Statistical identification of simultaneous equations models

(Former title: Consistent statistical identification of SVARs under (co)heteroskedasticity of unknown form)

Helmut Herwartz* Shu Wang *

August 26, 2024

Abstract

Recent development in identification methods utilizing higher-order moments have advanced structural analysis in macroeconomics. This paper reviews prevailing approaches, including those using time-varying volatility or assuming independence under non-Gaussianity, and highlights their limitations in addressing co-heteroskedasticity. We introduce a novel identification scheme that accommodates latent shocks with unknown second-order moment features and propose a new estimator, which is consistent regardless of whether shocks are (co-)heteroskedastic or homoskedastic. We study its asymptotic properties, and demonstrate superior finite-sample performance through simulations. Applying this method to US monetary policy model yields variance estimators that align well with benchmark macroeconomic uncertainty measures.

Keywords: Simultaneous equations models, structural VAR, heteroskedasticity, time-varying volatility, identification, monetary policy.
JEL Classification: C14, C32, E52.

*Chair of Econometrics, University of Göttingen, Humboldtallee 3, D-37073 Göttingen, Germany. Corresponding author: Shu Wang, e-mail: shu.wang@uni-goettingen.de. We thank Dante Amengual, Bertille Antoine, Jörg Breitung, James Duffy, Gabriele Fiorentini, Raffaella Giacomini, Carsten Jentsch, Sascha Keweloh, Christian M. Hafner, Helmut Lütkepohl and participants of the 34th (EC)^2 and the 2024 RCEA conferences for insightful discussion and comments. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (HE 2188/17-1). Any errors and omissions should be regarded as those of the authors.

1 Introduction

In recent years, a novel branch of literature has emerged, focusing on the statistical identification of structural simultaneous equations models (SEMs), such as vector autoregressive models (SVARs) and similar multivariate dynamic systems that face the problem of simultaneous causality. This line of research aims to achieve identification and subsequent structural analysis by imposing higher-order moment conditions on the data-generating process (DGP). These statistical identification approaches have garnered attention due to their intuitive appeal and flexibility, especially in scenarios with limited theoretical or institutional knowledge and when external instruments are either scant, weak, or unavailable. Within this literature, two main branches of statistical identification have emerged. The first branch encompasses identification techniques that build upon informative time-varying volatility of the data (see, e.g., Sentana and Fiorentini 2001, Rigobon 2003, Lewis 2021). The second branch of this literature explores the statistical independence of latent shocks and the potential of independent component analysis (ICA) as a promising tool for identification (see, e.g., Lanne et al. 2017, Gouriéroux et al. 2017, Guay 2021). Notably, statistical identifications have proven useful in analyzing small to medium-sized financial and macroeconometric models, contributing valuable insights into the causal relationships within these models and deepening the understanding of policy interventions and transmission mechanisms of structural innovations.

In this paper, we begin by reviewing the identifying assumptions underlying both approaches: While identifications based on time-varying volatility require that the sample paths of the volatility processes must be linearly independent and at most one path is constant, ICA-based identifications assume that the latent shocks are mutually independent and at most one shock is Gaussian. If one of these conditions is satisfied, the structural mixing parameters are locally identified, up to column permutations and sign flips.[1] We discuss the connection between both types of identification assumptions and when the violation of one assumption can invalidate the other. In particular, we show that *both* identification approaches exclude an empirically relevant scenario commonly observed in macroeconomic data, where shocks share common volatility processes, known as *co-heteroskedasticity* (Montiel Olea et al. 2022, Lewis 2024). Co-heteroskedasticity is rather a common specification in many financial and

[1]Statistically identified shocks do not necessarily allow for a direct economic interpretation. The so-called shock labeling problem is typically resolved through supplementary economic reasoning and careful assessment of estimated structural model implications, such as impact multipliers, impulse response functions and historical narratives (see, e.g., Herwartz and Lütkepohl 2014, Lewis 2021).

macroeconomic models, given the empirical evidence suggesting that volatilities of certain shocks may co-move, for instance, during the transition to the so-called *Great Moderation* era or periods of economic downturns and financial distress. Another example includes shocks to precautionary demand for oil, which may exhibit increased volatility amidst heightened uncertainty surrounding crude oil supply due to factors such as rising geopolitical tensions, warfare, and sanctions. In these cases, the second-order moment dependence among the shocks leads to violation of conditions for identification approaches both via time-varying volatility and independence under non-Gaussianity, and thus renders them under-identified.

