CAUSAL INFERENCE WITH (PARTIALLY) INDEPENDENT SHOCKS AND STRUCTURAL SIGNALS ON THE GLOBAL CRUDE OIL MARKET

Christian M. Hafner, Helmut Herwartz, Shu Wang

LIDAM Discussion Paper ISBA
2023 / 04



UCLouvain

ISBA | LIDAM
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.html

Causal inference with (partially) independent shocks and structural signals on the global crude oil market

Christian M. Hafner* Helmut Herwartz† Shu Wang†

January 25, 2023

Abstract

Independent component analysis has recently become a promising data-based approach to detect structural relations in multivariate dynamic systems in cases when a-priori knowledge about causal patterns are scant. This paper suggests a kernel-based ML estimation that is largely agnostic with regard to the distributional features of the structural origins of data variation and enables causal analysis under the assumption of having only a subset of independent shocks. In an empirical application to the global oil market model of Kilian (2009) we illustrate the benefits of allowing for an unmodelled higher-order dependence among the oil supply and speculative oil demand shocks.

Keywords: Structural VAR, structural MGARCH, Independent component analysis
JEL Classification: C14, C32, Q43

*Louvain Institute for Data Analysis and Modelling in Economics and Statistics (LIDAM), Université catholique de Louvain, Belgium. Financial support of the Belgian Federal Science Policy, contract ARC 18/23-089, is gratefully acknowledged.

†Chair of Econometrics, University of Göttingen, Humboldtallee 3, D-37073 Göttingen, Germany. Corresponding author: Shu Wang, e-mail: shu.wang@uni-goettingen.de. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (HE 2188/17-1). We thank Lutz Kilian and Helmut Lütkepohl for comments on an earlier version of the manuscript. Excellent computational assistance from Lasse Trienens is gratefully acknowledged.

1 Introduction

A paramount problem in causal analysis of multivariate dynamic systems – such as structural VAR (SVAR) and multivariate GARCH (MGARCH) models – is uncovering the structural representation of the data. Such a representation holds particular interest for both academics and policy-makers, as it reveals the underlying forces acting upon a financial or macroeconomic system and helps to understand the transmission mechanisms of certain policy actions. The identification of the structural form requires external information, which typically takes the form of instrumental variables (Mertens and Ravn 2013, Stock and Watson 2012) and theory-motivated restrictions, either on the magnitude (Sims 1980, Blanchard and Quah 1989) or the sign of the parameter (Canova and De Nicolo 2002, Faust 1998). To support structural analysis in cases where such information is scant or hardly available, a viable literature has emerged on data-based identification that relies on stochastic characteristics of latent structural shocks, in particular heteroskedasticity or independence of the shocks.[1]

Adapting the technique of so-called independent component analysis (ICA), which was originally developed for solving the problem of blind source separation, various independence-based identification approaches have been discussed in the literature. 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. Gouriéroux et al. (2017) develop the semi-parametric pseudo ML (PML) method that promises consistent estimation of structural relations under conditions that regulate, for instance, the 'extent' of misspecification. However, since the actual source distributions are rarely known to the analyst, in practice there always remains a risk of misspecification for approaches such as PML which rely on a set of pre-specified pseudo probability density functions (pdfs).

To guard against misspecification, in this paper we suggest an adaptive ML approach based on non-parametric density estimation as an alternative to PML, where kernel estimates (Rosenblatt 1971, Silverman 1978) are used to formulate the estimation equations. The proposed estimator is proved to be consistent under relatively weak conditions. We also provide extensive simulation-based evidence on its finite-sample performance. On the methodological side, we believe that this study contributes to the existing literature at two levels. First,

[1] For a comprehensive collection of data-driven identification strategies in SVARs, the reader may consult Chapter 14 of Kilian and Lütkepohl (2017) and references therein.

the paper provides the analyst with a robust identification approach, when there is a lack of prior information – either on the underlying economic theory or statistical properties – of the structural innovations, especially in very small samples. Second, we develop a less restrictive framework building upon partial independence, which also allows the analyst to introduce some external information into the model.

Recognizing that uniformly good estimate can only be obtained if the estimation function is properly adjusted to the distribution of the latent shocks, more flexible ICA routines have been developed (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 by means of B-spline 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. Most closely related to the method discussed in this paper are suggestions of Samarov and Tsybakov (2004) or Boscolo et al. (2004) which also build upon kernel density estimation. With this being said, the approach in this paper is developed within the context of time series models and fully adapted for causal inference in SVARs. In this sense, our study is also comparable to Braun (2021), who specifies the shocks by means of a Dirichlet process mixture and addresses the problem from a Bayesian perspective. Contributing to the existing literature, we conduct extensive simulation experiments and study the finite-sample performance of the proposed estimator under heterogeneous source distributions (symmetric vs. asymmetric, uni- vs. multimodal, and leptokurtic vs. platykurtic distributions). We also investigate the robustness of the estimator in the presence of various second-order-moment characteristics (heteroskedasticity and co-heteroskedasticity) and compare the estimator with its most prominent peers. In particular, we discuss the close link between the (P)ML approach and the generalized method of moments (GMM), which imposes a set of higher-order cross-moment conditions (co-skewness and co-kurtosis) and has been recently developed and discussed in Lanne and Luoto (2021), Keweloh (2021), Guay (2021) for inference in SVARs and in Hafner et al. (2022) who suggest a structural interpretation of multivariate GARCH models. In addition, the suggested approach is also compared with some fully non-parametric identification schemes, such as ICA by minimizing distance covariances, which has been successfully applied to small to medium-size macroeconometric models (Matteson and Tsay 2017,

Herwartz 2018, Herwartz et al. forthcoming).

Furthermore and as a novel concept, we show that the suggested approach can be easily extended towards a less restrictive framework for systems with dependent components. Allowing dependence among certain structural shocks appears to be relevant for modeling financial and macroeconomic variables, since mutual independence of all shocks might be considered as overly restrictive in many empirical applications (see Montiel Olea et al. 2022 for a critical assessment of ICA-based identification). We show that a substantial dimension reduction can be achieved based on some weaker forms of independence assumption such as partial independence and group-wise independence. For instance, assuming that there is only one independent component in a N-dimensional system ($N > 2$) identifies the structural quantities related to this component and reduces the number of free parameters by $N - 1$. We also illustrate that full identification can be accomplished by introducing only a small amount of additional restrictions. Therefore, the proposed approach combines the advantage of assigning a maximum information content to independent shocks, and the need to allow for unmodelled dependence whenever required from the perspective of economic theory. As such, it opens up new possibilities for identification in models of simultaneous causality by combining (only a few) external information with (relatively weak) statistical properties.

Monte-Carlo experiments show that the proposed kernel-based ML (KML) estimator outperforms, on average, a large variety of competing approaches under 18 different source distributions with a significant lead. It produces the most accurate structural estimates with a competitive level of estimation uncertainty in both small and large samples, low and high-dimensional systems, and in the presence of both heteroskedastic and co-heteroskedastic components. We find that the PML approach based on fixed pdfs bears a great risk of misspecification and could lead to 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 highly sensitive to outliers or extreme realizations in the data. It is particularly vulnerable if the source distribution has heavy tails. Nevertheless, owing to its weaker assumptions, the GMM approach is quite robust and shows similar performance as the proposed estimator in large samples or very large systems.

As an empirical test of concept, we employ the novel identification approach to unravel the structural effects in a stylized model of the global oil market (Kilian 2009, Kilian and

Murphy 2012). Mutually independent shocks show structural features that are fully in line with their counterparts derived by Kilian (2009). Without imposing a-priori bounds on the price elasticity of oil supply as suggested by Kilian and Murphy (2012), unrestricted KML estimates point to actually small and insignificant on-impact changes in the oil production after marginal changes in the oil price. We develop the oil market model further towards a system with only one independent shock (i.e. shocks to the aggregate flow demand for oil) and two partially dependent shocks (shocks to the precautionary oil demand and the flow supply of oil). The weaker form of independence assumption is then combined with a restriction on the price elasticity of oil supply, which produces soundly identified shocks. The system with partially dependent shocks finds support in terms of marked log-likelihood improvements and offers a sound interpretation of oil-specific demand shocks to capture mainly precautionary motives. In particular, the three-dimensional model with partially dependent shocks yields insights into the role of alternative motives for demanding oil that are in very close association with results documented in the sign-restricted four-dimensional model of Kilian and Murphy (2014).

The remainder of this work is organized as follows. We next introduce tools of KML estimation into the context of SVAR identification. Simulation results in Section 3 provide a comprehensive comparison of the suggested approach with alternative independence-based identification schemes. Evidence on structural effects in the global oil market is provided in Section 4. Section 5 concludes. A proof of consistency and discussion of local concavity of the log-likelihood is provided in the Appendix. In an Online Appendix (OA), we document further simulation results and supplementary materials for the empirical analysis.

2 Methodology

In this Section, we introduce our methodology. After stating the general setting of ICA as a tool for finding the structural data representations and its solution by means of ML estimation, we introduce the KML estimator and discuss its asymptotic properties. For the first time in the SVAR literature, we consider dynamic systems driven by a set of orthogonal shocks, some of which are independent and others are subject to higher-order dependence. We also discuss the connection between (P)ML estimation and 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}, the elements in ξ_t carry distinct structural information and are mutually orthogonal.[2]

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, Hafner and Herwartz 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 and the VAR model – in reduced and structural form, respectively – reads as

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

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

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)

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

[2]In general, the dimension of the observable system (N^*) can exceed the dimension of latent components ($N < N^*$). In this case, one can project the N^*-dimensional observable vector onto the N-dimensional space, where the projection matrix can be obtained from the singular value decomposition of Σ_{x}. The proposed approach can then be applied to the adjusted variables. However, noting that the covariance matrix will have reduced rank and such case might not be particularly relevant for the primary application in this study, we consider only non-singular systems.

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 resolved if the components of ξ_t are independent and not Gaussian.

Assumption 1. *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 1, 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, matrix C can be obtained from a spectral or Cholesky decomposition. Left-multiplying equation (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. Hyvärinen 1999, 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

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

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 (normalized) relative gradient (see e.g. Cardoso 1998a) 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. $Q \in \mathcal{O}(N)$.

Noticing that pdfs of the structural components f_i's are hardly known in practice, the MLE is infeasible in most cases. Lanne et al. (2017) assumed f_i to be the density of a Student-t distribution and Gouriéroux et al. (2017) generalized the approach by resorting to pseudo density functions $\tilde{f}_i$. PML estimation can be achieved in the framework described above, where the true pdfs in the log-likelihood functions are replaced by $\tilde{f}_i$ and S by $\tilde{S} = (\tilde{s}_1, \ldots, \tilde{s}_N)^\top$, where $\tilde{s}_i = \tilde{f}_i'/\tilde{f}_i$. For a detailed description of the PML estimator, we refer the reader to Gouriéroux et al. (2017). If $\tilde{f}_i$ coincides with f_i, the PML estimator will have the appealing properties of the MLE estimator. If this is not the case, the estimate could be inconsistent as we illustrate later in the Monte Carlo exercise. In Appendix B, we discuss the well-known condition for local concavity of the log-likelihood and show that such a condition cannot be uniformly satisfied by any pre-specified pseudo density. Furthermore, the ML approach is equivalent to the minimization of mutual information and PML can be considered as a Kullback-Leibler (KL) matching between the distributions $\tilde{f}(Q^\top \mathrm{z}_t)$ and $f(\xi_t)$ (Comon 1994, Cardoso 1998a). Thus, the risk of misspecification of the pdfs (more generally the non-linearities in $\tilde{s}_i$'s) can also lead to inconsistency for other ICA procedures such as FastICA (Hyvärinen and Oja 1997). This point has been demonstrated in Cardoso (1998a). It has been increasingly recognized that uniformly good estimates of the structural mixing matrix B can only be obtained if the estimation function is properly adjusted to the distribution of the latent shocks. Against this background, more adaptive methods for ICA have been developed (see e.g. Hastie and Tibshirani 2002, Bach and Jordan 2003). In this paper, we formulate the likelihood function based on non-parametric density estimation in a similar vein to Hastie and Tibshirani (2002), Samarov and Tsybakov (2004) and Boscolo et al. (2004). We describe our approach next.

