Barcelona School of Economics

WORKING PAPERS

Locally Robust Inference for Non-Gaussian SVAR Models

BSE Working Paper 1367
October 2022 (Revised: December 2023)

Lukas Hoesch, Adam Lee, Geert Mesters

Locally Robust Inference for Non-Gaussian SVAR models

Lukas Hoesch[*] Adam Lee[†] Geert Mesters[‡]

December 20, 2023

Abstract

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 semi-parametric 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 labour 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.

JEL classification: C32, C39, C51

Keywords: weak identification, semi-parametric inference, hypothesis testing, impulse responses, independent component analysis.

[*]Vrije Universiteit Amsterdam and Tinbergen Institute; `l.hoesch@vu.nl`
[†]BI Norwegian Business School; `adam.lee@bi.no`
[‡]Universitat Pompeu Fabra, Barcelona School of Economics and CREI; `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 nancial 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).

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 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 et al., 2013; Lanne et al., 2017; Kilian and Lütkepohl, 2017; Maxand, 2020; Lanne and Luoto, 2021; Gouriéroux et al., 2017, 2019; Tank et al., 2019; Herwartz, 2019; Bekaert et al., 2021, 2020; Fiorentini and Sentana, 2022; Braun, 2021; Sims, 2021; Guay, 2021; Brunnermeier et al., 2021; Drautzburg and Wright, 2023; Keweloh, 2021; Davis and Ng, 2022; Lanne et al., 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, e.g. 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 non-zero 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 standard (pseudo) maximum likelihood or generalised method of moments methodologies to yield correct coverage when the distance to the Gaussian distribution is not sufficiently large. As such we

[1] See Montiel Olea et al. (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 et al., 2006; Bonhomme and Robin, 2009; Wang and Drton, 2019).

[3] Simulation studies in, among others, Gouriéroux et al. (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 (2023a) for more discussion of the same issue in static ICA models.

[4] See e.g. the recent review by Andrews et al. (2019).

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 non-parametric part.

For this set-up we provide three main results. First, we adopt a semi-parametric generalisation of Neyman-Rao score statistic in order to test the possibly weakly identified (or under / unidentified) parameters of the SVAR. More precisely, the semi-parametric 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/non- identified 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 semi-parametric 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 utilise our proposed score test for the weakly identified parameters in a Bonferroni-based procedure (cf. Granziera et al., 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 semi-parametrically efficient for the finite dimensional parameters in the semi-parametric 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 semi-parametric 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 semi-parametric score test are always close to the nominal size. This is in contrast to several existing methods that are not 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 semi-parametric score test generally exceeds that of alternative robust methods such as weak identification robust GMM methods. Finally, we show that while the

[5]The general approach is applicable with other choices of log density score estimators, e.g. 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 A.1.

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 non-robust 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 non-parametric 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 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, i.e. the densities of the structural shocks. Despite this difference, conceptually our approach is similar to

[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 Montiel Olea et al. (2022).

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 et al. (1998) and van der Vaart (2002). While the majority of the statistical literature focuses on efficient estimation in semi-parametric models, a few papers have contributed to testing in well identified models (e.g. Choi et al., 1996; Bickel et al., 2006). The major difference with our paper is that in our setting, a subset of the parameters of interest are possibly weakly- or un- / under- identified, which violates a key regularity condition assumed in this literature. Lee and Mesters (2023a) 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 semi-parametric model and discusses the needed regularity conditions. Section 3 establishes a number of preliminary results that are of general interest. The semi-parametric 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.

2 Semi-parametric SVAR model

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

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

where $X_t := (1, Y'_{t-1}, \dots, Y'_{t-p})'$, $B := (c, B_1, \dots, B_p)$ 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 non-parametric part of model (2) we let $\eta = (\eta_1, \dots, \eta_K)$ correspond to the

[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 2.1). The supplementary material Section S1 provides more discussion and examples.

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 $b = \mathrm{vec}(B)$.

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

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

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 semi-parametric SVAR model, we now proceed to formulate the required assumptions. Broadly speaking, we split our assumptions into two parts: Assumption 2.1 details the main assumptions that allow us to establish the properties of the semi-parametric score test and Assumption 2.2 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.

Assumption 2.1: *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 $\delta > 0$);

[8] Lemma A.1 in the Appendix shows that, under Assumptions 2.1 and 2.2, 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.

(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) $(\alpha, \sigma) \to [D_{\alpha_l}(\alpha, \sigma)]_{k\bullet} A(\alpha, \sigma)^{-1}_{\bullet j}$ *and* $(\alpha, \sigma) \to [D_{\sigma_m}(\alpha, \sigma)]_{k\bullet} A(\alpha, \sigma)^{-1}_{\bullet j}$, *with* $D_{\alpha_l}(\alpha, \sigma) \coloneqq \partial A(\alpha, \sigma)/\partial \alpha_l$ *and* $D_{\sigma_m}(\alpha, \sigma) \coloneqq \partial A(\alpha, \sigma)/\partial \sigma_m$, *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) normalises 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. Whilst this may at first glance appear a strong condition, in Section S3 of the supplementary material 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 $A(\alpha, \sigma)$.

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 2.2: *Let* $\phi_{k,n} \coloneqq \phi_k \mathbf{1}_{[\Xi^L_{k,n}, \Xi^U_{k,n}]}$, $\Delta_{k,n} \coloneqq \Xi^U_{k,n} - \Xi^L_{k,n}$ *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* $[\Xi^L_{k,n}, \Xi^U_{k,n}] \uparrow \tilde{\Xi} \supset \mathrm{supp}(\eta_k)$ *and* $\delta_{k,n} \downarrow 0$ *it holds that*

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

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

(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* $\inf_{s \in [\Xi^L_{k,n}, \Xi^U_{k,n}]} |\eta_k(s)| \geq c\delta_{k,n}$.

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

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

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 semi-parametric SVAR models that we believe are useful more broadly. First, we provide a (uniform) local asymptotic normality [(U)LAN] result for the semi-parametric SVAR model in (2). The primary difference with existing results is that we explicitly perturb the non-parametric part of the model, i.e. 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 semi-parametric SVAR model, see e.g. van der Vaart (1998); Bickel et al. (1998) for general discussions on efficient score 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 $(k = 1, \ldots, K)$ and define

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

where ι is the identity funcion, $\kappa(z) = z^2 - 1$ (as defined above) and $\mathcal{C}^1_b(\lambda)$ 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 $\|(g, h)\| := \sqrt{\|g\|_2^2 + \sum_{k=1}^K \|h_k\|^2_{L_2(G_k)}}$ 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^n_θ the density of P^n_θ with respect to λ^n and $\Lambda^n_{\theta_n}$ the (conditional) log likelihood ratio:

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

where $\ell_\theta(Y_t, X_t)$ denotes the t-th contribution to the conditional log likelihood for the SVAR

[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 A.1 can be adopted.

model evaluated at θ. We note that this can be explicitly written as

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

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

$$\dot{\ell}_{\theta,\alpha_l}(Y_t, X_t) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j} \phi_k(A_{k\bullet}V_{\theta,t}) A_{j\bullet}V_{\theta,t} + \sum_{k=1}^{K} \zeta^{\alpha}_{l,k,k} \left(\phi_k(A_{k\bullet}V_{\theta,t})A_{k\bullet}V_{\theta,t} + 1\right), \tag{7}$$

$$\dot{\ell}_{\theta,\sigma_l}(Y_t, X_t) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{l,k,j} \phi_k(A_{k\bullet}V_{\theta,t}) A_{j\bullet}V_{\theta,t} + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k} \left(\phi_k(A_{k\bullet}V_{\theta,t})A_{k\bullet}V_{\theta,t} + 1\right), \tag{8}$$

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

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

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

Under assumption 2.1, we have the following ULAN result.[13]

Proposition 3.1 (ULAN): *Suppose that assumption 2.1 holds. Then as $n \to \infty$,*

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

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

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

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

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

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

Corollary 3.1: *If assumption 2.1 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 semi-parametric SVAR model can be locally asymp-

[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 $(g,h) \mapsto P^n_{\theta_n(g,h)}$.

totically 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 for the nuisance parameters (β, η) (e.g. Bickel et al., 1998; van der Vaart, 2002; Newey, 1990; Choi et al., 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 3.1: *Given Assumption 2.1 the efficient score function for γ in the semi-parametric SVAR model $\mathcal{P}^n_\Theta$ 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*

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

with components

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

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

$$\tau_k := M_k^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \varsigma_k := M_k^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{where } M_k := \begin{pmatrix} 1 & \mathbb{E}_\theta(A_{k\bullet} V_{\theta,t})^3 \\ \mathbb{E}_\theta(A_{k\bullet} V_{\theta,t})^3 & \mathbb{E}_\theta(A_{k\bullet} V_{\theta,t})^4 - 1 \end{pmatrix}.$$

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 (2023a). 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 $\tilde{\ell}_{\theta,\beta}(Y_t, X_t) = \left((\tilde{\ell}_{\theta,\sigma_l}(Y_t, X_t))_{l=1}^{L_\sigma}, (\tilde{\ell}_{\theta,b_l}(Y_t, X_t))_{l=1}^{L_b}\right)'$.

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

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

With Lemma 3.1 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)

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

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

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

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 et al. (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 (2023b). The singularity plays an important role in the construction of the semi-parametric score statistic below.

4 Inference for potentially non-identified 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 2.1 and 2.2 are imposed, under which α may not be (locally) identified.

Our approach is based on testing hypotheses of the form

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

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 semi-parametric setting (e.g. Choi et al., 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 3.1. These scores can be estimated by replacing the population quantities by sample equivalents. We have

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

with components

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

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 3.1 by their sample counterparts: $\hat{\tau}_k = \hat{M}_k(0, -2)'$ and $\hat{\varsigma}_k = \hat{M}_k(1, 0)'$, where

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

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

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

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

[15] In the simulation study below we fix the number of B-splines $B_{k,n} = 7$ and in the supplementary material we also investigate a data driven selection procedure.

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 supplementary material Section S4 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

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

and

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

Since the information matrix may be singular, we need to make an adjustment. Specifically, given the truncation rate ν_n defined in Assumption 2.2, we define a truncated eigenvalue version of the information matrix estimate as

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

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 non-increasing 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 semi-parametric score statistic for the SVAR model as follows.

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

where $\hat{\mathcal{I}}^{t,\dagger}_{n,\gamma}$ is the Moore-Penrose pseudo-inverse of $\hat{\mathcal{I}}^{t}_{n,\gamma}$. We note that the test statistic can be evaluated at any $\gamma = (\alpha,\beta)$. In practice we will set $\alpha = \alpha_0$, i.e. 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 4.1: *Suppose Assumptions 2.1 and 2.2 hold and that $\hat{\beta}_n$ is a $\sqrt{n}-$consistent estimator* *of β under P^n_θ, 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 $r_n = \operatorname{rank}(\hat{\mathcal{I}}^t_{n,\bar{\gamma}_n})$ and denote by c_n the $1-a$ quantile of the $\chi^2_{r_n}$ distribution, for any $a \in (0,1)$. Then if $\theta_n := (\alpha_0, \beta + b/\sqrt{n}, \eta(1 + h/\sqrt{n}))$,*

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

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,*

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

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]

Second, the score statistic is evaluated at the discretised 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 2.2, 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 $\hat{\mathcal{I}}^{t,\dagger}_{n,\gamma}$ is similar to $\hat{\mathcal{I}}^{\dagger}_{n,\gamma}$, 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 non-singular version that is computed with $\hat{\mathcal{I}}^{-1}_{n,\bar{\gamma}_n}$ instead of $\hat{\mathcal{I}}^{t,\dagger}_{n,\bar{\gamma}_n}$; 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) minimax optimal,

[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.

[18] Indeed, in practice, we always discretise at machine precision, see **Algorithm 1** below.

[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 amongst unbiased tests in the two-sided case can be shown similarly.

as can be shown by an argument analogous to that given in Lee (2022).

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 4.1: *Suppose that assumptions 2.1 and 2.2 hold. Let $\bar{\beta}_n$, $B^\star$, $H^\star$ and $\theta_n(b,h)$ be as in Theorem 4.1. Then,*

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

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 labour supply-demand model of Baumeister and Hamilton (2015) that is considered in Section 8.

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

Algorithm 1: Confidence set for α

(i) Choose a set $\mathcal{A}$;

(ii) For each $\alpha \in \mathcal{A}$:

1 Obtain estimates $\hat{\beta}_n = (\hat{\sigma}_n, \hat{b}_n)$, with $b_n = \operatorname{vec}(B_n)$, and set $\hat{V}_t = Y_t - \hat{B}_n X_t$;

2 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);

3 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 $\hat{\gamma}_n = (\alpha_0, \hat{\beta}_n)$;

4 Compute $\hat{\kappa}_{n,\hat{\gamma}_n}(Y_t, X_t)$ and $\hat{\mathcal{I}}_{n,\hat{\gamma}_n}$ from (20) and (21).

5 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 $r_n = \operatorname{rank}(\hat{\mathcal{I}}^t_{n,\hat{\gamma}_n})$.

(iii) Collect the accepted values for α to form $\hat{C}_{n,1-a}$.

The algorithm highlights that the computation costs for evaluating the score test, i.e. step (ii), 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, i.e. $L_\alpha = K(K-1)/2$. 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

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

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 5.1: *$g : D_g \to \mathbb{R}^{d_g}$ 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 5.1: *Consider the vector that collects all sIRF at horizon l*

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

where

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

In our general notation we have $d_g = K^2$ and we note that, given Assumption 2.1, 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*

$$J_{\gamma,1} := \left[(A(\alpha,\sigma)^{-1})' \otimes I_K\right] \left\{ \sum_{j=0}^{h-1} \left[D(\mathsf{B}(b)')^{h-1-j} \otimes (D\mathsf{B}(b)^j D')\right] \right\}$$
$$J_{\gamma,2} := \left[I_K \otimes D\mathsf{B}(b)^h D'\right] \nabla_\sigma \operatorname{vec}(A(\alpha,\sigma)^{-1}) \ .$$

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 utilised by Granziera et al. (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 4.1. 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

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

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 5.1: *Suppose that assumption 5.1 holds. Let $\hat{\beta}_n$ and $\hat{\Sigma}_n$ be sequences of estimates and $B^\star \subset \mathcal{B}$, $H^\star \subset \dot{\mathscr{H}}$ 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 $(b,h) \in B^\star \times H^\star$,*

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

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

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

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 2.1) 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 $g(\alpha, \beta)$, i.e. $\hat{C}_{n,g}$, 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

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

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

Corollary 5.1: *Let $\beta_n(b)$, $\theta_n(b,h)$ and $B^\star$, $H^\star$ be as in Proposition 5.1. If $\hat{C}_{n,1-q_1}$ satisfies (24) and $\hat{C}_{n,g,\alpha,1-q_2}$ satisfies (27), then*

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

This Corollary requires only the conclusions of Corollary 4.1 and Proposition 5.1.[20] For convenience we summarize the practical implementation in the following algorithm.

Algorithm 2: Robust confidence sets for smooth functions

(i) Obtain the confidence set $\hat{C}_{n,1-q_1}$ for α using **Algorithm 1**;

(ii) 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);

(iii) Construct $\hat{C}_{n,g}$ from (28).

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.

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 reduce the computational cost of **Algorithm 2**.

[20] These are proven under Assumptions 2.1 and 2.2 which, we re-iterate, do not impose that the structural shocks have non-Gaussian distributions.

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, e.g. the densities of the errors are non-Gaussian, such estimates have desirable efficiency properties as we document in this section.[21]

Assumption 6.1: *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 define the one-step efficient estimator as

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

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 discretised version of any $\sqrt{n}$–consistent estimator $\tilde{\gamma}_n = (\tilde{\alpha}_n, \tilde{\beta}_n)$. We note that under Assumption 6.1 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 6.1: *Suppose that Assumptions 2.1, 2.2 and 6.1 hold. Let $\tilde{\gamma}_n$ be a $\sqrt{n}$–consistent estimator of γ under P_θ^n. Let $\bar{\gamma}_n$ be a discretised version of $\tilde{\gamma}_n$ which which replaces its value with the closest point in $\mathscr{S}_n^* := n^{-1/2}C\mathbb{Z}^L$. Then,*

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

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 e.g. 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.

[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 $A(\alpha, \sigma)$.

Table 1: Distributions for Structural Shocks

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

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.

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 supplementary material Section S5.

7.1 Size of semi-parametric score test

We start by evaluating the empirical rejection frequencies of the score test $\hat{S}_{n,\hat{\gamma}_n}$ for the semi-parametric 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.

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

[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. The supplementary material Section S5 reports the results for the latter case.

burn-in periods to simulate data and, unless indicated differently, we use $M = 2,500$ Monte Carlo replications throughout the simulations.

Table 2 reports the empirical rejection frequencies of the semi-parametric score test defined in Section 4 for testing the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$. The test is implemented following steps 1-5 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 $L_\beta/n \to 0$.

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, i.e. 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 over-rejects 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 semi-parametric 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 (standardised) $t(7)$ density and a (standardised) $t(12)$ density for the second shock. The test is closely related

[24]Simulation evidence in Lee and Mesters (2023a) has shown that tests that do not fix α under the null often show severe over-rejection in static ICA models when the errors are close to Gaussian.

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 semi-parametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ 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 $M = 2,500$ Monte Carlo replications.

to the Wald test of Gouriéroux et al. (2017). We also consider the (psuedo –) likelihood ratio test ($\text{LR}^{\text{PML,t}}$). In addition, we consider two tests based on the work of Lanne and Luoto (2021) – the GMM Wald ($\text{W}^{\text{GMM,LL}}$) and likelihood ratio ($\text{LR}^{\text{GMM,LL}}$) tests which are based on higher (third & fourth) order moment conditions. We also include the closely related moment estimator from Keweloh (2021) for a Wald ($\text{W}^{\text{GMM,Kew}}$) and likelihood-ratio ($\text{LR}^{\text{GMM,Kew}}$) test.

In the second category we consider five tests. First, we have the pseudo maximum likelihood Lagrange Multiplier test ($\text{LM}^{\text{PML,t}}$) that is based on work of Gouriéroux et al. (2017). This test is based on the score of the pseudo log likelihood which we take, following Gouriéroux et al. (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] Secondly, we consider the LM test corresponding to the GMM setup of Lanne and Luoto (2021) ($\text{LM}^{\text{GMM,LL}}$). 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 non-parametric test which is based on Hoeffding (1948) and Blum et al. (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, similarly to our semi-parametric 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) — W^{PML}, W^{LL} and DM^{LL} — tend to perform very poorly, with the simulation results demonstrating both substantial over-rejection 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.

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 et al. (2017) (LM^{PML}): in the case with one Gaussian density, this test is able to control size for all choices of the second density considered. In the case where both shocks are drawn from the same distribution, this test is able to control size for most of the distributions, however over-rejects somewhat for the BM, SPB and TRI distributions. The LM test based on Lanne and Luoto (2021) (LM^{LL}) displays slightly worse performance, with over-rejections for about half of the distributions considered. Interestingly many of these over-rejections occur in the first panel, where we may expect that identification is somewhat stronger. The identification robust moment tests of Drautzburg and Wright (2023) (GMM^{DW} and BKR^{DW}) generally perform well, with the former always controlling size correctly

[25]Note that this test is not actually discussed in Gouriéroux et al. (2017), but the simulations in Lee and Mesters (2023a) 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).

TABLE 3: EMPIRICAL REJECTION FREQUENCIES FOR ALTERNATIVE TESTS

Test	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$\epsilon_{1,t} \sim \epsilon_{2,t}$										
$\hat{S}_{ols}$	5.3	5.7	6.1	5.7	4.4	5.7	4.0	3.7	4.8	3.9
$\hat{S}_{onestep}$	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
$\mathrm{BKR}^{\mathrm{DW}}$	3.8	3.9	4.3	4.9	5.5	28.5	5.9	5.5	7.9	6.1
$\mathrm{LM}^{\mathrm{PML,t}}$	5.1	5.1	5.7	5.3	4.9	6.8	16.6	22.2	5.4	21.7
$\mathrm{LM}^{\mathrm{GMM,LL}}$	1.8	1.5	4.2	12.1	16.2	10.6	3.5	3.7	2.3	3.9
$\mathrm{LM}^{\mathrm{GMM,Kew}}$	1.4	1.5	4.0	15.0	15.3	6.5	4.5	4.3	1.4	4.1
$\mathrm{LR}^{\mathrm{PML,t}}$	34.7	13.1	7.6	3.7	1.6	1.9	100.0	100.0	14.0	100.0
$\mathrm{LR}^{\mathrm{GMM,LL}}$	7.4	10.5	11.7	19.2	16.1	12.6	3.9	3.4	9.6	3.8
$\mathrm{LR}^{\mathrm{GMM,Kew}}$	9.8	10.3	13.7	20.0	16.9	12.6	4.6	4.7	9.4	4.3
$\mathrm{W}^{\mathrm{PML,t}}$	16.7	10.8	11.3	8.1	6.2	5.9	37.4	41.7	12.4	38.5
$\mathrm{W}^{\mathrm{GMM,LL}}$	20.5	24.5	23.5	27.2	22.2	17.9	4.4	4.8	22.8	4.5
$\mathrm{W}^{\mathrm{GMM,Kew}}$	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 $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$, and $\alpha_0 = 0.5594$. $\hat{S}_{ols}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. $\mathrm{LM}^{\mathrm{PML,t}}$, $\mathrm{W}^{\mathrm{PML,t}}$ and $\mathrm{LR}^{\mathrm{PML,t}}$ denote the pseudo-maximum likelihood tests based on Gouriéroux et al. (2017), assuming t-distributed shocks. $\mathrm{LM}^{\mathrm{GMM,LL}}$, $\mathrm{W}^{\mathrm{GMM,LL}}$ and $\mathrm{LR}^{\mathrm{GMM,LL}}$ denote the GMM-based tests based on Lanne and Luoto (2021) with one co-kurtosis condition based on $\epsilon_{1t}^3\epsilon_{2t}$. $\mathrm{LM}^{\mathrm{GMM,Kew}}$, $\mathrm{W}^{\mathrm{GMM,Kew}}$ and $\mathrm{LR}^{\mathrm{GMM,Kew}}$ 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 non-parametric 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 $M = 1,000$ Monte Carlo replications.

and the latter over-rejecting only in a few cases (e.g. the kurtotic unimodal distribution). This over-rejection is not due to identification failure but rather slow convergence due to the higher order moment conditions used.

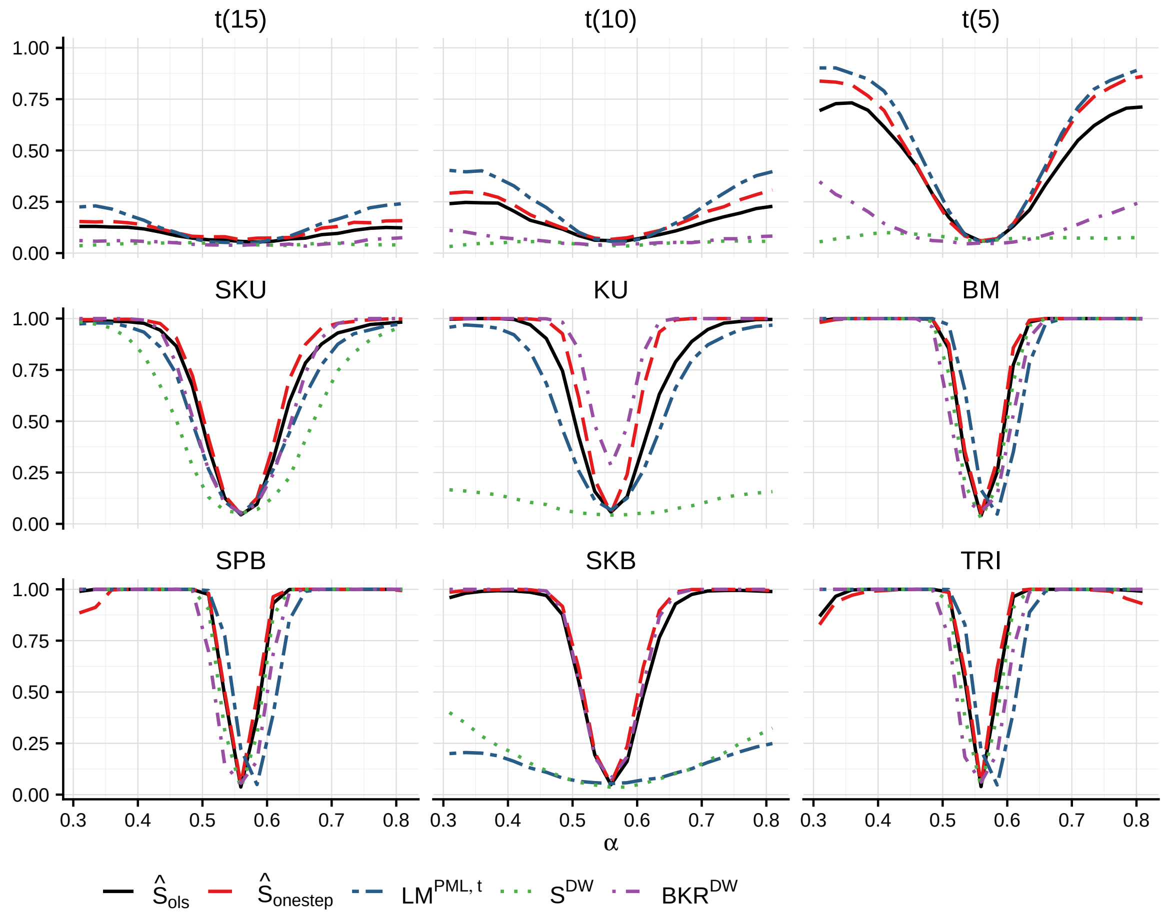
To summarize, most of the non-robust alternative procedures lead to incorrect inference if the distribution of the structural shocks is not "sufficiently" non-Gaussian. Furthermore, the identity of the best-performing alternative procedure crucially depends on which non-Gaussian distribution generated the data. In contrast, the semi-parametric 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 semi-parametric score test using one-step nuisance parameter estimates (red solid line), the semi-parametric score test using OLS nuisance parameter estimates (black sold line), the pseudo maximum likelihood LM

Figure 1: Power in the SVAR(1) model



Note: The figure reports unadjusted empirical power curves for tests of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ 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 $\alpha_0 = 0.5594$. $\hat{S}_{\text{ols}}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{\text{onestep}}$ uses one-step efficient estimates. $LM^{PML,t}$ denotes the pseudo-maximum likelihood test based on Gouriéroux et al. (2017), S^{DW} denotes the GMM-based test of Drautzburg and Wright (2023), BKR^{DW} denotes the non-parametric 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 $M = 1,000$ Monte Carlo replications.

test (dot - dashed blue line), the Drautzburg and Wright (2023) GMM test (dotted green line) and the non-parametric 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 non-parametric test has power curves which are not much below those of the efficient score test.[26]

7.3 Additional results

In the supplementary material we present 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 non-efficient competitors, i.e. 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]

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

[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.

[27] In the supplementary material, we provide additional results from an alternative parametrization of the model using a rotation matrix.

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 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.

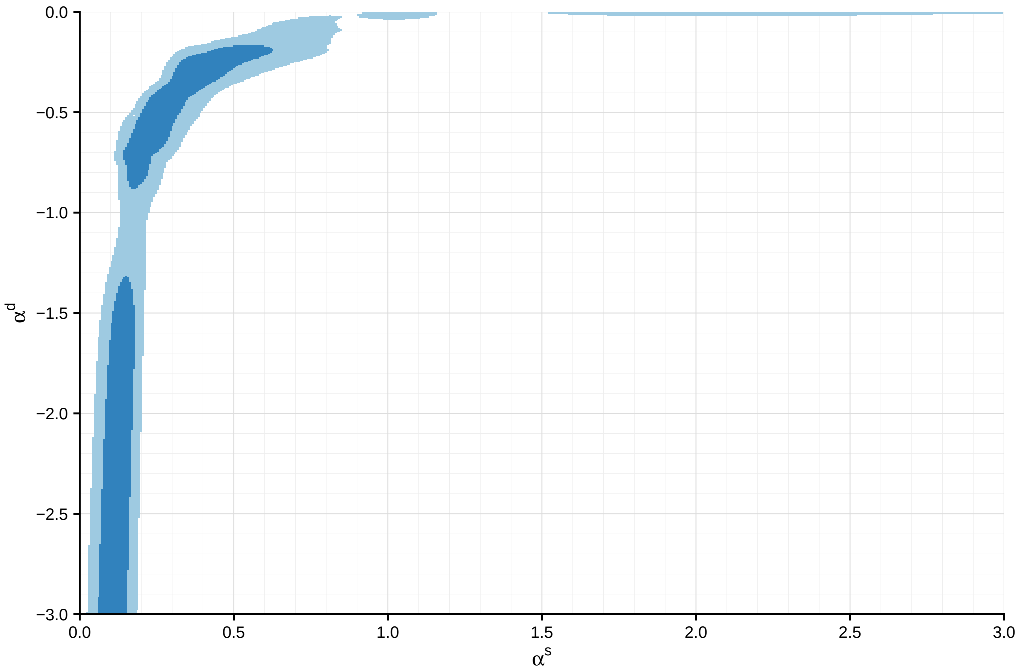
Recently, Lanne and Luoto (2022) adopted the methodology of Lanne and Luoto (2021) to assess identification of the model using non-Gaussianity, but this approach may yield incorrect coverage when the shocks are close to Gaussian (cf Section 7). Here we will adopt the robust score testing approach of 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 Montiel 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 Montiel 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 Section 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

Figure 2: Confidence Sets for Labor Demand and Supply Elasticities



Note: 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 $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$.

