Hoesch, Lukas; Lee, Adam; Mesters, Geert

Article

Locally robust inference for non-Gaussian SVAR models

Quantitative Economics

Provided in Cooperation with:
The Econometric Society

Suggested Citation: Hoesch, Lukas; Lee, Adam; Mesters, Geert (2024) : Locally robust inference for non-Gaussian SVAR models, Quantitative Economics, ISSN 1759-7331, The Econometric Society, New Haven, CT, Vol. 15, Iss. 2, pp. 523-570, https://doi.org/10.3982/QE2274

Quantitative Economics 15 (2024), 523–570 1759-7331/20240523

Locally robust inference for non-Gaussian SVAR models

Lukas Hoesch
Department of Econometrics and Data Science, Vrije Universiteit Amsterdam and Tinbergen Institute

Adam Lee
Department of Data Science and Analytics, BI Norwegian Business School

Geert Mesters
Department of Economics and Business, Universitat Pompeu Fabra, Barcelona School of Economics, and CREI

All parameters in structural vector autoregressive (SVAR) models are locally identified when the structural shocks are independent and follow non-Gaussian distributions. Unfortunately, standard inference methods that exploit such features of the data for identification fail to yield correct coverage for structural functions of the model parameters when deviations from Gaussianity are small. To this extent, we propose a locally robust semiparametric approach to conduct hypothesis tests and construct confidence sets for structural functions in SVAR models. The methodology fully exploits non-Gaussianity when it is present, but yields correct size/coverage for local-to-Gaussian densities. Empirically, we revisit two macroeconomic SVAR studies where we document mixed results. For the oil price model of Kilian and Murphy (2012), we find that non-Gaussianity can robustly identify reasonable confidence sets, whereas for the labor supply–demand model of Baumeister and Hamilton (2015) this is not the case. Moreover, these exercises highlight the importance of using weak identification robust methods to assess estimation uncertainty when using non-Gaussianity for identification.

Keywords. Weak identification, semiparametric inference, hypothesis testing, impulse responses, independent component analysis.

JEL classification. C32, C39, C51.

1. Introduction

In this paper, we develop locally robust inference methods for non-Gaussian structural vector autoregressive (SVAR) models. To outline our contribution, consider the SVAR

Lukas Hoesch: l.hoesch@vu.nl
Adam Lee: adam.lee@bi.no
Geert Mesters: geert.mesters@upf.edu

We thank Majid Al-Sadoon, Regis Barnichon, Christian Brownlees, Saskia ter Ellen, Juan Carlos Escanciano, Kirill Evdokimov, Eleonora Granziera, Katerina Petrova, Barbara Rossi, and Piotr Zwiernik, as well as seminar and conference participants from different institutions for helpful comments. Mesters acknowledges financial support from the European Research Council via Starting Grant 101041145—POLICYMETRICS, and from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S).

 https://doi.org/10.3982/QE2274

model

$$Y_t = c + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + A^{-1}\epsilon_t, \tag{1}$$

where Y_t is a $K \times 1$ vector of variables, c is an intercept, $B_1, \ldots, B_p$ are the autoregressive matrices, A is the invertible contemporaneous effect matrix, and ϵ_t is the $K \times 1$ vector of structural shocks with mean zero and unit variance.

It is well known that, without further restrictions, the first and second moments of $\{Y_t\}$ are insufficient to identify all parameters in A (e.g., Kilian and Lütkepohl (2017)). Instead, higher-order moments or non-Gaussian distributions can be exploited to (locally) identify A. The most well-known result follows from the Darmois–Skitovich theorem and is central to the literature on independent components analysis (ICA): if the components of ϵ_t are independent and at least $K-1$ have a non-Gaussian distribution, then A can be recovered up to sign and permutation of its rows; see Comon (1994). Based on such results, several recent works have exploited non-Gaussianity to improve identification and conduct inference in SVAR models (e.g., Lanne and Lütkepohl (2010), Moneta, Entner, Hoyer, and Coad (2013), Lanne, Meitz, and Saikkonen (2017), Kilian and Lütkepohl (2017), Maxand (2020), Lanne and Luoto (2021), Gouriéroux, Monfort, and Renne (2017, 2019), Tank, Fox, and Shojaie (2019), Herwartz (2019), Bekaert, Engstrom, and Ermolov (2021, 2020), Fiorentini and Sentana (2022), Braun (2021), Sims (2021), Guay (2021), Brunnermeier, Palia, Sastry, and Sims (2021), Drautzburg and Wright (2023), Keweloh (2021), Davis and Ng (2022), Lanne, Liu, and Luoto (2022)).[1,2]

Unfortunately, as we show in this paper, standard inference methods for non-Gaussian SVARs are not robust in situations where the densities of the structural shocks are too "close" to the Gaussian density. Intuitively, what matters for correctly sized inference is not non-Gaussianity per se, but a sufficient distance from the Gaussian distribution. When the true distributions of the structural shocks are close to the Gaussian distribution, local identification deteriorates and coverage distortions occur in confidence sets for structural functions, for example, structural impulse response functions or forecast error variance decompositions.[3] The problem is somewhat analogous to the weak instruments problem where it is well known that nonzero correlation between the instruments and the endogenous variables is not sufficient for standard inference methods to be reliable; the correlation must be sufficiently large in order for conventional IV asymptotic theory to provide an approximation, which accurately reflects the finite sample situation.[4] Similarly, in our setting, non-Gaussianity alone is not sufficient for

[1] See Olea, Luis, Plagborg-Møller, and Qian (2022) for a recent review of this approach and, for example, Lewis (2021) for a related approach based on heteroskedasticity.

[2] ICA type identification results have been applied/extended for various related models such as linear simultaneous equations models, graphical models, and factor models (e.g., Shimizu, Hoyer, Hyvärinen, and Kerminen (2006), Bonhomme and Robin (2009), Wang and Drton (2019)).

[3] Simulation studies in, among others, Gouriéroux, Monfort, and Renne (2017, 2019) and Lanne and Luoto (2021) have previously highlighted such coverage distortions for parameter estimates in the case of "weakly" non-Gaussian distributions; see also Lee and Mesters (2024a) for more discussion of the same issue in static ICA models.

[4] See, for example, the recent review by Andrews, Stock, and Sun (2019).

standard (pseudo)-maximum likelihood or generalized method of moments methodologies to yield correct coverage when the distance to the Gaussian distribution is not sufficiently large. As such, we refer to this phenomenon as "weak non-Gaussianity."

In this paper, we propose a solution to this problem by combining insights from the econometric literature on weak identification robust hypothesis testing as well as the statistical literature on semiparametric inference. Specifically, we treat the SVAR model with independent structural shocks as a semiparametric model where the densities of the structural shocks form the nonparametric part.

For this set-up, we provide three main results. First, we adopt a semiparametric generalization of Neyman–Rao score statistic in order to test the possibly weakly identified (or under/unidentified) parameters of the SVAR. More precisely, the semiparametric score statistic that we propose is based on a quadratic form of the efficient score function, which projects out all scores for the nuisance parameters, including the scores corresponding to the density functions of the structural shocks, from the conventional score function for the parameter of interest. This projection, along with the fact that the potentially weakly/nonidentified parameter is fixed under the null when conducting the test (as is standard in score-type testing procedures), enables us to circumvent the (weak-)identification problem and we show that the semiparametric score test has a χ^2 limit under local parameter sequences consistent with the null hypothesis.

Second, we propose a method for constructing confidence sets for smooth structural functions. Prominent examples of interest include structural impulse responses and forecast error variance decompositions. Specifically, we utilize our proposed score test for the weakly identified parameters in a Bonferroni-based procedure (cf. Granziera, Moon, and Schorfheide (2018), Drautzburg and Wright (2023)), which is guaranteed to provide correct coverage asymptotically.

Third, under the *additional* assumption that the errors of the SVAR model follow distributions that are different from the Gaussian distribution in the limit, we show that point estimates, constructed as one-step updates based on the efficient score function, are consistent and semiparametrically efficient for the finite-dimensional parameters in the semiparametric SVAR model. This implies that under strong identification and some regularity conditions such estimators are preferable to existing pseudo MLE and GMM estimators.

Overall, our methods are computationally simple as the estimation of the efficient scores typically only requires estimating regression coefficients, a covariance matrix, and the log density scores of the structural shocks. To estimate the log density scores, we use B-spline regressions as developed in Jin (1992) and also considered in Chen and Bickel (2006) for semiparametric independent component analysis. This approach is computationally convenient and allows our methodology to work under a wide variety of possible distributions for the structural shocks.[5]

We assess the finite-sample performance of our method in a large simulation study and find that the empirical rejection frequencies of the semiparametric score test are always close to the nominal size. This is in contrast to several existing methods that are not

[5]The general approach is applicable with other choices of log density score estimators, for example, the local polynomial estimators proposed in Pinkse and Schurter (2021). The main requirement is that the chosen estimator should satisfy the high-level conditions stated in Lemma 2.

robust to weak non-Gaussianity and show substantial size distortions for non-Gaussian distributions that are close to the Gaussian density. We also analyze the power of the proposed procedure and find that the power of the semiparametric score test generally exceeds that of alternative robust methods such as weak identification robust GMM methods. Finally, we show that while the Bonferroni approach for constructing confidence sets for structural functions is (by construction) conservative, it does often substantially reduce the length of the confidence bands for structural impulse responses when compared to alternative methods.

In our empirical study, we revisit two prominent macroeconomic SVAR applications and ask whether non-Gaussian distributions can help to robustly identify structural functions of interest. Specifically, we revisit (i) the labor supply–demand model of Baumeister and Hamilton (2015) and (ii) the oil price model of Kilian and Murphy (2012).[6] Our findings are mixed.

In the labor supply–demand model of Baumeister and Hamilton (2015), we find that allowing for non-Gaussian structural shocks does not lead to a tight confidence set for the supply and demand elasticities. In contrast, when nonrobust methods are used, as in Lanne and Luoto (2022) for instance, non-Gaussianity appears to pin down the elasticities in a narrow set. These findings strongly support the usage of robust confidence sets when assessing uncertainty around parameter estimates obtained using non-Gaussianity as an identifying assumption.

For the oil price model of Kilian and Murphy (2012), non-Gaussian structural shocks provide substantially more identifying power. In fact, we show that our robust methodology yields a finite confidence set for the short-run oil supply elasticities, thus avoiding the need to impose a priori bounds on these elasticities. For instance, the bounds imposed in Kilian and Murphy (2012) have been criticized for being too tight in Baumeister and Hamilton (2019) and have led to a large literature that assesses their importance; see Herrera and Rangaraju (2020) for an overview. We show that exploiting non-Gaussian shocks leads to finite bounds that are within the range of estimates documented in the literature, hence providing a data driven solution to the determination of appropriate bounds.

This paper relates to several strands of literature. First and foremost, the paper contributes to the SVAR literature that exploits non-Gaussianity of the structural shocks for identification (see the references above). Most related, Drautzburg and Wright (2023) are also concerned about identification when using higher-order moment restrictions for identification. To circumvent distortions in confidence sets, they exploit the identification robust S-statistic of Stock and Wright (2000) as well as a nonparametric independence test for conducting inference. The benefit of the S-statistic is that it is not necessary to assume full independence of the structural shocks. Instead, typically only the third- and fourth-order higher cross-moments are set to zero, leaving all higher-order moments unrestricted. A downside of such a robust moment approach is that it requires the existence of substantially higher-order moments. For instance, when using fourth-

[6]The assumption of independence among the structural shocks is maintained throughout this paper. Therefore, in each application we test for the existence of independent components following both Matteson and Tsay (2017) and Olea et al. (2022).

order moment restrictions, the convergence of the S-statistic requires the existence of at least eight moments. We provide a detailed comparison between the approaches in our simulation study.

Besides the non-Gaussian SVAR literature, we note that our approach is inspired by the identification robust inference literature in econometrics (e.g., Stock and Wright (2000), Kleibergen (2005), Andrews and Mikusheva (2015)). The crucial difference in our setting is that the nuisance parameters which determine identification status are infinite-dimensional, that is, the densities of the structural shocks. Despite this difference, conceptually our approach is similar to the score testing approach developed for weakly identified parametric models in Andrews and Mikusheva (2015). To handle infinite-dimensional nuisance parameters, we build on the general statistical theory discussed in Bickel, Klaasen, Ritov, and Wellner (1998) and van der Vaart (2002). While the majority of the statistical literature focuses on efficient estimation in semiparametric models, a few papers have contributed to testing in well-identified models (e.g., Choi, Hall, and Schick (1996), Bickel, Ritov, and Stoker (2006)). The major difference with our paper is that in our setting, a subset of the parameters of interest are possibly weakly—or un/underidentified, which violates a key regularity condition assumed in this literature. Lee and Mesters (2024a) consider a similar score testing approach, but their setting only considers static linear models, and hence their results cannot be applied to the SVAR models that are of interest in this paper.

The remainder of this paper is organized as follows. Section 2 casts the SVAR model as a semiparametric model and discusses the needed regularity conditions. Section 3 establishes a number of preliminary results that are of general interest. The semiparametric score testing approach is presented in Section 4 and inference for smooth structural functions is covered in Section 5. Section 6 discusses point estimation under strong identification. Section 7 evaluates the finite-sample performance of the proposed methodology and Section 8 discusses the results from the empirical studies. Section 9 concludes. Any references to sections, equations, lemmas, etc., which start with "S," refer to the Supplementary Material (Hoesch, Lee, and Mesters, 2024).

2. Semiparametric SVAR model

In this section, we cast the SVAR model as a semiparametric model and impose some primitive assumptions that will be maintained throughout the text. For convenience, we adopt the following notation for the SVAR model:

$$Y_t = BX_t + A^{-1}(\alpha, \sigma)\epsilon_t, \quad t \in \mathbb{Z}, \tag{2}$$

where $X_t := (1, Y_{t-1}', \ldots, Y_{t-p}')'$, $B := (c, B_1, \ldots, B_p)$, and $A(\alpha, \sigma)$ is a $K \times K$ invertible matrix that is parametrized by the vectors α and σ.

In general, we leave the choice for the specific parametrization of $A(\alpha, \sigma)$ open to the researcher. The key restriction is that σ should be recoverable from the variance of $Y_t - BX_t$ after α has been fixed, whereas α itself may be unidentified depending on the distribution of the structural shocks. One popular choice in this context sets

$A^{-1}(\alpha, \sigma) = \Sigma^{1/2}(\sigma)R(\alpha)$, where $\Sigma^{1/2}(\sigma)$ is a lower triangular matrix (with positive diagonal elements) parametrized by the vector σ and $R(\alpha)$ is a rotation matrix that is parametrized by the vector α. Alternatively, letting σ capture the lower triangular entries of $A^{-1}(\alpha, \sigma)$ and α the strictly upper triangular entries also defines an easy to interpret parametrization.[7]

To describe the nonparametric part of model (2), we let $\eta = (\eta_1, \ldots, \eta_K)$ correspond to the density functions of $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{K,t})'$. All parameters are summarized as follows:

$$\theta = (\gamma, \eta), \qquad \gamma = (\alpha, \beta), \qquad \beta = (\sigma, b), \tag{3}$$

where $b = \text{vec}(B)$.

Let $Y^n = (Y_1, \ldots, Y_n)'$ and let P_θ^n denote the distribution of Y^n conditional on the initial values $(Y_{1-p}, \ldots, Y_0)$. Throughout, we work with these conditional distributions; see Hallin and Werker (1999) for a similar setup. For a sample of size n, our semi-parametric SVAR model is the collection

$$\mathcal{P}_\Theta^n = \left\{P_\theta^n : \theta \in \Theta\right\}, \qquad \Theta = \underbrace{\mathcal{A} \times \mathcal{B}}_{\Gamma} \times \mathcal{H}, \tag{4}$$

where $\Gamma \subset \mathbb{R}^L$, with $L = L_\alpha + L_\sigma + L_b$ corresponding to the dimensions of (α, σ, b), $L_\beta = L_\sigma + L_b$, and $\mathcal{H} \subset \prod_{k=1}^K \mathsf{H}$ with

$$\mathsf{H} := \left\{ f \in L_1(\lambda) \cap \mathcal{C}^1 : \ f(z) \geq 0, \int f(z)\,\mathrm{d}z = 1, \int z f(z) dz = 0, \int \kappa(z) f(z)\,\mathrm{d}z = 0 \right\},$$

where λ denotes Lebesgue measure on $\mathbb{R}$, $\mathcal{C}^1$ is the class of real functions on $\mathbb{R}$, which are continuously differentiable and $\kappa(z) = z^2 - 1$. It is understood that $\gamma \in \Gamma$ and $\eta \in \mathcal{H}$, where the parameter space for the densities η_k is restricted such that $\epsilon_{k,t}$ has mean zero and variance one. Further restrictions are placed on the parameter space Θ in the assumptions below.

Assumptions

Having defined the semiparametric SVAR model, we now proceed to formulate the required assumptions. Broadly speaking, we split our assumptions into two parts: Assumption 1 details the main assumptions that allow us to establish the properties of the semiparametric score test and Assumption 2 defines a set of regularity conditions on densities η_k under which the log density scores can be consistently estimated using B-splines.[8] These scores are an important ingredient for the methodology discussed below.

The main assumption is stated as follows.

[7] In general, different parametrizations are often used in practice (cf. Section 8) and our general formulation allows for all sufficiently smooth choices (cf. Assumption 1). The Supplementary Material (Hoesch, Lee, and Mesters (2024)) provides more discussion and examples.

[8] Lemma 2 in the Appendix shows that, under Assumptions 1 and 2, the B-spline based estimator satisfies a particular high-level condition; the results of this paper will continue to apply if any alternative density score estimator, which also satisfies this high-level condition is used.

ASSUMPTION 1. *For model (2), we assume that*:

(i) *For all* $\beta \in \mathcal{B}$, $|I_K - \sum_{j=1}^{p} B_j z^j| \neq 0$ *for all* $|z| \leq 1$ *with* $z \in \mathbb{C}$.

(ii) *Conditional on the initial values* $(Y'_{-p+1}, \ldots, Y'_0)'$, $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{K,t})'$ *is independently and identically distributed across* t, *with independent components* $\epsilon_{k,t}$. *Each* $\eta = (\eta_1, \ldots, \eta_K) \in \mathcal{H}$ *is such that each* η_k *is nowhere vanishing, dominated by Lebesgue measure on* $\mathbb{R}$, *continuously differentiable with log density scores denoted by* $\phi_k(z) := \partial \log \eta_k(z)/\partial z$, *and for all* $k = 1, \ldots, K$:

(a) $\mathbb{E}\epsilon_{k,t} = 0$, $\mathbb{E}\epsilon_{k,t}^2 = 1$, $\mathbb{E}\epsilon_{k,t}^{4+\delta} < \infty$, $\mathbb{E}(\epsilon_{k,t}^4) - 1 > \mathbb{E}(\epsilon_{k,t}^3)^2$, *and* $\mathbb{E}\phi_k^{4+\delta}(\epsilon_{k,t}) < \infty$ (*for some* $\delta > 0$);

(b) $\mathbb{E}\,\phi_k(\epsilon_{k,t}) = 0$, $\mathbb{E}\,\phi_k^2(\epsilon_{k,t}) > 0$, $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t} = -1$, $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^2 = 0$, *and* $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^3 = -3$;

(iii) Γ *is an open subset of* $\mathbb{R}^L$ *and for all* $(\alpha, \beta) \in \Gamma$ *we have that*

(a) $A(\alpha, \sigma)$ *is nonsingular*

(b) $(\alpha, \sigma) \to A(\alpha, \sigma)$ *is continuously differentiable*

(c) $(\alpha, \sigma) \to [D_{\alpha_l}(\alpha, \sigma)]_{k\bullet} A(\alpha, \sigma)_{\bullet j}^{-1}$ *and* $(\alpha, \sigma) \to [D_{\sigma_m}(\alpha, \sigma)]_{k\bullet} A(\alpha, \sigma)_{\bullet j}^{-1}$, *with* $D_{\alpha_l}(\alpha, \sigma) := \partial A(\alpha, \sigma)/\partial \alpha_l$ *and* $D_{\sigma_m}(\alpha, \sigma) := \partial A(\alpha, \sigma)/\partial \sigma_m$, *are locally Lipschitz continuous for all* $l = 1, \ldots, L_\alpha$, $m = 1, \ldots, L_\sigma$, *and* $j, k = 1, \ldots, K$, *where the notation* $M_{\bullet j}$ *or* $M_{j\bullet}$ *denotes the* j*th column or row of a matrix* M.

Part (i) imposes that the SVAR model (2) admits a stationary and causal solution. Part (ii) imposes that the densities of the shocks are continuously differentiable and certain moment conditions hold. Specifically, part (a) normalizes the shocks to have mean zero, variance one, and finite four + δ moments.[9] Additionally, we require the log density scores $\phi_k(x) = \partial \log \eta_k(x)/\partial x$ evaluated at the shocks to have finite $4 + \delta$ moments. Part (b) simplifies the construction of the efficient score functions. While this may at first glance appear a strong condition, in the Supplementary Material (Hoesch, Lee, and Mesters (2024, Lemma S20)), we show that simple sufficient condition is that the tails of the densities η_k converge to zero at a polynomial rate. The final part (iii) of the assumption imposes that $A(\alpha, \sigma)$ is invertible and that the parametrization chosen by the researcher is sufficiently smooth.[10] These conditions can be easily verified for specific choices for $A(\alpha, \sigma)$.