To overcome the identification problem in the presence of co-heteroskedastic shocks, this paper suggests a novel identification scheme allowing latent shocks to have second-order moment features of unknown form. Sufficient conditions for local identification are provided. We demonstrate that the commonly adopted ICA-based methods relying on independent shocks are nested within our identification scheme. We give conditions under which the independence assumption imposed on the latent shocks remains valid and the standard techniques relying on it remain operational. We further propose a non-parametric kernel-based ML (KML) approach, which adjusts each individual orthogonalized observation by its corresponding standard deviation obtained from a kernel smoother, without imposing specific parametric forms for the underlying variance process. We study the asymptotic properties of the proposed estimator and investigate its finite-sample performance through Monte Carlo experiments. Specifically, by placing a continuity and mild integrability condition on the limiting variance processes, we show that the suggested estimator is consistent and asymptotically normal under standard regularity conditions.

Monte Carlo simulations highlight the superior finite-sample performance of the suggested estimator in various DGPs and a range of stochastic and deterministic volatility processes. In contrast, ML methods using fixed densities and approaches imposing higher-order moment restrictions might be vulnerable to pseudo density misspecification and outliers, respectively, in specific data generation scenarios. Particularly noteworthy is the comparison with an unfeasible estimator that employs shocks scaled with their true standard deviations, revealing that the non-parametric approach to variance estimation involves only a minimal loss in efficiency, which becomes negligible in large samples.

Our study contributes to the growing literature on identification approaches that rely on higher-order moments and is one of the first to explicitly address the issue of co-heteroskedasticity, which has been overlooked in most statistical identification procedures. It is noteworthy that certain higher-order moment or cumulant conditions implied by mutual independence remain valid even in the presence of co-heteroskedasticity, allowing methods based on these restrictions to remain applicable (see Lanne and Luoto 2021, Mesters and Zwiernik 2022). For instance, zero asymmetric co-skewness and co-kurtosis are valid restrictions. However, under co-heteroskedasticity, these conditions alone, without assuming independence, may not suffice for identification (see Keweloh 2021, Lewis 2024). Additionally, our simulation study demonstrates that estimation of higher-order moment statistics is challenging in small samples and can be sensitive to heavy-tailed source distributions and distorted by outliers in the data (see also Keweloh 2023 for discussion of GMM estimation). Importantly, the procedure proposed in this paper primarily utilizes information from non-Gaussianity rather than heteroskedasticity for identification. Therefore, the suggested estimator is consistent regardless of whether the shocks are heteroskedastic, co-heteroskedastic, or homoskedastic.

When applied to a small-scale US monetary policy analysis, the estimator identifies demand, supply, and monetary policy shocks in alignment with their well-established features regarding their effects on economic activity, inflation, and treasury yields. Our estimates indicate that an unexpected monetary tightening can have sustained effects on controlling medium to long-term inflation, while having only transitory adverse effects on economic activity. In addition, we demonstrate sizable co-movement in the second-order moments of the identified shocks. Furthermore, the estimated variance of structural shocks exhibits strong parallels with the benchmark information-rich measure of macroeconomic uncertainty proposed by Jurado et al. (2015). This outcome stands as an informal yet significant affirmation of the structural model's findings.

The remainder of this study is structured as follows. The next Section outlines the model, identification problems under (co-)heteroskedasticity and introduces the novel identification approach and a new estimator. Section 3 provides simulation-based insights into the finite-sample performance of the suggested estimator. In Section 4, we apply the approach to a small scale US monetary policy model. Section 5 summarizes and concludes. An Appendix collects proofs of the theorems. The Online Appendices document additional simulation results (OA A) and supplement the empirical analysis with model diagnostics (tests for remaining residual correlation, non-Gaussianity, and fundamentalness and independence of structural shocks, OA B), and further structural impulse response estimations (OA C). Throughout the paper, $X_T \xrightarrow{p} X$ denotes convergence in probability and $X_T \Rightarrow X$ denotes convergence in law as $T \to \infty$. We use $X_{\lfloor sT \rfloor} \Rightarrow X(s)$ to denote weak convergence in $\mathcal{C}[0,1]$ with respect to the uniform metric, where $\lfloor sT \rfloor$ is the integer part of sT. Moreover, $||\cdot||$ and $\|\cdot\|_p$ denotes the Euclidean and the L_p norm. The function δ_X takes value one if statement X is true and zero otherwise. The term 'volatilities' refers to both variances and standard deviations unless explicitly defined otherwise.

2 Methodology

In this section, we outline the general model framework, review the identifying assumptions underlying the two most popular identification methods via higher-order moments. We discuss their limitations under co-heteroskedasticity and propose suitable conditions allowing for (local) point identification and consistent estimation. We introduce a novel estimator is and study its asymptotic properties.