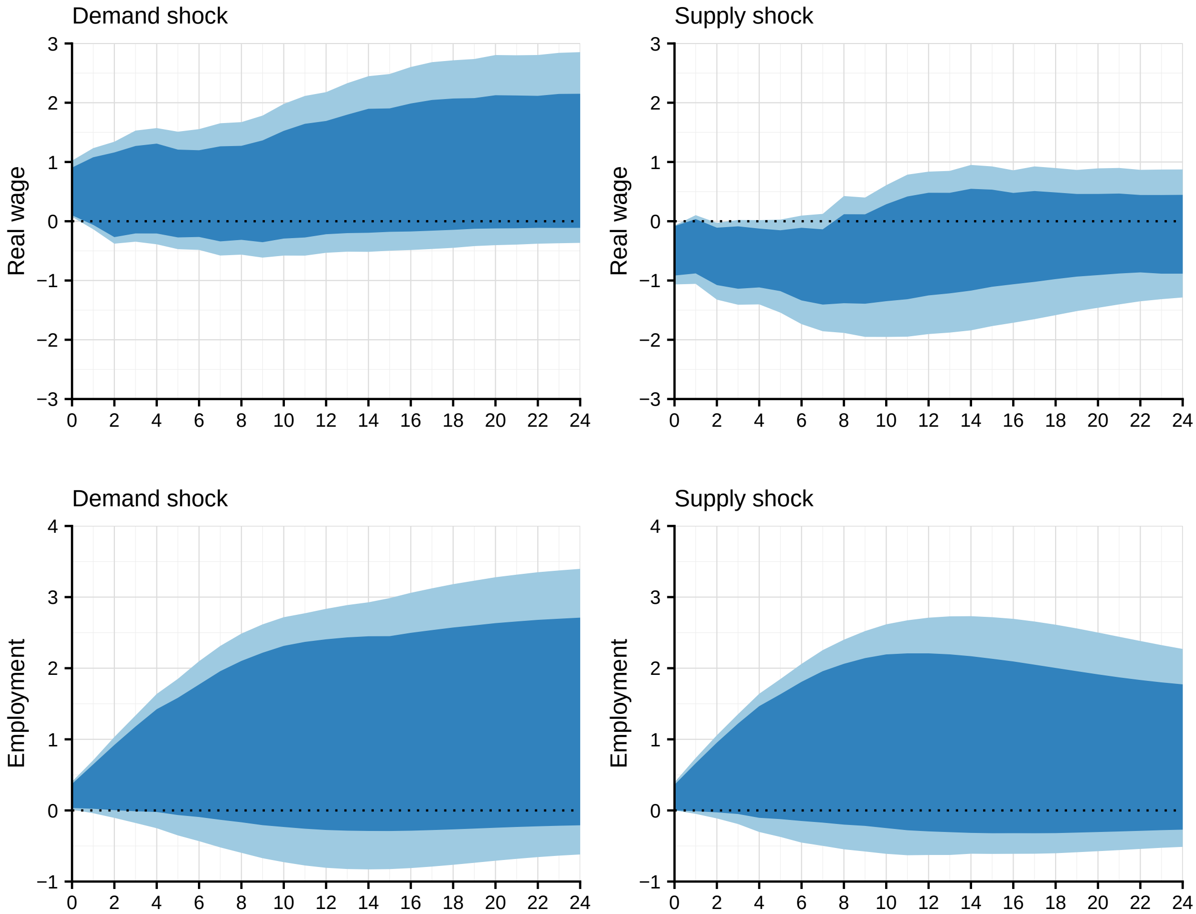
sets indicating that α^s lies in the 0 - 0.3 range for both 95% and 67% confidence level. In contrast, for values of α^d that are less negative (smaller absolute value), the confidence sets support a wide range of values for the supply elasticity, up to 0.6 at 67% confidence level and spanning almost 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

Figure 3: IRF confidence bands for labor demand and supply shocks



Note: 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 $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$.

Baumeister and Hamilton (2015)'s 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 tri-variate 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, 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 parameterised as follows

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

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 parametrisation 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]

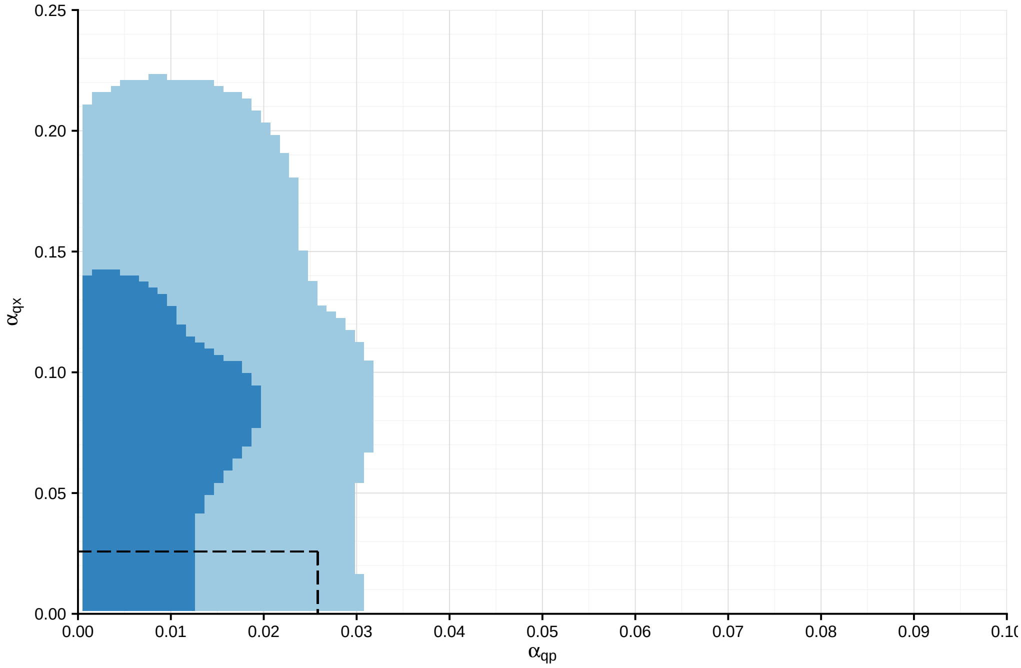
$$A^{-1}(\alpha, \sigma) = \begin{bmatrix} + & + & + \\ + & + & - \\ - & + & + \end{bmatrix} . \tag{32}$$

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 criticised 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 Montiel 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 Montiel Olea et al. (2022), hence we conclude this assumption is not unreasonable and proceed with constructing confidence sets for the elasticity parameters.

[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 $(\alpha_{qx}, \alpha_{qp})$



Note: 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).

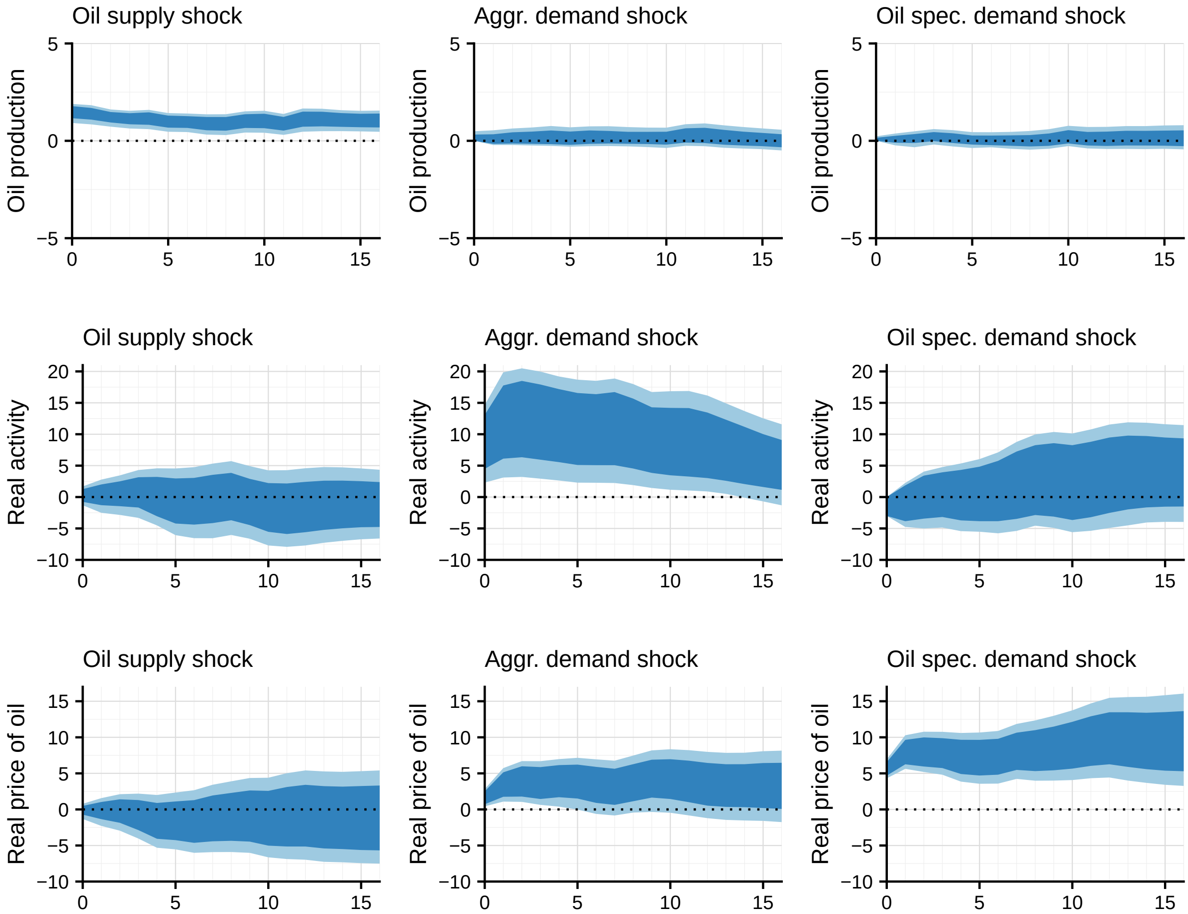
Confidence sets for oil supply elasticities $(\alpha_{qx}, \alpha_{qp})$

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 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

[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



Note: 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 $(\alpha_{qx}, \alpha_{qp}, \alpha_{xp}) \in (0, 0.25] \times (0, 0.1] \times [-3, 0)$.

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 can not 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).

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 semi-parametric model where the densities of the structural shocks form the non-parametric part and conduct inference on the possibly weakly identified or non identified parameters of the SVAR, using a semi-parametric 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 semi-parametric 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. Whilst 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 asses 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.

A.1 Notation

$x \coloneqq y$ 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: $Pf \coloneqq \int f \, \mathrm{d}P$, $\mathbb{P}_n f \coloneqq \frac{1}{n}\sum_{i=1}^n f(Y_i)$ and $\mathbb{G}_n f \coloneqq \sqrt{n}(\mathbb{P}_n - P)f$. 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, i.e. 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, i.e. $S^\perp = \{x \in V : \langle x, s\rangle = 0 \text{ for all } s \in S\}$. 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 $d_{TV}(P, Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)|$. $\rightsquigarrow$ denotes weak convergence.

A.2 Density score estimation

Lemma A.1: *Suppose Assumptions 2.1 and 2.2 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 $j, k = 1, \ldots, K$, $k \neq j$*

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

where $A_n \coloneqq A(\alpha_n, \beta_n)$, $B_n \coloneqq B(\beta_n)$ 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

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

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 (2023a); see Lee and Mesters (2023b) for the proof.[30] □

[30]Note that in the statement of Lemma 4 of Lee and Mesters (2023a) 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).

A.3 ULAN

To establish ULAN we establish LAN, as in Proposition A.1 directly below. Following this in Proposition A.2 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 A.1 (LAN): *Suppose that assumption 2.1 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 $n \to \infty$,*

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

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,

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

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

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

$\mathcal{F}_{n,t} := \sigma(Y_t, X_t)$ 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 2.1(ii),

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

almost surely, where the expectation is taken under P^n_θ.

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

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

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

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

We now verify conditions (1.2) – (1.6) of Lemma 1 in Swensen (1985), having shown (1.7) to hold above. (1.2), i.e. that $\mathbb{E}\sum_{t=1}^{n}[W_{n,t} - U_{n,t}]^2 \to 0$, where the expectation is taken under P^n_θ is shown to hold in Lemma S2.5. For (1.3) note that by Lemma S2.4, $P^n_\theta[|\sqrt{n}W_{n,t}|^{2+\rho}] \le C$ for some $\rho > 0$. Hence

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

For (1.4), by Lemma S2.4 and Markov's inequality,

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

(1.5) follows from Lemma S2.7. For (1.6), by Lemma S2.4 and the fact that conditional expectations are L_1 contractions we have for any $\varepsilon>0$

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

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)),

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

That this converges to τ^2 follows from Lemma S2.7 and the reverse triangle inequality. □

Proposition A.2: *Suppose that assumption 2.1 holds. Then, if $(g_n,h_n)\to(g,h)$,*

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

Proof. By Lemmas S2.8 and S2.9

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

whenever $(g_n,h_n)\to(g,h)$. Therefore, by Lemma S3.3, (i) $d_{TV}(P^n_{\theta_n(g_n,h_n)},P^n_{\theta_n(g_n,h)})\to0$ and (ii) $d_{TV}(P^n_{\theta_n(g_n,h)},P^n_{\theta_n(g,h)})\to0$. □

Proof of Proposition 3.1. The only conclusion of Proposition 3.1 which is not immediately implied by those of Proposition A.1 is that

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

By Proposition A.1,

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

and hence it suffices to show that

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

By Proposition A.2, $d_{TV}(P^n_{\theta_n(g_n,h_n)}, P^n_{\theta_n(g,h)}) \to 0$, 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 & Remark 18.3 in Strasser (1985)). By Proposition A.1 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 3.1. Combine Example 6.5 in van der Vaart (1998) with the fact that by Proposition 3.1, under P^n_θ

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

A.4 Scores

Proof of Lemma 3.1. 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) $\tilde{\ell}_\theta(Y_s,X_s) \in \left[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}\right]^\perp \subset L_2(P^n_\theta)$ (componentwise) and (b)

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

For (a), the fact that $\tilde{\ell}_\theta(Y_s,X_s) \in L_2(P^n_\theta)$ follows straightforwardly from its form and the moment conditions in assumption 2.1(ii). Next note that for any $h \in \dot{\mathscr{H}}$, $1 \le s \le n$,

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

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

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

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 $\mathbb{E}\left[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})\right] = \mathbb{E}\left[(X_s-\mu)\right]\mathbb{E}\left[\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})\right] = 0$. If $s > t$, then $\mathbb{E}\left[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})\right] = \mathbb{E}\left[(X_s-\mu)h_k(\epsilon_{k,t})\,\mathbb{E}\left[\phi_m(\epsilon_{m,s})|\sigma(\epsilon_1,\dots,\epsilon_{s-1})\right]\right] = 0$ again by independence.

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

$l = 1, \ldots, L_\sigma\}$ 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 2.1. 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 $x_l \in \{b_l : l = 1, \ldots, L_b\}$,

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

Again that this is mean zero and has finite second moment follows immediately from Assumption 2.1. 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 $\operatorname{cl}\dot{\mathscr{H}}$.[31] $\square$

A.5 Main Theorems

Proof of Theorem 4.1. Define

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

where $\gamma_\star := (\alpha, \beta_\star)$ and $\theta_\star := (\gamma_\star, \eta)$. By Corollary 3.1, $P_\theta^n \triangleleft\triangleright P^n_{\theta_n((0,b_n),0)}$ for any $b_n \to b \in \mathbb{R}^{L_\beta}$. It then follows by Lemmas S2.13, S2.15 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 S3.1 also

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

It follows that

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

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

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

[31] This follows by the argument of Lemma S8 in Lee and Mesters (2023b), noting that in the present context their $H_0, H_0^\star, \tilde{H}_0$ may be dropped.

We combine these to obtain

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

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

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

By (i) of Lemma S2.15 and Le Cam's third lemma (e.g. van der Vaart, 1998, Example 6.7)

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

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 non-singular matrix we obtain

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

Hence by (43) have $\left\|\hat{\mathcal{I}}_{n,\bar{\gamma}_n}-\tilde{\mathcal{I}}_{\theta}\right\|_2=o_{P_{\theta_n}^n}(\nu_n^{1/2})$. By Proposition S1 in Lee and Mesters (2023b)

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

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 e.g. 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

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

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 $\hat{\mathcal{I}}_{n,\bar{\gamma}_n}^t$ is the zero matrix, whose Moore-Penrose inverse is also the zero matrix. Hence on these sets we have $\hat{S}_{n,\bar{\gamma}_n}^{SR}=0$

and $c_n = 0$ and therefore do not reject, implying

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

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 A.2 and the fact that asymptotic uniform equicontinuity is implied by asymptotic equicontinuity on a compact set. □

Proof of Corollary 4.1. Apply Theorem 4.1 to conclude:

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

Proof of Proposition 5.1. By the uniform delta method (van der Vaart, 1998, Theorem 3.8), under $P^n_{\theta_n(b,h)}$,

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

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

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

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

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

The uniform statement then follows from Proposition A.2. □

Proof of Corollary 5.1. This follows directly from the hypotheses and the fact that

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

Proof of Theorem 6.1. Similarly to as in the Proof of Theorem 4.1, define

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

where $\theta_\star := (\gamma_\star, \eta)$. By Corollary 3.1, for any $g_n \to g \in \mathbb{R}^{L_\alpha + L_\beta}$, $P^n_\theta \lhd\rhd P^n_{\theta_n(g_n,0)}$. By Lemmas

S2.13, S2.15 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 S3.1 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

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

Combine this with Lemma S2.15 (i) and the continuous mapping theorem. □

References

Amari, S. and Cardoso, J.-F. (1997). Blind Source Separation - Semiparametric Statistical Approach. *IEEE Transactions On Signal Processing*, 45(11).

Andrews, I. and Mikusheva, A. (2015). Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics*, 6.

Andrews, I., Stock, J., and Sun, L. (2019). Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*, 11:727–753.

Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999.

Baumeister, C. and Hamilton, J. D. (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.

Bekaert, G., Engstrom, E., and Ermolov, A. (2020). Aggregate Demand and Aggregate Supply Effects of COVID-19: A Real-time Analysis. *Working paper.*

Bekaert, G., Engstrom, E., and Ermolov, A. (2021). Macro risks and the term structure of interest rates. *Journal of Financial Economics*, 141(2):479–504.

Bickel, P., Klaasen, C. A. J., Ritov, Y., and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York, NY, USA.

Bickel, P. J., Ritov, Y., and Stoker, T. M. (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann. Statist.*, 34(2):721–741.

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution Free Tests of Independence Based on the Sample Distribution Function. *The Annals of Mathematical Statistics*, 32(2):485 – 498.

Bonhomme, S. and Robin, J.-M. (2009). Consistent noisy independent component analysis. *Journal of Econometrics*, 149.

Braun, R. (2021). The importance of supply and demand for oil prices: evidence from non-gaussianity. *Working paper.*

Brunnermeier, M., Palia, D., Sastry, K. A., and Sims, C. A. (2021). Feedbacks: Financial markets and economic activity. *American Economic Review*, 111(6):1845–79.

Chen, A. and Bickel, P. J. (2006). Efficient Independent Component Analysis. *Annals of Statistics*, 34(6).

Choi, S., Hall, W. J., and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.*, 24(2):841–861.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, 36.

Davis, R. and Ng, S. (2022). Time Series Estimation of the Dynamic Effects of Disaster-Type Shocks. *Working paper.*

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

Dufour, J.-M. and Valery, P. (2016). Rank-robust regularized wald-type tests. *Working paper.*

Fiorentini, G. and Sentana, E. (2022). Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. *Journal of Econometrics*.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics*, 196.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2019). Identification and Estimation in Non-Fundamental Structural VARMA Models. *The Review of Economic Studies*, 87(4):1915–1953.

Granziera, E., Moon, H. R., and Schorfheide, F. (2018). Inference for vars identified with sign restrictions. *Quantitative Economics*, 9(3):1087–1121.

Guay, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics*, 225(1):27–46.

Hallin, M. and Saidi, A. (2007). Optimal tests of noncorrelation between multivariate time series. *Journal of the American Statistical Association*, 102(479):938–951.

Hallin, M. and Werker, B. J. M. (1999). Optimal testing for semi-parametric autoregressive models: From gaussian lagrange multipliers to regression rank scores and adaptive tests. In Ghosh, S., editor, *Asymptotics, Nonparametrics and Time Series*, pages 295 – 358. Marcel Dekker.

Herrera, A. M. and Rangaraju, S. K. (2020). The effect of oil supply shocks on us economic activity: What have we learned? *Journal of Applied Econometrics*, 35(2):141–159.

Herwartz, H. (2019). Long-run neutrality of demand shocks: Revisiting blanchard and quah (1989) with independent structural shocks. *Journal of Applied Econometrics*, 34(5):811–819.

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press, 2 edition.

Jin, K. (1992). Empirical smoothing parameter selection in adaptive estimation. *The Annals of Statistics*, pages 1844–1874.

Keweloh, S. A. (2021). A generalized method of moments estimator for structural vector autoregressions based on higher moments. *Journal of Business & Economic Statistics*, 39(3):772–782.

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

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

Kilian, L. and Murphy, D. P. (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.

Kilian, L. and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied econometrics*, 29(3):454–478.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4).

Lanne, M., Liu, K., and Luoto, J. (2022). Identifying structural vector autoregression via leptokurtic economic shocks. *Journal of Business & Economic Statistics*, 0(0):1–11.

Lanne, M. and Luoto, J. (2021). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, 39(1):69–81.

Lanne, M. and Luoto, J. (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.

Lanne, M. and Lütkepohl, H. (2010). Structural vector autoregressions with nonnormal residuals. *Journal of Business & Economic Statistics*, 28(1):159–168.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-Gaussian structual vector autoregressions. *Journal of Econometrics*, 196.

Le Cam, L. 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 Berkeley, Calif: University of California publications in statistics. University of California Press.

Le Cam, L. M. and Yang, G. L. (2000). *Asypmtotics in Statistics: Some Basic Concepts*. Springer, New York, NY, USA, 2 edition.

Lee, A. (2022). Robust and efficient inference for non-regular semiparametric models. *Working paper*.

Lee, A. and Mesters, G. (2023a). Robust inference for non-gaussian linear simultaneous equations models. *Working Paper*.

Lee, A. and Mesters, G. (2023b). Supplement to "robust inference for non-gaussian linear simultaneous equations models". *Working Paper*.

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

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics*, pages 712–736.

Matteson, D. S. and Tsay, R. S. (2017). Independent component analysis via distance covariance. *Journal of the American Statistical Association*, 112(518):623–637.

Maxand, S. (2020). Identification of independent structural shocks in the presence of multiple gaussian components. *Econometrics and Statistics*, 16:55–68.

Moneta, A., Entner, D., Hoyer, P. O., and Coad, A. (2013). Causal inference by independent component analysis: Theory and applications*. *Oxford Bulletin of Economics and Statistics*, 75(5):705–730.

Montiel Olea, J. L., Plagborg-Møller, M., and Qian, E. (2022). Svar identification from higher moments: Has the simultaneous causality problem been solved? *Prepared for AEA Papers and Proceedings.*

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2):99–135.

Pinkse, J. and Schurter, K. (2021). Estimates of derivatives of (log) densities and related objects. *Econometric Theory*, page 136.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications*. John Wiley & Sons, Inc., New York, NY, USA.

Rose, M. (1957). *Elementary Theory of Angular Momentum*. John Wiley & Sons.

Sen, A. (2012). On the Interrelation Between the Sample Mean and the Sample Variance. *The American Statistician*, 66(2).

Shapiro, M. D. and Watson, M. W. (1988). Sources of business cycle fluctuations. *NBER Macroeconomics Annual*, 3:111–148.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(72):2003–2030.

Sims, C. A. (2021). Svar identification through heteroskedasticity with misspecified regimes. *working paper*.

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5).

Strasser, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. De Gruyter studies in mathematics. W. de Gruyter.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis*, 16(1):54–70.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer.

Tank, A., Fox, E. B., and Shojaie, A. (2019). Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series. *Biometrika*, 106(2):433–452.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York, NY, USA, 1st edition.

van der Vaart, A. W. (2002). Semiparametric statistics. In Bernard, P., editor, *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999*. Springer, Berlin, Germany.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition.

Wang, Y. S. and Drton, M. (2019). High-dimensional causal discovery under non-Gaussianity. *Biometrika*, 107(1):41–59.

Zhou, X. (2020). Refining the workhorse oil market model. *Journal of Applied Econometrics*, 35(1):130–140.

Supplementary Material for: Locally Robust Inference for Non-Gaussian SVAR models

Lukas Hoesch*

Adam Lee†

Geert Mesters‡

This draft: December 20, 2023

Abstract

In this supplementary material we provide the following additional results.

S1: Choice for the parametrization

S2: Technical details for the main proofs

S3: Some technical tools

S4: Details log density score estimation

S5: Additional simulation results

S6: Additional empirical results

*Vrije Universiteit Amsterdam and Tinbergen Institute; l.hoesch@vu.nl

†BI Norwegian Business School; adam.lee@bi.no

‡Universitat Pompeu Fabra, Barcelona School of Economics and CREI; geert.mesters@upf.edu

S1 Parametrization of the semi-parametric SVAR model

Under the main assumptions of the paper (i.e. Assumptions 2.1 and 2.2) the parameters of the SVAR are generally not locally identified. Even under the *additional* assumption that the errors $\epsilon_{k,t}$ follow non-Gaussian distributions, we have that $A(\alpha, \sigma)$ can only be identified up to permutation and sign changes of its rows (e.g. Comon, 1994).

Therefore, to ensure that we study economically interesting permutations we typically need to impose additional identifying restrictions, such as zero or sign restrictions. The choice for such restrictions interacts with the chosen parametrization for $A(\alpha, \sigma)$ for which we give a few examples.

Example S1.1 (Supply and demand): *Following Baumeister and Hamilton (2015), when the SVAR defines a demand and a supply equation we can set*

$$A^{-1}(\alpha, \sigma) = \begin{pmatrix} -\alpha^d & 1 \\ -\alpha^s & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} , \tag{S1}$$

where $\alpha = (\alpha^d, \alpha^s)'$ are the short run demand and supply elasticities, and $\sigma = (\sigma_1, \sigma_2)'$ scales the structural shocks. With independent non-Gaussian errors A is identified up to permutation and sign changes of its rows. To pin down an economically interesting rotation we can impose the sign restrictions $\alpha^d \leq 0$, $\alpha^s \geq 0$ and $\sigma_1, \sigma_2 > 0$.

Example S1.2 (Rotation matrix): *A canonical choice sets*

$$A^{-1}(\alpha, \sigma) = \Sigma^{1/2}(\sigma) R(\alpha) , \tag{S2}$$

where $\Sigma^{1/2}(\sigma)$ is a lower triangular matrix (with positive diagonal elements) defined by the vector σ and $R(\alpha)$ is a rotation matrix that is parametrized by the vector α. Different parametrizations for the rotation matrix are possible, see Magnus et al. (2021) for a detailed discussion. Similar to in Example S1.1, even with independent non-Gaussian errors $R(\alpha)$ is not uniquely identified and additional zero-, sign-, or long-run-restrictions are needed to pin down the desired rotation.

As the above examples make clear, several commonly used parametrizations can be adopted. Three general comments apply.

First, pinning down a specific permutation, as in the first example, is necessary for the economic interpretation of the results, but it is not necessary for the score testing methodology of the paper which fixes α under the null.

Second, the robust non-Gaussian approach of this paper can be combined with any of the existing SVAR identification approaches to obtain an economically interesting specification. Besides zero and sign restrictions one can also think of combining with external instruments or more general prior information as in Baumeister and Hamilton (2015) or Braun (2021).

Third, often multiple parametrizations are possible. We recommend jointly testing the possibly weakly identified parameters when they are of direct economic interest (e.g. Example 1). In contrast, when the interest is in more general functions, such as impulse responses or forecast error variances, we suggest to parameterize A such that α is as low-dimensional as possible, e.g. via the rotation matrix specification as in Example 2. In this way the Bonferroni procedure of **Algorithm 2** can be executed over the smallest possible grid for α, which reduces the computational burden.

S2 Technical details for the main proofs

Here we establish some technical details utilised in the proofs in section A of the main text.

S2.1 Markov structure

Define $Z_t := (Y_t', Y_{t-1}', \ldots, Y_{t-p+1}')'$, $\mathsf{C}_\theta := (c_\theta', 0', \ldots, 0')'$,

$$\mathsf{B}_\theta := \begin{bmatrix} B_{\theta,1} & B_{\theta,2} & \cdots & B_{\theta,p-1} & B_{\theta,p} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \quad \mathsf{D}_\theta := \begin{bmatrix} A_\theta^{-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and note that we can write

$$Z_t = \mathsf{C}_\theta + \mathsf{B}_\theta Z_{t-1} + \mathsf{D}_\theta \epsilon_t. \tag{S3}$$

This can be re-written in de-meaned form as

$$\tilde{Z}_t = \mathsf{B}_\theta \tilde{Z}_{t-1} + \mathsf{D}_\theta \epsilon_t \tag{S4}$$

with $\tilde{Z}_t := Z_t - m_\theta$, for $m_\theta := \left(\sum_{i=0}^{\infty} \mathsf{B}_\theta\right) \mathsf{C}_\theta = (I - \mathsf{B}_\theta)^{-1}\mathsf{C}_\theta$.

Lemma S2.1: *Suppose that assumption 2.1 holds. Define $U_{\theta,t}$ as the (unique, strictly) stationary* *solution to* (S3)*. Then $U_{\theta,t}$ has the representation*

$$U_{\theta,t} = m_\theta + \sum_{j=0}^{\infty} \mathsf{B}_\theta^j \mathsf{D}_\theta \epsilon_{t-j}, \quad m_\theta := (I - \mathsf{B}_\theta)^{-1}\mathsf{C}_\theta, \quad \sum_{j=0}^{\infty} \|\mathsf{B}_\theta^j\| < \infty.$$

If ρ_θ is the largest absolute eigenvalue of the companion matrix B_θ and $\upsilon > 0$ is such that $\rho_\theta + \upsilon < 1$, then

$$\mathbb{E}\,\|U_{\theta,t} - m_\theta\|^{\mathsf{p}} \leq \frac{\mathbb{E}\,\|\mathsf{D}_\theta \epsilon_t\|^{\mathsf{p}}}{1 - (\rho_\theta + \upsilon)^{\mathsf{p}}}, \quad \mathsf{p} \in [1, 4+\delta].$$

Proof. Rewriting (S3) as (S4) and applying Theorem 11.3.1 in Brockwell and Davis (1991) yields the first part. For the second part,

$$\|U_{\theta,t} - m_\theta\| \leq \sum_{j=0}^{\infty} \|\mathsf{B}_\theta^j\| \|\mathsf{D}_\theta \epsilon_{t-j}\| \leq \sum_{j=0}^{\infty} \|\mathsf{B}_\theta\|^j \|\mathsf{D}_\theta \epsilon_{t-j}\| \leq \sum_{j=0}^{\infty} (\rho_\theta + \nu)^j \|\mathsf{D}_\theta \epsilon_{t-j}\|.$$

Since $\mathbb{E}\,\|\mathsf{D}_\theta \epsilon_{t-j}\|^{\mathsf{p}} = \mathbb{E}\,\|\mathsf{D}_\theta \epsilon_t\|^{\mathsf{p}} < \infty$ for all $t \in \mathbb{N}$, all $j \geq 0$ and $\rho \in [1, 4+\delta]$, it follows that

$$\mathbb{E}\,\|U_{\theta,t} - m_\theta\|^{\mathsf{p}} \leq \sum_{j=0}^{\infty} (\rho_\theta + \nu)^{j\mathsf{p}}\, \mathbb{E}\,\|\mathsf{D}_\theta \epsilon_{t-j}\|^{\mathsf{p}} = \frac{\mathbb{E}\,\|\mathsf{D}_\theta \epsilon_t\|^{\mathsf{p}}}{1 - (\rho_\theta + \nu)^{\mathsf{p}}}. \qquad \square$$

LEMMA S2.2: *Let $Q_{n,\theta}$ be the probability measure corresponding to $\bar{q}_{n,\theta} := \frac{1}{n}\sum_{t=1}^{n} q_{\theta,t}$, where $q_{\theta,t}$ is the density of X_t under P_θ^n $(1 \leq t \leq n)$.*[S1] *Then $Q_{n,\theta} \xrightarrow{TV} Q_\theta$, where Q_θ is the distribution of the (unique, strictly) stationary solution to* (1).