[9] $\mathbb{E}(\epsilon_{k,t}^4) - 1 \geq \mathbb{E}(\epsilon_{k,t}^3)^2$ always holds; this is known as Pearson's inequality. See, for example, result 1 in Sen (2012). Assuming that $\mathbb{E}(\epsilon_{k,t}^4) - 1 > \mathbb{E}(\epsilon_{k,t}^3)^2$ rules out (only) cases where 1, $\epsilon_{k,t}$, and $\epsilon_{k,t}^2$ are linearly dependent when considered as elements of L_2. See, for example, Theorem 7.2.10 in Horn and Johnson (2013).

[10] All of our results continue to hold without the restriction that Γ is open provided γ is an interior point of Γ.

Next, we impose a number of smoothness conditions on the densities η_k. These assumptions facilitate the estimation of the log density scores $\phi_k(z) = \nabla_z \log \eta_k(z)$, which are an important ingredient for the efficient score test discussed below.

ASSUMPTION 2. *Let* $\phi_{k,n} := \phi_k \mathbf{1}_{[\Xi^L_{k,n}, \Xi^U_{k,n}]}$, $\Delta_{k,n} := \Xi^U_{k,n} - \Xi^L_{k,n}$, *and* $\nu_n = \nu^2_{n,p}$ *with* $1 < p \leq 1 + \delta/4$ *and* $n^{-1/2(1-1/p)} = o(\nu_{n,p})$. *Suppose that for* $[\Xi^L_{k,n}, \Xi^U_{k,n}] \uparrow \tilde{\Xi} \supset \text{supp}(\eta_k)$ *and* $\delta_{k,n} \downarrow 0$ *it holds that*:

(i) $P(\epsilon_{k,t} \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]) = o(\nu^2_n)$;

(ii) *For some* $\iota > 0$, $n^{-1}\Delta^{2+2\iota}_{k,n}\delta^{-(8+2\iota)}_{k,n} = o(\nu_n)$;

(iii) η_k *is bounded* ($\|\eta_k\|_\infty < \infty$) *and differentiable, with a bounded derivative*: $\|\eta'_k\|_\infty < \infty$;

(iv) *For each* n, $\phi_{k,n}$ *is three-times continuously differentiable on* $[\Xi^L_{k,n}, \Xi^U_{k,n}]$ *and* $\|\phi^{(3)}_{k,n}\|^2_\infty \delta^6_{k,n} = o(\nu_n)$;[11]

(v) *There are* $c > 0$ *and* $N \in \mathbb{N}$ *such that for* $n \geq N$ *we have* $\inf_{s \in [\Xi^L_{k,n}, \Xi^U_{k,n}]} |\eta_k(s)| \geq c\delta_{k,n}$.

These assumptions are similar to those considered in Chen and Bickel (2006). They ensure that the log density scores can be estimated sufficiently accurately using B-spline regressions (as explained in Section 4).[12] Formally, we require that the support of the density η_k is contained with high probability in the interval $[\Xi^L_{k,n}, \Xi^U_{k,n}]$. These lower and upper points will correspond to the smallest and largest knots of the B-splines. Second, condition (ii) ensures that the number of knots does not increase too fast, relative to the sample size n. Conditions (iii) and (iv) impose that the density is sufficiently smooth, such that it can be well-fitted by B-splines. The final condition restricts the tails of the density.

3. PRELIMINARY RESULTS

In this section, we present two preliminary results for semiparametric SVAR models that we believe are useful more broadly. First, we provide a (uniform) local asymptotic normality [(U)LAN] result for the semiparametric SVAR model in (2). The primary difference with existing results is that we explicitly perturb the nonparametric part of the model, that is, the densities η_k, whereas existing (U)LAN results for VARs hold this fixed (e.g., Hallin and Saidi (2007)). This extension is necessary for deriving the form of the score test proposed in this paper and can be used in other applications. Second, we analytically derive the efficient score function for the semiparametric SVAR model; see, for example, van der Vaart (1998), Bickel et al. (1998) for general discussions on efficient score

[11] The differentiability and continuity requirements at the end-points are one-sided.

[12] These assumptions are tailored to the specific density score estimator we propose in this paper. Nevertheless, in principle, other density score estimators may be used. Inspection of the proofs reveals that any such estimator, which satisfies the conclusions of Lemma 2, can be adopted.

functions. Readers who are mainly interested in implementing the methodology of this paper can safely skip this section.

3.1 *Uniform local asymptotic normality*

Let G_k denote the law on $\mathbb{R}$ corresponding to η_k $(k = 1, \dots, K)$ and define

$$\dot{\mathscr{H}} := \prod_{k=1}^{K} \dot{\mathscr{H}}_k, \quad \dot{\mathscr{H}}_k := \left\{ h_k \in \mathcal{C}_b^1(\lambda) : \int h_k \, \mathrm{d}G_k = \int h_k \iota \, \mathrm{d}G_k = \int h_k \kappa \, \mathrm{d}G_k = 0 \right\}, \tag{5}$$

where ι is the identity function, $\kappa(z) = z^2 - 1$ (as defined above) and $\mathcal{C}_b^1(\lambda)$ denotes the class of real functions on $\mathbb{R}$ which are bounded, continuously differentiable, and have bounded derivatives. Note that $\mathbb{R}^L \times \dot{\mathscr{H}}$ is a linear subspace of $\mathbb{R}^L \times \prod_{k=1}^{K} L_2(G_k)$. Let this be normed by $\|(g, h)\| := \sqrt{\|g\|_2^2 + \sum_{k=1}^{K} \|h_k\|_{L_2(G_k)}^2}$ where $\|\cdot\|_2$ denotes the Euclidean norm.

For an arbitrary convergent sequence $(g_n, h_n) \to (g, h) \in \mathbb{R}^L \times \dot{\mathscr{H}}$, let $\theta_n := \theta_n(g_n, h_n) := (\gamma + g_n/\sqrt{n}, \eta(1 + h_n/\sqrt{n}))$. Denote by p_θ^n the density of P_θ^n with respect to λ^n and $\Lambda_{\theta_n}^n$ the (conditional) log likelihood ratio:

$$\Lambda_{\theta_n}^n := \log\left(\frac{p_{\theta_n}^n}{p_\theta^n}\right) = \sum_{t=1}^{n} \ell_{\theta_n}(Y_t, X_t) - \ell_\theta(Y_t, X_t), \tag{6}$$

where $\ell_\theta(Y_t, X_t)$ denotes the tth contribution to the conditional log likelihood for the SVAR model evaluated at θ. We note that this can be explicitly written as

$$\ell_\theta(Y_t, X_t) = \log\left|\det\big(A(\alpha, \sigma)\big)\right| + \sum_{k=1}^{K} \eta_k\big(A_{k\bullet}(\alpha, \sigma)(Y_t - BX_t)\big).$$

With this notation established, we first derive the scores for the full vector of finite-dimensional parameters $\gamma = (\alpha, \sigma, b)$. The score functions with respect to the components α_l, σ_l, and b_l are

$$\begin{aligned}\dot{\ell}_{\theta,\alpha_l}(Y_t, X_t) &= \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j}^{\alpha} \phi_k(A_{k\bullet} V_{\theta,t}) A_{j\bullet} V_{\theta,t} \\ &\quad + \sum_{k=1}^{K} \zeta_{l,k,k}^{\alpha}\big(\phi_k(A_{k\bullet} V_{\theta,t}) A_{k\bullet} V_{\theta,t} + 1\big), \end{aligned} \tag{7}$$

$$\begin{aligned}\dot{\ell}_{\theta,\sigma_l}(Y_t, X_t) &= \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j}^{\sigma} \phi_k(A_{k\bullet} V_{\theta,t}) A_{j\bullet} V_{\theta,t} \\ &\quad + \sum_{k=1}^{K} \zeta_{l,k,k}^{\sigma}\big(\phi_k(A_{k\bullet} V_{\theta,t}) A_{k\bullet} V_{\theta,t} + 1\big), \end{aligned} \tag{8}$$

$$\dot{\ell}_{\theta,b_l}(Y_t, X_t) = \sum_{k=1}^{K} \phi_k(A_{k\bullet} V_{\theta,t}) \times [-A_{k\bullet} D_{b_l} X_t], \tag{9}$$

where $V_{\theta,t} := Y_t - BX_t$, $A := A(\alpha, \sigma)$, $D_{\alpha_l}(\alpha, \sigma) := \nabla_{\alpha_l} A(\alpha, \sigma)$, $D_{\sigma_l}(\alpha, \sigma) := \nabla_{\sigma_l} A(\alpha, \sigma)$, $D_{b_l} = \nabla_{b_l} B$, $\zeta^{\alpha}_{l,k,j} := [D_{\alpha_l}(\alpha, \sigma)]_{k\bullet} A^{-1}_{\bullet j}$, $\zeta^{\sigma}_{l,k,j} := [D_{\sigma_l}(\alpha, \sigma)]_{k\bullet} A^{-1}_{\bullet j}$, and $\phi_k(z) := \nabla_z \log \eta_k(z)$.

We collect these scores in the vector

$$\dot{\ell}_\theta(Y_t, X_t) := \big(\big(\dot{\ell}_{\theta,\alpha_l}(Y_t, X_t)\big)_{l=1}^{L_\alpha}, \big(\dot{\ell}_{\theta,\sigma_l}(Y_t, X_t)\big)_{l=1}^{L_\sigma}, \big(\dot{\ell}_{\theta,b_l}(Y_t, X_t)\big)_{l=1}^{L_b} \big)'.$$

Under Assumption 1, we have the following ULAN result.[13]

PROPOSITION 1 (ULAN). *Suppose that Assumption 1 holds. Then as $n \to \infty$,*

$$\Lambda^n_{\theta_n}(Y^n) = \mathsf{g}_n(Y^n) - \frac{1}{2}\mathbb{E}_\theta\big[\mathsf{g}_n(Y^n)^2\big] + o_{P^n_\theta}(1), \tag{10}$$

where $\mathbb{E}_\theta$ indicates that the expectation is taken under P^n_θ and

$$\mathsf{g}_n(Y^n) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[g' \dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^{K} h_k(A_{k\bullet} V_{\theta,t}) \right],$$

with $A = A(\alpha, \sigma)$. Moreover, under P^n_θ,

$$\mathsf{g}_n(Y^n) \rightsquigarrow \mathcal{N}\big(0, \Psi_\theta(g, h)\big), \qquad \Psi_\theta(g, h) := \lim_{n\to\infty} \mathbb{E}_\theta\big[\mathsf{g}_n(Y^n)^2\big].$$

The corollary below follows from Le Cam's first lemma (e.g., van der Vaart (1998, Example 6.5)).

COROLLARY 1. *If Assumption 1 holds, then the sequences $(P^n_{\theta_n})_{n\in\mathbb{N}}$ and $(P^n_\theta)_{n\in\mathbb{N}}$ are mutually contiguous.*

The importance of this result is that the semiparametric SVAR model can be locally asymptotically approximated by a Gaussian shift experiment. This local approximation can be exploited to derive the form of the score test below as well as its limiting distribution under local alternatives, but can be more broadly used for other inference problems, such as building estimators.

3.2 *Efficient score function*

One of the key ingredients in our framework is the efficient score function for the parameter of interest, α. Loosely speaking, this is defined as the projection of the score function for α on the orthogonal complement (in L_2) of the space spanned by the score functions

[13]The proof of LAN is based on verifying the conditions of Lemma 1 in Swensen (1985). ULAN then follows by combining this with an asymptotic equicontinuity condition on $(g, h) \mapsto P^n_{\theta_n(g,h)}$.

for the nuisance parameters (β, η) (e.g., Bickel et al. (1998), van der Vaart (2002), Newey (1990), Choi, Hall, and Schick (1996)).

In the case of interest here, where the nuisance parameter contains both finite- (β) and infinite-dimensional (η) components, the efficient score function can be calculated in two steps: (1) compute the projection of the score for $\gamma = (\alpha, \beta)$ on the orthocomplement of the space spanned by the score functions for η, and (2) partition the resulting object into the components corresponding to α and β and project the former onto the orthocomplement of the latter.

We proceed according to this two-step procedure and now establish the form of the first projection.

LEMMA 1. *Given Assumption 1, the efficient score function for γ in the semiparametric SVAR model $\mathcal{P}_\Theta^n$ at any $\theta = (\gamma, \eta)$ with $\gamma = (\alpha, \beta)$, $\alpha \in \mathcal{A}$, $\beta = (\sigma, b) \in \mathcal{B}$, and $\eta \in \mathcal{H}$ is given by $\tilde{\ell}_{n,\theta}(Y^n) = \sum_{t=1}^n \tilde{\ell}_\theta(Y_t, X_t)$, where*

$$\tilde{\ell}_\theta(Y_t, X_t) = \big((\tilde{\ell}_{\theta,\alpha_l}(Y_t, X_t))_{l=1}^{L_\alpha}, (\tilde{\ell}_{\theta,\sigma_l}(Y_t, X_t))_{l=1}^{L_\sigma}, (\tilde{\ell}_{\theta,b_l}(Y_t, X_t))_{l=1}^{L_b}\big)'$$

with components

$$\begin{aligned}
\tilde{\ell}_{\theta,\alpha_l}(Y_t, X_t) &= \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta_{l,k,j}^\alpha \phi_k(A_{k\bullet} V_{\theta,t}) A_{j\bullet} V_{\theta,t} \\
&\quad + \sum_{k=1}^K \zeta_{l,k,k}^\alpha \big[\tau_{k,1} A_{k\bullet} V_{\theta,t} + \tau_{k,2} \kappa(A_{k\bullet} V_{\theta,t})\big], \\
\tilde{\ell}_{\theta,\sigma_l}(Y_t, X_t) &= \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta_{l,k,j}^\sigma \phi_k(A_{k\bullet} V_{\theta,t}) A_{j\bullet} V_{\theta,t} \\
&\quad + \sum_{k=1}^K \zeta_{l,k,k}^\sigma \big[\tau_{k,1} A_{k\bullet} V_{\theta,t} + \tau_{k,2} \kappa(A_{k\bullet} V_{\theta,t})\big], \\
\tilde{\ell}_{\theta,b_l}(Y_t, X_t) &= \sum_{k=1}^K -A_{k\bullet} D_{b_l} \big[(X_t - \mu)\phi_k(A_{k\bullet} V_{\theta,t}) - \mu\big(\varsigma_{k,1} A_{k\bullet} V_{\theta,t} + \varsigma_{k,2} \kappa(A_{k\bullet} V_{\theta,t})\big)\big],
\end{aligned}$$

where $V_{\theta,t} = Y_t - BX_t$, $\zeta_{l,k,j}^\alpha := [D_{\alpha_l}(\alpha, \sigma)]_{k\bullet} A_{\bullet j}^{-1}$ with $D_{\alpha_l}(\alpha, \sigma) := \partial A(\alpha, \sigma)/\partial \alpha_l$, $\zeta_{l,k,j}^\sigma := [D_{\sigma_l}(\alpha, \sigma)]_{k\bullet} A_{\bullet j}^{-1}$ with $D_{\sigma_l}(\alpha, \sigma) := \partial A(\alpha, \sigma)/\partial \sigma_l$, $D_{b_l} := \partial B/\partial b_l$, $\mu := (1, \mathrm{vec}(\iota_p \otimes (I_K - B_1 - \cdots - B_p)^{-1} c)')'$, and $\tau_k := (\tau_{1,k}, \tau_{2,k})'$ and $\varsigma_k := (\varsigma_{1,k}, \varsigma_{2,k})'$ are defined as

$$\tau_k := M_k^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix},$$

$$\varsigma_k := M_k^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{where } M_k := \begin{pmatrix} 1 & \mathbb{E}_\theta(A_{k\bullet} V_{\theta,t})^3 \\ \mathbb{E}_\theta(A_{k\bullet} V_{\theta,t})^3 & \mathbb{E}_\theta(A_{k\bullet} V_{\theta,t})^4 - 1 \end{pmatrix}.$$

The derivation of the efficient scores $\tilde{\ell}_\theta(Y_t, X_t)$ follows along the same lines as in Amari and Cardoso (1997), Chen and Bickel (2006), Lee and Mesters (2024a). The dependence on η comes through (a) the log density scores $\phi_k(z) = \nabla_z \log \eta_k(z)$, for $k = 1, \ldots, K$ and (b) the third- and fourth-order moments of ϵ_k in M_k.

For future reference, we partition

$$\tilde{\ell}_\theta(Y_t, X_t) = \begin{pmatrix} \tilde{\ell}_{\theta,\alpha}(Y_t, X_t) \\ \tilde{\ell}_{\theta,\beta}(Y_t, X_t) \end{pmatrix},$$

where $\tilde{\ell}_{\theta,\alpha}(Y_t, X_t) = (\tilde{\ell}_{\theta,\alpha_l}(Y_t, X_t))_{l=1}^{L_\alpha}$ and $\tilde{\ell}_{\theta,\beta}(Y_t, X_t) = ((\tilde{\ell}_{\theta,\sigma_l}(Y_t, X_t))_{l=1}^{L_\sigma}, (\tilde{\ell}_{\theta,b_l}(Y_t, X_t))_{l=1}^{L_b})'$.

Based on the efficient scores, we define the efficient information matrix for γ by

$$\tilde{I}_{n,\theta} := \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\,\tilde{\ell}_\theta(Y_t, X_t)\tilde{\ell}'_\theta(Y_t, X_t) \quad \text{with partitioning } \tilde{I}_{n,\theta} = \begin{pmatrix} \tilde{I}_{n,\theta,\alpha\alpha} & \tilde{I}_{n,\theta,\alpha\beta} \\ \tilde{I}_{n,\theta,\beta\alpha} & \tilde{I}_{n,\theta,\beta\beta} \end{pmatrix}. \tag{11}$$

With Lemma 1 and the efficient information matrix in place, we can compute the efficient score function for α with respect to β and η. In particular, this score can be computed by the second projection (e.g., Bickel et al. (1998, p. 74))

$$\tilde{\kappa}_{n,\theta}(Y_t, X_t) := \tilde{\ell}_{\theta,\alpha}(Y_t, X_t) - \tilde{I}_{n,\theta,\alpha\beta}\tilde{I}^{-1}_{n,\theta,\beta\beta}\tilde{\ell}_{\theta,\beta}(Y_t, X_t), \tag{12}$$

as long as $\tilde{I}_{\theta,\beta\beta}$ is positive definite. The corresponding efficient information matrix is given by

$$\tilde{\mathcal{I}}_{n,\theta} := \tilde{I}_{n,\theta,\alpha\alpha} - \tilde{I}_{n,\theta,\alpha\beta}\tilde{I}^{-1}_{n,\theta,\beta\beta}\tilde{I}_{n,\theta,\beta\alpha}. \tag{13}$$

We note that the efficient score function $\tilde{\kappa}_\theta(Y_t, X_t)$ and the efficient information matrix $\tilde{\mathcal{I}}_{n,\theta}$ can be evaluated at any parameters $\theta = (\alpha, \beta, \eta)$ and variables (Y_t, X_t).

Building tests or estimators based on the efficient score function is attractive as efficiency results are well established; see Choi, Hall, and Schick (1996), Bickel et al. (1998), and van der Vaart (2002). A crucial difference in our setting is that the efficient information matrix might be singular. For instance, if more than one component of ϵ_t follows an exact Gaussian distribution, $\tilde{\mathcal{I}}_{n,\theta}$ is singular; see Lemma S15 in Lee and Mesters (2024b). The singularity plays an important role in the construction of the semiparametric score statistic below.

4. Inference for potentially nonidentified parameters

In this section, we consider conducting inference on α without assuming that α is locally identified. Specifically, and in contrast to the existing literature, we do not assume that sufficiently many components of ϵ_t have a non-Gaussian distribution. Only Assumptions 1 and 2 are imposed, under which α may not be (locally) identified.

Our approach is based on testing hypotheses of the form

$$H_0 : \alpha = \alpha_0, \beta \in \mathcal{B}, \eta \in \mathcal{H} \quad \text{against} \quad H_1 : \alpha \neq \alpha_0, \beta \in \mathcal{B}, \eta \in \mathcal{H}. \tag{14}$$

The main idea is to consider test statistics whose computation does not require evaluation under the alternative H_1, thus avoiding the need to consistently estimate α. Clearly, based on the trinity of classical tests, the score test is the only viable candidate and we will proceed by constructing score tests in the spirit of Neyman–Rao, but adapted for the semiparametric setting (e.g., Choi, Hall, and Schick (1996)). Such test statistics can then be inverted to yield a confidence region for α with correct coverage. This confidence region then forms the basis for constructing confidence intervals for structural functions as we show in the next section.