2.1 General model specification

A fundamental challenge in causal analysis of multivariate time series lies in uncovering their structural representation from the data. Typically, identifying this representation involves solving a system of equations:

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

where ξ_t is a vector of N latent factors or shocks, which are serially uncorrelated and have a diagonal covariance matrix $\Sigma_t = \mathrm{diag}\,(\sigma_{1t}^2, \ldots, \sigma_{Nt}^2)$, which may or may not be time-varying. The latent shocks ξ_t are mapped to the N observables collected in x_t through a non-singular matrix B. The system in (1) can be written as:

$$\mathrm{x}_t = B\xi_t = B\Sigma_t^{1/2}\eta_t,\ \eta_t \overset{iid}{\sim} (\mathbf{0}, I_N), \tag{2}$$

where the random vector η_t is independent from the volatility processes in $\Sigma_t^{1/2}$. Let $\{\mathcal{F}_t\}_{t\geq 1}$ be a natural filtration generated by $\mathcal{F}_t = \boldsymbol{\sigma}(\eta_{t-j}, \Sigma_{t-j+1}, j \geq 0)$, to which ξ_t is measurable for all t. It follows that the structural shocks ξ_t are martingale difference sequences with $\mathbb{E}[\xi_t|\mathcal{F}_{t-1}] = \mathbf{0}$ and $\mathbb{E}[\xi_t\xi_t^\top|\mathcal{F}_{t-1}] = \Sigma_t$. This specification can accommodate a GARCH-type volatility process, where $\mathcal{F}_t$ simplifies to $\boldsymbol{\sigma}(\eta_{t-j}, j \geq 0)$. It also allows for a stochastic volatility process, where elements in Σ_t are driven by other innovations. Assume $\mathbb{E}[\Sigma_t] = I_N$, the vector of observables has a zero mean and non-diagonal covariance matrix given by

$$\mathbb{E}[\mathrm{x}_t\mathrm{x}_t^\top] = B\mathbb{E}[\Sigma_t^{1/2}\eta_t\eta_t^\top\Sigma_t^{1/2}]B^\top = B\mathbb{E}[\Sigma_t]B^\top = BB^\top = \Omega,$$

where the penultimate equality holds due to independence between Σ_t and η_t.[2]

[2] The imposition of unit variance for the latent shocks is made to normalize the size of the shocks and is applied without any loss of generality. Alternative normalizations have also been proposed. For

The SEM as outlined above finds many applications in empirical finance and macroeconomics. For example, x_t can represent the returns of N speculative assets, and the structural parameters in B provide insights into the contribution of each shock in ξ_t to the asset returns. In factor models, x_t is a potential rotation of the vector of orthogonalized factors. In structural VAR models (SVARs), x_t is the residual vector from the reduced-form model. In the following, we illustrate our approach by viewing SVAR analysis as the primary application and provide a short description of this framework.

Let y_t be the $N \times 1$ vector of observable variables. In its reduced form and adjusted for deterministic terms, the VAR of order p reads as

$$\mathrm{x}_t \equiv A(L)y_t, \tag{3}$$

where L is the lag-operator, $A(z) = \left(I_N - \sum_{j=1}^{p} A_j z^j\right),\ z \in \mathbb{C}$, is the reverse characteristic polynomial. Under the weak stationarity condition, assuming $\det A(z) \neq 0$ for $|z| \leq 1$, y_t has a Wold moving average (MA) representation. Such a representation is of great interest, as it reveals the dynamic effects of structural shocks on the observable system through structural impulse response functions (IRFs)

$$y_t = \sum_{j=0}^{\infty} \Phi_j \xi_{t-j} = \Phi(L)\xi_t,$$

where $\Phi(L) = A(1)^{-1}B$ with $\Phi_0 = B$. Hence, the 'impact multipliers' in B provide valuable insights into how the variables in y_t contemporaneously react to exogenous changes in ξ_t. While the reduced-form parameters in (3) (i.e., AR coefficients A_j's and unconditional covariance Ω) can be consistently estimated, it is well-known that structural parameters in B lack identification from the second moments of the data.

While classical identification techniques typically rely on economic theory and institutional knowledge that can be used to develop exclusion (e.g. Sims 1980, Blanchard and Quah 1989) and sign restrictions (e.g. Faust 1998, Uhlig 2005), or instruments for structural shocks (Mertens and Ravn 2013, Stock and Watson 2012), statistical identification approaches that exploit higher-order moment information have gained attention in recent years (see Montiel Olea et al. 2022, Lewis 2024 for a review of recent literature). We briefly outline these approaches and their underlying the identifying assumptions.

example, one approach involves restricting the diagonal elements of matrix B to be unity, denoted as $B^{\dagger}$, while allowing $\mathbb{E}[\Sigma_t] = \Sigma$ to have flexible diagonal elements. In this case, $B = B^{\dagger}\Sigma^{1/2}$ (see Gouriéroux et al. 2017 and references therein for a discussion).