2.3 Kernel-based maximum likelihood estimation

Denote the i-th component of ξ_t by $\xi_{it} = e_i^\top Q^\top \mathrm{z}_t = q_i^\top \mathrm{z}_t$. 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 consistently estimated as (Bhattacharya 1967)

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

where $\mathcal{H}_r(x)$ is the r-th Hermite polynomial. Let the orthonormal matrix Q be parameterized by a vector $\theta \in \Theta$ and $\xi_{it}(\theta) := e_i^\top Q(\theta)^\top \mathrm{z}_t$. 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, Hafner and Herwartz 2022). 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.[3] The structural orthogonal mixing matrix is then obtained by maximizing the non-parametric (pseudo) log-likelihood function, i.e.

$$\hat{Q} \equiv Q(\hat{\theta}) = \arg\min_{\theta \in \Theta} \hat{l}_T(\theta) = \arg\min_{\theta \in \Theta} \sum_{t=1}^{T} \sum_{i=1}^{N} \log \hat{f}_{ih}(\xi_{it}(\theta)). \tag{7}$$

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

The practical implementation of the estimator in (7) 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 robust with respect to the choice of the kernel (see e.g. Silverman 1998). For bandwidth selection we use – for simplicity – the rule-of-thumb

[3] For instance, in case $N = 3$, one has:

$$\begin{aligned} Q(\theta) &= (\mathcal{G}_{1,2}(\theta_3)\mathcal{G}_{1,3}(\theta_2)\mathcal{G}_{2,3}(\theta_1))^\top \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & 0 & -\sin\theta_2 \\ 0 & 1 & 0 \\ \sin\theta_2 & 0 & \cos\theta_2 \end{bmatrix} \begin{bmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{6}$$

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 show the consistency 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

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

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 $\xi_t(\theta)$, and

$$\log L(\theta) = \mathbb{E}\log f_{\theta}(\xi_{t}(\theta)).$$

Note that at the true parameter, pseudo densities and true densities coincide, so that $\log L_T(\theta_0) = l_T(\theta_0)$, and $\log L(\theta_0) = l(\theta_0)$. We make the following assumptions:

Assumption 2.

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

(A2) f_i *is uniformly continuous with bounded 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 1 and (A1)-(A4),*

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

A proof is provided in 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 discretization of the estimator sequence $\hat{\theta}_T$, asymptotic normality can be obtained at the usual parametric rate. Moreover, efficient estimators may be constructed that attain the semi-parametric lower bound (see e.g. Chen and Bickel 2006). For the main purpose of this study, however, 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 (i.e. $\hat{\theta}_T$), we evaluate the sampling distribution of $\hat{\theta}_T$ in finite samples based on a moving-block bootstrap approach (Brüggemann et al. 2016). This will be illustrated in the 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, even though it might not be fully efficient in a lower bound sense from a theoretic perspective.

2.5 System 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 Montiel Olea et al. 2022 for a critical assessment of ICA-based identification). 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 (1998b). 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 decomposition 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 (1998b) 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 (1998b) argues that a solution to the ICA problem exists if the estimation equations underlying the first step have certain specific forms.[4] 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 subspace containing mutually dependent ones in a single step.

[4]For instance, a solution to the ICA problem exists if the estimating equations have the form,

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

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 would use 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.

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, i.e.

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

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

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

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 instance, if there is one independent component in the system as outlined above for the global oil market, a unique (up to a sign flip) solution for the first column in the matrix Q can be obtained. After partially identifying the independent aggregate demand shock, one would have just one (instead of three) free parameter left unidentified. A full identification could then be achieved by resorting to additional information such as the imposition of an elasticity restriction. Moreover, if one assumes that $\xi_{C_\alpha^-,t}$ contains group-wise independent components – which is also weak compared with the assumption of mutual independence – the number of structural parameters of interest can be further reduced. For instance, in a five-dimensional model, say for an economic union consisting of two countries with an area-wide shock and two country-specific shocks for each member state, the imposition that the area-wide shock is independent identifies this shock (and its IRFs, FEVD, HD, etc.) and reduces the number of free parameters from 10 to 6. Imposing additionally that shocks in one country are independent of shocks in the other (while shocks stemming from the same country can be mutually dependent) reduces the number of free parameters further 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.

In this sense, the proposed approach opens up new possibilities for identification in models of simultaneous causality. A substantial dimension reduction can be achieved based on some

weaker forms of independence assumption and full identification could then be accomplished by introducing only a small amount of additional restrictions. We illustrate this in the empirical application. Next, we shortly discuss the connection of the ML-based approaches with the method of moments.

2.6 Method of moments

It is suggestive to write down the probability limit of the non-deterministic part of the relative gradient. Under our assumptions and by a law of large numbers, we have

$$T^{-1}\sum_{t=1}^{T}\tilde{S}(Q^\top \mathrm{z}_t)(Q^\top \mathrm{z}_t)^\top \xrightarrow{p} \mathbb{E}\left[\tilde{S}(Q^\top \mathrm{z}_t)(Q^\top \mathrm{z}_t)^\top\right]. \tag{8}$$

The $[i,j]$-th element of the right-hand side matrix 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 of the relative gradient are subject to the first-order condition

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

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.[5] 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 1, since independence implies zero co-moments but the inverse is not true. Nevertheless, GMM does require the existence of the moments imposed. Another significant merit of GMM is that it allows various forms of conditional heteroscedasticity – in

[5] Furthermore, the expression in the local concavity condition in Appendix B, i.e. $\mathbb{E}\left[\tilde{s}_i(x)x - \tilde{s}_i'(x)\right]$, becomes $\mathbb{E}[x^3]$ and $\mathbb{E}[x^4] - 3$, which are precisely the skewness and kurtosis of a random variable x with zero mean and unit variance. The Theorem stated in the Appendix implies that an estimate based on these $\tilde{f}$ is consistent if the true distribution is indeed right-tailed or leptokurtic, respectively.

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

This section compares the performance of the KML approach with some prominent alternatives for the estimation of independent components by means of simulation experiments. We first describe the stylized simulation design and performance criteria. After discussing respective simulation results in detail, we sketch further Monte-Carlo evidence retrieved from higher-dimensional systems (in small samples), and from experiments where structural shocks are subject to higher-order dependence in the form of non-modelled heteroskedasticity.

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, while noting that the KML performance is fairly robust against both kernel and bandwidth choice. As alternative approaches to ICA, we first consider PML estimation using a hyperbolic secant and a Gaussian mixture as pseudo densities with the respective non-linearities

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

for $i = 1, \ldots, N$.[6] Next, we consider GMM estimation based on a set of third- and fourth-order moment conditions taking, respectively, the form

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

Moreover, we employ an approach that has been put forward by Matteson and Tsay (2017), which uses the distance covariance statistics proposed by Székely et al. (2007) as a measure of dependence (dCovICA). The function for dCovICA is available in the `R` package `steadyICA`

[6] While the Student-t distribution appears to be more typical as a super-Gaussian pseudo pdf in (applied) econometrics (see e.g. Lanne et al. 2017), we follow the signal processing literature and consider the hyperbolic secant density as a convenient choice that does not require the specification of additional hyper-parameters (e.g. degrees of freedom). Employing a Student-t distribution in the PML procedure leads to a very similar results as those based on hyperbolic secant.

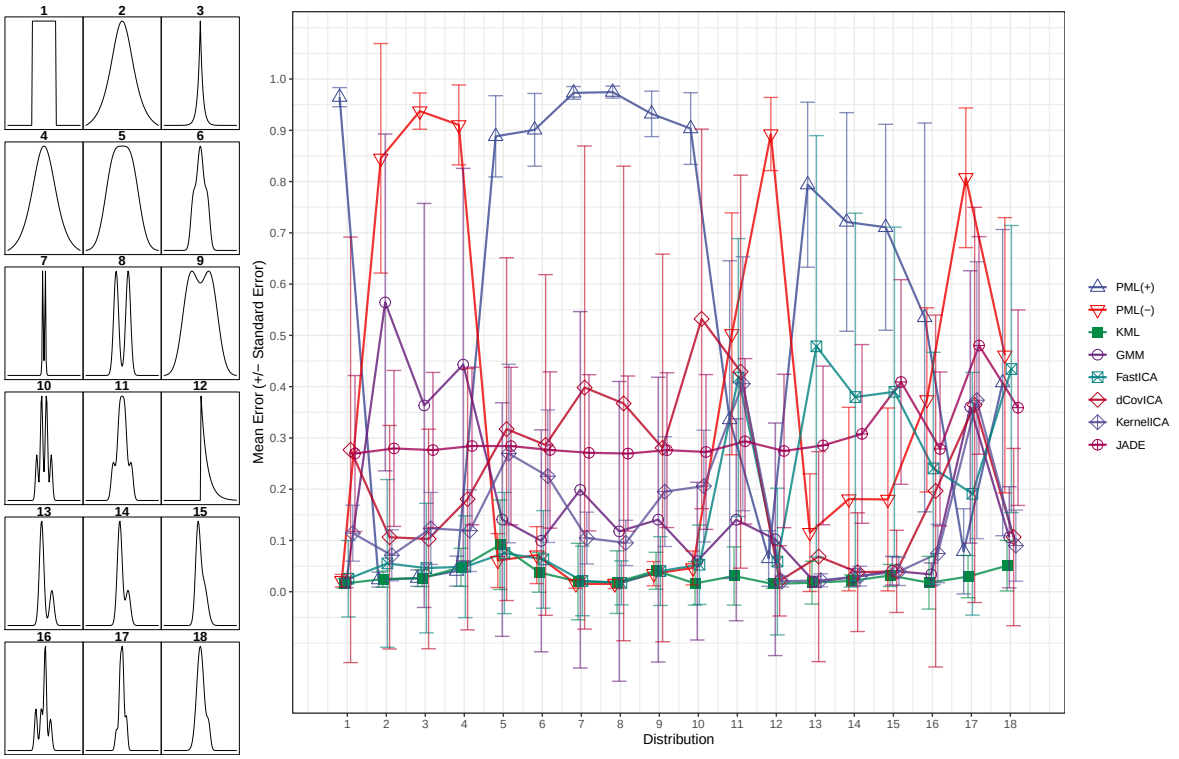


Figure 1: The left-hand-side panel shows 18 data-generating-distributions, including uniform, Student-t, exponential, symmetric and asymmetric mixtures of Gaussian, exponential distributions. Distributions 1–11 are symmetric and distributions 12–18 are asymmetric. The right-hand-side panel shows the average estimation error as given by (12) $\pm$ one estimated standard error for alternative methods and distributions based on 1000 simulation experiments with a sample size of $T = 1000$.

(Risk et al. 2015). The dCovICA has been successfully adopted, e.g. by Herwartz et al. (2022) for the purpose of identification of macroeconomic SVARs in the framework of a Hodges-Lehmann estimation. We further compare the performance of KML with some conventional and widespread ICA algorithms including the Joint Approximate Diagonalization of Eigenmatrices (JADE) (Cardoso 1999), the FastICA (Hyvärinen and Oja 1997) and the KernelICA algorithm (Bach and Jordan 2003). Both JADE and FastICA can be implemented by means of functions that are available from the `R` package `ica` (Helwig 2022). The implementation of KernelICA relies on the `R` package `KernelICA` (Koesner and Nordhausen 2021). Notice that both the FastICA and KernelICA are closely related to a PML approach under certain conditions.[7]

[7]For instance, if the non-linearity is correctly chosen, the FastICA approach can be interpreted as ML estimation (Hyvärinen 1999). Because ML estimation is equivalent to a maximization of mutual information between components in ξ_t (Cardoso 1998a), there exists also a relationship between the canonical correlations

We start with a bivariate model in the baseline experiment 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 $T = 200, 1000$ and 5000. 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, which are depicted in the left-hand-side panel of Figure 1. These distributions closely approximate those used by Matteson and Tsay (2017) and include uniform (1), Student-t with 3 and 5 degrees of freedom (2 and 4), exponential (12), Gaussian mixtures (symmetric: 5 – 6 and 8 – 11; asymmetric: 13 – 18) and mixtures of exponential distributions (3 and 7). The alternative distributions assumed for the structural components cover a wide range of stochastic origins that might be relevant in empirical practice: Symmetric vs. asymmetric shocks, uni- vs. multimodal and leptokurtic vs. platykurtic distributions. Underpinning their practical relevance, it is also conceivable that in practice the independent structural components exhibit distinct distributions, which might easily remain undetected by the analyst (noticing that the components are latent). Thus, we consider cases in which both components are generated from the same distribution as well as from randomly selected (without replacement) heterogeneous distributions.

