TI 2022-080/III
Tinbergen Institute Discussion Paper

Robust Inference for Non-Gaussian SVAR models

Lukas Hoesch[1]
Adam Lee[2]
Geert Mesters[3]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

2 BI Norwegian Business School

3 Universitat Pompeu Fabra, Barcelona School of Economics and CREI

Robust Inference for Non-Gaussian SVAR models

Lukas Hoesch* Adam Lee† Geert Mesters‡

October 24, 2022

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 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 regardless of the distance to the Gaussian distribution. 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 Regis Barnichon, Christian Brownlees, Kirill Evdokimov, Katerina Petrova, Barbara Rossi and Piotr Zwiernik, as well as seminar and conference participants from different institutions for helpful comments. Mesters acknowledges financial support from the European Research Council via Starting Grant 101041145 – POLICYMETRICS, and from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S).

1 Introduction

In this paper we develop 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 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; Brunnermeier et al., 2021; Drautzburg and Wright, 2021; Davis and Ng, 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 refer to this phenomenon as "weak non-Gaussianity".

[1] See Montiel Olea et al. (2022) for a recent review of this approach and 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 (2022); Lee and Mesters (2022a) for more discussion of the same issue in static ICA models.

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

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 two main results. First, we adopt a semi-parametric generalisation of Neyman's C(α) 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 the null hypothesis regardless of how close the shock distributions are to the Gaussian distribution (i.e. the point of identification failure). The choice to base the score test on the efficient score function is also beneficial in terms of power as such efficient score tests have been shown to possess power optimality properties in various settings (see e.g. Choi et al., 1996; van der Vaart, 1998; Lee, 2022).

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, 2021) which is guaranteed to provide correct coverage asymptotically, regardless of the degree of non-Gaussianity in the shock distribution.

Overall, our methods are computationally simple and efficient. The implementation of the semiparametric score test typically only requires estimating regression coefficients, a covariance matrix and the log density scores of the structural shocks, and does not require, for instance, bootstrap methods to obtain critical values. 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 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 analyse 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 Bonferroni approach for constructing confidence sets is (by construction) conservative, it does

[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.

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 labour 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 labour 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 (2019) 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 criticised 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 literature that exploits non-Gaussianity of the structural shocks for identification (see the references above). There are two papers that are specifically related to the current paper.

First, Drautzburg and Wright (2021) are similarly 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.

Second, this paper builds on Lee and Mesters (2022a) and Lee (2022) who consider a similar score testing approach in static ICA models. The crucial differences are that (a) those papers require that the observations are independent and identically distributed across time and (b) they focus on testing a hypothesis for a potentially weakly- or un-identified parameter and do not consider functions of identified and possibly unidentified parameters. Relaxing the independence assumption is non-trivial in this context; we show a new (uniform) local asymptotic normality

[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 Matteson and Tsay (2017); see also Montiel Olea et al. (2022).

result for semi-parametric SVARs. Further, in the SVAR setting the objects of economic interest, such as IRFs, are typically functions of both well-identified and possibly unidentified parameters. This paper provides a robust inference procedure for such objects.

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 Cheng, 2012). 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 the score testing approach developed for weakly identified parametric models in Andrews and Mikusheva (2016). 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.

The remainder of this paper is organised as follows. In Section 2, we briefly illustrate how non-Gaussian distributions can help with identification and how the weak identification problem arises. Section 3 casts the SVAR model as a semi-parametric model and Section 4 establishes a number of preliminary results. The semi-parametric score testing approach is presented in Section 5 and inference for smooth structural functions is covered in Section 6. Section 7 evaluates the finite-sample performance of the proposed methodology and Section 8 discusses the results from the empirical studies. Section 9 concludes.

2 An illustration of non-Gaussian identification

In this section, we briefly illustrate how non-Gaussian structural shocks can help to identify the parameters of the SVAR model. Furthermore, we provide an intuitive explanation for the weak identification problem that arises when the error distributions are close to Gaussian.

As an example, consider a bivariate SVAR model as defined in equation (1), but assume for simplicity that (i) the number of lags is zero ($p = 0$) and (ii) that the contemporaneous effect matrix A is a rotation matrix.[7] Under these assumptions, the matrix A can be parametrised by a scalar parameter α and the model can be written as

$$Y_t = A^{-1}\epsilon_t \ , \qquad \text{where} \qquad A^{-1} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} .$$

The parameter of interest is the scalar α that determines the rotation matrix A. If for example $\alpha = 0$, A equals the identity matrix and each of the structural shocks only impacts its respective component in Y_t. For $0 < \alpha < \pi$, or integer multiples thereof, the off-diagonal elements are non-zero so that the shocks affect all variables, with signs depending on the value of α.

[7] Note, that the assumption for A being a rotation matrix can be asymptotically justified if the data Y_t is jointly re-scaled to have mean zero and identity variance matrix (pre-whitening). For details, see Gouriéroux et al. (2017).

To illustrate how non-Gaussian distributions for ϵ_t may help to identify α, we study the expected log-likelihood $\mathbb{E}\ell_\alpha(Y_t)$ in the model above for different distributions of the structural shocks $\epsilon_{k,t}$. For instance, if $\epsilon_{k,t} \sim \mathcal{N}(0,1)$ for all k we have

$$\mathbb{E}\ell_\alpha(Y_t) \propto -\frac{1}{2}\,\mathbb{E}\left[(A^{-1}\epsilon_t)'A^{-1}\epsilon_t\right] = -1 \ ,$$

and the log likelihood takes the same value for all α. This reflects the standard identification problem: without additional identifying restrictions, the impact effects of the structural shocks are not identifiable when the shocks are Gaussian.

Figure 1 visually illustrates this result and shows how it changes when we move away from the Gaussian distribution in the direction of the student-t distribution. The left panels shows the expected log likelihood as a function of α, whereas the right panels show the contour plots of the log-likelihood together with a red and a blue line indicating the vector Y_t (i.e. a linear combination of the structural shocks ϵ_t), corresponding to two different choices for α.

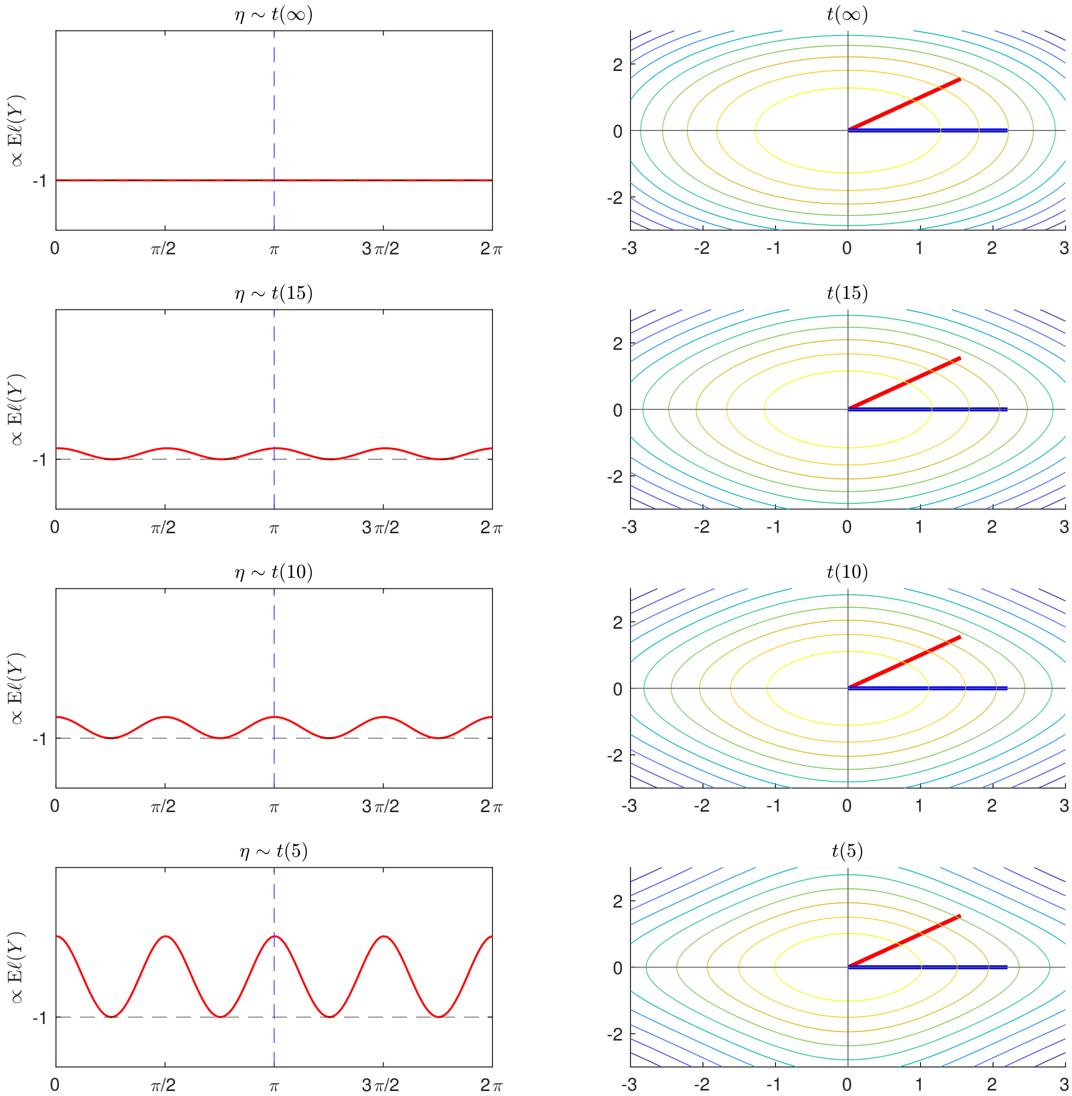
As we move away from the Gaussian distribution the expected gradient of $\ell_\alpha(Y_t)$ with respect to α becomes non-zero in the vicinity of the true parameter (here set as $\alpha = \pi$). Equivalently, in the contour plots the red and blue lines reach different level curves. This means that different choices of α lead to different values of the log-likelihood and hence, α is locally identifiable. That only local identification occurs is clear, as the same level curves are reached in each quadrant of the contour plot, with each quadrant corresponding to a permutation and/or sign change of the rows of A. These examples illustrate how non-Gaussianity of the structural shocks can help to identify parameters up to permutation and sign changes.

The problem of weak non-Gaussianity arises when the distance from the Gaussian distribution is not very large. In such scenarios, changes in α only imply minor changes in the level of the likelihood, so that the likelihood ends up being rather flat around the true parameter α. Compare for instance, the panels corresponding to the $t(5)$ density and the $t(15)$ density. In the case of the $t(5)$ density, the red and blue vectors end on clearly distinguishable contour lines of the log-likelihood and the value of the log likelihood varies substantially around $\alpha = \pi$. In contrast, for the $t(15)$ density, the differences are small and the red and blue vectors almost reach the same contour line. In the extreme case of Gaussian shocks (i.e. the upper panels of Figure 1) we find that α is completely unidentified.

Whilst in population we would always be able to locally identify α when the densities of the structural shocks differ from Gaussian, in finite sample, if the densities of the structural shocks are close to Gaussianity, the available identifying information may be small relative to sampling variability. This creates a problem when standard test statistics are used as, in such a setting, standard (i.e. fixed parameter) asymptotics provide a poor approximation to the finite sample behaviour of test statistics.

To illustrate this point, consider the standard Wald statistic that is based on the maximum likelihood estimate for α. When the densities are non-Gaussian α is (locally) identified and the Wald statistic converges to its usual $\chi^2(1)$ limit. In contrast, when the densities are Gaussian α is unidentified and the Wald statistic has a non-standard limit. Such a discontinuity implies that in finite sample the $\chi^2(1)$ distribution may provide a poor approximation to the finite sample behaviour of the test when the true densities are close to Gaussian relative to the sample size.

FIGURE 1: IDENTIFICATION WITH NON-GAUSSIAN DISTRIBUTIONS



Note: The figure shows the expected log-likelihood as a function of α (left panels) and the expected likelihood contours (right panels) for the SVAR(0) model with different distributions for the structural shocks $\epsilon_{k,t}$. The red and blue lines in the right plots denote the vector Y_t corresponding to two different choices for α.

In this paper we address this problem by developing a robust semi-parametric approach for constructing confidence bands for α. The methodology is based on inverting a score-type test for $\alpha = \alpha_0$. The key result is that the semi-parametric score statistic is asymptotically χ^2 both in the well identified case and when the structural shocks are close to Gaussianity. This ensures that the test is reliable for a large class of (true) densities: non-Gaussian, close to Gaussian, or even Gaussian densities are all permitted provided some regularity conditions are satisfied.[8]

[8] Alternatively, one could think of constructing some threshold statistic — analogous to the F-statistic used in instrumental variable studies — that indicates whenever the distance to the Gaussian distribution is sufficiently large such that standard methods can be used, see Guay (2021) for such approach when higher order moments are used for identification. While practically attractive, the results of this paper render developing this alternative unnecessary as the semi-parametric score test is efficient under strong identification.

3 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 parametrised by the vectors α and σ.

In general, we will leave the choice for the specific parametrisation of $A(\alpha,\sigma)$ open to the researcher. The key restriction is that σ should be recoverable from the variance of $Y_t - BX_t$, whereas α may be unidentified depending on the distribution of the structural shocks. A canonical 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) defined by the vector σ and $R(\alpha)$ is a rotation matrix that is parametrised by the vector α. That said, different parametrisations are often used in practice (cf Section 8) and our general formulation allows for such alternatives.

We let $\eta = (\eta_1, \dots, \eta_K)$ correspond to the density functions of $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{K,t})'$ and summarise the parameters in

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

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

Let $Y^n = (Y_1, \dots, Y_n)'$ and let P^n_θ denote the distribution of Y^n conditional on the initial values $(Y_{1-p}, \dots, Y_0)$. 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

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

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

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

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 $\kappa(z) = z^2 - 1$. The parameter space for the densities η_k is thus 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 3.1 details the main assumptions that allow us to establish the properties of the semi-parametric score test and Assumption 3.2 defines a set of regularity conditions on densities η_k under which the log density scores can be consistently estimated using B-splines.[9] These scores are an important ingredient for the methodology discussed below.

The main assumption is stated as follows.

Assumption 3.1: *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 $(Y'_{-p+1}, \dots, Y'_0)'$, $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{K,t})'$ is independently and identically distributed across t, with independent components $\epsilon_{k,t}$. Each $\eta = (\eta_1, \dots, \eta_K) \in \mathcal{H}$ 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 $k = 1, \dots, K$

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

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

(iii) 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 $D_{\alpha_l}(\alpha, \sigma) := \partial A(\alpha, \sigma)/\partial \alpha_l$ and $D_{\sigma_m}(\alpha, \sigma) := \partial A(\alpha, \sigma)/\partial \sigma_m$, are Lipschitz continuous for all $l = 1, \dots, L_\alpha$, $m = 1, \dots, L_\sigma$ and $j, k = 1, \dots, K$, where the notation $B_{\bullet j}$ or $B_{j\bullet}$ denotes the jth column or row of a matrix B.

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.[10] 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, lemma S12 in Lee and Mesters (2022b) shows that if (a) holds, then a simple sufficient condition is that the tails of the densities η_k converge to zero at a polynomial rate.[11] The final part (iii) of the assumption imposes that $A(\alpha, \sigma)$ is invertible and that the parametrisation chosen by the researcher is sufficiently smooth. For instance, for the canonical choice $A^{-1}(\alpha, \sigma) = \Sigma^{1/2}(\sigma)R(\alpha)$, when

[9] Lemma A.1 in the Appendix shows that, under Assumptions 3.1 and 3.2, 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.

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

[11] Alternatively, these moment conditions hold if one can interchange integration and differentiation appropriately.

we model $R(\alpha)$ by the Cayley or trigonometric transformation parts (b) and (c) can easily be verified to hold.

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

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

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

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

(iv) For each n, $\phi_{k,n}$ *is three-times continuously differentiable on* $[\Xi^L_{k,n}, \Xi^U_{k,n}]$ *and* $\|\phi^{(3)}_{k,n}\|^2_\infty \delta^6_{k,n} = o(\nu_n)$;[12]

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

These assumptions are similar to those considered in Chen and Bickel (2006). They ensure that the log density scores can be estimated sufficiently accurately using B-spline regressions (as explained in section 5).[13] 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.

4 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).[14] 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.

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

[13]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.

[14]The uniformity here is over the finite dimensional parameters, $\gamma = (\alpha, \beta)$. The results in the present paper only require uniformity over α, but the additional uniformity over β follows at essentially no additional cost.

4.1 Uniform Local Asymptotic Normality

Let $(\gamma_n)_{n\in\mathbb{N}} \subset \Gamma$ be such that $\gamma_n \to \gamma \in \Gamma$, fix $\eta \in \mathcal{H}$ and put $\theta := (\gamma, \eta)$. Let G_k denote the law on $\mathbb{R}$ corresponding to η_k $(k = 1, \dots, K)$ and define

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

where ι is the identity funcion, $\kappa(z) = z^2 - 1$ (as defined above) and $\mathcal{C}_b^1(\lambda)$ denotes the class of real functions on $\mathbb{R}$ which are bounded, continuously differentiable and have bounded derivatives λ-a.e.. Note that $\mathbb{R}^{L_\alpha + L_\beta} \times \dot{\mathscr{H}}$ is a linear subspace of $\mathbb{R}^{L_\alpha + L_\beta} \times L_\infty(\lambda)^K$. We make this into a normed space by equiping it with the norm $\|(c, h)\| := \|c\|_2 + \sum_{k=1}^{K} \|h_k\|_{\lambda,\infty}$ where $\|\cdot\|_2$ denotes the Euclidean norm.

For an arbitrary sequence $(c_n)_{n\in\mathbb{N}} \subset \mathbb{R}^{L_\alpha + L_\beta}$ such that $c_n := (a_n', d_n')' \to (a', d')' =: c$ let $\tilde{\gamma}_n := \gamma_n + c_n/\sqrt{n}$ and for an arbitrary $(h_n)_{n\in\mathbb{N}} \subset \dot{\mathscr{H}}$ with $h_n \to h \in \dot{\mathscr{H}}$ let $\tilde{\eta}_n := \eta(1 + h_n/\sqrt{n})$. Collect these parameters into $\theta_n := (\gamma_n, \eta)$ and $\tilde{\theta}_n := (\tilde{\gamma}_n, \tilde{\eta}_n)$ respectively. Denote by p_θ^n the density of P_θ^n with respect to λ^n and $\Lambda^n_{\tilde{\theta}_n/\theta_n}$ the (conditional) log likelihood ratio:

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

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

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

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

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

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

We collect these scores in the vector

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

Under assumption 3.1, we have the following ULAN result.[15]

PROPOSITION 4.1 (ULAN): *Suppose that assumption 3.1 holds. Then as $n \to \infty$,*

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

where the expectation is taken under $P^n_{\theta_n}$ and

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

with $A_n = A(\alpha_n, \sigma_n)$. Moreover, under $P^n_{\theta_n}$,

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

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

COROLLARY 4.1: *If assumption 3.1 holds, then the sequences $(P^n_{\theta_n})_{n\in\mathbb{N}}$ and $(P^n_{\tilde{\theta}_n})_{n\in\mathbb{N}}$ are mutually contiguous.*

The importance of this result is that the semi-parametric SVAR model can be locally asymptotically approximated by a Gaussian shift experiment, uniformly in γ. 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. One example is a setting where an initial $\sqrt{n}$-consistent estimate for α is available, say by imposing an additional identifying assumption, then the (U)LAN result may be used to obtain semi-parametrically efficient parameter estimates similarly to as was done in Chen and Bickel (2006) for the ICA model.[16]

4.2 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 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.

[15] The proof is based on verifying the conditions of Theorem 2.1.2 in Taniguchi and Kakizawa (2000), which is due to Swensen (1985, Lemma 1).

[16] Constructing such an estimator is suggested in Fiorentini and Sentana (2022) based on a non-Gaussian identifying assumption. This could be easily done based on the results in this paper, but in that set-up it would remain vulnerable to weak deviations from the Gaussian density. Therefore the score testing approach developed below should be preferred whenever assessing the uncertainty around parameters is required.

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

LEMMA 4.1: *Given Assumption 3.1 the efficient score function for γ in the semi-parametric SVAR model $\mathcal{P}_\Theta^n$ at any $\theta = (\gamma, \eta)$ with $\gamma = (\alpha, \beta)$, $\alpha \in \mathcal{A}$, $\beta = (\sigma, b) \in \mathcal{B}$ and $\eta \in \mathcal{H}$ is given by $\tilde{\ell}_{n,\theta}(Y^n) = \sum_{t=1}^n \tilde{\ell}_\theta(Y_t, X_t)$, where*

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

with components

$$\tilde{\ell}_{\theta,\alpha_l}(Y_t, X_t) = \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta_{l,k,j}^\alpha \phi_k(A_{k\bullet} V_{\theta,t}) A_{j\bullet} V_{\theta,t} + \sum_{k=1}^K \zeta_{l,k,k}^\alpha \left[\tau_{k,1} A_{k\bullet} V_{\theta,t} + \tau_{k,2} \kappa(A_{k\bullet} V_{\theta,t})\right]$$

$$\tilde{\ell}_{\theta,\sigma_l}(Y_t, X_t) = \sum_{k=1}^K \sum_{j=1, j\neq k}^K \zeta_{l,k,j}^\sigma \phi_k(A_{k\bullet} V_{\theta,t}) A_{j\bullet} V_{\theta,t} + \sum_{k=1}^K \zeta_{l,k,k}^\sigma \left[\tau_{k,1} A_{k\bullet} V_{\theta,t} + \tau_{k,2} \kappa(A_{k\bullet} V_{\theta,t})\right]$$

$$\tilde{\ell}_{\theta,b_l}(Y_t, X_t) = \sum_{k=1}^K -A_{k\bullet} D_{b_l} \left[(X_t - \mu)\phi_k(A_{k\bullet} V_{\theta,t}) - \mu(\varsigma_{k,1} A_{k\bullet} V_{\theta,t} + \varsigma_{k,2} \kappa(A_{k\bullet} V_{\theta,t}))\right]$$

where $V_{\theta,t} = Y_t - BX_t$, $\zeta_{l,k,j}^\alpha := [D_{\alpha_l}(\alpha,\sigma)]_{k\bullet} A_{\bullet j}^{-1}$ with $D_{\alpha_l}(\alpha,\sigma) := \partial A(\alpha,\sigma)/\partial \alpha_l$, $\zeta_{l,k,j}^\sigma := [D_{\sigma_l}(\alpha,\sigma)]_{k\bullet} A_{\bullet j}^{-1}$ with $D_{\sigma_l}(\alpha,\sigma) := \partial A(\alpha,\sigma)/\partial \sigma_l$, $D_{b_l} := \partial B/\partial b_l$, $\mu := \mathrm{vec}(\iota_K, (\iota_p \otimes (I_K - B_1 - \ldots - B_p)^{-1} c))$, 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 (2022a). 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

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

With Lemma 4.1 and the efficient information matrix in place, we can define 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 S11 in Lee and Mesters (2022b). The singularity plays an important role in the construction of the semi-parametric score statistic below.

5 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 3.1 and 3.2 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 the structural impulse responses 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 4.1. 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

$$\hat{\ell}_{\gamma,\alpha_l}(Y_t,X_t) = \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^{\alpha}_{l,k,j}\hat{\phi}_{k,n}(A_{k\bullet}V_{\gamma,t})A_{j\bullet}V_{\gamma,t} + \sum_{k=1}^{K}\zeta^{\alpha}_{l,k,k}\left[\hat{\tau}_{k,1}A_{k\bullet}V_{\gamma,t} + \hat{\tau}_{k,2}\kappa(A_{k\bullet}V_{\gamma,t})\right]$$

$$\hat{\ell}_{\gamma,\sigma_l}(Y_t,X_t) = \sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K} \zeta^{\sigma}_{l,k,j}\hat{\phi}_{k,n}(A_{k\bullet}V_{\gamma,t})A_{j\bullet}V_{\gamma,t} + \sum_{k=1}^{K}\zeta^{\sigma}_{l,k,k}\left[\hat{\tau}_{k,1}A_{k\bullet}V_{\gamma,t} + \hat{\tau}_{k,2}\kappa(A_{k\bullet}V_{\gamma,t})\right]$$

$$\hat{\ell}_{\gamma,b_l}(Y_t,X_t) = \sum_{k=1}^{K} -A_{k\bullet}D_{b_l}\left[(X_t-\bar{X}_n)\hat{\phi}_{k,n}(A_{k\bullet}V_{\gamma,t}) - \bar{X}_n(\hat{\zeta}_{k,1}A_{k\bullet}V_{\gamma,t} + \hat{\zeta}_{k,2}\kappa(A_{k\bullet}V_{\gamma,t}))\right]$$

where $V_{\gamma,t} = Y_t - BX_t$ and $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$.[17] The estimates for the τ_k's and ζ_k's are obtained by replacing the population moments defined in Lemma 4.1 by their sample counterparts: $\hat{\tau}_k = \hat{M}_k(0,-2)'$ and $\hat{\zeta}_k = \hat{M}_k(1,0)'$, 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 $i \in [B_{k,n}]$. The knots of the splines, $\xi_{k,n} = (\xi_{k,n,i})_{i=1}^{K_{k,n}}$ are taken as equally spaced in $[\Xi^L_{k,n}, \Xi^U_{k,n}]$.[18]

Our estimate for the log density score ϕ_k is given by

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

where

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

This shows that computing the log density score estimate (17) only requires computing the B-spline regression coefficients $\hat{\psi}_{k,n}$ in (18).

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

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

[18]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 $\gamma = (\alpha,\beta)$.

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 3.2, we define a truncated eigenvalue version of the information matrix estimate as

$$\hat{\mathcal{I}}^t_{n,\gamma} = \hat{U}_n\hat{\Lambda}_n(\nu_n)\hat{U}_n' \ , \tag{22}$$

where $\hat{\Lambda}_n(\nu_n)$ is a diagonal matrix with the ν_n-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 $\hat{\lambda}_{n,i}\mathbf{1}(\hat{\lambda}_{n,i} \geq \nu_n)$. 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 $\gamma = (\alpha, \beta)$. To evaluate the null hypothesis (14) we will use $\alpha = \alpha_0$, i.e. fixing the unidentified parameters under the null, and $\hat{\beta}_n$, some $\sqrt{n}$-consistent estimate for the finite dimensional nuisance parameters.

For such parameter choices, the limiting distribution of $\hat{S}_{n,\gamma}$ is derived in the following theorem.