Proof. By Lemma S2.1, (S4) has a (unique, strictly) stationary solution with finite second moments. Applying Theorem 2 in Saikkonen (2007) gives that the Markov chain $(\tilde{Z}_t)$ is V-geometrically ergodic with $\mathsf{V}(x) = 1 + \|x\|^2$. That is, for an invariant probability measure $\tilde{\pi}_\theta$, some $r \in (1, \infty)$ and some $R < \infty$

$$\sum_{n=1}^{\infty} r^n \|\tilde{P}_\theta^n(\cdot, \tilde{z}) - \tilde{\pi}_\theta\|_{TV} \leq \sum_{n=1}^{\infty} r^n \|\tilde{P}_\theta^n(\cdot, \tilde{z}) - \tilde{\pi}_\theta\|_{\mathsf{V}} \leq R\mathsf{V}(\tilde{z}) = R(\|\tilde{z}\|^2 + 1) < \infty, \tag{S5}$$

where $\tilde{P}_\theta^n(\cdot, \tilde{z})$ is the n-step transition probability and $\tilde{z}$ is the initial condition.[S2] $\tilde{\pi}_\theta$ is the distribution of $U_{\theta,t} - m_\theta$ as defined in Lemma S2.1 (Kallenberg, 2021, Theorem 11.11).

Let $f_\theta : \mathbb{R}^{Kp} \to \mathbb{R}^K$ be defined as

$$f_\theta(x) := \begin{bmatrix} I_K & 0_{K \times K(p-1)} \end{bmatrix} (x + m_\theta),$$

[S1] Here, and throughout the appendix, any reference to the density of X_t is to be understood as to the density of the non-deterministic parts of X_t.

[S2] The norm $\|\nu\|_{\mathsf{V}}$ is defined by $\|\nu\|_{\mathsf{V}} := \sup_{f \leq \mathsf{V}} \left|\int f \,\mathrm{d}\nu\right|$ where the supremum is taken over all measurable functions dominated by V for any probability measure ν.

i.e. the function which adds m_θ to its argument and then returns the first K elements. The distribution of X_t under P^n_θ (given the initial condition $\tilde{z}$) is then $Q^{t-1}_\theta(\cdot, \tilde{z}) = \tilde{P}^{t-1}_\theta(\cdot, \tilde{z}) \circ f^{-1}_\theta$, i.e. the pushforward of $\tilde{P}^{t-1}_\theta(\cdot, \tilde{z})$ under f_θ. Henceforth we shall omit the $\tilde{z}$ in the notation. Similarly let $Q_\theta = \tilde{\pi}_\theta \circ f^{-1}_\theta$, i.e.the pushforward of $\tilde{\pi}_\theta$ under f. That Q_θ is the distribution of the (unique, strictly) stationary solution to (1) can be seen by noting that the first K elements of $U_{\theta,t}$ form a (strictly) stationary time series and satisfy the defining equation (1); by Theorem 11.3.1 in Brockwell and Davis (1991) it is therefore the unique solution. Then by (S5),

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{t=1}^{n} Q^t_\theta - Q_\theta \right\|_{TV} &\leq \frac{1}{n} \sum_{t=1}^{n} \left\| Q^t_\theta - Q_\theta \right\|_{TV} \\
&\leq \frac{1}{n} \sum_{t=1}^{n} \left\| \tilde{P}^{t-1}_\theta - \tilde{\pi}_\theta \right\|_{TV} \\
&\leq \frac{1}{n} \sum_{t=1}^{n} \left\| \tilde{P}^{t}_\theta - \tilde{\pi}_\theta \right\|_{TV} + o(1) \\
&\to 0.
\end{aligned}$$

□

S2.2 Moment bounds

LEMMA S2.3: *Suppose that assumption 2.1 holds. Then for any sequence $\theta_n = (\gamma + g_n/\sqrt{n}, \eta)$ with $g_n \to g \in \mathbb{R}^L$, for some $\rho > 0$, under $P^n_{\theta_n}$*

(i) $\sup_{n\in\mathbb{N}} \mathbb{E}\left[\|\dot{\ell}_{\theta_n}\|^{2+\rho} \right] < \infty$;

(ii) $\sup_{n\in\mathbb{N}} \mathbb{E}\left[\|\tilde{\ell}_{\theta_n}\|^{2+\rho} \right] < \infty$.

Proof. Since the deterministic terms in $\dot{\ell}_{\theta_n}$ and $\tilde{\ell}_{\theta_n}$ are either constants or continuous functions of γ (by Assumption 2.1(iii)), they are uniformly bounded, since $\{\gamma + g_n/\sqrt{n} : n \in \mathbb{N}\} \cup \{\gamma\}$ is compact. It is therefore sufficient to show that under $P^n_{\theta_n}$, each of

$$\sup_{n\in\mathbb{N}, 1\leq t\leq n} \mathbb{E}\left[|A(\theta_n)_{k\bullet} V_{\theta_n,t}|^{4+\delta} \right], \sup_{n\in\mathbb{N}, 1\leq t\leq n} \mathbb{E}\left[|\phi_k(A(\theta_n)_{k\bullet} V_{\theta_n,t})|^{4+\delta} \right], \sup_{n\in\mathbb{N}, 1\leq t\leq n} \mathbb{E}\left[\|X_t\|^{4+\delta} \right],$$

is finite. Since under $P^n_{\theta_n}$, each $A(\theta_n)_{k\bullet} V_{\theta_n,t} \sim \eta_k$, finiteness of the first two follow directly from Assumption 2.1(ii). For the third, recurse equation (S3) backwards under $\theta = \theta_n$, to obtain

$$Z_t = \sum_{j=0}^{t-1} \mathsf{B}^j_{\theta_n} \mathsf{C}_{\theta_n} + \sum_{j=0}^{t-1} \mathsf{B}^j_{\theta_n} \mathsf{D}_{\theta_n} \epsilon_{t-j} + \mathsf{B}^t_{\theta_n} Z_0.$$

Each of B_θ, C_θ, D_θ (depend on θ only through γ and) are continuous functions of γ, hence

$$\varrho \coloneqq \sup_{n\in\mathbb{N}} \|\mathsf{B}_{\theta_n}\|_2 < 1, \; \sup_{n\in\mathbb{N}} \|\mathsf{C}_{\theta_n}\|_2 < C_1, \; \sup_{n\in\mathbb{N}} \|\mathsf{D}_{\theta_n}\|_2 < C_2,$$

where the first is due to Assumption 2.1(i). Since we condition on Z_0, by Assumption 2.1(ii),

$$\mathbb{E}\,\|Z_t\|^{4+\delta} \lesssim \left(\frac{C_1}{1-\varrho}\right)^{4+\delta} + \left(\frac{C_2}{1-\varrho}\right)^{4+\delta} \mathbb{E}\,|\epsilon_1|^{4+\delta} + \|Z_0\|^{4+\delta} < \infty. \tag{S6}$$

As the bound on the right hand side is independent of t or n, the claim follows. □

LEMMA S2.4: *Let $W_{n,t}$ be as in the Proof of Proposition A.1 and suppose the conditions of that Proposition hold. Then, $P^n_\theta[|\sqrt{n}W_{n,t}|^{2+\rho}]$ is uniformly bounded for some $\rho > 0$. In consequence, under P^n_θ, $W_{n,t}$ satisfies:*

$$\lim_{n\to\infty} \sum_{t=1}^n \mathbb{E}\left[W^2_{n,t}\mathbf{1}\{|\sqrt{n}W_{n,t}| > \varepsilon\sqrt{n}\}\right] = 0, \qquad \textit{for any } \varepsilon > 0. \tag{S7}$$

Proof. Uniform boundedness of $P^n_\theta[|\sqrt{n}W_{n,t}|^{2+\rho}]$ implies:

$$\lim_{n\to\infty} \sum_{t=1}^n W^{2+\rho}_{n,t} = 0,$$

which in turns implies (S7) (cf. Billingsley, 1995, page 362). For the uniform boundedness, as

$$2\sqrt{n}W_{n,t} = g'\dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^K h_k(A_{k\bullet}(\alpha,\sigma)V_{\theta,t}),$$

and the h_k are bounded, it suffices to note that by Lemma S2.3 $\mathbb{E}[(g'\dot{\ell}_\theta(X_t, Y_t))^{2+\rho}] \le C$ under P^n_θ for some $\rho > 0$. □

S2.3 Log-likelihood ratios

LEMMA S2.5 (DQM): *Suppose that assumption 2.1 holds. Then with $W_{n,t}$ and $U_{n,t}$ defined as in the proof of Proposition A.1,*

$$\lim_{n\to\infty} \mathbb{E}\sum_{t=1}^n (W_{n,t} - U_{n,t})^2 = 0,$$

where the expectation is taken under P^n_θ.

Proof. We argue similarly to Lemma 7.6 in van der Vaart (1998). Let $V_{\theta,t} \coloneqq Y_t - BX_t$ and

$\varphi(v) = (g, \eta_1 h_1, \ldots, \eta_K h_K)$ for $v = (g, h)$ with $g \in \mathbb{R}^L$, $h \in \dot{\mathscr{H}}$. Let

$$
\begin{aligned}
p_\theta(Y_t, X_t) &:= |A(\theta)| \prod_{k=1}^K \eta_k(A_{k\bullet}(\theta) V_{\theta,t}) \\
s_{\theta,u}(Y_t, X_t) &:= g' \dot{\ell}_{\theta+u\varphi(v)}(Y_t, X_t) + \sum_{k=1}^K \frac{h_k(A_{k\bullet}(\theta + u\varphi(v)) V_{\theta+u\varphi(v),t})}{1 + u h_k(A_{k\bullet}(\theta + u\varphi(v)) V_{\theta+u\varphi(v),t})} \\
&\quad + \sum_{k=1}^K \frac{u h_k'(A_{k\bullet}(\theta + u\varphi(v)) V_{\theta+u\varphi(v),t}) \left[\mathsf{D}_{1,k,u} V_{\theta+u\varphi(v),t} + \mathsf{D}_{2,k,u} X_t\right]}{1 + u h_k(A_{k\bullet}(\theta + u\varphi(v)) V_{\theta+u\varphi(v),t})},
\end{aligned}
$$

with

$$
\begin{aligned}
\mathsf{D}_{1,k,u} &:= e_k' \sum_{l=1}^{L_\alpha} g_{\alpha,l} D_{\alpha,l}(\theta + u\varphi(v)) + e_k' \sum_{l=1}^{L_\sigma} g_{\sigma,l} D_{\sigma,l}(\theta + u\varphi(v)) \\
\mathsf{D}_{2,k,u} &:= -A_{k\bullet}(\theta + u\varphi(v)) \sum_{l=1}^{L_b} D_{b,l}(\theta + u\varphi(v)).
\end{aligned}
$$

By Assumption 2.1 and standard computations, the derivative of $u \mapsto \sqrt{p_{\theta+u\varphi(v)}}$ at $u = \mathsf{u}$ is $\frac{1}{2} s_{\theta,\mathsf{u}} \sqrt{p_{\theta+\mathsf{u}\varphi(v)}}$ (everywhere). Inspection reveals that this is continuous in u.

For $q_{\theta,t}$ the density of X_t under P_θ^n and $s_\theta := s_{\theta,0}$,

$$
\begin{aligned}
\mathbb{E} \sum_{t=1}^n (W_{n,t} - U_{n,t})^2 &= \frac{1}{n} \sum_{t=1}^n \int \left(\sqrt{n} \left[\sqrt{\frac{p_{\theta_n}}{p_\theta}} - 1 \right] - \frac{1}{2} s_\theta \right)^2 p_\theta q_{\theta,t} \, \mathrm{d}\lambda \\
&= \int \left(\sqrt{n} \left[\sqrt{p_{\theta_n}} - \sqrt{p_\theta} \right] - \frac{1}{2} s_\theta \sqrt{p_\theta} \right)^2 \bar{q}_{n,\theta} \, \mathrm{d}\lambda,
\end{aligned}
$$

with $\bar{q}_{n,\theta} := \frac{1}{n} \sum_{t=1}^n q_{\theta,t}$. The integrand converges to zero as $n \to \infty$ by the differentiability of $u \mapsto \sqrt{p_{\theta+u\varphi(v)}}$ at $u = 0$.[S3] Let

$$
I_{\theta,u,n} := \int s_{\theta,u}^2 \, p_{\theta+u\varphi(v)} \, \bar{q}_{n,\theta} \, \mathrm{d}\lambda = \int s_{\theta,u}^2 \, \mathrm{d}G_{\theta,u,n},
$$

where $G_{\theta,u,n}$ is the distribution of (Y_t, X_t) corresponding to the density $p_{\theta+u\varphi(v)} \bar{q}_{n,\theta}$. By Lemma S3.2 $G_{\theta,u/\sqrt{n},n} \xrightarrow{TV} G_\theta$, defined by

$$
G_\theta(A) := \int_A p_\theta \, \mathrm{d}(\lambda(y) \otimes Q_\theta(x)).
$$

For any $(u_n) \subset [0,1]$ we have that $s_{\theta,u_n/\sqrt{n}}^2 \to s_\theta^2$ (pointwise). By Lemma S2.6 and Corollary

[S3] Note that $p_{\theta_n} = p_{\theta_n(g,h)} = p_{\theta+\varphi(v)/\sqrt{n}}$.

2.9 in Feinberg et al. (2016), $\lim_{n\to\infty} I_{\theta,u_n/\sqrt{n},n} = \int s_\theta^2 \,\mathrm{d}G_\theta < \infty$ and hence

$$\left|\int_0^1 I_{\theta,u/\sqrt{n},n}\,\mathrm{d}u - \int_0^1 \int s_\theta^2 \,\mathrm{d}G_\theta \,\mathrm{d}u\right| \leq \sup_{u\in[0,1]} \left| I_{\theta,u/\sqrt{n},n} - \int s_\theta^2 \,\mathrm{d}G_\theta\right| \to 0.$$

By absolute continuity, Jensen's inequality and the Fubini – Tonelli theorem,

$$\int \left(\sqrt{n}\left[\sqrt{p_{\theta_n}} - \sqrt{p_\theta}\right]\right)^2 \bar{q}_{n,\theta}\,\mathrm{d}\lambda \leq \frac{1}{4}\int\int_0^1 \left(s_{\theta,u/\sqrt{n}}\sqrt{p_{\theta+u\varphi(v)/\sqrt{n}}}\right)^2 \bar{q}_{n,\theta}\,\mathrm{d}u\,\mathrm{d}\lambda \leq \int_0^1 I_{\theta,u/\sqrt{n},n}\,\mathrm{d}u.$$

Combine these observations with Proposition 2.29 in van der Vaart (1998). □

Lemma S2.6: *Suppose that assumption 2.1 holds. Let $s_{\theta,u}$ and $G_{\theta,u,n}$ be as in the proof of Proposition S2.5. Then for any $(u_n)_{n\in\mathbb{N}} \subset [0,1]$, $s^2_{\theta,u_n/\sqrt{n}}$ is asymptotically uniformly $G_{\theta,u_n/\sqrt{n},n}$-integrable and $s_\theta \in L_2(G_\theta)$.*

Proof. That $s_\theta \in L_2(G_\theta)$ follows from the moment bounds in Assumption 2.1(ii), the boundedness of the h_k, the form of $\dot{\ell}_\theta$ given in equations (7) – (9) and Lemma S2.1 given that Q_θ is the law of the stationary solution to (1).

For the uniform integrability, let $\vartheta_n := \theta + u_n\varphi(v)/\sqrt{n} \to \theta$ and

$$\begin{aligned}
s_{\vartheta_n,1}(Y_t,X_t) &:= g'\dot{\ell}_{\vartheta_n}(Y_t,X_t)\\
s_{\vartheta_n,2}(Y_t,X_t) &:= \sum_{k=1}^K \frac{h_k(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t})}{1+u_n h_k(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t})/\sqrt{n}}\\
s_{\vartheta_n,3}(Y_t,X_t) &:= \sum_{k=1}^K \frac{u_n h_k'(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t})\left[\mathsf{D}_{1,k,u_n/\sqrt{n}}V_{\vartheta_n,t} + \mathsf{D}_{2,k,u_n/\sqrt{n}}X_t\right]/\sqrt{n}}{1+u_n h_k(A_{k\bullet}(\vartheta_n)V_{\vartheta_n,t})/\sqrt{n}}
\end{aligned}$$

It suffices to show that under $G_{\theta,u_n/\sqrt{n},n}$ each $s_{\vartheta_n,i}$ $(i=1,2,3)$ has uniformly bounded $2+\rho$ moments for some $\rho>0$ for all sufficiently large n.

We start with $s_{\vartheta_n,2}$: since each h_k is bounded, for all large enough n, each numerator is uniformly bounded above and each denominator is uniformly bounded below, away from zero. Thus there is a M such that $|s_{\vartheta_n,2}(Y_t,X_t)| \leq M$ for all such n.

For $s_{\vartheta_n,3}$, by assumption 2.1 part (iii), each $\mathsf{D}_{1,k,u_n/\sqrt{n}}$ and $\mathsf{D}_{2,k,u_n/\sqrt{n}}$ are uniformly bounded for all large enough n; the same is true of $\|A(\vartheta_n)^{-1}\|_2$. Using this, the fact that $V_{\vartheta_n,t} = A(\vartheta_n)^{-1}\epsilon_t$ and arguing similarly to as in the preceding paragraph we have that for some M and all large enough n, $|s_{\vartheta_n,3}(Y_t,X_t)| \leq M\left[\|\epsilon_t\| + \|X_t\|\right]$. Thus it is enough to verify that

$$\sup_{n\geq N, 1\leq t\leq n} G_{\theta,u_n/\sqrt{n},n}\|\epsilon_t\|^{4+\delta} < \infty, \qquad \sup_{n\geq N,1\leq t\leq n} G_{\theta,u_n/\sqrt{n},n}\|X_t\|^{4+\delta} < \infty. \tag{S8}$$

Under $G_{\theta,u_n/\sqrt{n},n}$, the elements $\epsilon_{t,k}$ are (independently across k) distributed according to $\eta_k(1+u_n h_k/\sqrt{n})$, so there are $c, C < \infty$ such that

$$G_{\theta,u_n/\sqrt{n},n}\|\epsilon_t\|^{4+\delta} \leq G_{\theta,u_n/\sqrt{n},n}\left[\sum_{k=1}^K \epsilon_{t,k}^2\right]^{\frac{4+\delta}{2}} \leq c\sum_{k=1}^K\left[\left(1+\frac{\bar{h}_k}{\sqrt{n}}\right)\int |x_k|^{4+\delta}\eta_k(x_k)\,\mathrm{d}x_k\right] \leq C,$$

where $|h_k(x)| \leq \bar{h}_k$. By arguing analogously to as in in Lemma S2.3, one has (cf. (S6))

$$G_{\theta,u_n/\sqrt{n},n}\|Z_t\|^{4+\delta} \lesssim \left(\frac{C_1}{1-\varrho}\right)^{4+\delta} + \left(\frac{C_2}{1-\varrho}\right)^{4+\delta} G_{\theta,u_n/\sqrt{n},n}|\epsilon_1|^{4+\delta} + \|Z_0\|^{4+\delta},$$

which is uniformly bounded given the penultimate display.

Finally consider $s_{\vartheta_n,1}$. It suffices to show that each component of $\dot{\ell}_{\vartheta_n}$ has $4+\delta$ moment bounded uniformly for all $n \geq N$.[S4] By Assumption 2.1(iii), by increasing N if necessary, $\sup_{\vartheta\in\mathsf{T}}|\zeta^x_{l,k,j}(\vartheta)| \leq M$ for all l,k,j and $x \in \alpha,\sigma$ and likewise $\sup_{\vartheta\in\mathsf{T}}\|A_{k\bullet}(\vartheta)D_{b_l}(\vartheta)\| \leq M$. Recall that $V_{\vartheta_n,t} = A(\vartheta_n)^{-1}\epsilon_t$. Given (S8) and the observations in footnote S4 to complete the proof it suffices to note that (for $\phi_k = \frac{\mathrm{d}\log\eta_k(x)}{\mathrm{d}x}$) and some $C < \infty$,

$$G_{\theta,u_n/\sqrt{n},n}|\phi_k|^{4+\delta} \leq \left(1+\frac{\bar{h}_k}{\sqrt{n}}\right)\int|\phi(x)|^{4+\delta}\eta_k(x)\,\mathrm{d}x \leq C. \qquad \square$$

Lemma S2.7: *Let $W_{n,t}$ be as in the Proof of Proposition A.1 and suppose the conditions of that Proposition hold. Let G_θ be defined as in the Proof of Lemma S2.5. Then, under P_θ^n,*

$$\lim_{n\to\infty}\mathbb{E}\left|\sum_{t=1}^n W_{n,t}^2 - \frac{\tau^2}{4}\right| = 0, \qquad \textit{with} \qquad \tau^2 := G_\theta\left(g'\dot{\ell}_\theta(Y,X) + \sum_{k=1}^K h_k(A_{k\bullet}(\theta)V_\theta)\right)^2.$$

Proof. Define

$$r_\theta(X_t) := \mathbb{E}[s_\theta(Y_t,X_t)^2|X_t], \quad s_\theta(Y,X) := g'\dot{\ell}_\theta(Y,X) + \sum_{k=1}^K h_k(A_{k\bullet}(\theta)V_\theta),$$

where the conditional expectation is taken under P_θ^n. Since conditional expectations are L_1 con-

[S4] The form each such component is that given in equations equations (7) – (9). Note here that each ϕ_k is (implicitly) a function of η_k and thus when evaluating equations (7) – (9) at ϑ_n, the ϕ_k that appear are $\phi_{k,u_n,n}$, defined as

$$\phi_{k,u,n} := \frac{\mathrm{d}(\log\eta_k(x) + \log(1+uh_k(x)/\sqrt{n}))}{\mathrm{d}x} = \phi_k + \frac{uh_k'/\sqrt{n}}{1+uh_k/\sqrt{n}}.$$

Since each h_k, and h_k' are bounded, increasing N if necessary, one has for $n \geq N$,

$$|\phi_{k,u_n,n}| \leq |\phi_k| + M.$$

tractions, by Lemma S2.4, we have that $P_\theta^n[|r_\theta(X_t)|^{1+\rho/2}] \lesssim C < \infty$ and hence $(|r_\theta(X_t)|^{1+\rho/2})_{t\in\mathbb{N}}$ is uniformly P_θ^n–integrable. Moreover we have for $\mathscr{F}_t := \sigma(\epsilon_1, \ldots, \epsilon_t)$,

$$r_\theta(X_t) = \mathbb{E}[s_\theta(Y_t, X_t)^2 | X_t] = \mathbb{E}[s_\theta(Y_t, X_t)^2 | \mathscr{F}_{t-1}],$$

as is clear from the definition of s_θ.[S5] Hence $(s_\theta(Y_t, X_t)^2 - r_\theta(X_t), \mathscr{F}_t)$ is a martingale difference squence and by Theorem 19.7 in Davidson (1994)

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{1}{n}\sum_{t=1}^n [s_\theta(Y_t, X_t)^2 - r_\theta(X_t)]\right|^{1+\rho/2} = 0.$$

Now define $u_\theta(X_t) := r_\theta(X_t) - \mathbb{E}[r_\theta(X_t)]$, which satisfies $P_\theta^n[|u_\theta(X_t)|^{1+\rho/2}] \lesssim C < \infty$ and is evidently mean zero. By Theorem 3 in Saikkonen (2007), Z_t and hence $u_\theta(X_t)$ (e.g. Davidson, 1994, Theorem 14.1) has geometrically decaying β-mixing coefficients. Therefore, by Theorem 14.2 in Davidson (1994), $(u_\theta(X_t)/n)_{n\in\mathbb{N}, 1\le t\le n}$ is an L_1–mixingale array with respect to the filtration formed by $\mathsf{F}_{n,t} := \sigma(X_1, \ldots, X_t)$ relative to the sequence of positive constants

$$n^{-1} \le c_{n,t} = \max\left\{1/n, \left(P_\theta^n\left[|u_\theta(X_t)/n|^{1+\rho/2}\right]\right)^{1/(1+\rho/2)}\right\} \le n^{-1}\max\{C, 1\}.$$

By Theorem 19.11 in Davidson (1994),

$$\lim_{n\to\infty} \mathbb{E}\left|\frac{1}{n}\sum_{t=1}^n u_\theta(Y_t, X_t)\right| = 0.$$

It remains to show that $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[r_\theta(X_t)] \to \tau^2$. Since $\mathbb{E}[r_\theta(X_t)] = \mathbb{E}[s_\theta(Y_t, X_t)]$,

$$\tau_n^2 := G_{\theta,0,n}\left[s_\theta(Y, X)^2\right] = \frac{1}{n}\sum_{t=1}^n \mathbb{E}\, s_\theta(Y_t, X_t)^2 = \frac{1}{n}\sum_{t=1}^n \mathbb{E}[r_\theta(X_t)],$$

where $G_{\theta,0,n}$ is as defined in the Proof of Lemma S2.5. That $\mathbb{E}\frac{1}{n}\sum_{t=1}^n s_\theta(Y_t, X_t)^2 \lesssim C$ follows from Lemma S2.4. Therefore, by Lemma S2.6, $s_\theta(Y, X)^2$ is uniformly $G_{\theta,0,n}$–integrable and also $\tau^2 < \infty$. Then, by Corollary 2.9 in Feinberg et al. (2016) and Lemma S3.2, $\tau_n^2 \to \tau$. □

Lemma S2.8: *In the setting of Proposition A.2,*

$$\log\frac{p^n_{\theta_n(g_n,h)}}{p^n_{\theta_n(g,h)}} = o_{P^n_{\theta_n(g,h)}}(1).$$

[S5] See e.g. Theorem 7.3.1 in Chow and Teicher (1997) for the (almost sure) equality of the conditional expectations.

Proof. Since by Proposition A.1 and Example 6.5 in van der Vaart (1998) $P^n_{\theta_n(g,h)} \triangleleft\triangleright P^n_\theta$ it suffices to show that the left hand side is $o_{P^n_\theta}(1)$. We first show that

$$\log \frac{p^n_{\theta_n(g_n,0)}}{p^n_\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^n g'\dot{\ell}_\theta(Y_t,X_t) - \mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n g'\dot{\ell}_\theta(Y_t,X_t)\right)^2 + o_{P^n_\theta}(1)$$
$$\log \frac{p^n_{\theta_n(g,0)}}{p^n_\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^n g'\dot{\ell}_\theta(Y_t,X_t) - \mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n g'\dot{\ell}_\theta(Y_t,X_t)\right)^2 + o_{P^n_\theta}(1)$$

For these log–likelihood expansions we may appeal to Lemma 1 in Swensen (1985). The required Conditions (1.3) - (1.7) and (iii) of his Theorem 1 are all established in the proof of Proposition A.1 (take each $h_k = 0$). It remains to show condition (1.2) for each of the cases in the above display. In particular, set

$$W_{n,t} := \frac{1}{2\sqrt{n}} g'\dot{\ell}_\theta(Y_t,X_t)$$

and (cf. equations (37), (38))

$$U_{n,t} := \left[\left(\frac{|A(\theta_n(g_n,h))|}{|A(\theta)|}\right) \times \prod_{k=1}^K \frac{\eta_k(A_{k\bullet}(\theta_n(g_n,h))V_{\theta_n(g_n,h),t})}{\eta_k(A_{k\bullet}(\theta)V_{\theta,t})}\right]^{1/2} - 1$$

where we note that $A(\theta) = A(\theta_n(0,h))$ and $V_\theta = V_{\theta_n(0,h)}$. We verify (1.2), i.e. that

$$\lim_{n\to\infty} \mathbb{E}\left[\sum_{t=1}^n (W_{n,t} - U_{n,t})^2\right] = 0,$$

under P^n_θ.[S6] The argument now follows similarly to that in Lemma S2.5. To simplify the notation, let $p_\gamma := p_{(\gamma,\eta)}$ and $\dot{\ell}_\gamma := \dot{\ell}_{(\gamma,\eta)}$ where $\eta = (\eta_1,\ldots,\eta_K)$ will remain fixed. By Assumption 2.1 and standard computations, the derivative of $\gamma \mapsto \sqrt{p_\gamma}$ is $\frac{1}{2}\dot{\ell}_\gamma\sqrt{p_\gamma}$ (everywhere). Inspection reveals that this is continuous in γ.