2.2 Statistical identifications via higher-order moments

Beginning with pioneering works by Sentana and Fiorentini (2001) and Rigobon (2003), which shed light on the identifiability under time-varying variances of latent shocks, subsequent studies have furthered the approach by exploring various forms for the volatility processes, including exogenous shifts (Lanne and Lütkepohl 2008), smooth transitions (Lütkepohl and Netšunajev 2017) and Markov switching (Lanne et al. 2010) between variance regimes, as well as generalized autoregressive conditional heteroskedasticity (GARCH) specifications (Normadin and Phaneuf 2004) and stochastic autoregressive volatility (Bertsche and Braun 2022). In addition, a growing body of work explores the potential of ICA as a promising tool for achieving identification. Within the realm of ICA-based identification, a range of methodologies has emerged, including methods of maximum likelihood (ML) assuming statistical independence among the structural shocks (Lanne et al. 2017, Gouriéroux et al. 2017, Fiorentini and Sentana 2023, Hafner et al. 2024, Jarociński 2024), generalized methods of moments (GMM) or classical minimum distance (CMD) estimations relying on (weaker) mean/moment independence (Lanne and Luoto 2021, Keweloh 2021, Hafner et al. 2022, Guay 2021), and CMD or Hodges-Lehmann (HL) methods utilizing non-parametric independence measures (Drautzburg and Wright 2023, Herwartz and Wang 2023, 2024).

While ICA-based methods often rely on similar assumptions, conditions for identifications via time-varying volatility may have different forms depending on the specific volatility processes assumed. We present here a general condition for the sample paths of the volatility processes, which covers special cases such as non-proportional covariance changes (Rigobon 2003) or linearly dependent stochastic autoregressive volatilities (Bertsche and Braun 2022) and is similar to the condition outlined in Sentana and Fiorentini (2001) for factor identifications. Let $\boldsymbol{\sigma}_t^2 = (\sigma_{1t}^2, \ldots, \sigma_{Nt}^2)^\top$ denote the vector collecting the marginal variances of the shocks and define a $(T \times N)$ matrix $\boldsymbol{S}_T = (\boldsymbol{\sigma}_1^2, \ldots, \boldsymbol{\sigma}_T^2)^\top$. The identifying assumptions underlying the two most prominent statistical identification schemes can be summarized as follows:

Theorem 1. *If one of the following conditions is satisfied:*

(i) the components in the random vector ξ_t are independent and at most one component is Gaussian;

(ii) the sample paths of the processes in $\boldsymbol{\sigma}_t^2$ are linearly independent and at most one path is constant over time, i.e., rank $\boldsymbol{S}_T = N$,

then the matrix B in (1) *and* (2) *is identified up to a right-multiplication by $\mathcal{DP}$, where*

$\mathcal{D}$ is diagonal and $\mathcal{P}$ is a permutation matrix. In other words, B is identified up to column permutations and sign flips.

Remark 1. *For a full (local) identification of the matrix B, it is sufficient to satisfy one of the conditions in Theorem 1. If both (i) and (ii) are imposed, the system in* (1) *and* (2) *is over-identified and condition (i) or (ii) can be tested conditional on the other identifying assumption.*

A subtle point contained in Remark 1 is that heteroskedasticity of the shocks does not prohibit them to be statistically independent. We will discuss this more explicitly in the next section. However, violation of condition (ii) can invalidate condition (i). Notably, the identifying assumption (ii) in Theorem 1 can be violated in two ways:

Remark 2. *rank $\boldsymbol{S}_T < N$, if one of the following statements holds:*

(a) $\exists$ at least two $j \in \{1, \ldots, N\}$ for $N > 2$, s.t. $\sigma^2_{jt} = 1$, a.s. $\forall t$;

(b) $\exists$ $i, j \in \{1, \ldots, N\}$, $i \neq j$ and $\mathbb{P}(\sigma^2_{jt} \neq 1) > 0$, s.t., $\exists \boldsymbol{c} \in \mathbb{R}^2$, $\boldsymbol{c} \neq \boldsymbol{0}$, with $(\sigma^2_{it}, \sigma^2_{jt})^\top \boldsymbol{c} = 0$, a.s. $\forall t$.