In our setting, we rely on the efficient score functions for the SVAR model to construct test statistics. The functional form of the efficient scores $\tilde{\ell}_\theta(y_t, x_t)$ was analytically derived in Lemma 1. These scores can be estimated by replacing the population quantities by sample equivalents. We have

$$\hat{\ell}_\gamma(Y_t, X_t) = \big((\hat{\ell}_{\gamma,\alpha_l}(Y_t, X_t))_{l=1}^{L_\alpha}, (\hat{\ell}_{\gamma,\sigma_l}(Y_t, X_t))_{l=1}^{L_\sigma}, (\hat{\ell}_{\gamma,b_l}(Y_t, X_t))_{l=1}^{L_b}\big)' \tag{15}$$

with components

$$\begin{aligned}
\hat{\ell}_{\gamma,\alpha_l}(Y_t, X_t) &= \sum_{k=1}^{K}\sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j}\hat{\phi}_{k,n}(A_{k\bullet}V_{\gamma,t})A_{j\bullet}V_{\gamma,t} \\
&\quad + \sum_{k=1}^{K} \zeta^{\alpha}_{l,k,k}\big[\hat{\tau}_{k,1}A_{k\bullet}V_{\gamma,t} + \hat{\tau}_{k,2}\kappa(A_{k\bullet}V_{\gamma,t})\big], \\
\hat{\ell}_{\gamma,\sigma_l}(Y_t, X_t) &= \sum_{k=1}^{K}\sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{l,k,j}\hat{\phi}_{k,n}(A_{k\bullet}V_{\gamma,t})A_{j\bullet}V_{\gamma,t} \\
&\quad + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k}\big[\hat{\tau}_{k,1}A_{k\bullet}V_{\gamma,t} + \hat{\tau}_{k,2}\kappa(A_{k\bullet}V_{\gamma,t})\big], \\
\hat{\ell}_{\gamma,b_l}(Y_t, X_t) &= \sum_{k=1}^{K} -A_{k\bullet}D_{b_l}\big[(X_t - \bar{X}_n)\hat{\phi}_{k,n}(A_{k\bullet}V_{\gamma,t}) \\
&\quad - \bar{X}_n\big(\hat{\varsigma}_{k,1}A_{k\bullet}V_{\gamma,t} + \hat{\varsigma}_{k,2}\kappa(A_{k\bullet}V_{\gamma,t})\big)\big],
\end{aligned}$$

where $V_{\gamma,t} = Y_t - BX_t$ and $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$.[14] The estimates for the τ_k's and ς_k's are obtained by replacing the population moments defined in Lemma 1 by their sample counterparts: $\hat{\tau}_k = \hat{M}_k(0, -2)'$ and $\hat{\varsigma}_k = \hat{M}_k(1, 0)'$, where

$$\hat{M}_k := \begin{pmatrix} 1 & \frac{1}{n}\sum_{t=1}^{n}(A_{k\bullet}V_{\gamma,t})^3 \\ \frac{1}{n}\sum_{t=1}^{n}(A_{k\bullet}V_{\gamma,t})^3 & \frac{1}{n}\sum_{t=1}^{n}(A_{k\bullet}V_{\gamma,t})^4 - 1 \end{pmatrix}. \tag{16}$$

[14]Note that the components are now indexed by γ as the score estimates no longer depend on η, recalling that $\theta = (\gamma, \eta)$.

Finally, the estimates of $\hat{\ell}_\gamma(Y_t, X_t)$ depend on $\hat{\phi}_{k,n}(\cdot)$, which is the estimate for the log density scores $\phi_k(z) = \nabla_z \log \eta_k(z)$. In practice, we estimate these density scores using B-splines following the methodology of Jin (1992) and Chen and Bickel (2006). To set this up, let $b_{k,n} = (b_{k,n,1}, \ldots, b_{k,n,B_{k,n}})'$ be a collection of $B_{k,n}$ cubic B-splines and let $c_{k,n} = (c_{k,n,1}, \ldots, c_{k,n,B_{k,n}})'$ be their derivatives: $c_{k,n,i}(x) := \frac{\mathrm{d}b_{k,n,i}(x)}{\mathrm{d}x}$ for each $i = 1, \ldots, B_{k,n}$. The knots of the splines, $\xi_{k,n} = (\xi_{k,n,i})_{i=1}^{K_{k,n}}$ are taken as equally spaced in $[\Xi_{k,n}^L, \Xi_{k,n}^U]$. In practice, we take these points as the 95th and 5th percentile of the samples $\{A_{k\bullet}V_t\}_{i=1}^n$ adjusted by $\log(\log(n))$, where $A = A(\alpha, \sigma)$ and $V_t = Y_t - BX_t$ for a given parameter choice $\gamma = (\alpha, \beta)$.[15]

With this our estimate for the log density score, ϕ_k is given by

$$\hat{\phi}_{k,n}(z) := \hat{\psi}_{k,n}' b_{k,n}(z), \tag{17}$$

where

$$\hat{\psi}_{k,n} := -\left[\frac{1}{n}\sum_{t=1}^n b_{k,n}(A_{k\bullet}V_{\gamma,t}) b_{k,n}(A_{k\bullet}V_{\gamma,t})'\right]^{-1} \frac{1}{n}\sum_{t=1}^n c_{k,n}(A_{k\bullet}V_{\gamma,t}). \tag{18}$$

This shows that computing the log density score estimate (17) only requires computing the B-spline regression coefficients $\hat{\psi}_{k,n}$ in (18). The working paper version of this paper (i.e., Hoesch, Lee, and Mesters (2023)) provides the exact expressions for the B-splines and more discussion.

Having defined all the components of the efficient score estimates, we may estimate the efficient information matrix for γ by

$$\hat{I}_{n,\gamma} = \frac{1}{n}\sum_{t=1}^n \hat{\ell}_\gamma(Y_t, X_t)\hat{\ell}_\gamma(Y_t, X_t)'. \tag{19}$$

With the estimates for the efficient scores and information for γ, we can estimate the efficient score and information for α. This amounts to replacing the population score $\tilde{\kappa}_{n,\theta}(Y_t, X_t)$ and information $\tilde{\mathcal{I}}_{n,\theta}$ in (12) and (13) by their sample counterparts. We have that

$$\hat{\kappa}_{n,\gamma}(Y_t, X_t) = \hat{\ell}_{\gamma,\alpha}(Y_t, X_t) - \hat{I}_{n,\gamma,\alpha\beta}\hat{I}_{n,\gamma,\beta\beta}^{-1}\hat{\ell}_{\gamma,\beta}(Y_t, X_t) \tag{20}$$

and

$$\hat{\mathcal{I}}_{n,\gamma} = \hat{I}_{n,\gamma,\alpha\alpha} - \hat{I}_{n,\gamma,\alpha\beta}\hat{I}_{n,\gamma,\beta\beta}^{-1}\hat{I}_{n,\gamma,\beta\alpha}. \tag{21}$$

Since the information matrix may be singular, we need to make an adjustment. Specifically, given the truncation rate ν_n defined in Assumption 2, we define a truncated eigenvalue version of the information matrix estimate as

$$\hat{\mathcal{I}}_{n,\gamma}^t = \hat{U}_n \hat{\Lambda}_n(\nu_n^{1/2})\hat{U}_n', \tag{22}$$

[15] In the simulation study below, we fix the number of B-splines $B_{k,n} = 7$, and in the working paper version (Hoesch, Lee, and Mesters (2023)) we also investigate a data driven selection procedure.

where $\hat{\Lambda}_n(\nu_n^{1/2})$ is a diagonal matrix with the $\nu_n^{1/2}$-truncated eigenvalues of $\hat{\mathcal{I}}_{n,\gamma}$ on the main diagonal and $\hat{U}_n$ is the matrix of corresponding orthonormal eigenvectors. To be specific, let $\{\hat{\lambda}_{n,i}\}_{i=1}^{L}$ denote the nonincreasing eigenvalues of $\hat{\mathcal{I}}_{n,\gamma}$, then the (i,i)th element of $\hat{\Lambda}_n(\nu_n)$ is given by $\hat{\lambda}_{n,i}\mathbf{1}(\hat{\lambda}_{n,i} \geq \nu_n^{1/2})$. Similar truncation schemes are discussed for reduced rank Wald statistics in Dufour and Valery (2016).

Based on this, we define the semiparametric score statistic for the SVAR model as follows:

$$\hat{S}_{n,\gamma} := \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{\kappa}_{n,\gamma}(Y_t, X_t)\right)' \hat{\mathcal{I}}_{n,\gamma}^{t,\dagger}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{\kappa}_{n,\gamma}(Y_t, X_t)\right), \tag{23}$$

where $\hat{\mathcal{I}}_{n,\gamma}^{t,\dagger}$ is the Moore–Penrose pseudo-inverse of $\hat{\mathcal{I}}_{n,\gamma}^{t}$. We note that the test statistic can be evaluated at any $\gamma = (\alpha, \beta)$. In practice, we will set $\alpha = \alpha_0$, that is, fixing the potentially unidentified parameters under the null (14), and $\hat{\beta}_n$, some $\sqrt{n}$-consistent estimate for the finite-dimensional nuisance parameters.

For such parameter choices, the limiting distribution of $\hat{S}_{n,\gamma}$ (under the null hypothesis $\alpha = \alpha_0$) is derived in the following theorem.

THEOREM 1. *Suppose Assumptions 1 and 2 hold and that $\hat{\beta}_n$ is a $\sqrt{n}$-consistent estimator of β under P_θ^n, for $\theta = (\alpha_0, \beta, \eta)$. Define $\mathscr{S}_n = n^{-1/2}C\mathbb{Z}^{L_\beta}$ for some $C > 0$ and let $\bar{\beta}_n$ be a discretized version of $\hat{\beta}_n$, which replaces its value with the closest point in $\mathscr{S}_n$; define $\bar{\gamma}_n = (\alpha_0, \bar{\beta}_n)$. Let $r_n = \operatorname{rank}(\hat{\mathcal{I}}_{n,\bar{\gamma}_n}^{t})$ and denote by c_n the $1-a$ quantile of the $\chi^2_{r_n}$ distribution, for any $a \in (0,1)$. Then if $\theta_n := (\alpha_0, \beta + b/\sqrt{n}, \eta(1 + h/\sqrt{n}))$,*

$$\lim_{n\to\infty} P_{\theta_n}^n(\hat{S}_{n,\bar{\gamma}_n} > c_n) \leq a,$$

with inequality only if $\operatorname{rank}(\tilde{\mathcal{I}}_\theta) = 0$. *Moreover, this size control is uniform over $(b,h) \in B^\star \times H^\star \subset \mathbb{R}^{L_\beta} \times \dot{\mathscr{H}}$, where $B^\star$ and $H^\star$ are compact.*[16] *That is,*

$$\lim_{n\to\infty} \sup_{(b,h)\in B^\star\times H^\star} P_{\theta_n(b,h)}^n(\hat{S}_{n,\bar{\gamma}_n} > c_n) \leq a.$$

The theorem shows that the efficient score test (23) is locally uniformly asymptotically correctly sized when we choose the critical value c_n to correspond to the $1-a$ quantile of the chi-squared distribution with degrees of freedom equal to the rank of the truncated (estimated) efficient information matrix. Several comments are in order.

First, we do not impose which estimator $\hat{\beta}_n$ should be adopted as the theorem holds for any $\sqrt{n}$-consistent estimator. In practice, standard estimators (e.g., GMM estimators) will satisfy this condition. Moreover, given that the efficient scores for γ need to be computed anyway, it is attractive to rely on one-step efficient estimates for $\beta = (\sigma, b)$ as discussed in van der Vaart (1998, Section 5.7). These estimates are guaranteed to satisfy the requirements of the theorem and typically improve the (finite-sample) power of the test.[17]

[16] $H^\star \subset \dot{\mathscr{H}} \subset \prod_{k=1}^{K} L_2(G_k)$ and is equipped with the $\prod_{k=1}^{K} L_2(G_k)$ norm.

[17] See the simulation results of Section 7.

Second, the score statistic is evaluated at the discretized estimator $\bar{\beta}_n$, which takes the estimate $\hat{\beta}_n$ and replaces its value with the closest point in $\mathscr{S}_n = n^{-1/2}C\mathbb{Z}^{L_2}$. Note that this changes each coordinate of $\hat{\beta}_n$ by a quantity which is at most $O_p(n^{1/2})$, hence the $\sqrt{n}$-consistency is retained by discretization. Since the constant C can be chosen arbitrarily small this change has no practical relevance for the implementation of the test.[18] Discretization is a technical device due to Le Cam (1960) that allows the proof to go through under weak conditions; see Le Cam and Yang (2000, p. 125) or van der Vaart (1998, pp. 72–73) for further discussion.

Third, the practical choice for the eigenvalue truncation rate $\nu_n^{1/2}$, which theoretically needs to satisfy Assumption 2, appears to have little effect on the finite-sample results. In our simulation studies and empirical applications, we always truncate at machine precision which implies that $\hat{\mathcal{I}}_{n,\gamma}^{t,\dagger}$ is similar to $\hat{\mathcal{I}}_{n,\gamma}^{\dagger}$, the Moore–Penrose inverse of $\hat{\mathcal{I}}_{n,\gamma}$. Experimenting with different, but small, truncation rates appears to show that this choice matters little in practice.

Fourth, if $\tilde{\mathcal{I}}_\theta$ has full rank, the singularity adjusted score statistic is asymptotically equivalent to its nonsingular version that is computed with $\hat{\mathcal{I}}_{n,\bar{\gamma}_n}^{-1}$ instead of $\hat{\mathcal{I}}_{n,\bar{\gamma}_n}^{t,\dagger}$; it is well known that the former is (locally asymptotically) optimal in a number of settings.[19] Moreover, if the rank of $\tilde{\mathcal{I}}_\theta$ is positive, the singularity adjusted score statistic is (locally asymptotically) maximin optimal, as can be shown by an argument analogous to that given in Lee (2024).

Confidence set A confidence set for the parameters α can be constructed by inverting the efficient score test $\hat{S}_{n,\gamma}$ over an arbitrarily fine grid of values for α. Formally, for any $a \in (0, 1)$ we define the $1-a$ confidence set estimate for α as

$$\hat{C}_{n,1-a} := \{\alpha \in \mathcal{A} : S_{n,(\alpha,\bar{\beta}_n)} \leq c_{n,\alpha}\},$$

where $c_{n,\alpha}$ the $1-a$ quantile of the $\chi^2_{r_{n,\alpha}}$ distribution and $r_{n,\alpha} = \operatorname{rank}(\hat{\mathcal{I}}^t_{n,(\alpha,\bar{\beta}_n)})$. The following corollary establishes that the confidence set $\hat{C}_{n,1-a}$ has asymptotically correct coverage, uniformly over local alternatives in the nuisance parameters.

COROLLARY 2. *Suppose that Assumptions 1 and 2 hold. Let $\bar{\beta}_n$, $B^\star$, $H^\star$ and $\theta_n(b,h)$ be as in Theorem 1. Then*

$$\liminf_{n\to\infty} \inf_{(b,h)\in B^\star\times H^\star} P^n_{\theta_n(b,h)}(\alpha \in \hat{C}_{n,1-a}) \geq 1-a. \tag{24}$$

The confidence set $\hat{C}_{n,1-a}$ is the main building block for constructing confidence bands for the structural functions in the next section. In addition, this set may be of interest in its own right as in some models the coefficients α have a direct structural interpretation; see, for instance, the labor supply-demand model of Baumeister and Hamilton (2015) that is considered in Section 8.

[18] Indeed, in practice, we always discretize at machine precision; see Algorithm 1.

[19] This can be seen by comparison of the asymptotic local power of this test with the power bound in the appropriate limit experiment. For example, see Theorem 25.44 in van der Vaart (1998) for the one-dimensional one-sided case; optimality among unbiased tests in the two-sided case can be shown similarly.

Algorithm 1 Confidence set for α

1. Choose a set $\mathcal{A}$;
2. For each $\alpha \in \mathcal{A}$:
   (a) Obtain estimates $\hat{\beta}_n = (\hat{\sigma}_n, \hat{b}_n)$, with $b_n = \text{vec}(B_n)$, and set $\hat{V}_t = Y_t - \hat{B}_n X_t$;
   (b) For $k = 1, \ldots, K$, compute the log density scores $\hat{\phi}_k(A(\alpha_0, \hat{\sigma}_n)_{k\bullet}\hat{V}_t)$ from (17);
   (c) Compute the efficient scores $\hat{\ell}_{\hat{\gamma}_n}(Y_t, X_t)$ from (15) and the information matrix $\hat{I}_{n,\hat{\gamma}_n}$ from (19) using $\hat{\gamma}_n = (\alpha_0, \hat{\beta}_n)$;
   (d) Compute $\hat{\kappa}_{n,\hat{\gamma}_n}(Y_t, X_t)$ and $\hat{\mathcal{I}}_{n,\hat{\gamma}_n}$ from (20) and (21).
   (e) Compute the score statistic $\hat{S}_{n,\hat{\gamma}_n}$ from (23) and accept $H_0 : \alpha = \alpha_0$ if $\hat{S}_{n,\hat{\gamma}_n} \leq c_n$, where c_n is the $1-a$ quantile of the $\chi^2_{r_n}$ distribution with $r_n = \text{rank}(\hat{\mathcal{I}}^t_{n,\hat{\gamma}_n})$.
3. Collect the accepted values for α to form $\hat{C}_{n,1-a}$.

We finish this section by summarizing the practical implementation for the construction of the confidence set, which naturally includes the implementation for the efficient score test.

The algorithm highlights that the computation costs for evaluating the score test, that is, step 2, are modest. Only K B-spline regressions and a few standard computations are needed. That said, for some applications the dimension of α may be large and, therefore, the grid over which the test needs to be computed is large as well leading to substantial computational costs. To avoid this somewhat, it is attractive to parameterize $A(\alpha, \sigma)$ such that α is as low-dimensional as possible, that is, $L_\alpha = K(K-1)/2$. In addition, it is attractive to incorporate additional restrictions, for example, in our empirical work we typically use sign restrictions to a priori shrink the set $\mathcal{A}$.

5. Robust inference for smooth functions

In this section, we discuss the methodology for conducting robust inference on smooth functions of the finite-dimensional parameters $\gamma = (\alpha, \beta)$. The main functions of interest are the structural impulse response functions (sIRF), but also forecast error variance decompositions and forecast scenarios can be considered within the general framework that we develop (e.g., Kilian and Lütkepohl (2017)). The main difference with the preceding section is that we are now explicitly interested in conducting inference on functions of *both* α and β, where we recall that the parameters β are $\sqrt{n}$-consistently estimable, but α may not be consistently estimable due to a potential lack of identification.

We define the general function of interest by

$$g(\alpha, \beta) : D_g \to \mathbb{R}^{d_g}, \quad \text{with } D_g \supset \mathcal{A} \times \mathcal{B}, \tag{25}$$

where D_g is the domain of g and d_g is some positive integer. The following assumption restricts the class of functions that we consider.

ASSUMPTION 3. *$g : D_g \to \mathbb{R}^{d_g}$ is continuously differentiable with respect to β and the Jacobian matrix $J_\gamma := \nabla_{\beta'} g(\alpha, \beta)$ has full column rank on D_g.*

The differentiability condition allows for the application of the (uniform) delta-method, whereas the rank condition ensures that no further degeneracy in the asymptotic distribution occurs, apart from that caused by α potentially suffering from identification problems.

For concreteness, the next example provides the details for a vector of structural impulse response functions.

EXAMPLE 1. Consider the vector that collects all sIRF at horizon l,

$$\text{IRF}(l) = g(\alpha, \beta) := \text{vec}\big(D\mathsf{B}(b)^l D' A(\alpha, \sigma)^{-1}\big),$$

where

$$D := \begin{bmatrix} I_K & 0_{K\times K(p-1)} \end{bmatrix}, \quad \text{and} \quad \mathsf{B}(b) := \begin{bmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}.$$

In our general notation, we have $d_g = K^2$ and we note that, given Assumption 1, this function is continuously differentiable with respect to β. The Jacobian $J_\gamma \in \mathbb{R}^{K^2 \times L_\beta}$ has the form $J_\gamma = [J_{\gamma,1}, J_{\gamma,2}]$ where

$$J_{\gamma,1} := \big[\big(A(\alpha, \sigma)^{-1}\big)' \otimes I_K\big] \left\{ \sum_{j=0}^{h-1} \big[D\big(\mathsf{B}(b)'\big)^{h-1-j} \otimes \big(D\mathsf{B}(b)^j D'\big)\big] \right\},$$

$$J_{\gamma,2} := \big[I_K \otimes D\mathsf{B}(b)^h D'\big] \nabla_\sigma \text{vec}\big(A(\alpha, \sigma)^{-1}\big).$$

Similar details can be worked out for forecast error variance decompositions and other structural functions of interest.

In general, our objective is to construct a valid $1-q$ confidence set for $g(\alpha, \beta)$. Intuitively, we proceed in two steps: first we construct a valid confidence set for α using the methodology of the previous section, and second, for each included α we construct a confidence set for $g(\alpha, \hat{\beta}_n)$. The union over the latter sets provides the final set. Overall, this two-step Bonferroni approach is similar to the approach utilized by Granziera, Moon, and Schorfheide (2018) and Drautzburg and Wright (2023).

Formally, let $q_1, q_2 \in (0, 1)$ such that $q_1 + q_2 = q \in (0, 1)$. In the first step, we construct a $1 - q_1$ confidence set $\hat{C}_{n,1-q_1}$ for α using Algorithm 1. The asymptotic validity of this set was proven in Corollary 2. Second, for each $\alpha \in \hat{C}_{n,1-q_1}$ we compute $\hat{\varrho}_{\alpha,n} := g(\alpha, \hat{\beta}_n)$. The confidence set for $\hat{\nu}_{\alpha,n}$ is given by

$$\hat{C}_{n,g,\alpha,1-q_2} := \big\{\varrho : n(\hat{\varrho}_{\alpha,n} - \varrho)' \hat{V}_{n,\alpha}^{-1} (\hat{\varrho}_{\alpha,n} - \varrho) \le c_{q_2}\big\}, \tag{26}$$

where $\varrho := g(\alpha, \beta)$ and $\hat{V}_{n,\alpha} = J_{\hat{\gamma}} \hat{\Sigma}_n J_{\hat{\gamma}}'$, with $\hat{\gamma} = (\alpha, \hat{\beta}_n)$ and $\hat{\Sigma}_n$ a consistent estimate for the asymptotic variance of $\hat{\beta}_n$. The critical value c_{q_2} corresponds to the $1 - q_2$ quantile of a $\chi^2_{1-q_2}$ random variable. The following proposition establishes the conditions on the estimates $\hat{\beta}_n$ that ensure that the confidence set (26) is valid.