THEOREM 5.1: *Let $\gamma_n = (\alpha_n, \beta) \to \gamma$ with each γ_n, γ in Γ and let $\theta_n := (\gamma_n, \eta) \to (\gamma, \eta) = \theta$ for some $\eta \in \mathcal{H}$. Suppose that under $P^n_{\theta_n}$, $\hat{\beta}_n$ is a $\sqrt{n}-$consistent estimator of β. Define $\mathscr{S}_n = n^{-1/2}C\mathbb{Z}^{L_2}$ for some $C > 0$ and let $\overline{\beta}_n$ be a discretized version of $\hat{\beta}_n$ which replaces its value with the closest point in $\mathscr{S}_n$. Define $\overline{\gamma}_n = (\alpha_n, \overline{\beta}_n)$, suppose that assumptions 3.1 and 3.2 hold. Let $r_n = \operatorname{rank}(\hat{\mathcal{I}}^t_{n,\overline{\gamma}_n})$ and denote by c_n the $1-a$ quantile of the $\chi^2_{r_n}$ distribution, for any $a \in (0,1)$. Then if $\tilde{\theta}_n := (\alpha_n, \tilde{\beta}_n, \tilde{\eta}_n)$ where $\sqrt{n}\|\tilde{\beta}_n - \beta\| = O(1)$ and $\tilde{\eta}_n = \eta(1 + h_n/\sqrt{n})$ with h_n in some compact $\dot{\mathscr{H}}_\star \subset \dot{\mathscr{H}}$,*

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

with inequality only if $\operatorname{rank}(\tilde{\mathcal{I}}_{\theta_0}) = 0$.

The theorem shows that the efficient score test (23) is 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 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. However, given that the efficient scores of 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), as this typically improves the (finite sample) power of the test.[19] That said, conventional OLS estimates for the regression coefficients $b = \text{vec}(B)$ and the variance parameters σ can also be used.

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.[20] 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 ν_n, which theoretically needs to satisfy Assumption 3.2, 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 matter 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.[21] Moreover, if the rank of $\tilde{\mathcal{I}}_\theta$ is positive, the singularity adjusted score statistic is (locally asymptotically) minimax optimal, as can be shown by an argument analogous to that given in Lee (2022).

Finally, the theorem is stated along (local) sequences of parameter values $\tilde{\theta}_n$. By standard arguments one can translate such limit statements along sequences to limit statements that hold uniformly over certain sets. In the present case a uniform statement would hold over, for example, sets of the form $\mathscr{P}_n := \{P^n_{\alpha,\beta+d/\sqrt{n},\eta(1+h/\sqrt{n})} : \alpha \in \mathcal{A}_\star, \|d\| \leq M, h \in \dot{\mathscr{H}}_\star\}$ where $\mathcal{A}_\star \subset \mathcal{A}$, $\dot{\mathscr{H}}_\star \subset \dot{\mathscr{H}}$ are compact, with $\dot{\mathscr{H}}$ as defined in (5), and $M \in (0, \infty)$.

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

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

[19] See the simulation results of section 7.

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

[21] 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.

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

COROLLARY 5.1: *Let γ_n, θ_n, $\tilde{\theta}_n$, $\hat{\beta}_n$, $\overline{\beta}_n$ and $\overline{\gamma}_n$ be as in Theorem 5.1 and suppose that assumptions 3.1 and 3.2 hold. Then,*

$$\lim_{n\to\infty} P^n_{\tilde{\theta}_n}\left(\alpha_n \in \hat{C}_{n,1-a}\right) \geq 1-a. \tag{24}$$

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 $b_n = \text{vec}(B_n)$, 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 $r_n = \text{rank}(\hat{\mathcal{I}}^t_{n,\hat{\gamma}_n})$.

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

The algorithm highlights that the computation costs for computing the confidence set are modest. In fact, the costs are similar to those for constructing standard weak instrument robust confidence sets, such as those based on the Anderson-Rubin statistic for instance (e.g. Andrews et al., 2019). The only difference is that we require K regression estimates (to estimate the log density scores) as opposed to one.

6 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 integer. The following assumption restricts the class of functions that we consider.

ASSUMPTION 6.1: *$g : D_g \to \mathbb{R}^{d_g}$ 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 a uniform delta-method (cf Theorem A.2 in the appendix), whereas the rank condition ensures that no further degeneracy in the asymptotic distribution occurs, apart from that caused by α being possibly non-identified.

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

EXAMPLE 6.1: *Consider the vector that collects all sIRF at horizon l*

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

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 3.1, 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

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

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 (2021).

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 5.1. Second, for each $\alpha \in \hat{C}_{n,1-q_1}$ we compute $\hat{\nu}_{\alpha,n} := g(\alpha, \hat{\beta}_n)$. The confidence set for $\hat{\nu}_{\alpha,n}$ is given by

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

where $\nu := g(\alpha, \beta)$ 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 6.1: *Suppose that assumption 6.1 holds and let γ_n, θ_n, $\tilde{\theta}_n$ be as in Theorem 5.1. Suppose $\hat{\beta}_n$ is a sequence of estimates such that*

$$\sqrt{n}(\hat{\beta}_n - \tilde{\beta}_n) \overset{P^n_{\tilde{\theta}_n}}{\rightsquigarrow} \mathcal{N}(0, \Sigma) , \quad \textit{with } \Sigma \textit{ positive definite,}$$

and $\hat{\Sigma}_n$ is a sequence of estimates such that $\hat{\Sigma}_n \xrightarrow{P^n_{\tilde{\theta}_n}} \Sigma$, then the confidence set $\hat{C}_{n,g,\alpha}$ in (26) *satisfies*

$$\lim_{n \to \infty} P^n_{\tilde{\theta}_n} \left(g(\alpha_n, \tilde{\beta}_n) \in \hat{C}_{n,g,\alpha_n,1-q_2} \right) = 1 - q_2 . \tag{27}$$

The proposition formally establishes that if $\hat{\beta}_n$ is asymptotically normal along the local sequences $\tilde{\theta}_n$, 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. Under Assumption 3.1 both OLS and one-step efficient estimates for the parameters β satisfy the required conditions on $\hat{\beta}_n$. That is, the proposition can be restated for such specific estimators after requiring Assumption 3.1 to hold. Moreover, conventional variance estimators for Σ can be adopted to satisfy the consistency of $\hat{\Sigma}_n$, see Kilian and Lütkepohl (2017, Chapter 9) for more details.

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

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

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

Corollary 6.1: *Let $\tilde{\theta}_n$ be as in Theorem 5.1, $\hat{C}_{n,1-q_1}$ satisfies Corollary 5.1 and $\hat{C}_{n,g,\alpha_n,1-q_2}$ satisfies Proposition 6.1, then*

$$\liminf_{n \to \infty} P^n_{\tilde{\theta}_n} \left(g(\alpha_n, \tilde{\beta}_n) \in \hat{C}_{n,g} \right) \geq 1 - q .$$

This Corollary requires only the conclusions of Corollary 5.1 and Proposition 6.1.[22] For convenience we summarise 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.

7 Finite sample performance

This section presents the results from a collection of simulation studies that were designed to evaluate the size and power of the proposed inference procedures. First, we evaluate the size and power of the score test for α, as discussed in Section 5, and compare its performance to existing approaches. Second, we evaluate the coverage and length of the confidence intervals for structural impulse responses using the methodology of Section 6.

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 $p = 1, 2, 4, 8, 12$ 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 $\epsilon_{k,t}$ with $k = 1, \ldots, K$. The density functions that we consider and their abbreviated names are reported in Table 1. We standardise the draws to have mean zero and unit variance.

We parametrise 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 parametrised using the trigonometric transformation as in Section 2. In the bivariate case, $L_\alpha = 1$ and we choose $\alpha_0 = \pi/5$ for the data-generating process. In the trivariate SVAR, $L_\alpha = 3$ and we use $\alpha_0 = (3\pi/5, 2\pi/5, -\pi/5)'$. Furthermore, we choose $\Sigma^{1/2}$ such that the diagonal elements of $\Sigma^{1/2}\Sigma^{1/2'}$ are equal to one, $\sigma_{ii} = 1$ for $i = 1, \ldots, K$, and its off-diagonal elements are equal to $\sigma_{ij} = 0.2$ for $|i - j| = 1$ and $\sigma_{ij} = 0.1$ for $|i - j| = 2$ (for $K = 3$). The SVAR coefficient matrices, $A_1, \ldots, A_p$ are generated

[22]These are proven under Assumptions 3.1 and 3.2 which, we re-iterate, do not impose that the structural shocks have non-Gaussian distributions.

Table 1: Distributions for Structural Shocks

Abbreviation	Name	Definition
$\mathcal{N}(0,1)$	Gaussian	$\frac{1}{\sqrt{2\pi}}\exp\big(-\frac{1}{2}x^2\big)$
$t(\nu)$, $\nu = 15, 10, 5$	Student's t	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\Big(1+\frac{x^2}{\nu}\Big)^{\left(-\frac{\nu+1}{2}\right)}$
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.

based on full $K \times K$ matrices with elements drawn from a $\mathcal{N}(0,1)$ distribution.[23] We use 400 burn-in periods to simulate the SVAR(p) model and use $M = 2,500$ Monte Carlo replications.

Tables 2-3 report the empirical rejection frequencies of the semi-parametric score test defined in Section 5 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 $B = 6$ cubic B-splines for the estimation of the log density scores. Table 2 reports the results when estimating the nuisance parameters β using OLS while Table 3 reports the results from using the one-step efficient estimates 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.

The first panel in Table 2 reports the OLS based results for $T = 200$. We find that for the SVAR(p) with $K = 2$ variables, the size of the test is generally very close to the nominal size of 5%. Importantly, this holds even when the shocks are normally distributed and α is not identified. Further, the size remains correct for all densities that are close to Gaussian, such as the $t(15)$ and the skewed-unimodal density. For more complicated densities such as the seperate bi-modal density, some under-rejection is observed.

As the number of parameters in the SVAR increases with the lag size p or the number of variables K, the rejection rates increase and the test starts to over-reject when $T = 200$. For an increase in the number of lags, rejection rates only increase slightly, but when the number of variables increases, the number of parameters grows quadratically and hence the rejection rates show a more substantial increase. Importantly, this holds regardless of the true underlying density considered and is caused by the rather imprecise OLS estimates that are plugged into the score test statistic.

[23] To ensure stationarity of the SVAR(p) model, the coefficient matrices are transformed using the approach of Ansley and Kohn (1986).

When we increase the sample size ($T = 500$, $T = 1000$) these size distortions quickly disappear and the rejection frequencies converge to the nominal size of the test. Thus, even in the case of an SVAR with a larger lag length, the testing procedure gives correct inference, as long as the sample size is not too small. We note that we continue to see under-rejection for some of the densities far from Gaussianity.

Table 3 reports the empirical rejection frequencies for the same simulation design, but now one-step efficient estimates are used for the nuisance parameters. The one-step efficient estimates of β are obtained by updating the OLS estimates of the nuisance parameters β towards the efficient estimates by one Gauss-Newton iteration. Comparing the rejection rates in Table 3 with those reported in the case of OLS estimates of the nuisance parameters in Table 2, shows that using the one-step estimates yields substantial improvements in the size of the test in small samples, especially when the number of lags is large. For example, for the case of an SVAR with three variables and 12 lags, the size of the rejection rates are very close to the nominal size of 5%. Further, we note that using one-step efficient updates of β also remedies the under-rejection observed for some of the Gaussian mixture distributions in Table 2. As the sample size grows, the difference between Tables 2-3 is less pronounced and the procedures yield comparable rejection rates.

Overall, we may conclude that the empirical size of the test is close to the nominal size regardless of the choice for the true densities, i.e. Gaussian, close to Gaussian, or far from Gaussian. Finite sample size distortions can be largely overcome by using one-step efficient estimates.

7.2 Comparison to alternative approaches

Next, we compare the performance of the semiparametric score test to a variety of alternative methods that have been proposed in the literature based on size and power. We distinguish between two types of 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 tests). Clearly, from the discussion in Section 2 it follows that we expect the tests in the first category to perform poorly as they are vulnerable to identification failures.[24]

In the first category, we consider the pseudo maximum likelihood Wald test ($\mathrm{W}^{\mathrm{PML}}$) of Gouriéroux et al. (2017) which we implement using one (standardised) $t(7)$ density and a (standardised) $t(12)$ density for the second shock, as in Gouriéroux et al. (2017). We additionally consider two tests based on the work of Lanne and Luoto (2021): these are the GMM Wald (W^{LL}) and distance metric ($\mathrm{DM}^{\mathrm{LL}}$) tests based on higher (third & fourth) order moment conditions.

In the second category we consider four tests. Firstly we have the pseudo maximum likelihood Lagrange Multiplier test ($\mathrm{LM}^{\mathrm{PML}}$) 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

[24]Simulation evidence in Lee and Mesters (2022a) has shown that tests that do not fix α under the null often show severe over-rejection in ICA models when the errors are close to Gaussian.

Table 2: Empirical rejection frequencies using OLS Estimates

K	p	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$T = 200$											
2	1	4.76	5.48	4.72	4.74	3.98	4.22	2.32	2.40	4.26	2.04
2	2	5.04	5.06	5.56	4.84	4.30	4.52	2.38	2.50	3.86	1.60
2	4	5.00	5.80	5.76	5.42	4.16	4.86	2.72	3.00	3.96	1.78
2	8	6.40	7.00	6.44	5.76	5.38	5.58	3.64	3.28	5.46	2.12
2	12	7.42	7.20	7.30	7.10	6.30	6.22	4.06	4.14	6.16	4.02
3	1	5.38	5.80	5.56	7.04	6.22	6.64	5.68	4.54	5.64	4.68
3	2	6.22	6.66	7.12	7.52	5.92	5.26	5.26	4.36	5.76	4.56
3	4	8.12	7.70	8.88	8.22	7.28	4.88	5.22	3.92	7.16	3.80
3	8	12.16	12.78	11.90	12.58	8.82	7.44	6.84	5.10	9.48	6.24
3	12	16.98	17.32	16.62	15.54	11.32	10.32	10.38	7.94	14.10	8.92
$T = 500$											
2	1	4.48	4.54	4.90	3.98	3.40	4.48	1.60	1.50	3.42	1.26
2	2	4.56	4.66	4.92	4.16	3.80	4.92	1.78	1.76	3.06	1.70
2	4	4.88	4.90	4.94	4.14	3.76	5.50	2.40	2.60	2.88	1.48
2	8	5.48	5.56	4.84	5.42	4.64	6.06	2.94	2.68	4.26	1.24
2	12	6.16	6.04	6.26	5.34	5.54	6.84	3.14	4.00	4.56	1.90
3	1	5.16	5.54	6.00	5.76	5.46	5.96	5.56	4.94	5.14	5.30
3	2	5.58	5.88	6.62	6.56	5.34	5.30	5.26	4.78	5.00	4.68
3	4	6.66	6.36	6.48	6.52	6.16	4.76	4.62	4.30	5.72	4.10
3	8	7.82	8.06	8.20	8.90	7.86	5.50	4.76	3.48	6.18	4.78
3	12	11.06	9.80	10.98	9.96	8.76	6.00	5.70	4.20	7.12	5.18
$T = 1,000$											
2	1	4.76	4.46	4.36	3.60	3.98	4.66	1.18	1.30	3.04	1.26
2	2	5.40	4.32	4.40	4.08	3.72	4.90	1.64	1.50	3.42	1.32
2	4	4.64	4.64	5.12	4.06	3.86	4.96	1.72	1.92	3.08	1.50
2	8	5.56	4.48	4.66	4.42	4.26	6.40	2.30	2.18	3.52	1.78
2	12	5.26	5.24	5.12	4.74	5.04	7.06	2.94	2.80	3.54	2.24
3	1	5.04	4.92	4.76	4.80	4.72	5.34	4.98	4.72	4.06	4.92
3	2	4.98	5.22	5.64	5.18	5.58	5.60	5.14	4.62	4.72	4.18
3	4	6.16	5.28	5.72	5.24	5.82	5.16	4.80	4.40	4.74	4.68
3	8	6.08	6.80	6.58	6.60	6.34	4.86	5.02	5.08	4.72	4.54
3	12	6.56	7.22	7.30	7.08	7.62	6.48	5.28	5.40	4.60	4.30

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 by OLS. 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. Rejection rates are computed based on $M = 5,000$ Monte Carlo replications.

Table 3: Empirical rejection frequencies using One-step Estimates

K	p	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$T = 200$											
2	1	5.92	6.56	5.80	5.36	5.34	4.62	4.84	5.46	4.86	4.88
2	2	5.96	5.18	5.74	5.48	4.90	4.86	5.04	5.38	4.20	4.70
2	4	5.06	4.68	4.88	4.92	4.20	4.08	5.18	5.08	4.44	5.16
2	8	4.36	3.98	4.62	4.18	4.36	4.02	5.76	6.18	4.08	5.76
2	12	3.76	3.88	4.08	3.68	4.24	3.24	5.44	6.64	3.62	5.98
3	1	7.26	7.70	7.62	7.36	7.04	6.66	6.50	6.06	6.26	5.60
3	2	7.20	7.74	8.08	8.10	6.24	7.50	6.22	6.86	6.78	6.08
3	4	6.30	7.02	7.48	7.36	6.92	6.30	6.20	6.44	6.12	5.54
3	8	3.96	4.68	4.80	5.40	4.72	4.32	3.30	4.28	4.42	3.82
3	12	2.30	2.42	2.22	2.90	2.08	2.16	2.64	2.16	2.44	2.96
$T = 500$											
2	1	6.40	6.06	6.44	5.26	5.08	4.72	5.62	5.70	5.02	4.46
2	2	5.98	5.98	6.30	5.60	5.06	4.44	5.90	6.12	4.62	5.96
2	4	6.30	5.62	5.50	5.36	4.86	4.86	5.86	6.38	4.04	6.04
2	8	5.72	4.98	5.62	5.78	4.96	5.18	6.40	6.76	4.92	6.64
2	12	6.00	5.34	6.02	5.02	4.98	5.00	6.18	7.54	4.54	7.78
3	1	8.50	8.36	8.86	7.04	5.82	6.06	5.78	6.02	5.98	6.06
3	2	7.80	8.22	8.22	7.32	6.04	6.70	5.94	5.38	5.70	5.80
3	4	8.60	8.20	7.62	7.50	6.58	5.98	5.98	6.34	6.38	6.40
3	8	8.20	7.62	8.34	8.24	6.86	7.30	7.02	7.96	6.62	6.22
3	12	7.98	8.00	8.04	7.98	6.02	6.90	6.86	7.70	6.14	6.44
$T = 1,000$											
2	1	6.30	6.10	6.00	5.40	5.34	4.42	5.26	5.92	5.04	5.18
2	2	6.94	5.90	5.80	5.90	5.02	4.68	5.52	6.40	5.02	5.96
2	4	5.90	6.22	6.12	5.90	4.74	4.70	5.68	5.94	4.22	5.88
2	8	6.56	5.66	5.44	5.88	4.70	5.06	5.80	6.08	4.78	6.64
2	12	5.98	6.20	5.86	5.78	4.98	4.60	6.38	6.28	4.52	7.34
3	1	8.16	7.34	7.58	6.64	5.08	5.46	5.48	4.90	5.22	4.86
3	2	8.02	7.82	8.12	6.48	6.02	6.12	5.60	5.56	5.00	4.76
3	4	9.30	7.78	7.94	6.62	5.94	6.36	6.10	5.68	5.56	6.12
3	8	8.24	8.56	7.64	8.10	6.12	6.74	6.58	6.92	5.80	7.10
3	12	7.50	8.32	8.42	7.46	6.56	8.32	6.96	7.96	5.80	6.90

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 by the one-step efficient procedure. 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. Rejection rates are computed based on $M = 5,000$ Monte Carlo replications.

first and second shocks respectively.[25] Secondly, we consider the LM test corresponding to the GMM setup of Lanne and Luoto (2021) (LM^{LL}). Lastly, we compare to the recently proposed robust GMM methods of Drautzburg and Wright (2021). 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 (2021).

To evaluate the finite-sample performance, we focus on an SVAR(1) model with $K = 2$ variables and a sample size of $T = 500$. We use the same parametrisation and parameter values as described in the previous subsection to generate the data. However, different to the previous simulation study evaluating the size of the score test, we report results both for the case where the structural shocks $\epsilon_{1,t}, \epsilon_{2,t}$ are identically distributed, but also for the case where the first shock is fixed to have a Gaussian distribution while the distribution of the second structural shock varies. Note that in the latter case, theoretically non-Gaussianity can still be used to identify the parameters of the SVAR if the second structural shock does not follow a Gaussian distribution. However, identification is generally weaker in this case.

Size comparison

Table 4 compares the size of the different testing procedures. The first panel reports the results for the case where the two structural shocks, $\epsilon_{1,t}, \epsilon_{2,t}$ are drawn from the same (non-Gaussian) distribution while the second panel reports the results where $\epsilon_{1,t}$ is fixed to have a Gaussian distribution.

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 3). This holds regardless whether the second structural shock is Gaussian or not. 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

[25]Note that this test is not actually discussed in Gouriéroux et al. (2017), but the simulations in Lee and Mesters (2022a) show that it has reliable size for ICA models.

is somewhat stronger. The identification robust moment tests of Drautzburg and Wright (2021) (GMM^{DW} and BKR^{DW}) generally perform well, with the former always controlling size correctly 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 summarise, 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 and whether one or both shocks are non-Gaussian.

Table 4: Empirical rejection frequencies for alternative tests

Test	N(0,1)	t(15)	t(10)	t(5)	SKU	KU	BM	SPB	SKB	TRI
$\epsilon_{1,t} \sim \epsilon_{2,t}$										
$\hat{S}_{ols}$	4.56	6.24	4.72	4.56	5.16	5.16	4.28	4.40	4.16	4.56
$\hat{S}_{onestep}$	5.88	7.28	6.28	4.92	5.28	5.20	4.92	4.48	4.64	5.20
LM^{PML}	4.48	4.84	4.96	4.84	6.36	5.76	20.44	31.68	5.68	32.36
LM^{LL}	6.04	9.88	13.20	25.88	22.36	14.96	5.64	4.72	11.32	5.28
GMM^{DW}	3.40	4.04	3.92	5.24	4.88	4.36	3.04	2.36	3.56	2.96
BKR^{DW}	5.00	4.64	4.00	5.24	6.76	30.56	4.80	4.76	6.44	4.80
W^{PML}	20.44	3.16	1.60	2.40	3.36	3.32	100.00	100.00	3.12	100.00
W^{LL}	74.96	44.08	22.64	1.00	0.44	2.40	0.00	0.00	50.00	0.00
DM^{LL}	11.80	12.56	13.60	14.28	11.96	10.68	5.48	4.92	13.72	4.28
$\epsilon_{1,t} \sim \mathcal{N}(0,1)$										
$\hat{S}_{ols}$	5.12	4.52	4.64	4.40	4.16	4.36	1.60	1.12	3.48	1.88
$\hat{S}_{onestep}$	6.72	6.32	6.20	5.76	5.08	4.56	5.04	5.00	5.24	6.00
LM^{PML}	5.56	6.28	5.68	6.08	9.04	6.80	5.68	6.68	5.04	5.68
LM^{LL}	7.36	6.12	6.40	6.56	7.12	8.08	12.36	13.60	6.24	12.36
GMM^{DW}	3.00	3.84	4.36	5.56	3.60	3.20	3.04	4.52	3.32	4.08
BKR^{DW}	4.52	5.24	5.28	5.88	9.84	49.72	7.56	9.20	13.44	9.32
W^{PML}	22.20	10.40	7.64	2.04	1.88	1.44	95.08	97.68	11.20	97.92
W^{LL}	74.88	67.40	58.64	24.64	14.80	43.84	56.08	50.88	72.36	54.28
DM^{LL}	12.04	11.96	11.48	9.08	9.24	11.64	6.20	5.04	12.72	5.20

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 $\alpha_0 = \pi/5$. $\hat{S}_{ols}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. LM^{LL}, W^{LL} and DM^{LL} denote the GMM-based LM, Wald and distance metric tests of Lanne and Luoto (2021). LM^{PML} and W^{PML} denote the pseudo-maximum likelihood LM and Wald tests of Gouriéroux et al. (2017), GMM^{DW} denotes the GMM-based test of Drautzburg and Wright (2021), BKR^{DW} denotes the non-parametric test of Drautzburg and Wright (2021). 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 (2021) use 500 bootstrap replications to simulate the null distribution of the test statistics. Rejection rates are computed based on $M = 2,500$ Monte Carlo replications.

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 $T = 500$, and two independent shocks drawn from the same distribution.

Figure 2 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 test (dot - dashed blue line), the Drautzburg and Wright (2021) GMM test (dotted green line) and the non-parametric Drautzburg and Wright (2021) 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. Nevertheless, the efficient score tests are not far behind, offering almost comparable power. 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 (2021) 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 Coverage and length of confidence sets

Next, we consider evaluating our methodology for constructing confidence sets for smooth functions of the SVAR parameters as discussed in Section 6. We focus on evaluating the coverage and length of the confidence sets for structural impulse response functions, see Example 6.1 for the details.

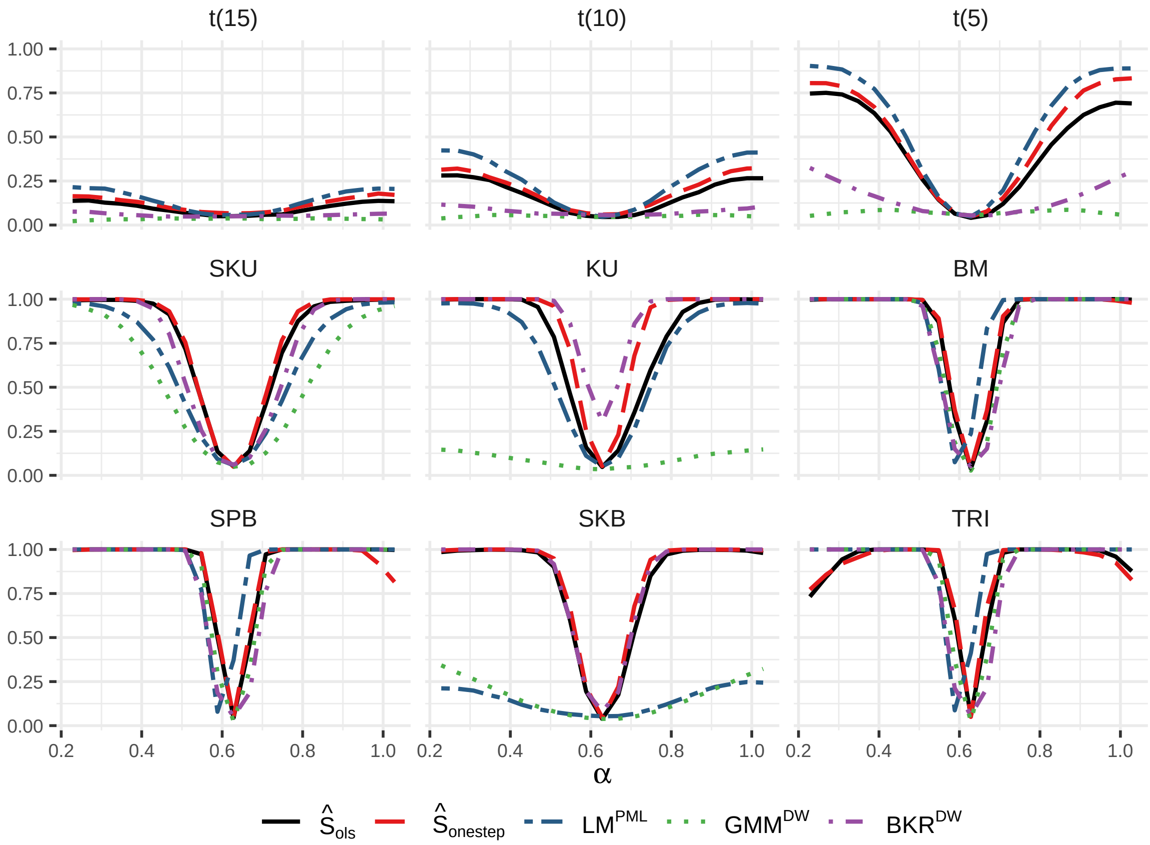
We consider a similar simulation set up as above and discuss the results for the SVAR(1) model with $K = 2$, $T = 500$, and two independent shocks drawn from the same distribution, as listed in Table 1. In each case, the confidence set is calculated using **Algorithm 2** for the structural impulse response 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 (2021), for which we change step (i) in **Algorithm 2** and replace the efficient score test by the tests of Drautzburg and Wright (2021).