Let $\gamma_n := \gamma + g_n/\sqrt{n}$. For $q_{\theta,t}$ the density of X_t under P^n_θ,

$$\begin{aligned}\mathbb{E}\sum_{t=1}^n (W_{n,t} - U_{n,t})^2 &= \frac{1}{n}\sum_{t=1}^n \int \left(\sqrt{n}\left[\sqrt{\frac{p_{\gamma_n}}{p_\gamma}} - 1\right] - \frac{1}{2}g'\dot{\ell}_\gamma\right)^2 p_\gamma q_{\theta,t}\,\mathrm{d}\lambda \\ &= \int \left(\sqrt{n}\left[\sqrt{p_{\gamma_n}} - \sqrt{p_\gamma}\right] - \frac{1}{2}g'\dot{\ell}_\gamma\sqrt{p_\gamma}\right)^2 \bar{q}_{n,\theta}\,\mathrm{d}\lambda,\end{aligned}$$

with $\bar{q}_{n,\theta} := \frac{1}{n}\sum_{t=1}^n q_{\theta,t}$. The term inside the parentheses converges to zero as $n \to \infty$ by the

[S6]This suffices as the second expansion is just the special case $g_n = g$ for each $n \in \mathbb{N}$.

differentiability of $\gamma \mapsto \sqrt{p_\gamma}$ and that $(g_n - g)'\dot{\ell}_\gamma \sqrt{p_\gamma} \to 0$ pointwise. Let

$$I_{\theta,u,n} \coloneqq \int (g'\dot{\ell}_{\gamma+ug_n})^2 \, p_{\gamma+ug_n} \, \bar{q}_{n,\theta} \, \mathrm{d}\lambda = \int (g'\dot{\ell}_{\gamma+ug_n})^2 \, \mathrm{d}G_{\theta,u,n},$$

where $G_{\theta,u,n}$ is the distribution of (Y_t, X_t) corresponding to the density $p_{\gamma+ug_n} \, \bar{q}_{n,\theta}$. By Lemma S3.2 $G_{\theta,u_n/\sqrt{n},n} \xrightarrow{TV} G_\theta$, defined as in the proof of Lemma S2.5. For any $(u_n) \subset [0,1]$ we have that $(g'\dot{\ell}_{\gamma+u_n g_n/\sqrt{n}})^2 \to (g'\dot{\ell}_\gamma)^2$ (pointwise). Each component of $\dot{\ell}_\gamma \in L_2(G_\theta)$ by Lemma S2.6 and moreover $\sup_{n\geq N} G_{\theta,u_n/\sqrt{n},n} \|\dot{\ell}_{\gamma+u_n g_n/\sqrt{n}}\|^{2+\rho} \leq C$ for some $\rho > 0$.[S7] Therefore, by Corollary 2.9 in Feinberg et al. (2016), $\lim_{n\to\infty} I_{\theta,u_n/\sqrt{n},n} = \int (g'\dot{\ell}_\gamma)^2 \, \mathrm{d}G_\theta < \infty$ and hence

$$\left| \int_0^1 I_{\theta,u/\sqrt{n},n} \, \mathrm{d}u - \int_0^1 \int s_\theta^2 \, \mathrm{d}G_\theta \, \mathrm{d}u \right| \leq \sup_{u\in[0,1]} \left| I_{\theta,u/\sqrt{n},n} - \int (g'\dot{\ell}_\gamma)^2 \, \mathrm{d}G_\theta \right| \to 0.$$

By the continuous differentiability of $\sqrt{p_\gamma}$, Jensen's inequality and the Fubini – Tonelli theorem,

$$\begin{aligned} \int \left(\sqrt{n}\left[\sqrt{p_{\gamma_n}} - \sqrt{p_\gamma}\right]\right)^2 \bar{q}_{n,\theta} \, \mathrm{d}\lambda &\leq \frac{1}{4} \int \int_0^1 \left((g'\dot{\ell}_{\gamma+ug_n/\sqrt{n}}) \sqrt{p_{\gamma+ug_n/\sqrt{n}}} \right)^2 \bar{q}_{n,\theta} \, \mathrm{d}u \, \mathrm{d}\lambda \\ &\leq \int_0^1 I_{\theta,u/\sqrt{n},n} \, \mathrm{d}u. \end{aligned}$$

Combining these observations with Proposition 2.29 in van der Vaart (1998) verifies (1.2) and hence the claimed log – likelihood expansions follow from Lemma 1 in Swensen (1985).

To complete the proof set

$$\tilde{u}_{k,n,t} \coloneqq A_{k\bullet}(\theta_n(g_n,h)) V_{\theta_n(g_n,h),t}, \qquad u_{k,n,t} \coloneqq A_{k\bullet}(\theta_n(g,h)) V_{\theta_n(g,h),t},$$

and observe that

$$\begin{aligned} &\log \frac{p^n_{\theta_n(g_n,h)}}{p^n_{\theta_n(g,h)}} - \left[\log \frac{p^n_{\theta_n(g_n,0)}}{p^n_\theta} - \log \frac{p^n_{\theta_n(g,0)}}{p^n_\theta} \right] \\ &\qquad = \sum_{k=1}^K \sum_{i=1}^n \log\left(1 + \frac{h_k(\tilde{u}_{k,n,t})}{\sqrt{n}}\right) - \log\left(1 + \frac{h_k(u_{k,n,t})}{\sqrt{n}}\right), \end{aligned}$$

where the bracketed term is $o_{P^n_\theta}(1)$ by the preceding argument. Hence it suffices to show that an arbitrary k-th element of the outer sum on the right hand side is also $o_{P^n_\theta}(1)$. Let $\varepsilon \in (0,1)$

[S7] This follows from (a) the continuity requirements in Assumption 2.1(iii), (b) under $G_{\theta,u_n/\sqrt{n},n}$ we have that $e_k' A(\theta_n(u_n g_n, 0))^{-1} V_{\theta_n(u_n g_n,0)} = \epsilon_k \sim \eta_k$ and (c) $\sup_{n\geq N, 1\leq t\leq n} G_{\theta,u_n/\sqrt{n},n} \|X_t\|^{4+\delta} < \infty$, which can be shown by an argument analogous to that which is established in the proof of Lemma S2.6.

be fixed and define

$$E_n := \left\{ \max_{1\le i\le n} |h_k(\tilde{u}_{k,n,t})|/\sqrt{n} \le \varepsilon \right\}, \quad F_n := \left\{ \max_{1\le i\le n} |h_k(u_{k,n,t})|/\sqrt{n} \le \varepsilon \right\}.$$

Since h_k is bounded $P_\theta^n(E_n \cap F_n) \to 1$. On this set we may perform a two-term Taylor expansion of $\log(1+x)$ to obtain

$$\log\left(1+\frac{h_k(\tilde{u}_{k,n,t})}{\sqrt{n}}\right) - \log\left(1+\frac{h_k(u_{k,n,t})}{\sqrt{n}}\right)$$
$$= \frac{h_k(\tilde{u}_{k,n,t}) - h_k(u_{k,n,t})}{\sqrt{n}} - \frac{1}{2}\frac{h_k(\tilde{u}_{k,n,t})^2 - h_k(u_{k,n,t})^2}{n} + R\left(\frac{h_k(\tilde{u}_{k,n,t})}{\sqrt{n}}\right) - R\left(\frac{h_k(u_{k,n,t})}{\sqrt{n}}\right),$$

where $|R(x)| \le |x|^3$. For the remainder terms one has for any u_i,

$$\sum_{i=1}^n \left| R\left(\frac{h_k(u_i)}{\sqrt{n}}\right)\right| \le \max_{1\le i\le n} \frac{h_k(u_i)}{\sqrt{n}} \frac{1}{n}\sum_{i=1}^n h_k(u_i)^2 \lesssim \frac{1}{\sqrt{n}},$$

since h_k is bounded. For the first term in Taylor expansion, note that the derivative (in θ, σ) of $A(\theta,\sigma)$ is bounded on a neighbourhood of (θ,σ) (by Assumption 2.1). Combine this with the boundedness of h_k' and the mean value theorem to conclude that

$$|h_k(\tilde{u}_{k,n,t}) - h_k(u_{k,n,t})| \lesssim n^{-1/2}\|g_n - g\| \left[\|\epsilon_t\| + \|X_t\|\right].$$

Using this, since h_k is bounded,

$$|h_k(\tilde{u}_{k,n,t})^2 - h_k(u_{k,n,t})^2| \lesssim n^{-1/2}\|g_n - g\| \left[\|\epsilon_t\| + \|X_t\|\right].$$

Therefore, using (S6) and Assumption 2.1(ii)

$$\sum_{i=1}^n \left| \frac{h_k(\tilde{u}_{k,n,t}) - h_k(u_{k,n,t})}{\sqrt{n}} - \frac{1}{2}\frac{h_k(\tilde{u}_{k,n,t})^2 - h_k(u_{k,n,t})^2}{n}\right|$$
$$\lesssim \|g_n - g\| \left(1+\frac{1}{\sqrt{n}}\right)\frac{1}{n}\sum_{i=1}^n \left[\|\epsilon_t\| + \|X_t\|\right] = o_{P_\gamma^n}(1). \qquad \square$$

Lemma S2.9: *In the setting of Proposition A.2,*

$$\log \frac{p^n_{\theta_n(g_n,h_n)}}{p^n_{\theta_n(g_n,h)}} = o_{P^n_{\theta_n(g_n,h)}}(1).$$

Proof. For notational ease, set

$$u_{k,n,t} \coloneqq e_k' A(\theta_n(g_n, h)) V_{\theta_n(g_n,h),t} = e_k' A(\theta_n(g_n, h_n)) V_{\theta_n(g_n,h_n),t}.$$

One has that

$$\log \frac{p^n_{\theta_n(g_n,h_n)}}{p^n_{\theta_n(g_n,h)}} = \sum_{k=1}^K \sum_{t=1}^n \log(1 + h_{k,n}(u_{k,n,t})/\sqrt{n}) - \log(1 + h_k(u_{k,n,t})/\sqrt{n}),$$

hence it suffices to show that each

$$l_{n,k} \coloneqq \sum_{t=1}^n \log(1 + h_{k,n}(u_{k,n,t})/\sqrt{n}) - \log(1 + h_k(u_{k,n,t})/\sqrt{n}) \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

Let $\varepsilon \in (0, 1)$ be fixed and define

$$\begin{aligned} E_n &\coloneqq \left\{ \max_{1\le t\le n} |h_{k,n}(u_{k,n,t})|/\sqrt{n} \le \varepsilon \right\}; \\ F_n &\coloneqq \left\{ \max_{1\le t\le n} |h_k(u_{k,n,t})|/\sqrt{n} \le \varepsilon \right\}. \end{aligned}$$

Since h_k is bounded, $P^n_{\theta_n(g_n,h)} F_n \to 1$; $P^n_{\theta_n(g_n,h)} E_n \to 1$ follows from Lemma S2.11. Hence $P^n_{\theta_n(g_n,h)} F_n \cap E_n \to 1$. On $E_n \cap F_n$ we can perform a two-term Taylor expansion of $\log(1 + x)$ to obtain

$$\begin{aligned} \log(1+&h_{k,n}(u_{k,n,t})/\sqrt{n}) - \log(1 + h_k(u_{k,n,t})/\sqrt{n}) \\ &= \frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}} - \frac{1}{2}\frac{h_{k,n}(u_{k,n,t})^2}{n} - \frac{h_k(u_{k,n,t})}{\sqrt{n}} + \frac{1}{2}\frac{h_k(u_{k,n,t})^2}{n} \\ &\quad + R\left(\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\right) - R\left(\frac{h_k(u_{k,n,t})}{\sqrt{n}}\right), \end{aligned}$$

where $|R(x)| \le |x|^3$. It follows that

$$\begin{aligned} l_{n,k} = \frac{1}{\sqrt{n}} \sum_{t=1}^n h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t}) - \frac{1}{2}\frac{1}{n} \sum_{t=1}^n [h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2] \\ + \sum_{t=1}^n R\left(\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\right) - R\left(\frac{h_k(u_{k,n,t})}{\sqrt{n}}\right). \end{aligned}$$

We will show that the remainder terms vanish. In particular, one has

$$\sum_{t=1}^n \left| R\left(\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\right)\right| \le \sum_{t=1}^n \left|\frac{h_{k,n}(u_{k,n,t})}{\sqrt{n}}\right| \left|\frac{h_{k,n}(u_{k,n,t})^2}{n}\right| \le \max_{1\le t\le n} \frac{|h_{k,n}(u_{k,t,n})|}{\sqrt{n}} \frac{1}{n} \sum_{t=1}^n h_{k,n}(u_{k,n,t})^2.$$

By Markov's inequality with Lemmas S2.10 and S2.11, this converges to zero in $P^n_{\theta_n(g_n,h)}$ probability. The same evidently holds for the case where $h_{k,n} = h_k$ for each $n \in \mathbb{N}$. Thus,

$$l_{n,k} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t}) - \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2] + o_{P^n_{\theta_n(g_n,h)}}(1),$$

and it remains to show that $\frac{1}{\sqrt{n}}\sum_{t=1}^n h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})$ and $\frac{1}{n}\sum_{t=1}^n[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2]$ also converge to zero in probability. The second of these follows directly from Lemma S2.10, Markov's inequality and the reverse triangle inequality since

$$\begin{aligned} P^n_{\theta_n(g_n,h)}\left(\left|\frac{1}{n}\sum_{t=1}^{n}[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2]\right| > \varepsilon\right) &\leq \varepsilon^{-1}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\right] \\ &= \varepsilon^{-1}\,\mathbb{E}\left[h_{k,n}(u_{k,n,t})^2 - h_k(u_{k,n,t})^2\right] \\ &\to 0. \end{aligned}$$

For the remaining term, we start by noting that

$$\mathbb{E}[h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})] = \frac{\mathbb{E}[(h_{k,n}(\epsilon_k) - h_k(\epsilon_k))h_k(\epsilon_k)]}{\sqrt{n}}$$

so

$$\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}[h_{k,n}(u_{k,n,t})] - \mathbb{E}[h_k(u_{k,n,t})]\right| \leq \frac{1}{n}\sum_{t=1}^{n}\|h_{k,n} - h_k\|_{L_2(P^n_\theta)}\|h_k\|_{L_2(P^n_\theta)} \to 0.$$

Thus it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t}) \xrightarrow{P^n_{\theta_n(g_n,h)}} 0,$$

for $\tilde{h}_{k,n}(u_{k,n,t}) \coloneqq \tilde{h}_{k,n}(u_{k,n,t}) - \mathbb{E}\left[\tilde{h}_{k,n}(u_{k,n,t})\right]$ and $\tilde{h}_k(u_{k,n,t}) \coloneqq \tilde{h}_{k,n}(u_{k,n,t}) - \mathbb{E}\left[\tilde{h}_k(u_{k,n,t})\right]$. By the reverse triangle inequality and Lemma S2.10,

$$\mathbb{E}\left[\left(\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t})\right)^2\right] \to 0, \quad \text{uniformly in } t.$$

Using this, the independence of the $u_{k,t,n}$ and Markov's inequality:

$$P^n_{\theta_n(g_n,h)}\left(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t})\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\tilde{h}_{k,n}(u_{k,n,t}) - \tilde{h}_k(u_{k,n,t})\right)^2\right] \to 0.$$

This establishes that $\sum_{k=1}^K l_{n,k} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0$, as required. □

Lemma S2.10: *In the setting of Proposition A.2, let* $u_{k,n,t} := e_k' A_{\theta_n(g_n,h)} V_{\theta_n(g_n,h),t}$. *Under* $P^n_{\theta_n(g_n,h)}$,

$$\mathbb{E}\left[h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})\right]^2 \leq \|h_{n,k} - h_k\|_{L_2(P^n_\theta)} \left(1 + \frac{\|h_k\|_{L_\infty(P^n_\theta)}}{\sqrt{n}}\right).$$

Proof. Under $P^n_{\theta_n(g_n,h)}$, $u_{k,n,t} \sim \eta_k(1 + h_k/\sqrt{n})$, so for $\epsilon_k \sim \eta_k$, since h_k is bounded,

$$\begin{aligned}
&\mathbb{E}\left[h_{k,n}(u_{k,n,t}) - h_k(u_{k,n,t})\right]^2 \\
&= \int \left[h_{n,k}(x) - h_k(x)\right]^2 \eta_k(x)(1 + h_k(x)/\sqrt{n})\,\mathrm{d}x \\
&\leq \mathbb{E}[h_{k,n}(\epsilon_k) - h_k(\epsilon_k)]^2 + \frac{1}{\sqrt{n}}\mathbb{E}[h_{k,n}(\epsilon_k) - h_k(\epsilon_k)]^2 \|h_k\|_{L_\infty(P^n_\theta)} \\
&\leq \|h_{n,k} - h_k\|_{L_2(P^n_\theta)} + \|h_{n,k} - h_k\|_{L_2(P^n_\theta)} \|h_k\|_{L_\infty(P^n_\theta)}/\sqrt{n}.
\end{aligned}$$

□

Lemma S2.11: *In the setting of Proposition A.2, let* $u_{k,n,t} := e_k' A_{\theta_n(g_n,h)} V_{\theta_n(g_n,h),t}$. *Then*

$$\max_{1\leq t\leq n} \frac{|h_{k,n}(u_{k,n,t})|}{\sqrt{n}} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

Proof. Under $P^n_{\theta_n(g_n,h)}$, $u_{k,n,t} \sim \eta_k(1 + h_k/\sqrt{n})$. By Lemma S2.10, $h_{k,n}(u_{k,n,t})$ is uniformly square $P^n_{\theta_n(g_n,h)}$–integrable and hence the Lindeberg condition holds for $h_{k,n}(u_{k,n,t})/\sqrt{n}$:

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{t=1}^n \mathbb{E}\left[\frac{h_{k,n}(u_{k,n,t})^2}{n} \mathbf{1}\left\{|h_{n,k}(u_{k,n,t})| > \delta\sqrt{n}\right\}\right] \\
&\quad = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}\left[h_{k,n}(u_{k,n,t})^2 \mathbf{1}\left\{|h_{n,k}(u_{k,n,t})| > \delta\sqrt{n}\right\}\right] \\
&\quad = \lim_{n\to\infty} \mathbb{E}\left[h_{k,n}(u_{k,n,t})^2 \mathbf{1}\left\{|h_{n,k}(u_{k,n,t})| > \delta\sqrt{n}\right\}\right] \\
&\quad = 0,
\end{aligned}$$

for any $\delta > 0$. This implies the claimed uniform asymptotic negligability condition (e.g. Gut, 2005, Remark 7.2.4):

$$\max_{1\leq t\leq n} \frac{|h_{k,n}(u_{k,n,t})|}{\sqrt{n}} \xrightarrow{P^n_{\theta_n(g_n,h)}} 0.$$

□

S2.4 Scores

Lemma S2.12: *Suppose Assumption 2.1 holds. Let p_θ and $\bar{q}_{n,\theta}$ be as in the Proof of Proposition S2.5 and suppose that $\theta_n = (\gamma_n, \eta) \to (\gamma, \eta) = \theta$. Then*

$$\lim_{n\to\infty} \int \left\| \tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2} \bar{q}_{n,\theta}^{1/2} - \tilde{\ell}_\theta p_\theta^{1/2} \bar{q}_{n,\theta}^{1/2} \right\|^2 \mathrm{d}\lambda = 0. \tag{S9}$$

Proof. The integral in (S9) can be re-written as

$$\sum_{l=1}^{L} \int \left(\tilde{\ell}_{\theta_n,l}(y,x) p_{\theta_n}(y,x)^{1/2} - \tilde{\ell}_{\theta,l}(y,x) p_\theta(y,x)^{1/2} \right)^2 \mathrm{d}(\lambda(y) \otimes Q_{n,\theta}(x))$$

Inspection of the forms of $\tilde{\ell}_\vartheta$ and p_ϑ reveals that each integrand in the preceding display converges to zero as $n \to \infty$. If we show that

$$\limsup_{n\to\infty} \int \tilde{\ell}_{\theta_n,l}^2 p_{\theta_n} \, \mathrm{d}(\lambda \otimes Q_{n,\theta}) \le \int \tilde{\ell}_{\theta,l}^2 p_\theta \, \mathrm{d}(\lambda \otimes Q_\theta) < \infty, \tag{S10}$$

the proof will be complete in view of Lemma S2.2, Proposition S3.1 and Remark S3.1.[S8] The preceding display is equivalent to

$$\limsup_{n\to\infty} \int \tilde{\ell}_{\theta_n,l}^2 \, \mathrm{d}G_{\theta_n,\theta,n} \le \int \tilde{\ell}_{\theta,l}^2 \, \mathrm{d}G_\theta < \infty,$$

for $G_{\vartheta,\theta,n}$ the distribution of (Y, X) corresponding to the density $p_\vartheta \bar{q}_{n,\theta}$ and G_θ as defined in the proof of Lemma S2.5. That $\tilde{\ell}_{\theta_n,l}^2 \to \tilde{\ell}_{\theta,l}^2$ pointwise is clear from its form, as given in Lemma 3.1. The finiteness of each of the integrals in the above display along with the fact that for some $N \in \mathbb{N}$ and some $\rho > 0$,

$$\sup_{n\ge N} \int \tilde{\ell}_{\theta_n,l}^{2+\rho} \, \mathrm{d}G_{\theta_n,\theta,n} < \infty$$

follows from the form of $\tilde{\ell}_{\vartheta,l}^2$ (as given in Lemma 3.1) along with Assumption 2.1.[S9] □

Lemma S2.13 (Smoothness): *Suppose that Assumption 2.1 holds. Then for any sequence $\theta_n = (\gamma + g_n/\sqrt{n}, \eta)$ with $g_n \to g \in \mathbb{R}^L$,*

$$R_n := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\tilde{\ell}_{\theta_n}(Y_t, X_t) - \tilde{\ell}_\theta(Y_t, X_t) \right] + \tilde{I}_{\theta,n} g_n \xrightarrow{P_\theta^n} 0.$$

[S8] Note that the product structure of $\lambda \otimes Q_{n,\theta}$ and Lemma S2.2 ensure that $\lambda \otimes Q_{n,\theta} \to \lambda \otimes Q_\theta$ setwise.

[S9] Cf. the proof of Lemma S2.3: arguing in essentially the same manner as there allows one to obtain uniform boundedness of the $4+\delta$ moments of ϵ_k, $\phi_k(\epsilon_k)$, X_t (uniformly in t) and all the non-stochastic terms in $\tilde{\ell}_{\theta_n,l}^2$.

Proof. From (the proof of) Lemma S2.8 we have

$$\lim_{n\to\infty}\int\left[\sqrt{n}\left(p_{\theta_n}^{1/2}-p_\theta^{1/2}\right)\bar{q}_{n,\theta}^{1/2}-\frac{1}{2}g'\dot{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{1/2}\right]^2\mathrm{d}\lambda=0, \tag{S11}$$

whilst by Lemma S2.12 we have

$$\lim_{n\to\infty}\int\left\|\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2}\bar{q}_{n,\theta}^{1/2}-\tilde{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{1/2}\right\|^2\mathrm{d}\lambda=0. \tag{S12}$$

Define

$$c_n^{-1}:=\int p_{\theta_n}^{1/2}p_\theta^{1/2}\bar{q}_{n,\theta}\,\mathrm{d}\lambda=1-\frac{1}{2}\int(p_\theta^{1/2}-p_{\theta_n}^{1/2})^2\bar{q}_{n,\theta}\,\mathrm{d}\lambda.$$

We have

$$\begin{aligned}-n\left(p_\theta^{1/2}-p_{\theta_n}^{1/2}\right)^2=&-\left(\sqrt{n}\left[p_{\theta_n}^{1/2}-p_\theta^{1/2}\right]-\frac{1}{2}g'\dot{\ell}_\theta p_\theta^{1/2}\right)^2+\left(\frac{1}{2}g'\dot{\ell}_\theta p_\theta^{1/2}\right)^2\\&-g'\dot{\ell}_\theta p_\theta^{1/2}\sqrt{n}\left(p_{\theta_n}^{1/2}-p_\theta^{1/2}\right),\end{aligned}$$

and so by (S11) and the continuity of the inner product

$$\begin{aligned}\int(p_\theta^{1/2}-p_{\theta_n}^{1/2})^2\bar{q}_{n,\theta}\,\mathrm{d}\lambda&=\frac{1}{n}\int g'\dot{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{1/2}\sqrt{n}\left(p_{\theta_n}^{1/2}-p_\theta^{1/2}\right)\bar{q}_{n,\theta}^{1/2}\,\mathrm{d}\lambda\\&\quad-\frac{1}{n}\int\left(\frac{1}{2}g'\dot{\ell}_\theta p_\theta^{1/2}\right)^2\bar{q}_{n,\theta}\,\mathrm{d}\lambda+o(n^{-1})\\&=\frac{1}{4}(n^{-1/2}g)'\dot{I}_{n,\theta}(n^{-1/2}g)+o(n^{-1}),\end{aligned}$$

where $\dot{I}_{n,\theta}:=\int\dot{\ell}_\theta\dot{\ell}_\theta' p_\theta\bar{q}_{n,\theta}\,\mathrm{d}\lambda=O(1)$.[S10] It follows that $c_n^{-1}=1-a_n$ with $a_n\to 0$ and $na_n=\frac{1}{4}g'\dot{I}_\theta g+o(1)$.

R_n is equal to the sum of

$$\begin{aligned}R_{1,n}'&:=\frac{1}{\sqrt{n}}\sum_{t=1}^n\left[\tilde{\ell}_{\theta_n}(Y_t,X_t)\left(1-\frac{p_{\theta_n}(Y_t,X_t)^{1/2}}{p_\theta(Y_t,X_t)^{1/2}}\right)\right]+\frac{1}{2}\tilde{I}_{n,\theta}g_n\ ;\\R_{2,n}'&:=\frac{1}{\sqrt{n}}\sum_{t=1}^n\left[\tilde{\ell}_{\theta_n}(Y_t,X_t)\frac{p_{\theta_n}(Y_t,X_t)^{1/2}}{p_\theta(Y_t,X_t)^{1/2}}-\tilde{\ell}_\theta(Y_t,X_t)\right]+\frac{1}{2}\tilde{I}_{n,\theta}g_n\ .\end{aligned}$$

Since $\tilde{I}_{n,\theta}$ is $O(1)$ by Lemma S2.3 it suffices to prove that these converge in probability to zero with g_n replaced by g; let the corresponding expressions be called $R_{i,n}$ for $i=1,2$.

[S10]This follows by noting that $\|\dot{\ell}_\theta\|^2$ is uniformly integrable under $p_\theta\bar{q}_{n,\theta}$ which is a consequence of Lemma S2.3.

For $R_{1,n}$ we note that (omitting the arguments of the functions)

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\left(1-\frac{p_{\theta_n}^{1/2}}{p_{\theta}^{1/2}}\right)+\frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta}'g=\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\sqrt{n}\left(1-\frac{p_{\theta_n}^{1/2}}{p_{\theta}^{1/2}}+\frac{1}{2\sqrt{n}}\dot{\ell}_{\theta}'g\right)$$
$$\leq\frac{1}{n}\sum_{t=1}^{n}\|\tilde{\ell}_{\theta_n}\|^2\times\frac{1}{n}\sum_{t=1}^{n}\left[\sqrt{n}\left(1-\frac{p_{\theta_n}^{1/2}}{p_{\theta}^{1/2}}+\frac{1}{2\sqrt{n}}\dot{\ell}_{\theta}'g\right)\right]^2.$$

The first term on the second line is $O_{P_{\theta_n}^n}(1)$ hence $O_{P_{\theta}^n}(1)$ (by contiguity). The second has $L_1(P_\theta^n)$ norm

$$\mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\sqrt{n}\left(1-\frac{p_{\theta_n}^{1/2}}{p_{\theta}^{1/2}}+\frac{1}{2\sqrt{n}}\dot{\ell}_{\theta}'g\right)\right]^2\right|\leq\int\left[\sqrt{n}\left(p_{\theta}^{1/2}-p_{\theta_n}^{1/2}+\frac{1}{2\sqrt{n}}\dot{\ell}_{\theta}'gp_{\theta}^{1/2}\right)\right]^2\bar{q}_{n,\theta}\,\mathrm{d}\lambda\to 0,$$

where the convergence is by (S11). Therefore, it suffices to show that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta}'-\tilde{I}_{n,\theta}\xrightarrow{P_\theta^n}0. \tag{S13}$$

We may replace $\tilde{I}_{n,\theta}$ in (S13) with $\tilde{I}_\theta:=\int\tilde{\ell}_\theta\dot{\ell}_\theta'\,\mathrm{d}G_\theta$ with G_θ as defined in the proof of Lemma S2.5. In particular, let $G_{\theta,n}:=G_{\theta,0,n}$ as defined in the proof of Lemma S2.5. Then, since $\|\tilde{\ell}_\theta(Y_t,X_t)\dot{\ell}_\theta(Y_t,X_t)'\|^{1+\rho/2}$ is uniformly $L_1(P_\theta^n)$ bounded (Lemma S2.3) one has

$$\sup_{n\in\mathbb{N}}\int\|\tilde{\ell}_\theta\dot{\ell}_\theta'\|^{1+\rho/2}\,\mathrm{d}G_{n,\theta}<\infty,$$

and so $\|\tilde{\ell}_\theta\dot{\ell}_\theta'\|$ is uniformly $G_{\theta,n}$–integrable. By Lemma S3.2 and Theorem 2.8 of Serfozo (1982),

$$\tilde{I}_{n,\theta}=\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\tilde{\ell}_\theta(Y_t,X_t)\dot{\ell}_\theta(Y_t,X_t)'\right]=\int\tilde{\ell}_\theta\dot{\ell}_\theta'\,\mathrm{d}G_{n,\theta}\to\int\tilde{\ell}_\theta\dot{\ell}_\theta'\,\mathrm{d}G_\theta=\tilde{I}_\theta. \tag{S14}$$

For any $M>0$, one has the decompositions

$$E_{n,1}^M:=\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\dot{\ell}_\theta'-\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\|\leq M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\|\leq M\}$$
$$=\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\|>M\}\dot{\ell}_\theta'+\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\|\leq M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\|>M\}$$

and

$$\begin{aligned}
E_2^M &:= \tilde{I}_\theta - \int \tilde{\ell}_\theta \dot{\ell}_\theta' \mathbf{1}\{\|\tilde{\ell}_\theta\| \le M\}\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\,\mathrm{d}G_\theta \\
&= \int \tilde{\ell}_\theta \dot{\ell}_\theta' \mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\,\mathrm{d}G + \int \tilde{\ell}_\theta \dot{\ell}_\theta' \mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\mathbf{1}\{\|\dot{\ell}_\theta\| > M\}\,\mathrm{d}G_\theta.
\end{aligned}$$

Additionally, for $\mathbb{E}$ taken under P_θ^n, define

$$\begin{aligned}
E_{n,3}^M &:= \frac{1}{n}\sum_{t=1}^n \tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\} - \mathbb{E}\left[\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\right];\\
E_{n,4}^M &:= \mathbb{E}\,\frac{1}{n}\sum_{t=1}^n \tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\} - \int \tilde{\ell}_\theta \dot{\ell}_\theta'\mathbf{1}\{\|\tilde{\ell}_\theta\| \le M\}\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\,\mathrm{d}G_\theta.
\end{aligned}$$

Since $\|\tilde{\ell}_\theta \dot{\ell}_\theta' \mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\| \le \|\tilde{\ell}_\theta\dot{\ell}_\theta'\|$, $\|\tilde{\ell}_\theta \dot{\ell}_\theta' \mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\mathbf{1}\{\|\dot{\ell}_\theta\| > M\}\| \le \|\tilde{\ell}_\theta\dot{\ell}_\theta'\|$ and $\|\tilde{\ell}_\theta\dot{\ell}_\theta'\|$ is G_θ–integrable by Lemma S2.3, by the dominated convergence theorem, for any $\delta > 0$ there is an M such that $E_2^{M'} < \delta$ for $M' \ge M$. For any $M > 0$, by Theorem 3 in Saikkonen (2007), Theorem 14.1 in Davidson (1994) and Theorem 2 in Kanaya (2017) one has (cf. Lemma S2.14 below)

$$E_{n,3}^M = O_{P_\theta^n}(M^2/\sqrt{n}).$$

For $E_{n,4}^M$ we introduce a new measure: define μ_n as

$$\mu_n(A) := \int_A c_n p_{\theta_n}(x,y)^{1/2} p_\theta(x,y)^{1/2}\,\mathrm{d}(\lambda(y)\otimes Q_n(x)).$$