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.

Simulated observations are prewhitened before being subjected to ICA. To evaluate the estimates obtained from these alternative procedures, we document the estimation error as measured by the Amari distance (Bach and Jordan 2003), i.e.

$$d_{Amari}(B, \hat{B}) = \sum_{i=1}^{2} \left(\frac{\sum_{j=1}^{2} |r_{ij}|}{\max_j |r_{ij}|} - 1 \right) + \sum_{j=1}^{2} \left(\frac{\sum_{i=1}^{2} |r_{ij}|}{\max_i |r_{ij}|} - 1 \right), \tag{12}$$

with $r_{ij} = (B^{-1}\hat{B})_{ij}$, as well as the standard error of $d_{Amari}(B, \hat{B})$. Notice that the Amari distance is invariant to sign slips and column permutations.

used in KernelICA and the mutual information.

3.2 Simulation results

Figure 1 shows simulation results for alternative distributions with a sample size $T = 1000$, which is – to some extent – informative for the asymptotic properties of the alternative estimators under specific distributions of the structural components.[8] Overall, the suggested KML estimator is unique in the sense that it is among the best-performing estimators irrespective of the underlying distributional model. Taking the uncertainty of simulation outcomes into account, there is no distributional model for which we can diagnose a significant lead of any other estimator over the KML approach. The performance of PML depends crucially 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 deterioration once $\tilde{f}$ fails to capture the higher-order characteristics of f, and thereby the condition (13) is violated. Notably, PML(+) (PML(−)) suffers from severe 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 particularly weak. Similar to KML the performance of the JADE estimator is rather stable over all considered distributional scenarios, however, at a higher level of average absolute estimation errors. It is also interesting to see that the performance of the FastICA estimator appears weaker in presence of asymmetric structural components in comparison with sampling from symmetric distributions, while KernelICA and dCovICA tend to produce less accurate estimates under multimodal distributions.

Simulation results documented in Table 1 are informative for the (realistic) case where an analyst has no or little information about the distributional origin of the structural components of the data. In fact, structural shocks might exhibit any of the considered 18 alternative distributions, while an analyst pursues the structural analysis by means of a specific ICA approach owing to an ad-hoc selection or experience and familiarity with the selected approach. The mean estimation errors ($\bar{d}_{Amari}$) documented in Table 1 can be considered to summarize the average performance of a given estimation procedure over all alternative distributional models employed for drawing the structural components. Unsurprisingly and in the light of results from Figure 1, the suggested KML estimator outperforms all alternative

[8] Similar figures for cases $T = 200$ and 5000 can be found in OA-A.

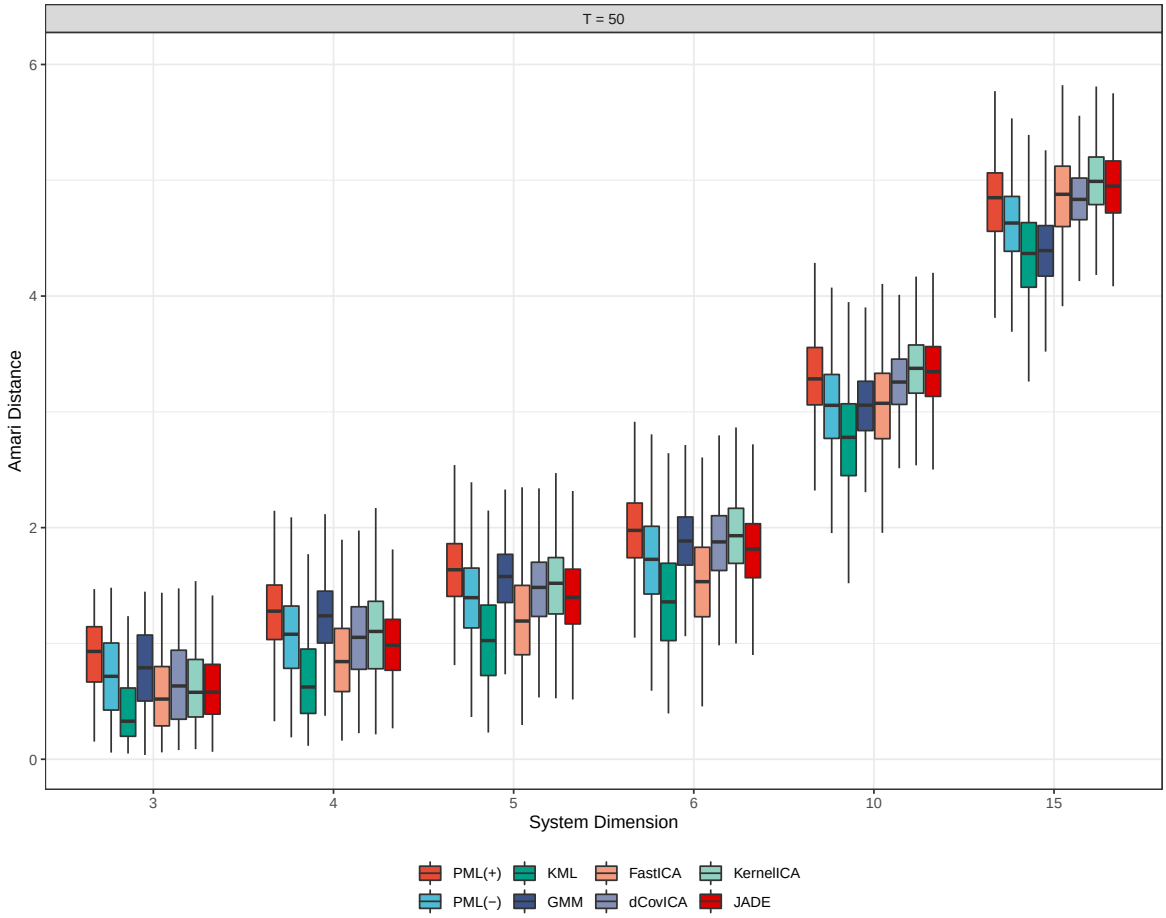


Figure 2: Distribution of the Amari distance via boxplots (with horizontal lines marking the medians) for alternative methods, system dimensions and shocks with distinct distributions based on a sample size of $T = 50$.

estimators on average under both homogeneous and heterogeneous source distributions with a significant lead compared with other methods. Results in Table 1 are also indicative for estimator performance in very large samples ($T = 5000$) and conditional on sample sizes that might be considered more realistic for empirical practice in economic studies ($T = 200$). With regard to the former, the consistency of the KML approach is translated into the convergence of both the estimation error and the associated standard deviation towards zero. Some alternative estimators clearly suffer from inconsistency, as the results suggest that the condition (13) for PML(+) and PML(−) is likely to be violated (on average) when structural components are randomly drawn from alternative scenarios of distributions. Accordingly, these procedures show largest average estimation errors and respective standard deviations. Regarding small-sample-performances ($T = 200$), the KML approach shows its merit that the remaining ICA procedures require considerably larger sample sizes to practically benefit

from consistency of the estimators. With limited sample information, the mean errors for the second (KernelICA) and third best performing estimators (FastICA, GMM) are 2.6 and 2.7 times larger than the loss measure documented for the KML approach, respectively. The overall lead of KML in finite samples is also reflected in the documented standard deviations of estimation errors, which is the smallest among all approaches considered (with JADE and KernelICA performing second and third best in this regard).

3.3 Small samples and higher dimensions

Allowing the estimation equation to adjust according to the source distribution appears crucial for an accurate estimation of the causal scheme, if there is only limited data information. We further compare the relative performance of alternative identification approaches in a small-sample setting ($T = 50$) and systems with a higher dimension ($N = 3$, 4, 5, 6, 10 and 15). In specific, the structural mixing matrix B is randomly generated in a similar fashion as in the baseline experiment and the distributions of the N shocks are uniformly drawn (with replacement) from the 18 distributions displayed in Figure 1.[9] Results displayed in Figure 2 show that the estimation error measured by the Amari distance increases for all methods as the number of independent components increases. In terms of the median estimation error, the suggested KML estimator outperforms its rivals in both small and large systems. The GMM approach has also a quite competitive simulation performance in very large systems ($N = 10, 15$).

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 previous experiment, 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 $\exp w_t$.[10] Results depicted in Figure 3 underline that the performance of KML, GMM and KernelICA is markedly robust under (co-)heteroskedasticity,

[9] To simulate B in a 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 $[-\pi; \pi]$.

[10] Note that $\mathbb{E}[w_t] = \psi_{w_t}(1) = \exp(-g^2/2 + g^2/2) = 1$ such that the unconditional second-order moment is preserved, where $\psi_{w_t}(x)$ is the moment generating function of w_t.

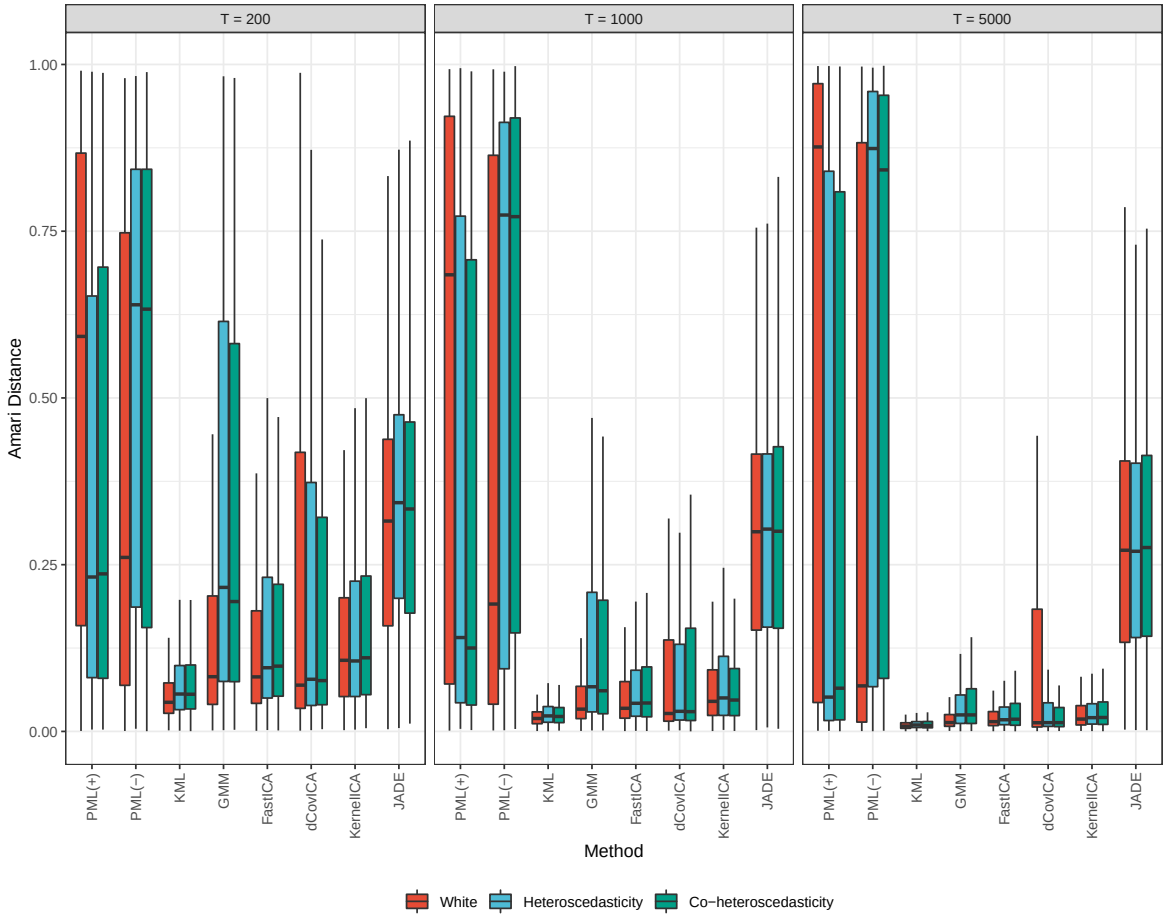


Figure 3: Distribution of the Amari distance via boxplots for alternative methods, sample sizes and shocks with distinct distributions and second-order-moment-features (white noise, heteroscedasticity and co-heteroscedasticity).

while the KML approach has a substantial performance lead in both small and large samples. Similar results are documented in the OA-A, where a stochastic volatility process is simulated from a stationary AR(1) model.