PROPOSITION 2. *Suppose that Assumption 3 holds. Let $\hat{\beta}_n$ and $\hat{\Sigma}_n$ be sequences of estimates and $B^\star \subset \mathcal{B}$, $H^\star \subset \mathscr{H}$ be compact. Let $\beta_n(b) := \beta + b/\sqrt{n}$. If, for any $\theta_n(b, h) := (\alpha, \beta_n(b), \eta(1 + h/\sqrt{n}))$ with $(b, h) \in B^\star \times H^\star$,*

$$\sqrt{n}\big(\hat{\beta}_n - \beta_n(b)\big) \overset{P^n_{\theta_n}}{\rightsquigarrow} \mathcal{N}(0, \Sigma), \quad \text{and,} \quad \hat{\Sigma}_n \overset{P^n_{\theta_n}}{\longrightarrow} \Sigma,$$

where Σ is positive definite, then the confidence set $\hat{C}_{n,g,\alpha}$ in (26) satisfies

$$\lim_{n\to\infty} \inf_{(b,h)\in B^\star \times H^\star} P^n_{\theta_n(b,h)}\big(g(\alpha, \beta_n(b)) \in \hat{C}_{n,g,\alpha,1-q_2}\big) = 1 - q_2. \tag{27}$$

The proposition formally establishes that if $\hat{\beta}_n$ is asymptotically normal along the local sequences $\theta_n(b, h)$, then the confidence set $\hat{C}_{n,g,\alpha}$ is valid. The proof of this proposition is a straightforward application of the uniform delta method.

The condition imposed on the estimator $\hat{\beta}_n$ is satisfied by most typical estimators (e.g., GMM estimators) under appropriate regularity conditions. Additionally, it can always be ensured (under Assumption 1) by taking $\hat{\beta}_n$ as a one-step efficient estimator based on any initial $\sqrt{n}$—consistent estimator (cf. Section 6).

The final confidence set for $g(\alpha, \beta)$, that is, $\hat{C}_{n,g}$, is formed by taking the union of the sets $\hat{C}_{n,g,\alpha,1-q_2}$ over $\alpha \in \hat{C}_{n,1-q_1}$. Formally, we consider

$$\hat{C}_{n,g} := \bigcup_{\alpha \in \hat{C}_{n,1-q_1}} \hat{C}_{n,g,\alpha,1-q_2}. \tag{28}$$

The confidence set $\hat{C}_{n,g}$ is a valid $1 - q$ confidence set as we formally establish in the following corollary.

COROLLARY 3. *Let $\beta_n(b)$, $\theta_n(b, h)$, and $B^\star$, $H^\star$ be as in Proposition 2. If $\hat{C}_{n,1-q_1}$ satisfies (24) and $\hat{C}_{n,g,\alpha,1-q_2}$ satisfies (27), then*

$$\liminf_{n\to\infty} \inf_{(b,h)\in B^\star \times H^\star} P^n_{\theta_n(b,h)}\big(g(\alpha, \beta_n(b)) \in \hat{C}_{n,g}\big) \geq 1 - q.$$

This corollary requires only the conclusions of Corollary 2 and Proposition 2.[20] For convenience, we summarize the practical implementation in the following algorithm.

As is demonstrated in the subsequent section, for structural impulse responses this approach often provides confidence sets with shorter average length when compared to alternative robust confidence set constructions proposed in the literature.

[20]These are proven under Assumptions 1 and 2, which we reiterate, do not impose that the structural shocks have non-Gaussian distributions.

Algorithm 2 Robust confidence sets for smooth functions

1. Obtain the confidence set $\hat{C}_{n,1-q_1}$ for α using Algorithm 1;
2. For each $\alpha \in \hat{C}_{n,1-q_1}$,
   (a) Estimate $\hat{\beta}_n$ and $\hat{\Sigma}_n$;
   (b) Compute $\hat{V}_{n,\alpha} = J_{\hat{\gamma}} \hat{\Sigma} J_{\hat{\gamma}}'$ with $J_{\hat{\gamma}}$ and $\hat{\gamma} = (\alpha, \hat{\beta}_n)$
   (c) Construct the confidence set $\hat{C}_{n,g,\alpha,1-q_2}$ as in (26);
3. Construct $\hat{C}_{n,g}$ from (28).

The structure of Algorithm 2 implies that different parametrizations for $A(\alpha, \sigma)$ can lead to different confidence sets for the structural functions. For example, suppose that $K = 2$: we could choose $A(\alpha, \sigma) = \Sigma^{1/2}(\sigma)R(\alpha)$ such that α is a scalar, or we could set $\alpha = (\alpha_1, \alpha_2)$ as the off-diagonal elements of $A(\alpha, \sigma)$ and let $\sigma = (\sigma_1, \sigma_2)$ capture the diagonal elements. The stated results hold for both options, but which approach results in the smallest confidence sets for a given structural function depends on the true data generating process. In practice, unless the researcher is interested in jointly testing specific entries of A, we recommend choosing α as small as possible, this reduces the computational burden of searching over the set $\mathcal{A}$ in Algorithm 1 and, therefore, immediately reduces the computational cost of Algorithm 2.

6. Point estimation under strong identification

While the main emphasis of this paper is on providing robust confidence sets for (functions of) possibly weakly identified parameters in non-Gaussian SVAR models, the results from Section 3 can also be exploited to construct point estimates for the finite-dimensional parameters $\gamma = (\alpha, \sigma, b)$. Under an additional strong identification assumption, for example, that the densities of the errors are non-Gaussian, such estimates have desirable efficiency properties as we document in this section.[21]

Assumption 4. *The limiting efficient information matrix for γ, $\tilde{I}_\theta = \lim_{n\to\infty} \tilde{I}_{n,\theta}$ is non-singular, where $\tilde{I}_{n,\theta}$ is as in* (11).

A necessary underlying condition for this assumption is that at most one of the structural shocks can follow a Gaussian distribution (e.g., Comon (1994)).[22] Under this assumption the literature has developed a variety of $\sqrt{n}$—consistent estimators for this case; see the references cited in the Introduction. Based on any of such estimators, we

[21]These efficiency properties transfer to smooth functions of γ (e.g., IRFs) in the usual way (cf. Section 25.7 in van der Vaart (1998)).

[22]We note that primitive sufficient conditions depend also on the specific parametrization that is chosen for $A(\alpha, \sigma)$.

define the one-step efficient estimator as

$$\hat{\gamma}_n = \bar{\gamma}_n + \hat{I}_{n,\bar{\gamma}_n}^{-1} \bar{\ell}_{n,\bar{\gamma}_n}, \quad \text{where } \bar{\ell}_{n,\bar{\gamma}_n} = \frac{1}{n}\sum_{t=1}^{n} \hat{\ell}_{n,\bar{\gamma}_n}(Y_t, X_t), \tag{29}$$

with $\hat{\ell}_{n,\gamma}(Y_t, X_t)$ and $\hat{I}_{n,\gamma}$ defined in (15) and (19), respectively, and $\bar{\gamma}_n$ a discretized version of any $\sqrt{n}$–consistent estimator $\tilde{\gamma}_n = (\tilde{\alpha}_n, \tilde{\beta}_n)$. We note that under Assumption 4 and the regularity conditions stated above $\hat{I}_{n,\bar{\gamma}_n}^{-1}$ exists with probability approaching one. See van der Vaart (1998) for a more elaborate discussion on one-step efficient estimators.

The following theorem summarizes the main result.

THEOREM 2. *Suppose that Assumptions 1, 2, and 4 hold. Let $\tilde{\gamma}_n$ be a $\sqrt{n}$–consistent estimator of γ under P_θ^n. Let $\bar{\gamma}_n$ be a discretized version of $\tilde{\gamma}_n$, which replaces its value with the closest point in $\mathscr{S}_n^* := n^{-1/2}C\mathbb{Z}^L$. Then*

$$\sqrt{n}(\hat{\gamma}_n - \gamma) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \tilde{I}_\theta^{-1}\tilde{\ell}_\theta(Y_t, X_t) + o_{P_\theta^n}(1) \rightsquigarrow \mathcal{N}\big(0,\ \tilde{I}_\theta^{-1}\big) \tag{30}$$

and, moreover,

$$\tilde{I}_\theta^{1/2}\sqrt{n}(\hat{\gamma}_n - \gamma) \rightsquigarrow \mathcal{N}(0,\ I).$$

The theorem reveals that the estimator $\hat{\gamma}_n$ is asymptotically efficient in the sense that it is locally regular and achieves the asymptotic semiparametric efficiency bound for locally regular estimators given by an infinite-dimensional version of the Hájek–Le Cam convolution theorem; see, for example, Theorem 3.11.2 in van der Vaart and Wellner (1996) for a version of this theorem, which applies to the present setting. The estimator in (29) can be iterated to achieve finite-sample improvements.

7. FINITE-SAMPLE PERFORMANCE

This section discusses the results from a collection of simulation studies that were designed to evaluate the size and power of the proposed inference procedures. Additional results are presented in the working paper version of this paper (e.g., Hoesch, Lee, and Mesters (2023)).

7.1 *Size of semiparametric score test*

We start by evaluating the empirical rejection frequencies of the score test $\hat{S}_{n,\hat{\gamma}_n}$ for the semiparametric SVAR model. We consider SVAR(p) specifications with $p = 1, 4, 12$ lags, $K = 2, 3$ variables, and sample sizes $T = 200, 500, 1000$. We simulate the SVAR(p) model for ten different choices for the distributions of the structural shocks $\epsilon_{k,t}$. The density functions that we consider and their abbreviated names are reported in Table 1. We normalize each $\epsilon_{k,t}$ to have mean zero and variance one by standardizing by the population mean and variance implied by the densities in Table 1.

TABLE 1. Distributions for structural shocks.

Abbreviation	Name	Definition
$\mathcal{N}(0,1)$	Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$
$t(\nu)$, $\nu = 15, 10, 5$	Student's t	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1+\frac{x^2}{\nu})^{(-\frac{\nu+1}{2})}$
SKU	Skewed Unimodal	$\frac{1}{5}\mathcal{N}(0,1)+\frac{1}{5}\mathcal{N}(\frac{1}{2},(\frac{2}{3})^2)+\frac{3}{5}\mathcal{N}(\frac{13}{12},(\frac{5}{9})^2)$
KU	Kurtotic Unimodal	$\frac{2}{3}\mathcal{N}(0,1)+\frac{1}{3}\mathcal{N}(0,(\frac{1}{10})^2)$
BM	Bimodal	$\frac{1}{2}\mathcal{N}(-1,(\frac{2}{3})^2)+\frac{1}{2}\mathcal{N}(1,(\frac{2}{3})^2)$
SPB	Separated Bimodal	$\frac{1}{2}\mathcal{N}(-\frac{3}{2},(\frac{1}{2})^2)+\frac{1}{2}\mathcal{N}(\frac{3}{2},(\frac{1}{2})^2)$
SKB	Skewed Bimodal	$\frac{3}{4}\mathcal{N}(0,1)+\frac{1}{4}\mathcal{N}(\frac{3}{2},(\frac{1}{3})^2)$
TRI	Trimodal	$\frac{9}{20}\mathcal{N}(-\frac{6}{5},(\frac{3}{5})^2)+\frac{9}{20}\mathcal{N}(\frac{6}{5},(\frac{3}{5})^2)+\frac{1}{10}\mathcal{N}(0,(\frac{1}{4})^2)$

Note: The table reports the distributions that are used in the simulation studies in Section 7 to draw the structural shocks. The mixture distributions are taken from Marron and Wand (1992); see their Table 1.

For the purpose of the simulation study, we parametrize the contemporaneous effect matrix by $A(\alpha,\sigma)^{-1} = \Sigma^{1/2}(\sigma)R(\alpha)'$ where $\Sigma^{1/2}(\sigma)$ is lower triangular and the rotation matrix $R(\alpha)$ is parametrized using the Cayley transform: $R(\alpha) = [I_K - \Gamma(\alpha)][I_K + \Gamma(\alpha)]^{-1}$, where $\Gamma(\alpha)$ is a skew-symmetric matrix with elements α.[23] The true structural parameters α_0 are fixed at randomly sampled values. Furthermore, we choose $\Sigma^{1/2}$ to be lower triangular with ones on the main diagonal and zeros elsewhere. The coefficient matrices, A_j, $j = 1,\ldots,p$ are parametrized as $A_j = \phi_j I_K$ where ϕ_j are fixed at values that ensure the SVAR is stationary. We use 400 burn-in periods to simulate data, and unless indicated differently, we use $M = 2500$ Monte Carlo replications throughout the simulations.

Table 2 reports the empirical rejection frequencies of the semiparametric score test defined in Section 4 for testing the hypothesis $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$. The test is implemented following steps 2.1–2.5 of Algorithm 1 for $\alpha = \alpha_0$ and using $B = 7$ cubic B-splines for the estimation of the log density scores. The nuisance parameters β are estimated using either OLS or using a one-step efficient estimator for β which update the OLS estimates using one Gauss–Newton iteration (van der Vaart (1998, Section 5.7)). All tests are conducted at 5% nominal size.

For the one-step efficient estimates (top panel), we find that the size of the test is generally very close to the nominal size of 5%, regardless of the dimension of the SVAR or the number of lags. Only for SVARs with a large number of parameters (high K and high p), do we see minor size distortions. Most notably for $K = 3$, $p = 12$, and $n = 200$ the empirical size of the test is often below the nominal level. We note that such settings, where the number of nuisance parameters L_β is proportional to the sample size is not covered by our theory, which imposes $L_\beta/n \to 0$.

[23]Our results are robust to using different parametrizations such as parametrizing $R(\alpha)$ by Euler angles (e.g., Rose (1957)) or directly parametrizing $A^{-1}(\alpha,\sigma) = L(\sigma) + U(\alpha)$ where $L(\sigma)$ is a lower triangular matrix and $U(\alpha)$ is an upper triangular matrix excluding the main diagonal. Hoesch, Lee, and Mesters (2023) reports the results for the latter case.

TABLE 2. Empirical rejection frequencies.

K	p	n	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
						One-Step Efficient Estimates						
2	1	200	5.4	6.3	5.8	5.4	5.8	5.3	4.6	4.3	4.7	4.8
2	1	500	6.6	6.3	6.1	6.1	5.4	5.4	4.2	4.2	5.5	4.9
2	1	1000	5.9	6.3	5.7	5.2	4.8	6.1	4.4	4.1	4.8	5.2
2	4	200	4.3	6.0	6.0	4.4	4.5	4.2	5.2	5.4	3.8	4.6
2	4	500	6.0	5.7	6.0	5.3	4.6	5.5	5.6	5.9	4.6	4.8
2	4	1000	5.8	5.8	6.6	4.7	4.8	5.3	4.3	4.0	4.8	4.4
2	12	200	4.7	4.3	5.0	4.7	4.4	3.9	4.6	6.5	3.3	5.4
2	12	500	6.2	6.9	5.0	4.9	4.2	4.5	5.2	5.8	4.7	4.6
2	12	1000	6.8	5.7	5.5	5.4	4.4	4.8	4.3	4.4	5.0	5.6
3	1	200	7.2	7.6	7.6	8.4	7.4	7.2	4.8	4.4	4.8	5.7
3	1	500	7.4	8.3	8.1	6.6	6.1	5.6	5.6	5.4	5.2	4.9
3	1	1000	7.4	7.8	6.5	5.6	5.0	5.5	4.6	4.2	5.3	4.1
3	4	200	6.2	7.6	7.5	8.3	6.0	5.9	3.6	4.1	5.5	3.6
3	4	500	9.5	7.2	8.0	7.7	6.4	6.2	5.9	5.6	4.7	4.5
3	4	1000	7.8	6.7	7.9	6.2	5.3	6.7	5.7	5.5	5.0	5.0
3	12	200	2.4	2.7	3.3	4.5	3.1	2.7	3.2	2.0	2.3	3.4
3	12	500	8.4	8.5	9.4	9.4	6.6	4.7	3.9	3.5	5.3	2.4
3	12	1000	8.4	8.0	8.5	8.1	5.8	6.6	6.7	6.3	5.3	4.6
						OLS Estimates						
2	1	200	4.0	4.4	4.8	5.7	4.4	5.0	3.8	3.5	4.0	3.8
2	1	500	4.6	5.0	5.5	6.6	5.0	5.0	3.6	4.0	4.2	4.6
2	1	1000	4.7	5.4	4.9	5.0	4.8	6.2	3.9	3.8	5.1	4.9
2	4	200	4.6	6.1	5.1	5.1	3.5	4.0	3.0	2.8	4.0	3.0
2	4	500	5.0	5.3	5.5	5.9	5.1	4.0	3.5	3.7	4.1	3.6
2	4	1000	4.8	5.4	5.4	4.8	5.0	4.6	3.9	3.3	4.0	3.5
2	12	200	8.2	6.9	8.5	9.6	4.7	5.4	4.8	4.0	5.6	3.5
2	12	500	6.7	7.8	6.4	6.7	5.4	3.4	3.5	2.7	4.4	3.6
2	12	1000	6.3	5.3	5.8	5.6	6.2	3.7	3.9	3.0	4.6	4.2
3	1	200	5.6	6.9	7.1	10.2	5.7	5.7	3.4	2.5	4.4	3.0
3	1	500	5.4	6.4	6.7	8.6	5.6	5.9	3.3	3.0	4.2	3.0
3	1	1000	5.0	5.9	5.5	6.6	4.8	6.0	3.6	3.2	4.4	3.2
3	4	200	7.7	8.9	10.1	11.5	5.7	3.7	2.4	1.3	4.7	1.9
3	4	500	6.9	6.3	7.7	9.0	5.9	3.0	2.5	1.8	3.5	2.2
3	4	1000	6.1	5.7	7.5	6.7	5.0	4.2	3.0	2.4	3.9	2.6
3	12	200	16.0	18.5	19.7	20.6	11.0	9.7	6.1	4.8	13.6	4.7
3	12	500	12.7	13.6	13.7	14.5	7.2	2.5	2.5	1.4	6.2	1.5
3	12	1000	8.5	8.8	8.7	8.4	7.0	2.5	2.9	1.4	4.3	2.0

Note: The table reports empirical rejection frequencies for the semiparametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$ in the K-variable SVAR(p) model with nominal size 5%. The nuisance parameter estimates $\hat{\beta}$ are either one-step efficient or OLS estimates. The columns correspond to the dimension K, the number of lags p, the sample size n, and the different choices for the distributions of the structural shocks, $\epsilon_{k,t}$ for $k = 1, \ldots, K$. The distributions are reported in Table 1. Rejection rates are computed based on $M = 2500$ Monte Carlo replications.

Most importantly, however, and central to the main objective of this paper, the results are similar across the different densities for $\epsilon_{k,t}$. Regardless whether the density is Gaussian, close-to-Gaussian, or far away from the Gaussian density the behavior of the test is similar, and we do not see an increase in the rejection frequency around the point of no-identification, that is, the Gaussian density.

For the test that is based on OLS estimates (bottom panel), the results are quite similar. The only difference is that for small sample sizes with K and p large the test overrejects substantially more when compared to the test based on one-step efficient estimates. The reason is that OLS estimates are considerably more noisy and biased in settings where the number of parameters is proportional to the number of observations.

7.2 *Comparison to alternative approaches*

Next, we compare the performance of the semiparametric score test to a variety of alternative methods that have been proposed in the literature based on size and power. We focus on an SVAR(1) model with $K = 2$ variables and a sample size of $T = 500$. We use the same parametrization and parameter values as described in the previous subsection to generate the data.

We distinguish between two types of alternative tests: (i) tests that do not fix α under the null (e.g., Wald and likelihood ratio type tests) and (ii) tests that fix α under the null (e.g., score type or Lagrange multiplier tests). We expect the tests in the first category to perform poorly as they are more vulnerable to identification failures.[24] In the first category, we consider three different Wald and three different likelihood ratio tests. The first test ($\mathrm{W}^{\mathrm{PML,t}}$) is a pseudo-maximum likelihood test based on the t-distribution, implemented using one (standardized) $t(7)$ density and a (standardized) $t(12)$ density for the second shock. The test is closely related to the Wald test of Gouriéroux, Monfort, and Renne (2017). We also consider the (pseudo-)likelihood ratio test ($\mathrm{LR}^{\mathrm{PML,t}}$). In addition, we consider two tests based on the work of Lanne and Luoto (2021)—the GMM Wald ($\mathrm{W}^{\mathrm{GMM,LL}}$) and likelihood ratio ($\mathrm{LR}^{\mathrm{GMM,LL}}$) tests, which are based on higher (third- and fourth-) order moment conditions. We also include the closely related moment estimator from Keweloh (2021) for a Wald ($\mathrm{W}^{\mathrm{GMM,Kew}}$) and likelihood ratio ($\mathrm{LR}^{\mathrm{GMM,Kew}}$) test.

