conditions. As a particular merit, the KML design can be adapted to take possible dependence among components into account.

In the empirical analysis of structural sources of the global oil market, we illustrate the case of partial dependence with a prime example. As a counterpart of the three-dimensional model of Kilian (2009), an identified representation of three mutually independent structural shocks already yields a soundly identified oil supply and two oil demand shocks. Allowing for unmodelled dependence in shocks originating in speculative demand for oil stocks on the one hand and oil supply on the other hand, however, resembles the historical contributions of demand shocks to real oil prices that are in an outstandingly close relation to implications of an informationally richer model of Kilian and Murphy (2014).

6 Supplementary material

In the online supplements of this work (OA), we provide additional theoretical discussions (OA A), further simulation results (OA B), and supplementary materials complementing empirical analysis (OA C).

7 Acknowledgements

We gratefully acknowledge financial support by the Belgian Federal Science Policy (contract ARC 18/23-089) and the Deutsche Forschungsgemeinschaft (HE 2188/17-1). We thank the editor Ivan Canay, two anonymous referees, Dante Amengual, Gabriele Fiorentini, Markus Fülle, Lutz Kilian, Helmut Lütkepohl and participants at the Financial Econometrics Conference (University of Cambridge, May 2024) for helpful discussions and comments. Excellent computational assistance from Lasse Trienens is gratefully acknowledged. Any errors and omissions should be regarded as those of the authors.

8 Disclosure statement

The authors report there are no competing interests to declare.

A Proofs of theorems

Proof of Theorem 1: First, $\arg\max_\theta \log L(\theta) = \theta_0$ by assumption, and $l_T(\theta) \xrightarrow{p} l(\theta)$ by a law of large numbers. Let $f_{i\theta}$ denote the true density of $\xi_{it}(\theta)$, Assumption (A2) implies that

$$\sup_\theta |\hat{f}_{i\theta h}(\xi_{it}(\theta)) - f_{i\theta}(\xi_{it}(\theta))| = O_p\left\{(\log(T)/Th)^{1/2} + h^2\right\} = o_p(1),$$

see Masry (1996). The true density $f_{i\theta}(\xi_{it}(\theta))$ is bounded away from zero by assumption (A2), while the estimate $\hat{f}_{i\theta h}(\xi_{it}(\theta))$ is bounded away from zero with probability 1 because $K(0) \geq \delta/Th > 0$ by assumption (A4). Thus, for each $\theta \in \Theta$, $\Delta(\theta) := \log \hat{f}_{i\theta h}(\xi_{it}(\theta)) - \log f_{i\theta}(\xi_{it}(\theta)) = o_p(1)$ by the continuous mapping theorem. Then, by the mean value theorem and the Cauchy-Schwarz inequality,

$$|\Delta(\theta') - \Delta(\theta)| \leq B||\theta' - \theta||,$$

almost surely, where $||\cdot||$ is the Euclidean vector norm, $B = \sup_{\theta^* \in \Theta^*} ||\partial\Delta(\theta^*)/\partial\theta||$, Θ^* is the line segment joining θ and θ', and B is bounded in probability because $f_{i\theta}(\xi_{it}(\theta))$ and $\hat{f}_{i\theta h}(\xi_{it}(\theta))$ have bounded derivatives on Θ, and both are bounded away from zero. This implies, by Theorem 21.10 of Davidson (1994), that $\Delta(\theta)$ is stochastically equicontinuous, which in turn implies that $\sup_\theta |\log \hat{f}_{i\theta h}(\xi_{it}(\theta)) - \log f_{i\theta}(\xi_{it}(\theta))| = o_p(1)$ by Theorem 21.9 of Davidson (1994). Therefore,

$$\begin{aligned} \sup_\theta |\hat{l}_T(\theta) - l(\theta)| &= \sup_\theta |\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} \log \hat{f}_{i\theta h}(\xi_{it}(\theta)) - \log f_{i\theta}(\xi_{it}(\theta))| \\ &\leq \sup_\theta \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} |\log \hat{f}_{i\theta h}(\xi_{it}(\theta)) - \log f_{i\theta}(\xi_{it}(\theta))| = o_p(1). \end{aligned}$$