In other words, if at least two shocks are either (a) homoskedastic or (b) co-heteroskedastic, condition (ii) in Theorem 1 is violated. While case (a) does not invalidate the independence assumption (i) in Theorem 1, co-heteroskedasticity of the shocks will render them dependent in their second-order moment:

Remark 3. *In case (b) of Remark 2, both conditions (i) and (ii) in Theorem 1 are violated and matrix B is under-identified.*

Therefore, co-heteroskedasticity presents a critical challenge to identification approaches based on (i) assuming statistical independence under non-Gaussianity and (ii) time-varying volatility of the shocks. Specifically, co-movement of shock variances during periods of economic recession or financial distress can lead to weak identification due to weak dependence in second-order moments of the shocks or near rank deficiency of the matrix $\boldsymbol{S}_T$. In the following section, we discuss the identification problem in the presence of co-heteroskedasticity in more derail and propose a novel identification scheme that accommodates latent shocks with second-order moment features of unknown form.

2.3 Identification under (co-)heteroskedasticity

To illustrate the identification problem, consider the scenario of two volatility regimes, where $\Sigma_t = \Sigma_I$ for $t \in$ regime I and $\Sigma_t = \Sigma_{II} \neq \Sigma_I$ for $t \in$ regime II. Condition (ii) in Theorem 1 implies that $\nexists c \in \mathbb{R}$ such that $\Sigma_{II} = c\Sigma_I$. In other words, changes in

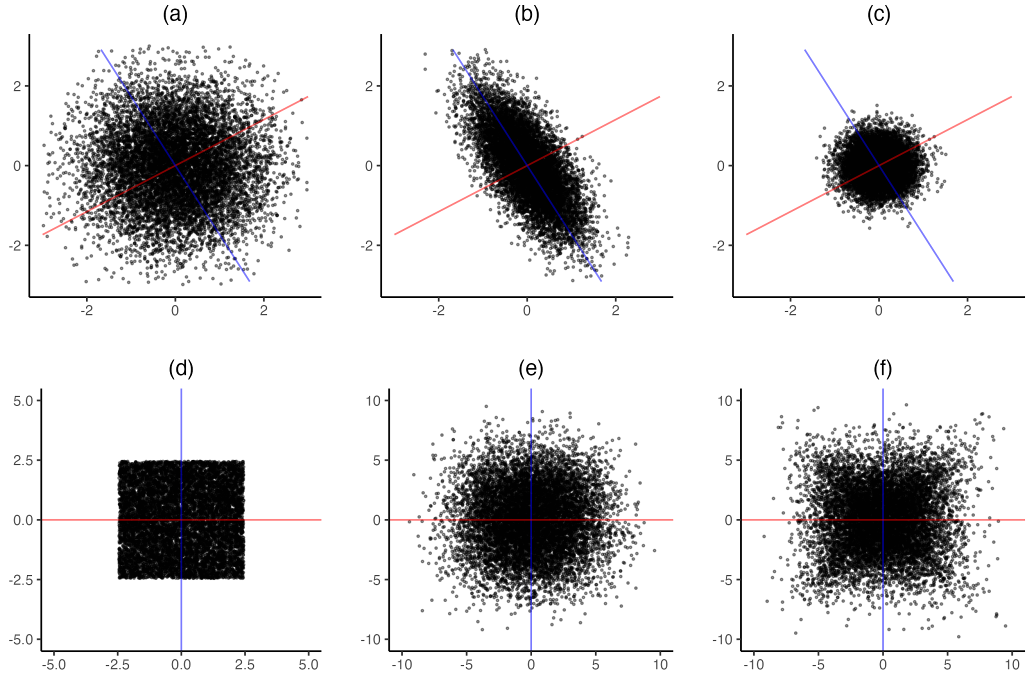

Figure 1: Joint distribution of $(\xi_{1t}, \xi_{2t})'$ with time-varying variances. Red and blue lines indicate orientations of marginals of ξ_{1t} and ξ_{2t}. Upper panel: Gaussian shocks under variance change; lower: uniform shocks with χ^2 random variances. From left to right: Homoskedasticity, (independent) heteroskedasticity, and co-heteroskedasticity.

variances can not be proportional. As depicted in panels (a) – (b) of Figure 1, a non-proportional change reveals orientations of the marginal distributions. However, if both structural shocks undergo a common variance change, the joint distribution will expand or contract in its scale along all directions, rendering the shocks under-identified. This is illustrated in panel (c) of Figure 1.[3]