In the second category, we consider five tests. First, we have the pseudo-maximum likelihood Lagrange multiplier test ($\mathrm{LM}^{\mathrm{PML,t}}$) that is based on work of Gouriéroux, Monfort, and Renne (2017). This test is based on the score of the pseudo-log likelihood, which we take, following Gouriéroux, Monfort, and Renne (2017), to be the Student's t with degrees of freedom fixed at $\nu = 7$ and $\nu = 12$ for the first and second shocks, respectively.[25]

[24] Simulation evidence in Lee and Mesters (2024a) has shown that tests that do not fix α under the null often show severe overrejection in static ICA models when the errors are close to Gaussian.

[25] Note that this test is not actually discussed in Gouriéroux, Monfort, and Renne (2017), but the simulations in Lee and Mesters (2024a) show that it has reliable size for ICA models. Moreover, the same idea could be implemented using mixtures of normals instead of the Student's t density (Fiorentini and Sentana (2022)).

Second, we consider the LM tests corresponding to the GMM setups of Lanne and Luoto (2021) ($\text{LM}^{\text{GMM,LL}}$) and Keweloh (2021) ($\text{LM}^{\text{GMM,Kew}}$). Lastly, we compare to the recently proposed robust GMM methods of Drautzburg and Wright (2023). We include both tests that they propose. The first is based on the S-statistic of Stock and Wright (2000), which sets the cross- (third- and fourth-order) moments to zero (S^{DW}). Second, we include their nonparametric test, which is based on Hoeffding (1948) and Blum, Kiefer, and Rosenblatt (1961) and sets all higher-order cross-moments to zero (BKR^{DW}). The S^{DW} has the benefit that it does not require a full independence assumption, whereas the BKR^{DW} test, similar to our semiparametric score test, requires full independence of the structural shocks. We implement the S^{DW} and BKR^{DW} tests using the bootstrap procedure described in Drautzburg and Wright (2023).

Size comparison Table 3 compares the size of the different testing procedures.

First as expected, the tests in group (i) —Wald and likelihood ratio type tests—tend to perform very poorly, with the simulation results demonstrating both substantial overrejection and extremely conservative performance, depending on the test and distribution pair. This leads to the strong recommendation to avoid tests that are not robust to weak deviations from Gaussian densities.

TABLE 3. Empirical rejection frequencies for alternative tests.

Test	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
				$\epsilon_{1,t} \sim \epsilon_{2,t}$						
$\hat{S}_{\text{ols}}$	5.3	5.7	6.1	5.7	4.4	5.7	4.0	3.7	4.8	3.9
$\hat{S}_{\text{onestep}}$	7.8	6.5	6.7	6.0	4.7	5.3	5.2	5.0	5.5	4.7
S^{DW}	3.9	3.8	3.7	5.8	5.3	4.3	2.6	2.9	3.6	3.7
BKR^{DW}	3.8	3.9	4.3	4.9	5.5	28.5	5.9	5.5	7.9	6.1
$\text{LM}^{\text{PML,t}}$	5.1	5.1	5.7	5.3	4.9	6.8	16.6	22.2	5.4	21.7
$\text{LM}^{\text{GMM,LL}}$	1.8	1.5	4.2	12.1	16.2	10.6	3.5	3.7	2.3	3.9
$\text{LM}^{\text{GMM,Kew}}$	1.4	1.5	4.0	15.0	15.3	6.5	4.5	4.3	1.4	4.1
$\text{LR}^{\text{PML,t}}$	34.7	13.1	7.6	3.7	1.6	1.9	100.0	100.0	14.0	100.0
$\text{LR}^{\text{GMM,LL}}$	7.4	10.5	11.7	19.2	16.1	12.6	3.9	3.4	9.6	3.8
$\text{LR}^{\text{GMM,Kew}}$	9.8	10.3	13.7	20.0	16.9	12.6	4.6	4.7	9.4	4.3
$\text{W}^{\text{PML,t}}$	16.7	10.8	11.3	8.1	6.2	5.9	37.4	41.7	12.4	38.5
$\text{W}^{\text{GMM,LL}}$	20.5	24.5	23.5	27.2	22.2	17.9	4.4	4.8	22.8	4.5
$\text{W}^{\text{GMM,Kew}}$	33.0	29.7	28.7	24.1	21.1	14.5	5.0	5.3	27.6	4.8

Note: The table reports empirical rejection frequencies for tests of the hypothesis $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$, and $\alpha_0 = 0.5594$. $\hat{S}_{\text{ols}}$ denotes the semiparametric score test using OLS estimates for β, $\hat{S}_{\text{onestep}}$ uses one-step efficient estimates. $\text{LM}^{\text{PML,t}}$, $\text{W}^{\text{PML,t}}$, and $\text{LR}^{\text{PML,t}}$ denote the pseudo-maximum likelihood tests based on Gouriéroux, Monfort, and Renne (2017), assuming t-distributed shocks. $\text{LM}^{\text{GMM,LL}}$, $\text{W}^{\text{GMM,LL}}$, and $\text{LR}^{\text{GMM,LL}}$ denote the GMM-based tests based on Lanne and Luoto (2021) with one co-kurtosis condition based on $\epsilon_{1t}^3\epsilon_{2t}$. $\text{LM}^{\text{GMM,Kew}}$, $\text{W}^{\text{GMM,Kew}}$, and $\text{LR}^{\text{GMM,Kew}}$ denote the corresponding GMM-based tests of Keweloh (2021) using both co-kurtosis conditions. Finally, S^{DW} and BKR^{DW} denote the bootstrapped GMM-based and nonparametric test of Drautzburg and Wright (2023), respectively. The columns correspond to different choices for the distributions of the structural shocks, $\epsilon_{k,t}$ for $k = 1, \ldots, K$. The distributions are reported in Table 1. The tests of Drautzburg and Wright (2023) use 500 bootstrap replications to simulate the null distribution of the test statistics. Rejection rates are computed based on $M = 1000$ Monte Carlo replications.

Overall, all tests in group (ii) perform much better, yet there are some differences that are worth noting. First, similarly as before, the rejection rates for the two efficient score tests ($\hat{S}$) are close to the nominal size of 5%, regardless of the distribution of the structural shocks (as in Table 2).

Next, consider the LM test based on Gouriéroux, Monfort, and Renne (2017) ($\text{LM}^{\text{PML,t}}$): this test is able to control size for most of the distributions, however overrejects for the BM, SPB, and TRI distributions. The LM tests based on Lanne and Luoto (2021) ($\text{LM}^{\text{GMM,LL}}$) and Keweloh (2021) ($\text{LM}^{\text{GMM,Kew}}$) display mixed performance, with milder overrejections for about a third of the distributions and strong conservativeness for other distributions. The identification robust moment tests of Drautzburg and Wright (2023) (S^{DW} and BKR^{DW}) generally perform well, with the former always controlling size correctly and the latter overrejecting only in a few cases (e.g., the kurtotic unimodal distribution). This overrejection is not due to identification failure but rather slow convergence due to the higher- order moment conditions used.

To summarize, most of the nonrobust alternative procedures lead to incorrect inference if the distribution of the structural shocks is not "sufficiently" non-Gaussian. Furthermore, the identity of the best-performing alternative procedure crucially depends on which non-Gaussian distribution generated the data. In contrast, the semiparametric score test proposed in this paper gives correct inference regardless of the distribution of the structural shocks.

Power comparison Next, we compare power among the identification robust tests. We again focus on an SVAR(1) model with $K = 2$ variables a sample size of $T = 500$.

Figure 1 reports the raw (i.e., not size-adjusted) power for the semiparametric score test using one-step nuisance parameter estimates (red solid line), the semiparametric score test using OLS nuisance parameter estimates (black solid line), the pseudo-maximum likelihood LM test (dot-dashed blue line), the Drautzburg and Wright (2023) GMM test (dotted green line), and the nonparametric Drautzburg and Wright (2023) test (dot-dashed purple line).

For the t-distributions in the first row of the figure, the best performing test is the pseudo-maximum likelihood LM test. This is not surprising as this test is based on the t—density and, therefore, is close to correctly specified. The efficient score tests show greater power than either of the other tests considered. Moreover, in the other panels, the efficient score tests are typically the most powerful tests (that also control size), with the one-step update version performing slightly better. The quality of the other three tests depends to a large extent on the underlying density. For example, the tests of Drautzburg and Wright (2023) offer very little power in the t-distribution cases, but for the other distributions their nonparametric test has power curves, which are not much below those of the efficient score test.[26]

[26]For the kurtotic unimodal distribution, the power curve of this test is higher; however, this test is substantially oversized for this density. It should also be noted that the tests of Drautzburg and Wright (2023) are substantially more computationally demanding than the efficient score based approaches, as they use a bootstrap approach to obtain the critical value. Relying on asymptotic critical values for these tests yields substantially worse performance.

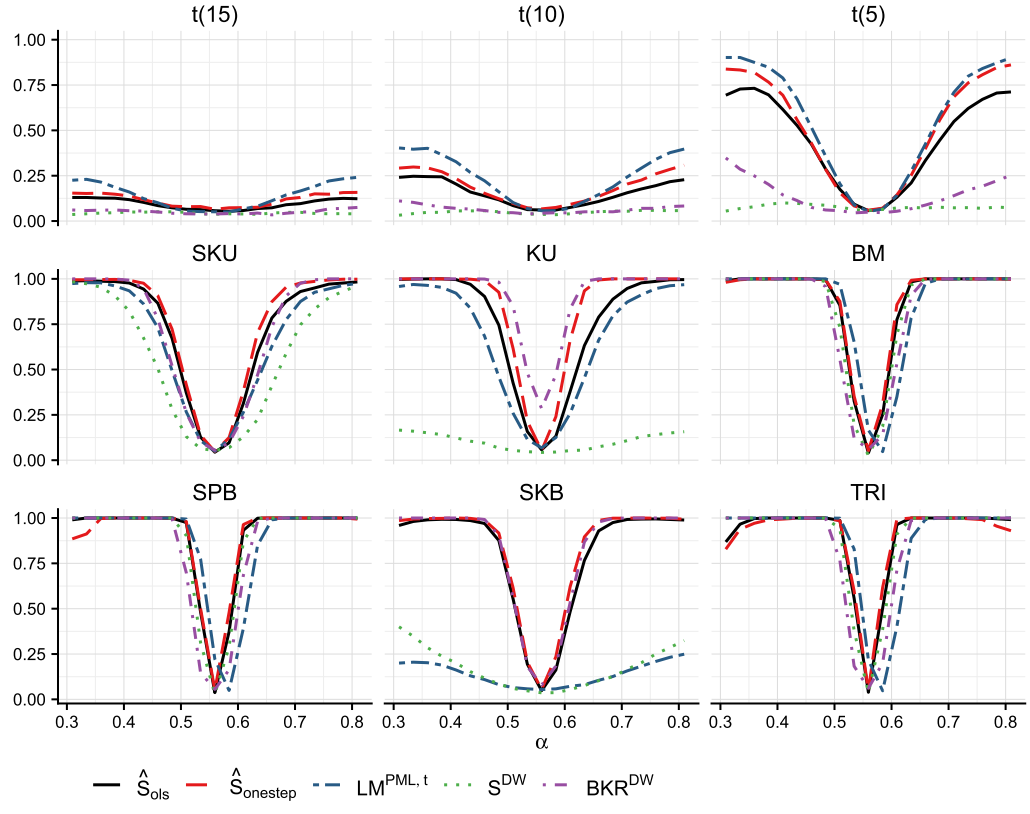

FIGURE 1. Power in the SVAR(1) model. The figure reports unadjusted empirical power curves for tests of the hypothesis $H_0 : \alpha = \alpha_0$ versus $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$. The x-axis corresponds to different alternatives for α around $\alpha_0 = 0.5594$. $\hat{S}_{ols}$ denotes the semiparametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. $LM^{PML,t}$ denotes the pseudo-maximum likelihood test based on Gouriéroux, Monfort, and Renne (2017), S^{DW} denotes the GMM-based test of Drautzburg and Wright (2023), BKR^{DW} denotes the nonparametric test of Drautzburg and Wright (2023). The tests of Drautzburg and Wright (2023) use 500 bootstrap replications to obtain critical values. Rejection frequencies are computed using $M = 1000$ Monte Carlo replications.

7.3 *Additional results*

Hoesch, Lee, and Mesters (2023) show additional results that evaluate (i) the score test under alternative parametrizations, (ii) the score test for higher dimensions, (iii) the score test with cross-validation for selecting the number of B-splines as in Chen and Bickel (2006), (iv) the confidence sets for smooth functions of the SVAR parameters as discussed in Section 5 (both coverage and confidence set length), and (v) the point estimates introduced in Section 6. The results show that the finite sample properties of the score test are invariant to the specific parametrization chosen. The cross-validation procedure leads to rejection frequencies that are generally closer to the nominal level. In higher dimensions, the performance of the test deteriorates; similar to other SVAR

studies, a bootstrap implementation of our test is likely to be preferable in such settings. The evaluation of the impulse responses shows that the two-step Bonferroni approach is conservative; but if the efficient score test, based on one-step efficient estimates, is used as the first step the coverage becomes much closer to the nominal size. Also, the efficient score approach gives the smallest length among all procedures considered and for all densities. Finally, the one-step efficient point estimates are generally more accurate when compared to nonefficient competitors, that is, their root mean-squared error is lower when compared to existing estimators.

8. Empirical studies

In this section, we discuss the results from two empirical studies: one for labor supply and demand and the other for the oil market. We investigate the consequences of replacing some of the identifying information used in previous studies with identification based on non-Gaussianity and illustrate the calculation of confidence sets based on the methodology of this paper.

8.1 *Labor supply–demand model of Baumeister and Hamilton (2015)*

We revisit the bivariate SVAR(p) model of the U.S. labor market as considered in Baumeister and Hamilton (2015). We have $Y_t = (\Delta w_t, \Delta \eta_t)'$, where Δw_t is the growth rate of real compensation per hour and $\Delta \eta_t$ is the growth rate of total U.S. employment. The SVAR model for Y_t is defined by (2) with parametrization[27]

$$A^{-1}(\alpha, \sigma) = \begin{pmatrix} -\alpha^d & 1 \\ -\alpha^s & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}.$$

It follows that here the parameter α^d is the short-run wage elasticity of demand, and α^s is the short-run wage elasticity of supply. The number of lags used is $p = 8$, the sample is from 1970:Q1 through 2014:Q2, and conventional sign restrictions are imposed on the supply and demand elasticities ($\alpha^d \leq 0$, $\alpha^s \geq 0$). These restrictions ensure that we test economically interesting permutations of the impact matrix.

Without further identifying information, any fixed point that satisfies the sign restrictions is a valid point and nothing more can be learned. To improve identification, Baumeister and Hamilton (2015) introduce carefully motivated priors on the short-run labor supply and demand elasticities, based on estimates from the micro-econometric and macroeconomic literature, as well as a long-run restriction on the effect of labor-demand shocks on employment (e.g., Shapiro and Watson (1988)). We investigate whether such additional identifying assumptions can be avoided by exploiting possible non-Gaussianity in the supply and demand shocks. For the purpose of our analysis, we consider a wide grid of potential elasticities, $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$, which covers the majority of elasticity estimates reported in the microeconometric literature, as well

[27] In Hoesch, Lee, and Mesters (2023), we provide additional results from an alternative parametrization of the model using a rotation matrix.

as findings from theoretical macroeconomic models (see the discussion in Baumeister and Hamilton (2015)). We confine our analysis to this grid, which can be regarded as an additional identification restriction.

Recently, Lanne and Luoto (2022) adopted the methodology of Lanne and Luoto (2021) to assess identification of the model using non-Gaussianity, but this approach may yield incorrect coverage when the shocks are close to Gaussian (cf. Section 7). Here, we will adopt the robust score testing approach of Sections 4 and 5 to construct confidence sets for the elasticity parameters as well as impulse responses to labor supply and labor demand shocks. Specifically, we construct confidence sets for α using Algorithm 1 and confidence bands for the impulse responses using Algorithm 2. For both algorithms, we make use of one-step efficient parameter estimates $\hat{\beta}_n$.

Before getting there, we recall that our methodology relies on the assumption that the demand and supply shocks are independent and not merely uncorrelated. Therefore, we start by testing for independent components using the permutation tests of Matteson and Tsay (2017) and Olea et al. (2022). To compute the test, we obtain an initial GMM estimate of α using the moment conditions of Keweloh (2021). For the given sample period, we obtain a p-value of 0.12 for the test of Matteson and Tsay (2017) and a p-value of 0.55 for the test of Olea et al. (2022), hence we conclude this assumption is not unreasonable and proceed with constructing confidence sets for the elasticity parameters.

Confidence sets for (α^d, α^s) Figure 2 shows the 95% and 67% joint confidence sets for labor demand (α^d) and labor supply (α^s) parameters obtained using Algorithm 1 of Sec-

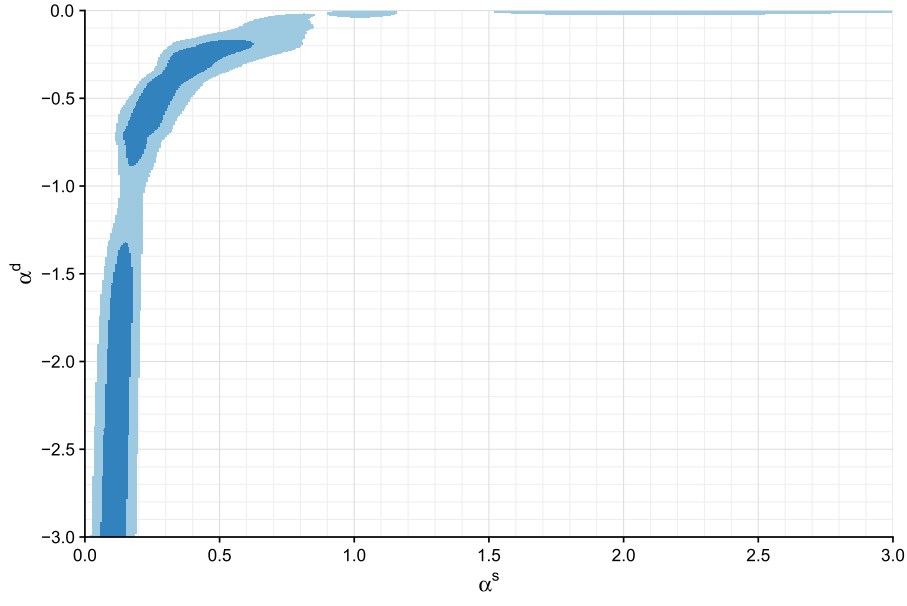

FIGURE 2. Confidence sets for labor demand and supply elasticities. 95% (light blue) and 67% (dark blue) confidence regions for labor demand and supply elasticities obtained using Algorithm 1 with 250,000 equally-spaced grid points for $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$.

tion 4. The confidence sets are constructed based on a grid of 250,000 equally spaced points spanning the elasticity ranges discussed above. The figure shows that overall, non-Gaussianity is not sufficient to pin down a precise region for the elasticities, though it does rule out parts of the parameter space, which would be accepted under Gaussianity. For sufficiently negative values of the short-run demand elasticity, the short-run supply elasticity is reasonably well identified from non-Gaussianity with confidence sets indicating that α^s lies in the 0–0.3 range for both 95% and 67% confidence level. In contrast, for values of α^d that are less negative (smaller absolute value), the confidence sets support a wide range of values for the supply elasticity, up to 0.6 at 67% confidence level and spanning almost all values in the inspected grid at 95% confidence level. Our results match the findings of Baumeister and Hamilton (2015) who report that the main posterior mass for α^s lies in the 0–0.5 range while the posterior for α^d indicates that demand elasticities between -3 and 0 are well supported by the model.

Note that the estimate of Lanne and Luoto (2022) obtained using non-Gaussianity identification ($\alpha^d = -0.317$, $\alpha^s = 0.514$) falls within our confidence set at 95% level. However, they find narrow confidence sets for the elasticity parameters while our weak-identification robust approach results in much wider confidence sets, similar to the credible sets of Baumeister and Hamilton (2015).

Confidence sets for impulse responses Figure 3 shows our identification-robust 95% and 67% confidence sets for the impulse responses to labor-demand and labor-supply shocks. Comparing the impulse response bands to the posterior credible sets reported by Baumeister and Hamilton (2015), we note that the implied impulse responses are overall very similar and show long and persistent responses to the supply and demand shocks. The main differences are that our 95% identification-robust bands support slightly negative long-run responses of the real wage and employment to a demand shock, as well as a more pronounced negative long-run response of employment to a supply shock while Baumeister and Hamilton's (2015) credible sets contain only (weakly) positive responses. Comparing our results to Lanne and Luoto (2022), we note several differences. First, Lanne and Luoto (2022) find a significant negative long-run response of the real wage to a supply shock while our confidence sets do not rule out that the long-run response is weakly positive. Second, and most important, they find a strong and significant dynamic response of both the real wage and employment to the labor demand shock, inconsistent with the tight prior variance Baumeister and Hamilton (2015) impose on the long-run response of employment to a demand shock. In contrast to their findings, both our 67% and 95% identification-robust confidence bands do not rule out that the long-run response of either variable to the demand shock is zero. This evidence suggests that the long-run restriction of Baumeister and Hamilton (2015) cannot be rejected solely on the basis of non-Gaussianity.