Figure 3 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.

[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 (2021) are substantially more computationally demanding than the efficient score based approaches, as they use a bootstrap approach to obtain the critical value. Relying on asymptotic critical values for these tests yields substantially worse performance.

Figure 2: Power in the SVAR(1) model

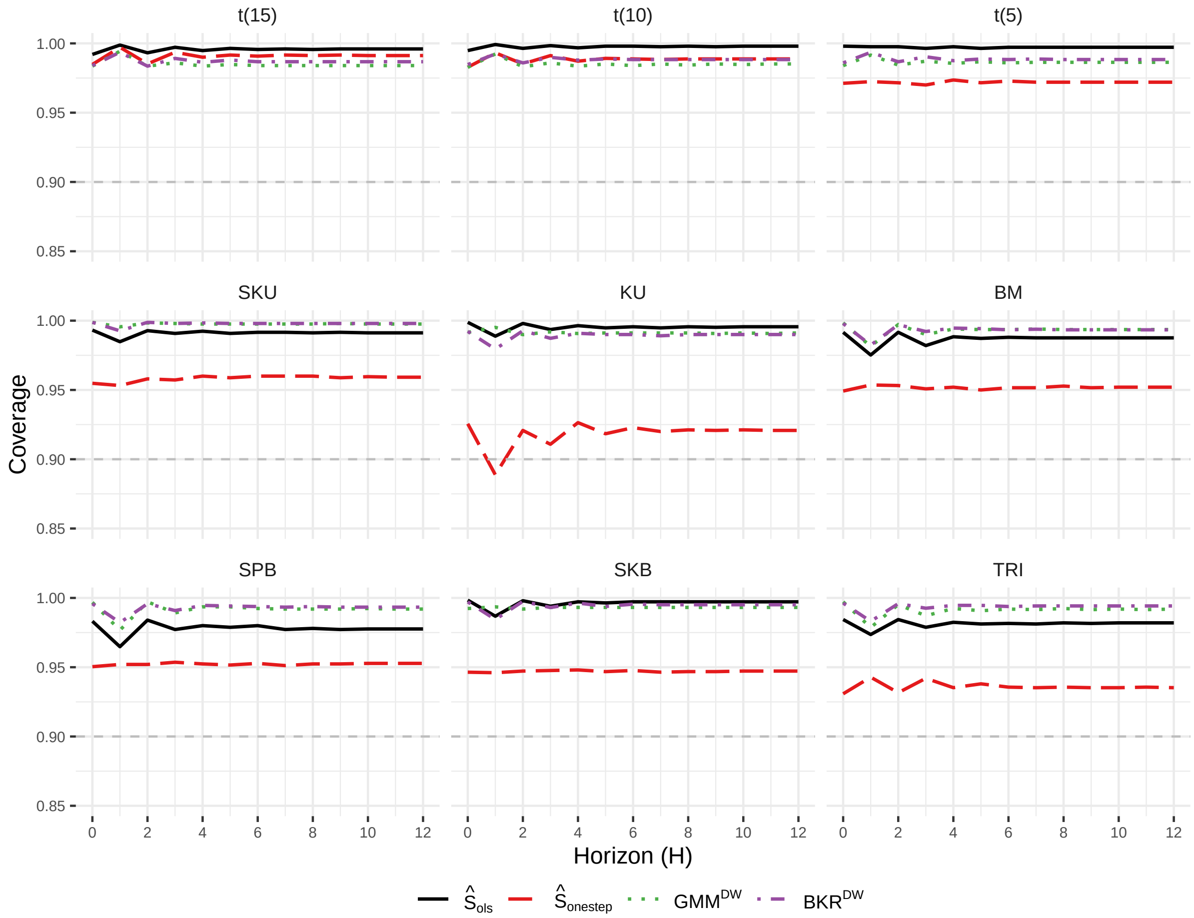


Note: The figure reports unadjusted empirical power curves for tests of the hypothesis $H_0 : \alpha = \alpha_0$ vs. $H_1 : \alpha \neq \alpha_0$ with 5% nominal size for the SVAR(1) model with $K = 2$ and $T = 500$. The x-axis corresponds to different alternatives for α around $\alpha_0 = \pi/5$. $\hat{S}_{ols}$ denotes the semi-parametric score test using OLS estimates for β, $\hat{S}_{onestep}$ uses one-step efficient estimates. LM^{PML} denotes the pseudo-maximum likelihood test of Gouriéroux et al. (2017), GMM^{DW} denotes the GMM-based test of Drautzburg and Wright (2021), BKR^{DW} denotes the non-parametric test of Drautzburg and Wright (2021). The tests of Drautzburg and Wright (2021) use 500 bootstrap replications to obtain critical values. Rejection frequencies are computed using $M = 2,500$ Monte Carlo replications.

Figure 4 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.

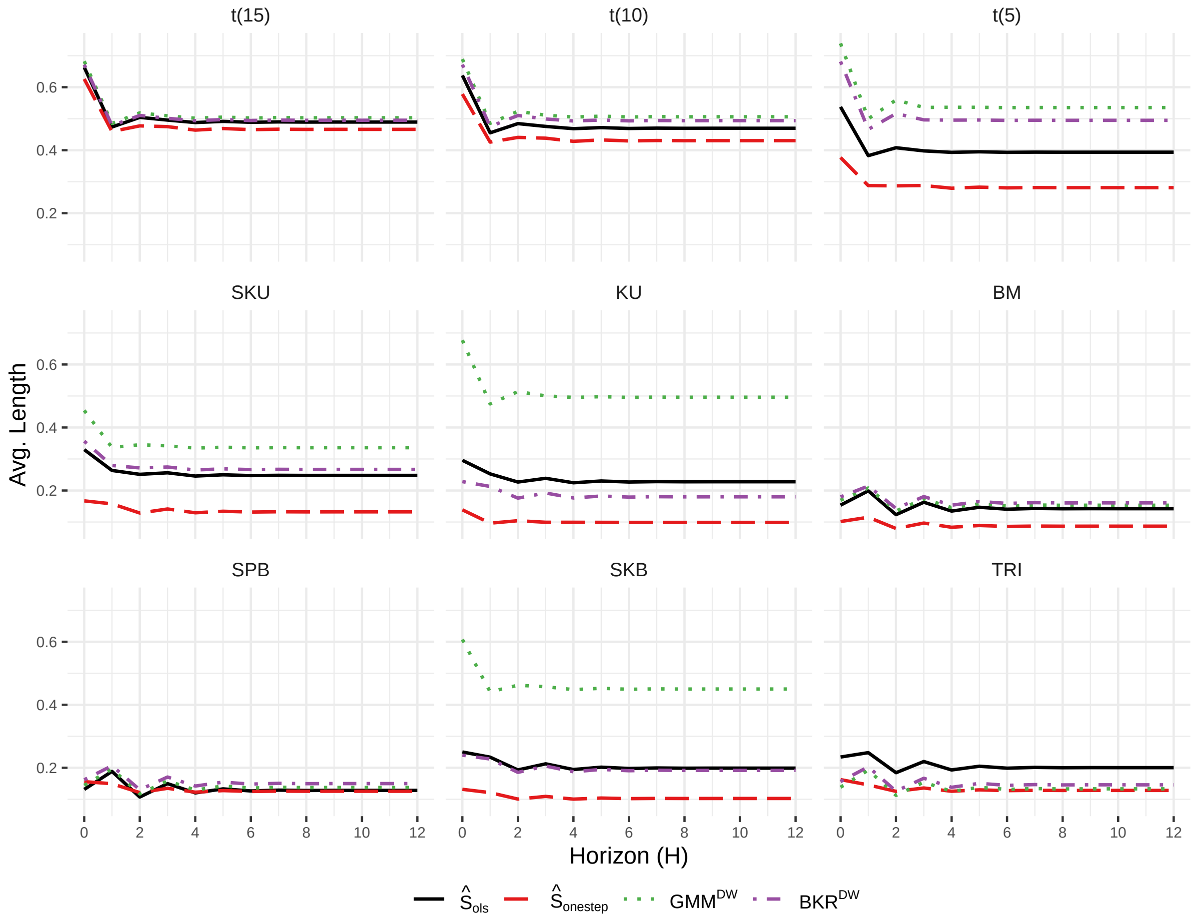
We conclude that the two-step Bonferroni method, where the first step is based on the efficient score test, gives substantial efficiency improvements when compared to existing methods.

Figure 3: 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 (2021) and BKR^{DW} denotes the non-parametric test of Drautzburg and Wright (2021). The tests of Drautzburg and Wright (2021) use 500 bootstrap replications to obtain critical values. Coverage is computed using $M = 2,500$ Monte Carlo replications.

Figure 4: 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 (2021) and BKR^{DW} denotes the non-parametric test of Drautzburg and Wright (2021). The tests of Drautzburg and Wright (2021) use 500 bootstrap replications to obtain critical values. Average length is computed using $M = 2,500$ Monte Carlo replications.

8 Empirical studies

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 parametrised as

$$Y_t = c + B_1 Y_{t-1} + \cdots + B_p Y_{t-p} + B_0^{-1} \Sigma^{1/2} \epsilon_t \ , \tag{29}$$

with

$$B_0 = \begin{bmatrix} -\alpha^d & 1 \\ -\alpha^s & 1 \end{bmatrix} , \qquad \text{and} \qquad \Sigma^{1/2} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} .$$

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; $\alpha^s > 0$ and $\alpha^d < 0$.

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.

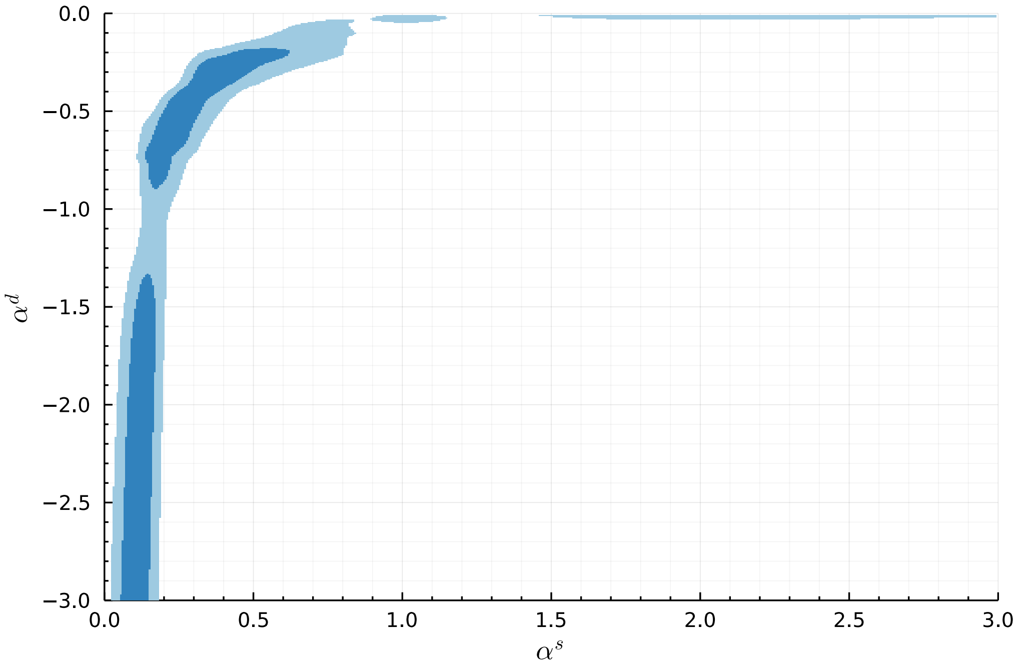
Recently, Lanne and Luoto (2019) adopted the methodology of Lanne and Luoto (2021) to assess this possibility, but this approach may yield incorrect coverage when the shocks are close to Gaussian (cf Section 7). Here we will adopt the robust score testing approach of Sections 5 and 6 to construct confidence 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 test of Matteson and Tsay (2017), see also Montiel Olea et al. (2022). For the given sample period, the test does not reject that ϵ_t has independent components (p-value = 0.248), 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 5 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 5. The confidence sets are constructed based on a grid of 250,000 equally spaced points for $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$ which covers the

Figure 5: Confidence Sets for Labor Demand and Supply Elasticities

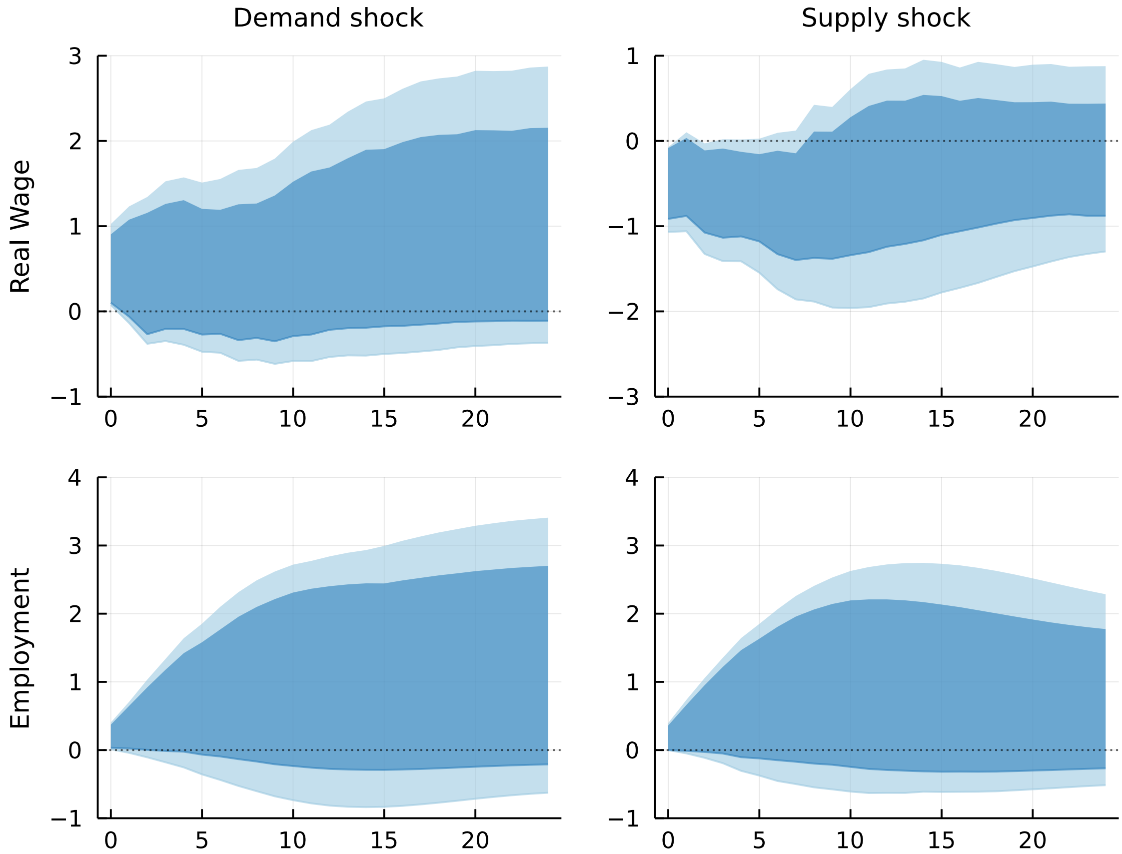


Note: 95% (light blue) and 67% (dark blue) confidence regions for labor demand and supply elasticities obtained using **Algorithm 1** with 250,000 equally-spaced grid points for $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$.

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

Note that the estimate of Lanne and Luoto (2019) obtained using non-Gaussianity identification ($\alpha^d = -0.197, \alpha^s = 0.765$) falls within our confidence set at 95% level. However, they find narrow confidence sets for the elasticity parameters (asymptotic standard errors of 0.057 for α^d and 0.196 for α^s, respectively) while our weak-identification robust approach results in much wider confidence sets, similar to the credible sets of Baumeister and Hamilton (2015).

FIGURE 6: 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 $(\alpha^d, \alpha^s) \in [-3, 0) \times (0, 3]$.

Confidence Sets for impulse responses

Figure 6 shows our identification-robust 95% and 67% confidence sets for the impulse responses to labor-demand and labor-supply shocks. Comparing the impulse response bands to the posterior credible sets reported by Baumeister and Hamilton (2015), we note that the implied impulse responses are, overall, very similar and show long and persistent responses to the supply and demand shocks. The main differences are that our 95% identification-robust bands support slightly negative long-run responses of the real wage and employment to a demand shock, as well as a more pronounced negative long-run response of employment to a supply shock while Baumeister and Hamilton (2015)'s credible sets contain only (weakly) positive responses. Comparing our results to Lanne and Luoto (2019), we note several differences. First, Lanne and Luoto (2019) 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

$$Y_t = c + B_1 y_{t-1} + \cdots + B_p Y_{t-p} + A^{-1}(\alpha,\sigma)\,\epsilon_t, \quad A^{-1}(\alpha,\sigma) = \begin{bmatrix} \sigma_1 & \alpha_{qx}\cdot\sigma_5 & \alpha_{qp}\cdot\sigma_6 \\ \sigma_2 & \sigma_4 & \alpha_{xp} \\ \sigma_3 & \sigma_5 & \sigma_6 \end{bmatrix} \tag{30}$$

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:[27]

$$A^{-1}(\alpha,\sigma) = \begin{bmatrix} + & + & + \\ + & + & - \\ - & + & + \end{bmatrix}\,. \tag{31}$$

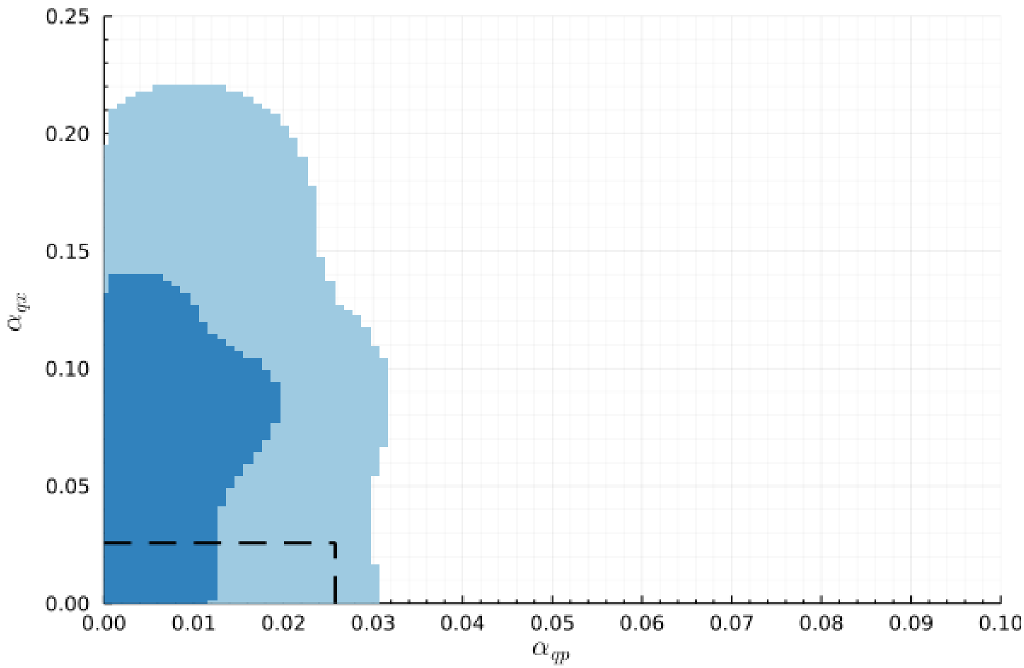
In addition, Kilian and Murphy (2012) impose a set of upper bounds on the short-run oil supply elasticities implied by the model to shrink the identified set for the impulse responses. Specifically, they assume that $\alpha_{qp} < 0.0258$, $\alpha_{qx} < 0.0258$ and that $\alpha_{xp} > -1.5$. These restrictions, 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 consider the robust score testing approach of Sections 5 and 6 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.[28]

[27] 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.

[28] We started again by testing for independent components using the permutation test of Matteson and Tsay (2017). The test does not reject that ϵ_t has independent components (p-value = 0.24).

Figure 7: 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).

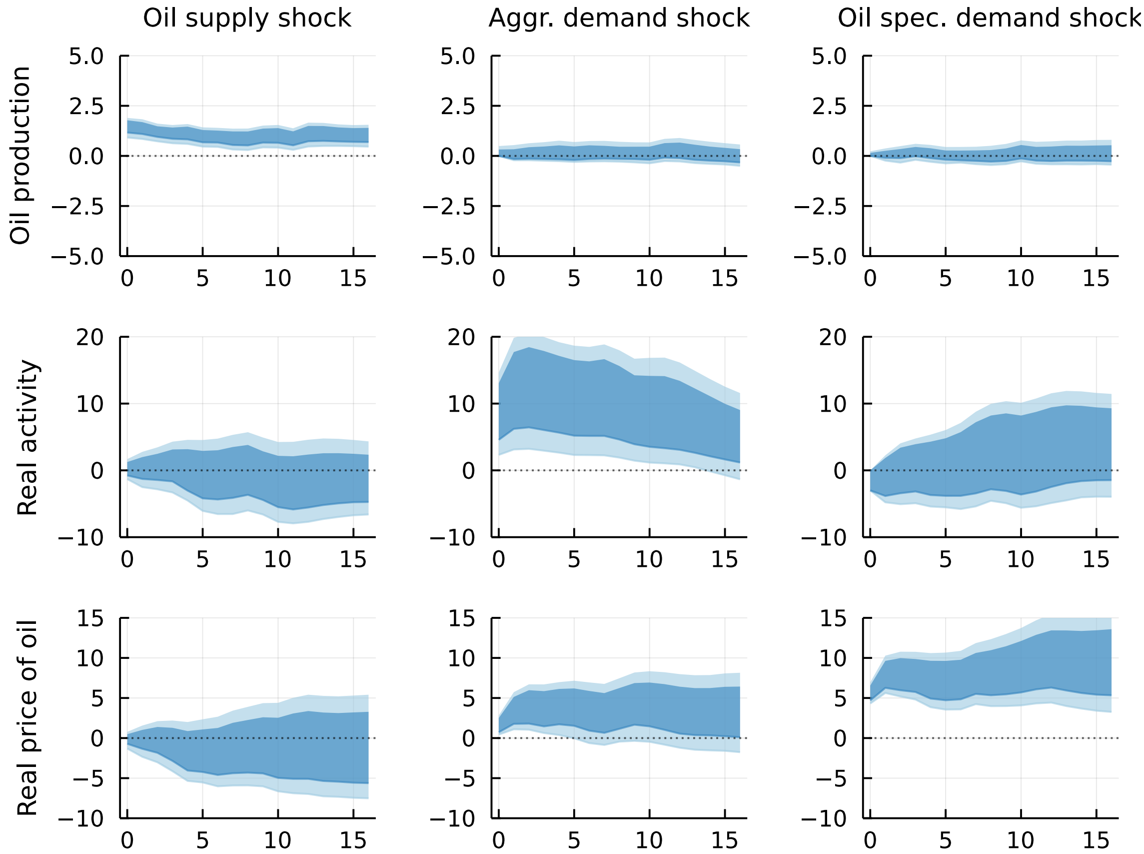
Confidence sets for oil supply elasticities $(\alpha_{qx}, \alpha_{qp})$

Figure 7 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 5 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 (31), we impose these signs in the estimation of the nuisance parameters σ.[29]

Inspecting the confidence set depicted in Figure 7, 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 8: 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 8. 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 generalisation of Neyman's C(α) 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

A.1 Density score estimation

LEMMA A.1: *Suppose Assumptions 3.1 and 3.2 hold. Let $\tilde{\theta}_n = (\alpha_n, \tilde{\beta}_n, \eta) \to \theta$ where $\sqrt{n}\|\tilde{\beta}_n - \beta\| = O(1)$. Then the log density score estimates $\hat{\phi}_{k,n}$ defined as in* (17) *satisfy for $j, k = 1, \ldots, K$, $k \neq j$*

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}(Y_t - B_n X_t)) - \phi_k(A_{n,k\bullet}(Y_t - B_n X_t))\right] W_{n,t} = o_{P^n_{\tilde{\theta}_n}}(n^{-1/2}), \tag{32}$$

where $A_n \coloneqq A(\alpha_n, \tilde{\beta}_n)$ and $B_n \coloneqq B(\tilde{\beta}_n)$ and 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

$$\frac{1}{n}\sum_{t=1}^{n}\left(\left[\hat{\phi}_{k,n}(A_{n,k\bullet}(Y_t - B_n X_t)) - \phi_k(A_{n,k\bullet}(Y_t - B_n X_t))\right] W_{n,t}\right)^2 = o_{P^n_{\theta'_n}}(\nu_n). \tag{33}$$

where $W_{n,t}$ are any random variables independent from all $A_{n,k\bullet}(Y_s - c_n - B_n X_s)$ with $s > t$ and such that $\sup_{n\in\mathbb{N}, 1\leq t\leq n} \mathbb{E}_{\tilde{\theta}_n} W_{n,t}^2 < \infty$ and $\frac{1}{n}\sum_{t=1}^{n} W_{n,t}^2 - \mathbb{E}_{\tilde{\theta}_n} W_{n,t}^2 \xrightarrow{P^n_{\tilde{\theta}_n}} 0$.