Instead, if the analyst pursues a statistical identification relying on independence under non-Gaussianity, the result on local identifiability under condition (i) in Theorem 1 holds under homoskedasticity (panel (d)) and heteroskedasticity (panel (e) of Figure 1), under the additional assumption on the variance processes stated in Theorem 3. However, if the shocks exhibit co-heteroskedasticity, condition (i) will be violated. For simplicity, suppose $\boldsymbol{c} = (1, -1)^\top$, case (b) of Remark 2 implies – for a scalar positive random variable σ_t –

$$\xi_{it} = \sigma_t \eta_{it} \quad \text{and} \quad \xi_{jt} = \sigma_t \eta_{jt}.$$

[3]More broadly, common variance shifts may result in the matrix $\Sigma_{II}\Sigma_I^{-1}$ having diagonal elements that are close to each other or even identical, which leads to weak or under-identification of the corresponding shocks and their structural impact multipliers in the respective columns of matrix B (see also Rigobon 2003, Lanne et al. 2010).

In such a case, ξ_{it} and ξ_{jt} remain uncorrelated obeying an interpretation as structural shocks. They are mean independent in the first but no longer in the second order moment:

$$\mathbb{E}[\xi_{it}\xi_{jt}] = \mathbb{E}[\sigma_t^2]\mathbb{E}[\eta_{it}\eta_{jt}] = 0,\ \mathbb{E}[\xi_{it}|\xi_{jt}] = \sigma_t\mathbb{E}[\eta_{it}] = 0,\ \mathbb{E}[\xi_{it}^2|\xi_{jt}] = \sigma_t^2\mathbb{E}[\eta_{it}^2] = \sigma_t^2,$$

which follows from the fact that η_{it} and η_{jt} have a diagonal covariance matrix and are independent of σ_t. This is illustrated in panel (f) of Figure 1, where the common variance σ_t^2 is simulated from a χ^2 distribution.

While co-heteroskedasticity invalidates both conditions (i) and (ii), some weaker higher-order moment conditions (e.g., asymmetric co-skewness $\mathbb{E}[\xi_{it}^2\xi_{jt}] = 0$) are still valid, and thus, approaches that exploit these restrictions for identification purposes will remain operational (Lanne and Luoto 2021, Mesters and Zwiernik 2022). However, estimation of higher-order moment statistics and especially the asymptotically efficient GMM weighting matrix can be challenging due to outliers in the data and may suffer from poor finite-sample performance for heavy-tailed source distributions as discussed in Keweloh (2023) and further illustrated in Section 3.[4]

If it is possible to appropriately adjust the structural shocks in ξ_t based on their specific or shared variance processes, i.e., $\eta_{it} = \xi_{it}/\sigma_{it},\ i = 1, \dots, N$, the elements of the resulting vector η_t can be identified under the assumption that they are statistically independent and at most one element is Gaussian. We replace (i) in Theorem 1 with the following assumption on η_t:

Assumption A. *The components in the random vector η_t are independent and at most one component is Gaussian.*

We establish the identifiability through the following theorem:

Theorem 2. *Under Assumption A, the matrix B in* (1) *and* (2) *is identified up to a right-multiplication by $\mathcal{DP}$.*

As discussed above, independence of components in η_t (i.e., Assumption A) does not in general imply independence of structural shocks in ξ_t (i.e., (i) in Theorem 1). However, in cases where the system is almost homoskedastic or the variance processes of distinct shocks are independent, which is stronger than condition (ii) in Theorem 1, Assumption A does imply independent shocks. In these scenarios, the identifying assumption (i) in Theorem 1 is valid and statistical identifications relying on it remain operational.

[4]Keweloh (2023) illustrates that imposing mutual and serial independence of the shocks can lead to a reduction of both the order and the number of higher-order moments in the weighting matrix and thus improve the finite-sample performance of GMM estimation.

Theorem 3. *Under Assumption A, if one of the following conditions holds:*

(i)* $\exists$ *at most one* $j \in \{1, \dots, N\}$*, s.t.* $\mathbb{P}(\sigma_{jt}^2 \neq 1) > 0$*,* $\forall t$*;*