8.2 *Oil price model of Kilian and Murphy (2012)*

Next, we revisit the trivariate oil market SVAR(p) model of Kilian and Murphy (2012). We have $Y_t = (\Delta q_t, x_t, p_t)'$ where Δq_t is the percent change in global crude oil production,

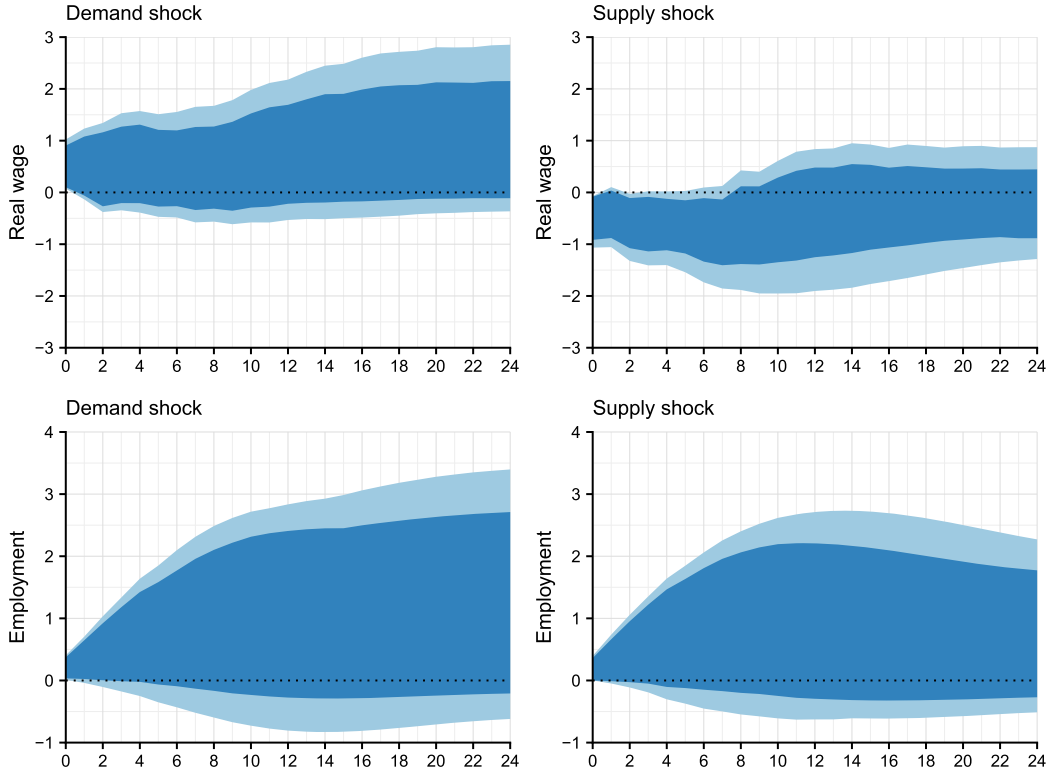

FIGURE 3. IRF confidence bands for labor demand and supply shocks. 95% (light blue) and 67% (dark blue) identification-robust confidence bands for impulse responses to labor supply and labor demand shocks, obtained using 250,000 equally-spaced grid points for $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$.

x_t is an index of real economic activity representing the global business cycle, and p_t is the log of the real price of oil. The SVAR model is parameterized as follows:

$$Y_t = c + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + A^{-1}(\alpha, \sigma)\epsilon_t,$$
$$A^{-1}(\alpha, \sigma) = \begin{bmatrix} \sigma_1 & \alpha_{qx} \cdot \sigma_5 & \alpha_{qp} \cdot \sigma_6 \\ \sigma_2 & \sigma_4 & \alpha_{xp} \\ \sigma_3 & \sigma_5 & \sigma_6 \end{bmatrix}, \tag{31}$$

where following Baumeister and Hamilton (2019) we use $p = 12$. In this model, ϵ_t includes a shock to the world production of crude oil (*"oil supply shock"*), a shock to the demand for crude oil and other industrial commodities associated with the global business cycle (*"aggregate demand shock"*), and a shock to demand for oil that is specific to the oil market (*"oil-market-specific demand shock"*). In the parametrization above, α_{qx} is the short-run (impact) demand elasticity of oil supply while α_{qp} captures the short-run (impact) price elasticity of oil supply.

The baseline model of Kilian and Murphy (2012) makes use of the following sign restrictions on the impact responses in A^{-1} to identify impulse responses:[28]

$$A^{-1}(\alpha, \sigma) = \begin{bmatrix} + & + & + \\ + & + & - \\ - & + & + \end{bmatrix}. \tag{32}$$

In addition, Kilian and Murphy (2012) impose a set of upper bounds on the short-run oil supply elasticities implied by the model to shrink the identified set for the impulse responses. Specifically, they assume that $\alpha_{qp} < 0.0258$, $\alpha_{qx} < 0.0258$ and that $\alpha_{xp} > -1.5$. These restrictions, in particular the elasticity bound on α_{qp}, have been criticized by Baumeister and Hamilton (2019) as being too tight and there is an active debate around which values for these bounds are reasonable (see Herrera and Rangaraju (2020) for an overview).

We investigate whether the bounds on the elasticities can be avoided by exploiting non-Gaussian features of the structural shocks. We base our analysis on the monthly data sample considered in Zhou (2020), which spans February 1973–August 2009. This data corresponds to the original data of Kilian and Murphy (2014), but includes the correction to the index of global economic activity discussed in Kilian (2019). We consider the robust score testing approach of Sections 4 and 5 to construct confidence sets for the elasticity parameters as well as the impulse responses to the oil supply shock, the aggregate demand shock, and the oil-market-specific demand shock. Our implementation is similar as in the previous application. We start by testing for independent components using the permutation tests of Matteson and Tsay (2017) and Olea et al. (2022). As before, we base the test on a GMM estimate of α obtained using the moment conditions of Keweloh (2021). For the given sample period, we obtain a p-value of 0.35 for the test of Matteson and Tsay (2017) and a p-value of 0.47 for the test of Olea et al. (2022), hence we conclude this assumption is not unreasonable and proceed with constructing confidence sets for the elasticity parameters.

Confidence sets for oil supply elasticities $(\alpha_{qx}, \alpha_{qp})$ Figure 4 shows the 95% and 67% joint confidence sets for the price elasticity of oil supply (α_{qp}) and the demand elasticity of oil supply (α_{qx}) obtained using Algorithm 1 of Section 4 from a grid of 500,000 points for $(\alpha_{qx}, \alpha_{qp}, \alpha_{xp}) \in (0, 0.25] \times (0, 0.1] \times [-3, 0)$ with 100 points for α_{qx} and α_{qp} each and 50 points for α_{xp}. The confidence set for $(\alpha_{qx}, \alpha_{qp})$ is obtained by projecting over all values of α_{xp} in the grid. The end points of the grid were chosen by (i) doubling the bound on α_{xp} imposed by Kilian and Murphy (2012), (ii) allowing for a large range of values for α_{qx}, and (iii) substantially relaxing the bound on the price elasticity of oil supply (α_{qp}) in Kilian and Murphy (2012) to address the critique of Baumeister and Hamilton (2019). In particular, the grid end-point of 0.1 for α_{qp} matches the largest supply elasticity bound considered in the sensitivity analysis of Baumeister and Hamilton (2019)'s model carried out in Herrera and Rangaraju (2020) and nests the relaxed supply elasticity bound

[28] Kilian and Murphy (2012) normalize the first shock to be an oil supply disruption, leading to inverted signs in the first column of A^{-1}. Following Baumeister and Hamilton (2019), we consider a positive oil supply shock.

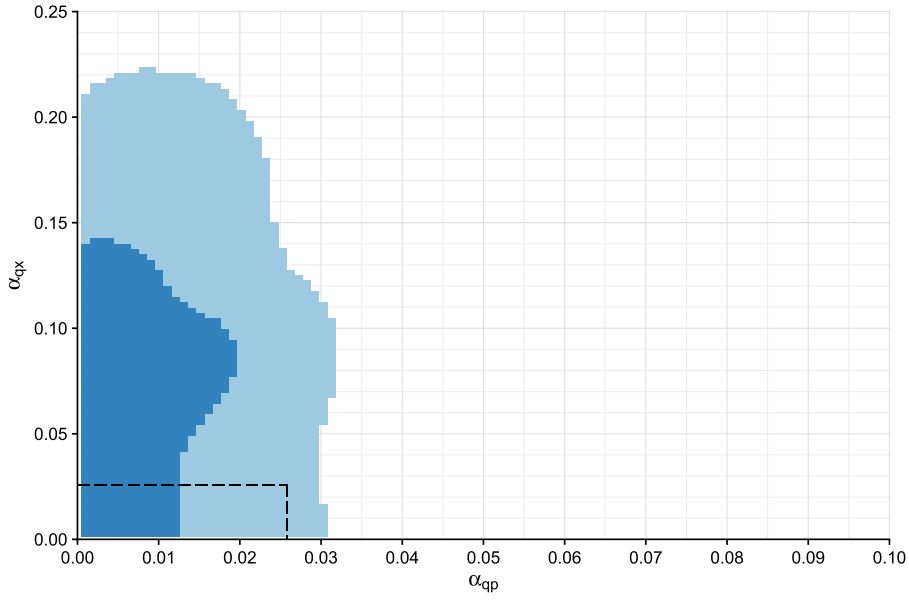

FIGURE 4. Confidence sets for $(\alpha_{qx}, \alpha_{qp})$. 95% (light blue) and 67% (dark blue) confidence regions for supply elasticities $(\alpha_{qx}, \alpha_{qp})$ obtained using Algorithm 1 using 500,000 grid points for $(\alpha_{qx}, \alpha_{qp}, \alpha_{xp}) \in (0, 0.25] \times (0, 0.1] \times [-3, 0)$ by projection across accepted values for α_{xp}. The black-dashed lines denote the original supply elasticity bounds of 0.0258 imposed by Kilian and Murphy (2012).

considered in Zhou (2020). To ensure that our robust confidence set is compatible with the sign restrictions in (32), we impose these signs in the estimation of the nuisance parameters σ.[29]

Inspecting the confidence set depicted in Figure 4, we note that non-Gaussianity significantly helps to identify the price elasticity of the oil supply, but is less able to accurately pin down the demand elasticity of oil supply. In particular, while the considered grid allows for supply elasticities up to 0.1, the bound on the price elasticity of oil supply implied by the 95% and 67% confidence set for α_{qp} falls within the relaxed bound of 0.04 considered by Zhou (2020). In addition, at the 67% level, the elasticity lies within the bound of 0.0258 originally considered in Kilian and Murphy (2012). At the 95% level, non-Gaussianity cannot rule out that α_{qp} falls outside this bound. For the demand elasticity of oil supply (α_{qx}), the confidence set spans a large range of values between zero and 0.22, depending on the value for α_{qp}.

Overall, our results suggest that non-Gaussianity is informative about the oil supply elasticities α_{qx}, α_{qp} in the model of Kilian and Murphy (2012). However, it is not able to justify the bounds considered in Kilian and Murphy (2012).

[29]Note that the set of sign restrictions on A^{-1} does not merely pin down a signed permutation of A^{-1}, but also imposes additional restrictions on the magnitudes of elasticities; see the discussion in Baumeister and Hamilton (2019, p. 1881).

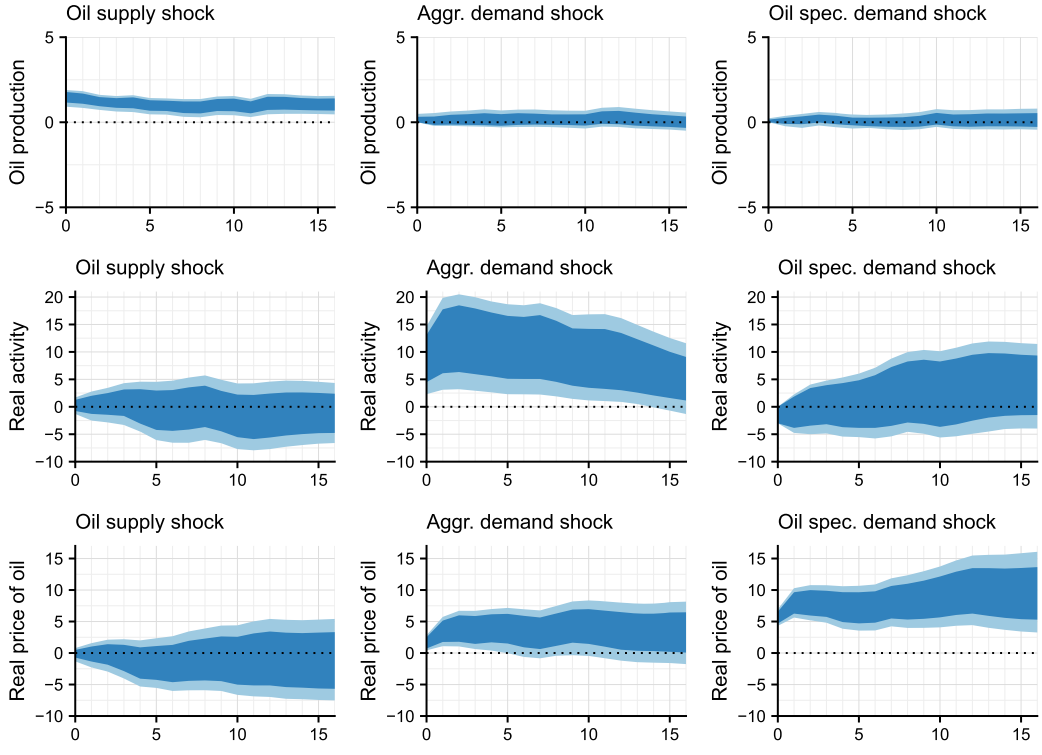

FIGURE 5. IRF confidence bands in the oil market model. 95% (light blue) and 67% (dark blue) identification-robust confidence bands for the impulse responses to oil supply, aggregate demand, and oil-specific demand shocks, obtained using 500,000 equally-spaced grid points for $(\alpha_{qx}, \alpha_{qp}, \alpha_{xp}) \in (0, 0.25] \times (0, 0.1] \times [-3, 0)$.

Confidence sets for impulse responses Finally, we turn to inspecting the 95% and 67% confidence bands for impulse responses to oil supply, aggregate demand, and oil-specific supply shocks, which are depicted in Figure 5. We note that our confidence bands overall exhibit response patterns that are similar to the results reported in Kilian and Murphy (2012) based on sign restrictions and the elasticity bound of 0.0258. However, our procedure results in substantially wider confidence bands for the responses of global real activity and the real price of oil than the ones originally reported in Kilian and Murphy (2012). In particular, while the responses of oil production are identified precisely, the responses of global real activity and of the real price of oil exhibit large uncertainty with insignificant and flat responses to the oil supply shock, significant positive hump-shaped responses to the aggregate demand shock, and mixed response patterns to the oil-specific demand shock.

9. CONCLUSION

This paper develops robust inference methods for structural vector autoregressive (SVAR) models that are identified via non-Gaussianity in the distributions of the structural shocks. We treat the SVAR model as a semiparametric model where the densities of

the structural shocks form the nonparametric part and conduct inference on the possibly weakly identified or nonidentified parameters of the SVAR, using a semiparametric score statistic. We additionally provide a two-step Bonferroni-based approach to conduct inference on smooth functions of all the finite-dimension parameters of the model.

We assess the finite-sample performance of our method in a large simulation study and find that the empirical rejection frequencies of the semiparametric score test are always close to the nominal size, regardless of the true distribution of the shocks. Moreover, the power of the test is typically higher than alternative methods that have been proposed in the literature.

Finally, we employ the proposed approach in a number of empirical studies. Overall our findings are mixed. While non-Gaussianity does provide some identifying information for the structural parameters of interest, it is unable to always pin down the parameter values or impulse responses precisely. These exercises also highlight the importance of using weak identification robust methods to assess estimation uncertainty when using non-Gaussianity for identification.

Appendix A: Proofs and additional results

Here, we prove the main results of the paper. Only the main arguments are given here, with the verification of technical details relegated to lemmas, which can be found in the Supplementary Material (Hoesch, Lee, and Mesters (2024)).

A.1 *Notation*

$x := y$ means that x is defined to be y. The Lebesgue measure on $\mathbb{R}^K$ is denoted by λ_K or λ if the dimension is clear from context. The standard basis vectors in $\mathbb{R}^K$ are $e_1, \ldots, e_K$. We make use of the empirical process notation: $Pf := \int f \, dP$, $\mathbb{P}_n f := \frac{1}{n}\sum_{i=1}^n f(Y_i)$, and $\mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f$. For any two sequence of probability measures $(Q_n)_{n\in\mathbb{N}}$ and $(P_n)_{n\in\mathbb{N}}$ (where Q_n and P_n are defined on a common measurable space for each $n \in \mathbb{N}$), $Q_n \triangleleft P_n$ indicates that $(Q_n)_{n\in\mathbb{N}}$ is contiguous with respect to $(P_n)_{n\in\mathbb{N}}$. $Q_n \triangleleft\triangleright P_n$ indicates that both $Q_n \triangleleft P_n$ and $P_n \triangleleft Q_n$ hold; see van der Vaart (1998, Section 6.2) for formal definitions. $X \perp\!\!\!\perp Y$ indicates that random vectors X and Y are independent; $X \simeq Y$ indicates that they have the same distribution. $a \lesssim b$ means that a is bounded above by Cb for some constant $C \in (0, \infty)$; the constant C may change from line to line. $\operatorname{cl} X$ means the closure of X. vec^{-1} is the inverse vec operator, that is, if $b = \operatorname{vec}(B)$ then $B = \operatorname{vec}^{-1}(b)$. If S is a subset of an inner product space $(V, \langle\cdot,\cdot\rangle)$, $S^\perp$ is its orthogonal complement, that is, $S^\perp = \{x \in V : \langle x, s\rangle = 0 \text{ for all } s \in S\}$. If $S \subset V$ is complete (hence a Hilbert space), the orthogonal projection of $x \in V$ onto S is $\Pi(x|S)$. The total variation distance between measures P and Q defined on the measurable space $(\Omega, \mathcal{F})$ is $d_{\mathrm{TV}}(P, Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)|$. $\rightsquigarrow$ denotes weak convergence.

A.2 *Density score estimation*

Lemma 2. *Suppose Assumptions 1 and 2 hold. Let $\theta_n = (\alpha_n, \beta_n, \eta) \to \theta$ be a deterministic sequence with $\sqrt{n}\|\beta_n - \beta\| = O(1)$. Then the log density score estimates $\hat{\phi}_{k,n}$ defined as in*

(17) *satisfy for* $j, k = 1, \ldots, K$, $k \neq j$,

$$\frac{1}{n}\sum_{t=1}^{n}\big[\hat{\phi}_{k,n}\big(A_{n,k\bullet}(Y_t - B_n X_t)\big) - \phi_k\big(A_{n,k\bullet}(Y_t - B_n X_t)\big)\big]W_{n,t} = o_{P^n_{\tilde{\theta}_n}}\big(n^{-1/2}\big), \tag{33}$$

where $A_n := A(\alpha_n, \beta_n)$, $B_n := B(\beta_n)$ *and* $W_{n,t}$ *are any mean-zero random variables independent from all* $A_{n,k\bullet}(Y_s - B_n X_s)$ *with* $s \geq t$ *and such that* $\sup_{n\in\mathbb{N}, 1\leq t\leq n}\mathbb{E}_{\tilde{\theta}_n} W_{n,t}^2 < \infty$. *Additionally, for* $\nu_n = \nu_{n,p}^2$ *with* $1 < p \leq 1 + \delta/4$ *and* $n^{-1/2(1-1/p)} = o(\nu_{n,p})$ *we have*

$$\frac{1}{n}\sum_{t=1}^{n}\big(\big[\hat{\phi}_{k,n}\big(A_{n,k\bullet}(Y_t - B_n X_t)\big) - \phi_k\big(A_{n,k\bullet}(Y_t - B_n X_t)\big)\big]W_{n,t}\big)^2 = o_{P^n_{\theta'_n}}(\nu_n). \tag{34}$$

where $W_{n,t}$ *are any random variables independent from all* $A_{n,k\bullet}(Y_s - B_n X_s)$ *with* $s \geq t$ *and such that* $\sup_{n\in\mathbb{N}, 1\leq t\leq n}\mathbb{E}_{\tilde{\theta}_n} W_{n,t}^2 < \infty$.

Proof. The claim follows by an argument analogous to that used to prove Lemma 4 of Lee and Mesters (2024a); see Lee and Mesters (2024b) for the proof.[30] □

A.3 *ULAN*

To establish ULAN, we establish LAN, as in Proposition 3 directly below. Following this in Proposition 4, we show that $(g, h) \mapsto P^n_{\theta_n(g,h)}$ is asymptotically equicontinuous in total variation. These properties are together equivalent to ULAN.