3.5 Summary

In summary, we conclude that KML has outstandingly accurate and robust performance within a family of alternative ICA-based methods for identifying structural relationships. Our comparative analysis covers cases of small sample sizes, large system dimensions, and non-modelled higher-order dependence structures. In particular, our simulation-based evidence highlights that 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 distribution 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 approach might be considered as the first choice for causal estimation and inference.

		PML(+)	PML(−)	KML	GMM	FastICA	dCovICA	KernelICA	JADE
Homogeneous distributions	$T = 200$								
	$\bar{d}_{Amari}$	.543	.372	.099	.254	.226	.444	.240	.329
	$s.e.(d_{Amari})$	.375	.353	.171	.308	.268	.337	.222	.186
	$CI^{95\%}(d_{Amari})$	[.524, .563]	[.354, .390]	[.090, .108]	[.238, .270]	[.212, .239]	[.427, .462]	[.229, .252]	[.319, .338]
	$T = 1000$								
	$\bar{d}_{Amari}$	.571	.361	.031	.169	.158	.221	.144	.305
	$s.e.(d_{Amari})$	.404	.379	.043	.286	.254	.342	.154	.179
	$CI^{95\%}(d_{Amari})$	[.550, .592]	[.341, .381]	[.028, .033]	[.154, .184]	[.145, .172]	[.203, .238]	[.136, .152]	[.295, .314]
	$T = 5000$								
	$\bar{d}_{Amari}$	.592	.339	.011	.121	.138	.212	.081	.286
	$s.e.(d_{Amari})$	.440	.395	.010	.272	.271	.372	.085	.167
	$CI^{95\%}(d_{Amari})$	[.569, .615]	[.318, .359]	[.011, .012]	[.107, .135]	[.124, .152]	[.192, .231]	[.076, .085]	[.278, .295]
Heterogeneous distributions	$T = 200$								
	$\bar{d}_{Amari}$	.532	.376	.070	.188	.150	.216	.158	.315
	$s.e.(d_{Amari})$	.348	.335	.100	.243	.193	.289	.165	.167
	$CI^{95\%}(d_{Amari})$	[.514, .550]	[.358, .393]	[.064, .075]	[.175, .200]	[.140, .160]	[.201, .231]	[.149, .166]	[.307, .324]
	$T = 1000$								
	$\bar{d}_{Amari}$	.539	.396	.024	.096	.077	.181	.076	.276
	$s.e.(d_{Amari})$	.397	.389	.039	.193	.132	.300	.093	.162
	$CI^{95\%}(d_{Amari})$	[.519, .560]	[.375, .416]	[.022, .026]	[.086, .106]	[.070, .084]	[.166, .197]	[.071, .081]	[.267, .284]
	$T = 5000$								
	$\bar{d}_{Amari}$	.586	.365	.010	.047	.044	.169	.031	.266
	$s.e.(d_{Amari})$	.427	.414	.008	.144	.115	.307	.038	.150
	$CI^{95\%}(d_{Amari})$	[.563, .608]	[.343, .386]	[.010, .010]	[.039, .054]	[.038, .050]	[.153, .185]	[.029, .033]	[.258, .274]

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 $\bar{d}_{Amari} \pm 1.96 s.e.(d_{Amari})/\sqrt{1000}$).

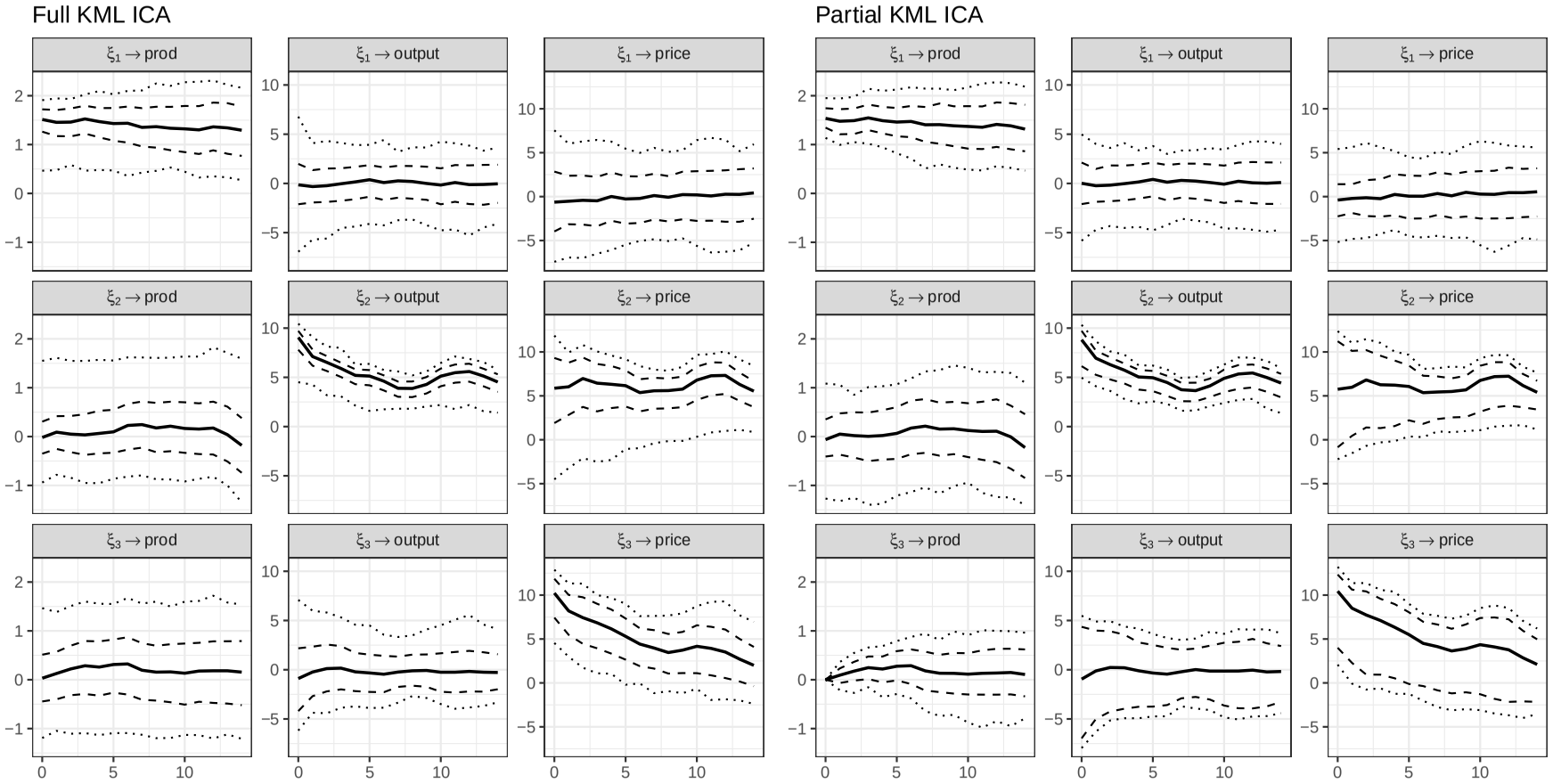


Figure 4: Impulse responses to one-standard-deviation shocks. The left- (right-) hand side panel shows results obtained from imposing mutual independence (partial dependence and an elasticity restriction). The structural shocks are oil supply shock (ξ_1), aggregate demand shock (ξ_2) and speculative 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 1000 replications, respectively.

4 Structural signals underlying the global crude oil market

Oil price changes are an important determinant of 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. As an empirical application of the KML estimator, we reconsider the SVAR model for the global oil market of Kilian (2009). The empirical model consists of 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) by means of 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.

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 assumption of mutual independence by means of the suggested KML approach. To shed light on the performance of alternative ICA-based estimators, we subsequently consider alternative estimators of the price elasticity of oil supply, for which Kilian and Murphy (2012) have powerfully proposed a specific upper bound. Furthermore, we exploit the potential of the KML approach to evaluate joint densities under the assumption of partial dependence. Both KML-based models and the benchmark model of Kilian (2009) are then compared in terms of implied historical decompositions of actual real oil prices. In these exercises, we also compare model-specific outcomes with those communicated in Kilian and Murphy (2014).

4.1 Impulse responses and shock labelling

The left-hand-side panel of Figure 4 displays IRFs as implied by the structural parameter matrix estimated by means of KML under the assumption of mutual independence. To account for estimation uncertainty, we employ the residual-based moving block bootstrap developed by Brüggemann et al. (2016) and follow the rule $\ell = 5.03T^{1/4}$ suggested by these authors to choose a block length of 23. 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.[11] As the structural model is purely data-based, it is unclear if the detected independent components could be reasonably considered as structural shocks in an economic sense. In directional terms, KML structural IRFs are in full analogy to results in Figure 3 of Kilian (2009). In particular, unlike the effects of both demand shocks, the effects of oil supply shocks on the real price of oil lack significance at all horizons. In terms of estimator precision, benchmark results of Kilian (2009) are more precise in comparison with 90% joint confidence intervals displayed in the left-hand side of Figure 4.[12] As an interesting confirmation of benchmark results, the price response to oil-specific demand shocks shows indications of overshooting. Noticing a slight difference to benchmark results, the price responses to shocks in the flow demand for oil are largely stable over the displayed horizons. 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).

[11] Using a slightly different block length does not lead to a notable change of the structural outcomes. The procedure is described in further detail in OA-B. We also provide in the OA-C so-called 'shotgun' plots suggested by Inoue and Kilian (2016) which is designed to graphically display the trajectories of IRFs in a more explicit way.

[12] The determination of joint confidence bands comes at the cost of additional estimation uncertainty. The precision of the structural IRFs in Kilian (2009) benefits from the imposition of three identifying exclusion restrictions. Initial impacts of both demand shocks are estimated quite distinctly under ICA-based identification in comparison with imposing a-priori exclusion restrictions. Both impact effect estimates are sizably larger, if all model parameters are left agnostically unrestricted.

4.2 Estimating the short-run price elasticity of oil supply

Kilian and Murphy (2012) powerfully argue for imposing an upper bound of 0.0258 to 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).[13] 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 $\varepsilon_f = b_{23}/b_{23}$ (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 $\tilde{\varepsilon}_s := -\tilde{b}_{13}/\tilde{b}_{11}$.