Proof of Lemma A.1. The proof follows by an argument analogous to that used to prove Lemma 3 of Lee and Mesters (2022a); see Lee and Mesters (2022b) for the proof. □

A.2 Main proofs

Proof of Proposition 4.1. Throughout we work conditional on $(Y_{-p+1}, \ldots, Y_0)'$. Define

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

where $A_n \coloneqq A(\alpha_n, \sigma_n)$, $\mathcal{F}_{n,t} \coloneqq \sigma(Y_t, X_t)$, $\mathcal{F}_n \coloneqq \mathcal{F}_{n,n}$ and note that $(W_{n,t}, \mathcal{F}_{n,t})_{n\in\mathbb{N},\, t\in[n]}$ forms an adapted stochastic process. Moreover it is clear that given assumption 3.1(ii),

$$\mathbb{E}\left[W_{n,t}|\mathcal{F}_{n,t-1}\right] = \frac{1}{2\sqrt{n}}\left[c'\,\mathbb{E}\left[\dot{\ell}_\theta(Y_t, X_t)|\mathcal{F}_{n,t-1}\right] + \sum_{k=1}^{K}\mathbb{E}[h_k(A_{n,k\bullet}V_{\theta_n,t})|\mathcal{F}_{n,t-1}]\right] = 0, \tag{34}$$

almost surely, where the expectation is taken under $P^n_{\theta_n}$.

Next define $Z_{n,t} \coloneqq (z_{n,t}/z_{n,t-1})^{1/2} - 1$ where $z_{n,0} = 1$ and else

$$z_{n,j} \coloneqq \left(\frac{|\tilde{A}_n|}{|A_n|}\right)^j \times \prod_{t=1}^{j}\prod_{k=1}^{K}\frac{\eta_k(\tilde{A}_{n,k\bullet}\tilde{V}_{n,t})}{\eta_k(A_{n,k\bullet}V_{n,t})}\left(1 + h_{n,k}(\tilde{A}_{n,k\bullet}\tilde{V}_{n,t})/\sqrt{n}\right),$$

i.e.,

$$Z_{n,t} \coloneqq \left[\frac{|\tilde{A}_n|}{|A_n|}\prod_{k=1}^{K}\frac{\eta_k(\tilde{A}_{n,k\bullet}\tilde{V}_{n,t})}{\eta_k(A_{n,k\bullet}V_{n,t})}\left(1 + h_{n,k}(\tilde{A}_{n,k\bullet}\tilde{V}_{n,t})/\sqrt{n}\right)\right]^{1/2} - 1.$$

We now verify conditions (S2) – (S6) of Theorem 2.1.2 in Taniguchi and Kakizawa (2000), having shown (S1) to hold above. (S2), i.e. that $\mathbb{E}\sum_{t=1}^{n}[W_{n,t} - Z_{n,t}]^2 \to 0$, where the expectation is taken under $P^n_{\theta_n}$ is shown to hold in Lemma A.3 below. (S3) – (S6) follow from Lemmas A.7 and A.8. (S3) follows immediately from Lemma A.7; (S5) follows from Lemma A.8 by Markov's inequality. For (S4), use the uniform integrability given by Lemma A.7 and Markov's inequality to obtain that for any $\varepsilon > 0$, as $n \to \infty$

$$\begin{aligned} P^n_{\theta_n}\left(\max_{1\leq t\leq n}|W_{n,t}| > \varepsilon\right) &\leq P^n_{\theta_n}\left(\sum_{t=1}^n W^2_{n,t}\mathbf{1}\{|W_{n,t}|>\varepsilon\} > \varepsilon^2\right) \\ &\leq \varepsilon^{-2}\frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[nW^2_{n,t}\mathbf{1}\{\sqrt{n}|W_{n,t}| > \varepsilon\sqrt{n}\}\right] \\ &\to 0. \end{aligned}$$

For (S6), note that the same UI argument as just used yields that

$$\lim_{n\to\infty}\sum_{t=1}^n \mathbb{E}\left[W^2_{n,t}\mathbf{1}\{|W_{n,t}|>\delta\}\right] = 0,$$

for some $\delta > 0$ and hence as conditional expectations are contractions in L_1,

$$\lim_{n\to\infty}\mathbb{E}\left|\sum_{t=1}^n \mathbb{E}\left[W^2_{n,t}\mathbf{1}\{|W_{n,t}|>\delta\}|\mathcal{F}_{n,t-1}\right]\right| = 0,$$

implying (S6). (L3) of Theorem 2.1.1 in Taniguchi and Kakizawa (2000) holds since the relevant measures are both absolutely continuous with respect to Lebesgue measure (cf. Taniguchi and Kakizawa, 2000, p. 34). By Theorem 2.1.2 of Taniguchi and Kakizawa (2000), under $P^n_{\theta_n}$:

$$\Lambda^n_{\tilde{\theta}_n/\theta_n}(Y^n) \rightsquigarrow \mathcal{N}(-\tau^2/2, \tau^2). \tag{35}$$

In view of Lemma A.8 and (S1) we have that $\Psi_\theta(c,h) := \lim_{n\to\infty}\mathbb{E}\left[\mathrm{g}_n(Y^n)^2\right] = \tau^2$ (in which the dependence on c, h is notationally supressed on the right hand side). Let $\varepsilon \in (0,1)$ be fixed and define $E_n := \{\max_{1\leq t\leq n}|Z_{n,t}| \leq \varepsilon\}$ and note that by Theorem 2.1.2 of Taniguchi and Kakizawa (2000) $P^n_{\theta_n}E_n \to 1$. By Taylor expansion of $\log(1+x)$, on E_n we have

$$\log(1+Z_{n,t}) = Z_{n,t} - \frac{1}{2}Z^2_{n,t} + Z^2_{n,t}R(Z_{n,t}),$$

where $R(x) \leq M|x|$ for some $M \in [0,\infty)$ and so by (S2), on E_n

$$\begin{aligned} \Lambda^n_{\tilde{\theta}_n/\theta_n}(Y^n) &= 2\sum_{t=1}^n \log(Z_{n,t}+1) \\ &= \sum_{t=1}^n 2Z_{n,t} - \frac{1}{2}\sum_{t=1}^n 2Z^2_{n,t} + \sum_{t=1}^n Z^2_{n,t}R(Z_{n,t}). \end{aligned}$$

Moreover, by Theorem 2.1.2 of Taniguchi and Kakizawa (2000),

$$\sum_{t=1}^n Z^2_{n,t}R(Z_{n,t}) \leq M\max_{1\leq t\leq n}|Z_{n,t}|\sum_{t=1}^n W^2_{n,t} = o_{P^n_{\theta_n}}(1),$$

and so using also Lemma A.4

$$\Lambda^n_{\tilde{\theta}_n/\theta_n}(Y^n) = \sum_{t=1}^n 2W_{n,t} - \tau^2/4 - \frac{1}{2}\sum_{t=1}^n 2W^2_{n,t} + o_{P^n_{\theta_n}}(1).$$

Lemma A.8, comparison of $W_{n,t}$ and $\mathsf{g}_n(Y^n)$ and the fact that the above display holds with $P^n_{\theta_n}$–probability approaching 1 yields the asymptotic expansion (10). The weak convergence of $\mathsf{g}_n(Y^n)$ follows by combining (10), (35) and (S5). □

Proof of Corollary 4.1. Combine (35) with Example 6.5 in van der Vaart (1998). □

Proof of Lemma 4.1. Define

$$\mathcal{T}^{\eta|\gamma}_{P_\theta,H} := \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,t}) : h = (h_1,\ldots,h_K) \in \dot{\mathscr{H}} \right\}, \quad V_{\theta,t} := Y_t - B_\theta X_t. \tag{36}$$

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) under P^n_θ

$$\mathbb{E}\left[\left(\dot{\ell}_\theta(Y_s,X_s) - \tilde{\ell}_\theta(Y_s,X_s)\right)\sum_{t=1}^{n}\sum_{k=1}^{K} h_k(A_{k\bullet}V_{\theta,t})\right] = 0 \quad \text{for all } h \in \dot{\mathscr{H}}.$$

For (a), the fact that $\tilde{\ell}_\theta(Y_s,X_s) \in L_2(P^n_\theta)$ follows straightfowardly from its form and the moment conditions in assumption 3.1(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 obtain under P^n_θ if we have that 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 $\mathbb{E}\left[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})\right] = \mathbb{E}\left[(X_s-\mu)\phi_m(\epsilon_{m,s})\right]\mathbb{E}\left[h_k(\epsilon_{k,t})\right] = 0$. If $s > t$, then $\mathbb{E}\left[(X_s-\mu)\phi_m(\epsilon_{m,s})h_k(\epsilon_{k,t})\right] = \mathbb{E}\left[(X_s-\mu)h_k(\epsilon_{k,t})\,\mathbb{E}\left[\phi_m(\epsilon_{m,s})|\sigma(\epsilon_1,\ldots,\epsilon_{s-1})\right]\right] = 0$ again by independence.

For (b), that $\dot{\ell}_\theta(Y_s,X_s) - \tilde{\ell}_\theta(Y_s,X_s) \in L_2(P^n_\theta)$ follows from $\tilde{\ell}_\theta(Y_s,X_s) \in L_2(P^n_\theta)$ (as noted above) and Lemma A.7. Note that for any $h \in \dot{\mathscr{H}}$, $1 \le s \le n$,

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

will obtain under P^n_θ if we have that for any $m \in [K]$, $1 \le t \le n$, $1 \le s \le n$ and

$$\begin{aligned}
&\mathbb{E}\left[(\phi_m(\epsilon_{m,s})\epsilon_{m,s} + 1 - \tau_{m,1}\epsilon_{m,s} - \tau_{m,2}\kappa(\epsilon_{m,s}))\sum_{k=1}^{K} h_k(\epsilon_{k,t})\right] = 0\\
&\mathbb{E}\left[(\phi_m(\epsilon_{m,s}) + \varsigma_{m,1}\epsilon_{m,s} + \varsigma_{m,2}\kappa(\epsilon_{m,s}))\sum_{k=1}^{K} h_k(\epsilon_{k,t})\right] = 0.
\end{aligned}$$

If $s \neq t$, both terms follows by independence (over t) of $(\epsilon_t)_{t\ge1}$ and the definition of $\dot{\mathscr{H}}$. If $s = t$ the first term follows from the fact that the projection of $\phi_m(\epsilon_{m,t})\epsilon_{k,t} + 1$ on $[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}]^\perp$ is

$\tau_{k,1}\epsilon_{k,t} + \tau_{k,2}\kappa(\epsilon_{k,t})$ as follows from the analogous result in the proof of Lemma 2 of Lee and Mesters (2022a).[30] For the second term, if $s \neq t$, then this follows by independence (over t) of $(\epsilon_t)_{t\geq 1}$ and the definition of $\dot{\mathscr{H}}$. If $s = t$, then define $q(e) \coloneqq \phi_m(e) + \varsigma_{m,1}e + \varsigma_{m,2}\kappa(e)$. $q(\epsilon_m)$ belongs to $\operatorname{cl}\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ as $q(\epsilon_{m,t}) \in L_2(P^n_\theta)$ and the choice of ς ensures that

$$\mathbb{E}[q(\epsilon_{m,t})] = \mathbb{E}[q(\epsilon_{m,t})\epsilon_{m,t}] = \mathbb{E}[q(\epsilon_{m,t})\kappa(\epsilon_{m,t})] = 0,$$

as is easily verified.[31] Define also $r(e) \coloneqq \varsigma_{m,1}e + \varsigma_{m,2}\kappa(e)$. Then, by definition of $\dot{\mathscr{H}}$ we have that $r(\epsilon_{m,t}) \in [\mathcal{T}^{\eta|\gamma}_{P_\theta,H}]^\perp$. Hence we can write

$$\phi_m(\epsilon_{m,t}) = q(\epsilon_{m,t}) - r(\epsilon_{m,t})$$

where the first right hand side term belongs to $\operatorname{cl}\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ and the second to its orthogonal complement. Therefore, by e.g. Theorem 4.11 of Rudin (1987), $-r(\epsilon)_{m,t}$ is the orthogonal projection of $\phi_m(\epsilon_{m,t})$ onto $[\mathcal{T}^{\eta|\gamma}_{P_\theta,H}]^\perp$ which implies that $\mathbb{E}\left[(\phi_m(\epsilon_{m,t}) - (-r(\epsilon_{m,t})))\sum_{k=1}^K h_k(\epsilon_{k,t})\right] = 0$. $\square$

Proof of Theorem 5.1. Define

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

where $\gamma_\star \coloneqq (\alpha_n, \beta_\star)$ and $\theta_\star \coloneqq (\gamma_\star, \eta)$. We show that we have

$$R_{n,i}(\tilde{\theta}_n) \xrightarrow{P^n_{\tilde{\theta}_n}} 0 \quad \text{for } i = 1, 2, 3. \tag{37}$$

Define $b_n \coloneqq \sqrt{n}(\beta'_n - \beta)$. We may assume without loss of generality that $b_n \to b$ and $h_n \to h$.[32]

Let Q_n denote the law of $(Y_t)_{t=1}^n$ corresponding to $\tilde{\theta}_n \coloneqq (\alpha_n, \beta + b_n/\sqrt{n}, \eta(1 + h_n/\sqrt{n}))$ and P_n that corresponding to $\check{\theta}_n \coloneqq (\alpha_n, \beta + b_n/\sqrt{n}, \eta)$ (both conditional on the initial observations). By Corollary 4.1 $Q_n \lhd\rhd P_n$ and hence (37) follows by Lemma A.10 and Le Cam's first Lemma (e.g. van der Vaart, 1998, Lemma 6.4).

Next we show that (37) continues to hold if the argument of the remainders $R_{n,i}$ is replaced by $\bar{\beta}_n$ as defined in the theorem. Since $\bar{\beta}_n$ remains $\sqrt{n}$-consistent there is an $M > 0$ such that $P^n_{\tilde{\theta}_n}\left(\sqrt{n}\|\bar{\beta}_n - \beta\| > M\right) < \varepsilon$. If $\sqrt{n}\|\bar{\beta}_n - \beta\| \leq M$ then $\bar{\beta}_n$ is equal to one of the values in the finite set $\mathscr{S}^c_n = \{\beta' \in n^{-1/2}C\mathbb{Z}^{L_2} : \|\beta' - \beta\| \leq n^{-1/2}M\}$. For each M this set has finite number of elements bounded independently of n, call this upper bound $\overline{B}$. Letting R_n denote any of $R_{n,1}$, $R_{n,2}$ or $R_{n,3}$ we have that for any $\upsilon > 0$

$$\begin{aligned}
P^n_{\tilde{\theta}_n}\left(\|R_n(\bar{\beta}_n)\| > \upsilon\right) &\leq \varepsilon + \sum_{\beta_n \in \mathscr{S}^c_n} P^n_{\tilde{\theta}_n}\left(\{\|R_n(\beta_n)\| > \upsilon\} \cap \{\bar{\beta}_n = \beta_n\}\right) \\
&\leq \varepsilon + \sum_{\beta_n \in \mathscr{S}^c_n} P^n_{\tilde{\theta}_n}\left(\|R_n(\beta_n)\| > \upsilon\right) \\
&\leq \varepsilon + \overline{B}P^n_{\tilde{\theta}_n}(\|R_n(\beta^*_n)\| > \upsilon),
\end{aligned}$$

where $\beta^*_n \in B_n$ maximises $\beta \mapsto P^n_{\tilde{\theta}_n}\left(\|R_n(\beta)\| > \upsilon\right)$. As $(\beta^*_n)_{n\in\mathbb{N}}$ is a deterministic $\sqrt{n}$-consistent

[30] See Lee and Mesters (2022b) for the proof.

[31] That $\operatorname{cl}\mathcal{T}^{\eta|\gamma}_{P_\theta,H}$ is the set of L_2 random variables satisfying these equations can be shown by an argument analogous to that in footnote S5 of Lee and Mesters (2022b).

[32] Otherwise the same argument can proceed along appropriately chosen subsequences.

sequence for β we have that $P^n_{\tilde{\theta}_n}(\|R_n(\beta^*_n)\| > \upsilon) \to 0$ by (37).

It follows that

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

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

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

We combine these to obtain

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

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

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

By (i) of Lemma A.10 and Le Cam's third lemma (e.g. van der Vaart, 1998, Example 6.7)

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

and hence under $P_{\tilde{\theta}_n}$

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

We additionally have

$$\left\|\hat{\mathcal{I}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{I}}_\theta\right\|_2 \leq \left\|\hat{I}_{n,\bar{\gamma}_n,\alpha\alpha} - \tilde{I}_{\theta,\alpha\alpha}\right\|_2 + \left\|\hat{I}_{n,\bar{\gamma}_n,\alpha\beta}\hat{I}^{-1}_{n,\bar{\gamma}_n,\beta\beta}\hat{I}_{n,\bar{\gamma}_n,\beta\alpha} - \tilde{I}_{\theta,\alpha\beta}\tilde{I}^{-1}_{\theta,\beta\beta}\tilde{I}_{\theta,\beta\alpha}\right\|_2.$$

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 equation (37) with $\bar{\gamma}_n$ replacing γ_n we have $P_{\tilde{\theta}_n}\left(\left\|\hat{\mathcal{I}}_{n,\bar{\gamma}_n} - \tilde{\mathcal{I}}_\theta\right\|_2 < \check{\nu}_n\right) \to 1$ where $\check{\nu}_n = C\nu_n$ for some positive constant $C \geq 1$. By Proposition 3.13 and Lemma C.6 of Lee (2022)

$$\hat{\mathcal{I}}^{t,\dagger}_{n,\bar{\gamma}_n} \xrightarrow{P^n_{\tilde{\theta}_n}} \tilde{\mathcal{I}}^\dagger_\theta \qquad \text{and} \qquad P^n_{\tilde{\theta}_n}R_n \to 1,$$

where $R_n := \{\operatorname{rank}(\tilde{\mathcal{I}}^t_{n,\bar{\gamma}_n}) = \operatorname{rank}(\tilde{\mathcal{I}}_\theta)\}$.

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

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

where the distributional result $X := Z'\tilde{\mathcal{I}}^\dagger_\theta Z \sim \chi^2_r$, 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^2_r distribution, which we will call c. Hence, we have $c_n \xrightarrow{P^n_{\tilde{\theta}_n}} c$ and as a result,$\hat{S}^{SR}_{n,\bar{\gamma}_n} - c_n \rightsquigarrow X - c$ where $X \sim \chi^2_r$. Since the χ^2_r distribution is continuous, we have by the Portmanteau theorem

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

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

It remains to handle the case with $r = 0$. We first note that $Z_n \rightsquigarrow Z \sim \mathcal{N}(0, \tilde{\mathcal{I}}_\theta)$ continues to hold by our assumptions, though in this case $\tilde{\mathcal{I}}_\theta$ is the zero matrix and hence the limiting distribution is degenerate: $Z = 0$.

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

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

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

Proof of Corollary 5.1. Apply Theorem 5.1 to conclude:

$$\lim_{n\to\infty} P^n_{\tilde{\theta}_n}(\alpha_n \in \hat{C}_n) \geq 1 - \lim_{n\to\infty} P^n_{\tilde{\theta}_n}(\hat{S}^{SR}_{n,\bar{\gamma}_n} > c_n) \geq 1 - \alpha. \qquad \square$$

Proof of Proposition 6.1. Let G be a convex, compact set with $G \supset \{\gamma_n : n \geq N_0\}$ for some $N_0 \in \mathbb{N}$. Since g is continuously differentiable and G is compact, $\{\|g'_\gamma\| : \gamma \in G\}$ is bounded and hence $\{g'_{\gamma_n} : n \in \mathbb{N}\}$ is uniformly equicontinuous (cf. Remark A.2). By compactness, $\gamma \mapsto g'_\gamma$ is uniformly continuous on G. Combined with the mean-value theorem (e.g. Drabek and Milota, 2007, Theorem 3.2.7) this implies that g is uniformly differentiable along $(\gamma_n)_{n\in\mathbb{N}}$. By Theorem A.2 and the fact that $\mathcal{N}(0, M_n) \xrightarrow{TV} \mathcal{N}(0, M)$ if $M_n \to M \succ 0$,

$$\sqrt{n}\left(g(\alpha_n, \hat{\beta}_n) - g(\alpha_n, \tilde{\beta}_n)\right) \overset{P^n_{\tilde{\theta}_n}}{\rightsquigarrow} \mathcal{N}\left(0, J_\gamma V_\theta \Sigma'\right).$$

This and the fact that $\hat{V}_{n,\alpha} \xrightarrow{P^n_{\tilde{\theta}_n}} J_\gamma \Sigma J'_\gamma \succ 0$ by our hypotheses and the continuous mapping theorem imply that

$$n g(\alpha_n, \hat{\beta}_n)' \hat{V}^{-1}_{n,\alpha} g(\alpha_n, \hat{\beta}_n) \rightsquigarrow \chi^2_{d_g} \quad \text{under } P^n_{\tilde{\theta}_n}.$$

It follows that

$$\lim_{n\to\infty} P^n_{\tilde{\theta}_n}(g(\alpha_n, \tilde{\beta}_n) \in \hat{C}_{n,g,\alpha_n,1-a}) = \lim_{n\to\infty} P^n_{\tilde{\theta}_n}\left(n g(\alpha_n, \hat{\beta}_n)' \hat{V}^{-1}_{n,\alpha} g(\alpha_n, \hat{\beta}_n) \leq c_a\right) = 1 - a. \qquad \square$$

Proof of Corollary 6.1. This follows directly from the hypotheses and the fact that

$$\begin{aligned} P^n_{\tilde{\theta}_n}\left(g(\alpha_n, \tilde{\beta}_n) \in \hat{C}_{n,g}\right) &\geq P^n_{\tilde{\theta}_n}\left(\left\{g(\alpha_n, \hat{\beta}_n) \in \hat{C}_{n,g,\alpha_n,1-q_2}\right\} \cap \left\{\alpha_n \in \hat{C}_{n,1-q_1}\right\}\right) \\ &\geq P^n_{\tilde{\theta}_n}\left(g(\alpha_n, \hat{\beta}_n) \in \hat{C}_{n,g,\alpha_n,1-q_2}\right) + P^n_{\tilde{\theta}_n}\left(\alpha_n \in \hat{C}_{n,1-q_1}\right) - 1. \qquad \square \end{aligned}$$

A.3 Auxilliary results

Here we record results relating to the model under study in the main text, which are used in establishing the main results which are proven above.

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

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

and note that we can write

$$Z_t = \mathsf{C}_\theta + \mathsf{B}_\theta Z_{t-1} + \mathsf{D}_\theta \epsilon_t. \tag{38}$$

Proposition A.1: *Suppose that assumption 3.1 holds. Then $\Phi := (Z_t)_{t\geq 0}$ (with initial value $Z_0 = z$) is a uniformly ergodic Markov chain on $\mathbb{R}^{Kp}$. Moreover for any compact set $\mathsf{K} \subset \mathbb{R}^{d_\gamma}$, we have that for any (initial value) $z \in \mathbb{R}^{Kp}$,*

$$\sup_{\theta=(\gamma,\eta):\ \gamma\in\mathsf{K}} \|Q_\theta^n(z,\cdot) - \pi_\theta(\cdot)\|_{TV} \leq (M_1 + \|z\|^2)\gamma^n, \quad \text{for some } \gamma < 1,\ M_1 < \infty$$

and π_θ an invariant probability distribution for Φ (under θ) and for $M_2 < \infty$

$$\sup_{\theta=(\gamma,\eta):\ \gamma\in\mathsf{K}} \beta_\theta(n) \leq (4M_1 + 3\|z\|^2 + M_2)\gamma^{\lfloor n/2 \rfloor},$$

where $\beta_\theta(n)$ are the β-mixing coefficients of Φ.

Proof. That Φ is a Markov chain follows from Proposition 11.6 in Kallenberg (2021). Explicit computation of the rank of the controllability matrix (Meyn and Tweedie, 2009, equation (4.13)) demonstrates that the associated linear control model is controllable. Moreover under assumption 3.1, (LSS4) and (LSS5) of Meyn and Tweedie (2009) hold and hence by Proposition 6.3.5 in Meyn and Tweedie (2009), Φ is a ψ-irreducible T-chain and every compact subset is a small set. Aperiodicity of Φ follows from the assumptions on the densities.

The 1-step transition probability is given by the density on $\mathbb{R}^{Kp} \times \mathbb{R}^{Kp}$ defined as

$$q_\theta(y,x) := |A_\theta| \prod_{k=1}^{K} \eta_k(A_{\theta,k} V_\theta), \quad V_\theta := y_1 - c_\theta - \sum_{l=1}^{p} B_{\theta,l} x_l,$$

where e.g. y_1 denotes the first K elements of y and similarly for x. By assumption 3.1, the map $(\gamma, y, x) \mapsto q_{(\gamma,\eta)}(y,x)$ is continuous and positive everywhere on $\Gamma \times \mathbb{R}^{Kp} \times \mathbb{R}^{Kp}$. For any compact $B \subset \mathbb{R}^{Kp}$ put $\varepsilon := \int \inf_{(\gamma,x)\in K\times B} q_{(\gamma,\eta)}(y,x)\,\mathrm{d}y$ and $\rho(y) := \inf_{(\gamma,x)\in K\times B} q_{(\gamma,\eta)}(y,x)/\varepsilon$.[33] Then for any $A \in \mathcal{B}(\mathbb{R}^{Kp})$ and any $x \in B$,

$$\int_A q_\theta(y,x)\,\mathrm{d}y \geq \varepsilon \int_A \rho(y)\,\mathrm{d}y.$$

Under assumption 3.1 the eigenvalues of B_θ are bounded above by some $\overline{\rho} < 1$ for all $\theta \in \mathsf{T} := \{(\gamma,\eta) : \gamma \in \mathsf{K}\}$. Using this and the Gelfand formula (e.g. Horn and Johnson, 2013, Corollary 5.6.14) there exists a $\rho_\star < 1$ with $\|\mathsf{B}_\theta^n\| \leq \rho_\star^n$ on T. Since we can re-write (38) as

$$Z_t - m_\theta = \mathsf{B}_\theta(Z_{t-1} - m_\theta) + \mathsf{D}_\theta \epsilon_t, \tag{39}$$

[33] Note that $\epsilon > 0$ by the positivity and continuity.

with $m_\theta \coloneqq \left(\sum_{i=0}^{\infty} \mathsf{B}_\theta^i\right) \mathsf{C}_\theta$, we have

$$\mathsf{V}_\theta(Z_t) \coloneqq \|\mathsf{B}_\theta(Z_{t-1} - m_\theta)\|^2 + \|\mathsf{D}_\theta \epsilon_t\|^2 + 2[\mathsf{B}_\theta(Z_{t-1} - m_\theta)]' \mathsf{D}_\theta \epsilon_t + 1,$$

and since ϵ_t is independent of Z_{t-1}, and $\|\mathsf{D}_\theta\| \le D_\star < \infty$ on T,

$$\mathbb{E}[\mathsf{V}_\theta(Z_t)|Z_{t-1}] \le \rho_\star^2 \|Z_{t-1} - m_\theta\|^2 + D_\star^2 \le \rho_\star^2 \mathsf{V}_\theta(Z_{t-1}) + D_\star^2.$$

This, in conjunction with Proposition 5.5.3 and Lemmas 15.2.8 of Meyn and Tweedie (2009) establishes that the Markov chain satisfies the drift condition (10) in Roberts and Rosenthal (2004) with $\lambda = (1+\rho_\star^2)/2 < 1$, $b = D_\star^2 < \infty$ and $C = C_\theta = \{z : \mathsf{V}_\theta(z) \le 2D_\star^2/(1-\rho_\star^2)\}$. By Proposition 11 in Roberts and Rosenthal (2004) their bivariate drift condition (11) is satisfied with $h(x,y) = [\mathsf{V}_\theta(x) + \mathsf{V}_\theta(y)]/2$ and $\alpha^{-1} = \lambda + b/(d+1) < 1$. Moreover $b_{0,\theta} \coloneqq \max\{1, \alpha(1-\varepsilon) \sup_{(x,y)\in C_\theta \times C_\theta} \bar{R}_\theta h_\theta(x,y)\}$ is bounded above by $(1-\varepsilon)D_\star^2/(1-\rho_\star^2) < \infty$, where $\bar{R}_\theta h_\theta(x,y)$ is defined analogously to $\bar{R}h(x,y)$ on p. 41 of Roberts and Rosenthal (2004). By Theorem 16.0.2 of Meyn and Tweedie (2009) there exists an invariant π_θ with

$$\|Q_\theta^n(z,\cdot) - \pi_\theta\| \le R r^{-n}, \quad R < \infty,\ r > 1,$$

where $Q_\theta(z,\cdot)$ is the transition probability. That is, Φ is uniformly ergodic.