Proposition 3 (LAN). *Suppose that Assumption 1 holds. Then for any* $g, h \in \mathbb{R}^L \times \dot{\mathscr{H}}$ *such that* $\theta_n(g, h) = (\gamma + g/\sqrt{n}, \eta(1 + h/\sqrt{n}))$, *as* $n \to \infty$,

$$\Lambda^n_{\theta_n(g,h)}(Y^n) = \mathsf{g}_n(Y^n) - \frac{1}{2}\mathbb{E}\big[\mathsf{g}_n(Y^n)^2\big] + o_{P^n_\theta}(1), \tag{35}$$

where the expectation is taken under P^n_θ *and*

$$\mathsf{g}_n(Y^n) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[g'\dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,t})\right],$$

with $A = A(\alpha, \sigma)$. *Moreover, under* P^n_θ,

$$\mathsf{g}_n(Y^n) \rightsquigarrow \mathcal{N}\big(0, \Psi_\theta(g, h)\big), \qquad \Psi_\theta(g, h) := \lim_{n\to\infty}\mathbb{E}\big[\mathsf{g}_n(Y^n)^2\big].$$

Proof. Throughout, we work conditional on $(Y_{-p+1}, \ldots, Y_0)'$. Define $V_{\theta,t} := Y_t - BX_t$,

$$W_{n,t} := \frac{1}{2\sqrt{n}}\left[g'\dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,t})\right],$$

[30] Note that in the statement of Lemma 4 of Lee and Mesters (2024a) the object corresponding to $W_{n,t}$ here (their $Z_{n,i}$) is assumed to be mean zero in the equations corresponding to both (33) and (34). Inspection of the proof reveals that this is unnecessary for the equation corresponding to (34).

and $\mathcal{F}_{n,t} := \sigma(Y_t, X_t)$ and note that $(W_{n,t}, \mathcal{F}_{n,t})_{n\in\mathbb{N},\ 1\le t\le n}$ forms an adapted stochastic process. By Assumption 1 part 2,

$$\mathbb{E}[W_{n,t}|\mathcal{F}_{n,t-1}] = \frac{1}{2\sqrt{n}}\left[g'\,\mathbb{E}\big[\dot{\ell}_\theta(Y_t, X_t)|\mathcal{F}_{n,t-1}\big] + \sum_{k=1}^{K}\mathbb{E}\big[h_k(A_{k\bullet}V_{\theta,t})|\mathcal{F}_{n,t-1}\big]\right] = 0, \quad (36)$$

almost surely, where the expectation is taken under P_θ^n.

Next, define $U_{n,t} := (u_{n,t}/u_{n,t-1})^{1/2} - 1$ where $u_{n,0} = 1$ and else

$$u_{n,j} := \left(\frac{|A_n|}{|A|}\right)^j \times \prod_{t=1}^{j}\prod_{k=1}^{K}\frac{\eta_k(A_{n,k\bullet}V_{\theta_n,t})}{\eta_k(A_{k\bullet}V_{\theta,t})}\left(1 + \frac{h_k(A_{n,k\bullet}V_{\theta_n,t})}{\sqrt{n}}\right), \quad (37)$$

with $A := A(\alpha, \sigma)$ and $A_n := A(\alpha + g_\alpha/\sqrt{n}, \sigma + g_\sigma/\sqrt{n})$. That is,

$$U_{n,t} := \left[\left(\frac{|A_n|}{|A|}\right) \times \prod_{k=1}^{K}\frac{\eta_k(A_{n,k\bullet}V_{\theta_n,t})}{\eta_k(A_{k\bullet}V_{\theta,t})}\left(1 + \frac{h_k(A_{n,k\bullet}V_{\theta_n,t})}{\sqrt{n}}\right)\right]^{1/2} - 1. \quad (38)$$

We now verify conditions (1.2)–(1.6) of Lemma 1 in Swensen (1985), having shown (1.7) to hold above. (1.2), that is, that $\mathbb{E}\sum_{t=1}^{n}[W_{n,t} - U_{n,t}]^2 \to 0$, where the expectation is taken under P_θ^n is shown to hold in Lemma S5. For (1.3), note that by Lemma S4, $P_\theta^n[|\sqrt{n}W_{n,t}|^{2+\rho}] \le C$ for some $\rho > 0$. Hence,

$$\sup_{n\in\mathbb{N}} P_\theta^n\left[\sum_{t=1}^{n} W_{n,t}^2\right] \le \sup_{n\in\mathbb{N}}\frac{1}{n}\sum_{t=1}^{n} P_\theta^n(\sqrt{n}W_{n,t})^2 \lesssim C.$$

For (1.4), by Lemma S4 and Markov's inequality,

$$\begin{aligned} P_\theta^n\Big(\max_{1\le t\le n}|W_{n,t}| > \varepsilon\Big) &\le P_\theta^n\left(\sum_{t=1}^{n} W_{n,t}^2\mathbf{1}\big\{|W_{n,t}| > \varepsilon\big\} > \varepsilon^2\right) \\ &\le \varepsilon^{-2}\sum_{t=1}^{n}\mathbb{E}\big[W_{n,t}^2\mathbf{1}\big\{\sqrt{n}|W_{n,t}| > \sqrt{n}\varepsilon\big\}\big] \\ &\to 0. \end{aligned}$$

Condition (1.5) follows from Lemma S7. For (1.6), by Lemma S4 and the fact that conditional expectations are L_1 contractions we have for any $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E}\left|\sum_{t=1}^{n}\mathbb{E}\big[W_{n,t}^2\mathbf{1}\big\{|W_{n,t}| > \varepsilon\big\}|\mathcal{F}_{n,t-1}\big]\right| &\le \sum_{t=1}^{n}\mathbb{E}\big|\mathbb{E}\big[W_{n,t}^2\mathbf{1}\big\{\sqrt{n}|W_{n,t}| > \sqrt{n}\varepsilon\big\}|\mathcal{F}_{n,t-1}\big]\big| \\ &\le \sum_{t=1}^{n}\mathbb{E}\big[W_{n,t}^2\mathbf{1}\big\{\sqrt{n}|W_{n,t}| > \sqrt{n}\varepsilon\big\}\big] \\ &\to 0. \end{aligned}$$

under P_θ^n. Additionally (iii) of Theorem 1 in Swensen (1985) holds since the relevant measures are both absolutely continuous with respect to Lebesgue measure (cf. Taniguchi and Kakizawa (2000, p. 34)). Therefore, by Lemma 1 in Swensen (1985), under P_θ^n,

$$\Lambda^n_{\theta_n(g,h)}(Y^n) = 2\sum_{t=1}^n W_{n,t} - \tau^2/2 + o_{P_\theta^n}(1) \rightsquigarrow \mathcal{N}\left(-\frac{\tau^2}{2}, \tau^2\right).$$

Given the form of $W_{n,t}$, it remains only to show that $\mathbb{E}[\mathsf{g}_n(Y^n)^2] \to \tau^2$. Since $\mathsf{g}_n(Y^n) = 2\sum_{t=1}^n W_{n,t}$ and $W_{n,t}$ forms a martingale difference array with respect to $\mathcal{F}_{n,t}$ (equation (36)),

$$\mathbb{E}\big[\mathsf{g}_n(Y^n)^2\big] = 4\,\mathbb{E}\left[\sum_{t=1}^n W_{n,t}\right]^2 = 4\,\mathbb{E}\sum_{t=1}^n W_{n,t}^2.$$

That this converges to τ^2 follows from Lemma S7 and the reverse triangle inequality. □

PROPOSITION 4. *Suppose that Assumption 1 holds. Then, if* $(g_n, h_n) \to (g, h)$,

$$\lim_{n\to\infty} d_{\mathrm{TV}}\big(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g,h)}\big) = 0.$$

PROOF. By Lemmas S8 and S9,

$$\log\frac{p^n_{\theta_n(g_n,h_n)}}{p^n_{\theta_n(g_n,h)}} = o_{P^n_{\theta_n(g_n,h_n)}}(1) \quad \text{and} \quad \log\frac{p^n_{\theta_n(g_n,h)}}{p^n_{\theta_n(g,h)}} = o_{P^n_{\theta_n(g,h)}}(1), \tag{39}$$

whenever $(g_n, h_n) \to (g, h)$. Therefore, by Lemma S23, (i) $d_{\mathrm{TV}}(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g_n,h)}) \to 0$ and (ii) $d_{\mathrm{TV}}(P^n_{\theta_n(g_n,h)}, P^n_{\theta_n(g,h)}) \to 0$. □

PROOF OF PROPOSITION 1. The only conclusion of Proposition 1, which is not immediately implied by those of Proposition 3, is that

$$\Lambda^n_{\theta_n(g_n,h_n)}(Y^n) - \mathsf{g}_n(Y^n) + \frac{1}{2}\mathbb{E}\big[\mathsf{g}_n(Y^n)^2\big] = o_{P_\theta^n}(1).$$

By Proposition 3,

$$\Lambda^n_{\theta_n(g,h)}(Y^n) - \mathsf{g}_n(Y^n) + \frac{1}{2}\mathbb{E}\big[\mathsf{g}_n(Y^n)^2\big] = o_{P_\theta^n}(1),$$

and hence it suffices to show that

$$\Lambda^n_{\theta_n(g_n,h_n)}(Y^n) - \Lambda^n_{\theta_n(g,h)}(Y^n) = o_{P_\theta^n}(1). \tag{40}$$

By Proposition 4, $d_{\mathrm{TV}}(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g,h)}) \to 0$, hence $(P^n_{\theta_n(g_n,h_n)})_{n\in N}$ and $(P^n_{\theta_n(g,h)})_{n\in\mathbb{N}}$ are mutually contiguous (e.g., Lemma 2.15 and Remark 18.3 in Strasser (1985)). By Proposition 3 and Example 6.5 in van der Vaart (1998), the same is true of $(P^n_{\theta_n(g,h)})_{n\in\mathbb{N}}$ and $(P^n_\theta)_{n\in\mathbb{N}}$. By the transitivity of mutual contiguity, we conclude that $(P^n_{\theta_n(g_n,h_n)})_{n\in\mathbb{N}}$ and $(P^n_\theta)_{n\in\mathbb{N}}$ are mutually contiguous. Combine this with equation (39) to conclude that (40) holds. □

PROOF OF COROLLARY 1. Combine Example 6.5 in van der Vaart (1998) with the fact that by Proposition 1, under P_θ^n,

$$\Lambda^n_{\theta_n(g_n,h_n)} \rightsquigarrow \mathcal{N}\left(-\frac{1}{2}\Psi(g,h), \Psi(g,h)\right). \qquad \square$$

A.4 *Scores*

PROOF OF LEMMA 1. Define

$$\mathcal{T}^{\eta|\gamma}_{P_\theta,H} := \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,t}) : h = (h_1,\dots,h_K) \in \dot{\mathscr{H}} \right\}, \quad V_{\theta,t} := Y_t - B_\theta X_t. \tag{41}$$

It suffices to show that (a) $\tilde{\ell}_\theta(Y_s, X_s) \in [\mathcal{T}^{\eta|\gamma}_{P_\theta,H}]^\perp \subset L_2(P_\theta^n)$ (componentwise) and (b)

$$\dot{\ell}_\theta(Y_s,X_s) - \tilde{\ell}_\theta(Y_s,X_s) \in \left\{ \sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,s}) : h = (h_1,\dots,h_K) \in \operatorname{cl}\dot{\mathscr{H}} \right\}, \quad s = 1,\dots,n. \tag{42}$$

For (a), the fact that $\tilde{\ell}_\theta(Y_s, X_s) \in L_2(P_\theta^n)$ follows straightforwardly from its form and the moment conditions in Assumption 1. Next, note that for any $h \in \dot{\mathscr{H}}$, $1 \le s \le n$,

$$\sum_{t=1}^{n}\sum_{k=1}^{K} \mathbb{E}\big[\tilde{\ell}_\theta(Y_s,X_s)h_k(A_{k\bullet}V_{\theta,t})\big] = 0$$

will be obtained under P_θ^n if for all $k, j, m \in [K]$ with $m \neq j$ and all $1 \le s \le n$, $1 \le t \le n$,

$$\begin{aligned}
\mathbb{E}\big[\phi_l(\epsilon_{m,s})\epsilon_{j,s}h_k(\epsilon_{k,t})\big] &= 0,\\
\mathbb{E}\big[\epsilon_{m,s}h_k(\epsilon_{k,t})\big] &= 0,\\
\mathbb{E}\big[\kappa(\epsilon_{m,s})h_k(\epsilon_{k,t})\big] &= 0,\\
\mathbb{E}\big[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})\big] &= 0,
\end{aligned}$$

the first three of which follow from the independence between components and across time of $(\epsilon_t)_{t\ge1}$. If $s \le t$, then by independence $\mathbb{E}[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})] = \mathbb{E}[(X_s-\mu)]\,\mathbb{E}[\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})] = 0$. If $s > t$, then $\mathbb{E}[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})] = \mathbb{E}[(X_s-\mu)h_k(\epsilon_{k,t})\,\mathbb{E}[\phi_m(\epsilon_{m,s})|\sigma(\epsilon_1,\dots,\epsilon_{s-1})]] = 0$ again by independence.

For (b), we note that for the components corresponding to a $x_l \in \{\alpha_l : l = 1,\dots,L_\alpha\} \cup \{\sigma_l : l = 1,\dots,L_\sigma\}$ and we have

$$\dot{\ell}_{\theta,x_l}(Y_s,X_s) - \tilde{\ell}_{\theta,x_l}(Y_s,X_s) = \sum_{k=1}^{K} \phi_k(A_{k\bullet}V_{\theta,s})A_{k\bullet}V_{\theta,s} + 1 - \tau_{k,1}A_{k\bullet}V_{\theta,s} - \tau_{k,2}\kappa(A_{k\bullet}V_{\theta,s}).$$

That this is mean zero and has finite second moment follows immediately from Assumption 1. That it has covariance zero with $A_{k\bullet}V_{\theta,s}$ and $\kappa(A_{k\bullet}V_{\theta,s})$ is ensured by the choice of τ_k.

For the components $x_l \in \{b_l : l = 1, \dots, L_b\}$,

$$\dot{\ell}_{\theta,x_l}(Y_s, X_s) - \tilde{\ell}_{\theta,x_l}(Y_s, X_s) = \sum_{k=1}^{K} \big(\phi_k(A_{k\bullet}V_{\theta,s}) + \varsigma_{k,1}A_{k\bullet}V_{\theta,s} + \varsigma_{k,2}\kappa(A_{k\bullet}V_{\theta,s})\big)\big[-e_k'\mu\big].$$

Again that this is mean zero and has finite second moment follows immediately from Assumption 1. That it has covariance zero with $A_{k\bullet}V_{\theta,s}$ and $\kappa(A_{k\bullet}V_{\theta,s})$ is ensured by the choice of ς_k.

This establishes that (42) holds since these are the defining properties of $\mathrm{cl}\,\dot{\mathscr{H}}$.[31] □

A.5 *Main theorems*

Proof of Theorem 1. Define

$$\begin{aligned}
R_{n,1}(\gamma_\star) &:= \big\|\sqrt{n}\mathbb{P}_n[\hat{\ell}_{\gamma_\star} - \tilde{\ell}_{\theta_\star}]\big\|,\\
R_{n,2}(\gamma_\star) &:= \big\|\sqrt{n}\mathbb{P}_n[\tilde{\ell}_{\theta_\star} - \tilde{\ell}_{\theta}] + \sqrt{n}\tilde{I}_{n,\theta}(\gamma_\star - \gamma)\big\|,\\
R_{n,3}(\gamma_\star) &:= \nu_n^{-1/2}\|\hat{I}_{n,\gamma_\star} - \tilde{I}_\theta\|,
\end{aligned}$$

where $\gamma_\star := (\alpha, \beta_\star)$ and $\theta_\star := (\gamma_\star, \eta)$. By Corollary 1, $P_\theta^n \triangleleft\triangleright P_{\theta_n((0,b_n),0)}^n$ for any $b_n \to b \in \mathbb{R}^{L_\beta}$. It then follows by Lemmas S13, S15, and Le Cam's first lemma (e.g., van der Vaart (1998, Lemma 6.4)) that

$$R_{n,i}(\gamma_n) \xrightarrow{P_\theta^n} 0 \quad \text{for } i = 1, 2, 3,$$

for any sequence $\gamma_n = (\alpha, \beta + b_n/\sqrt{n})$ with $b_n \to b \in \mathbb{R}^{L_\beta}$. Hence, by Lemma S21 also

$$R_{n,i}(\bar{\gamma}_n) \xrightarrow{P_\theta^n} 0 \quad \text{for } i = 1, 2, 3. \tag{43}$$

It follows that

$$\sqrt{n}\mathbb{P}_n[\hat{\ell}_{\bar{\gamma}_n} - \tilde{\ell}_\theta] = \sqrt{n}\mathbb{P}_n[\hat{\ell}_{\bar{\gamma}_n} - \tilde{\ell}_{\bar{\theta}_n}] + \sqrt{n}\mathbb{P}_n[\tilde{\ell}_{\bar{\theta}_n} - \tilde{\ell}_\theta] = -\tilde{I}_{n,\theta}\big(0, \sqrt{n}(\bar{\beta}_n - \beta)'\big)' + o_{P_\theta^n}(1),$$

and $\hat{I}_{n,\bar{\theta}_n} \xrightarrow{P_\theta^n} \tilde{I}_\theta$ and so $\hat{\mathcal{K}}_{\bar{\theta}_n,n} \xrightarrow{P_\theta^n} \tilde{\mathcal{K}}_\theta$ for

$$\tilde{\mathcal{K}}_\theta := \begin{bmatrix} I & -\tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1}\end{bmatrix}, \qquad \hat{\mathcal{K}}_{n,\theta} := \begin{bmatrix} I & -\hat{I}_{n,\theta,\alpha\beta}\hat{I}_{n,\theta,\beta\beta}^{-1}\end{bmatrix}.$$

We combine these to obtain

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n[\hat{\kappa}_{\bar{\gamma}_n,n} - \tilde{\kappa}_{n,\theta}]\\
&\quad = (\hat{\mathcal{K}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{K}}_\theta)\sqrt{n}\mathbb{P}_n[\hat{\ell}_{\bar{\gamma}_n} - \tilde{\ell}_\theta] + \tilde{\mathcal{K}}_{\theta_n}\sqrt{n}\mathbb{P}_n[\hat{\ell}_{\bar{\gamma}_n} - \tilde{\ell}_\theta] + (\hat{\mathcal{K}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{K}}_\theta)\sqrt{n}\mathbb{P}_n\tilde{\ell}_\theta\\
&\quad = -\tilde{\mathcal{K}}_\theta\tilde{I}_\theta\big(0, \sqrt{n}(\bar{\beta}_n - \beta)'\big)' + o_{P_\theta^n}(1)
\end{aligned}$$

[31] This follows by the argument of Lemma S8 in Lee and Mesters (2024b), noting that in the present context their H_0, $H_0^\star$, $\tilde{H}_0$ may be dropped.

$$= -\begin{bmatrix} I & -\tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{I}_{\theta,\alpha\alpha} & \tilde{I}_{\theta,\alpha\beta} \\ \tilde{I}_{\theta,\beta\alpha} & \tilde{I}_{\theta,\beta\beta} \end{bmatrix} \begin{bmatrix} 0 \\ \sqrt{n}(\bar{\beta}_n - \beta) \end{bmatrix} + o_{P_\theta^n}(1)$$

$$= o_{P_\theta^n}(1).$$

Next, let $Z_n := \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{\kappa}_{n,\bar{\gamma}_n}(Y_t, X_t)$ and rewrite it as

$$Z_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{\kappa}_\theta(Y_t, X_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^n \big(\hat{\kappa}_{\bar{\gamma}_n,n}(Y_t, X_t) - \tilde{\kappa}_\theta(Y_t, X_t)\big)$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{\kappa}_\theta(Y_t, X_t) + o_{P_\theta^n}(1).$$

By (i) of Lemma S15 and Le Cam's third lemma (e.g., (van der Vaart, 1998, Example 6.7)),

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{\ell}_\theta(Y_t, X_t) \rightsquigarrow \mathcal{N}\big(\tilde{I}_\theta(0', b')', \tilde{I}_\theta\big) \quad \text{under } P_{\theta_n},$$

and hence under P_{θ_n}

$$Z_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{\ell}_{\theta,\alpha}(Y_t, X_t) - \tilde{I}_{\theta,\alpha\beta}\tilde{I}_{\theta,\beta\beta}^{-1}\tilde{\ell}_{\theta,\beta}(Y_t, X_t) + o_{P_{\theta_n}^n}(1) \rightsquigarrow Z \sim \mathcal{N}(0, \tilde{\mathcal{I}}_\theta).$$

By repeated addition and subtraction along with the observations that any submatrix has a smaller operator norm than the original matrix, we obtain and the matrix inverse is Lipschitz continuous at a nonsingular matrix and we obtain

$$\|\hat{\mathcal{I}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{I}}_\theta\|_2 \lesssim \|\hat{I}_{n,\bar{\gamma}_n} - \tilde{I}_\theta\|_2.$$

Hence, by (43) have $\|\hat{\mathcal{I}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{I}}_\theta\|_2 = o_{P_{\theta_n}^n}(\nu_n^{1/2})$. By Proposition S1 in Lee and Mesters (2024b),

$$\hat{\mathcal{I}}_{n,\bar{\gamma}_n}^{t,\dagger} \xrightarrow{P_{\theta_n}^n} \tilde{\mathcal{I}}_\theta^\dagger \quad \text{and} \quad P_{\theta_n}^n R_n \to 1, \quad \text{where } R_n := \big\{\operatorname{rank}\big(\tilde{\mathcal{I}}_{n,\bar{\gamma}_n}^t\big) = \operatorname{rank}(\tilde{\mathcal{I}}_\theta)\big\}.$$

Suppose first that $r := \operatorname{rank}(\tilde{\mathcal{I}}_\theta) > 0$. By Slutsky's lemma and the continuous mapping theorem, we have that

$$\hat{S}_{n,\bar{\gamma}_n}^{SR} = Z_n' \hat{\mathcal{I}}_{n,\bar{\gamma}_n}^{t,\dagger} Z_n \rightsquigarrow Z' \tilde{\mathcal{I}}_\theta^\dagger Z \sim \chi_r^2,$$

where the distributional result $X := Z'\tilde{\mathcal{I}}_\theta^\dagger Z \sim \chi_r^2$, follows from, for example, Theorem 9.2.2 in Rao and Mitra (1971). On R_n c_n is the $1-a$ quantile of the χ_r^2 distribution, which we will call c. Hence, we have $c_n \xrightarrow{P_{\theta_n}^n} c$ and as a result, $\hat{S}_{n,\bar{\gamma}_n}^{SR} - c_n \rightsquigarrow X - c$ where $X \sim \chi_r^2$. Since the χ_r^2 distribution is continuous, we have by the Portmanteau theorem

$$P_{\theta_n}^n\big(\hat{S}_{n,\bar{\gamma}_n}^{SR} > c_n\big) = 1 - P_{\theta_n}^n\big(\hat{S}_{n,\bar{\gamma}_n}^{SR} - c_n \le 0\big) \to 1 - \mathrm{P}(X - c \le 0) = 1 - \mathrm{P}(X \le c) = a,$$

which completes the proof in the case that $r > 0$.