Given an a-priori consensus that the price elasticity of oil supply is likely small, it is interesting to assess in how far the alternative implementations of point identification by means of ICA methods (encountered in Section 3) produce elasticity estimates of realistic magnitudes. From estimation results documented in Table 2 (i.e. average outcomes from 1000 moving-block bootstrap replications, their standard deviations and 90% confidence bands) two points are worth remarking. First, the alternative definitions of $\varepsilon_s = b_{13}/b_{33}$ (Kilian and Murphy 2012) and $\tilde{\varepsilon}_s = -\tilde{b}_{13}/\tilde{b}_{11}$ (Baumeister and Hamilton 2022) allow for conclusions that are rather similar in qualitative and quantitative terms. Second, alternative estimates of oil supply elasticities are useful to assess the empirical performance of rival ICA-based implementations of SVAR identification. For instance, one might conclude that GMM estimation fails to provide precise assessments of both elasticities with average estimates being the largest in

[13] Unlike imposing exclusion restrictions for point identification as in Kilian (2009), Kilian and Murphy (2012) adopt a less restrictive approach to identification in the form of sign restrictions. As a result, the set of admissible models turned out to be quite large and in specific, the model-implied response of real oil prices to oil supply shocks lacked sufficient precision. Moreover, as argued by the authors an elicited estimate of the impact price elasticity of oil supply given a surprise change of one standard deviation in the oil-specific demand (of ≈ 1.89) appeared excessive in light of the evidence provided in the related literature and according to conventional wisdom shared by oil industry experts. Kilian and Murphy (2012) suggest imposing an elasticity bound of 0.0258 for set identification, and show that this restriction is effective to sharpen inferential analysis within the set of admissible models.

absolute value. Specifically, the bootstrap mean of ε_s exceeds the upper bound of 0.0258 suggested in Kilian and Murphy (2012) by a factor of 4.5, and the bootstrap mean of ε_f points to a *sizeable negative* oil supply elasticity. Since the underlying structural shocks all exhibit heavy tails, the GMM estimation is likely to underperform in comparison with other competing methods as shown in simulation experiments.[14] While elasticities ε_s obtained from KML, PML and FastICA show similar precision (with standard deviations between 0.035 and 0.041), KML is outstanding in the relatively precise assessment of the elasticity ε_s showing a bootstrap standard deviation of 0.085 which is by a factor of 0.7 smaller than the second best uncertainty estimate documented for dCovICA. KML and PML are unique in providing throughout bootstrap mean elasticities of less than 1% in absolute value. Moreover, the documented 90% bootstrap intervals allow for the conclusion that both elasticities do not significantly differ from zero. To exploit an a-priori knowledge that actual elasticities ε_s and ε_f are negligibly small, we also document the frequencies of bootstrap outcomes with positive elasticity of less than 3%. In this regard and for the case of ε_f, KML most often produces small positive implied elasticities and outperforms all alternative estimators with conventional significance. While GMM-based interval estimates are almost uninformative, interval precision achieved by KML is second best for both ε_s (PML most precise) and ε_f (dCovICA most precise). In summary, the empirical application of the KML estimator yields results that align with theoretical considerations of Kilian and Murphy (2012).

4.3 A model with partially dependent shocks

Applying the KML principle to a system of three mutually independent shocks (i.e. maximizing the sum of three marginal log-densities) produces a log likelihood statistic of -1569.193. As an alternative device, we adopt the KML approach to maximizing the joint density as the product of the densities of a univariate independent shock and of a bivariate vector of orthogonal shocks that are allowed to be dependent.

In general, for a N-dimensional system with one independent shock, we obtain an estimate for $\hat{Q}(\theta)$ by maximizing

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

[14] The (bias-corrected) estimates for kurtoses of supply, aggregate demand and oil-specific demand shocks as obtained from unrestricted KML (restricted partial KML) model are 6.06 (6.35), 8.49 (8.94) and 4.62 (3.88), respectively. Since all shocks exhibit excessive kurtosis, we use a super-Gaussian pseudo density for PML estimation. To allow potential (co-)heteroscedasticity, GMM results are obtained on the basis of odd cross-moments (i.e. $\mathbb{E}[\xi_{it}^2\xi_{jt}]=0$ and $\mathbb{E}[\xi_{it}^3\xi_{jt}]=0$).

		KML	PML	GMM	FastICA	dCovICA	KernelICA
$\varepsilon_s := b_{13}/b_{33}$	$\bar{\varepsilon}_s^*$	0.003	0.000	0.132	0.003	0.014	0.030
	$\hat{\sigma}_s^*(\varepsilon_s)$	0.036	0.035	4.935	0.041	0.087	0.089
	$CI_s^*(\varepsilon_s)$	[-0.051, 0.054]	[-0.046, 0.048]	[-0.971, 0.531]	[-0.054, 0.058]	[-0.066, 0.117]	[-0.052, 0.207]
	$\overline{\varepsilon_s^* \in [0, 0.03]}$	0.426	0.427	0.122	0.424	0.316	0.260
$\varepsilon_f := b_{12}/b_{32}$	$\bar{\varepsilon}_f^*$	-0.003	-0.006	-0.160	-0.018	-0.006	0.025
	$\hat{\sigma}_f^*(\varepsilon_f)$	0.085	0.526	4.886	0.234	0.121	0.752
	$CI_f^*(\varepsilon_f)$	[-0.082, 0.058]	[-0.136, 0.130]	[-1.592, 0.390]	[-0.087, 0.065]	[-0.063, 0.038]	[-0.081, 0.122]
	$\overline{\varepsilon_f^* \in [0, 0.03]}$	0.322	0.276	0.014	0.268	0.273	0.235
$\tilde{\varepsilon}_s := -\tilde{b}_{13}/\tilde{b}_{11}$	$\bar{\tilde{\varepsilon}}_s^*$	0.003	-0.001	-0.165	0.003	0.003	0.023
	$\hat{\sigma}_s^*(\tilde{\varepsilon}_s)$	0.033	0.028	3.003	0.031	0.036	0.074
	$CI_s^*(\tilde{\varepsilon}_s)$	[-0.040, 0.049]	[-0.038, 0.035]	[-1.091, 0.155]	[-0.047, 0.050]	[-0.039, 0.048]	[-0.043, 0.188]
	$\overline{\tilde{\varepsilon}_s^* \in [0, 0.03]}$	0.453	0.445	0.073	0.437	0.371	0.312

Table 2: Average bootstrap estimates and standard deviations for three variants of the impact price elasticities of the oil supply obtained from alternative ICA procedures based on moving block bootstraps with 1000 replications. Numbers in brackets show a bootstrap confidence intervals with 90% coverage (determined from empirical quantiles). Rows labelled, for instance, $\overline{\varepsilon_s^* \in [0, 0.03]}$ document the share of respective bootstrap draws with positive elasticity estimates which are at most 3%.

where $C_i^- = \{1, \ldots, N\} \setminus \{i\}$. For expository purposes, we display the independent (here: the aggregate demand shock) as the first shock in the system. Notice that matrix $\hat{Q}$ is identified up to a right-multiplication by

$$\begin{bmatrix} \pm 1 & 0 \\ 0 & Q^* \end{bmatrix},$$

where Q^* is an orthonormal matrix of dimension $(N-1) \times (N-1)$. In the current application, the subsystem has a dimension of two and can be parameterized with a scalar rotation angle θ^*. We obtain a log-likelihood of -1549.105 and, hence, can conclude that the data are well in line with the suggested refinement of the distribution of structural shocks.[15] By assumption, we consider the aggregate demand shock to be independent of the oil supply and precautionary demand shock, which are allowed to depend on each other in a higher-order sense. While considering these shocks as dependent sacrifices the possibility to identify them uniquely, it is

[15] A formal comparison of the models of mutual independence with partial dependence through likelihood ratio criteria is not possible without further assumptions that enable a nesting of the former model within the latter (e.g. by means of orthogonalized copula distributions). However, we notice that as twice the likelihood difference amounts to a statistic of more than 40 one can consider this improvement as significant without knowledge of the actual degrees of freedom.

not without merit to allow that (in terms of orthogonal components) precautionary demand responds to news on unexpected changes in oil supply. Agents might immediately worry about potential disruptions of supply chains after announcements of large suppliers (e.g. OPEC) to reduce their oil production. As a response, to enhanced uncertainty about the future oil supply agents might react with additional precautionary demand (Alquist and Kilian 2010). Concerning the actual relevance of such dependence patterns, it is worth noticing that the structural IRFs in the left-hand side of Figure 4 have already revealed some indication of overshooting price effects of precautionary demand shocks.

Without loss of generality, we fix the sign of the independent shock and order it first (temporarily) for expository purposes ξ_{1t}. Later when discussing structural outcomes such as IRFs, we use the same ordering scheme (i.e. supply, aggregate and precautionary demand shock) as in the previous section. As we have left the two dependent shocks unidentified, identification of both shocks requires further information. To identify the precautionary demand shock, we impose the 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

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

where e_1 and e_3 are the first and third column of I_3. The right-hand-side panel of Figure 4 shows the structural IFRs that are implied by the model specification with partially dependent shocks. Two observations are worth making. First, with one exception, the structural IRFs are very similar both in terms of shape and magnitude for the models evaluated under mutual independence or partial dependence of structural shocks. Second, owing to the imposed impact restriction, the dynamic effect of flow demand shocks on oil production is of similar precision as in Kilian (2009).

4.4 Historical decompositions

As a natural complement of average structural effects, historical decompositions shown in Figure 5 allow a more specific view at the time-varying sources of actual real oil prices 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 historical decompositions that allow for a comparison of alternative structural models, i.e. (i) the recursive model of Kilian (2009), (ii) the model implied by the evaluation of mutual independence (KML), and (iii) the model of partial dependence.

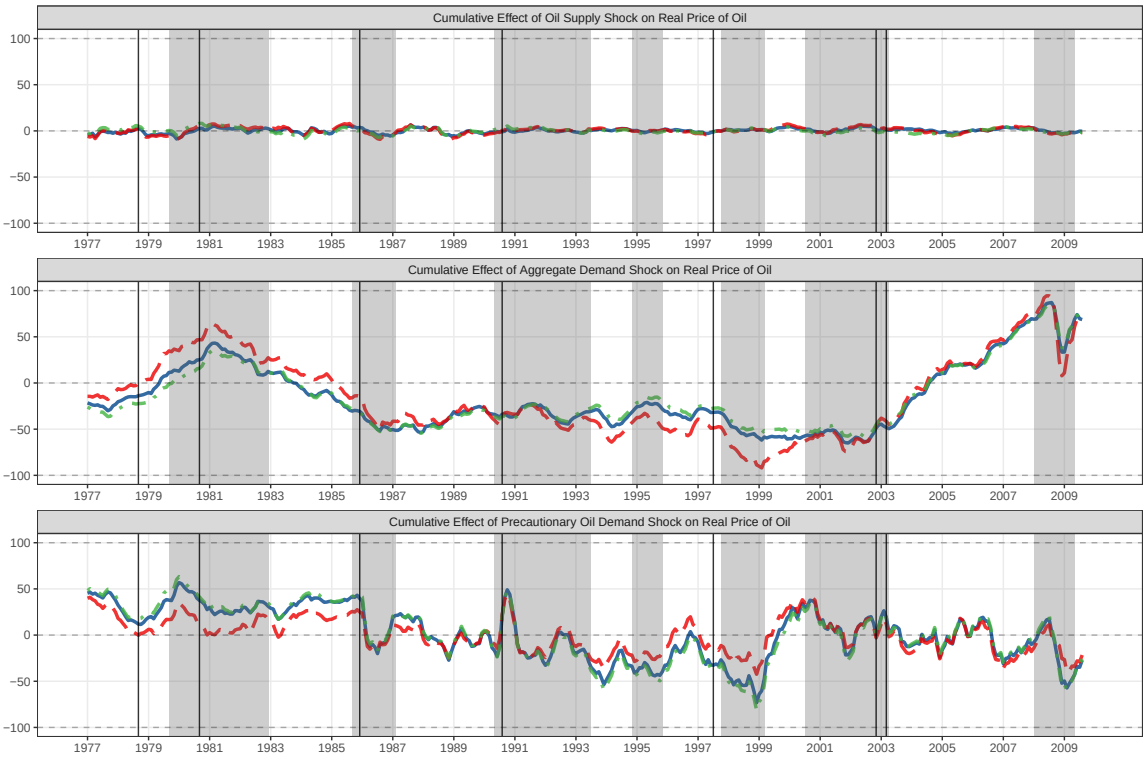


Figure 5: Historical decompositions of the real price of oil for the period 1977M1 – 2009M8. Solid (blue), dashed (red) and dot-dashed (green) lines indicate results obtained from an unrestricted model imposing mutual independence, a model with partial dependence and one elasticity restriction, and the Cholesky scheme (i.e. two elasticity restrictions), 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 collapse of OPEC 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 composed in full analogy to Figure 2 of Kilian and Murphy (2014).