By Lemma S3.2 one has that $\mu_n \to G$, as well as $G_{n,\theta} \to G$, in TV. Then, by Cauchy – Schwarz and Lemma S2.3

$$\begin{aligned}
c_n^{-1}&\int \tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\,\mathrm{d}\mu_n - \int \tilde{\ell}_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| \le M\}\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}\,\mathrm{d}G_{n,\theta}\\
&= \int \left(\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \le M\}p_{\theta_n}^{1/2} - \tilde{\ell}_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| \le M\}p_\theta^{1/2}\right)\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}p_\theta^{1/2}\,\mathrm{d}(\lambda\otimes Q_{\theta,n})\\
&= \int \left(\tilde{\ell}_{\theta_n}\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\}p_{\theta_n}^{1/2} - \tilde{\ell}_\theta\mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}p_\theta^{1/2}\right)\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}p_\theta^{1/2}\,\mathrm{d}(\lambda\otimes Q_{\theta,n})\\
&\quad + \int \left(\tilde{\ell}_{\theta_n}p_{\theta_n}^{1/2} - \tilde{\ell}_\theta p_\theta^{1/2}\right)\dot{\ell}_\theta'\mathbf{1}\{\|\dot{\ell}_\theta\| \le M\}p_\theta^{1/2}\,\mathrm{d}(\lambda\otimes Q_{\theta,n})\\
&\lesssim o(1) + \sup_{n\in\mathbb{N}}\mathbb{E}_{\theta_n}\left[\|\tilde{\ell}_{\theta_n}\|^2\mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\}\right] + \sup_{n\in\mathbb{N}}\mathbb{E}_\theta\left[\|\tilde{\ell}_\theta\|^2\mathbf{1}\{\|\tilde{\ell}_\theta\| > M\}\right].
\end{aligned}$$

The last two right hand side terms can be made arbitrarily small, uniformly in n, by taking M

large enough; the $o(1)$ term follows from (S12) and is uniform in M. Now, by $G_{n,\theta} \xrightarrow{TV} G_\theta$,

$$\left| \int \tilde{\ell}_\theta \mathbf{1}\{\|\tilde{\ell}_\theta\| \leq M\} \dot{\ell}_\theta' \mathbf{1}\{\|\dot{\ell}_\theta\| \leq M\} \, \mathrm{d}G_{\theta,n} - \int \tilde{\ell}_\theta \mathbf{1}\{\|\tilde{\ell}_\theta\| \leq M\} \dot{\ell}_\theta' \mathbf{1}\{\|\dot{\ell}_\theta\| \leq M\} \, \mathrm{d}G_\theta \right|$$
$$\leq M^2 \|G_{n,\theta} - G_\theta\|_{TV}.$$

Since $\mu_n \to G_\theta$ and $G_{n,\theta} \to G_\theta$ in total variation, one has that $\|\mu_n - G_{n,\theta}\|_{TV} \to 0$. Since $\tilde{\ell}_{\theta_n} \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \leq M\} \dot{\ell}_\theta' \mathbf{1}\{\|\dot{\ell}_\theta\| \leq M\}$ is uniformly bounded, one has that

$$\left| \int \tilde{\ell}_{\theta_n} \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \leq M\} \dot{\ell}_\theta' \mathbf{1}\{\|\dot{\ell}_\theta\| \leq M\} \, \mathrm{d}\mu_n - \int \tilde{\ell}_{\theta_n} \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| \leq M\} \dot{\ell}_\theta' \mathbf{1}\{\|\dot{\ell}_\theta\| \leq M\} \, \mathrm{d}G_{n,\theta} \right|$$
$$\leq M^2 \|\mu_n - G_{n,\theta}\|_{TV}.$$

As $c_n^{-1} - 1 = -a_n \to 0$, it follows that

$$E_{n,4}^M \leq M^2 \left[\|\mu_n - G_{n,\theta}\|_{TV} + \|G_{n,\theta} - G_\theta\|_{TV} \right] + e_n + M^2 |a_n| + r(M),$$

where $0 \leq r(M) \coloneqq \sup_{n \in \mathbb{N}} \mathbb{E}_{P_{\theta_n}^n} \left[\|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\} \right] + \sup_{n \in \mathbb{N}} \mathbb{E}_{P_\theta^n} \left[\|\tilde{\ell}_\theta\|^2 \mathbf{1}\{\|\tilde{\ell}_\theta\| > M\} \right] \to 0$ as $M \to \infty$ and r does not depend on n and $e_n = o(1)$. For $E_{n,1}^M$ note that since $\|\dot{\ell}_\theta\|^2$ is uniformly P_θ^n–integrable (Lemma S2.3), $\frac{1}{n} \sum_{t=1}^n \|\dot{\ell}_\theta\|^2 = O_{P_\theta^n}(1)$. By Markov's inequality, for any $\delta > 0$

$$\begin{aligned}
P_{\theta_n}^n \left(\left| \frac{1}{n} \sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\} \right| > \delta \right) &\leq \delta^{-1} \, \mathbb{E} \left[\left| \frac{1}{n} \sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\} \right| \right] \\
&\leq \delta^{-1} \sup_{n \in \mathbb{N}} \mathbb{E} \, \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\} \\
&\leq \delta^{-1} r(M).
\end{aligned}$$

Thus by taking $M \to \infty$, the probability on the left hand side of the preceding display vanishes. Therefore, the same is true of

$$P_\theta^n \left(\left| \frac{1}{n} \sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M\} \right| > \delta \right),$$

by contiguity. That is, we can take a large enough M such that the probability in the display above is arbitrarily small (for all large enough $n \in \mathbb{N}$).

Now, fix $\varepsilon > 0, \delta > 0$. By Lemma S2.3, $\frac{1}{n} \sum_{t=1}^n \|\tilde{\ell}_\theta\|^2 = O_{P_\theta^n}(1)$ and also $\frac{1}{n} \sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 =$

$O_{P^n_{\theta_n}}(1)$. By this and contiguity, we can choose $R > 0$ be such that for all $n \geq N_1$,

$$P^n_\theta\left(\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_\theta\|^2 > R\right) < \varepsilon/4, \qquad P^n_\theta\left(\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 > R\right) < \varepsilon/4.$$

Take M large enough that $\|E_2^M\| < \delta$, $r(M) < \delta$ and for all $n \geq N_2$

$$\begin{aligned}
&P^n_\theta\left(\left|\frac{1}{n}\sum_{t=1}^n \|\tilde{\ell}_{\theta_n}\|^2 \mathbf{1}\{\|\tilde{\ell}_{\theta_n}\| > M_n\}\right| > \delta/R\right) < \varepsilon/4\\
&P^n_\theta\left(\left|\frac{1}{n}\sum_{t=1}^n \|\dot{\ell}_\theta\|^2 \mathbf{1}\{\|\dot{\ell}_\theta\| > M_n\}\right| > \delta/R\right) < \varepsilon/4
\end{aligned}$$

where $M_n \geq M$ and $M_n \to \infty$ slowly. This ensures that $\|E_2^{M_n}\| < \delta$, $P^n_\theta(\|E^{M_n}_{n,1}\| > 2\delta) < \varepsilon$ for all $n \geq \max\{N_1, N_2\}$. Then, let N be large enough such that $N \geq \max\{N_1, N2\}$, and for all $n \geq N$, $P^n_\theta(\|E^{M_n}_{n,3}\| > \delta) < \varepsilon$ and $\|E^{M_n}_{n,4}\| \leq 3\delta$.[S11] Combining these ensures that for all such n,

$$P^n_\theta\left(\left\|\frac{1}{n}\sum_{t=1}^n \tilde{\ell}_{\theta_n}\dot{\ell}_\theta' - \tilde{I}_\theta\right\| > 7\delta\right) < 2\varepsilon.$$

In conjunction with (S14) this establishes (S13).

We next show that $R_{2,n}$ converges to zero in P^n_θ–probability. Define

$$Z_{n,t} := \tilde{\ell}_{\theta_n}(Y_t, X_t)\frac{p_{\theta_n}(Y_t, X_t)^{1/2}}{p_\theta(Y_t, X_t)^{1/2}}, \quad m_n(X_t) := \int \tilde{\ell}_{\theta_n}(y, X_t) p_{\theta_n}(y, X_t)^{1/2} p_\theta(y, X_t)^{1/2} dy,$$

and note that $m_n(X_t) = \mathbb{E}[Z_{n,t}|X_t]$ (P^n_θ–a.s.). Since $\mathbb{E}[\tilde{\ell}_{\theta_n}(Y_t, X_t)|X_t] = 0$ under $P^n_{\theta_n}$ (which is clear from its form),

$$\begin{aligned}
m_n(X_t) &= \int \tilde{\ell}_{\theta_n}(y, X_t) p_{\theta_n}(y, X_t)^{1/2} p_\theta(y, X_t)^{1/2}\, \mathrm{d}y\\
&= \int \tilde{\ell}_{\theta_n}(y, X_t) p_{\theta_n}(y, X_t)^{1/2}\left[p_\theta(y, X_t)^{1/2} - p_{\theta_n}(y, X_t)^{1/2}\right] \mathrm{d}y.
\end{aligned} \tag{S15}$$

Using (S11), (S12) and Cauchy-Schwarz yields

$$\lim_{n\to\infty}\left|\left\langle \tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2}\bar{q}_{\theta,n}^{-1/2}, \sqrt{n}\left(p_\theta^{1/2} - p_{\theta_n}^{1/2}\right)\bar{q}_{n,\theta}^{-1/2}\right\rangle_\lambda - \left\langle \tilde{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{-1/2}, -\frac{1}{2}g'\dot{\ell}_\theta p_\theta^{1/2}\bar{q}_{n,\theta}^{-1/2}\right\rangle_\lambda\right| = 0,$$

which implies that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n m_n(X_t) + \frac{1}{2}\tilde{I}_{n,\theta} g \xrightarrow{P^n_\theta} 0,$$

[S11] I.e. n such that $M_n^2|a_n| < \delta$, $|e_n| < \delta$, $M_n^2\left[\|\mu_n - G_{n,\theta}\|_{TV} + \|G_{n,\theta} - G_\theta\|_{TV}\right] < \delta$. Here one needs to take $M_n \to \infty$ slowly enough that these sequences still converge to zero and $M_n^2/\sqrt{n} \to 0$.

given the representation of m_n in (S15). In consequence it remains to show that

$$R_{2,n}^* \coloneqq \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t,n} - m_n(X_t) - \tilde{\ell}_\theta(Y_t, X_t) \xrightarrow{P_\theta^n} 0.$$

Put $\mathcal{F}_{n,t} = \sigma(Y_t, X_t)$. Then, as is straightforward to verify, $(Z_{t,n} - m_n(X_t) - \tilde{\ell}_\theta(Y_t, X_t), \mathcal{F}_{n,t})_{n\in\mathbb{N},1\le t\le n}$ forms a martingale difference array. Hence it suffices to show that

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left\| Z_{t,n} - m_n(X_t) - \tilde{\ell}_\theta(Y_t, X_t) \right\|^2 \xrightarrow{P_\theta^n} 0.$$

The left hand side of this display can be written as

$$\int \left\| \tilde{\ell}_{\theta_n} \frac{p_{\theta_n}^{1/2}}{p_\theta^{1/2}} - m_n - \tilde{\ell}_\theta \right\|^2 p_\theta \bar{q}_{n,\theta} \,\mathrm{d}\lambda \le 2 \int \left\| \tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2} \bar{q}_{n,\theta}^{1/2} - \tilde{\ell}_\theta p_\theta^{1/2} \bar{q}_{n,\theta}^{1/2} \right\|^2 \mathrm{d}\lambda + 2 \int \|m_n\|^2 \,\mathrm{d}Q_{n,\theta},$$

and so, given (S12) it suffices to show that the second term on the right hand side converges to zero. For this note that by Fubini's theorem and the Cauchy-Schwarz inequality

$$\begin{aligned}\int \|m_n\|^2 \,\mathrm{d}Q_{n,\theta} &\le \int \left\| \tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2} \left[p_\theta^{1/2} - p_{\theta_n}^{1/2} \right] \right\|^2 \bar{q}_{n,\theta} \,\mathrm{d}\lambda \\ &\le \int \left\| \tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2} \bar{q}_{n,\theta}^{1/2} \right\|^2 \mathrm{d}\lambda \int \left[\left(p_{\theta_n}^{1/2} - p_\theta^{1/2} \right) \bar{q}_{n,\theta}^{1/2} \right]^2 \mathrm{d}\lambda.\end{aligned}$$

The first term on the right hand side is $O(1)$ by equation (S10), whilst the second converges to zero by (S11) and the uniform $G_{\theta,0,n}$ – integrability of $g' \dot{\ell}_\theta$ as established in Lemma S2.6. □

S2.4.1 Estimation

Lemma S2.14: *Suppose that Assumption 2.1 holds and g_n are ϱ – integrable functions for some $\varrho > 2$ such that $\max_{t=1,\ldots,n} \|g_n(Y_t, X_t)\|_{L_\varrho} \le M_n$ (all under P_θ^n). Then,*

$$\frac{1}{n} \sum_{t=1}^n g_n(Y_t, X_t) - \mathbb{E}\left[g_n(Y_t, X_t)\right] = O_{P_\theta}(M_n/\sqrt{n}).$$

Proof. Let $\alpha_n(m)$ be the α – mixing coefficients of the array $\{g_n(Y_t, X_t) - \mathbb{E}[g_n(Y_t, X_t)] : n \in \mathbb{N},\ 1 \le t \le n\}$. By (the proof of) Theorem 14.1 in Davidson (1994), $\alpha_n(m) \le \tilde{\alpha}(m - p)$ (for $m \ge p$) where $\tilde{\alpha}(m)$ are the mixing coefficients of $\{Y_t : t \in \mathbb{N}\}$. By Theorem 3 in Saikkonen (2007) and Proposition 1.1.1 in Doukhan (1994) $\tilde{\alpha}(m) = O(a^m)$ for some $a \in (0, 1)$. Condition A1 in Kanaya (2017) then holds (with $\Delta = 1$) with $\beta > \varrho/(\varrho - 2)$. To see this note that for all

$m \geq M_1$ we have $\tilde{\alpha}(m-p) \leq Ca^m$ whilst $Ca^m \leq Am^{-\beta}$ whenever

$$\beta \leq \frac{\log(A) - \log(C) + m|\log(a)|}{\log(m)}.$$

As the right hand side diverges as $m \to \infty$, for all m larger than some $M \geq M_1$, the inequality will hold for some $\beta > \varrho/(\varrho - 2)$. Noting that the inequality above continues to hold if we increase A, we may then choose A such that each $\tilde{\alpha}(m) \leq Am^{-\beta}$ for all $1 \leq m \leq M$. The result then follows by Theorem 2 in Kanaya (2017). $\square$

Lemma S2.15: *Suppose that Assumptions 2.1 and 2.2 hold. Then*

(i) If $Z_{n,1} := \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{\ell}_\theta(Y_t, X_t)$ *and* $Z_{n,2} := \Lambda^n_{\theta_n(g,h)}(Y^n)$*, then under* P^n_θ*,*

$$Z_n \rightsquigarrow Z \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\sigma^2_{g,h} \end{pmatrix}, \begin{pmatrix} \tilde{I}_\theta & \tilde{I}_\theta g \\ g'\tilde{I}_\theta & \sigma^2_{g,h} \end{pmatrix} \right).$$

Additionally, let $\theta_n := \theta_n(g_n, 0) = (\gamma + g_n/\sqrt{n}, \eta)$ *for* $g_n \to g \in \mathbb{R}^L$*. Then*

(ii) We have that

$$\frac{1}{n}\sum_{t=1}^n \left(\hat{\ell}_{\theta_n}(Y_t, X_t) - \tilde{\ell}_{\theta_n}(Y_t, X_t) \right) = o_{P^n_{\theta_n}}(n^{-1/2}).$$

(iii) $\|\hat{I}_{n,\theta_n} - \tilde{I}_\theta\| = o_{P^n_{\theta_n}}(\nu_n^{1/2})$ *where* ν_n *is defined in Assumption 2.2, and* $\tilde{I}_\theta := G_\theta \tilde{\ell}_\theta \tilde{\ell}'_\theta$ *with* G_θ *as in the proof of Lemma S2.5.*

Proof. For part (i), let z_t be

$$z_t := \left(\tilde{\ell}_\theta(Y_t, X_t)', g'\dot{\ell}_\theta(Y_t, X_t) + \sum_{k=1}^K h_k(A_{k\bullet}V_{\theta,t}) \right)',$$

and $\mathcal{F}_t := \sigma(\epsilon_1, \ldots, \epsilon_t)$. Under P^n_θ, $\{z_t, \mathcal{F}_t : t \in \mathbb{N}\}$ is a martingale difference sequence such that

$$\frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[z_t z_t'\right] = \begin{bmatrix} \tilde{I}_{n,\theta} & \tilde{I}_{\theta,\theta} g \\ g'\tilde{I}_{n,\theta} & \sigma^2_{g,h,n} \end{bmatrix} \to \begin{bmatrix} \tilde{I}_\theta & \tilde{I}_\theta g \\ g'\tilde{I}_\theta & \sigma^2_{g,h} \end{bmatrix},$$

noting Lemma 3.1 and Theorem 12.14 of Rudin (1991). That $\sigma^2_{g,h,n}$ converges to a $\sigma^2_{g,h}$ is part of the conclusion of Proposition A.1. That $\tilde{I}_{\theta,n} \to \tilde{I}_\theta$ follows by combining Lemma S2.3, the fact that $G_{\theta,0,n}$ as defined in the proof of Lemma S2.5 converges in total variation to G_θ (cf. Lemma S3.2), and Corollary 2.9 in Feinberg et al. (2016). Lindeberg's condition is satisfied since $\{\|z_t\|^2 : t \in \mathbb{N}\}$ is uniformly P^n_θ-integrable (by Lemma S2.3 and the fact that each h_k

is bounded) and the variance convergence in the preceding display. Part (i) then follows from Proposition A.1 and the central limit theorem for martingale differences.

Define $A_n := A(\theta_n)$, $B_n := B(\theta_n)$, and $\zeta^x_{n,l,k,j} := \zeta^x_{l,k,j}(\theta_n)$ for each triple (l,j,k) of indicies and $x \in \{\alpha, \sigma\}$. Note that each $A_{n,k}(Y_t - B_n X_t) \approx \epsilon_{k,t} \sim \eta_k$ under $P^n_{\theta_n}$. Hence

$$\tilde{\ell}_{\theta_n,\alpha_l}(Y_t, X_t) \approx \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\alpha}_{l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t} + \sum_{k=1}^{K} \zeta^{\alpha}_{n,l,k,k}\left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\right] \tag{S16}$$

$$\tilde{\ell}_{\theta_n,\sigma_l}(Y_t, X_t) \approx \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^{\sigma}_{n,l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t} + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k}\left[\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})\right] \tag{S17}$$

$$\tilde{\ell}_{\theta_n,b_l}(Y_t, X_t) \approx \sum_{k=1}^{K} -A_{n,k\bullet}D_{b,l}\left[\phi_k(\epsilon_{k,t})(X_t - \mathbb{E}\,X_t) - \mathbb{E}\,X_t\left(\varsigma_{k,1}\epsilon_{k,t} + \varsigma_{k,2}\kappa(\epsilon_{k,t})\right)\right] \tag{S18}$$

By Assumption 2.1(iii), $\zeta^x_{n,l,k,j} \to \zeta^{\alpha}_{\infty,l,k,j} := [D_{x_l}(\alpha,\sigma)]_{k\bullet}A(\alpha,\sigma)^{-1}_{\bullet j}$ for $x \in \{\alpha,\sigma\}$. Note that the entries of $D_{b,l}$ are all zero except for entry l (corresponding to b_l) which is equal to one.

We verify (ii) for each component of the efficient score (S16) – (S18). For components (S16) and (S17), we define for x either of α, σ

$$\varphi_{1,n,t} := \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^x_{l,k,j,n}\phi_k(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t}\ ,$$

and

$$\hat{\varphi}_{1,n,t} := \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta^x_{l,k,j,n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t}\ ,$$

with $V_{n,t} = Y_t - B_n X_t$, and let $\overline{\zeta}_n := \max_{l\in[L],j\in[K],k\in[K]} |\zeta^x_{l,j,k,n}|$ which converges to $\overline{\zeta} := \max_{l\in[L],j\in[K],k\in[K]} |\zeta^x_{l,j,k,\infty}| < \infty$. We have that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\hat{\varphi}_{1,n,t} - \varphi_{1,n,t}) \leq \sqrt{n}\sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \overline{\zeta}_n \left|\frac{1}{n}\sum_{t=1}^{n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t} - \phi_k(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t}\right|,$$

Each $\left|\frac{1}{n}\sum_{t=1}^{n}\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t} - \phi_k(A_{n,k\bullet}V_{n,t})A_{n,j\bullet}V_{n,t}\right| = o_{P_{\theta_n}}(n^{-1/2})$ by applying Lemma A.1 with $W_{n,t} = A_{n,j\bullet}V_{n,t}$ (noting that $A_{n,k\bullet}V_{n,s} \simeq \epsilon_{k,s}$ and $A_{n,j\bullet}V_{n,t} \simeq \epsilon_{j,t}$ with are independent for any s,t with $\mathbb{E}_{\theta_n}(A_{n,j\bullet}V_{n,t})^2 = 1$ by Assumption 2.1(ii)), and the outside summations are finite, it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\hat{\varphi}_{1,n,t} - \varphi_{1,n,t}) = o_{P^n_{\tilde{\theta}_n}}(1)\ . \tag{S19}$$

That $\hat{\tau}_{k,n} \xrightarrow{P^n_{\theta_n}} \tau_k$ follows from Lemma S2.16. Now, consider $\varphi_{2,\tau,n,t}$ defined by

$$\varphi_{2,\tau,n,t} := \sum_{k=1}^{K} \zeta^z_{n,l,k,k} \left[\tau_{k,1} A_{n,k\bullet} V_{n,t} + \tau_{k,2} \kappa(A_{n,k\bullet} V_{n,t})\right],$$

for x equal to either α or σ. Since sum is finite and each $|\zeta^x_{n,l,k,k}| \to |\zeta^x_{\infty,l,k,k}| < \infty$ it is sufficient to consider the convergence of the summands. In particular we have that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\hat{\tau}_{k,n,1} - \tau_{k,1}\right] A_{n,k\bullet} V_{n,t} = \left[\hat{\tau}_{k,n,1} - \tau_{k,1}\right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{n,k\bullet} V_{n,t} \to 0,$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\hat{\tau}_{k,n,2} - \tau_{k,2}\right] \kappa(A_{n,k\bullet} V_{n,t}) = \left[\hat{\tau}_{k,n,2} - \tau_{k,2}\right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \kappa(A_{n,k\bullet} V_{n,t}) \to 0,$$

in probability, since $A_{n,k\bullet} V_{n,t} \approx \epsilon_{k,t} \sim \eta_k$ and $(\epsilon_{k,t})_{t\geq 1}$ and $(\kappa(\epsilon_{k,t}))_{t\geq 1}$ are i.i.d. mean-zero sequences with finite second moments such that the central limit theorem holds.

Together these yield that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varphi_{2,\hat{\tau}_n,n,t} - \varphi_{2,\tau,n,t}) \xrightarrow{P^n_{\theta_n}} 0. \tag{S20}$$

Combination of (S19) and (S20) yields (ii) for components of the type (S16), (S17).

For components (S18) let $a_{n,k,l} := -A_{n,k\bullet} D_{b_l}$, $\tilde{\varsigma}_{k,n} := \hat{\varsigma}_{k,n} - \varsigma_k$, $c_{n,t} := \mathbb{E}_{\theta_n} X_t$ and $\bar{c}_n := \frac{1}{n}\sum_{t=1}^n c_{n,t}$. Since $a_{n,k,l} \to a_{\infty,k,l} := A(\alpha,\sigma)_{k\bullet} D_{b_l}(\alpha,\sigma)$, it suffices to show that

(i) $\frac{1}{n}\sum_{t=1}^n \left[\phi_k(A_{n,k\bullet} V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet} V_{n,t})\right](X_t - c_{n,t}) = o_{P^n_{\theta_n}}(n^{-1/2})$;

(ii) $\frac{1}{n}\sum_{t=1}^n \left[\phi_k(A_{n,k\bullet} V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet} V_{n,t})\right]\left(\bar{X}_n - \bar{c}_n\right) = o_{P^n_{\theta_n}}(n^{-1/2})$;

(iii) $\frac{1}{n}\sum_{t=1}^n \left[\phi_k(A_{n,k\bullet} V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet} V_{n,t})\right](\bar{c}_n - c_{n,t}) = o_{P^n_{\theta_n}}(n^{-1/2})$;

(iv) $\frac{1}{n}\sum_{t=1}^n \phi_k(A_{n,k\bullet} V_{n,t})\left(\bar{X}_n - \bar{c}_n\right) = o_{P^n_{\theta_n}}(n^{-1/2})$;

(v) $\frac{1}{n}\sum_{t=1}^n \phi_k(A_{n,k\bullet} V_{n,t})(\bar{c}_n - c_{n,t}) = o_{P^n_{\theta_n}}(n^{-1/2})$;

(vi) $\frac{1}{n}\sum_{t=1}^n \bar{X}_n\left[\tilde{\varsigma}_{k,n,1} A_{n,k\bullet} V_{n,t} + \tilde{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet} V_{n,t})\right] = o_{P^n_{\theta_n}}(n^{-1/2})$;

(vii) $\frac{1}{n}\sum_{t=1}^n (\bar{X}_n - \bar{c}_n)\left[\varsigma_{k,1} A_{n,k\bullet} V_{n,t} + \varsigma_{k,2}\kappa(A_{n,k\bullet} V_{n,t})\right] = o_{P^n_{\theta_n}}(n^{-1/2})$;

(viii) $\frac{1}{n}\sum_{t=1}^n (\bar{c}_n - c_{n,t})\left[\varsigma_{k,1} A_{n,k\bullet} V_{n,t} + \varsigma_{k,2}\kappa(A_{n,k\bullet} V_{n,t})\right] = o_{P^n_{\theta_n}}(n^{-1/2})$

(i) follows by (the first part of) Lemma A.1 applied with $W_{n,t} = X_t - c_{n,t}$. This is mean-zero, independent of all $A_{n,k\bullet} V_{n,s}$ with $s \geq t$ and has uniformly bounded second moments (cf. (S6)).

(ii) follows by Jensen's inequality, (the second part of) Lemma A.1 applied with $W_{n,t} = 1$, (S6), Lemma S2.14 and Corollary 3.1.

(iii) follows by Cauchy – Schwarz, (the second part of) Lemma A.1 applied with $W_{n,t} = 1$ and Lemma S2.17.

For (iv), $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \phi_k(A_{n,k\bullet} V_{n,t}) = O_{P_{\theta_n}^n}(1)$ by the central limit theorem and $\bar{X}_n - \bar{c}_n = \frac{1}{n}\sum_{t=1}^{n}[X_t - c_{n,t}] \xrightarrow{P_{\theta_n}} 0$, which follows by (S6), Lemma S2.14 and Corollary 3.1.

(v) follows by Cauchy – Schwarz, the fact that $\mathbb{E}\,\phi_k(A_{n,k\bullet}V_{n,t})^2 = \mathbb{E}\,\phi_k(\epsilon_{k,t})^2$ is uniformly bounded hence $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2 = O_{P_{\theta_n}^n}(1)$ by Markov's inequality and Lemma S2.17.

For (vi), $\bar{X}_n = O_{P_{\theta_n}}(1)$ by e.g. Markov's inequality and (S6). By the central limit theorem also $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t = O_{P_{\theta_n}^n(1)}$ for U_t equal to either $A_{n,k\bullet}V_{n,t}$ or $\kappa(A_{n,k\bullet}V_{n,t})$. The result therefore follows from Lemma S2.16.

For (vii), as for (vi), $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t = O_{P_{\theta_n}^n(1)}$ for U_t equal to either $A_{n,k\bullet}V_{n,t}$ or $\kappa(A_{n,k\bullet}V_{n,t})$. Therefore it suffices to note that $\bar{X}_n - \bar{c}_n \xrightarrow{P_{\theta_n}} 0$, as noted for (iv).

For (viii), for U_t equal to either $\varsigma_{k,1}A_{n,k\bullet}V_{n,t}$ or $\varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})$, by Markov's inequality

$$P_{\theta_n}^n\left(\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\bar{c}_n - c_{n,t})U_t\right\| > \varepsilon\right) \leq \varepsilon^{-2}\,\mathbb{E}\,U_t^2\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2 \lesssim \frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2 \to 0,$$

by Lemma S2.17.

To verify (iii) we note that

$$\left\|\hat{I}_{n,\theta_n} - \tilde{I}_\theta\right\|_2 \leq \left\|\hat{I}_{n,\theta_n} - \breve{I}_{n,\theta_n}\right\|_2 + \left\|\breve{I}_{n,\theta_n} - \tilde{I}_{n,\theta_n}\right\|_2 + \left\|\tilde{I}_{n,\theta_n} - \tilde{I}_\theta\right\|_2 \tag{S21}$$

where $\tilde{I}_\theta := \mathbb{E}[\tilde{\ell}_\theta(Y_t,X_t)\tilde{\ell}_\theta(Y_t,X_t)'] = \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[\tilde{\ell}_\theta(Y_t,X_t)\tilde{\ell}_\theta(Y_t,X_t)']$ with the expectation taken under G_θ, $\hat{I}_{n,\theta} := \frac{1}{n}\sum_{t=1}^{n}\hat{\ell}_\theta(Y_t,X_t)\hat{\ell}_\theta(Y_t,X_t)'$ and $\breve{I}_{n,\theta} := \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_\theta(Y_t,X_t)\tilde{\ell}_\theta(Y_t,X_t)'$. We will show each right hand side term is $o_{P_{\theta_n}^n}(\nu_n^{1/2})$.

For the first right hand side term in (S21) let $r \in \{\alpha,\sigma,b\}$ and let l denote an index, we write $\hat{U}_{n,t,r_l} := \hat{\ell}_{\theta_n,r_l}(Y_t,X_t)$, $\tilde{U}_{t,r_l} := \tilde{\ell}_{\theta_n,r_l}(Y_t,X_t)$ and $D_{n,t,r_l} := \hat{\ell}_{\theta_n,r_l}(Y_t,X_t) - \tilde{\ell}_{\theta_n,r_l}(Y_t,X_t)$.