Therefore, $\arg\max_\theta \hat{l}_T(\theta) - \arg\max_\theta l(\theta) = o_p(1)$. Since $l(\theta) \leq \log L(\theta)$, with equality if and only if $\theta = \theta_0$, we have $\arg\max_\theta \hat{l}_T(\theta) - \arg\max_\theta \log L(\theta) = \hat{\theta} - \theta_0 = o_p(1)$, which proves the consistency. □

Proof of Theorem 2: First, let

$$\begin{aligned} \hat{I}(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\hat{s}_{it}^2(\theta)\frac{\partial\xi_{it}(\theta)}{\partial\theta}\frac{\partial\xi_{it}(\theta)}{\partial\theta^\top}\right], \\ \hat{J}(\theta) &= -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\hat{s}_{it}(\theta)\frac{\partial^2\xi_{it}(\theta)}{\partial\theta\partial\theta^\top} + \frac{\partial\hat{s}_{it}}{\partial\xi_{it}}\frac{\partial\xi_{it}(\theta)}{\partial\theta}\frac{\partial\xi_{it}(\theta)}{\partial\theta^\top}\right]. \end{aligned}$$

By a Taylor expansion, the estimator satisfies

$$\hat{l}'_T(\hat{\theta}) = \hat{l}'_T(\theta_0) - \hat{J}(\theta_0)(\hat{\theta} - \theta_0) + o_p(1) = 0$$

where

$$\hat{l}'_T(\theta_0) = \frac{\partial\hat{l}_T(\theta_0)}{\partial\theta} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{s}_{it}(\theta_0)\frac{\partial\xi_{it}(\theta_0)}{\partial\theta}.$$

Hence,

$$\sqrt{T}(\hat{\theta} - \theta_0) = \hat{J}^{-1}(\theta_0)\sqrt{T}\hat{l}'_T(\theta_0) + o_p(1).$$

We want to show that $\sqrt{T}\hat{l}'_T(\theta_0) \xrightarrow{d} N(0, I(\theta_0))$. First, note that

$$\sqrt{T}\hat{l}'_T(\theta_0) = \sqrt{T}l'_T(\theta_0) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{s}_{it}(\theta_0) - s_{it}(\theta_0))\frac{\partial\xi_{it}}{\partial\theta}$$

and by a Lindeberg-Levy CLT

$$\sqrt{T}l'_T(\theta_0) \xrightarrow{d} N(0, I(\theta_0)).$$

It remains to show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{s}_{it}(\theta_0)-s_{it}(\theta_0))\frac{\partial\xi_{it}}{\partial\theta}=o_p(1) \tag{13}$$

such that $\sqrt{T}\hat{l}'_T(\theta_0)$ and $\sqrt{T}l'_T(\theta_0)$ have the same asymptotic distribution. We show (13) using the approach of Andrews (1994b). A sufficient condition for (13) is that $\nu_{iT}(\tau_i)$ is stochastically equicontinuous at the true parameter τ_{i0} for all $i=1,\dots,N$, where

$$\nu_{iT}(\tau_i)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{m_i(\tau_i(\xi_{it}))-\mathbb{E}m_i(\tau_i(\xi_{it}))\right\},\quad \tau_i\in\mathcal{T}, \tag{14}$$

$\hat{\tau}_i$ is consistent for τ_i with respect to the pseudo-metric $\rho_{\mathcal{T}}$, $P(\hat{\tau}_i\in\mathcal{T})\to 1$, and additionally,

$$\int m_i(\hat{\tau}_i(x))f_i(x)dx=o_p(T^{-1/2}). \tag{15}$$

Stochastic equicontinuity of $\nu_{iT}(\tau_i)$ is obtained by the sufficient conditions given by Assumption SE of Andrews (1994b). His proposition establishes that $\nu_{iT}(\tau_i)$ is stochastically equicontinuous at each $\tau_0\in\mathcal{T}$ and $\mathcal{T}$ is totally bounded under the pseudo-metric $\rho_{\mathcal{T}}$ given by (7). Note that ξ_{it} is i.i.d., and hence it also satisfies Andrews (1994b) strong mixing property.