We next handle the case with $r = 0$. On the sets R_n, we have that $\hat{\mathcal{I}}^t_{n,\bar{\gamma}_n}$ is the zero matrix, whose Moore–Penrose inverse is also the zero matrix. Hence, on these sets we have $\hat{S}^{SR}_{n,\bar{\gamma}_n} = 0$ and $c_n = 0$ and, therefore, do not reject, implying

$$P^n_{\theta_n}\big(\hat{S}^{SR}_{n,\bar{\gamma}_n} > c_n\big) \leq 1 - P^n_{\theta_n} R_n \to 0.$$

It follows that $P^n_{\theta_n}(\hat{S}^{SR}_{n,\bar{\gamma}_n} > c_n) \to 0$.

This completes the demonstration of the pointwise convergence

$$\lim_{n\to\infty} P^n_{\theta_n(b,h)}(\hat{S}_{n,\bar{\gamma}_n} > c_n) = \begin{cases} \alpha & \text{if } \operatorname{rank}(\tilde{\mathcal{I}}_\theta) > 0, \\ 0 & \text{if } \operatorname{rank}(\tilde{\mathcal{I}}_\theta) = 0. \end{cases}$$

Finally, to complete the proof, note that the norm on $B \times \dot{\mathscr{H}}$ induces the product topology, hence $B^\star \times H^\star$ is compact. The uniformity then follows from the asymptotic uniform equicontinuity in total variation of $(b, h) \mapsto P^n_{\theta_n(b,h)}$ on $B^\star \times H^\star$, which is an immediate consequence of Lemma 4 and the fact that asymptotic uniform equicontinuity is implied by asymptotic equicontinuity on a compact set. □

Proof of Corollary 2. Apply Theorem 1 to conclude:

$$\lim_{n\to\infty} \inf_{(b,h)\in B^\star\times H^\star} P^n_{\theta_n(b,h)}(\alpha \in \hat{C}_n) \geq 1 - \lim_{n\to\infty} \sup_{(b,h)\in B^\star\times H^\star} P^n_{\theta_n(b,h)}\big(\hat{S}^{SR}_{n,\bar{\gamma}_n} > c_n\big) \geq 1 - \alpha.$$ □

Proof of Proposition 2. By the uniform delta method (van der Vaart (1998, Theorem 3.8)), under $P^n_{\theta_n(b,h)}$,

$$\sqrt{n}\big(g(\alpha, \hat{\beta}_n) - g\big(\alpha, \beta_n(b)\big)\big) \overset{P^n_{\theta_n(b,h)}}{\rightsquigarrow} \mathcal{N}\big(0, J_\gamma \Sigma J_\gamma'\big).$$

Combine with $\hat{V}_{n,\alpha} \xrightarrow{P^n_{\theta_n(b,h)}} J_\gamma \Sigma J_\gamma' \succ 0$ and the continuous mapping theorem to obtain

$$n g(\alpha, \hat{\beta}_n)' \hat{V}^{-1}_{n,\alpha} g(\alpha, \hat{\beta}_n) \overset{P^n_{\theta_n(b,h)}}{\rightsquigarrow} \chi^2_{d_g}.$$

Hence, pointwise in $(b, h) \in B^\star \times H^\star$,

$$\begin{aligned} \lim_{n\to\infty} P^n_{\theta_n(b,h)}\big(g\big(\alpha, \beta_n(b)\big) \in \hat{C}_{n,g,\alpha_n,1-a}\big) &= \lim_{n\to\infty} P^n_{\theta_n(b,h)}\big(n g(\alpha, \hat{\beta}_n)' \hat{V}^{-1}_{n,\alpha} g(\alpha, \hat{\beta}_n) \leq c_a\big) \\ &= 1 - a. \end{aligned}$$

The uniform statement then follows from Proposition 4. □

Proof of Corollary 3. This follows directly from the hypotheses and the fact that

$$\begin{aligned} &P^n_{\theta_n(b,h)}\big(g\big(\alpha, \beta_n(b)\big) \in \hat{C}_{n,g}\big) \\ &\quad \geq P^n_{\theta_n(b,h)}\big(\big\{g(\alpha, \hat{\beta}_n) \in \hat{C}_{n,g,\alpha,1-q_2}\big\} \cap \{\alpha \in \hat{C}_{n,1-q_1}\}\big) \\ &\quad \geq P^n_{\theta_n(b,h)}\big(g(\alpha, \hat{\beta}_n) \in \hat{C}_{n,g,\alpha,1-q_2}\big) + P^n_{\theta_n(b,h)}(\alpha \in \hat{C}_{n,1-q_1}) - 1. \end{aligned}$$ □

PROOF OF THEOREM 2. Similar to as in the proof of Theorem 1, define

$$R_{n,1}(\gamma_\star) := \big\| \sqrt{n}\mathbb{P}_n[\hat{\ell}_{\gamma_\star} - \tilde{\ell}_{\theta_\star}] \big\|,$$
$$R_{n,2}(\gamma_\star) := \big\| \sqrt{n}\mathbb{P}_n[\tilde{\ell}_{\theta_\star} - \tilde{\ell}_{\theta}] + \sqrt{n}\tilde{I}_{n,\theta}(\gamma_\star - \gamma) \big\|,$$
$$R_{n,3}(\gamma_\star) := \|\hat{I}_{\gamma_\star,n} - \tilde{I}_\theta\|,$$

where $\theta_\star := (\gamma_\star, \eta)$. By Corollary 1, for any $g_n \to g \in \mathbb{R}^{L_\alpha + L_\beta}$, $P_\theta^n \triangleleft\triangleright P^n_{\theta_n(g_n,0)}$. By Lemmas S13, S15, and Le Cam's first lemma (e.g., van der Vaart (1998, Lemma 6.4))

$$R_{n,i}(\gamma_n) \xrightarrow{P_\theta^n} 0 \quad \text{for } i = 1, 2, 3,$$

where $\gamma_n = \gamma + g_n/\sqrt{n}$. Hence, by Lemma S21 also

$$R_{n,i}(\bar{\gamma}_n) \xrightarrow{P_\theta^n} 0 \quad \text{for } i = 1, 2, 3. \tag{44}$$

Combine these and (29) to yield

$$\begin{aligned}
\sqrt{n}\tilde{I}_\theta(\hat{\gamma}_n - \gamma) &= \sqrt{n}\tilde{I}_\theta(\bar{\gamma}_n - \gamma) + \sqrt{n}\tilde{I}_\theta \hat{I}^{-1}_{n,\bar{\gamma}_n} \bar{\ell}_{n,\bar{\gamma}_n} \\
&= \sqrt{n}\tilde{I}_\theta(\bar{\gamma}_n - \gamma) + \sqrt{n}\mathbb{P}_n(\hat{\ell}_{n,\bar{\gamma}_n} - \tilde{\ell}_{\bar{\theta}_n}) + \sqrt{n}\mathbb{P}_n(\tilde{\ell}_{\bar{\theta}_n} - \tilde{\ell}_\theta) \\
&\quad + \sqrt{n}\mathbb{P}_n\tilde{\ell}_\theta + o_{P_\theta^n}(1) \\
&= \sqrt{n}\mathbb{P}_n\tilde{\ell}_\theta + o_{P_\theta^n}(1).
\end{aligned}$$

Combine this with Lemma S15(i) and the continuous mapping theorem. □

REFERENCES

Amari, Shun-Ichi and Jean-Francois Cardoso (1997), "Blind source separation—semiparametric statistical approach." IEEE Transactions On Signal Processing, 45 (11). [534]

Andrews, Isaiah and Anna Mikusheva (2015), "Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models." Quantitative Economics, 6. [527]

Andrews, Isaiah, James Stock, and Liyang Sun (2019), "Weak instruments in iv regression: Theory and practice." *Annual Review of Economics*, 11, 727–753. [524]

Baumeister, Christiane and James D. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83 (5), 1963–1999. [523, 526, 538, 550, 551, 552]

Baumeister, Christiane and James D. Hamilton (2019), "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks." *American Economic Review*, 109 (5), 1873–1910. [526, 553, 554, 555]

Bekaert, Geert, Eric Engstrom, and Andrey Ermolov (2020), "Aggregate demand and aggregate supply effects of Covid-19: A real-time analysis." FEDS Working Paper No. 2020-49, Available at SSRN: https://ssrn.com/abstract=3633961. [524]

Bekaert, Geert, Eric Engstrom, and Andrey Ermolov (2021), "Macro risks and the term structure of interest rates." *Journal of Financial Economics*, 141 (2), 479–504. [524]

Bickel, Peter J., Chris A. J. Klaasen, Ya'acov Ritov, and Jon A. Wellner (1998), *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York, NY, USA. [527, 530, 533, 534]

Bickel, Peter J., Ya'acov Ritov, and Thomas M. Stoker (2006), "Tailor-made tests for goodness of fit to semiparametric hypotheses." *Ann. Statist.*, 34 (2), 721–741. [527]

Blum, Julius R., John Kiefer, and Murray Rosenblatt (1961), "Distribution free tests of independence based on the sample distribution function." *The Annals of Mathematical Statistics*, 32 (2), 485–498. [547]

Bonhomme, Stephane and Jean-Marc Robin (2009), "Consistent noisy independent component analysis." Journal of Econometrics, 149. [524]

Braun, Robin (2023), "The importance of supply and demand for oil prices: Evidence from non-Gaussianity." *Quantitative Economics*, 14 (4), 1163–1198. [524]

Brunnermeier, Markus, Darius Palia, Karthik A. Sastry, and Christopher A. Sims (2021), "Feedbacks: Financial markets and economic activity." *American Economic Review*, 111 (6), 1845–1879. [524]

Chen, Aiyou and Peter J. Bickel (2006), "Efficient independent component analysis." Annals of Statistics, 34 (6). [525, 530, 534, 536, 549]

Choi, Sungsub, W. Jackson Hall, and Anton Schick (1996), "Asymptotically uniformly most powerful tests in parametric and semiparametric models." *Ann. Statist.*, 24 (2), 841–861. [527, 533, 534, 535]

Comon, Pierre (1994), "Independent component analysis, a new concept?" Signal Processing, 36. [524, 542]

Davis, Richard and Serena Ng (2023), "Time series estimation of the dynamic effects of disaster-type shocks." *Journal of Econometrics*, 235 (1), 180–201. [524]

Drautzburg, Thorsten and Jonathan H. Wright (2023), "Refining set-identification in vars through independence." *Journal of Econometrics*, 235 (2), 1827–1847. [524, 525, 526, 540, 547, 548, 549]

Dufour, Jean-Marie and Pascale Valery (2016), "Rank-robust regularized Wald-type tests." Working paper. [537]

Fiorentini, Gabriele and Enrique Sentana (2022), "Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions." Journal of Econometrics. [524, 546]

Gouriéroux, Christian, Alain Monfort, and Jean-Paul Renne (2017), "Statistical inference for independent component analysis: Application to structural VAR models." Journal of Econometrics, 196. [524, 546, 547, 548, 549]

Gouriéroux, Christian, Alain Monfort, and Jean-Paul Renne (2019), "Identification and estimation in non-fundamental structural VARMA models." *The Review of Economic Studies*, 87 (4), 1915–1953. [524]

Granziera, Eleonora, Hyungsik Roger Moon, and Frank Schorfheide (2018), "Inference for vars identified with sign restrictions." *Quantitative Economics*, 9 (3), 1087–1121. [525, 540]

Guay, Alain (2021), "Identification of structural vector autoregressions through higher unconditional moments." *Journal of Econometrics*, 225 (1), 27–46. [524]

Hallin, Marc and Abdessamad Saidi (2007), "Optimal tests of noncorrelation between multivariate time series." *Journal of the American Statistical Association*, 102 (479), 938–951. [530]

Hallin, Marc and Bas J. M. Werker (1999), "Optimal testing for semi-parametric autoregressive models: From Gaussian Lagrange multipliers to regression rank scores and adaptive tests." In *Asymptotics* (Subir Ghosh, ed.), Nonparametrics and Time Series, 295–358, Marcel Dekker. [528]

Herrera, Ana Maria and Sandeep Kumar Rangaraju (2020), "The effect of oil supply shocks on us economic activity: What have we learned?" *Journal of Applied Econometrics*, 35 (2), 141–159. [526, 554]

Herwartz, Helmut (2019), "Long-run neutrality of demand shocks: Revisiting Blanchard and Quah (1989) with independent structural shocks." *Journal of Applied Econometrics*, 34 (5), 811–819. [524]

Hoeffding, Wassily (1948), "A class of statistics with asymptotically normal distribution." *The Annals of Mathematical Statistics*, 19 (3), 293–325. [547]

Hoesch, Lukas, Adam Lee, and Geert Mesters (2023), "Locally robust inference for non-Gaussian svar models." Barcelona School of Economics Working Paper 1367. [536, 543, 544, 549, 550]

Hoesch, Lukas, Adam Lee, and Geert Mesters (2024), "Supplement to 'Locally robust inference for non-Gaussian SVAR models'." *Quantitative Economics Supplemental Material*, 15, https://doi.org/10.3982/QE2274. [527, 528, 529, 557]

Horn, Roger A. and Charles R. Johnson (2013), *Matrix Analysis*, second edition. Cambridge University Press. [529]

Jin, Kun (1992), "Empirical smoothing parameter selection in adaptive estimation." The Annals of Statistics, 1844–1874. [525, 536]

Keweloh, Sascha Alexander (2021), "A generalized method of moments estimator for structural vector autoregressions based on higher moments." *Journal of Business & Economic Statistics*, 39 (3), 772–782. [524, 546, 547, 548, 551, 554]

Kilian, Lutz (2019), "Measuring global real economic activity: Do recent critiques hold up to scrutiny?" *Economics Letters*, 178, 106–110. [554]

Kilian, Lutz and Helmut Lütkepohl (2017), *Structural Vector Autoregressive Analysis*. Cambridge University Press. [524, 539]

Kilian, Lutz and Daniel 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 (5), 1166–1188. [523, 526, 552, 554, 555, 556]

Kilian, Lutz and Daniel P. Murphy (2014), "The role of inventories and speculative trading in the global market for crude oil." *Journal of Applied econometrics*, 29 (3), 454–478. [554]

Kleibergen, Frank (2005), "Testing parameters in GMM without assuming that they are identified." Econometrica, 73 (4). [527]

Lanne, Markku, Keyan Liu, and Jani Luoto (2022), "Identifying structural vector autoregression via leptokurtic economic shocks." *Journal of Business & Economic Statistics*, 0 (0), 1–11. [524]

Lanne, Markku and Jani Luoto (2021), "Gmm estimation of non-Gaussian structural vector autoregression." *Journal of Business & Economic Statistics*, 39 (1), 69–81. [524, 546, 547, 548, 551]

Lanne, Markku and Jani Luoto (2022), "Statistical identification of economics shocks by signs in structural vector autoregression." *Essays is Honor of Fabio Canova. Advances in Econometrics, A*, 44, 165–175. [526, 551, 552]

Lanne, Markku and Helmut Lütkepohl (2010), "Structural vector autoregressions with nonnormal residuals." *Journal of Business & Economic Statistics*, 28 (1), 159–168. [524]

Lanne, Markku, Mika Meitz, and Pentti Saikkonen (2017), "Identification and estimation of non-Gaussian structual vector autoregressions." Journal of Econometrics, 196. [524]

Le Cam, Lucien M. (1960), *Locally Asymptotically Normal Families of Distributions: Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses*. University of California Press. Berkeley, CA. [538]

Le Cam, Lucien M. and Grace L. Yang (2000), *Asypmtotics in Statistics: Some Basic Concepts*, second edition. Springer, New York, NY, USA. [538]

Lee, Adam (2024), "Locally Regular and Efficient Tests in Non-Regular Semiparametric Models." ArXiv preprint, arXiv:2403.05999. [538]

Lee, Adam and Geert Mesters (2024a), "Locally robust inference for non-Gaussian linear simultaneous equations models." *Journal of Econometrics*, 240 (1), 105647. [524, 527, 534, 546, 558]

Lee, Adam and Geert Mesters (2024b), "Supplement to 'locally robust inference for non-Gaussian linear simultaneous equations models'." Journal of Econometrics Supplementary Material, 240. [534, 558, 562, 563]

Lewis, Daniel J. (2021), "Identifying shocks via time-varying volatility." *The Review of Economic Studies*, 88 (6), 3086–3124. [524]

Marron, J. Steve and Matt P. Wand (1992), "Exact mean integrated squared error." The Annals of Statistics, 712–736. [544]

Matteson, David S. and Ruey S. Tsay (2017), "Independent component analysis via distance covariance." *Journal of the American Statistical Association*, 112 (518), 623–637. [526, 551, 554]

Maxand, Simone (2020), "Identification of independent structural shocks in the presence of multiple Gaussian components." *Econometrics and Statistics*, 16, 55–68. [524]

Moneta, Alessio, Doris Entner, Patrik O. Hoyer, and Alex Coad (2013), "Causal inference by independent component analysis: Theory and applications*." *Oxford Bulletin of Economics and Statistics*, 75 (5), 705–730. [524]

Newey, Whitney K. (1990), "Semiparametric efficiency bounds." *Journal of Applied Econometrics*, 5 (2), 99–135. [533]

Montiel Olea, José Luis, Mikkel Plagborg-Møller, and Eric Qian (2022), "Svar identification from higher moments: Has the simultaneous causality problem been solved?" *AEA Papers and Proceedings*, 112, 481–485. [524, 526, 551, 554]

Pinkse, Joris and Karl Schurter (2021), "Estimates of derivatives of (log) densities and related objects." Econometric Theory, 1–36. [525]

Rao, C. Radhakrishna and Sujit Kumar Mitra (1971), *Generalized Inverse of Matrices and Its Applications*. John Wiley & Sons, Inc., New York, NY, USA. [563]

Rose, Morris E. (1957), *Elementary Theory of Angular Momentum*. John Wiley & Sons. [544]

Sen, Ananda (2012), "On the interrelation between the sample mean and the sample variance." The American Statistician, 66 (2). [529]

Shapiro, Matthew D. and Mark W. Watson (1988), "Sources of business cycle fluctuations." *NBER Macroeconomics Annual*, 3, 111–148. [550]

Shimizu, Shohei, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen (2006), "A linear non-Gaussian acyclic model for causal discovery." *Journal of Machine Learning Research*, 7 (72), 2003–2030. [524]

Sims, Christopher A. (2021), "Svar identification through heteroskedasticity with misspecified regimes." Working Paper. [524]

Stock, James H. and Jonathan H. Wright (2000), "GMM with weak identification." Econometrica, 68 (5). [526, 527, 547]

Strasser, Helmut (1985), *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. De Gruyter Studies in Mathematics. W. de Gruyter. [560]

Swensen, Anders Rygh (1985), "The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend." *Journal of Multivariate Analysis*, 16 (1), 54–70. [532, 559, 560]

Taniguchi, Masanobu and Yoshihide Kakizawa (2000), *Asymptotic Theory of Statistical Inference for Time Series*. Springer. [560]

Tank, Alex, Emily B. Fox, and Ali Shojaie (2019), "Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series." *Biometrika*, 106 (2), 433–452. [524]

van der Vaart, Aad W. (1998), *Asymptotic Statistics*, first edition. Cambridge University Press, New York, NY, USA. [530, 532, 537, 538, 542, 543, 544, 557, 560, 561, 562, 563, 564, 565]

van der Vaart, Aad W. (2002), "Semiparametric statistics." In *Lectures on Probability Theory and Statistics: Ecole D'Eté de Probabilités de Saint-Flour XXIX—1999* (Pierre Bernard, ed.). Springer, Berlin, Germany. [527, 533, 534]

van der Vaart, Aad W. and Jon A. Wellner (1996), *Weak Convergence and Empirical Processes*, first edition. Springer-Verlag, New York, Inc., New York, NY, USA. [543]

Wang, Y. Samuel and Mathias Drton (2019), "High-dimensional causal discovery under non-Gaussianity." *Biometrika*, 107 (1), 41–59. [524]

Zhou, Xiaoqing (2020), "Refining the workhorse oil market model." *Journal of Applied Econometrics*, 35 (1), 130–140. [554, 555]

Co-editor James Hamilton handled this manuscript.

Manuscript received 24 October, 2022; final version accepted 24 January, 2024; available online 1 February, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.10539210. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.