Since it is the absolute value of the $(r,l)-(s,m)$ component of $\hat{I}_{n,\theta_n} - \breve{I}_{n,\theta_n}$, it is sufficient to show that $\left|\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m} + \frac{1}{n}\sum_{t=1}^{n}D_{n,t,r_l}\tilde{U}_{t,s_m}\right| = o_{P_{\theta_n}^n}(\nu_n^{1/2})$ as $n \to \infty$ for any $r,s \in \{\alpha,\sigma,b\}$ and l,m. By Cauchy-Schwarz and Lemma S2.19

$$\left|\frac{1}{n}\sum_{t=1}^{n}D_{n,t,r_l}\tilde{U}_{t,s_m}\right| \leq \left(\frac{1}{n}\sum_{t=1}^{n}\tilde{U}_{t,s_m}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}D_{n,t,r_l}^2\right)^{1/2} = O_{P_{\theta_n}^n}(1)\times o_{P_{\theta_n}^n}(\nu_n^{1/2}) = o_{P_{\theta_n}^n}(\nu_n^{1/2}),$$

$$\left|\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m}\right| \leq \left(\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}D_{n,t,s_m}^2\right)^{1/2} = O_{P_{\theta_n}^n}(1)\times o_{P_{\theta_n}^n}(\nu_n^{1/2}) = o_{P_{\theta_n}^n}(\nu_n^{1/2}),$$

for any $(r,l)-(s,m)$. It follows that

$$\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m}+D_{n,t,r_l}\tilde{U}_{t,s_m}\right]^2\le 2\left[\frac{1}{n}\sum_{t=1}^{n}\hat{U}_{n,t,r_l}D_{n,t,s_m}\right]^2+2\left[\frac{1}{n}\sum_{t=1}^{n}D_{n,t,r_l}\tilde{U}_{t,s_m}\right]^2=o_{P_{\theta_n}^n}(\nu_n)$$

and hence $\|\hat{I}_{n,\theta_n}-\breve{I}_{n,\theta_n}\|_2\le\|\hat{I}_{n,\theta_n}-\breve{I}_{n,\theta_n}\|_F=o_{P_{\theta_n}^n}(\nu_n^{1/2})$

For the second right hand side term in (S21), Let $Q_{l,m,t,n}^{r,ls}=\tilde{\ell}_{\theta_n,r_l}(Y_t,X_t)\tilde{\ell}_{\theta_n,s_m}(Y_t,X_t)$, where $r,s\in\{\alpha,\sigma,b\}$ and l,m denote the indices of the components of the efficient scores. Fix any r,s and l,m and note that by the fact that $\tilde{\ell}_{\theta_n}$ has uniformly bounded $2+\delta/2$ moments under $P_{\theta_n}^n$, Theorem 3 of Saikkonen (2007) and Theorem 1 of Kanaya (2017) together imply that (cf. Lemma S2.14)

$$\frac{1}{n}\sum_{t=1}^{n}Q_{l,m,t,n}^{r,s}-\mathbb{E}_{\theta_n}Q_{l,m,t,n}^{r,s}=O_{P_{\theta_n}^n}\left(n^{(1/p-1)/2}\right)=o_{P_{\theta_n}^n}(\nu_n^{1/2}),\quad p\in(1,1+\delta/4],$$

hence $\|\breve{I}_{n,\theta_n}-\tilde{I}_{n,\theta_n}\|_2=o_{P_{\theta_n}^n}(\nu_n^{1/2})$.

That the last right hand side term in (S21) is $o(\nu_n^{1/2})$ follows from the assumed local Lipschitz continuity of the map defining the ζ's, that of each $\beta\mapsto A(\alpha,\sigma)_{k\bullet}$, Theorem 11.11 of Kallenberg (2021) and Lemma S2.18. □

Lemma S2.16: *If assumption 2.1 holds, then* $\|\hat{\varrho}_{k,n}-\varrho_{k,n}\|_2=o_{P_{\tilde{\theta}_n}^n}(\nu_{n,p})=o_{P_{\theta_n}^n}(\nu_n^{1/2})$, *where* $\tilde{\theta}_n$ *is as in Lemma S2.15 and* $\varrho\in\{\tau,\varsigma\}$.

Proof. Under $P_{\theta_n}^n$, $A_{n,k\bullet}V_{n,t}\approx\epsilon_{k,t}\sim\eta_k$, for $V_{n,t}:=Y_t-B_nX_t$ and $A_n:=A(\theta_n)$. Let $w\in\{(0,-2)',(1,0)'\}$ Since the map $M\mapsto M^{-1}$ is Lipschitz at a positive definite matrix M_0, then for large enough n, with probability approaching one

$$\|\hat{\varrho}_{k,n}-\varrho_{k,n}\|_2=\|(\hat{M}_{k,n}^{-1}-M_k^{-1})w\|_2\le 2\|\hat{M}_{k,n}^{-1}-M_k^{-1}\|_2\le 2C\|\hat{M}_{k,n}-M_k\|_2, \tag{S22}$$

for some positive constant C. By Theorem 2.5.11 in Durrett (2019)

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}[(A_{n,k\bullet}V_{n,t})^3-\mathbb{E}(A_{n,k\bullet}V_{n,t})^3]&=o_{P_{\theta_n}^n}\left(n^{\frac{1-p}{p}}\right)\\ \frac{1}{n}\sum_{t=1}^{n}[(A_{n,k\bullet}V_{n,t})^4-\mathbb{E}(A_{n,k\bullet}V_{n,t})^4]&=o_{P_{\theta_n}^n}\left(n^{\frac{1-p}{p}}\right).\end{aligned}$$

These together imply that

$$\|\hat{M}_{k,n}-M_k\|_2\le\|\hat{M}_{k,n}-M_k\|_F=o_{P_{\theta_n}^n}\left(n^{\frac{1-p}{p}}\right)=o_{P_{\theta_n}^n}(\nu_{n,p}).$$

Combining these convergence rates with equation (S22) yields the result. $\square$

LEMMA S2.17: *In the setting of Lemma S2.15, let $c_{n,t} := \mathbb{E}_{\theta_n} X_t$ and $\bar{c}_n := \frac{1}{n}\sum_{t=1}^n c_{n,t}$. Then*

$$\frac{1}{n}\sum_{t=1}^n \|\bar{c}_n - c_{n,t}\|^2 = O(n^{-1}).$$

Proof. Since $X_t = (1, Z'_{t-1})'$, it suffices to show that $\frac{1}{n}\sum_{t=1}^n \left\|\tilde{c}_{n,t} - \frac{1}{n}\sum_{t=1}^n \tilde{c}_{n,t}\right\|^2 = O(n^{-1})$ for $\tilde{c}_{n,t} := \mathbb{E}_{\theta_n} Z_{t-1}$. Let $\tilde{c}_{n,\infty} := \sum_{j=0}^\infty \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n}$. This converges uniformly in n since under Assumption 2.1 parts (i) & (iii), the sets $\{\|\mathsf{B}_{\theta_n}\|_2 : n \in \mathbb{N}\} \cup \{\|\mathsf{B}_\theta\|_2\}$ and $\{\|\mathsf{C}_{\theta_n}\|_2 : n \in \mathbb{N}\} \cup \{\|\mathsf{C}_\theta\|_2\}$ are bounded above by $\rho_\star < 1$ and $C_\star < \infty$ respectively. By Jensen's inequality

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^n \left\|\tilde{c}_{n,t} - \frac{1}{n}\sum_{t=1}^n \tilde{c}_{n,t}\right\|^2 &\lesssim \frac{1}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2 + \frac{1}{n}\sum_{t=1}^n \left\|\frac{1}{n}\sum_{t=1}^n [\tilde{c}_{n,\infty} - \tilde{c}_{n,t}]\right\|^2 \\
&\le \frac{2}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2
\end{aligned}$$

so it suffices to show that $n/2$ times the last term is uniformly bounded above. One has:

$$\begin{aligned}
\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2 &= \sum_{t=1}^n \left\|\sum_{j=t-1}^\infty \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n} - \mathsf{B}_{\theta_n}^{t-1} Z_0\right\|^2 \\
&\lesssim \sum_{t=1}^n \left\|\sum_{j=t-1}^\infty \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n}\right\|^2 + \sum_{t=1}^n \left\|\mathsf{B}_{\theta_n}^{t-1} Z_0\right\|^2 \\
&\le \sum_{t=1}^n \left[\sum_{j=t-1}^\infty \|\mathsf{B}_{\theta_n}\|_2^j \|\mathsf{C}_{\theta_n}\|_2\right]^2 + \sum_{t=1}^n \|B_{\theta_n}\|_2^{2(t-1)} \|Z_0\|^2 \\
&\le C_\star^2 \sum_{t=1}^n \left[\frac{\rho_\star^{t-1}}{1-\rho_\star}\right]^2 + \|Z_0\|^2 \sum_{t=1}^n \rho_\star^{2(t-1)} \\
&\le \left[\frac{C_\star^2}{(1-\rho_\star)^2} + \|Z_0\|^2\right] \frac{1}{1-\rho_\star^2}.
\end{aligned}$$

$\square$

LEMMA S2.18: *In the setting of Lemma S2.15, let $\tilde{X}_t = (1, \tilde{Y}'_{t-1}, \dots, \tilde{Y}'_{t-p})'$ where $\tilde{Y}_t$ is a stationary solution to (1). Then,*

(i) $\frac{1}{n}\sum_{t=1}^n \mathbb{E}_{\theta_n} X_t - \mathbb{E}_\theta \tilde{X}_t = o(\nu_n^{1/2})$,

(ii) $\frac{1}{n}\sum_{t=1}^n [\mathbb{E}_{\theta_n} X_t][\mathbb{E}_{\theta_n} X_t]' - [\mathbb{E}_\theta \tilde{X}_t][\mathbb{E}_\theta \tilde{X}_t]' = o(\nu_n^{1/2})$.

(iii) $\frac{1}{n}\sum_{t=1}^n \mathbb{E}_{\theta_n}[X_t - \mathbb{E}_{\theta_n} X_t][X_t - \mathbb{E}_{\theta_n} X_t]' - \mathbb{E}_\theta[X_t - \mathbb{E}_\theta X_t][X_t - \mathbb{E}_\theta X_t]' = o(\nu_n^{1/2})$.

Proof. Note that $\|\mathbb{E}_{\theta_n} X_t - \mathbb{E}_{\theta_n} \tilde{X}_t\|^2 \le \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2$ in the notation of (the proof of) Lemma

S2.17, which shows that $\frac{1}{n}\sum_{t=1}^n \|\tilde{c}_{n,t} - \tilde{c}_{n,\infty}\|^2 = O(n^{-1})$. Hence by Jensen's inequality,

$$\frac{1}{n}\sum_{t=1}^{n} \| \mathbb{E}_{\theta_n} X_t - \mathbb{E}_{\theta_n} \tilde{X}_t\| = O(n^{-1/2}) = o(\nu_n^{1/2}),$$

Since $\beta \mapsto \mathbb{E}_\theta \tilde{X}_t = \operatorname{vec}(\iota_K, (\iota_p \otimes (I_K - B_1 - \ldots - B_p)^{-1} c))$ is locally Lipschitz,

$$\frac{1}{n}\sum_{t=1}^{n} \| \mathbb{E}_{\theta_n} \tilde{X}_t - \mathbb{E}_{\theta} \tilde{X}_t\| = \| \mathbb{E}_{\theta_n} \tilde{X}_t - \mathbb{E}_{\theta} \tilde{X}_t\| = O(n^{-1/2}) = o(\nu_n^{1/2}).$$

Combination of the above two displays yields that $\frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\theta_n} X_t - \mathbb{E}_\theta \tilde{X}_t\| = O(n^{-1/2}) = o(\nu_n^{1/2})$ which implies (i). Moreover, combined with the uniform moment bounds given in (S6) and Lemma S2.1 this yields

$$\frac{1}{n}\sum_{t=1}^{n} \|[\mathbb{E}_{\theta_n} X_t][\mathbb{E}_{\theta_n} X_t]' - [\mathbb{E}_{\theta} \tilde{X}_t][\mathbb{E}_{\theta} \tilde{X}_t]'\| \lesssim \frac{1}{n}\sum_{t=1}^{n} \| \mathbb{E}_{\theta_n} X_t - \mathbb{E}_{\theta} \tilde{X}_t\| = O(n^{-1/2}) = o(\nu_n^{1/2}),$$

which implies (ii).

For (iii) let $U_{\vartheta,t} := X_t - \mathbb{E}_\vartheta X_t$ and $\tilde{U}_{\vartheta,t} := \tilde{X}_t - \mathbb{E}_\vartheta \tilde{X}_t$. Note that as $U_{\vartheta,t} = \sum_{j=0}^{t-2} \mathsf{B}_\vartheta^j \mathsf{D}_\vartheta \epsilon_{t-j}$ and $\tilde{U}_{\vartheta,t} = \sum_{j=0}^{\infty} \mathsf{B}_\vartheta^j \mathsf{D}_\vartheta \epsilon_{t-j}$, $U_{\theta_n,t} - \tilde{U}_{\theta_n,t}$ and $U_{\theta_n,t}$ are independent. Additionally by Assumption 2.1 parts (i) and (iii) the sets the sets $\{\|\mathsf{B}_{\theta_n}\|_2 : n \in \mathbb{N}\}$ and $\{\|\mathsf{D}_{\theta_n}\|_2 : n \in \mathbb{N}\}$ are bounded above by $\rho_\star < 1$ and $D_\star < \infty$ respectively. Hence

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \left\|\mathbb{E}_{\theta_n}\left[U_{\theta_n,t}U_{\theta_n,t}' - \tilde{U}_{\theta_n,t}\tilde{U}_{\theta_n,t}'\right]\right\| \\
&\quad \le \frac{1}{n}\sum_{t=1}^{n} \left\|\mathbb{E}_{\theta_n}\left[\left(U_{\theta_n,t} - \tilde{U}_{\theta_n,t}\right) U_{\theta_n,t}'\right]\right\| + \frac{1}{n}\sum_{t=1}^{n} \left\|\mathbb{E}_{\theta_n}\left[\left(U_{\theta_n,t} - \tilde{U}_{\theta_n,t}\right) \tilde{U}_{\theta_n,t}'\right]\right\| \\
&\quad \le \frac{1}{n}\sum_{t=1}^{n} \left\|\mathbb{E}_{\theta_n} \sum_{k=0}^{\infty}\sum_{j=t-1}^{\infty} \mathsf{B}_{\theta_n}^j \mathsf{D}_{\theta_n} \epsilon_{t-j}\epsilon_{t-k}' \mathsf{D}_{\theta_n}' (\mathsf{B}_{\theta_n}^j)'\right\| \\
&\quad \le \frac{1}{n}\sum_{t=1}^{n}\sum_{j=t-1}^{\infty} \|\mathsf{B}_{\theta_n}\|_2^{2j} \|\mathsf{D}_{\theta_n}\|_2^2 \\
&\quad \le D_\star^2 \times \frac{1}{n}\sum_{t=1}^{n}\sum_{j=t-1}^{\infty} \rho_\star^{2j} \\
&\quad \le \frac{D_\star^2}{1-\rho_\star^2} \times \frac{1-\rho_\star^{2n}}{1-\rho_\star^2} \times \frac{1}{n} \\
&\quad = O(n^{-1}).
\end{aligned}$$

Additionally, we can write $\operatorname{vec}(\mathbb{E}_\vartheta \tilde{U}_{\vartheta,t}\tilde{U}_{\vartheta,t}') = (I - \mathsf{B}_\vartheta \otimes \mathsf{B}_\vartheta)^{-1} \operatorname{vec}(\mathsf{D}_\vartheta \mathsf{D}_\vartheta')$, which is locally

Lipschitz in β at θ. This implies that

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\theta_n}\tilde{U}_{\theta_n,t}\tilde{U}_{\theta_n,t}' - \mathbb{E}_{\theta}\tilde{U}_{\theta,t}\tilde{U}_{\theta,t}' = O(n^{-1/2}) = o(\nu_n^{1/2}).$$

The previous two displays suffice for (iii). $\square$

Lemma S2.19: *In the setting of Lemma S2.15, for each $r \in \{\alpha, \sigma, b\}$ and l*

$$\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\tilde{\theta}_n,r_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n,r_l}(Y_t, X_t)\right)^2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$

Proof. We start by considering elements in $\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t, X_t)\right)^2$. Define $\tilde{\tau}_{k,n,q} := \hat{\tau}_{k,n,q} - \tau_{k,q}$ and $V_{n,t} = Y_t - B_n X_t$. Since each $|\zeta^{\alpha}_{n,l,k,j}| < \infty$ and the sums over k, j are finite, it is sufficient to demonstrate that for every $k, j, m, s \in [K]$, with $k \neq j$ and $s \neq m$,

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right]\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\right]A_{n,j\bullet}V_{t,n}A_{n,m\bullet}V_{n,t} \tag{S23}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right]A_{n,j\bullet}V_{n,t}\left[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{n,t} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{n,t})\right] \tag{S24}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{n,t} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{n,t})\right]\left[\tilde{\tau}_{k,n,1}A_{n,k\bullet}V_{n,t} + \tilde{\tau}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})\right] \tag{S25}$$

are each $o_{P^n_{\tilde{\theta}_n}}(\nu_n)$.

For (S25), let $\xi_1(x) = x$ and $\xi_2(x) = \kappa(x)$. Then, we can split the sum into 4 parts, each of which has the following form for some $q, w \in \{1, 2\}$

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w}\xi_q(A_{n,s\bullet}V_{n,t})\xi_w(A_{n,k\bullet}V_{n,t}) = \tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w}\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{n,t})\xi_w(A_{n,k\bullet}V_{n,t}) = o_{P^n_{\tilde{\theta}_n}}(\nu_n),$$

since we have that each $\tilde{\tau}_{s,n,q}\tilde{\tau}_{k,n,w} = o_{P^n_{\tilde{\theta}_n}}(\nu_n)$ by lemma S2.16.[S12] For (S24) we can argue similarly. Again let $\xi_1(x) = x$ and $\xi_2(x) = \kappa(x)$. Then, we can split the sum into 2 parts, each

[S12] The fact that $\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{n,t})\xi_w(A_{n,k\bullet}V_{n,t}) = O_{P^n_{\tilde{\theta}_n}}(1)$ can be seem to hold using the moment and i.i.d. assumptions from assumption 2.1 and Markov's inequality, noting once more that $A_{n,k\bullet}V_{n,t} \simeq \epsilon_{k,t}$ under $P^n_{\tilde{\theta}_n}$.

of which has the following form for some $q \in \{1, 2\}$

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right] A_{n,j\bullet}V_{n,t}\tilde{\tau}_{s,n,q}\xi_q(A_{n,s\bullet}V_{n,t}) \\
&\qquad \leq \tilde{\tau}_{s,n,q}\left(\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right]^2 (A_{n,j\bullet}V_{n,t})^2\right)^{1/2} \left(\frac{1}{n}\sum_{t=1}^{n}\xi_q(A_{n,s\bullet}V_{n,t})^2\right)^{1/2} \\
&\qquad = o_{P^n_{\tilde{\theta}_n}}(\nu_n).
\end{aligned}
$$

by Lemma A.1 applied with $W_{n,t} = A_{n,j\bullet}V_{n,t}$ and $\tilde{\tau}_{s,n,q} = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$.[S13] For (S23) use Cauchy-Schwarz with Lemma A.1

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right]\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\right] A_{n,j\bullet}V_{n,t}A_{n,m\bullet}V_{n,t} \\
&\qquad \leq \left(\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right]^2 (A_{n,j\bullet}V_{n,t})^2\right)^{1/2} \\
&\qquad\qquad \times \left(\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\right]^2 (A_{n,m\bullet}V_{n,t})^2\right)^{1/2} \\
&\qquad = o_{P^n_{\tilde{\theta}_n}}(\nu_n).
\end{aligned}
$$

This completes the proof for the components corresponding to α_l. We note that the components corresponding to σ_l follow analogously.

Finally, we consider the elements in $\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\theta_n,b_l}(Y_t, X_t) - \tilde{\ell}_{\theta_n,b_l}(Y_t, X_t)\right)^2$. Let $a_{n,k,l} := -A_{n,k\bullet}D_{b_l}$, $\tilde{\varsigma}_{k,n} := \hat{\varsigma}_{k,n} - \varsigma_k$, $c_{n,t} := \mathbb{E}_{\theta_n} X_t$ and $\bar{c}_n := \frac{1}{n}\sum_{t=1}^{n} c_{n,t}$. Since $a_{n,k,l} \to a_{\infty,k,l} := A(\alpha, \sigma)_{k\bullet}D_{b_l}(\alpha, \sigma)$, it suffices to show that

(i) $\frac{1}{n}\sum_{t=1}^{n}\left[\phi_k(A_{n,k\bullet}V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})\right]^2 \|X_t - c_{n,t}\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(ii) $\frac{1}{n}\sum_{t=1}^{n}\left[\phi_k(A_{n,k\bullet}V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})\right]^2 \|\bar{X}_n - \bar{c}_n\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(iii) $\frac{1}{n}\sum_{t=1}^{n}\left[\phi_k(A_{n,k\bullet}V_{n,t}) - \hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t})\right]^2 \|\bar{c}_n - c_{n,t}\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(iv) $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2 \left\|\bar{X}_n - \bar{c}_n\right\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(v) $\frac{1}{n}\sum_{t=1}^{n}\phi_k(A_{n,k\bullet}V_{n,t})^2 \|\bar{c}_n - c_{n,t}\|^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(vi) $\frac{1}{n}\sum_{t=1}^{n}\|\bar{X}_n\|^2 [\tilde{\varsigma}_{k,n,1}A_{n,k\bullet}V_{n,t} + \tilde{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})]^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(vii) $\frac{1}{n}\sum_{t=1}^{n}\|\bar{X}_n - \bar{c}_n\|^2 [\varsigma_{k,1}A_{n,k\bullet}V_{n,t} + \varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})]^2 = o_{P^n_{\theta_n}}(\nu_n)$;

(viii) $\frac{1}{n}\sum_{t=1}^{n}\|\bar{c}_n - c_{n,t}\|^2 [\varsigma_{k,1}A_{n,k\bullet}V_{n,t} + \varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})]^2 = o_{P^n_{\theta_n}}(\nu_n)$.

[S13] See footnote S12.

(i) follows from repeated application of Lemma A.1 with $W_{n,t} = e_j'(X_t - c_{n,t})$.

(ii) follows from application of Lemma A.1 with $W_{n,t} = 1$ and $\bar{X}_n - \bar{c}_n = \frac{1}{n}\sum_{t=1}^n [X_t - c_{n,t}] \xrightarrow{P_{\theta_n}} 0$, which follows by (S6), Lemma S2.14 and Corollary 3.1.

(iii) follows by Lemma A.1 applied repeatedly with $W_{n,t} = e_j'(\bar{c}_n - c_{n,t})$.[S14]

For (iv), $\frac{1}{n}\sum_{t=1}^n \phi_k(A_{n,k\bullet}V_{n,t})^2 = O_{P_{\theta_n}^n}(1)$ since $\phi_k(A_{n,k\bullet}V_{n,t})^2$ has uniformly bounded second moments and $\bar{X}_n - \bar{c}_n = O_{P_{\theta_n}^n}(n^{-1/2})$, by (S6), Lemma S2.14 and Corollary 3.1.

For (v) use Markov's inequality and Lemma S2.17 to conclude

$$P_{\theta_n}^n\left(\frac{1}{n}\sum_{t=1}^n \phi_k(A_{n,k\bullet}V_{n,t})^2\,\|\bar{c}_n - c_{n,t}\|^2 > \nu_n\varepsilon\right) \leq \nu_n^{-1}\varepsilon^{-1}\,\mathbb{E}\left[\phi_k(\epsilon_k)^2\right]\frac{1}{n}\sum_{t=1}^n \|\bar{c}_n - c_{n,t}\|^2 \to 0.$$

For (vi), $\bar{X}_n = O_{P_{\theta_n}}(1)$ by e.g. Markov's inequality and (S6). Similarly, $\frac{1}{n}\sum_{t=1}^n U_{t,i}U_{t,j} = O_{P_{\theta_n}^n}(1)$ for $i, j \in \{1, 2\}$ with $U_{t,1} = A_{n,k\bullet}V_{n,t}$ and $U_{t,2} = \kappa(A_{n,k\bullet}V_{n,t})$. The result then follows from Lemma S2.16.

For (vii), $\frac{1}{n}\sum_{t=1}^n U_{t,i}U_{t,j} = O_{P_{\theta_n}^n}(1)$ for $i, j \in \{1, 2\}$ with $U_{t,1}$ and $U_{t,2}$ as in the preceding paragraph. Therefore it suffices to note that $\bar{X}_n - \bar{c}_n = O_{P_{\theta_n}}(n^{-1/2})$, as noted for (iv).

For (viii), for $U_{t,1}$ and $U_{t,2}$ as in the preceding paragraph and $i, j \in \{1, 2\}$,

$$\begin{aligned} P_{\theta_n}^n\left(\left|\frac{1}{n}\sum_{t=1}^n \|\bar{c}_n - c_{n,t}\|^2 \varsigma_{k,i}U_{t,i}\varsigma_{k,j}U_{t,j}\right| > \nu_n\varepsilon\right) &\leq \nu_n^{-1}\varepsilon^{-1}|\varsigma_{k,i}\varsigma_{k,j}|[\mathbb{E}\,U_{t,i}^2]^{1/2}[\mathbb{E}\,U_{t,j}^2]^{1/2}\frac{1}{n}\sum_{t=1}^n \|\bar{c}_n - c_{n,t}\|^2 \\ &\lesssim \nu_n^{-1}\frac{1}{n}\sum_{t=1}^n \|\bar{c}_n - c_{n,t}\|^2 \to 0, \end{aligned}$$

by Markov's inequality and Lemma S2.17. □

S2.5 Assumption 2.1-(ii)-(b)

We provide a sufficient condition under which Assumption 2.1 part (ii)-(b) holds, given part (ii)-(a). For convenience recall that part (ii) reads as

(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 $\delta > 0$);

[S14] That this is uniformly bounded follows from (S6).

(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 $\mathbb{E}\,\phi_k(\epsilon_{k,t})\epsilon_{k,t}^3 = -3$;

In this assumption part (a) is standard — only imposes that the shocks are mean zero with unit variance, and that certain $4+\delta$ moments are finite —. In contrast, part (b) may seem strong at first sight.

An important observation is that (b) should not be understood independently from (a). Indeed, the following lemma shows that given (a), condition (b) follows if the structural shocks have densities that decays to zero at a polynomial rate.

Lemma S2.20: *Let $a_k = \inf\{x \in \mathbb{R} \cup \{-\infty\} : \eta_k(x) > 0\}$ and $b_k = \sup\{x \in \mathbb{R} \cup \{\infty\} : \eta_k(x) > 0\}$. Suppose that, for $r = 0,1,2,3$: (i) if $a_k = -\infty$ then $\eta_k(x) = o(x^{-3})$ as $x \to -\infty$, else $a_k^r \lim_{x \downarrow a_k} \eta_k(x) = 0$, and (ii) if $b_k = \infty$ then $\eta_k(x) = o(x^{-3})$ as $x \to \infty$, else $b_k^r \lim_{x \uparrow b_k} \eta_k(x) = 0$. Then, if part (a) of assumption 2.1-(ii) holds, part (b) is also satisfied.*

Proof. Let $r \in \{0,1,2,3\}$, $b_k = \sup\{x \in \mathbb{R} : \eta_k(x) > 0\}$ and $a_k = \inf\{x \in \mathbb{R} : \eta_k(x) > 0\}$. We have, by integration by parts, with G_k denoting the measure on $\mathbb{R}$ corresponding to η_k,

$$\int \phi_k(z) z^r \,\mathrm{d}G_k = \int \frac{\eta_k'(z)}{\eta_k(z)} \eta_k(z) z^r \,\mathrm{d}z = \int \eta_k'(z) z^r \,\mathrm{d}z = \eta_k(z) z^r \Big|_{a_k}^{b_k} - \int \eta_k(z) \frac{\mathrm{d}z^r}{\mathrm{d}z} \,\mathrm{d}z.$$

Our hypothesis ensures that $z^r \eta_k(z)\big|_{a_k}^{b_k} = 0$. Therefore we have $G_k \phi_k(z) z^r = -G_k \frac{\mathrm{d}}{\mathrm{d}z} z^r$. For $r = 0$ this equals zero as $\frac{\mathrm{d}}{\mathrm{d}z} z^0 = \frac{\mathrm{d}}{\mathrm{d}z} 1 = 0$. For $r \in \{1,2,3\}$ we have $\frac{\mathrm{d}z^r}{\mathrm{d}z} = r z^{r-1}$ and hence $G_k \phi_k(z) z^r = -r G_k z^{r-1}$. Since $G_k 1 = 1$, $G_k z = 0$, and $G_k z^2 = 1$, the result follows. □

We now provide two examples. The first is a mixture of normals. We directly verify the moment conditions in (a) and (b) are satisfied.

The second example is a normalised χ_2^2 distribution. We show that this does satisfy the moment conditions in (a) but not those in (b) (nor the conditions of Lemma S2.20).[S15]

Example S2.1 (Normal mixtures): *Suppose that ϵ_k has the density function*

$$\eta_k(z) = \sum_{m=1}^{M} p_m f_m(z, \mu_m, \sigma_m^2), \quad p_m \geq 0, \quad \sum_{m=1}^{M} p_m = 1, \quad \sum_{m=1}^{M} p_m \mu_m = 0, \quad \sum_{m=1}^{M} p_m(\sigma_m^2 + \mu_m^2) = 1,$$

where $f_m(z, \mu_m, \sigma_m^2)$ is the density function of a $e_m \sim \mathcal{N}(\mu_m, \sigma_m^2)$.