(ii) stochastic processes in* $\boldsymbol{\sigma}_t^2$ *are independent,*

then condition (i) *in Theorem 1 is satisfied.*

2.4 Estimation under (co-)heteroskedasticity

Let $\Omega^{1/2} := \Gamma \Lambda^{1/2} \Gamma^\top$ denote the matrix square root of Ω. Here, $\Lambda^{1/2}$ is a diagonal matrix with the square root of the eigenvalues of Ω on the main diagonal, and Γ is the matrix of the eigenvectors. Left-multiplying equation (2) by the inverse of $\Omega^{1/2}$ yields

$$\mathrm{z}_t := \Omega^{-1/2} \mathrm{x}_t = Q \Sigma_t^{1/2} \eta_t, \tag{4}$$

where Q is orthonormal such that $QQ^\top = I_N$ and $|\det Q| = 1$. It can be easily verified that z_t is unconditionally white, i.e., $\mathrm{z}_t \sim (\mathbf{0}, I_N)$ and the structural parameter matrix B can be expressed as $B = \Omega^{1/2} Q$. Prewhitening is a necessary step for consistent estimation in many ICA techniques (see, e.g., Hyvärinen 1999, Gouriéroux et al. 2017), since it effectively reduces the number of free parameters from N^2 to $N(N-1)/2$. The space of the parameter of interest in Q is restricted to be the set of all N-dimensional orthonormal matrices $\mathcal{O}(N) = \{Q = (q_1, \dots, q_N) : q_i \in \mathbb{R}^N, q_i^\top q_j = \delta_{i=j}, \forall i, j = 1, \dots, N\}$. Let the orthonormal matrix Q be parameterized by a vector $\theta \in \boldsymbol{\Theta}$, we define $\xi_t(\theta) := Q(\theta)^\top \mathrm{z}_t$, $\xi_{it}(\theta) := e_i^\top \xi_t(\theta)$, $\eta_{it}(\theta) = \xi_{it}(\theta)/\sigma_{it}$, with e_i being the i-th column of an identity matrix with conformable dimension. One particularly interesting parameterization of Q under the constraint $Q \in \mathcal{O}(N)$ is given by a sequence of Givens rotation matrices (see Gouriéroux et al. 2017). A typical choice in the literature has the following form:

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

where θ_n for $n = 1, \dots, N(N-1)/2$ are rotation angles and $\mathcal{G}_{i,j}(\theta_n)$ is the Givens matrix that rotates the subspace spanned by axes i and j while holding other axes fixed.[5] Let $\theta_0 \in \boldsymbol{\Theta}$ be the true parameter vector, $\eta_t := \eta_t(\theta_0)$ be the true independent components and f_{i,θ_0} its pdf. Under the identifiability stated in Theorem 2, the log-likelihood is given by $\sum_{t=1}^T \sum_{i=1}^N (\log f_{i,\theta_0}(\eta_{it}(\theta)) - \log \sigma_{it})$. Because the true density

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

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

functions f_{i,θ_0}'s are rarely known and estimation based on fixed pseudo densities can suffer from missspecification, Hafner et al. (2024) propose the KML approach that maximizes an estimated log-likelihood

$$\tilde{l}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\log \hat{f}_{i,\theta}\left(\eta_{it}(\theta)\right) - \log \sigma_{it}\right), \tag{5}$$

where $\hat{f}_{i,\theta}$ is a kernel density estimate

$$\hat{f}_{i,\theta}(\eta_{it}(\theta)) = \frac{1}{Th_f}\sum_{k=1}^{T} K_f\left(\frac{\eta_{it}(\theta) - \eta_{ik}(\theta)}{h_f}\right),\ i = 1, \ldots, N,$$

where $K_f : \mathbb{R} \mapsto \mathbb{R}_0^+$ is a bounded kernel function and h_f is a bandwidth parameter. Essentially, the KML approach replaces $f_{i,\theta}$ in the estimation equation (see equation (8) and Lemma 1) by a kernel estimate, which is consistent under mild assumptions about the true density function and standard conditions for K_f and h_f. Hafner et al. (2024) shown by simulation that the KML approach has favorable finite-sample performance and is robust to a large variety of distributions that exhibit different forms of non-Gaussianity, including those close to a Gaussian distribution.

Notably, both identification and estimation hinge on the knowledge about Σ_t. Unfortunately, the structural shocks are latent and their variance processes are typically unknown, rendering the aforementioned identification procedure infeasible in practice. This is especially challenging without assuming that variance processes follow specific parametric forms. However, consistent non-parametric variance estimates can be obtained under the mild assumption that the limiting variance process is continuous.If that process is bounded away from zero, by a continuous mapping theorem, the independent components adjusted based on the estimated variance $\hat{\eta}_{it}(\theta) = \xi_{it}(\theta)/\hat{\sigma}_{it}$ will be close to $\eta_{it}(\theta)$ as n becomes large. In order to consistently estimate the variance process through a non-parametric kernel method, we make the following assumptions concerning unconditional moments of components in η_t and the volatility processes.