For the second claim, by Theorem 12 in Roberts and Rosenthal (2004) we have that for any (initial) $z \in \mathbb{R}^{Kp}$ and some $\gamma < 1$, for all $\theta \in \mathsf{T}$,

$$\|Q_\theta^n(z,\cdot) - \pi_\theta\|_{TV} \le (M_1 + \|z\|^2)\gamma^n,$$

where[34]

$$M_1 = 1 + \sup_{\theta\in\mathsf{T}} \|m_\theta\|^2 + \sup_{\theta\in\mathsf{T}} \int \|z - m_\theta\|^2 \, \mathrm{d}\pi_\theta(z) < \infty.$$

The claim regarding the β-mixing coefficients then follows directly from Proposition 3 in Liebscher (2005), with $M_2 \coloneqq \sup_{\theta\in\mathsf{T}} \int \|z\|^2 \, \mathrm{d}\pi_\theta(z) < \infty$.[35] □

Lemma A.2: *Suppose that assumption 3.1 holds. Define $U_{\theta,t}$ as the (unique, strictly) stationary solution to* (38) *(under θ). Then $U_{\theta,t}$ has the representation*

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

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

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

Proof. Rewriting (38) as (39) and applying Theorem 11.3.1 in Brockwell and Davis (1991) yields the first part. For the second part, let $\mathcal{U}_\theta^* \mathcal{J}_\theta \mathcal{U}_\theta$ be a Schur decomposition of B_θ. Then

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

[34] That the first supremum is finite is clear since $m_\theta = (I - \mathsf{B}_\theta)^{-1}\mathsf{C}_\theta$ which is evidently continuous. For the second supremum note that the integral is taking an expectation with respect to the distribution of the stationary solution of a VAR model. This is bounded uniformly over $\theta \in \mathsf{T}$ by Lemma A.2, the fact that $\|\mathsf{D}_\theta\|$ is uniformly bounded on T and the observation that since $M \mapsto \rho(M)$ is continuous and T is compact, there is a ρ and υ with $\rho + \upsilon < 1$ such that $\rho \ge \rho(\mathsf{B}_\theta)$ for all $\theta \in \mathsf{T}$.

[35] The uniform boundedness of M_2 follows by an analogous argument as given in footnote 34.

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

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

Corollary A.1: *Suppose that assumption 3.1 holds and $\theta_n = (\gamma_n, \eta) \to (\gamma, \eta) = \theta$. Define π_θ as in Proposition A.1 and let $\mathscr{G}_{\theta_n,n}$ be the measure corresponding to the density $\frac{1}{n}\sum_{t=1}^{n} \rho_{\theta_n,t}$ where $\rho_{\theta_n,t}$ is the density of the non-deterministic parts of X_t under $P_{\theta_n}^n$ $(1 \leq t \leq n)$. Then $\mathscr{G}_{\theta_n,n} \rightsquigarrow \pi_\theta$.*

Proof. By Proposition A.1, $\mathscr{G}_{\theta,n} \xrightarrow{TV} \pi_\theta$ uniformly on $\mathsf{T} := \{\theta_n : n \in \mathbb{N}\} \cup \{\theta\}$. We also have that $\pi_{\theta_n} \rightsquigarrow \pi_\theta$. To see this, use the representation in Lemma A.2 and the fact that we can uniformly bound $\|\mathsf{B}_\vartheta^j\|$ and $\|\mathsf{D}_\vartheta\|$ for $\vartheta \in \mathsf{T}$ and $j \in \mathbb{N}$ to obtain

$$\begin{aligned}
\mathbb{E}\,\|U_{\theta_n,t} - U_{\theta,t}\| &\leq \|m_{\theta_n} - m_\theta\| + \mathbb{E}\left\|\sum_{j=0}^{\infty} \mathsf{B}_{\theta_n}^j \mathsf{D}_{\theta_n} \epsilon_{t-j} - \mathsf{B}_\theta^j \mathsf{D}_\theta \epsilon_{t-j}\right\| \\
&= o(1) + \mathbb{E}\,\|\epsilon_t\| \sum_{j=0}^{\infty} \left(\|\mathsf{B}_{\theta_n}^j\|\|\mathsf{D}_{\theta_n} - \mathsf{D}_\theta\| + \|\mathsf{D}_\theta\|\|\mathsf{B}_{\theta_n}^j - \mathsf{B}_\theta^j\|\right) \\
&= o(1)
\end{aligned}$$

where the second equality uses the fact the ϵ_t are identically distributed and the third equality uses the dominated convergence theorem.[36] This implies that $U_{\theta_n,t} \rightsquigarrow U_{\theta,t}$ as $n \to \infty$, i.e. $\pi_{\theta_n} \rightsquigarrow \pi_\theta$. Combination of these results yields the claim. $\square$

Lemma A.3 (UDQM): *Suppose that assumption 3.1 holds. Then, with $W_{n,t}$ and $Z_{n,t}$ defined as in the proof of Proposition 4.1,*

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

Proof. Write $Y_{n,t}$ and $X_{n,t}$ for random elements which have the same law as $Y_t,\ X_t$ (respectively) under $P_{\theta_n}^n$. Recall $V_{n,t} := Y_{n,t} - BX_{n,t}$ and define

$$q_\theta(Y_{n,t}, X_{n,t}) := |A| \prod_{k=1}^{K} \eta_k(A_{k\bullet} V_t), \qquad g_\theta(Y_{n,t}, X_{n,t}) := c'\dot{\ell}_\theta(Y_{n,t}, X_{n,t}) + \sum_{k=1}^{K} h_k(A_{k\bullet} V_{n,t}). \quad (40)$$

Let $\varphi(u) = (c, \eta_1 h_1, \ldots, \eta_K h_K)$ for $u = (c, h)$ with $c \in \mathbb{R}^{L_\alpha + L_\beta}$, $h \in \mathscr{H}$. We initially suppose that $\theta_n = \theta$ for all $n \in \mathbb{N}$ and argue similarly to Lemma 7.6 in van der Vaart (1998). By Assumption 3.1 and standard computations, the derivative of $s \mapsto \sqrt{q_{\theta + s\varphi(u)}}$ at $s = \mathsf{s}$ is $\frac{1}{2} g_{\theta + \mathsf{s}\varphi(u)} \sqrt{q_{\theta + \mathsf{s}\varphi(u)}}$ (everywhere). Inspection reveals that this is continuous in s. Let $\rho_{\theta,t}$ be as defined in Corollary A.1. Define

$$I_{\theta,t} := \int g_\theta^2 q_\theta \rho_{\theta,t}\, \mathrm{d}\lambda.$$

By the mean-value theorem and Jensen's inequality we can write

$$\int \left(\frac{\sqrt{q_{\vartheta_{1,n}}} - \sqrt{q_\theta}}{1/\sqrt{n}}\right)^2 \rho_{\theta,t}\, \mathrm{d}\lambda \leq \frac{1}{4} \int \int_0^1 (g_{\vartheta_{v,n}} \sqrt{q_{\vartheta_{v,n}}})^2 \rho_{\theta,t}\, \mathrm{d}v\, \mathrm{d}\lambda = \frac{1}{4} \int_0^1 I_{\vartheta_{v,n},t}\, \mathrm{d}v \quad (41)$$

[36]Note that $\|\mathsf{B}_{\theta_n}^j - \mathsf{B}_\theta^j\| \to 0$ pointwise in j and is dominated by $2\rho_\star^j$ where $\rho_\star < 1$ is a uniform upper bound on $\|B_\vartheta\|$ for $\vartheta \in \mathsf{T}$ and $\sum_{j=0}^{\infty} 2\rho_\star^j = 2\sum_{j=0}^{\infty} \rho_\star^j < \infty$.

where $\vartheta_{v,n} \coloneqq \theta + \frac{v}{\sqrt{n}}\varphi(u)$ and the last step follows by Tonelli's theorem.

It is shown in Lemma A.5 that as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\int_0^1 I_{\vartheta_{v,n},t}\,\mathrm{d}v = \int_0^1\int g^2_{\vartheta_{v,n}}\,\mathrm{d}G_{\theta_{v,n},n}\,\mathrm{d}v \to \int g_\theta^2\,\mathrm{d}G_\theta < \infty, \tag{42}$$

where $G_{\theta,n}$ is as defined in Lemma A.9. Using this, we can re-write

$$\sum_{t=1}^{n}\int\left(\sqrt{q_{\vartheta_{1,n}}} - \sqrt{q_\theta} - \frac{1}{2\sqrt{n}}g_\theta\sqrt{q_\theta}\right)^2 p_{\theta,t}\,\mathrm{d}\lambda = \int\left(\sqrt{n}\left[\frac{\sqrt{q_{\vartheta_{1,n}}}}{\sqrt{q_\theta}} - 1\right] - \frac{1}{2}g_\theta\right)^2\mathrm{d}G_{\theta,n}. \tag{43}$$

By the assumed differentiability, the integrand in the last integral converges pointwise to zero. Combining this with (41), (42) and (43) with Proposition A.2 we have

$$\lim_{n\to\infty}\int\left(\sqrt{n}\left[\sqrt{q_{\vartheta_{1,n}}} - \sqrt{q_\theta}\right] - \frac{1}{2}\Delta_\theta(u)\sqrt{q_\theta}\right)^2\bar{\rho}_{\theta,n}\,\mathrm{d}\lambda = 0, \tag{44}$$

where $\bar{\rho}_{\theta,n} \coloneqq \frac{1}{n}\sum_{t=1}^{n}\rho_{\theta,t}$ and $\Delta_\theta(u) \coloneqq g_\theta$, to emphasise the linearity in u of g_θ. We next show that any $u_n \to u$, $\mathsf{u}_n \to \mathsf{u}$ (all in U), and any $(v_n)_{n\in\mathbb{N}} \subset [0,\infty)$ with $v_n \downarrow 0$,

$$\lim_{n\to\infty}\int\left[\Delta_{\theta_n+v_n\varphi(\mathsf{u}_n)}(u_n)\sqrt{q_{\theta_n+v_n\varphi(\mathsf{u}_n)}} - \Delta_{\theta_n}(u)\sqrt{q_{\theta_n}}\right]^2\bar{\rho}_{\theta_n,n}\,\mathrm{d}\lambda = 0. \tag{45}$$

We first note that for any (deterministic) convergent sequence $x_n \to x$, we have

$$[\Delta_{\theta_n+v_n\varphi(\mathsf{u}_n)}(u_n)\sqrt{q_{\theta_n+v_n\varphi(\mathsf{u}_n)}}](\cdot,x_n) - [\Delta_\theta(u)\sqrt{q_\theta}](\cdot,x) \to 0,$$

pontwise in y. This follows by the continuity of the relevant functions and that, for $\check{\vartheta}_n \coloneqq \theta_n + v_n\varphi(\mathsf{u}_n)$, (i)

$$(y - B_{\check{\vartheta}_n}x_n) - (y - B_\theta x) = B_{\check{\vartheta}_n}(x_n - x) + (B_{\check{\vartheta}_n} - B_\theta)x \to 0,$$

since $\vartheta \mapsto B_\vartheta$ is continuous and (ii), since $\vartheta \mapsto A_\vartheta$ is continuous,

$$A_{\check{\vartheta}_n,k\bullet}D_{b_l}x_n - A_{\theta,k\bullet}D_{b_l}x = A_{\check{\vartheta}_n,k\bullet}D_{b_l}(x_n - x) + (A_{\check{\vartheta}_n,k\bullet} - A_{\theta,k\bullet})D_{b_l}x \to 0.$$

The form of $\dot{\ell}_{\check{\vartheta}_n}$ is the same as that given in (7) – (9) once each ϕ_k is replaced by

$$\tilde{\phi}_{k,n} \coloneqq \phi_k + \frac{v_n h_k/\sqrt{n}}{1 + v_n h_k/\sqrt{n}}, \tag{46}$$

and, moreover, since $\check{\vartheta}_n \to \theta$, the continuity and continuous differentiability conditions in assumption 3.1 ensure that all non-random terms in the expressions (7) – (9) converge and are thus bounded.[37] Noting this and directly integrating, it follows that

$$\lim_{n\to\infty}\int\left([\Delta_{\theta_n+v_n\varphi(\mathsf{u}_n)}(u_n)\sqrt{q_{\theta_n+v_n\varphi(\mathsf{u}_n)}}](y,x_n)\right)^2\mathrm{d}y = \int([\Delta_\theta(u)\sqrt{q_\theta}](y,x))^2\,\mathrm{d}y < \infty,$$

and hence by Proposition 2.29 in van der Vaart (1998),

$$\int\left([\Delta_{\theta_n+v_n\varphi(\mathsf{u}_n)}(u_n)\sqrt{q_{\theta_n+v_n\varphi(\mathsf{u}_n)}}](y,x_n) - \Delta_\theta(u)\sqrt{q_\theta}(y,x)\right)^2\mathrm{d}y \to 0.$$

[37] Cf. footnote 43.

Taking $v_n = 0$, $u_n = u$ and $\theta_n = \theta$ in the above yields also

$$\int ([\Delta_\theta(u_n)\sqrt{q_\theta}](y, x_n) - \Delta_\theta(u)\sqrt{q_\theta}(y, x))^2 \,\mathrm{d}y \to 0,$$

and hence we have that

$$\mathscr{Q}_n(x) := \int \Big([\Delta_{\theta_n + v_n\varphi(\mathsf{u}_n)}(u_n)\sqrt{q_{\theta_n + v_n\varphi(\mathsf{u}_n)}}](y, x) - [\Delta_\theta(u_n)\sqrt{q_\theta}](y, x)\Big)^2 \,\mathrm{d}y$$

converges continuously to 0. Using the form given in (54) for the (non-deterministic) parts of X_t and noting (as discussed following (54)) that $\{\rho(\mathsf{B}_\vartheta) : \vartheta \in \{\theta_n : n \in \mathbb{N}\} \cup \theta\}$ is bounded, and similarly that $\{\|A_\vartheta^{-1}\| : \vartheta \in \{\theta_n : n \in \mathbb{N}\} \cup \{\theta\}\}$ is bounded, it is easy to see that $\sup_{n\in\mathbb{N}, 1\le t\le n} \mathbb{E}\,\|X_t\| < \infty$. Hence by Markov's inequality for any $\varepsilon > 0$, there is an M such that $\sup_{n\in\mathbb{N},1\le t\le n} P^n_{\theta_n}(\|X_t\| \le M) \ge 1 - \epsilon$ and so the family $\{X_{n,t} : n \in \mathbb{N}, 1 \le t \le n\}$ is uniformly tight, where each $X_{n,t}$ is a random variable (defined on a common probability space) with law $\mathscr{L}(X_t|P^n_{\theta_n})$. Let $(t_n)_{n\in\mathbb{N}}$ be an arbitrary sequence of positive integers satisfying $t_n \le n$ and put $\tilde{X}_n := X_{n,t_n}$. The sequence $(\tilde{X}_n)_{n\in\mathbb{N}}$ is uniformly tight. It follows by Prohorov's theorem that any subsequence $(\tilde{X}_{k_n})_{n\in\mathbb{N}}$ contains a further subsequence $(\tilde{X}_{m_n})_{n\in\mathbb{N}}$ with $X_{m_n} \rightsquigarrow X$ for some random variable X. Since $(\mathscr{Q}_n)_{n\in\mathbb{N}}$ is continuously convergent to the zero function, it follows by the extended continuous mapping theorem (van der Vaart and Wellner, 1996, Theorem 1.11.1) that $\mathscr{Q}_{m_n}(\tilde{X}_{m_n}) \rightsquigarrow 0$. Equation (45) will then follow provided we show that $(\mathscr{Q}_n(\tilde{X}_n))_{n\in\mathbb{N}}$ is uniformly integrable. For this, dominate the n-th term by

$$\begin{aligned}\mathscr{Q}_n(\tilde{X}_n) &\le 2\left[\int [\Delta_{\check{\vartheta}_n}(u_n)\sqrt{q_{\check{\vartheta}_n}}](y, \tilde{X}_n)^2 \,\mathrm{d}y + \int [\Delta_\theta(u)\sqrt{q_\theta}](y, \tilde{X}_n)^2 \,\mathrm{d}y\right] \\ &= 2\left[\mathbb{E}\left[\Delta_{\check{\vartheta}_n}(u_n)(\tilde{Y}_n, \tilde{X}_n)^2 \middle| \tilde{X}_n\right] + \mathbb{E}\left[\Delta_\theta(u)(\tilde{Y}, \tilde{X}_n)^2 \middle| \tilde{X}_n\right]\right],\end{aligned}$$

where $\tilde{Y}_n$ and $\tilde{Y}$ have laws such that their conditional density given $\tilde{X}_n$ is $q_{\check{\vartheta}_n}$ and q_θ, respectively. Lemma A.7 and (46) ensure that $(\Delta_{\check{\vartheta}_n}(u_n)(\tilde{Y}_n, \tilde{X}_n)^2)_{n\in\mathbb{N}}$ is UI. Combining this with the conditional Jensen inequality and the de la Valée Poussin criterion for uniform integrability (e.g. Bogachev, 2007, Theorem 4.5.9) yields that the first conditional expectation in the preceeding display is UI. That the second conditional expectation is also UI follows similarly.

To complete the proof, first let $\theta \in \Theta$ be arbitrary, $s_n := n^{-1/2}$, $u_n \to u$, and use (44), the mean-value theorem (e.g. Drabek and Milota, 2007, Theorem 3.2.7(i)) and (45) to obtain

$$\begin{aligned}&\left\|\left(\frac{\sqrt{q_{\theta+s_n\varphi(u_n)}} - \sqrt{q_\theta}}{s_n} - \frac{1}{2}\Delta_\theta(u)\sqrt{q_\theta}\right)\sqrt{\bar{\rho}_{\theta,n}}\right\|_{\lambda,2} \\ &\quad\le \left\|\left(\frac{\sqrt{q_{\theta+s_n\varphi(u_n)}} - \sqrt{q_{\theta+s_n\varphi(u)}}}{s_n}\right)\sqrt{\bar{\rho}_{\theta,n}}\right\|_{\lambda,2} \\ &\qquad + \left\|\left(\frac{\sqrt{q_{\theta+s_n\varphi(u)}} - \sqrt{q_\theta}}{s_n} - \frac{1}{2}\Delta_\theta(u)\sqrt{q_\theta}\right)\sqrt{\bar{\rho}_{\theta,n}}\right\|_{\lambda,2} \\ &\quad\le \sup_{\delta\in[0,1]} \left\|\frac{1}{2}\Delta_{\theta+s_n\varphi(u)+s_n\delta\varphi(u_n-u)}(u_n - u)\sqrt{q_{\theta+s_n\varphi(u)+s_n\delta\varphi(u_n-u)}}\sqrt{\bar{\rho}_{\theta,n}}\right\|_{\lambda,2} + o(1) \\ &\quad= o(1).\end{aligned}$$

Now return to our original setting with $\theta_n = (\gamma_n, \eta) \to \theta = (\gamma, \eta)$. By the preceding display, applying the mean-value theorem (e.g. Drabek and Milota, 2007, Theorem 3.2.7(ii)) at each n

gives

$$\sum_{t=1}^{n}\mathbb{E}\left[(W_{n,t}-Z_{n,t})^2\right]=\left\|\left(\frac{\sqrt{q_{\theta_n+s_n\varphi(u_n)}}-\sqrt{q_{\theta_n}}}{s_n}-\frac{1}{2}\Delta_{\theta_n}(u)\sqrt{q_{\theta_n}}\right)\sqrt{\bar{\rho}_{\theta_n,n}}\right\|_{\lambda,2}$$

$$\leq \sup_{\delta\in[0,1]}\left\|\frac{1}{2}\left(\Delta_{\theta_n+\delta s_n\varphi(u_n)}(u_n)\sqrt{q_{\theta_n+\delta s_n\varphi(u_n)}}-\Delta_{\theta_n}(u)\sqrt{q_{\theta_n}}\right)\sqrt{\bar{\rho}_{\theta_n,n}}\right\|_{\lambda,2}$$

$$=o(1),$$

where the convergence in the last line is due to (45). □

Lemma A.4: *In the setting of Proposition 4.1, it holds that*

$$2\sum_{t=1}^{n}Z_{n,t}=2\sum_{t=1}^{n}W_{n,t}-\tau^4/2+o_{P^n_{\theta_n}}(1).$$

Proof. Throughout expectations are taken under $P^n_{\theta_n}$. Let $m_n(X_t):=\mathbb{E}[Z_{n,t}|X_t]=\mathbb{E}[Z_{n,t}|\mathcal{F}_{n,t-1}]$ with $\mathcal{F}_{n,t}:=\sigma(\epsilon_i : i=1,\dots,t)$.[38] Form $U_{n,t}:=Z_{n,t}-m_n(X_t)-W_{n,t}$ and note that $(U_{n,t},\mathcal{F}_{n,t})_{n\in\mathbb{N},1\leq t\leq n}$ is a martingale difference array (by (34)). Hence

$$\mathbb{V}\left[\sum_{t=1}^{n}U_{n,t}\right]=\sum_{t=1}^{n}\mathbb{E}\left[Z_{n,t}-W_{n,t}\right]^2+\sum_{t=1}^{n}\mathbb{E}[m_n(X_t)^2]-2\sum_{t=1}^{n}\mathbb{E}\left[(Z_{n,t}-W_{n,t})m_n(X_t)\right].$$

Observe that

$$\mathbb{E}\left[(Z_{n,t}-W_{n,t})m_n(X_t)\right]=\mathbb{E}\left[\mathbb{E}\left[(Z_{n,t}-W_{n,t})m_n(X_t)|X_t\right]\right]=\mathbb{E}\left[m_n(X_t)\,\mathbb{E}\left[Z_{n,t}|X_t\right]\right]=\mathbb{E}\left[m_n(X_t)^2\right],$$

and so by Lemma A.3

$$0\leq\mathbb{V}\left[\sum_{t=1}^{n}U_{n,t}\right]=\sum_{t=1}^{n}\mathbb{E}\left[Z_{n,t}-W_{n,t}\right]^2-\sum_{t=1}^{n}\mathbb{E}[m_n(X_t)^2]\leq\sum_{t=1}^{n}\mathbb{E}\left[Z_{n,t}-W_{n,t}\right]^2\to 0,$$

which, in combination with (L3) of Theorem 2.1.1 in Taniguchi and Kakizawa (2000) (which is noted to hold in the proof of Proposition 4.1), yields

$$2\sum_{t=1}^{n}Z_{n,t}-2\sum_{t=1}^{n}W_{n,t}+\sum_{t=1}^{n}\mathbb{E}[Z_{n,t}^2|\mathcal{F}_{n,t-1}]=o_{P^n_{\theta_n}}(1).$$

It therefore suffices to show that $\sum_{t=1}^{n}\mathbb{E}[Z_{n,t}^2|\mathcal{F}_{n,t-1}]\xrightarrow{P^n_{\theta_n}}\tau^2/4$. For this, first observe that by Lemma A.8,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{2}\Delta_{\theta_n}(u)(Y_t,X_t)\right)^2\xrightarrow{P^n_{\theta_n}}\frac{\tau^2}{4}.$$

Next, since the $\left(\frac{1}{2}\Delta_{\theta_n}(u)\right)^2$ are UI by Lemma A.7, applying Theorem 2.22 in Hall and Heyde (1980), Jensen's inequality for conditional expectations and the de la Vallée Poussin criterion for uniform integrability (e.g. Bogachev, 2007, Theorem 4.5.9) we have that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{2}\Delta_{\theta_n}(u)(Y_t,X_t)\right)^2-\mathbb{E}\left[\left(\frac{1}{2}\Delta_{\theta_n}(u)(Y_t,X_t)\right)^2|\mathcal{F}_{n,t-1}\right]\xrightarrow{L_1}0.$$

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

To complete the proof it therefore suffices show that

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}([\sqrt{n}Z_{n,t}]^2|\mathcal{F}_{n,t-1}) - \mathbb{E}\left[\left(\frac{1}{2}\Delta_{\theta_n}(u)\right)^2|\mathcal{F}_{n,t-1}\right] \xrightarrow{L_1} 0. \tag{47}$$

Since $\mathbb{E}[\mathscr{U}_{n,t}|\mathcal{F}_{n,t-1}] = \mathbb{E}[\mathscr{U}_{n,t}|X_t]$ for $\mathscr{U}_{n,t} \in \left\{[\sqrt{n}Z_{n,t}]^2, \left(\frac{1}{2}\Delta_{\theta_n}(u)(Y_t,X_t)\right)^2\right\}$,[39] we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}([\sqrt{n}Z_{n,t}]^2|\mathcal{F}_{n,t-1}) - \mathbb{E}\left[\left(\frac{1}{2}\Delta_{\theta_n}(u)(Y_t,X_t)\right)^2|\mathcal{F}_{n,t-1}\right]\right| \\
&\quad\leq \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left|\mathbb{E}([\sqrt{n}Z_{n,t}]^2|\mathcal{F}_{n,t-1}) - \mathbb{E}\left[\left(\frac{1}{2}\Delta_{\theta_n}(u)(Y_t,X_t)\right)^2|\mathcal{F}_{n,t-1}\right]\right| \\
&\quad\leq \frac{1}{n}\sum_{t=1}^{n}\int\int\left|\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) - \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right|\left|\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) + \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right| \mathrm{d}y\rho_{\theta_n,t}\,\mathrm{d}x \\
&\quad= \left\langle\left|\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) - \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right|\bar{\rho}_{\theta_n,n}^{1/2}, \left|\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) + \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right|\bar{\rho}_{\theta_n,n}^{1/2}\right\rangle_\lambda \\
&\quad\leq \left\|\left[\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) - \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right]\bar{\rho}_{\theta_n,n}^{1/2}\right\|_{\lambda,2}\left\|\left[\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) + \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right]\bar{\rho}_{\theta_n,n}^{1/2}\right\|_{\lambda,2},
\end{aligned}$$

by Cauchy-Schwarz. The proof of (47) (and hence the Lemma) is completed by applying Lemmas A.3, A.7 and noting

$$\begin{aligned}
&\left\|\left[\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) + \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right]\bar{\rho}_{\theta_n,n}^{1/2}\right\|_{\lambda,2} \\
&\qquad\leq \left(\left\|\left[\sqrt{n}(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}) - \frac{1}{2}\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\right]\bar{\rho}_{\theta_n,n}^{1/2}\right\|_{\lambda,2} + \left\|\Delta_{\theta_n}(u)q_{\theta_n}^{1/2}\bar{\rho}_{\theta_n,n}\right\|_{\lambda,2}\right). \qquad \square
\end{aligned}$$

Lemma A.5: *Suppose that assumption 3.1 holds. Then* (42) *in the proof of Lemma A.3 holds.*

Proof. The finiteness of the integral on the right hand side follows by direct evaluation using the moment bounds in assumption 3.1 along with the fact that under π_θ, $\mathbb{E}\|X_t\|^{4+\delta} < \infty$ which can be seen on combining Lemma A.2 with the fact that π_θ is the law of a stationary solution to the defining VAR equation (see e.g. Kallenberg, 2021, Theorem 11.11).