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 the four-dimensional model of Kilian and Murphy (2014).[16] As a general impression, it is

[16]Kilian and Murphy (2014) decompose the oil-specific demand shock further and distinguish oil storage

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 model 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 much less to speculative demand shocks. These findings are in full analogy 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).[17]

demand and other-oil-demand, with the latter being 'a complement to the first three structural shocks' which comprises, e.g. shocks to inventory techniques, changes in preferences. Accordingly, Kilian and Murphy (2014) refrain from assigning an explicit structural interpretation to this shock. Pointing to the overall negligible effect of oil supply shocks for the real price of oil, the results in the upper panel of Figure 2 of Kilian and Murphy (2014) are slightly more volatile than those in Figure 5, without signifying any systematic contribution of flow supply shocks to oil prices.

[17]Kilian and Murphy (2014) discuss several further events (e.g. Iran-Iraq war, collapse of OPEC in 1985, Venezuelan crisis). In general, the historical contributions of the oil-specific demand shock (which is a hybrid shock in our model, see footnote 16) are well in line with results provided by Kilian and Murphy (2014). While the general pattern of historical decompositions implied by the KML model under mutual independence and the recursive model of Kilian (2009) align with results of Kilian and Murphy (2014), it is interesting to see

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 show the consistency of the structural estimates under much weaker conditions as required for (pseudo) ML estimation in the vein of suggestions of Lanne et al. (2017) or Gouriéroux et al. (2017). As a particular merit, the KML design can be straightforwardly 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) that has been identified by means of sign restrictions.

References

Alquist, R. and L. Kilian (2010): "What do we learn from the price of crude oil futures?" *Journal of Applied Econometrics*, 25, 539–573.

Bach, F. R. and M. I. Jordan (2003): "Kernel independent component analysis," *J.*

how close the historical decompositions implied by the model with partially dependent shocks match results of Kilian and Murphy (2014) in terms of numerical estimates. A prime example, for instance, could be the relative assessment of the contribution of both demand shocks for the oil price movements around 1999. Both Kilian and Murphy (2014) and the partial dependence model imply (unlike Cholesky and full KML) a much stronger contribution of aggregate (flow) demand than oil-specific (speculative) demand.

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.

Blanchard, O. J. and D. Quah (1989): "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 655–673.

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. (2021): "The importance of supply and demand for oil prices: evidence from non-Gaussianity," Bank of England working papers 957, Bank of England.

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

Canova, F. and G. De Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

Cardoso, J.-F. (1998a): "Blind signal separation: statistical principles," *Proceedings of the IEEE*, 86, 2009–2025.

——— (1998b): "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.

——— (1999): "High-order contrasts for independent component analysis," *Neural Computation*, 11, 157–192.

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.

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

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.

——— (2022): "Dynamic Score-Driven Independent Component Analysis," *Journal of Business & Economic Statistics*, 0, 1–11.

Hafner, C. M., H. Herwartz, and S. Maxand (2022): "Identification of structural multivariate GARCH models," *Journal of Econometrics*, 227, 212–227, annals Issue: Time Series Analysis of Higher Moments and Distributions of Financial Data.

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.

Helwig, N. E. (2022): *ica: Independent Component Analysis*, R package version 1.0-3.

Herwartz, H. (2018): "Hodges–Lehmann Detection of Structural Shocks – An Analysis of Macroeconomic Dynamics in the Euro Area," *Oxford Bulletin of Economics and Statistics*, 80, 736–754.

Herwartz, H., S. Maxand, and H. Rohloff (2022): "The link between monetary policy, stock prices, and house prices—Evidence from a statistical identification approach," *International Journal of Central Banking*, 18, 111–164.

——— (forthcoming): "Lean against the wind or float with the storm? Revisiting the monetary policy asset price nexus by means of a novel statistical identification approach," *International Journal of Central Banking*.

Hyvärinen, A. (1999): "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, 10, 626–634.

Hyvärinen, A. (1999): "The Fixed-Point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis," *Neural Processing Letters*, 10, 1–5.

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.

Koesner, C. L. and K. Nordhausen (2021): *KernelICA: Kernel Independent Component Analysis*, R package version 0.1.0.

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.

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

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

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.

Stock, J. H. and M. W. Watson (2012): "Disentangling the channels of the 2007-2009 recession," Working Paper 18094, National Bureau of Economic Research.

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

A 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. Assumption (A2) implies that

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

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

$$|\Delta(\theta') - \Delta(\theta)| \leq B||\theta' - \theta|| \quad a.s.$$

where $B = \sup_{\theta^* \in \Theta^*} ||\partial\Delta(\theta^*)/\partial\theta||$, Θ^* is the line segment joining θ and θ', and B is bounded in probability because $f_{ih}(\xi_{it}(\theta))$ and $\hat{f}_{ih}(\xi_{it}(\theta))$ 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 $\sup_\theta |\log \hat{f}_{ih}(\xi_{it}(\theta)) - \log f_{ih}(\xi_{it}(\theta))| = o_p(1)$ by Theorem 21.9 of Davidson (1994). Therefore,

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

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

B Local concavity

The following condition ensuring the local consistency of the (P)ML estimator is widely discussed in the literature (e.g. Cardoso 1998a, Hyvärinen 1999, Th. 1, see also Hyvärinen et al. 2001, Th. 9.1):

Theorem 2. *Denoting Q_0 the true orthogonal mixing matrix, the (pseudo) log-likelihood function $\sum_{i=1}^{N} \mathbb{E} \log \tilde{f}_i \left(e_i^\top Q^\top \mathrm{z}_t \right)$ is locally concave at Q_0, if $\tilde{f}_i$ fulfills*

$$\mathbb{E}\left[\tilde{s}_i(x)x - \tilde{s}_i'(x) \right] > 0,\ i = 1, \ldots, N-1. \tag{13}$$

Proof. Transform the the observations $\tilde{\mathrm{z}}_t = (C^{-1}B)^\top \mathrm{z}_t$. The true mixing matrix based on observations $\tilde{\mathrm{z}}_t$ is $Q_0 = I_N$. We expand the asymptotic normalized log-likelihood in a neighbourhood of Q_0, i.e. $l(Q_0 + \boldsymbol{\epsilon}) \coloneqq \lim_{T\to\infty} T^{-1} l_T(Q_0 + \boldsymbol{\epsilon})$, where $Q_0 + \boldsymbol{\epsilon} \in \mathcal{O}(N)$. Without loss of generality, we focus on the log-likelihood with respect to the i-th column of Q_0, which gives the i-th independent component $\xi_{it} = e_i^\top Q^\top \tilde{\mathrm{z}}_t = q_i^\top \tilde{\mathrm{z}}_t$ and

$$l(q_i + \epsilon_i) \simeq l(q_i) + \mathbb{E}\left[e_i^\top \tilde{\mathrm{z}}_t \tilde{s}_i(e_i^\top \tilde{\mathrm{z}}_t) \right] \epsilon_{ii} + \frac{1}{2} \epsilon_i^\top \mathbb{E}\left[\nabla^2 l(q_i)|_{q_i = e_i} \right] \epsilon_i,$$

where

$$\mathbb{E}\left[\nabla^2 l(q_i)|_{q_i=e_i} \right] = \begin{bmatrix} \mathbb{E}[\xi_{1t}^2 \tilde{s}_i'(\xi_{it})] & \cdots & \mathbb{E}[\xi_{1t}\xi_{Nt}\tilde{s}_i'(\xi_{it})] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\xi_{Nt}\xi_{1t}\tilde{s}_i'(\xi_{it})] & \cdots & \mathbb{E}[\xi_{Nt}^2 \tilde{s}_i'(\xi_{it})] \end{bmatrix}.$$

Notice that $\mathbb{E}[\xi_{it}\xi_{jt}\tilde{s}_n'(\xi_{nt})] = \mathbb{E}[\xi_{it}\xi_{jt}]\mathbb{E}[\tilde{s}_n'(\xi_{nt})], \forall i, j \neq n$ and $\mathbb{E}[\xi_{it}\xi_{jt}] = \delta_{ij}$, where δ_{ij} is the Kronecker delta. Thus, $\mathbb{E}\left[\nabla^2 l(q_i)|_{q_i=e_i} \right]$ is a diagonal matrix with i-th diagonal element being $\mathbb{E}[\xi_{it}^2]\mathbb{E}[\tilde{s}_j'(\xi_{jt})] = \mathbb{E}[\tilde{s}_j'(\xi_{jt})]$ for $i \neq j$ and $\mathbb{E}[\xi_{it}^2 \tilde{s}_j'(\xi_{jt})]$ for $i = j$. Therefore, we obtain

$$l(q_i + \epsilon_i) \simeq l(q_i) + \mathbb{E}\left[\xi_{it}\tilde{s}_i(\xi_{it}) \right] \epsilon_{ii} + \frac{1}{2}\mathbb{E}[\tilde{s}_i'(\xi_{it})] \sum_{n\neq i} \epsilon_{ni}^2 + \frac{1}{2}\mathbb{E}[\xi_{it}^2 \tilde{s}_i'(\xi_{it})]\epsilon_{ii}^2$$

Moreover, since the parameter space is restricted to $\mathcal{O}(N)$, the parameter vector under scrutiny $q_i + \epsilon_i$ must have unit length, which implies $\sum_{j=1}^{N}(q_{ij} + \epsilon_{ij})^2 = 1$. Notice that $(q_{ij} + \epsilon_{ij})^2 = \epsilon_{ij}^2$ for $i \neq j$ and $(q_{ii} + \epsilon_{ii})^2 = (1 + \epsilon_{ii})^2$. Accordingly, the unit length restriction leads to the approximation (following a simple first-order Taylor expansion of the square root) $\epsilon_{ii} = \sqrt{1 - \sum_{n\neq i} \epsilon_{ni}^2} - 1 \simeq -\frac{1}{2}\sum_{n\neq i} \epsilon_{ni}^2$. As a result, we obtain

$$\begin{aligned} l(q_i + \epsilon_i) &\simeq l(q_i) - \frac{1}{2}\mathbb{E}\left[\xi_{it}\tilde{s}_i(\xi_{it}) \right] \sum_{n\neq i} \epsilon_{ni}^2 + \frac{1}{2}\mathbb{E}[\tilde{s}_i'(\xi_{it})] \sum_{n\neq i} \epsilon_{ni}^2 \\ &=\simeq l(q_i) - \frac{1}{2}\mathbb{E}\left[\xi_{it}\tilde{s}_i(\xi_{it}) - \tilde{s}_i'(\xi_{it}) \right] \sum_{n\neq i} \epsilon_{ni}^2. \end{aligned}$$

□

Remark 1 (Identifiability)**.** *The condition* (13) *is violated, if the i-th structural component, with* $i \in \{1, \ldots, N-1\}$*, has a standard Gaussian distribution with pdf* $\varphi(x)$*.*

Proof. For a standard Gaussian random variable X, $s_\varphi(x) = -x$ and the left-hand-side of (13) becomes[18]

$$\int \left[\frac{\varphi'(x)}{\varphi(x)} x - \frac{\varphi''(x)\varphi(x) - \varphi'(x)^2}{\varphi(x)^2} \right] \varphi(x) dx = \int (1 - x^2)\varphi(x) dx = 0.$$

□

Notice that at most one element in ξ_t can be Gaussian. If the condition (13) is satisfied for $N-1$ components, the last column in Q is self-identified. Hence, the identifiability of the system requires that at most one element in ξ_t exhibits a Gaussian distribution (see also Comon 1994). In the borderline case, in which the random vector ξ_t follows a N-dimensional Gaussian distribution, the asymptotic relative gradient becomes $-Q^\top \mathbb{E}\left[\mathrm{z}_t \mathrm{z}_t^\top\right] Q = -I_N$, which holds for all $Q \in \mathcal{O}(N)$. As a result, it is not possible to disentangle z_t from any orthogonal rotation of z_t and the system is not identifiable. Moreover and in presence of more than one Gaussian component, Theorem 2 allows for partial identification of the n-th structural component, with $n \in \{1, \ldots, N\}$, as long as the condition (13) holds for such component. A similar conclusion was also drawn by Maxand (2018).