Similar to Andrews (1994a), we can show that the conditions $P(\hat{\tau}_i\in\mathcal{T})\to 1$ and $\rho_{\mathcal{T}}(\hat{\tau}_i,\tau_{i0})\xrightarrow{p}0$ hold based on the following results: (i) $\tau_{i0}\in\mathcal{T}$ by assumption, (ii) $\sup_{x\in\Xi}|\hat{\tau}_i(x)-\tau_{i0}(x)|=o_p(1)$ which holds using Assumptions (A3) and (A7), see Silverman (1978), (iii) $\sup_{x\in\Xi}|\partial\hat{\tau}_i(x)/\partial x-\partial\tau_{i0}(x)/\partial x|=o_p(1)$, which holds by Theorem 1 of Andrews (1995), and (iv) $\sup_{x\in\Xi}|\partial\tau_{i0}(x)/\partial x|<\infty$ by Assumption (A2).

We now show (15). At the true score function, $\sum_{i=1}^{N}\int m_i(\tau_{i0}(x))f_i(x)dx=0$, and $\sum_{i=1}^{N}\mathbb{E}\left[\frac{\partial}{\partial\tau_i}m_i(\tau_{i0}(\xi_{it}))|\xi_{it}=x\right]=0$, $\forall t,\forall x$, which is condition (4.12) of Andrews (1994b). His condition (4.13) is satisfied provided that $T^{1/4}(\hat{\tau}_i(x)-\tau_{i0}(x))=o_p(1)$ for all $i=1,\dots,N$. By classical kernel arguments, our estimator is such that $\hat{\tau}_{1i}(x)-\tau_{1i0}(x)=O(h^2)+O_p((Th^3)^{-1/2})$ and $\hat{\tau}_{2i}(x)-\tau_{2i0}(x)=O(h^2)+O_p((Th)^{-1/2})$. Hence, $T^{1/4}(\hat{\tau}_{1i}(x)-\tau_{1i0}(x))=o_p(1)$ provided that $T^{1/4}h^2\to 0$ and $T^{1/2}h^3\to\infty$, while $T^{1/4}(\hat{\tau}_{2i}(x)-\tau_{2i0}(x))=o_p(1)$ provided that $T^{1/4}h^2\to 0$ and $T^{1/2}h\to\infty$. These conditions hold under Assumption (A7), which proves (15). Hence, we have established (13), which implies that $\sqrt{T}\hat{l}'_T(\theta_0)\xrightarrow{d}N(0,I(\theta_0))$.

Finally, $\hat{I}(\hat{\theta})\xrightarrow{p}I(\theta_0)$ and $\hat{J}(\hat{\theta})\xrightarrow{p}J(\theta_0)$ by a WLLN and Assumption (A2). Thus, $\hat{J}(\hat{\theta})^{-1}\hat{I}(\hat{\theta})\hat{J}(\hat{\theta})^{-1}$ is a consistent estimator of the asymptotic covariance matrix. The theorem follows. □

References

Amengual, D., G. Fiorentini, and E. Sentana (2022): "Moment tests of independent components," *SERIEs: Journal of the Spanish Economic Association*, 13, 429–474.

Andrews, D. W. (1994a): "Chapter 37 Empirical process methods in econometrics," Elsevier, vol. 4 of *Handbook of Econometrics*, 2247–2294.

——— (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–586.

Andrews, D. W. K. (1994b): "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica*, 62, 43–72.

Bach, F. R. and M. I. Jordan (2003): "Kernel independent component analysis," *J. Mach. Learn. Res.*, 3, 1–48.

Baumeister, C. and J. D. Hamilton (2022): "Advances in using vector autoregressions to estimate structural magnitudes," *Econometric Theory*, 1–39.

Bhattacharya, P. K. (1967): "Estimation of a probability density function and its derivatives," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 29, 373–382.

Bickel, P. J. (1982): "On adaptive estimation," *The Annals of Statistics*, 10, 647–671.

Boscolo, R., H. Pan, and V. Roychowdhury (2004): "Independent component analysis based on nonparametric density estimation," *IEEE Transactions on Neural Networks*, 15, 55–65.