ϵ_k has mean zero and unit variance. We first establish that each of the conditions in (a) are

[S15] Additionally, the (normalised) χ_2^2 distribution does not have a nowhere vanishing Lebesgue density.

satisfied. In particular we first note that $\mathbb{E}[|\epsilon_k|^r]$ *is finite for any positive integer* r *as*

$$\mathbb{E}[|\epsilon_k|^r] = \sum_{m=1}^{M} p_m \, \mathbb{E}[|e_m|^r] < \infty, \tag{S26}$$

since the Normal distribution has finite moments of all orders. To establish that $\mathbb{E}[\epsilon_k^3]^2 < \mathbb{E}[\epsilon_k^4] - 1$ *note that this is equivalent to the linear independence in* L_2 *of* $1, \epsilon_k, \epsilon_k^2$ *(e.g. Horn and Johnson, 2013, Theorem 7.2.10). This is equivalent to the condition that*

$$a_1^2 + 2a_1a_3 + a_2^2 + a_3^2\,\mathbb{E}[\epsilon_k^4] = 0 \implies a_1 = a_2 = a_3 = 0.$$

This holds since $\mathbb{E}[\epsilon_k^4] \geq 1 = \mathbb{E}[\epsilon_k^2]$ *by the fact that* L_p *norms are increasing and so*

$$a_1^2 + 2a_1a_3 + a_2^2 + a_3^2\,\mathbb{E}[\epsilon_k^4] \geq a_1^2 + 2a_1a_3 + a_3^2 = (a_1 + a_3)^2 \geq 0,$$

where equality is possible only if $a_1 = a_2 = a_3 = 0$*. Next, note that*

$$\phi_k(z) = -\frac{\sum_{m=1}^{M} p_m \sigma_m^{-2}(z - \mu_m) f_m(z, \mu_m, \sigma_m^2)}{\eta_k(z)}, \tag{S27}$$

and for any integer r *and some* $\mu \in \mathbb{R}$

$$|\phi_k(z)|^r \lesssim |\phi_k(z)|^{r-1} \left| \eta_k(z)^{-1}(|z| + |\mu|) \sum_{m=1}^{M} p_m f_m(z, \mu_m, \sigma_m^2) \right| = |\phi_k(z)|^{r-1}(|z| + |\mu|).$$

Recursively using this inequality from $r = 0$*, yields (for some constant* $C_r \in (0, \infty)$*)*

$$|\phi_k(z)|^r \leq C_r(|z|^r + |\mu|^r).$$

That $\mathbb{E}\,|\phi(\epsilon_k)|^r < \infty$ *for any integer* r *then follows from* (S26).

For the conditions in (b), note that by (S27),

$$\begin{aligned}
\mathbb{E}[\phi_k(\epsilon_k)\epsilon_k^r] &= -\sum_{m=1}^{M} p_m \int z^r \frac{\sigma_m^{-2}(z - \mu_m) f_m(\epsilon_k, \mu_m, \sigma_m^2)}{\eta_k(z)} \eta_k(z)\,\mathrm{d}z \\
&= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \int z^r (z - \mu_m) f_m(\epsilon_k, \mu_m, \sigma_m^2)\,\mathrm{d}z \\
&= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \left(\mathbb{E}\left[e_m^{r+1}\right] - \mathbb{E}\left[e_m^r\right] \mu_m \right).
\end{aligned}$$

Taking $r = 0, 1, 2, 3$ in the right hand expression respectively gives:

$$\begin{aligned}
\mathbb{E}[\phi_k(\epsilon_k)] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2} (\mu_m - \mu_m) = 0 \ , \\
\mathbb{E}[\phi_k(\epsilon_k)\epsilon_k] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \left(\sigma_m^2 + \mu_m^2 - \mu_m^2\right) = -1 \ , \\
\mathbb{E}[\phi_k(\epsilon_k)\epsilon_k^2] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \left(\mu_m^3 + 3\mu_m\sigma_m^2 - (\sigma_m^2 + \mu_m^2)\mu_m\right) = 0 \ , \\
\mathbb{E}[\phi_k(\epsilon_k)\epsilon_k^3] &= -\sum_{m=1}^{M} p_m \sigma_m^{-2} \left(\mu_m^4 + 6\mu_m^2\sigma_m^2 + 3\sigma_m^4 - \mu_m^4 - 3\mu_m^2\sigma_m^2\right) = -3 \ .
\end{aligned}$$

Example S2.2 (The normalised χ_2^2 distribution): *Suppose that $\tilde{\epsilon}_k \sim \chi_2^2$ and let $\epsilon_k = (\tilde{\epsilon}_k - 2)/2$. Then ϵ_k has mean zero, variance one and density function $\eta_k(z) = \exp(-z-1)$ on its support $[-1, \infty)$ on which we also have that $\phi_k(z) = -1$. The χ_2^2 distribution has finite moments of all orders and has moment generating function (e.g. Johnson et al., 1995, p. 420)*

$$M_{\tilde{\epsilon}}(t) = (1-2t)^{-1}, \quad t < 1/2.$$

Hence ϵ_k has finite moments of all orders. The same is evidently true of $\phi_k(\epsilon_k) = -1$. Using the above display, we have

$$M_{\epsilon}(t) = e^{-t}(1-t)^{-1}, \quad t < 1,$$

and therefore may directly calculate $\mathbb{E}[\epsilon_k^3] = 2$ and $\mathbb{E}[\epsilon_k^4] = 9$, hence $\mathbb{E}[\epsilon_k^3]^2 < \mathbb{E}[\epsilon_k^4] - 1$ holds. The moment conditions in part (a) are therefore all satisfied.

However, $\mathbb{E}\,\phi_k(z) = -1 \neq 0$, hence part (b) does not hold. Note also that this example does not satisfy the requirements of Lemma S2.20: we have $a_k = -1, b_k = \infty$ and

$$\lim_{z \downarrow a_k} \eta_k(x) = \lim_{z \downarrow -1} \exp(-z-1) = 1 \neq 0,$$

and hence the required condition is violated for $r = 0$.

S3 Technical tools

This section records some technical tools used in the proofs for ease of reference.

Lemma S3.1 (Discretisation): *Suppose that P_n is a sequence of probability measures and f_n :*

$\Gamma \to \mathbb{R}$, $\Gamma \subset \mathbb{R}^L$, *is a sequence of functions which satisfy*

$$f_n(\gamma_n) \xrightarrow{P_n} 0 \tag{S28}$$

for any $\gamma_n := \gamma + g_n/\sqrt{n}$, $g_n \to g \in \mathbb{R}^L$. *Suppose that the estimator sequence* $\bar{\gamma}_n$ *satisfies* $\sqrt{n}\|\bar{\gamma}_n - \gamma\| = O_{P_n}(1)$ *and* $\bar{\gamma}_n$ *takes values in* $\mathscr{S}_n := \{CZ/\sqrt{n} : Z \in \mathbb{R}^L\}$ *for some* $L \times L$ *matrix* C. *Then*

$$f_n(\bar{\gamma}_n) \xrightarrow{P_n} 0.$$

Proof. Since $\bar{\gamma}_n$ is $\sqrt{n}$-consistent there is an $M > 0$ such that $P_n\left(\sqrt{n}\|\bar{\gamma}_n - \gamma\| > M\right) < \varepsilon$. If $\sqrt{n}\|\bar{\gamma}_n - \gamma\| \le M$ then $\bar{\gamma}$ is equal to one of the values in the finite set $\mathscr{S}_n^c = \{\gamma^* \in \mathscr{S}_n : \|\gamma^* - \gamma\| \le n^{-1/2}M\}$. For each M this set has finite number of elements bounded independently of n, call this upper bound $\overline{B}$. For any $\upsilon > 0$

$$\begin{aligned}
P_n\left(|f_n(\bar{\gamma}_n)| > \upsilon\right) &\le \varepsilon + \sum_{\gamma_n \in \mathscr{S}_n^c} P_n\left(\{|f_n(\gamma_n)| > \upsilon\} \cap \{\bar{\gamma}_n = \gamma_n\}\right) \\
&\le \varepsilon + \sum_{\gamma_n \in \mathscr{S}_n^c} P_n\left(|f_n(\gamma_n)| > \upsilon\right) \\
&\le \varepsilon + \overline{B} P_n(|f_n(\gamma_n^\star)| > \upsilon),
\end{aligned}$$

where $\gamma_n^\star \in \mathscr{S}_n^c$ maximises $\gamma \mapsto P_n\left(|f_n(\gamma)| > \upsilon\right)$. As $\gamma_n^\star \in \mathscr{S}_n^c$, $\|\gamma^\star - \gamma\| \le n^{-1/2}M$. Hence letting $g_n := \sqrt{n}(\gamma_n^\star - \gamma)$, $\|g_n\| \le M$. Arguing along subsequences if necessary, we may therefore assume that $g_n \to g \in \mathbb{R}^L$ and hence $f_n(\gamma_n^\star) \xrightarrow{P_n} 0$ by (S28). The proof is complete on combining this with the previously established bound on $P_n\left(|f_n(\bar{\gamma}_n)| > \upsilon\right)$. $\square$

Lemma S3.2: *Let* $(X, \mathcal{B}(X))$ *be a measurable space, and* Q_n *a sequence of probability measures on* $(X, \mathcal{B}(X))$ *which converges to a probability measure* Q *in total variation. Let* $(Y, \mathcal{B}(Y), \lambda)$ *be a measure space and suppose that* $p_n : X \times Y \to [0, \infty)$ *is a sequence of functions and* $p : X \times Y \to [0, \infty)$ *a function such that (i)* $\int p_n(x, y)\,\mathrm{d}\lambda(y) = 1 = \int p(x, y)\,\mathrm{d}\lambda(y)$ *for each* $n \in \mathbb{N}$ *and each* $x \in X$ *and (ii)* $p_n \to p$ *pointwise. Then, if* G_n *and* G_n *are defined according to*

$$\begin{aligned}
G_n(A) &:= \int_A p_n(x, y)\,\mathrm{d}(\lambda(y) \otimes Q_n(x)); \\
G(A) &:= \int_A p(x, y)\,\mathrm{d}(\lambda(y) \otimes Q(x)),
\end{aligned}$$

it follows that $G_n \xrightarrow{TV} G$.

Proof. For any x, $p_n(x, \cdot) \to p(x, \cdot)$ pointwise and since each $p_n(\cdot, x)$, $p(\cdot, x)$ has integral one under λ, by Proposition 2.29 in van der Vaart (1998),

$$\mathscr{Q}_n(x) := \int |p_n(x, y) - p(x, y)| \, \mathrm{d}\lambda(y) \to 0,$$

pointwise. Let $(\psi_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions on $X \times Y$ with $\psi_n \in [0, 1]$. Then

$$\left| \int \int \psi_n(x, y)(p_n(x, y) - p(x, y)) \, \mathrm{d}\lambda(y) \, \mathrm{d}Q_n(x) \right| \leq \int \mathscr{Q}_n(x) \, \mathrm{d}Q_n(x).$$

Since $\mathscr{Q}_n(x) \leq \int p_n(x, y) \, \mathrm{d}\lambda(y) + \int p(x, y) \, \mathrm{d}\lambda(y) = 2$, the $\mathscr{Q}_n(x)$ are uniformly Q_n – integrable and uniformly Q – integrable. By Theorem 2.8 of Serfozo (1982), $\int \mathscr{Q}_n(x) \, \mathrm{d}Q_n(x) \to 0$. $\square$

Lemma S3.3: *Suppose that P_n and Q_n are probability measures (each pair (P_n, Q_n) is defined on a common measurable space) with corresponding densities p_n and q_n (with respect to some σ-finite measure ν_n). Let $l_n = \log q_n/p_n$ be the log-likelihood ratio.*[S16] *If*

$$l_n = o_{P_n}(1),$$

then $d_{TV}(P_n, Q_n) \to 0$.

Proof. By the continuous mapping theorem

$$\frac{q_n}{p_n} = \exp{(l_n)} \xrightarrow{P_n} 1.$$

Le Cam's first lemma (e.g. van der Vaart, 1998, Lemma 6.4) then implies that $Q_n \triangleleft P_n$. Let ϕ_n be arbitrary measurable functions valued in $[0, 1]$. Since the ϕ_n are uniformly tight, Prohorov's theorem ensures that for any arbitrary subsequence $(n_j)_{j\in\mathbb{N}}$ there exists a further subsequence $(n_m)_{m\in\mathbb{N}}$ such that $\phi_{n_m} \rightsquigarrow \phi \in [0, 1]$ under P_{n_m}. Therefore,

$$(\phi_{n_m}, \exp(l_{n_m})) \rightsquigarrow (\phi, 1) \quad \text{ under } P_{n_m}.$$

By Le Cam's third Lemma (e.g. van der Vaart, 1998, Theorem 6.6), under Q_{m_n} the law of ϕ_{n_m} converges weakly to the law of ϕ. Since each $\phi_n \in [0, 1]$

$$\lim_{m\to\infty} [Q_{n_m}\phi_{n_m} - P_{n_m}\phi_{n_m}] = 0.$$

As $(n_j)_{j\in\mathbb{N}}$ was arbitrary, the preceding display holds also along the original sequence. $\square$

[S16] l_n may be defined arbitrarily when $p_n = 0$.

PROPOSITION S3.1 (Cf. Proposition 2.29 in van der Vaart, 1998): *Suppose that on a measureable space* $(S, \mathcal{S})$, $(\mu_n)_{n\in\mathbb{N}}$ *is a sequence of measures and* μ *a measure such that* $\mu(A) \leq \liminf_{n\to\infty} \mu_n(A)$ *for each* $A \in \mathcal{S}$. *If* $(f_n)_{n\in\mathbb{N}}$ *and* f *are (real-valued) measurable functions such that* $f_n \to f$ *in* μ*-measure and* $\limsup_{n\to\infty} \int |f_n|^p \,\mathrm{d}\mu_n \leq \int |f|^p \,\mathrm{d}\mu < \infty$ *for some* $p \geq 1$, *then* $\int |f_n - f|^p \,\mathrm{d}\mu_n \to 0$.

Proof. $(a + b)^p \leq 2^p(a^p + b^p)$ for any $a, b \geq 0$ and hence, under our hypotheses,

$$0 \leq 2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p \to 2^{p+1}|f|^p \quad \text{in } \mu \text{ - measure.}$$

By Lemma 2.2 of Serfozo (1982) and $\limsup_{n\to\infty} \int |f_n|^p \,\mathrm{d}\mu_n \leq \int |f|^p \,\mathrm{d}\mu < \infty$,

$$\begin{aligned} \int 2^{p+1}|f|^p \,\mathrm{d}\mu &\leq \liminf_{n\to\infty} \int 2^p|f_n|^p + 2^p|f|^p - |f_n - f|^p \,\mathrm{d}\mu_n \\ &\leq 2^{p+1} \int |f|^p \,\mathrm{d}\mu - \limsup_{n\to\infty} \int |f_n - f|^p \,\mathrm{d}\mu_n. \end{aligned}$$ □

REMARK S3.1: *The condition that* $\mu(A) \leq \liminf_{n\to\infty} \mu_n(A)$ *for each* $A \in \mathcal{S}$ *in Propositions S3.1 is clearly satisfied if* $\mu_n \to \mu$ *setwise or in total variation.*

S4 Log density score estimation and optimal knot selection

In this section we provide more details for the estimation of the log density scores. Further, we discuss a data-driven way for selecting the number of knots following the approach of Chen and Bickel (2006). We evaluate the size and power of the test under optimal knot selection in some additional simulations that are presented below.

S4.1 B-spline based log density score estimation

For $\xi_1 < \cdots < \xi_N$ a knot sequence, the first order B-splines are defined according to $b_i^{(1)}(x) := \mathbf{1}_{[\xi_i, \xi_{i+1})}(x)$. Subsequent order B-splines can be computed according to the recurrence relation

$$b_i^{(l)}(x) = \frac{x - \xi_i}{\xi_{i+l-1} - \xi_i} b_i^{(l-1)}(x) + \frac{\xi_{i+l} - x}{\xi_{i+l} - \xi_{i+1}} b_{i+1}^{(l-1)}(x), \tag{S29}$$

for $l > 1$ and $i = 1, \ldots, N - l$. A l-th order B-spline is $l - 2$ times differentiable in x with first derivative

$$c_i^{(l)}(x) = \frac{l - 1}{\xi_{i+l-1} - \xi_i} b_i^{(l-1)}(x) - \frac{l - 1}{\xi_{i+l} - \xi_{i+1}} b_{i+1}^{(l-1)}(x). \tag{S30}$$

See de Boor (2001) for more details on B-splines.

Let $b_{k,n} = (b_{k,n,1}, \ldots, b_{k,n,\mathsf{B}_{k,n}})'$ be a collection of $B_{k,n}$ cubic (i.e. 4-th order) 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 $i \in \{1, \ldots, B_{k,n}\}$. The knots of the splines, $\xi_{k,n} = (\xi_{k,n,i})_{i=1}^{K_{k,n}}$ are equally spaced in $[\Xi^L_{k,n}, \Xi^U_{k,n}]$ with $\delta_{k,n} := \xi_{k,n,i+1} - \xi_{k,n,i} > 0$.[S17] For each (k,n) pair the relationships between the number of knots ($K_{k,n}$), the number of spline functions ($B_{k,n}$) and $\delta_{k,n}$ are given by $B_{k,n} = K_{k,n} - 4$ and $K_{k,n} = 1 + (\Xi^U_{k,n} - \Xi^L_{k,n})/\delta_{k,n}$.

Since the B-splines vanish at infinity for any $n \in \mathbb{N}$, integration by parts gives that

$$\begin{aligned}
&\int (\phi_k(z) - \psi'_{k,n} b_{k,n}(z))^2 \eta_k(z)\,\mathrm{d}z \\
&= \int \phi_k(z)^2 \eta_k(z)\,\mathrm{d}z + \int (\psi'_{k,n} b_{k,n})^2 \eta_k(z)\,\mathrm{d}z + 2\int \psi'_{k,n} c_{k,n}(z)\eta_k(z)\,\mathrm{d}z \\
&= \mathbb{E}\,\phi_k(\epsilon_k)^2 + \psi'_{k,n}\,\mathbb{E}[b_{k,n}(\epsilon_k) b_{k,n}(\epsilon_k)']\psi_{k,n} + 2\psi'_{k,n}\,\mathbb{E}\,c_{k,n}(\epsilon_k),
\end{aligned} \tag{S31}$$

where we integrate over the support of $\phi_{k,n}$ (which is also the support of $b_{k,n}$ and $c_{k,n}$). This mean-squared error is minimized by:[S18]

$$\psi_{k,n} := -\,\mathbb{E}[b_{k,n}(\epsilon_k) b_{k,n}(\epsilon_k)']^{-1}\,\mathbb{E}[c_{k,n}(\epsilon_k)]. \tag{S32}$$

Replace the population expectations with sample counterparts to define the estimator of $\psi_{k,n}$

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

Our estimate for the log density score ϕ_k is given by

$$\hat{\phi}_{k,n}(z) := \hat{\psi}'_{k,n} b_{k,n}(z)\ . \tag{S33}$$

As discussed in the main text, the knots of the splines, $\xi_{k,n} = (\xi_{k,n,i})_{i=1}^{K_{k,n}}$ are taken as equally spaced in $[\Xi^L_{k,n}, \Xi^U_{k,n}]$. 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 $\gamma = (\alpha, \beta)$. In our main simulations we used $B_{k,n} = 12$ splines.

S4.2 Data driven B-spline selection

The number of B-spline basis functions $B_{k,n}$ is a tuning parameter. In practice we can use cross-validation to choose $B_{k,n}$ for each k. A possible approach is as follows

[S17]For each $k = 1, \ldots, K$ the sequences $(\Xi^L_{k,n})_{n\in\mathbb{N}}$, $(\Xi^U_{k,n})_{n\in\mathbb{N}}$, $(B_{k,n})_{n\in\mathbb{N}}$ and $(\delta_{k,n})_{n\in\mathbb{N}}$ are deterministic.
[S18]This differs from the expression in Chen and Bickel (2006) by a factor of -1 as they estimate $-\phi_k$.

(i) Split the sample $A_{k\bullet}V_{\gamma,t}$ randomly into two halves, say n_1 and n_2.

(ii) For $B_{k,n} = 1, 2, \ldots$, use n_1 to estimate γ based on (S33), say $\hat{\phi}_{k,n_1}(z)$, and use n_2 to evaluate (S31) empirically, but omitting the first term $\mathbb{E}\,\phi_k(\epsilon_k)^2$, say $c_{n_2|n_1}(B_{k,n})$. Similarly calculate $c_{n_1|n_2}(B_{k,n})$.

(iii) Select the optimal $B_{k,n}$ as the largest value such that $\frac{1}{2}(c_{n_2|n_1}(B_{k,n})+c_{n_1|n_2}(B_{k,n}))$ strictly decreases until $B_{k,n}$.

This method is taken from Jin (1992) and Chen and Bickel (2006). Jin (1992) proved its validity under an iid assumption. In the additional simulations of Section S5 we experiment with this cross-validation algorithm.

S5 Additional simulation results

S5.1 Alternative parametrizations

We show that the parametrization of $A(\alpha, \sigma)$ does not affect the size of the score test nor the alternative tests considered. Specifically, we repeat Tables 2 and 3 from the main text, respectively, for an upper triangular parameterization of A. Tables S1 and S2 below show that rejection rates are not affected by the change in parameterization.

S5.2 Data driven B-spline selection

In this section we evaluate the performance of the score test when the number of B-splines is selected using cross-validation following the approach of Jin (1992), see the discussion in Section S4. All specifications are the same as in the main text and we use the one-step efficient estimates to estimate the nuisance parameters β. The results are shown in Table S3.

We find that with cross validation the test becomes closer to the nominal size. The empirical rejection frequencies for $n = 200, 500$ are nearly always close to the nominal level. Only when $n = 200$, $K = 3$ and $p = 12$ the test over-rejects. A possible route for improving this result is by adjusting the selection criteria from Jin (1992); Chen and Bickel (2006) — which is based on minimizing the mean squared error of the log density score estimate — to directly target the size of the test.

S5.3 Size for larger SVARs

In the main text we presented simulation results for SVAR models of dimensions $K = 2$ and $K = 3$. Here we explore higher dimensional SVAR models. In such settings two bottlenecks

Table S1: Empirical rejection frequencies: Triangular A

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	5.0	5.8	5.8	6.0	5.2	7.2	4.2	4.6	4.6	4.6
2	1	500	7.2	5.8	6.0	5.9	5.1	6.1	4.8	5.1	6.2	5.0
2	1	1000	6.9	6.2	6.1	6.2	4.8	5.3	4.8	4.8	5.2	5.3
2	4	200	4.7	4.8	5.4	5.8	5.7	5.8	4.9	5.6	4.8	3.8
2	4	500	7.0	5.6	6.8	6.1	4.6	5.2	4.0	4.7	4.6	4.6
2	4	1000	6.1	6.5	6.4	5.8	5.0	5.1	4.3	4.6	4.5	5.3
2	12	200	6.1	5.6	5.2	6.4	5.8	4.5	4.4	4.7	4.7	4.2
2	12	500	6.6	7.0	6.2	7.1	6.2	5.3	4.7	5.1	5.6	4.6
2	12	1000	7.0	6.0	5.8	6.4	5.4	5.5	4.6	5.8	6.2	4.9
3	1	200	5.6	6.8	7.0	8.0	7.2	9.9	5.1	5.5	6.1	4.6
3	1	500	7.6	6.8	7.0	7.1	5.9	6.6	4.2	5.1	6.0	4.9
3	1	1000	7.5	7.2	6.1	6.2	5.0	6.2	4.8	5.2	4.9	5.2
3	4	200	5.4	7.4	8.2	8.9	7.1	6.8	4.1	4.6	5.7	3.8
3	4	500	8.0	6.4	7.2	8.8	6.8	7.7	6.4	6.1	5.8	4.9
3	4	1000	7.9	6.6	8.0	6.7	5.8	6.2	6.0	5.8	5.3	6.3
3	12	200	3.1	3.9	3.0	4.2	2.5	3.6	3.0	2.0	2.8	2.6
3	12	500	8.5	9.4	8.8	10.2	9.6	6.2	3.8	4.1	6.0	2.3
3	12	1000	8.8	7.8	8.2	8.7	7.4	6.6	5.4	5.5	6.2	4.7
OLS Estimates												
2	1	200	3.6	4.2	4.2	6.5	4.8	7.2	3.2	2.8	5.6	3.2
2	1	500	4.4	4.1	4.5	6.3	4.7	7.5	3.7	4.1	5.6	4.3
2	1	1000	4.5	5.0	4.8	6.0	4.8	6.2	4.2	4.1	5.0	5.1
2	4	200	3.6	5.4	6.1	6.4	5.2	4.6	3.6	2.9	5.0	3.1
2	4	500	4.8	4.5	5.5	6.6	4.8	4.6	3.4	3.2	4.3	3.5
2	4	1000	4.1	5.2	5.2	5.2	5.3	4.8	3.1	3.2	4.8	4.7
2	12	200	7.2	7.4	8.4	9.4	6.4	4.8	4.5	3.0	6.6	4.4
2	12	500	5.6	7.3	6.1	8.3	6.9	3.7	4.3	3.3	4.9	3.8
2	12	1000	5.2	5.1	5.2	6.4	8.2	4.2	3.4	2.6	5.5	3.4
3	1	200	3.6	5.2	6.0	9.5	6.0	7.4	2.8	2.4	6.0	2.5
3	1	500	3.8	5.0	5.3	9.4	5.2	6.8	3.2	3.0	5.5	3.2
3	1	1000	3.6	4.9	4.5	7.6	5.9	7.1	3.6	3.6	4.4	3.9
3	4	200	7.1	8.8	8.4	12.2	7.0	5.0	2.8	0.8	4.9	2.2
3	4	500	5.0	5.4	6.7	10.9	7.0	4.9	2.2	1.4	5.6	1.2
3	4	1000	4.7	4.8	6.0	7.8	7.6	4.7	3.6	2.8	4.2	2.8
3	12	200	15.6	14.6	17.1	21.6	11.1	10.7	8.2	6.3	14.4	6.7
3	12	500	9.7	10.4	11.5	16.0	11.0	3.5	2.8	1.5	7.5	2.8
3	12	1000	6.4	6.8	8.2	10.5	10.4	3.3	3.2	1.6	5.4	3.3

Note: The table reports empirical rejection frequencies for the semi-parametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ 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 $M = 2,500$ Monte Carlo replications.

Table S2: Empirical rejection frequencies for alternative tests: Triangular A

Test	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$\epsilon_{1,t} \sim \epsilon_{2,t}$										
$\hat{S}_{ols}$	4.8	4.9	5.4	7.1	4.2	7.1	3.7	3.9	6.3	4.7
$\hat{S}_{onestep}$	8.0	6.7	6.9	6.9	4.7	6.3	4.9	5.5	7.1	5.3
S^{DW}	4.2	3.8	4.0	6.6	4.7	3.7	3.2	3.8	4.1	4.0
$\mathrm{BKR}^{\mathrm{DW}}$	4.4	4.1	4.4	4.2	5.8	28.4	5.1	5.7	6.5	5.2
$\mathrm{LM}^{\mathrm{PML,t}}$	5.1	4.9	4.9	6.5	9.1	8.3	66.7	81.4	5.9	85.0
$\mathrm{LM}^{\mathrm{GMM,LL}}$	1.7	1.6	3.6	11.1	6.9	7.2	6.4	6.1	1.9	5.1
$\mathrm{LM}^{\mathrm{GMM,Kew}}$	1.4	2.1	3.8	17.2	9.9	8.0	6.1	5.9	1.2	5.2
$\mathrm{LR}^{\mathrm{PML,t}}$	25.9	11.3	6.5	4.8	4.9	1.6	100.0	100.0	11.5	100.0
$\mathrm{LR}^{\mathrm{GMM,LL}}$	3.5	7.4	8.9	15.9	12.4	9.8	5.9	5.8	7.5	4.6
$\mathrm{LR}^{\mathrm{GMM,Kew}}$	6.3	7.7	12.3	22.2	16.3	12.9	6.4	6.2	6.8	4.9
$\mathrm{W}^{\mathrm{PML,t}}$	4.5	7.4	9.2	10.4	11.2	8.0	66.2	69.9	8.3	69.6
$\mathrm{W}^{\mathrm{GMM,LL}}$	12.0	17.9	18.8	22.5	17.5	14.6	6.6	6.8	15.3	5.6
$\mathrm{W}^{\mathrm{GMM,Kew}}$	19.3	21.8	25.1	23.0	16.5	15.8	7.1	7.0	19.1	5.8

Note: The table reports empirical rejection frequencies for tests of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$, and $\alpha_0 = 0.5594$. $\hat{S}_{ols}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. $\mathrm{LM}^{\mathrm{PML,t}}$, $\mathrm{W}^{\mathrm{PML,t}}$ and $\mathrm{LR}^{\mathrm{PML,t}}$ denote the pseudo-maximum likelihood tests based on Gouriéroux et al. (2017), assuming t-distributed shocks. $\mathrm{LM}^{\mathrm{GMM,LL}}$, $\mathrm{W}^{\mathrm{GMM,LL}}$ and $\mathrm{LR}^{\mathrm{GMM,LL}}$ denote the GMM-based tests based on Lanne and Luoto (2021) with one co-kurtosis condition based on $\epsilon_{1t}^3\epsilon_{2t}$. $\mathrm{LM}^{\mathrm{GMM,Kew}}$, $\mathrm{W}^{\mathrm{GMM,Kew}}$ and $\mathrm{LR}^{\mathrm{GMM,Kew}}$ 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 non-parametric 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 $M = 1,000$ Monte Carlo replications.

Table S3: Empirical rejection frequencies: optimal knot selection

K	p	n	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
2	1	200	5.2	4.6	5.8	6.1	4.4	5.5	5.9	6.6	5.0	7.3
2	1	500	4.7	4.1	5.0	5.7	4.7	5.6	5.5	6.4	4.9	7.5
2	1	1000	4.6	4.9	4.4	3.7	4.6	5.7	5.3	6.7	4.6	8.3
2	4	200	5.0	5.9	5.7	5.6	3.7	5.1	4.6	4.6	4.6	5.3
2	4	500	4.8	4.8	4.8	5.6	4.4	5.7	5.8	5.8	4.2	6.0
2	4	1000	4.2	5.0	5.0	4.5	4.8	5.2	5.1	6.1	4.8	6.4
2	12	200	7.0	6.5	7.4	7.8	5.2	4.8	4.6	4.0	6.8	3.5
2	12	500	5.7	6.6	6.4	6.2	4.6	4.9	4.1	4.8	5.3	4.5
2	12	1000	5.4	5.2	5.5	5.0	5.6	4.9	4.3	4.7	5.7	5.4
3	1	200	5.3	6.5	7.1	9.8	7.6	6.9	4.8	5.2	4.9	5.6
3	1	500	5.0	5.3	5.9	7.3	5.0	6.1	5.8	7.6	5.2	6.8
3	1	1000	5.0	5.8	5.3	5.9	4.7	5.8	6.3	9.1	4.9	9.0
3	4	200	6.1	8.4	9.2	11.0	6.2	6.0	2.9	2.3	5.8	3.1
3	4	500	5.7	5.7	6.8	8.6	5.6	5.0	4.6	4.7	4.6	3.9
3	4	1000	5.4	5.2	5.7	5.7	5.2	5.2	5.0	6.1	4.4	6.1
3	12	200	13.0	14.0	14.8	15.6	12.7	8.3	7.0	5.5	12.8	5.8
3	12	500	9.4	10.3	10.2	12.4	8.3	4.5	3.3	2.5	6.8	3.2
3	12	1000	6.8	7.3	7.7	8.2	7.3	5.1	4.2	4.3	5.6	4.4

Note: The table reports empirical rejection frequencies for the semi-parametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ in the K-variable SVAR(p) model with nominal size 5%. The nuisance parameter estimates $\hat{\beta}$ are OLS estimates. For each density score the number of B-splines is determined by cross-validation. 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 $M = 2,500$ Monte Carlo replications.

can arise. First, the computational costs for constructing the confidence sets in **Algorithm 1** increase substantially as one must evaluate the test at each point of the grid. Even in the most parsimonious specification for $K = 5$ such a grid is 10 dimensional. We note that this bottleneck is not specific to our approach but arises for most weak identification robust tests when constructing confidence sets.