By Lemma A.9 and Corollary 2.9 in Feinberg et al. (2016) it is enough to prove the uniform $G_{\vartheta_{v_n,n},n}$ – integrability of $(g^2_{\vartheta_{v_n,n}})_{n\in\mathbb{N}}$ for an arbitrary $(v_n)_{n\in\mathbb{N}} \subset [0,1]$. As each h_k is bounded, it suffices to show $\sup_{n\in\mathbb{N}}\int\left|c'\dot{\ell}_{\vartheta_{v_n,n}}\right|^{2+\delta/2}\mathrm{d}G_{\vartheta_{v_n,n},n} < \infty$ for some $\delta > 0$. The form of $\dot{\ell}_{\vartheta_{v_n,n}}$ is the same as that given in equations (7) – (9) once each ϕ_k is replaced by

$$\tilde{\phi}_{k,n} \coloneqq \phi_k + \frac{v_n h_k/\sqrt{n}}{1 + v_n h_k/\sqrt{n}},$$

where, since each h_k is bounded, the second term is bounded for large enough n. Since $\vartheta_{v_n,n} \to \theta$, the continuity and continuous differentiability conditions in assumption 3.1 ensure that all non-random terms in the expressions (7) – (9) converge and are thus bounded.[40] The required bound then follows as, under $G_{\vartheta_{v_n,n},n}$ we have that $V_{\vartheta_{v_n,n},t} \sim \epsilon_t$, with independent components and also independent of X_t, and $\sup_{n\in\mathbb{N}}\mathbb{E}[|\epsilon_t|^{4+\delta}] < \infty$, $\sup_{n\in\mathbb{N}}\mathbb{E}[|\phi_k(\epsilon_t)|^{4+\delta}] < \infty$ and $\sup_{n\in\mathbb{N}}\mathbb{E}\|X_t\|^{4+\delta} < \infty$. The first two moment bounds are immediate from assumption 3.1. The

[39] Cf. footnote 38.
[40] Cf. footnote 43.

latter follows since under each $\rho_{\theta,t}$, $\sup_{n\in\mathbb{N},1\leq t\leq n}\mathbb{E}\|X_t\|^{4+\delta}<\infty$ which follows as in the proof of Lemma A.7 and hence

$$\sup_{n\in\mathbb{N}}\int\|x\|^{4+\delta}\frac{1}{n}\sum_{t=1}^{n}\rho_{\theta,t}(x)\,\mathrm{d}\lambda\leq\sup_{n\in\mathbb{N},1\leq t\leq n}\int\|x\|^{4+\delta}\rho_{\theta,t}(x)\,\mathrm{d}\lambda<\infty.\qquad\square$$

LEMMA A.6 (Cf. Lemma A.10 in van der Vaart (1988)): *Suppose that Assumption 3.1 holds. Then for any $\tilde{\theta}_n$ which takes the form $\tilde{\theta}_n=(\alpha_n,\beta_n+b_n/\sqrt{n},\eta)$ with $b_n\to b\in\mathbb{R}^{L_\beta}$ a convergent sequence,*

$$R_n:=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\tilde{\ell}_{\tilde{\theta}_n}(Y_t,X_t)-\tilde{\ell}_{\theta_n}(Y_t,X_t)\right]+\tilde{I}_{n,\theta_n}(0',b_n')'\xrightarrow{P^n_{\theta_n}}0.$$

Proof. Let q_θ be as defined in Lemma A.3 and note that the sequence which is to be shown to converge to zero (in probability) can be written as the sum of the following two terms

$$R_{1,n}:=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\tilde{\ell}_{\tilde{\theta}_n}(Y_t,X_t)\left(1-\frac{q_{\tilde{\theta}_n}(Y_t,X_t)^{1/2}}{q_{\theta_n}(Y_t,X_t)^{1/2}}\right)\right]+\frac{1}{2}\tilde{I}_{n,\theta_n}(0',b_n')'$$
$$R_{2,n}:=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\tilde{\ell}_{\tilde{\theta}_n}(Y_t,X_t)\frac{q_{\tilde{\theta}_n}(Y_t,X_t)^{1/2}}{q_{\theta_n}(Y_t,X_t)^{1/2}}-\tilde{\ell}_{\theta_n}(Y_t,X_t)\right]+\frac{1}{2}\tilde{I}_{n,\theta_n}(0',b_n')'$$

To simplify notation, let $Z_{n,t,1}:=\tilde{\ell}_{\tilde{\theta}_n}(Y_t,X_t)\frac{q_{\tilde{\theta}_n}(Y_t,X_t)^{1/2}}{q_{\theta_n}(Y_t,X_t)^{1/2}}$ and $Z_{n,t,2}:=\tilde{\ell}_{\theta_n}(Y_t,X_t)$. Define $m_n(x):=\int\tilde{\ell}_{\tilde{\theta}_n}(y,x)q_{\tilde{\theta}_n}(y,x)^{1/2}q_{\theta_n}(y,x)^{1/2}\,\mathrm{d}y$. Evaluated at X_t, this is the conditional (on X_t) expectation of $Z_{n,t,1}$. Observe that since $\mathbb{E}[\tilde{\ell}_{\theta_n}(Y_t,X_t)|X_t]=0$ under $P^n_{\tilde{\theta}_n}$,

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

Let $\rho_{\theta_n,t}$ be the density of (the non-deterministic parts of) X_t under $P^n_{\theta_n}$, $\bar{\rho}_{\theta_n,n}:=\frac{1}{n}\sum_{t=1}^{n}\rho_{\theta_n,t}$ and $G_{\theta_n,n}$ be the measure corresponding to $\bar{\rho}_{\theta_n,n}$. By Lemma A.3,

$$\lim_{n\to\infty}\int\left[\sqrt{n}\left(q_{\theta_n}^{1/2}-q_{\tilde{\theta}_n}^{1/2}\right)\bar{\rho}_{\theta_n,n}^{1/2}+\frac{1}{2}b_n'\dot{\ell}_{\theta_n}q_{\theta_n}^{1/2}\bar{\rho}_{\theta_n,n}^{1/2}\right]^2\mathrm{d}\lambda=0.\tag{48}$$

Additionally,

$$\lim_{n\to\infty}\int\left\|\tilde{\ell}_{\theta_n}q_{\theta_n}^{1/2}\bar{\rho}_{\theta_n,n}^{1/2}-\tilde{\ell}_{\tilde{\theta}_n}q_{\tilde{\theta}_n}^{1/2}\bar{\rho}_{\theta_n,n}^{1/2}\right\|^2\mathrm{d}\lambda=0.\tag{49}$$

To demonstrate this we first note that by inspection of their forms, it is clear that for ϑ_n equal to either θ, θ_n or $\tilde{\theta}_n$ and any $x_n\to x$, $\tilde{\ell}_{\vartheta_n}(y,x_n)q_{\vartheta_n}(y,x_n)^{1/2}\to\tilde{\ell}_\theta(y,x)q_\theta(y,x)^{1/2}$ (pointwise in y). Moreover, noting the fact that these integrals are expectations conditional on X and using the forms given in Lemma 4.1 along with the continuity given by Assumption 3.1 we have that

$$\lim_{n\to\infty}\int\left\|\tilde{\ell}_{\vartheta_n}(y,x_n)q_{\vartheta_n}^{1/2}(y,x_n)\right\|^2\mathrm{d}y=\int\left\|\tilde{\ell}_\theta(y,x)q_\theta^{1/2}(y,x)\right\|^2\mathrm{d}y<\infty.\tag{50}$$

Hence by Proposition 2.29 in van der Vaart (1998) we have that

$$\lim_{n\to\infty}\int\left\|\tilde{\ell}_{\vartheta_n}(y,x_n)q_{\vartheta_n}^{1/2}(y,x_n)-\tilde{\ell}_\theta(y,x)q_\theta^{1/2}(y,x)\right\|^2\mathrm{d}y=0.\tag{51}$$

Since this also applies with $\vartheta_n = \theta$ we may conclude that

$$\mathscr{Q}_n(x) := \int \left\| \tilde{\ell}_{\vartheta_n}(y,x) q_{\vartheta_n}^{1/2}(y,x) - \tilde{\ell}_{\theta}(y,x) q_{\theta}^{1/2}(y,x) \right\|^2 \,\mathrm{d}y \tag{52}$$

converges continuously to the zero function. By Corollary A.1 and the extended continuous mapping theorem (van der Vaart and Wellner, 1996, Theorem 1.11.1) it follows that $\mathscr{Q}_n(\tilde{X}_n) \rightsquigarrow 0$ where $\tilde{X}_n$ has law $G_{\vartheta_n,n}$. We next show that $\mathscr{Q}_n(\tilde{X}_n)_{n\in\mathbb{N}}$ is uniformly integrable. Dominate the n-th term by

$$\begin{aligned}
\mathscr{Q}_n(\tilde{X}_n) &\le 2\left[\int \left\|\tilde{\ell}_{\vartheta_n}(y,\tilde{X}_n)\right\|^2 q_{\vartheta_n}(y,\tilde{X}_n)\,\mathrm{d}y + \int \left\|\tilde{\ell}_{\theta}(y,\tilde{X}_n)\right\|^2 q_{\theta}(y,\tilde{X}_n)\,\mathrm{d}y\right] \\
&\le 2\left[\mathbb{E}\left[\left\|\tilde{\ell}_{\vartheta_n}(\tilde{Y}_n,\tilde{X}_n)\right\|^2 \middle| \tilde{X}_n\right] + \mathbb{E}\left[\left\|\tilde{\ell}_{\theta}(\tilde{Y},\tilde{X}_n)\right\|^2 \middle| \tilde{X}_n\right]\right],
\end{aligned}$$

where $\tilde{Y}_n$ and $\tilde{Y}$ have laws such that their conditional density given $\tilde{X}_n$ is q_{ϑ_n} and q_θ respectively. Under Assumption 3.1(ii) and using Lemma A.2 and the forms given in Lemma 4.1 it is easily seen that each of $\left(\left\|\tilde{\ell}_{\vartheta_n}(\tilde{Y}_n,\tilde{X}_n)\right\|^2\right)_{n\in\mathbb{N}}$ and $\left(\left\|\tilde{\ell}_{\theta}(\tilde{Y},\tilde{X}_n)\right\|^2\right)_{n\in\mathbb{N}}$ are uniformly integrable. The uniform integrability of the corresponding conditional expectations above then follows from Jensen's inequality for conditional expectations and the de la Vallée - Poussin criterion for uniform integrability (e.g. Bogachev, 2007, Theorem 4.5.9). We may now conclude that

$$\lim_{n\to\infty} \int \left\| \tilde{\ell}_{\vartheta_n} q_{\vartheta_n}^{1/2} \bar{\rho}_{\vartheta_n,n}^{1/2} - \tilde{\ell}_{\theta} q_{\theta}^{1/2} \bar{\rho}_{\vartheta_n,n}^{1/2} \right\|^2 \,\mathrm{d}\lambda = 0, .$$

Using this result twice (once with $\vartheta_n = \theta_n$ and once with $\vartheta_n = \tilde{\theta}_n$) we obtain (49). Combination of (48) and (49) with the continuity of the inner product yields

$$\lim_{n\to\infty} \left\langle \tilde{\ell}_{\tilde{\theta}_n} q_{\tilde{\theta}_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2},\ \sqrt{n}\left(q_{\theta_n}^{1/2} - q_{\tilde{\theta}_n}^{1/2}\right)\bar{\rho}_{\theta_n,n}^{1/2} \right\rangle_\lambda - \left\langle \tilde{\ell}_{\theta_n} q_{\theta_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2},\ -\frac{1}{2} b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2} \right\rangle_\lambda = 0.$$

Since

$$\int \sqrt{n} m_n \bar{\rho}_{\theta_n,n}\,\mathrm{d}\lambda = \left\langle \tilde{\ell}_{\tilde{\theta}_n} q_{\tilde{\theta}_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2},\ \sqrt{n}\left(q_{\theta_n}^{1/2} - q_{\tilde{\theta}_n}^{1/2}\right)\bar{\rho}_{\theta_n,n}^{1/2} \right\rangle_\lambda$$

and

$$\left\langle \tilde{\ell}_{\theta_n} q_{\theta_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2},\ -\frac{1}{2} b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2} \right\rangle_\lambda = -\frac{1}{2}\tilde{I}_{n,\theta_n}(0', b_n')'.$$

Combining these displays allows us to conclude that to establish that $R_{2,n} \to 0$ in $P_{\theta_n}^n$-probability it will suffice to show the same for $R_{2,n}' := \frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{n,t,1} - m_n(X_t) - Z_{n,t,2}$. As is easy to verify, $(Z_{n,t,1} - m_n(X_t) - Z_{n,t,2}, \mathcal{F}_{n,t})_{n\in\mathbb{N}, 1\le t\le n}$ forms a martingale difference array with $\mathcal{F}_{n,t} := \sigma(\epsilon_1, \ldots, \epsilon_t)$. It follows that it suffices to show that (under $P_{\theta_n}^n$)

$$\mathbb{V}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{n,t,1} - m_n(X_t) - Z_{n,t,2}\right] = \frac{1}{n}\sum_{t=1}^n \mathbb{V}\left[Z_{n,t,1} - m_n(X_t) - Z_{n,t,2}\right] \to 0.$$

In view of (49) for this, it suffices to show that

$$\frac{1}{n}\sum_{t=1}^n \mathbb{E}[\|m_n(X_t)\|^2] = \int \|m_n(x)\|^2 \bar{\rho}_{\theta_n,n}(x)\,\mathrm{d}x \to 0.$$

For this, define $\mathsf{m}_n(y,x) := \tilde{\ell}_{\tilde{\theta}_n}(y,x) q_{\tilde{\theta}_n}(y,x)^{1/2}\left[q_{\theta_n}(y,x)^{1/2} - q_{\tilde{\theta}_n}(y,x)^{1/2}\right]$ and note that $\mathsf{m}_n(y,x_n) \to$

0 pointwise for any convergent $x_n \to x$. We additionally have by Cauchy-Schwarz that

$$\left\| \int \mathsf{m}_n(y, x_n) \, dy \right\| \leq \left[\int \|\tilde{\ell}_{\tilde{\theta}_n}\|^2 q_{\tilde{\theta}_n}(y, x_n) \, dy \right]^{1/2} \left[\int \left(q_{\theta_n}(y, x_n)^{1/2} - q_{\tilde{\theta}_n}(y, x_n)^{1/2} \right)^2 dy \right]^{1/2}.$$

As can be easily verified, $\int \|\tilde{\ell}_{\tilde{\theta}_n}\|^2 q_{\tilde{\theta}_n}(y, x_n) \, dy$ is upper bounded by a term of the form $M_1 + M_2 \|x_n\|^2$ (with M_1, M_2 finite positive constants which do not depend on n). Additionally $(q_{\theta_n}(y, x_n)^{1/2} - q_{\tilde{\theta}_n}(y, x_n)^{1/2})^2 \to 0$ pointwise in y and is upper bounded by $2q_{\theta_n}(y, x_n) + 2q_{\tilde{\theta}_n}(y, x_n)$ which satisisfies $\int 2q_{\theta_n}(y, x_n) + 2q_{\tilde{\theta}_n}(y, x_n) \, dy = 4 = \int 4q_\theta(y, x) \, dy$ for each $n \in \mathbb{N}$. Therefore, by the generalised Lebesgue dominated convergence theorem

$$\int \left(q_{\theta_n}(y, x_n)^{1/2} - q_{\tilde{\theta}_n}(y, x_n)^{1/2} \right)^2 dy \to 0.$$

It follows that $\|m_n(x_n)\|^2 \to 0$ pointwise for any $x_n \to x$. Re-using the bound from above, we note that

$$\|m_n(X_t)\|^2 \leq 4(M_1 + M_2 \|X_t\|^2)$$

and hence $\|m_n(X_t)\|^2$ is $G_{\theta_n,n}$-uniformly integrable by Lemma A.2.[41] Moreover, by corollary A.1, $G_{\theta_n,n} \rightsquigarrow \pi_\theta$ and hence by Theorem 3.5 in Serfozo (1982)

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[\|m_n(X_t)\|^2] = \int \|m_n\|^2 \, dG_{\theta_n,n} \to \int 0 \, d\pi_\theta = 0.$$

This establishes that $R_{2,n} \xrightarrow{P^n_{\theta_n}} 0$. For $R_{1,n}$, define $f_n(y, x) := c_n q_{\theta_n}(y, x)^{1/2} q_{\tilde{\theta}_n}(y, x)^{1/2} \bar{\rho}_{\theta_n,n}(x)$, where

$$c_n^{-1} := \int q_{\theta_n}^{1/2} q_{\tilde{\theta}_n}^{1/2} \bar{\rho}_{\theta_n,n} \, d\lambda = 1 - \frac{1}{2} \int (q_{\theta_n}^{1/2} - q_{\tilde{\theta}_n}^{1/2})^2 \bar{\rho}_{\theta_n,n} \, d\lambda.$$

We have

$$-n \left(q_{\theta_n}^{1/2} - q_{\tilde{\theta}_n}^{1/2} \right)^2 = \left(\sqrt{n} \left[q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2} \right] - \frac{1}{2} b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \right)^2 + \left(\frac{1}{2} b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \right)^2 - b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \sqrt{n} \left(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2} \right),$$

and so by Lemma A.3 and the continuity of the inner product

$$\begin{aligned} \int (q_{\theta_n}^{1/2} - q_{\tilde{\theta}_n}^{1/2})^2 \bar{\rho}_{\theta_n,n} \, d\lambda &= \frac{1}{n} \int b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \bar{\rho}_{\theta_n,n}^{1/2} \sqrt{n} \left(q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2} \right) \bar{\rho}_{\theta_n,n}^{1/2} \, d\lambda \\ &\quad - \frac{1}{n} \int \left(\frac{1}{2} b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2} \right)^2 \bar{\rho}_{\theta_n,n} \, d\lambda + o(n^{-1}) \\ &= \frac{1}{2} (n^{-1/2} b_n)' \dot{I}_{\theta_n} (n^{-1/2} b_n) + o(n^{-1}), \end{aligned}$$

where $\dot{I}_{\theta_n} := \int \dot{\ell}_{\theta_n} \dot{\ell}_{\theta_n}' q_{\theta_n} \bar{\rho}_{\theta_n,n} \, d\lambda$.[42] It follows that $c_n^{-1} = 1 - a_n$ with $a_n \to 0$ and $n a_n = \frac{1}{4} b_n' \dot{I}_{\theta_n} b_n + o(1)$. By Taylor's theorem $\log(1 - a_n) = -a_n + R(1 - a_n) a_n^2$ with $R_n(1 - x) \to 0$ as $x \to 0$. Hence $\log c_n^n = -n \log(1 - a_n) = n a_n - n R(1 - a_n) a_n^2 = \frac{1}{4} b_n' \dot{I}_{\theta_n} b_n + o(1)$. $P^n_{\theta_n}$ is the measure corresponding to the product density $\prod_{t=1}^n q_{\theta_n} \rho_{\theta_n,t}$. Let Q_n^n be the measure corresponding to the product density $\prod_{t=1}^n c_n q_{\theta_n}^{1/2} q_{\tilde{\theta}_n}^{1/2} \bar{\rho}_{\theta_n,n}$. Writing $\Lambda_n := \Lambda_n(Q_n^n, P^n_{\theta_n})$ for their

[41] Lemma A.2 bounds the moments of the (demeaned) stationary solution; it is easy to see that this provides a uniform (in t, n) upper bound for our process (conditional on the initial values).

[42] This sequence of matrices is bounded (see e.g. intermediate results used in the proof of Proposition 4.1).

log-likelihood ratio and using notation from the proof of Proposition 4.1, by (35)

$$\Lambda_n = \log c_n^n + 2\sum_{t=1}^{n} \log(Z_{n,t}+1) \overset{P_{\theta_n}^n}{\rightsquigarrow} Z,$$

where Z has a normal distribution. By Example 6.5 in van der Vaart (1998) $P_{\theta_n}^n \triangleleft Q_n^n$ and so by Le Cam's first lemma (e.g. van der Vaart, 1998, Lemma 6.4) it suffices to show that $R_{n,1} \xrightarrow{Q_n^n} 0$. For this we first note that if

$$\frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_{\theta_n}\dot{\ell}_{\theta_n}' - \tilde{I}_{n,\theta_n} \xrightarrow{Q_n^n} 0, \tag{53}$$

then we have $R_{n,1} \xrightarrow{Q_n^n} 0$ as

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\int \left\| \tilde{\ell}_{\tilde{\theta}_n}\sqrt{n}\left(1-\frac{q_{\tilde{\theta}_n}^{1/2}}{q_{\theta_n}^{1/2}}\right) + \frac{1}{2}\sqrt{n}\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta_n}'(b_n/\sqrt{n})\right\| c_n q_{\theta_n}^{1/2} q_{\tilde{\theta}_n}^{1/2}\bar{\rho}_{\theta_n,n}\,\mathrm{d}\lambda \\ \le c_n \int \left\|\tilde{\ell}_{\tilde{\theta}_n} q_{\tilde{\theta}_n}^{1/2}\right\| \sqrt{n}\left| q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2} - \frac{1}{2\sqrt{n}} b_n' \dot{\ell}_{\theta_n} q_{\theta_n}^{1/2}\right| \bar{\rho}_{\theta_n,n}\,\mathrm{d}\lambda \\ = o(1),\end{aligned}$$

where the convergence follows from Lemma A.3 and the continuity of the inner product. It remains to prove (53). For this it suffices to observe that

$$\begin{aligned}Q_n^n\left\|\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta_n}' - \tilde{I}_{n,\theta_n}\right\| &\le |c_n - 1|\int \left\|\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta_n}'\right\| q_{\theta_n}^{1/2} q_{\tilde{\theta}_n}^{1/2}\bar{\rho}_{\theta_n,n}\,\mathrm{d}\lambda + \int \left\|\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta_n}'\right\| q_{\theta_n}^{1/2}|q_{\tilde{\theta}_n}^{1/2} - q_{\theta_n}^{1/2}|\bar{\rho}_{\theta_n,n}\,\mathrm{d}\lambda \\ &= o(1),\end{aligned}$$

by Cauchy-Schwarz and the facts that $\sup_{n\in\mathbb{N}}\|\|\tilde{\ell}_{\theta_n}\dot{\ell}_{\theta_n}'\| q_{\theta_n}^{1/2}\bar{\rho}_{\theta_n,n}^{1/2}\|_{\lambda,2} < \infty$ (under assumption 3.1), $\|q_{\tilde{\theta}_n}^{1/2}\bar{\rho}_{\theta_n,n}^{1/2}\|_{\lambda,2} = 1$, $c_n \to 1$ and $\||q_{\theta_n}^{1/2} - q_{\tilde{\theta}_n}^{1/2}|\bar{\rho}_{\theta_n,n}^{1/2}\|_{\lambda,2} \to 0$ by Lemma A.3. □

Lemma A.7: *Suppose assumption 3.1 holds and let $\Delta_\theta(u)$ be as defined as in Lemma A.3. If $\theta_n = (\gamma_n, \eta) \to (\gamma, \eta) = \theta$, the sequence $(\Delta_{\theta_n}(u))_{n\in\mathbb{N}}$ has uniformly bounded $4+\delta$ moments under $P_{\theta_n}^n$, i.e.*

$$\sup_{n\in\mathbb{N}, 1\le t\le n}\int |\Delta_{\theta_n}(u)|^{4+\delta}\,\mathrm{d}P_{\theta_n}^n < \infty.$$

In consequence, it is uniformly square $P_{\theta_n}^n$-integrable.

Proof. The continuity and continuous differentiability conditions in assumption 3.1 ensure that all non-random terms in the expressions (7) – (9) converge and are thus bounded.[43] Note that under $P_{\theta_n}^n$, $A_{k\bullet}V_{\theta_n,t} \sim \eta_k$ and is independent of both X_t and $A_{j\bullet}V_{\theta_n,t}$ for $j \neq k$. Given this independence and the forms given in (7) – (9) it suffices to show that

$$\mathbb{E}[|\phi_k(\epsilon_k)|^{4+\delta}] < \infty, \quad \mathbb{E}[|\epsilon_k|^{4+\delta}] < \infty, \quad \sup_{n\in\mathbb{N}, 1\le t\le n}\mathbb{E}\,\|X_t\|^{4+\delta} < \infty,$$

where the expectation is taken under $P_{\theta_n}^n$ and we note that η_k does not depend on n. The first two of these follow immediately from the moment assumptions in part (ii) of assumption 3.1.