Braun, R. (2023): "The importance of supply and demand for oil prices: Evidence from non-Gaussianity," *Quantitative Economics*, 14, 1163–1198.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016): "Inference in VARs with conditional heteroskedasticity of unknown form," *Journal of Econometrics*, 191, 69 – 85.

Cardoso, J.-F. (1998): "Multidimensional independent component analysis," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, vol. 4, 1941–1944 vol.4.

Chen, A. and P. J. Bickel (2006): "Efficient independent component analysis," *The Annals of Statistics*, 34, 2825 – 2855.

Comon, P. (1994): "Independent component analysis, A new concept?" *Signal Processing*, 36, 287 – 314.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press.

Faust, J. (1998): "The robustness of identified VAR conclusions about money," *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

Fiorentini, G. and E. Sentana (2023): "Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions," *Journal of Econometrics*, 235, 643–665.

Gouriéroux, C., A. Monfort, and J. 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 of Econometrics*, 225, 27–46, themed Issue: Vector Autoregressions.

Hafner, C. M. and H. Herwartz (2006): "Volatility impulse responses for multivariate GARCH models: An exchange rate illustration," *Journal of International Money and Finance*, 25, 719–740.

Hafner, C. M., H. Herwartz, and S. Maxand (2022): "Identification of structural multivariate GARCH models," *Journal of Econometrics*, 227, 212–227.

Hastie, T. and R. Tibshirani (2002): "Independent components analysis through product density estimation," in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA: MIT Press, NIPS'02, 665–672.

Herwartz, H. and S. Wang (2023): "Point estimation in sign-restricted SVARs based on independence criteria with an application to rational bubbles," *Journal of Economic Dynamics and Control*, 151, 104630.

Hyvärinen, A., J. Karhunen, and E. Oja (2001): *Independent Component Analysis*, John Wiley & Sons.

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

Inoue, A. and L. Kilian (2016): "Joint confidence sets for structural impulse responses," *Journal of Econometrics*, 192, 421–432, innovations in Multiple Time Series Analysis.

Jones, M. C., J. S. Marron, and S. J. Sheather (1996): "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, 91, 401–407.

Keweloh, S. A. (2021): "A generalized method of moments estimator for structural vector autoregressions based on higher moments," *Journal of Business & Economic Statistics*, 39, 772–782.

Kilian, L. (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99, 1053–1069.

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, Cambridge University Press.

Kilian, L. and D. P. Murphy (2012): "Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models," *Journal of the European Economic Association*, 10, 1166–1188.

——— (2014): "The role of inventories and speculative trading in the global market for crude oil," *Journal of Applied Econometrics*, 29, 454–478.

Kilian, L. and X. Zhou (2023): "The econometrics of oil market VAR models," *Advances in Econometrics*, 45B, 65–95.

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.

Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2015): "Confidence bands for impulse responses: Bonferroni vs. Wald," *Oxford Bulletin of Economics and Statistics*, 77, 800–821.

Masry, E. (1996): "Multivariate local polynomial regression for time series: Uniform strong consistency and rates," *Journal of Time Series Analysis*, 17, 571–599.

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

Mertens, K. and M. O. Ravn (2013): "The dynamic effects of personal and corporate income tax changes in the United States," *American Economic Review*, 103, 1212–1247.

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

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, 481–85.

Ramey, V. A. (2016): "Macroeconomic shocks and their propagation," Working Paper 21978, National Bureau of Economic Research.

Risk, B. B., N. A. James, and D. S. Matteson (2015): *steadyICA: ICA and tests of independence via multivariate distance covariance*, R package version 1.0.

Robinson, P. M. and L. Taylor (2017): "Adaptive estimation in multiple time series with independent component errors," *Journal of Time Series Analysis*, 38, 191–203.

Rosenblatt, M. (1971): "Curve estimates," *The Annals of Mathematical Statistics*, 42, 1815–1842.

Samarov, A. and A. Tsybakov (2004): "Nonparametric independent component analysis," *Bernoulli*, 10, 565 – 582.

Silverman, B. W. (1978): "Weak and strong uniform consistency of the kernel estimate of a density and its derivatives," *The Annals of Statistics*, 6, 177–184.

——— (1998): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC.

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

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007): "Measuring and testing dependence by correlation of distances," *The Annals of Statistics*, 35, 2769–2794.