Second the number of finite dimensional nuisance parameters β increases rapidly when the dimension of the SVAR model increases. For instance for $K = 5$ and $p = 12$ the number of nuisance parameters L_β is around 300. This has several consequences. First, when n is smaller than the number of nuisance parameters the test does not exist anymore as the inverse of $\hat{I}_{n,\gamma,\beta\beta}$ is not defined. Second, even when the number of nuisance parameters is proportional (but smaller) than the sample size the asymptotic theory of our paper may not provide a good approximation to the finite sample performance. The reason is that our theory is developed for L_β fixed (hence $L_\beta/n \to 0$). Extending the theory to the case where L_β may increase with n is an interesting topic for future work.

That said, it is of interest to explore the finite sample performance of the test in these settings. Table S4 reports the empirical rejection frequencies for the score test for larger SVARs with $K = 5$. All other settings for the simulation design are similar as above. We exclude $n = 200$ as the test is not defined for all specifications for this sample size. We find that the test generally behaves poorly and we recommend keeping the dimensions and lag length modest when evaluating the test based on asymptotic critical values; similar to other SVAR studies, a bootstrap implementation of our test is likely to be preferable for higher dimensional SVAR models.

S5.4 Coverage and length of confidence sets

In this section, we consider evaluating our methodology for constructing confidence sets for smooth functions of the SVAR parameters as discussed in Section 5. We focus on evaluating the coverage and length of the confidence sets for structural impulse response functions, see Example 5.1 for the details.

We consider a similar simulation set up as above and discuss the results for the SVAR(1) model with $K = 2$, $T = 500$, and two independent shocks drawn from the same distribution, as listed in Table 1. In each case, the confidence set is calculated using **Algorithm 2** for the structural impulse response of the first variable to the second shock and we report the coverage rate and length for horizons 0-12. Further, we compare our approach to the identification robust methods of Drautzburg and Wright (2023), for which we change step (i) in **Algorithm 2** and

Table S4: Empirical rejection frequencies for larger SVARs

K	p	n	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
One-Step Efficient Estimates												
5	1	500	14.1	14.6	14.0	13.6	7.8	8.7	6.2	5.5	7.0	4.4
5	1	1000	10.9	10.9	11.2	10.7	6.0	7.4	6.3	5.3	5.5	4.5
5	4	500	14.9	18.2	17.2	17.9	9.8	8.4	5.0	4.0	7.0	2.3
5	4	1000	14.3	14.9	14.9	12.8	6.8	8.9	6.4	5.6	7.6	5.1
5	12	500	0.4	0.3	0.4	0.5	0.6	0.4	0.0	0.0	0.4	0.0
5	12	1000	17.6	18.8	18.6	15.1	9.1	7.1	2.8	1.9	5.6	1.1
OLS Estimates												
5	1	500	9.4	11.2	11.7	16.1	6.6	4.5	1.9	1.0	4.0	1.2
5	1	1000	6.8	8.3	9.1	12.4	6.1	4.8	3.2	2.0	4.2	1.7
5	4	500	16.4	20.7	21.1	25.2	7.4	1.6	0.5	0.0	4.2	0.0
5	4	1000	11.8	13.5	14.7	13.6	7.2	2.4	1.1	0.3	3.2	0.6
5	12	500	56.4	60.5	59.9	55.0	16.4	8.6	2.2	0.2	26.7	0.6
5	12	1000	28.5	27.8	30.7	27.6	10.8	2.3	0.9	0.1	5.5	0.4

Note: The table reports empirical rejection frequencies for the semi-parametric score test of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ in the K-variable SVAR(p) model with nominal size 5%. The nuisance parameters β are estimated using either one-step efficient estimates or OLS. For each density score the number of B-splines is fixed at $B = 6$. The columns correspond to different choices for the distributions of the structural shocks, $\epsilon_{k,t}$ for $k = 1, \dots, K$. The distributions are reported in Table 1. Rejection rates are computed based on $M = 2,500$ Monte Carlo replications.

replace the efficient score test by the tests of Drautzburg and Wright (2023).

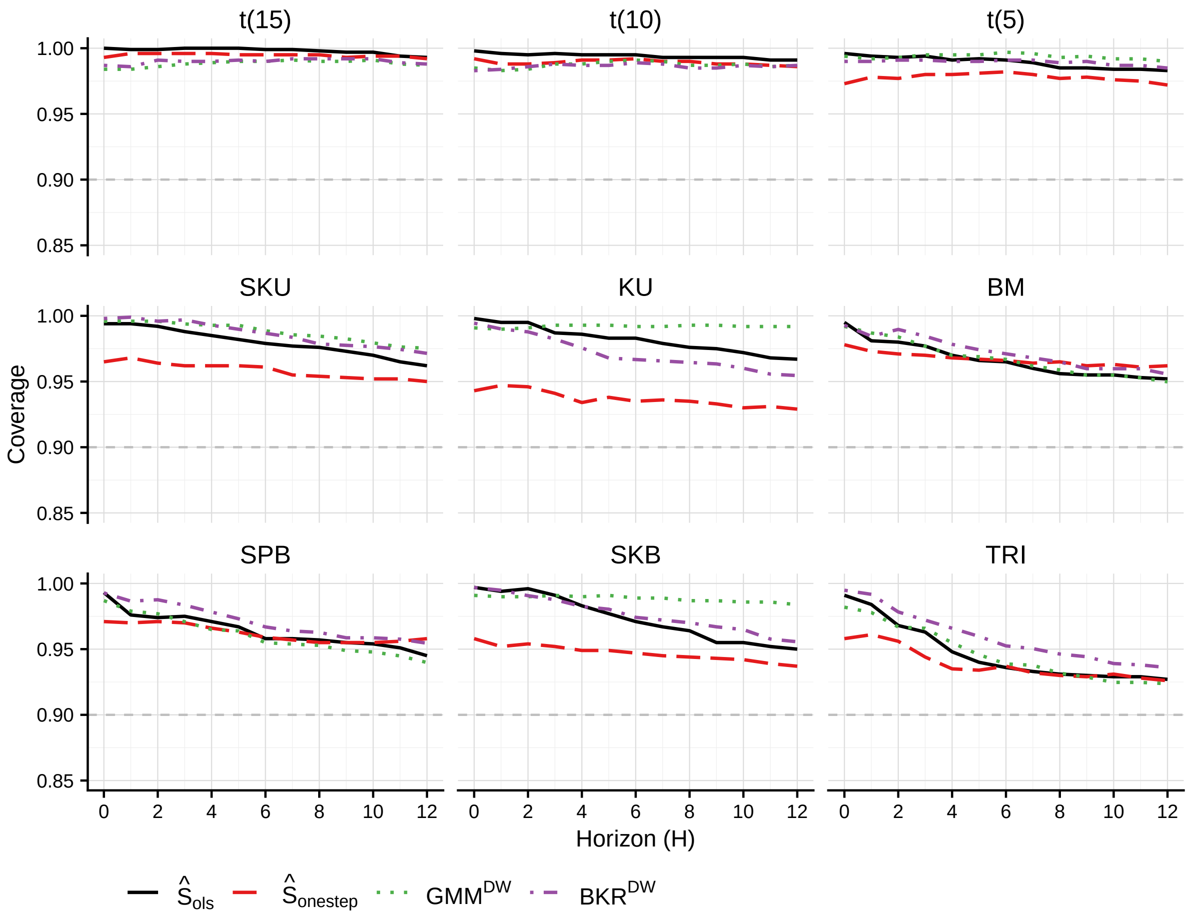
Figure S1 shows the empirical coverage rates. Not surprising we generally find that the two-step Bonferroni approach is conservative; all empirical coverage rates are above the nominal 90% level. This holds for all horizons, densities and methods considered.

That said, we find that if the efficient score test, based on one-step efficient estimates, is used as the first step in the Bonferroni method the coverage becomes much closer to the nominal size. This holds for nearly all densities, the exception being the t densities that are very close to Gaussian, where there is generally very low power.

Figure S2 shows the length of the confidence intervals. We find that efficient score approach gives the smallest length among all procedures considered and for all densities. The differences between the methods varies; for some densities all methods give comparable intervals, but for others the efficient score approach can give intervals that are up to 30% shorter in length. This holds especially at longer horizons.

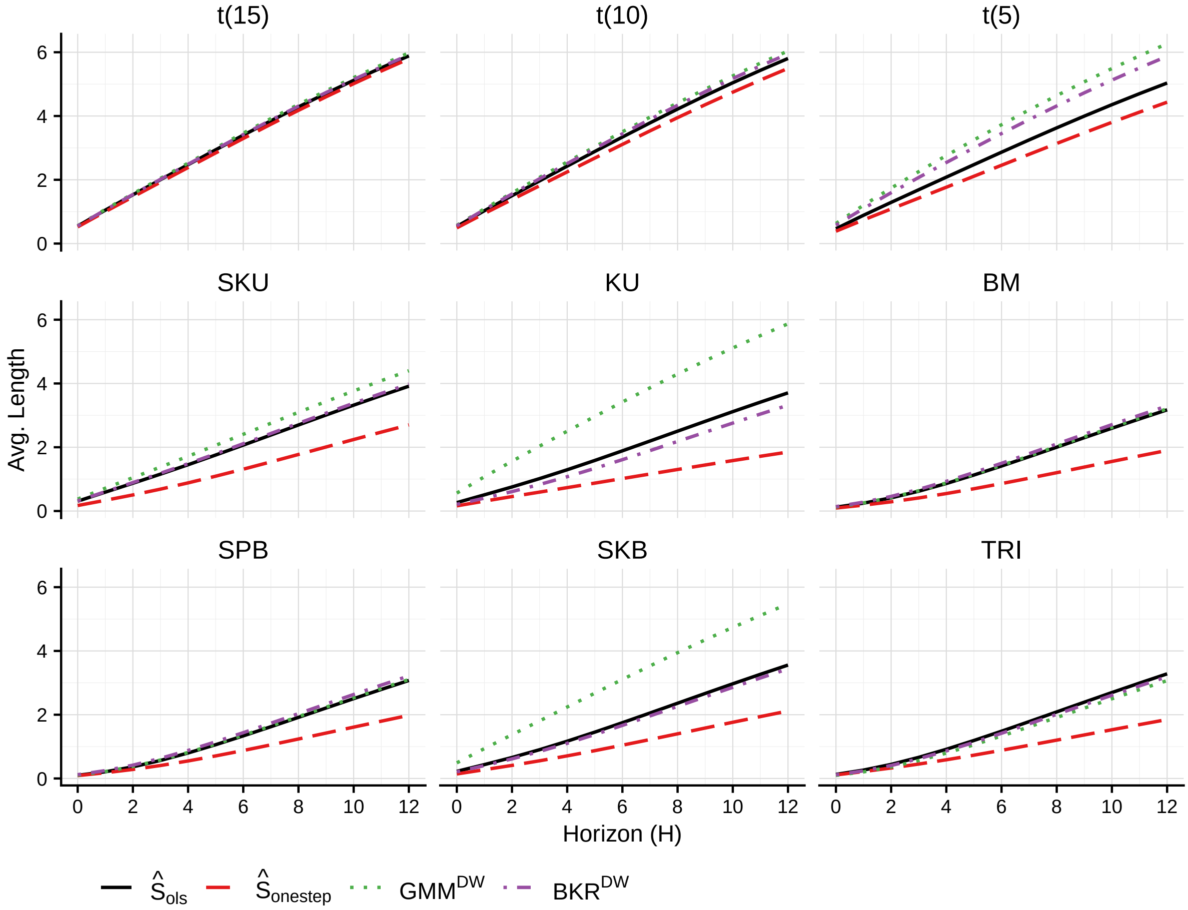
We conclude that the two-step Bonferroni method, where the first step is based on the efficient score test, gives substantial efficiency improvements when compared to existing methods.

Figure S1: Coverage rates of $\hat{C}_{n,g,\alpha,0.9}$



Note: The figure reports empirical coverage rates of confidence intervals at individual horizons for the impulse response of the first variable to the second shock with 90% nominal coverage for the SVAR(1) model with $K = 2$ and $T = 500$. $\hat{S}_{ols}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. GMM^{DW} denotes the GMM-based test of Drautzburg and Wright (2023) and BKR^{DW} denotes the non-parametric test of Drautzburg and Wright (2023). The tests of Drautzburg and Wright (2023) use 500 bootstrap replications to obtain critical values. Coverage is computed using $M = 1,000$ Monte Carlo replications.

Figure S2: Average length of $\hat{C}_{n,g,\alpha,0.9}$



Note: The figure reports average length of confidence intervals at individual horizons for the impulse response of the first variable to the second shock with 90% nominal coverage for the SVAR(1) model with $K = 2$ and $T = 500$. $\hat{S}_{ols}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. GMM^{DW} denotes the GMM-based test of Drautzburg and Wright (2023) and BKR^{DW} denotes the non-parametric test of Drautzburg and Wright (2023). The tests of Drautzburg and Wright (2023) use 500 bootstrap replications to obtain critical values. Average length is computed using $M = 1,000$ Monte Carlo replications.

S5.5 Point estimation results

Table S5 shows the Root Mean Squared Errors (RMSEs) for parameter estimates $\hat{\alpha}$ in the K-variable SVAR(1) model with $K = 2, T = 500$. We compare the performance of different estimators and their one-step efficient counterparts as discussed in Section 6. Specifically, we consider the psuedo maximum likelihood estimator of Gouriéroux et al. (2017) and the moment estimators of Lanne and Luoto (2019) and Keweloh (2021) as initial estimators. For each of these we compute the corresponding one-step efficient estimate from (29).

The results show that if the true density is Gaussian or close to Gaussian there is no advantage in doing a one-step efficient update. Intuitively, in these settings the efficient scores are noisy and add little additional information to the initial estimate, implying that the mean squared errors do not improve. In contrast, when the underlying density is away from the Gaussian (as imposed asymptotically by Assumption 6.1) the one-step efficient estimates always have lower RMSEs. The gains can be large, and appear to outweigh the small relative losses that are sometimes incurred for densities close to Gaussian.

Table S5: Efficiency of one-step updated estimates $\hat{\alpha}$

	PML^t		GMM^{LL}		GMM^{Kew}	
η	$\hat{\alpha}_n$	$\hat{\alpha}_n^{onestep}$	$\hat{\alpha}_n$	$\hat{\alpha}_n^{onestep}$	$\hat{\alpha}_n$	$\hat{\alpha}_n^{onestep}$
N(0,1)	0.207	0.235	0.188	0.194	0.194	0.194
t(15)	0.137	0.146	0.156	0.147	0.154	0.148
t(10)	0.103	0.108	0.129	0.113	0.120	0.114
t(5)	0.051	0.056	0.082	0.061	0.070	0.061
SKU	0.042	0.032	0.071	0.037	0.058	0.035
KU	0.041	0.026	0.082	0.041	0.068	0.038
BM	0.250	0.070	0.030	0.016	0.016	0.015
SPB	0.250	0.090	0.027	0.012	0.013	0.012
SKB	0.138	0.067	0.163	0.074	0.160	0.074
TRI	0.250	0.113	0.025	0.012	0.012	0.012

Note: The table reports Root Mean Squared Errors (RMSEs) for parameter estimates $\hat{\alpha}$ in the K-variable SVAR(1) model with $K = 2, T = 500$. The rows 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. RMSEs are computed based on $M = 2,500$ Monte Carlo replications.

S6 Additional empirical results

S6.1 Alternative parametrization of Baumeister and Hamilton (2015) model

This section presents results from an alternative parametrization of the Baumeister and Hamilton (2015) model where A^{-1} is expressed as follows:

$$A^{-1}(\xi,\sigma) = \begin{pmatrix} A^{-1}_{11} & A^{-1}_{12} \\ A^{-1}_{21} & A^{-1}_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2 & \sigma_3 \end{pmatrix} \begin{pmatrix} \cos(\xi) & -\sin(\xi) \\ \sin(\xi) & \cos(\xi) \end{pmatrix}, \qquad \xi \in [0, 2\pi) \tag{S34}$$

The sign restrictions of the original model, which are formulated on demand and supply elasticities as well as scaling parameters (see Example S1.1), imply that $A^{-1}_{11} \geq 0, A^{-1}_{12} \leq 0, A^{-1}_{21} \geq 0$ and $A^{-1}_{22} \geq 0$. The parametrization in (S34) translates the sign restrictions into constraints imposed on (ξ, σ). For the alternative parameterization, we assume that $\sigma_1, \sigma_2, \sigma_3 > 0$, which corresponds to imposing the identification restriction $\xi \in (0, \pi/2)$.[S19] Hence, for **Algorithm 1**, we set up a grid of 500 grid-points in $\xi \in (0, \pi/2)$. Note that there is a direct mapping between the two parametrizations that given (ξ, σ) lets us recover the elasticities (α^d, α^s) from the main parametrization discussed in the paper. Specifically, we can define $g(\xi, \sigma)$ as the following (smooth) vector-valued function which recovers the structural elasticities (α^d, α^s) from the rotation angle α.

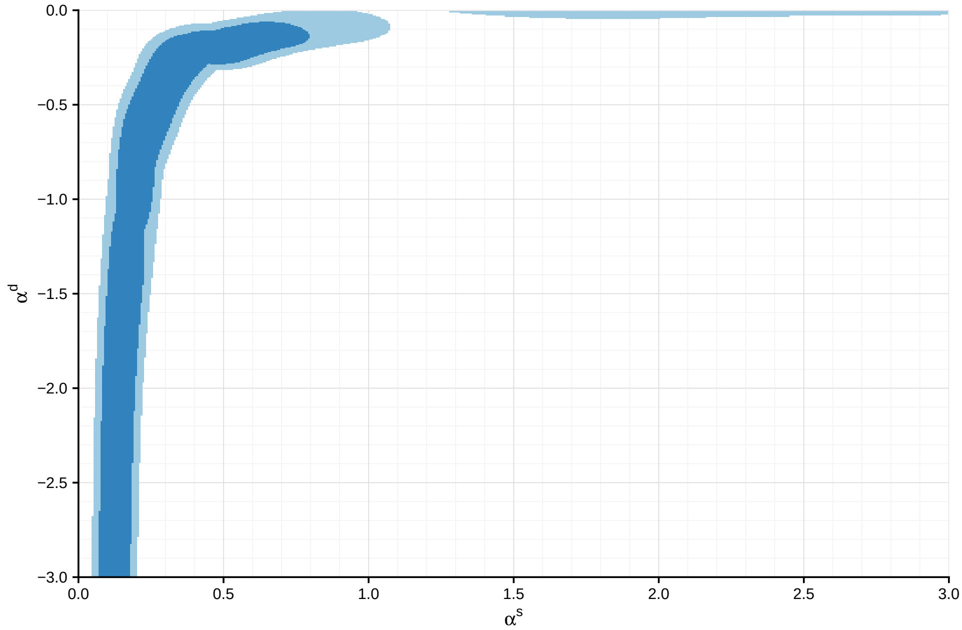
$$g(\xi,\sigma) = \begin{pmatrix} \alpha^d, & \alpha^s \end{pmatrix}' := \begin{pmatrix} \frac{\sigma_2 \cdot \sin(\xi) - \sigma_3 \cdot \cos(\xi)}{\sigma_1 \cdot \sin(\xi)}, & \frac{\sigma_2 \cdot \cos(\xi) + \sigma_3 \cdot \sin(\xi)}{\sigma_1 \cdot \cos(\xi)} \end{pmatrix}' \tag{S35}$$

Since $g(\xi, \sigma)$ is a smooth function, we can use **Algorithm 2** to recover confidence sets for the structural elasticities (α^d, α^s). To this purpose, we define a grid with 250,000 equally-spaced grid points for $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$, similar to the grid used in the main parametrisation. Similarly, we can use **Algorithm 2** to directly recover confidence bands for the impulse response functions.

Figures S3 and S4 report the confidence sets for labor demand and labor supply elasticities, as well as confidence bands for the impulse response functions, respectively, obtained using the alternative parametrization. Overall, the results are very close to the ones reported for the main parametrization. Due to the Bonferroni procedure of **Algorithm 2**, the confidence set for the elasticites is slightly wider than the one reported in the main text of the paper based on the alternative parametrization. For the IRF bands, there are also slight differences in the widths of the impulse response bands.

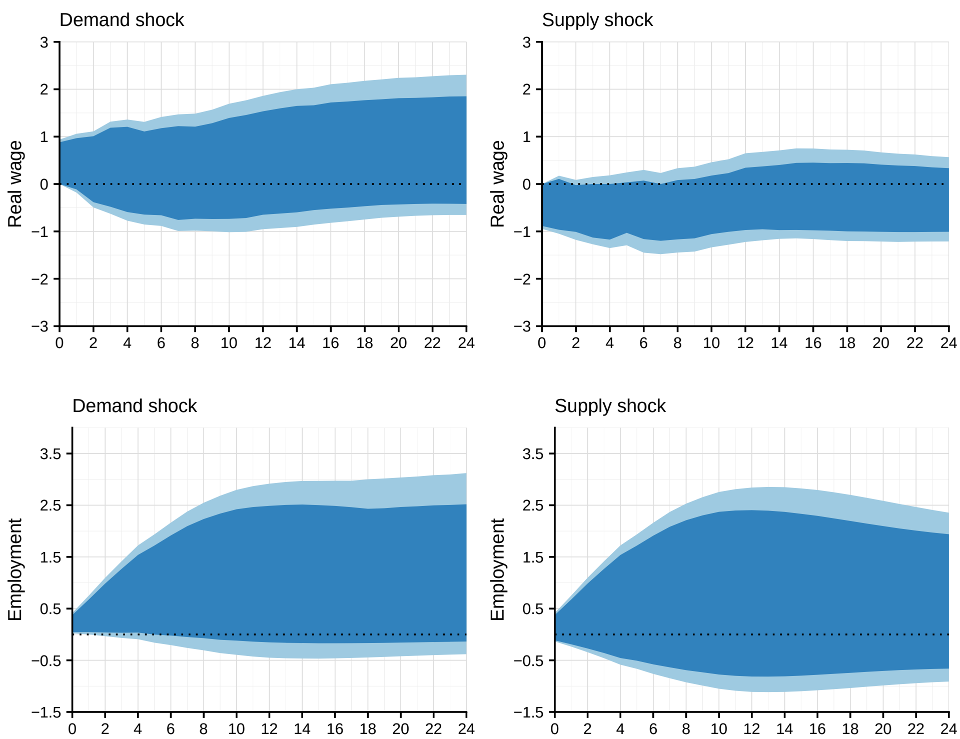
[S19] $\sigma_1, \sigma_3 > 0$ is trivial, since these capture standard deviations of the reduced form SVAR residuals. $\sigma_2 > 0$ can be verified from a Cholesky decomposition of the estimated reduced-form errors of the SVAR.

Figure S3: Confidence Sets for Labor Demand and Supply Elasticities



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

Figure S4: IRF confidence bands for labor demand and supply shocks

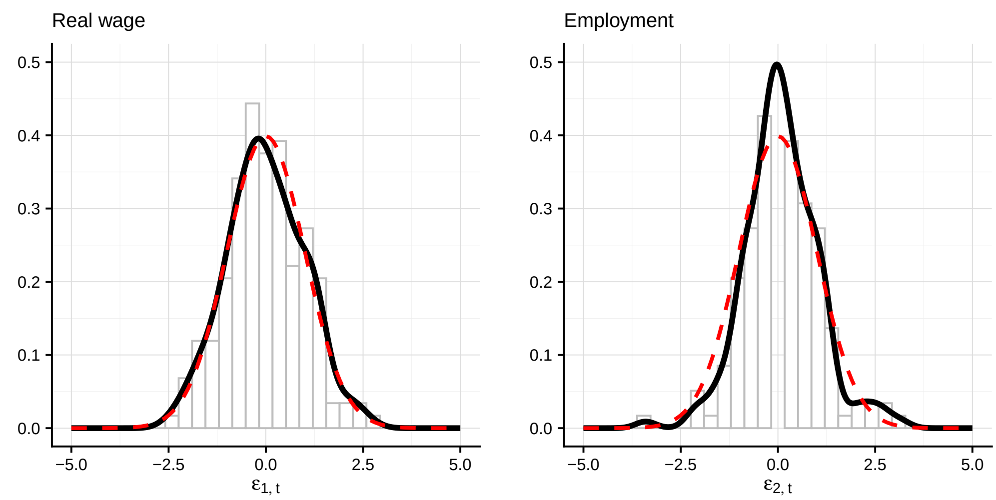


Note: 95% (light blue) and 67% (dark blue) confidence bands for impulse responses to labor supply and labor demand shocks, obtained using using **Algorithm 2** with 500 equally-spaced grid points for $\xi \in [0, \pi/2]$.

S6.2 Distributions of recovered structural shocks

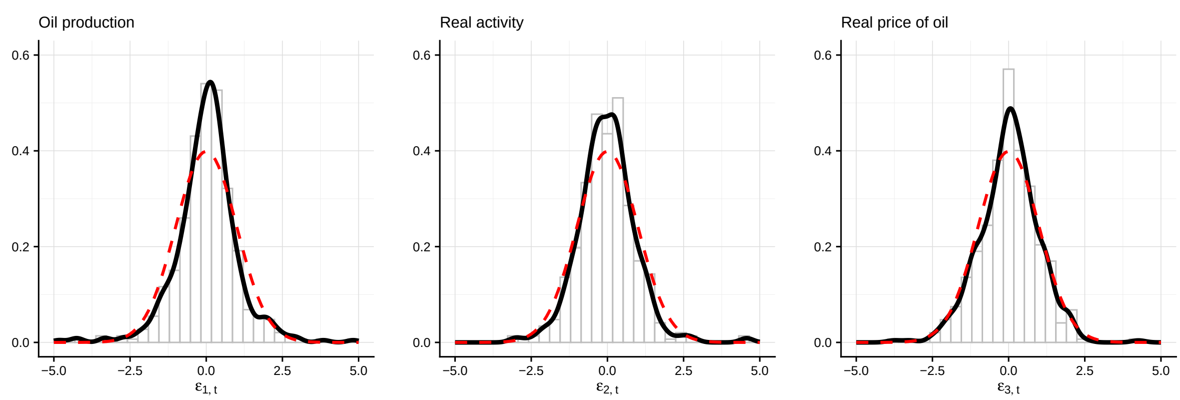
In this section, we present kernel density estimates for the structural errors recovered from the empirical studies in the paper. To obtain estimates of the structural shocks, we require an estimate of α, which we obtained using a GMM estimator employing the moment conditions of Keweloh (2021). Using the estimate, we can recover the structural shocks $\hat{\epsilon}_{k,t}(\hat{\alpha}, \hat{\beta})$ for $k = 1, \ldots, K$. We plot histograms of the structural errors in Figure S6.2 for the model of Baumeister and Hamilton (2015) and on Figure S6.2 for the model of Kilian and Murphy (2012), together with their kernel density estimates and an overlaid standard gaussian density.

FIGURE S5: DISTRIBUTIONS OF SHOCKS – BAUMEISTER AND HAMILTON (2015) MODEL



Note: Histogram (light gray) and kernel density estimates (black solid) of the recovered structural shocks $\hat{\epsilon}_{k,t}$ for $k = 1, 2$ from the Baumeister and Hamilton (2015) model, overlaid with a standard normal density (red dashed).

FIGURE S6: DISTRIBUTIONS OF SHOCKS – KILIAN AND MURPHY (2012) MODEL



Note: Histogram (light gray) and kernel density estimates (black solid) of the recovered structural shocks $\hat{\epsilon}_{k,t}$ for $k = 1, 2, 3$ from the Kilian and Murphy (2012) model, overlaid with a standard normal density (red dashed).

References

Baumeister, C. and Hamilton, J. D. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999.

Billingsley, P. (1995). *Probability and Measure*. Wiley.

Braun, R. (2021). The importance of supply and demand for oil prices: evidence from non-gaussianity. *Working paper*.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, 2 edition.

Chen, A. and Bickel, P. J. (2006). Efficient Independent Component Analysis. *Annals of Statistics*, 34(6).

Chow, Y. S. and Teicher, H. (1997). *Probability Theory*. Springer Texts in Statstics. Springer, 3 edition.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, 36.

Davidson, J. (1994). *Stochastic limit theory*. Oxford University Press.

de Boor, C. (2001). *A Practical Guide to Splines*. Springer, New York, NY, USA.

Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer.

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

Durrett, R. (2019). *Probability Theory and Examples*. Cambridge University Press, Cambridge, UK, 5th edition.

Feinberg, E. A., Kasyanov, P. O., and Zgurovsky, M. Z. (2016). Uniform fatou's lemma. *Journal of Mathematical Analysis and Applications*, 444(1):550–567.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics*, 196.

Gut, A. (2005). *Probability: A Graduate Course*. Springer Texts in Statistics. Springer.

Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press, 2 edition.

Jin, K. (1992). Empirical smoothing parameter selection in adaptive estimation. *The Annals of Statistics*, pages 1844–1874.

Johnson, N., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions, Volume 2*. Wiley Series in Probability and Statistics. Wiley.

Kallenberg, O. (2021). *Foundations of Modern Probability*. Probability Theory and Stochastic Modelling. Springer.

Kanaya, S. (2017). Convergence rates of sums of α-mixing triangular arrays: With an application to nonparametric drift function estimation of continuous-time processes. *Econometric Theory*, 33(5):11211153.

Keweloh, S. A. (2021). A generalized method of moments estimator for structural vector autoregressions based on higher moments. *Journal of Business & Economic Statistics*, 39(3):772–782.

Kilian, L. and Murphy, D. P. (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.

Lanne, M. and Luoto, J. (2019). Useful prior information in sign-identified structural vector autoregression: Replication of baumeister and hamilton (2015). *Working paper*.

Lanne, M. and Luoto, J. (2021). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, 39(1):69–81.

Magnus, J. R., Pijls, H. G., and Sentana, E. (2021). The jacobian of the exponential function. *Journal of Economic Dynamics and Control*, 127:104122.

Rudin, W. (1991). *Functional analysis*. McGraw Hill, 2 edition.

Saikkonen, P. (2007). Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity. *Statistica Sinica*, 17(1):221–239.

Serfozo, R. (1982). Convergence of lebesgue integrals with varying measures. *Sankhy: The Indian Journal of Statistics, Series A (1961-2002)*, 44(3):380–402.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis*, 16(1):54–70.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York, NY, USA, 1st edition.