[43] These terms are of the form 1, $A(\alpha_n,\sigma_n)D_{b_l}$ (with l an integer) or $[D_{x_l}(\alpha_n,\sigma_n)]_{k\bullet}[A(\alpha_n,\sigma_n)^{-1}]_{\bullet j}$ for $k, j \in \{1,\ldots,K\}$ and $x \in \{\alpha,\sigma\}$ (with l an integer).

For the last term, by recursing backwards we obtain

$$Z_t = \sum_{j=0}^{t-1} \mathsf{B}_\theta^j \mathsf{C}_\theta + \sum_{j=0}^{t-1} \mathsf{B}_\theta^j \mathsf{D}_\theta \epsilon_{t-j} + \mathsf{B}_\theta^t Z_0. \tag{54}$$

Assumption 3.1(i) ensures that the matrices B_θ^j are absolutely summable and $\sum_{j=0}^\infty \mathsf{B}_\theta^j = (I - \mathsf{B}_\theta)^{-1}$ exists (e.g. Lütkepohl, 2005, Section A.9.1). Moreover, $\mathsf{T} := \{\theta_n : n \in \mathbb{N}\} \cup \{\theta\}$ is compact, and the spectral radius $M \mapsto \rho(M)$ is a continuous function, then $\{\rho(\mathsf{B}_\vartheta) : \vartheta \in \mathsf{T}\}$ is compact, which ensures that this set is bounded above by some $\upsilon < 1$. Let $m_1, m_2 \in \mathbb{N}$ with $m_2 \geq m_1$ and let $\mathsf{B}_\vartheta = U_\vartheta^* J_\vartheta U_\vartheta$ be a Schur decomposition of B_ϑ (see e.g. Horn and Johnson, 2013, Theorem 2.3.1). Let $\|\cdot\|$ be the spectral norm and note that we have $\|U_\vartheta\| = 1$ and hence by Lemma 5.6.10 in Horn and Johnson (2013), for any $\varepsilon > 0$ with $\upsilon + \varepsilon < 1$ we have

$$\left\| \sum_{j=0}^{m_2} \mathsf{B}_\vartheta^j - \sum_{j=0}^{m_1} \mathsf{B}_\vartheta^j \right\| \leq \sum_{j=m_1}^{m_2} \|\mathsf{B}_\vartheta^j\| \leq \sum_{j=m_1}^{m_2} \|J_\vartheta\|^j \leq \sum_{j=m_1}^{m_2} (\upsilon + \varepsilon)^j.$$

Since $\sum_{j=0}^\infty (\upsilon+\varepsilon)^j = \frac{1}{1-\upsilon-\varepsilon} < \infty$, in view of the preceding display, the convergence $\sum_{j=0}^\infty \mathsf{B}_\theta^j = (I - \mathsf{B}_\theta)^{-1}$ is uniform in $\theta \in \mathsf{T}$. Since $\theta \mapsto \mathsf{C}_\theta$ is continuous, this immediately implies that $\sup_{n\in\mathbb{N}, 1\leq t\leq n} \left\| \sum_{j=0}^{t-1} \mathsf{B}_{\theta_n}^j \mathsf{C}_{\theta_n} \right\|^{4+\delta} < \infty$. Similarly using the same uniform bound, that $(\epsilon_t)_{t\geq 1}$ are i.i.d. and since $\theta \mapsto \|\mathsf{D}_\theta\|$ is continuous we have that

$$\sup_{n\in\mathbb{N}, 1\leq t\leq n} \mathbb{E} \left\| \sum_{j=0}^{t-1} \mathsf{B}_{\theta_n}^j \mathsf{D}_{\theta_n} \epsilon_{t-j} \right\|^{4+\delta} \leq \sup_{n\in\mathbb{N}} \|\mathsf{D}_{\theta_n}\|^{4+\delta}\, \mathbb{E}\, \|\epsilon_t\|^{4+\delta} \sup_{n\in\mathbb{N}, 1\leq t\leq n} \sum_{j=0}^{t-1} \|B_{\theta_n}^j\|^{4+\delta} < \infty.$$

Hence by Minkowski's inequality we have that $\sup_{n\in\mathbb{N},1\leq t\leq n} \mathbb{E}\, \|X_t\|^{4+\delta} < \infty$, where the expectation is taken under $P_{\theta_n}^n$. □

Lemma A.8: *Suppose assumption 3.1 holds and let $\Delta_\theta(u)$ be as defined as in Lemma A.3. If $\theta_n = (\gamma_n, \eta) \to (\gamma, \eta) = \theta$ and G_θ is defined as in Lemma A.9, then under $P_{\theta_n}^n$,*

$$\lim_{n\to\infty} \mathbb{E} \left| \frac{1}{n} \sum_{t=1}^n \Delta_{\theta_n}(u)(Y_t, X_t)^2 - \tau^2 \right|^2 = 0, \qquad \text{with } \tau^2 := G_\theta \Delta_\theta(u)^2 < \infty.$$

Proof. Let ϑ_n indicate either θ_n or θ. By inspection of their forms it is clear that for any $x_n \to x$, $[\Delta_{\vartheta_n}(u)(y, x_n)] q_{\vartheta_n}(y, x_n)^{1/2} \to [\Delta_\theta(u)(y, x)] q_\theta(y, x)^{1/2}$ pointwise in y. By inspection of their form, the continuity given by Assumption 3.1 we have

$$\lim_{n\to\infty} \int [\Delta_{\vartheta_n}(u)(y, x_n)]^2 q_{\vartheta_n}(y, x_n)\, \mathrm{d}y = \int [\Delta_\theta(u)(y, x)]^2 q_\theta(y, x)\, \mathrm{d}y < \infty,$$

i.e. $\mathscr{Q}_n$ converges continuously to $\mathscr{Q}$ where

$$\mathscr{Q}_n(x) := \int [\Delta_{\vartheta_n}(u)(y, x)]^2 q_{\vartheta_n}(y, x)\, \mathrm{d}y, \quad \mathscr{Q}(x) := \int [\Delta_\theta(u)(y, x)]^2 q_\theta(y, x)\, \mathrm{d}y.$$

We can bound

$$\mathscr{Q}_n(x) \leq 2 \left[\mathbb{E} \left[\Delta_{\theta_n}(u)(\tilde{Y}_n, \tilde{X}_n)^2 \middle| \tilde{X}_n \right] + \mathbb{E} \left[\Delta_\theta(u)(\tilde{Y}, \tilde{X}_n)^2 \middle| \tilde{X}_n \right] \right],$$

where $\tilde{Y}_n$ and $\tilde{Y}$ have laws such that their conditional density given $\tilde{X}_n$ is q_{ϑ_n} and q_θ respectively.

Under Assumption 3.1(ii) and an argument similar to that of Lemma A.7 it is easily seen that each of $\left(\Delta_{\theta_n}(u)(\tilde{Y}_n, \tilde{X}_n)^2\right)_{n\in\mathbb{N}}$ and $\left(\Delta_{\theta}(u)(\tilde{Y}, \tilde{X}_n)^2\right)_{n\in\mathbb{N}}$ are uniformly integrable. The uniform integrability of the corresponding conditional expectations above then follows from Jensen's inequality for conditional expectations and the de la Vallée - Poussin criterion for uniform integrability (e.g. Bogachev, 2007, Theorem 4.5.9). Corollary A.1 in conjunction with Theorem 3.5 of Serfozo (1982) then yields that $G_{\theta_n,\theta_n,n}\Delta_{\theta_n}(u)^2 \to G_\theta\Delta_\theta(u)^2 < \infty$, where $G_{\vartheta,\theta,n}(A) := \int\int \mathbf{1}_A(y,x) q_\vartheta(y,x)\,\mathrm{d}y\,\mathrm{d}\bar{\rho}_{\theta,n}(x)$ and the finiteness follows from the form of $\Delta_\theta(u)$, assumption 3.1(ii) and Lemma A.2.[44] The convergence follows on combining Lemma A.7, Proposition A.1 and Corollary 19.3(ii) of Davidson (1994). □

Lemma A.9: *Suppose that assumption 3.1 holds. Let $\rho_{\theta,t}$ be the density of $\mathsf{X}_t := \mathrm{vec}(Y_{t-1},\ldots,Y_{t-p})$ (i.e. the non-deterministic parts of X_t) under θ. Let $G_{\vartheta,n}$ be the measure defined by $G_{\vartheta,n}(A) := \int_A q_\vartheta \frac{1}{n}\sum_{t=1}^n \rho_{\theta,t}\,\mathrm{d}\lambda$ and G_θ the measure defined by $G_\theta(A) := \int_A q_\theta(y,x)\,\mathrm{d}(\lambda(y)\otimes\pi_\theta(x))$ for q_ϑ as defined as in (40). Let $(\vartheta_n)_{n\in\mathbb{N}} \subset \Theta$ be an sequence with $\vartheta_n = (\gamma_n,\eta) \to (\gamma,\eta) = \theta$. Then, $G_{\vartheta_n,n} \xrightarrow{TV} G_\theta$.*

Proof. By the form of $\theta \mapsto q_\theta$ we have that $q_{\vartheta_n} \to q_\theta$ (pointwise) as $n\to\infty$. Hence, for any x, $q_{\vartheta_n}(\cdot,x) \to q_\theta(\cdot,x)$ pointwise and since each $q_{\vartheta_n}(\cdot,x)$ and $q_\theta(\cdot,x)$ is a probability density with respect to Lebesgue measure, by Proposition 2.29 in van der Vaart (1998),

$$\mathscr{Q}_n(x) := \int |q_{\vartheta_n}(y,x) - q_\theta(y,x)|\,\mathrm{d}y \to 0,$$

pointwise in x. Let $(\psi_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions on $\mathbb{R}^{Kp}$ with $\psi_n \in [0,1]$ and $\pi_{\theta,n}$ the probability measure corresponding to the density $\frac{1}{n}\sum_{t=1}^n \rho_{\theta,t}$. Then

$$\left|\int\int \psi_n(y,x)(q_{\vartheta_n}(y,x) - q_\theta(y,x))\,\mathrm{d}y\,\mathrm{d}\pi_{\theta,n}(x)\right| \le \int \mathscr{Q}_n(x)\,\mathrm{d}\pi_{\theta,n}(x).$$

Since $\mathscr{Q}_n(x) \le \int q_{\vartheta_n}(y,x)\,\mathrm{d}y + \int q_\theta(y,x)\,\mathrm{d}y$ and $\int\left[\int q_{\vartheta_n}(y,x)\,\mathrm{d}y + \int q_\theta(y,x)\,\mathrm{d}y\right]\mathrm{d}\pi_{\theta,n}(x) = 2 = \int 2\int q_\theta(y,x)\,\mathrm{d}y\,\mathrm{d}\pi_\theta(x)$, the $\mathscr{Q}_n(x)$ are uniformly $\pi_{\theta,n}$ – integrable.[45] Hence by Corollary 2.9 of Feinberg et al. (2016), $\int \mathscr{Q}_n(x)\,\mathrm{d}\pi_{\theta,n}(x) \to 0$. The proof is completed by noting that by Proposition A.1,

$$\left|\int\int \psi_n(y,x) q_\theta(y,x)\,\mathrm{d}y\,\mathrm{d}\pi_{\theta,n}(x) - \int \psi_n\,\mathrm{d}G_\theta\right| \to 0.$$

□

Lemma A.10: *Let $\gamma_n = (\alpha_n,\beta) \to (\alpha,\beta) = \gamma$ and $\theta_n = (\gamma_n,\eta) \to (\gamma,\eta) = \theta$ for $\gamma_n,\gamma \in \Gamma$. Let $\tilde{\gamma}_n = (\alpha_n,\beta_n) \to (\alpha,\beta) = \gamma$, $\tilde{\theta}_n := (\tilde{\gamma}_n,\eta) \to (\gamma,\eta) = \theta$ with $b_n := \sqrt{n}(\beta_n - \beta) \to b$ and $\check{\theta}_n := (\tilde{\gamma}_n,\tilde{\eta}_n) \to \theta$ with $\tilde{\eta}_n := \eta(1 + h_n/\sqrt{n})$ for $h_n \to h$ in $\mathscr{H}$. Then, under the conditions of Theorem 5.1,*

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

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

[44] $\bar{\rho}_{\theta,n} := \frac{1}{n}\sum_{t=1}^n \rho_{\theta,t}$.

[45] The uniform integrability follows since, by the integral equality, Proposition A.1 and Proposition A.2,

$$\int\left|\int q_{\vartheta_n}(y,x)\,\mathrm{d}y + \int q_\theta(y,x)\,\mathrm{d}y - 2\int q_\theta(y,x)\,\mathrm{d}y\right|\mathrm{d}\pi_{\theta,n}(x) \to 0.$$

(ii) *We have that*

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

(iii) $\tilde{I}_{n,\theta_n}\to\tilde{I}_\theta:=G_\theta\dot{\ell}_\theta\dot{\ell}'_\theta$ *and* $P^n_{\tilde{\theta}_n}\left(\|\hat{I}_{n,\tilde{\theta}_n}-\tilde{I}_\theta\|_2<\nu_n\right)\to 1$ *where* ν_n *is defined in Assumption 3.2 and* G_θ *in Lemma A.9.*

(iv) *We have that*

$$R_n:=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\tilde{\ell}_{\tilde{\theta}_n}(Y_t,X_t)-\tilde{\ell}_{\theta_n}(Y_t,X_t)\right]+\tilde{I}_{n,\theta_n}(0',\sqrt{n}(\beta_n-\beta)')'\xrightarrow{P^n_{\theta_n}}0.$$

Proof. For part (i), let $z_{n,t}$ be

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

and $\mathcal{F}_{n,t}:=\sigma(\epsilon_1,\ldots,\epsilon_t)$. Under assumption 3.1(ii), $\mathbb{E}\|z_{n,t}\|_2^2<\infty$ and $\{z_{n,t},\mathcal{F}_{n,t}:1\le t\le n,n\in\mathbb{N}\}$ is a martingale difference array (all under $P^n_{\theta_n}$) such that

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[z_{n,t}z'_{n,t}\right]=\begin{bmatrix}\tilde{I}_{n,\theta_n} & \tilde{I}_{n,\theta_n}(0',b'_n)'\\(0',b'_n)\tilde{I}_{n,\theta_n} & \sigma^2_{n,b,h}\end{bmatrix}\to\begin{bmatrix}\tilde{I}_\theta & \tilde{I}_\theta(0',b')'\\(0',b')\tilde{I}_\theta & \sigma^2_{b,h}\end{bmatrix},$$

noting Lemma 4.1 and Theorem 12.14 of Rudin (1991). That $\sigma^2_{n,b,h}$ converges to a $\sigma^2_{b,h}$ is part of the conclusion of Proposition 4.1. That $\tilde{I}_{n,\theta_n}\to\tilde{I}_\theta$ follows from Lemma A.8. Moreover, the Lindeberg condition in (67) is satisfied since $\{\|z_{n,t}\|^2:1\le t\le n,n\in\mathbb{N}\}$ is uniformly $P^n_{\theta_n}$-integrable. That this is true for $\|z_{n,t,2}\|^2$ follows from A.7. That it is also true for $\|z_{n,t,1}\|^2$ can be shown by an analogous argument. Part (i) then follows from Propositions 4.1, A.3 and Lemma A.8.

Next, define $A_n:=A_{\tilde{\theta}_n}$ and $B_n:=B_{\tilde{\theta}_n}$ and note that each $A_{n,k}(Y_t-c_n-B_nX_t)\eqsim\epsilon_{k,t}\sim\eta_k$ under $P^n_{\tilde{\theta}_n}$. Hence we can compute certain properties of the efficient score using the equality in distribution:

$$\tilde{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t,X_t)\eqsim\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\zeta^\alpha_{l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t}+\sum_{k=1}^{K}\zeta^\alpha_{n,l,k,k}\left[\tau_{k,1}\epsilon_{k,t}+\tau_{k,2}\kappa(\epsilon_{k,t})\right]\tag{55}$$

$$\tilde{\ell}_{\tilde{\theta}_n,\sigma_l}(Y_t,X_t)\eqsim\sum_{k=1}^{K}\sum_{j=1,j\neq k}^{K}\zeta^\sigma_{n,l,k,j}\phi_k(\epsilon_{k,t})\epsilon_{j,t}+\sum_{k=1}^{K}\zeta^\sigma_{l,k,k}\left[\tau_{k,1}\epsilon_{k,t}+\tau_{k,2}\kappa(\epsilon_{k,t})\right]\tag{56}$$

$$\tilde{\ell}_{\tilde{\theta}_n,b_l}(Y_t,X_t)\eqsim\sum_{k=1}^{K}-A_{n,k\bullet}D_{b,l}\left[\phi_k(\epsilon_{k,t})(X_t-\mathbb{E}X_t)-\mathbb{E}X_t\left(\varsigma_{k,1}\epsilon_{k,t}+\varsigma_{k,2}\kappa(\epsilon_{k,t})\right)\right]\tag{57}$$

where we note that the same observation implies that $\tau_{k,n}=\tau_k$ and $\varsigma_{k,n}=\varsigma_k$ for each n.[46] By our assumptions on the map $(\alpha,\sigma)\mapsto A(\alpha,\sigma)$, we have $\zeta^\alpha_{n,l,k,j}\to\zeta^\alpha_{\infty,l,k,j}:=[D_{\alpha_l}(\alpha_0,\sigma)]_{k\bullet}A(\alpha,\sigma)^{-1}_{\bullet j}$ and $\zeta^\sigma_{n,l,k,j}\to\zeta^\sigma_{\infty,l,k,j}:=[D_{\sigma_l}(\alpha,\sigma)]_{k\bullet}A(\alpha,\sigma)^{-1}_{\bullet j}$. 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 (55)-(57). Components (55) and (56)

[46] In the preceding display we have written $\zeta^\alpha_{n,l,k,k}$ and $\zeta^\sigma_{n,l,k,k}$ rather than $\zeta^\alpha_{l,k,k}$ and $\zeta^\sigma_{l,k,k}$ to indicate their dependence on $\tilde{\theta}_n$.

follow similarly and we focus on (55). We define

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

and

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

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

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

Since 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 the Lemma A.1 with $W_{n,t} = A_{n,j\bullet}V_{n,t}$ (noting that $A_{n,k\bullet}V_{n,t} \simeq \epsilon_{k,t}$ and $A_{n,j\bullet}V_{n,t} \simeq \epsilon_{j,t}$ are independent with $\mathbb{E}_{\theta_n}(A_{n,j\bullet}V_{n,t})^2 = 1$ by Assumption 3.1(ii), hence the WLLN implies the required convergence) and the outside summations are finite, it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\hat{\varphi}_{1,n,t} - \varphi_{1,n,t}) = o_{P^n_{\tilde{\theta}_n}}(1) \ . \tag{58}$$

That $\hat{\tau}_{k,n} \xrightarrow{P^n_{\tilde{\theta}_n}} \tau_k$ follows from Lemma A.12. Now, consider $\varphi_{2,\tau,n,t}$ defined by

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

Since sum is finite and each $|\zeta^{\alpha}_{n,l,k,k}| \to |\zeta^{\alpha}_{\infty,l,k,k}| < \infty$ 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 $A_{n,k\bullet}V_{n,t} \eqsim \epsilon_{k,t} \sim \eta_k$ 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 CLT holds.

Together these yield that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varphi_{2,\hat{\tau}_n,n,t} - \varphi_{2,\tau,n,t}) \to 0 \quad \text{in } P^n_{\tilde{\theta}_n}\text{-probability.} \tag{59}$$

Putting (58) and (59) together yields the required convergence for components of the type (55). We note that the required convergence for components of type (56) follows using identical steps.

For components (57) let $a_{n,k,l} \coloneqq -A_{n,k\bullet} D_{b_l}$ and note for $\tilde{\varsigma}_{k,n,1} \coloneqq \hat{\varsigma}_{k,n} - \varsigma_k$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\hat{\ell}_{\tilde{\theta}_n, b_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n, b_l}(Y_t, X_t)) \\
&\quad = \sum_{k=1}^{K} a_{n,k,l} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[(X_t - \mathbb{E}\, X_t) \left[\hat{\phi}_k(A_{n,k\bullet} V_{n,t}) - \phi_k(A_{n,k\bullet} V_{n,t}) \right] \right] \\
&\quad\quad + \sum_{k=1}^{K} a_{n,k,l} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[(\mathbb{E}\, X_t - \bar{X}_n) \left(\phi_k(A_{n,k\bullet} V_{n,t}) + \hat{\varsigma}_{k,n,1} A_{n,k\bullet} V_{n,t} + \hat{\varsigma}_{k,n,2} \kappa(A_{n,k\bullet} V_{n,t}) \right) \right] \\
&\quad\quad - \sum_{k=1}^{K} a_{n,k,l} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\mathbb{E}\, X_t \left(\tilde{\varsigma}_{k,n,1} A_{n,k\bullet} V_{n,t} + \tilde{\varsigma}_{k,n,2} \kappa(A_{n,k\bullet} V_{n,t}) \right) \right].
\end{aligned}$$

Noting first that $a_{n,k,l} \to a_{\infty,k,l} \coloneqq A_{k\bullet} D_{b_l}$, each of the terms on the right hand side converges to zero in probability. For the first term, this follows from Lemma A.1 applied with $W_{n,t} \coloneqq a_{n,k,l}(X_t - \mathbb{E}\, X_t)$, noting that this is independent of $A_{n,k\bullet} V_{n,t}$ by Assumption (ii).[47] For the second term, this follows from A.12, the CLT applied to $A_{n,k\bullet} V_{n,t} \simeq \epsilon_{k,t}$, $\kappa(A_{n,k\bullet} V_{n,t}) \simeq \kappa(\epsilon_{k,t})$ and $\phi_k(A_{n,k\bullet} V_{n,t}) \simeq \phi_k(\epsilon_{k,t})$ and the fact that $\bar{X}_n - \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\, X_t$ converges to zero in probability by e.g. Corollary 19.3 in Davidson (1994), Lemma A.2 (which provides a uniform upper bound for the $4+\delta$ moments of the X_t) and Proposition A.1. For the third term, this follows from A.12 and the CLT applied to $A_{n,k\bullet} V_{n,t} \simeq \epsilon_{k,t}$ and $\kappa(A_{n,k\bullet} V_{n,t}) \simeq \kappa(\epsilon_{k,t})$.

The first part of (iii) follows from Lemma A.8. To verify the second part of (iii) we will show that

$$\left\| \hat{I}_{n,\tilde{\theta}_n} - \tilde{I}_{\theta} \right\|_2 \leq \left\| \hat{I}_{n,\tilde{\theta}_n} - \breve{I}_{n,\tilde{\theta}_n} \right\|_2 + \left\| \breve{I}_{n,\tilde{\theta}_n} - \tilde{I}_{\theta} \right\|_2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}). \tag{60}$$

where $\tilde{I}_\theta \coloneqq \mathbb{E}[\tilde{\ell}_\theta(Y_t, X_t) \tilde{\ell}_\theta(Y_t, X_t)'] = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[\tilde{\ell}_\theta(Y_t, X_t) \tilde{\ell}_\theta(Y_t, X_t)']$ with the expectation taken under G_θ, $\hat{I}_{n,\theta} \coloneqq \frac{1}{n} \sum_{t=1}^{n} \hat{\ell}_\theta(Y_t, X_t) \hat{\ell}_\theta(Y_t, X_t)'$ and $\breve{I}_{n,\theta} \coloneqq \frac{1}{n} \sum_{t=1}^{n} \tilde{\ell}_\theta(Y_t, X_t) \tilde{\ell}_\theta(Y_t, X_t)'$.

To obtain the rates we start with $\|\breve{I}_{\theta_n} - \tilde{I}_\theta\|_2$, for which we show that each component satisfies the required rate. To set this up, let $Q^{r,s}_{l,m,t,n} = \tilde{\ell}_{\tilde{\theta}_n, r_l}(Y_t, X_t) \tilde{\ell}_{\tilde{\theta}_n, s_m}(Y_t, X_t)$, 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 it suffices to show

$$\frac{1}{n} \sum_{t=1}^{n} Q^{r,s}_{l,m,t,n} - \mathbb{E}_{\tilde{\theta}_n} Q^{r,s}_{l,m,t,n} + \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}_{\tilde{\theta}_n}[Q^{r,s}_{l,m,t,n}] - \mathbb{E}_\theta[Q^{r,s}_{l,m,t,\infty}] = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}).$$

For the first term, by the fact that $\tilde{\ell}_{\tilde{\theta}_n}$ has uniformly bounded $4+\delta$ moments,[48] Proposition A.1 and Theorem 1 of Kanaya (2017) we obtain

$$\frac{1}{n} \sum_{t=1}^{n} Q^{r,s}_{l,m,t,n} - \mathbb{E}_{\tilde{\theta}_n} Q^{r,s}_{l,m,t,n} = O_{P^n_{\tilde{\theta}_n}}\left(n^{(1/p-1)/2} \right) = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}), \quad p \in (1, 1+\delta/4].$$

That the second term is $o(\nu_n^{1/2})$ follows by the assumed Lipschitz continuity of the map defining the ζ's, that of each $\beta \mapsto A(\alpha, \sigma)_{k\bullet}$ (which holds locally at θ) and Lemma A.11.

For the other component of the sum, let $r \in \{\alpha, \sigma, b\}$ and let l denote an index, we write $\hat{U}_{n,t,r_l} \coloneqq \hat{\ell}_{\tilde{\theta}_n, r_l}(Y_t, X_t)$, $\tilde{U}_{t,r_l} \coloneqq \tilde{\ell}_{\tilde{\theta}_n, r_l}(Y_t, X_t)$ and $D_{n,t,r_l} \coloneqq \hat{\ell}_{\tilde{\theta}_n, r_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n, r_l}(Y_t, X_t)$.

Since it is the absolute value of the $(r,l) - (s,m)$ component of $\hat{I}_{n,\tilde{\theta}_n} - \breve{I}_{n,\tilde{\theta}_n}$, it is sufficient to show that $\left| \frac{1}{n} \sum_{t=1}^{n} \hat{U}_{n,t,r_l} D_{n,t,s_m} + \frac{1}{n} \sum_{t=1}^{n} D_{n,t,r_l} \tilde{U}_{t,s_m} \right| = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$ as $n \to \infty$ for any $r, s \in$

[47] The convergence condition follows by combining Proposition A.1, Lemma A.2 (which provides a uniform upper bound for the $4+\delta$ moments of the X_t) and Corollary 19.3 of Davidson (1994).

[48] Argue as in Lemma A.7.