Gouriéroux et al. (2017) have suggested and employed some standardized (i.e. with zero mean and unit variance) pseudo pdf that can be used to construct the pseudo likelihood function. These are Student $t(\nu)$ distribution with degrees of freedom $\nu > 4$, hyperbolic secant and Gaussian mixture. As an example, we shall examine the concavity condition for the hyperbolic secant and the Gaussian mixture. The term within the expectation on the left-hand-side of (13) for these pseudo pdfs read as[19]

$$-\frac{\pi}{2} x \tanh \frac{\pi}{2} x + \left(\frac{\pi}{2} \frac{1}{\cosh \frac{\pi}{2} x} \right)^2 \quad \text{and} \quad \frac{\pi}{2} x \tanh \frac{\pi}{2} x - \left(\frac{\pi}{2} \frac{1}{\cosh \frac{\pi}{2} x} \right)^2,$$

respectively. It becomes evident that the non-polynomial moments of these two pseudo pdfs have opposite signs. Therefore, one may expect that PML based on either of these pseudo pdfs will work in some cases, but fail in others. Put differently, in a particular scenario, only

[18] Notice that the r-th derivative of Gaussian pdf, denoted $\varphi^{(r)}(x)$, is given by $\varphi^{(r)}(x) = (-1)^r \mathcal{H}_r(x)\varphi(x)$, where $\mathcal{H}_r(x)$ is the r-th Hermite polynomial. It follows that $\varphi'(x) = -x\varphi'(x)$ and $\varphi''(x) = (x^2 - 1)\varphi(x)$.

[19] For a detailed description of the log-densities and the first two derivatives, we refer the reader to Table 1 of Gouriéroux et al. 2017.

one of these pdfs can satisfy the concavity condition (13), but not both. While choosing the hyperbolic secant (Gaussian mixture) as $\tilde{f}_i$ will lead to consistent estimation of q_i, if the true distribution of the underlying component f_i is leptokurtic (platykurtic), the condition (13) is violated once f_i is in fact platykurtic (leptokurtic).

Supplementary materials for 'Causal inference with (partially) independent shocks and structural signals on the global crude oil market'

January 25, 2023

Contents

A Further simulation results **2**

B Joint Wald confidence region for impulse responses **5**

C Shotgun trajectories **5**

References **8**

A Further simulation results

In this Section, we display simulation results from the baseline setting with alternative sample sizes T=200 and 5000 in Figure A.1 and A.2. In addition, we compare the performance of these ICA routines in the presence of heteroscedastic and co-heteroscedastic source signals with stochastic volatilities in Figure A.3. In specific, we simulate a stochastic volatility (log-variance) process – either independently for each shock (heteroskedasticity) or commonly for all shocks (co-heteroscedasticity) – from a stationary AR(1) process (without intercept) and a centered Gaussian innovation with a standard deviation of 0.5, where the autoregressive coefficient in each experiment is uniformly drawn from the interval $[-0.99; 0.99]$. All Monte Carlo experiments are performed with 1000 replications. Results depicted in Figure A.3 underline that the performance of KML, GMM and KernelICA is markedly robust under (co-)heteroscedasticity, while the KML approach has a substantial performance lead in both small and large samples.

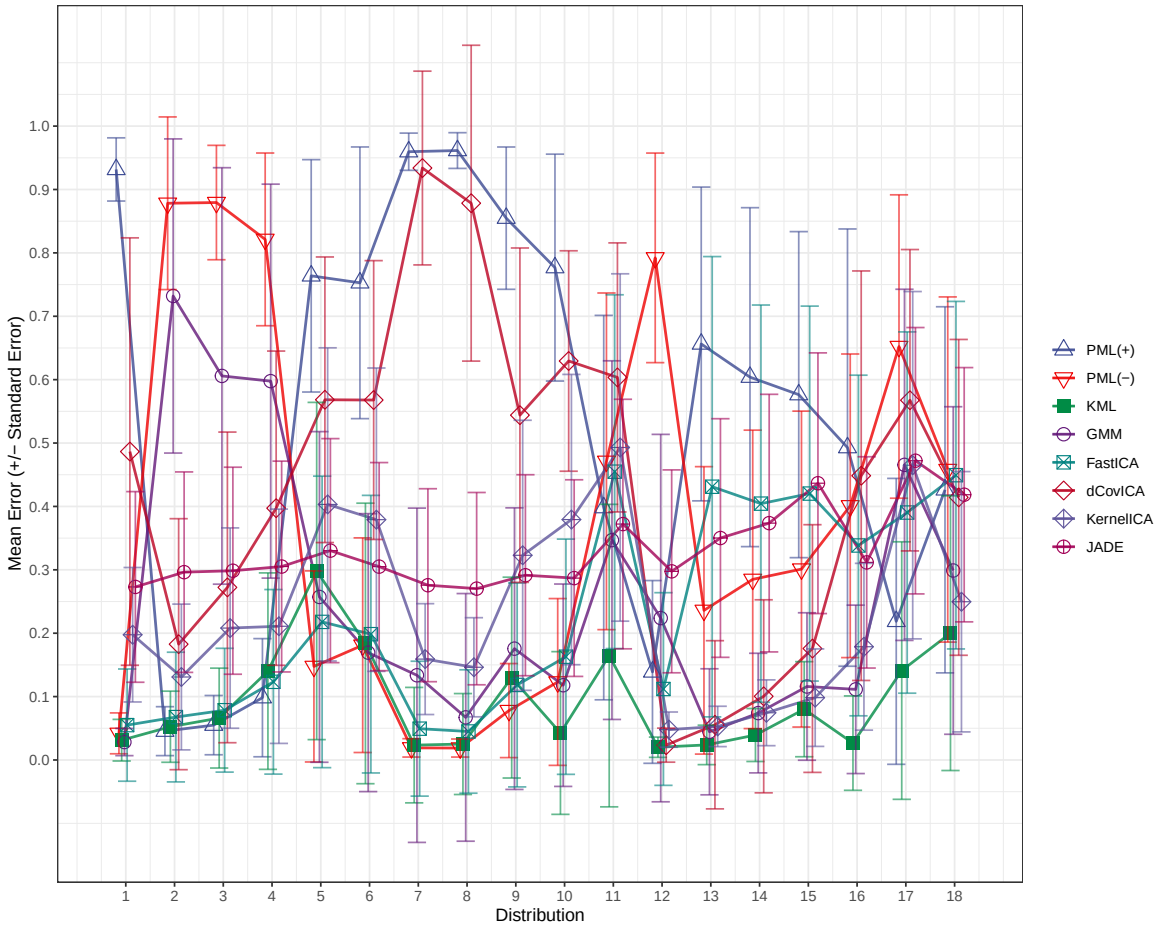


Figure A.1: Average estimation error ± one estimated standard error for alternative methods and distributions based on 1000 simulation experiments with a sample size of $T = 200$.

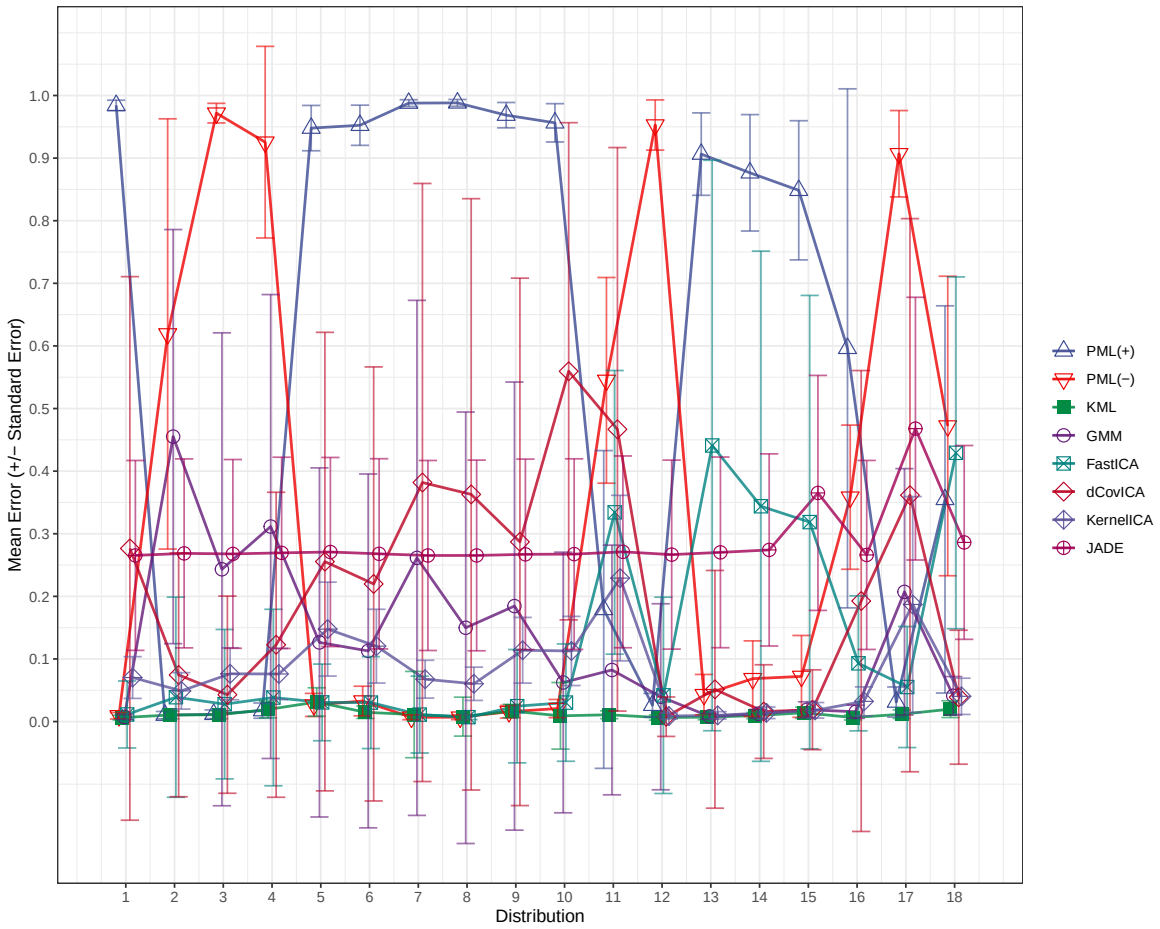


Figure A.2: Average estimation error ± one estimated standard error for alternative methods and distributions based on 1000 simulation experiments with a sample size of $T = 5000$.

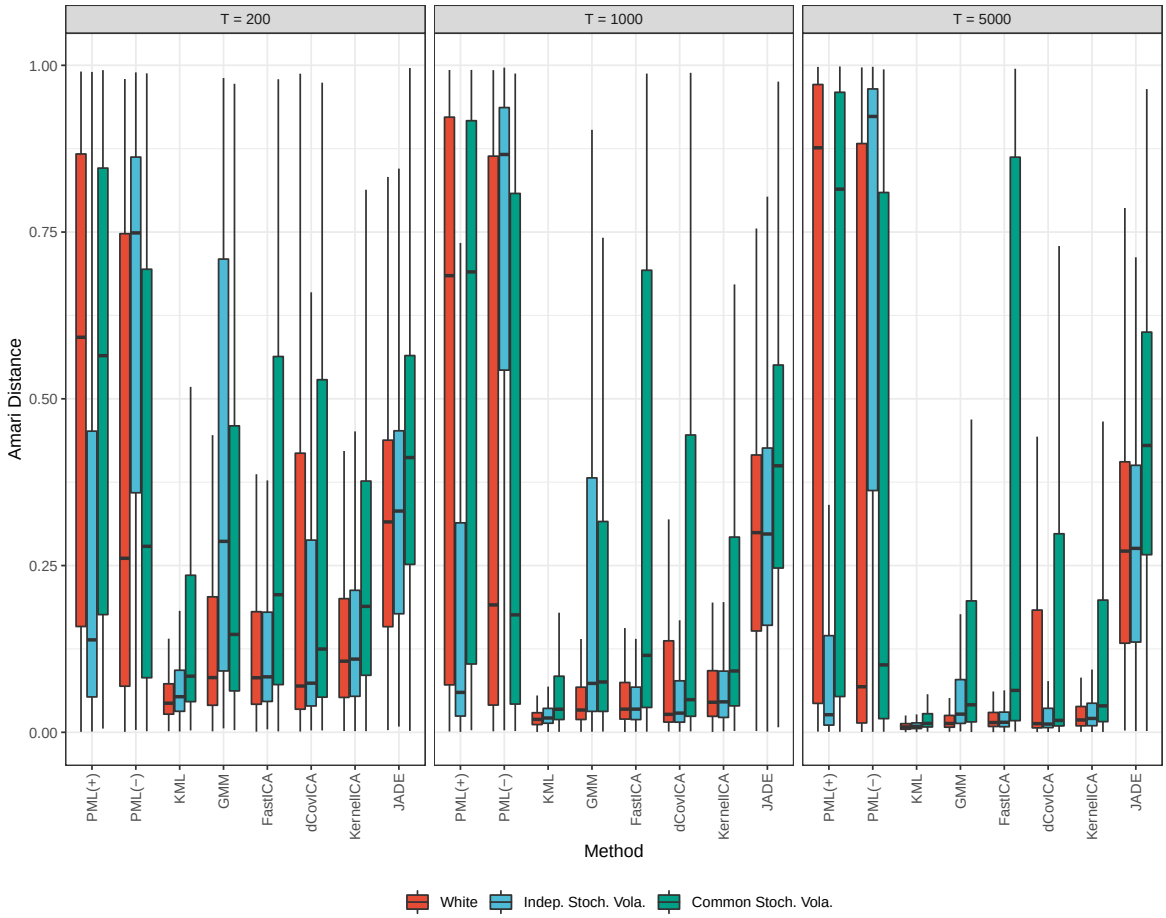


Figure A.3: Distribution of the Amari distance via boxplots (with horizontal lines marking the medians) for alternative methods, sample sizes and shocks with distinct distributions (randomly drawn with replacement from the 18 distributions displayed in Figure 1) and second-order-moment-features (white noise, independent and common stochastic volatility) based on 1,000 simulation experiments.

B Joint Wald confidence region for impulse responses

Inference of structural IRFs based on point-wise confidence intervals has been criticized for its lack of coverage accuracy, since it ignores the fact that estimates of IRFs are typically dependent across horizons and different response functions (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. Let $\eta = (\boldsymbol{a}, \boldsymbol{\sigma})$ denote the vector that collects the AR parameters $\boldsymbol{a} = \text{vec}([A_1, \ldots, A_p])$ and the lower triangular part of the residual covariance matrix $\boldsymbol{\sigma} = \text{vech}(\Sigma_\text{x})$. Let $\phi = g(\eta)$ denote the vector of structural impulse responses of interest and $g(\cdot)$ is a nonlinear function. Given the estimates $\hat{\eta}_T$, the Wald statistic is

$$W_T = T(\hat{\eta}_T - \eta)^\top \hat{\Sigma}_\eta^{*-1}(\hat{\eta}_T - \eta),$$

where $\hat{\Sigma}_\eta^{*-1} = 1/M \sum_{r=1}^{M}(\hat{\eta}_T^{*r} - \hat{\eta}_T)(\hat{\eta}_T^{*r} - \hat{\eta}_T)^\top$ and $\hat{\eta}_T^{*r}$ is the vector of estimates obtained from the r-th bootstrap for $r = 1, \ldots, M$. Define

$$W_r^* = T(\hat{\eta}_T^{*r} - \hat{\eta}_T)^\top \hat{\Sigma}_\eta^{*-1}(\hat{\eta}_T^{*r} - \hat{\eta}_T),\ r = 1, \ldots, M,$$

and order the bootstrap estimates of the statistic such that $W_1^* \leq \cdots \leq W_M^*$. The $1 - \alpha$ Wald confidence set is chosen such that

$$W_{1-\alpha}^\eta = \{\eta | W_T \leq W_{(1-\alpha)M}^*\}.$$

As shown by Lütkepohl et al. (2015), as T and M tend to ∞, and under the usual conditions attached to the bootstrap procedure, the confidence set $W_{1-\alpha}^\eta$ is asymptotically valid, i.e.

$$\mathbb{P}(\eta_0 \in W_{1-\alpha}^\eta) \to 1 - \alpha,$$

where η_0 is the true parameter. Let $\hat{\phi}_i^{*r} = g(\hat{\eta}_T^{*r})$ denote the i-th structural IRFs obtained in the r-th bootstrap realization. The joint confidence region for ϕ can be constructed based on the lower and upper bounds for each element in ϕ determined as

$$\min\{\hat{\phi}_i^{*r} | r = 1, \ldots, (1-\alpha)M\} \quad \text{and} \quad \max\{\hat{\phi}_i^{*r} | r = 1, \ldots, (1-\alpha)M\},$$

respectively.

C Shotgun trajectories

Inoue and Kilian (2016) propose a shotgun trajectory plot by presenting IRFs obtained from each individual structural model covered by the confidence region. Compared with

the conventional graphical representation of confidence regions, the shotgun plot can provide additional information about the shape of the IRFs over the horizons. For this reason, we display the shotgun plot for IRFs obtained from the unrestricted model with full ICA, the restricted counterpart and the responses to an aggregate demand shock obtained from partial identification.

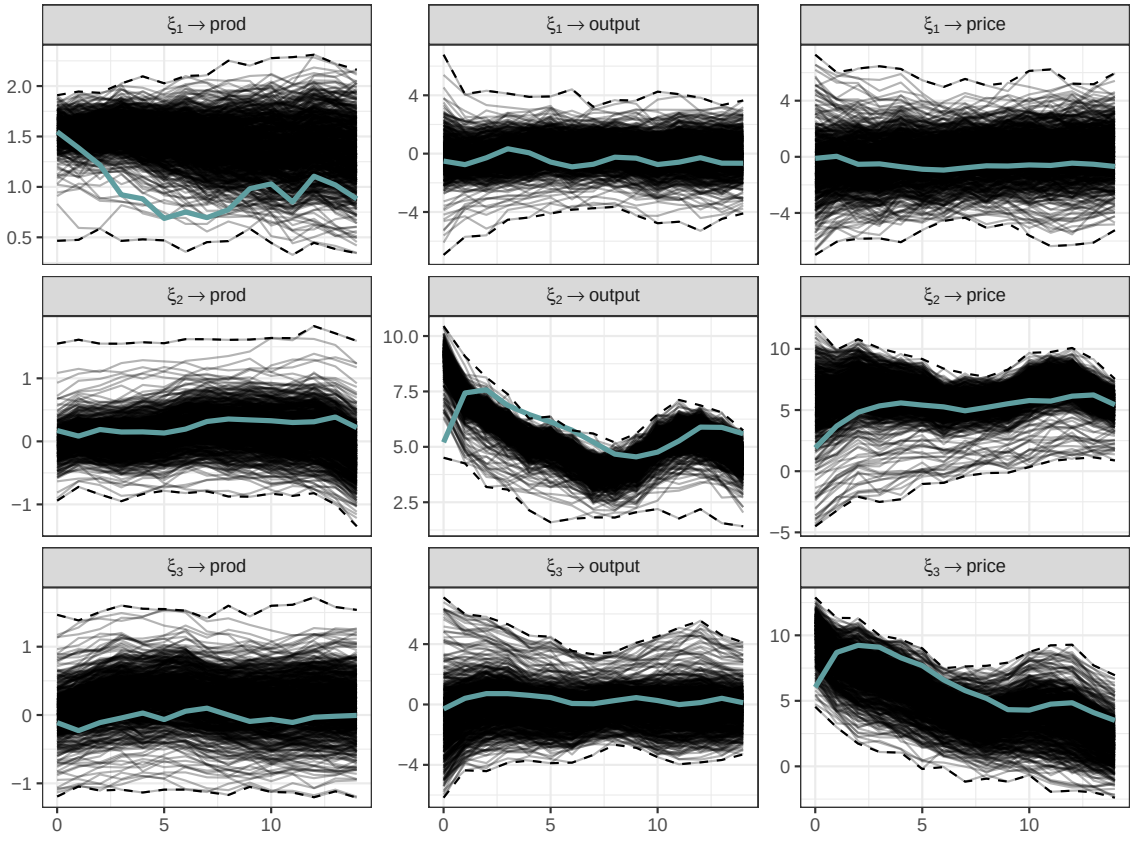


Figure C.1: Impulse responses to one-standard-deviation shocks identified from unrestricted model by imposing mutual independence. The structural shocks are oil supply shock (ξ_1), aggregate demand shock (ξ_2) and speculative oil demand shock (ξ_3). Shotgun plot implied by joint 68% Wald confidence set approximated by the moving block bootstrap with 1000 replications. Thick solid (light blue) line indicates estimates from the original data.

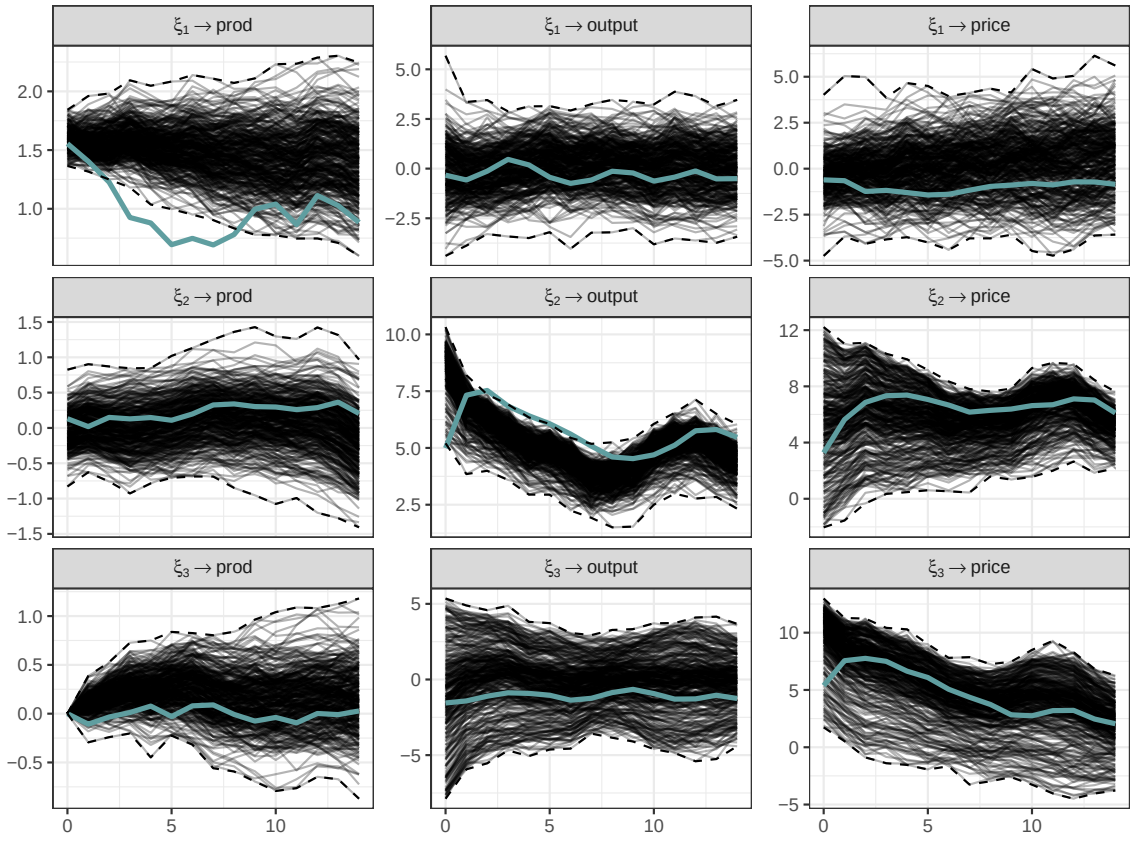


Figure C.2: Impulse responses to one-standard-deviation shocks identified by imposing partial independence and the elasticity restriction. The structural shocks are oil supply shock (ξ_1), aggregate demand shock (ξ_2) and speculative oil demand shock (ξ_3). Shotgun plot implied by joint 68% Wald confidence set approximated by the moving block bootstrap with 1000 replications. Thick solid (light blue) line indicates estimates from the original data.

References

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.

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.