$\{\alpha, \sigma, b\}$ and l, m. By Cauchy-Schwarz and lemma A.13

$$\left|\frac{1}{n}\sum_{t=1}^n D_{n,t,r_l}\tilde{U}_{t,s_m}\right| \le \left(\frac{1}{n}\sum_{t=1}^n \tilde{U}_{t,s_m}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n D_{n,t,r_l}^2\right)^{1/2} = O_{P^n_{\tilde{\theta}_n}}(1)\times o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}) = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}),$$

$$\left|\frac{1}{n}\sum_{t=1}^n \hat{U}_{n,t,r_l}D_{n,t,s_m}\right| \le \left(\frac{1}{n}\sum_{t=1}^n \hat{U}_{n,t,r_l}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n D_{n,t,s_m}^2\right)^{1/2} = O_{P^n_{\tilde{\theta}_n}}(1)\times o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}) = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}),$$

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

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

and hence $\|\hat{I}_{n,\tilde{\theta}_n} - \breve{I}_{n,\tilde{\theta}_n}\|_2 \le \|\hat{I}_{n,\tilde{\theta}_n} - \breve{I}_{n,\tilde{\theta}_n}\|_F = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$. We can combine these results to obtain:

$$\|\hat{I}_{n,\tilde{\theta}_n} - \tilde{I}_\theta\|_2 \le \|\hat{I}_{n,\tilde{\theta}_n} - \breve{I}_{n,\tilde{\theta}_n}\|_2 + \|\breve{I}_{n,\tilde{\theta}_n} - \tilde{I}_\theta\|_2 = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}) + o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}) = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2}).$$

Part (iv) follows directly from Lemma A.6. □

Lemma A.11: *In the setting of Lemma A.10*

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

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

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

Proof. For (i) we decompose as

$$\mathbb{E}_{\tilde{\theta}_n} X_t - \mathbb{E}_\theta X_t = [\mathbb{E}_{\tilde{\theta}_n} X_t - \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t] + [\mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \mathbb{E}_\theta \tilde{X}_t] + [\mathbb{E}_\theta \tilde{X}_t - \mathbb{E}_\theta X_t]$$

where $\tilde{X}_t$ denotes a stationary solution to the VAR equation. Note that for all $\vartheta \in \{\tilde{\theta}_n : n \in \mathbb{N}\} \cup \{\theta\}$ and some $\rho_\star < 1$

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^n \|\mathbb{E}_\vartheta X_t - \mathbb{E}_\vartheta \tilde{X}_t\|^2 &= \frac{1}{n}\sum_{t=1}^n \|\mathbb{E}_\vartheta Z_{t-1} - \mathbb{E}_\vartheta \tilde{Z}_{t-1}\|^2 \\
&\le \|\mathsf{C}_\vartheta\|^2 \times \frac{1}{n}\sum_{t=1}^n \left(\sum_{j=t}^\infty \|\mathsf{B}_\vartheta^j\|\right)^2 \\
&\le \|\mathsf{C}_\vartheta\|^2 \times \frac{1}{n}\sum_{t=1}^n \left(\sum_{j=t}^\infty \rho_\star^j\right)^2 \\
&= \frac{\|\mathsf{C}_\vartheta\|^2}{(1-\rho_\star)^2} \times \frac{1}{n}\sum_{t=1}^n \rho_\star^{2t} \\
&= \frac{\|\mathsf{C}_\vartheta\|^2\left(1-\rho_\star^{2(n+1)}\right)}{(1-\rho_\star)^2(1-\rho_\star^2)} \times \frac{1}{n} \\
&= O(n^{-1}),
\end{aligned} \tag{61}$$

and hence by Jensen's inequality $\frac{1}{n}\sum_{t=1}^{n} \| \mathbb{E}_\vartheta X_t - \mathbb{E}_\vartheta \tilde{X}_t\| = O(n^{-1/2})$. The middle term satisfies

$$\frac{1}{n}\sum_{t=1}^{n} \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \mathbb{E}_\theta \tilde{X}_t\| = \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \mathbb{E}_\theta \tilde{X}_t\| = (I-\mathsf{B}_{\tilde{\theta}_n})^{-1}\mathsf{C}_{\tilde{\theta}_n} - (I-\mathsf{B}_\theta)^{-1}\mathsf{C}_\theta = O(n^{-1/2}),$$

since $\beta \mapsto (I-\mathsf{B}_\theta)^{-1}\mathsf{C}_\theta$ is locally Lipschitz at θ.

For (ii), note that combination of the preceding displays yields that

$$\frac{1}{n}\sum_{t=1}^{n} \| \mathbb{E}_{\tilde{\theta}_n} X_t - \mathbb{E}_\theta X_t\|^2 = O(n^{-1}),$$

which, in conjunction with the Cauchy-Schwarz inequality and Lemma A.2 yields (ii).

For (iii) let $U_{\vartheta,t} \coloneqq X_t - \mathbb{E}_\vartheta X_t$ and $\tilde{U}_{\vartheta,t} \coloneqq \tilde{X}_t - \mathbb{E}_\vartheta \tilde{X}_t$. We note that for all $\vartheta \in \{\tilde{\theta}_n : n \in \mathbb{N}\} \cup \{\theta\}$, some $\rho_\star < 1$ and some finite, positive M

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}_\vartheta\left(U_{\vartheta,t}U_{\vartheta,t}'\right) - \mathbb{E}_\vartheta\left(\tilde{U}_{\vartheta,t}\tilde{U}_{\vartheta,t}'\right) &= \frac{1}{n}\sum_{t=1}^{n}\sum_{j=t}^{\infty} \mathsf{B}_\vartheta^j \mathsf{D}_\vartheta \mathsf{D}_\vartheta' (\mathsf{B}_\vartheta^j)' \\
&\leq M\frac{1}{n}\sum_{t=1}^{n}\sum_{j=t}^{\infty} \rho_\star^{2j} \\
&= \frac{M}{1-\rho_\star^2}\frac{1}{n}\sum_{t=1}^{n}\rho_\star^{2t} \\
&= \frac{M\left(1-\rho_\star^{2(n+1)}\right)}{(1-\rho_\star^2)^2}\frac{1}{n} \\
&= O(n^{-1}).
\end{aligned}$$

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 under our assumptions. This implies that

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

By using a similar decomposition as in (i), the previous two displays suffice for (iii). □

Lemma A.12: *If assumption 3.1 holds, then* $\|\hat{\varrho}_{k,n} - \varrho_{k,n}\|_2 = o_{P^n_{\tilde{\theta}_n}}(\nu_{n,p}) = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$*, where* $\tilde{\theta}_n$ *is as in Lemma A.10 and* $\varrho \in \{\tau, \varsigma\}$*.*

Proof. Under $P^n_{\tilde{\theta}_n}$, $A_{n,k\bullet}V_{n,t} \approx \epsilon_{k,t} \sim \eta_k$, for $V_{n,t} \coloneqq Y_t - c_n - B_n X_t$. Let $w \in \{(0,-2)', (1,0)'\}$ By the fact that the map $M \mapsto M^{-1}$ is Lipschitz at a positive definite matrix M_0 we have that for a positive constant C then for large enough n, with probability approaching one

$$\|\hat{\varrho}_{k,n} - \varrho_{k,n}\|_2 = \|(\hat{M}_{k,n}^{-1} - M_k^{-1})w\|_2 \leq 2\|\hat{M}_{k,n}^{-1} - M_k^{-1}\|_2 \leq 2C\|\hat{M}_{k,n} - M_k\|_2. \tag{62}$$

By Theorem 2.5.11 in Durrett (2019)

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

These together imply that

$$\|\hat{M}_{k,n} - M_k\|_2 \leq \|\hat{M}_{k,n} - M_k\|_F = o_{P^n_{\tilde{\theta}_n}}\left(n^{\frac{1-p}{p}}\right) = o_{P^n_{\tilde{\theta}_n}}(\nu_{n,p}).$$

Combining these convergence rates with equation (62) yields the result. □

LEMMA A.13: *Suppose assumptions 3.1 and 3.2 hold and $\tilde{\theta}_n = (\alpha_n, \beta_n, \eta)$ where $\sqrt{n}(\beta_n - \beta) = O(1)$ is a deterministic sequence. Then 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 $\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t, X_t) - \tilde{\ell}_{\tilde{\theta}_n,\alpha_l}(Y_t, X_t)\right)^2$. We define $\tilde{\tau}_{k,n,q} := \hat{\tau}_{k,n,q} - \tau_{k,q}$ and $V_{n,t} = Y_t - c_n - B_n X_t$. 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 $s \neq m$,

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right]\left[\hat{\phi}_{s,n}(A_{n,s\bullet}V_{n,t}) - \phi_s(A_{n,s\bullet}V_{n,t})\right] A_{n,j\bullet}V_{t,n}A_{n,m\bullet}V_{n,t} = o_{P^n_{\tilde{\theta}_n}}(\nu_n), \tag{63}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\phi}_{k,n}(A_{n,k\bullet}V_{n,t}) - \phi_k(A_{n,k\bullet}V_{n,t})\right] A_{n,j\bullet}V_{n,t}\left[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{n,t} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{n,t})\right] = o_{P^n_{\tilde{\theta}_n}}(\nu_n), \tag{64}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left[\tilde{\tau}_{s,n,1}A_{n,s\bullet}V_{n,t} + \tilde{\tau}_{s,n,2}\kappa(A_{n,s\bullet}V_{n,t})\right]\left[\tilde{\tau}_{k,n,1}A_{n,k\bullet}V_{n,t} + \tilde{\tau}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})\right] = o_{P^n_{\tilde{\theta}_n}}(\nu_n). \tag{65}$$

For (65), 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 A.12.[49] For (64) 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 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 $\tilde{\tau}_{s,n,q} = o_{P^n_{\tilde{\theta}_n}}(\nu_n^{1/2})$.[50] For (63) use Cauchy-

[49] 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 3.1 and Markov's inequality, noting once more that $A_{n,k\bullet}V_{n,t} \simeq \epsilon_{k,t}$ under $P^n_{\tilde{\theta}_n}$.

[50] See footnote 49.

Schwarz with Lemma A.1

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

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

Finally, we consider the elements in $\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\theta_n,b_l}(Y_t,X_t)-\tilde{\ell}_{\theta_n,b_l}(Y_t,X_t)\right)^2$, where we note that with $\tilde{\varsigma}_{k,n}:=\hat{\varsigma}_{k,n}-\varsigma_k$,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(\hat{\ell}_{\theta_n,b_l}(Y_t,X_t)-\tilde{\ell}_{\theta_n,b_l}(Y_t,X_t)\right)^2\\
&\lesssim\sum_{k=1}^{K}\frac{1}{n}\sum_{t=1}^{n}\left[[a_{n,k,l}(X_t-\mathbb{E}\,X_t)]^2\left[\hat{\phi}_k(A_{n,k\bullet}V_{n,t})-\phi_k(A_{n,k\bullet}V_{n,t})\right]^2\right]\\
&\quad+\sum_{k=1}^{K}\frac{1}{n}\sum_{t=1}^{n}\left[[a_{n,k,l}(\mathbb{E}\,X_t-\bar{X}_n)]^2\left(\phi_k(A_{n,k\bullet}V_{n,t})+\hat{\varsigma}_{k,n,1}A_{n,k\bullet}V_{n,t}+\hat{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})\right)^2\right]\\
&\quad+\sum_{k=1}^{K}\frac{1}{n}\sum_{t=1}^{n}\left[[a_{n,k,l}\,\mathbb{E}\,X_t]^2\left(\tilde{\varsigma}_{k,n,1}A_{n,k\bullet}V_{n,t}+\tilde{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})\right)^2\right]
\end{aligned}$$

That the first right hand side term is $o_{P^n_{\tilde{\theta}_n}}(\nu_n)$ follows by Lemma A.1.[51] and the Cauchy-Schwarz inequality. The third follows from Lemma A.12 since $[a_{n,k,l}\,\mathbb{E}\,X_t]^2$ is uniformly (in t) bounded (cf. Lemma A.2).

For the second, let $\tilde{X}_t$ denote a random vector which has the stationary distribution of X_t and note that by equation (61) we have

$$\frac{1}{n}\sum_{t=1}^{n}\|\,\mathbb{E}_{\tilde{\theta}_n}X_t-\mathbb{E}_{\tilde{\theta}_n}\tilde{X}_t\|^2=O(n^{-1}).$$

Now let

$$\begin{aligned}
U_{n,t}&:=\left(\phi_k(A_{n,k\bullet}V_{n,t})+\varsigma_{k,1}A_{n,k\bullet}V_{n,t}+\varsigma_{k,2}\kappa(A_{n,k\bullet}V_{n,t})\right)^2\\
\tilde{U}_{n,t}&:=\left(\tilde{\varsigma}_{k,n,1}A_{n,k\bullet}V_{n,t}+\tilde{\varsigma}_{k,n,2}\kappa(A_{n,k\bullet}V_{n,t})\right)^2.
\end{aligned}$$

By Theorem 1 in Arnold (1985) and Markov's inequality, we have that $\max_{1\le t\le n}U_{n,t}=O_{P^n_{\tilde{\theta}_n}}(n^{1/p})$. Then

$$\frac{1}{n}\sum_{t=1}^{n}\|\,\mathbb{E}_{\tilde{\theta}_n}X_t-\mathbb{E}_{\tilde{\theta}_n}\tilde{X}_t\|^2U_{n,t}\le\max_{1\le t\le n}U_{n,t}\frac{1}{n}\sum_{t=1}^{n}\|\,\mathbb{E}_{\tilde{\theta}_n}X_t-\mathbb{E}_{\tilde{\theta}_n}\tilde{X}_t\|^2=O_{P^n_{\tilde{\theta}_n}}(n^{-1+1/p})=o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$

[51] Cf. footnote 47.

Additionally, by equation (61), Jensen's inequality, Lemma A.2 and Theorem 2 of Kanaya (2017)

$$\| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \bar{X}_n\|^2 \leq 2 \left[\left\| \frac{1}{n}\sum_{t=1}^n (X_t - \mathbb{E}_{\tilde{\theta}_n} X_t) \right\|^2 + \left\| \frac{1}{n}\sum_{t=1}^n (\mathbb{E}_{\tilde{\theta}_n} X_t - \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t) \right\|^2 \right] = O_{P^n_{\tilde{\theta}_n}}(n^{-1}) + O(n^{-1}),$$

hence

$$\frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \bar{X}_n\|^2 U_{n,t} = \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \bar{X}_n\|^2 \frac{1}{n}\sum_{t=1}^n U_{n,t} = O_{P^n_{\tilde{\theta}_n}}(n^{-1}) = o_{P^n_{\tilde{\theta}_n}}(\nu_n).$$

To complete the proof, it suffices to combine the above results with the observation that by Lemma A.12 and Theorem 1 of Arnold (1985)

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\tilde{\theta}_n} X_t - \bar{X}_n\|^2 \tilde{U}_{n,t} \\
&\quad \lesssim \tilde{\tau}^2_{k,n,1} \frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \bar{X}_n\|^2 (A_{n,k\bullet} V_{n,t})^2 + \tilde{\tau}^2_{k,n,1} \max_{1\leq t\leq n} (A_{n,k\bullet} V_{n,t})^2 \frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \mathbb{E}_{\tilde{\theta}_n} X_t\|^2 \\
&\qquad + \tilde{\tau}^2_{k,n,2} \frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \bar{X}_n\|^2 \kappa (A_{n,k\bullet} V_{n,t})^2 + \tilde{\tau}^2_{k,n,2} \max_{1\leq t\leq n} \kappa (A_{n,k\bullet} V_{n,t})^2 \frac{1}{n}\sum_{t=1}^n \| \mathbb{E}_{\tilde{\theta}_n} \tilde{X}_t - \mathbb{E}_{\tilde{\theta}_n} X_t\|^2 \\
&\quad = o_{P^n_{\tilde{\theta}_n}}(\nu_n).
\end{aligned}$$

□

A.4 Miscellaneous results

The results in this subsection are general results, which are useful in establishing the main results of the paper, but are not specific to the model under study.

PROPOSITION A.2 (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 finite measures such that* $\mu_n \xrightarrow{TV} \mu$ *(with* μ *a finite measure on* $(S, \mathcal{S})$*. 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 Corollary 2.3 of Feinberg et al. (2016) and the hypothesis that $\limsup_{n\to\infty} \int |f_n|^p \,\mathrm{d}\mu_n \leq \int |f|^p \,\mathrm{d}\mu < \infty$,

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

□

PROPOSITION A.3: *Let* $\{Z_{n,k}, \mathcal{F}_{n,k} : k \leq n, n \in \mathbb{N}\}$ *be a martingale difference array of* $L-$*dimensional random vectors, such that* $\Sigma_{n,k} := \mathbb{E}\left[Z_{n,k}Z'_{n,k}\right]$ *exists. Suppose that*

$$\frac{1}{n}\sum_{k=1}^n \Sigma_{n,k} \to \Sigma_\star, \tag{66}$$

with $\Sigma_\star$ positive semi-definite (and finite) and that for each $\varepsilon > 0$

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[\|Z_{n,k}\|^2\mathbf{1}\{\|Z_{n,k}\| \geq \varepsilon\sqrt{n}\}\right] \to 0. \tag{67}$$

Then

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Z_{n,k} \rightsquigarrow \mathcal{N}(0, \Sigma_\star).$$

Proof. Put $\xi_{n,k} \coloneqq Z_{n,k}/\sqrt{n}$ for $k \leq n$ and $\xi_{n,k} \coloneqq 0$ otherwise. Fix $a \in \mathbb{R}^L$. The adapted sequence $(a'\xi_{n,k}, \mathcal{F}_{n,k})_{k\in\mathbb{N}}$ is clearly a martingale difference sequence under our hypotheses. Moreover, the sums $\sum_{k=1}^{\infty} a'\xi_{n,k} = \sum_{k=1}^{n} a'\xi_{n,k}$ and $\sum_{k=1}^{\infty}\mathbb{E}[(a'\xi_{n,k})^2] = \sum_{k=1}^{n}\mathbb{E}[(a'\xi_{n,k})^2]$ trivially converge with probability 1 for each $n \in \mathbb{N}$. By linearity and continuity we have that

$$\sum_{k=1}^{\infty}\mathbb{E}[(a'\xi_{n,k})^2] = \sum_{k=1}^{n}\mathbb{E}[(a'\xi_{n,k})^2] = a'\left[\frac{1}{n}\sum_{k=1}^{n}\Sigma_{n,k}\right]a \to a'\Sigma_\star a \geq 0.$$

Next, suppose that $a \neq 0$ and let $\varepsilon > 0$. We have that $\{|a'Z_{n,k}| \geq \varepsilon\sqrt{n}\} \subset \{\|Z_{n,k}\| \geq \varepsilon\sqrt{n}/\|a\|\}$ and therefore

$$\sum_{k=1}^{\infty}\mathbb{E}\left[(a'\xi_{n,k})^2\mathbf{1}\{|a'\xi_{n,k}| \geq \varepsilon\}\right] \leq \|a\|^2\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[\|Z_{n,k}\|^2\mathbf{1}\{\|Z_{n,k}\| \geq \varepsilon\sqrt{n}/\|a\|\}\right] \to 0,$$

by assumption.[52] This establishes that the conditions of Theorem 18.1 of Billingsley (1999) are satisfied and hence

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} a'Z_{n,k} = \sum_{k=1}^{\infty} a'\xi_{n,k} \rightsquigarrow \mathcal{N}(0, a'\Sigma_\star a).$$

The claimed result then follows by an application of the Cramér-Wold theorem. □

Remark A.1: *Proposition A.3 is completely standard. It is recorded here because we have been unable to find a reference for a multivariate CLT for martingale difference arrays which permits a positive* semi-*definite limiting variance matrix.*

Theorem A.1 (Extended uniformly equicontinuous mapping): *Let (X, d_X) and (Y, d_Y) be separable metric spaces and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions from $X \to Y$ and $(g_n)_{n\in\mathbb{N}}$ a uniformly equicontinuous sequence of functions from $X \to Y$. Suppose that $x \mapsto d_Y(f_n(x), g_n(x))$ converges compactly to 0. If $(P_n)_{n\in\mathbb{N}}$ and $(Q_n)_{n\in\mathbb{N}}$ are sequences of laws on X such that (i) $(P_n)_{n\in\mathbb{N}}$ is uniformly tight and (ii) $d_{BL}(P_n, Q_n) \to 0$, then $d_{BL}(\tilde{P}_n, \tilde{Q}_n) \to 0$ for $\tilde{P}_n \coloneqq P_n \circ f_n^{-1}$ and $\tilde{Q}_n \coloneqq Q_n \circ g_n^{-1}$.*

Proof. By Theorem 11.7.1 in Dudley (2002), there exist on some probability space X-valued random variables X_n and Y_n such that $X_n \sim P_n$ and $Y_n \sim Q_n$ and $d_X(X_n, Y_n) \to 0$ in probability. By the triangle inequality

$$d_Y(f_n(X_n), g_n(Y_n)) \leq d_Y(f_n(X_n), g_n(X_n)) + d_Y(g_n(X_n), g_n(Y_n)).$$

By uniform equicontinuity of $(g_n)_{n\in\mathbb{N}}$, $d_Y(g_n(X_n), g_n(Y_n)) \to 0$ in probability. Let $\delta,\ \varepsilon > 0$ be given and choose a compact K such that (each) $P_nK > 1-\varepsilon$. The compact convergence ensures that for all sufficiently large n, $\sup_{x\in K} d_Y(f_n(x), g_n(x)) < \delta$. Hence, for all such n,

$$\mathrm{P}\left(d_Y(f_n(X_n), g_n(X_n)) > \delta\right) \leq \mathrm{P}\left(X_n \notin K\right) = P_nK^{\complement} \leq \epsilon.$$

[52] In the case that $a = 0$ this limit trivially holds.

It follows that $d_Y(f_n(X_n), g_n(Y_n)) \to 0$ in probability and the conclusion follows by applying Theorem 11.7.1 in Dudley (2002) once more. □

THEOREM A.2 (Uniform Delta-method): *Let U and V be normed linear spaces and $\phi : U_\phi \to V$ (with $U_\phi \subset U$). Let $(r_n)_{n\in\mathbb{N}}$ be a sequence of constants with $r_n \to \infty$, $(X_n)_{n\in\mathbb{N}}$ a sequence of U_ϕ-valued random variables, $(\theta_n)_{n\in\mathbb{N}} \subset U_\phi$ and $(P_n)_{n\in\mathbb{N}}$, $(Q_n)_{n\in\mathbb{N}}$ sequences of laws on U with (each) $P_n U_0 = 1$ for a separable $U_0 \subset U$. Suppose that (i) ϕ is Hadamard differentiable tangentially to U_0, uniformly along $(\theta_n)_{n\in\mathbb{N}}$, with derivative ϕ'_θ, (ii) $T_n := r_n(X_n - \theta_n) \sim P_n$ where $(P_n)_{n\in\mathbb{N}}$ is uniformly tight, (iii) $d_{BL}(P_n, Q_n) \to 0$ and (iv) $(\phi'_{\theta_n})_{n\in\mathbb{N}}$ is uniformly equicontinuous. Then,*

$$d_{BL}\left(\mathscr{L}\left(r_n\left(\phi(X_n) - \phi(\theta_n)\right)\right),\ Q_n \circ [\phi'_{\theta_n}]^{-1}\right) \to 0. \tag{68}$$

Proof. Define $f_n(h) := r_n\left(\phi(\theta_n + r_n^{-1}h) - \phi(\theta_n)\right)$ and $g_n(h) := \phi'_{\theta_n}(h)$. By our uniform differentiability assumption, for any compact $K \subset U_0$ we have

$$\sup_{h\in K}\left\|r_n\left(\phi(\theta_n + r_n^{-1}h) - \phi(\theta_n)\right) - \phi'_{\theta_n}(h)\right\| \to 0,$$

and so $h \mapsto \|f_n(g) - g_n(h)\|$ converges compactly to 0 on U_0.[53] This fact and (ii) - (iv) allows the application of Theorem A.1 to conclude (68).[54] □

REMARK A.2: *Since Hadamard derivatives are bounded linear maps by definition, a sufficient condition for the uniform equicontinuity of $(\phi'_{\theta_n})_{n\in\mathbb{N}}$ is that $\sup_{n\in\mathbb{N}}\|\phi'_{\theta_n}\| < \infty$, i.e. their operator norms are uniformly bounded. This ensures that each ϕ'_{θ_n} is Lipschitz with Lipschitz constant $\sup_{n\in\mathbb{N}}\|\phi'_{\theta_n}\|$ and hence the collection is uniformly equicontinuous.*

[53] Cf. e.g. pp. 453 – 455 in Bickel et al. (1998)

[54] The image of a separable space under a continuous function is separable, cf. Theorem 16.4(a) in Willard (1970).

References

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

Andrews, D. W. K. and Cheng, X. (2012). Estimation and inference with weak, semi-strong and strong identification. *Econometrica*, 80(5).

Andrews, I. and Mikusheva, A. (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, 84(4).

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

Ansley, C. F. and Kohn, R. (1986). A note on reparameterizing a vector autoregressive moving average model to enforce stationarity. *Journal of Statistical Computation and Simulation*, 24(2):99–106.

Arnold, B. C. (1985). p-norm bounds on the expectation of the maximum of a possibly dependent sample. *Journal of Multivariate Analysis*, 17(3):316–332.

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.

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley.

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.

Bogachev, V. I. (2007). *Measure Theory.* Springer Berlin Heidelberg.

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 a svar identified by non-gaussianity. *working paper.*

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer Series in Statistics. Springer, 2 edition.

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.

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.

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

Drabek, P. and Milota, J. (2007). *Methods of Nonlinear Analysis: Applications to Differential Equations.* Birkhäuser Advanced Texts Basler Lehrbücher. Birkhäuser Basel.

Drautzburg, T. and Wright, J. H. (2021). Refining set-identification in vars through independence. *Working paper.*

Dudley, R. M. (2002). *Real Analysis and Probability.* Cambridge University Press, Cambridge, UK.

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

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.

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.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York, NY, USA.

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.

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):1121–1153.

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.

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

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.

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 structural 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. (2022a). Robust inference for non-gaussian linear simultaneous equations models. *Working Paper*.

Lee, A. and Mesters, G. (2022b). Supplement to “robust inference for non-gaussian linear simultaneous equations models”. *Working Paper*.

Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis*, 26(5):669–689.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.

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.

Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press, 2 edition.

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 1–36.

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

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20 – 71.

Rudin, W. (1987). *Real & Complex Analysis*. McGraw Hill.

Rudin, W. (1991). *Functional analysis*. McGraw Hill, 2 edition.

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

Serfozo, R. (1982). Convergence of lebesgue integrals with varying measures. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 44(3):380–402.

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).

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. (1988). *Statistical Estimation in Large Parameter Spaces.* Number 44 in CWI Tracts. Centrum voor Wiskunde en Informatica, Amsterdam.

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.

Willard, S. (1970). *General Topology.* Addison-Wesley.

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