Assumption B.
(i) For some $p > 1$ and for all $i, j, k, l \in \{1, \ldots, N\}$, $\mathbb{E}[\eta_{it}^p \eta_{jt}^p \eta_{kt}^p \eta_{lt}^p] \leq M_\eta < \infty$, $\mathbb{E}[\sigma_{it}^p \sigma_{jt}^p \sigma_{kt}^p \sigma_{lt}^p] \leq M_\sigma < \infty$,
(ii) For all $i \in \{1, \ldots, N\}$, $\sigma_{i,\lfloor sT \rfloor} \Rightarrow \sigma_i(s)$ as $T \to \infty$ with $s \in [0, 1]$, where $\sigma_i(s)$ has continuous sample paths a.s. and $\exists \sigma_-, \sigma_+$, s.t. $\sup_{s\in[0,1]} \sigma_i(s) \leq \sigma_+ < \infty$, $\inf_{s\in[0,1]} \sigma_i(s) \geq \sigma_- > 0$, $\forall s$ and $\int_0^1 \sigma_i^{-2}(s)ds < \infty$.

The moment conditions in $B(i)$ ensure that components in $\eta_t(\theta)$ are bounded in L_{4p} for all θ. It also implies that fourth moments of structural shocks in ξ_t exist, which is commonly required for inference about covariance matrix Ω in VAR models.[6] The continuity and boundedness of volatility process in $B(ii)$ is essential for consistent estimation of σ_{it} and its inverse with kernel smoothing (see Hansen 1995). Define

$$\hat{\sigma}_{it}^2(\theta) := \frac{\sum_{k=1}^{T} K_\sigma((k-t)/h_\sigma)\xi_{ik}^2(\theta)}{\sum_{k=1}^{T} K_\sigma((k-t)/h_\sigma)},$$

where $K_\sigma : [-1, 1] \mapsto [0, 1]$ represents a standard kernel function satisfying $\int K_\sigma(x)dx > 0$ and h_σ denotes the bandwidth parameter controlling the smoothness of the estimated volatility processes, which increases as $T \to \infty$. Substituting η_{it} and σ_{it} in (5) by $\hat{\eta}_{it}$ and $\hat{\sigma}_{it}$, respectively, the objective function is given by

$$\hat{l}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\log \hat{f}_{i,\theta}(\hat{\eta}_{it}(\theta)) - \log \hat{\sigma}_{it}(\theta) \right), \tag{6}$$

where the density estimate is $\hat{f}_{i,\theta}(\hat{\eta}_{it}(\theta)) = \frac{1}{Th_f} \sum_{k=1}^{T} K_f \left(\frac{\hat{\eta}_{it}(\theta) - \hat{\eta}_{ik}(\theta)}{h_f} \right)$. The heteroskedasticity-consistent KML (HC-KML) estimator of the orthogonal mixing matrix is then obtained by maximizing the non-parametric pseudo log-likelihood function (6), i.e.,

$$\hat{Q} \equiv Q(\hat{\theta}), \text{ with } \hat{\theta} = \arg\max_{\theta \in \boldsymbol{\Theta}} \hat{l}_T(\theta) \tag{7}$$

In implementing the estimator given in equation (7), the analyst faces the task of selecting suitable kernel functions, namely K_f and K_σ, as well as bandwidth parameters, h_f and h_σ, for variance and density estimation. Notably, Hafner et al. (2024) show the robust performance of the KML estimator by utilizing a Gaussian kernel with $K_f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and the rule-of-thumb bandwidth estimator $h_f = 1.06T^{1/5}$ proposed by Silverman (1998). This choice minimizes the mean integrated squared error under a Gaussian density. As shown by Hafner et al. (2024), the kernel choice has minimal impact on the performance of the KML approach. For variance estimation, we employ the Epanechnikov kernel $K_\sigma(x) = 3/4(1 - x^2)\delta_{|x|\leq 1}$ due to its favorable simulation performance documented by Herwartz et al. (2023). The utilization of a symmetric kernel is also motivated by simulation-based evidence presented in Boswijk (2005). While the impact of the kernel choice is generally less significant in contexts with unbounded support, the compact support of variance smoothing means that fixed kernel functions can introduce boundary biases in small samples. Such biases can be mitigated by utilizing a more flexible kernel such as the beta kernel smoother (see, e.g., Bouezmarni and Scaillet 2005).

[6] If analyst directly observes the mixed independent components, the moment condition on σ_{it} can be dropped. We illustrate this in Section 3 with a volatility scenario (II).

Regarding the bandwidth parameter h_σ, it is essential for it to increase at a rate slower than T, though not too slowly. Detailed assumptions for h_σ will be presented later. A smaller value of h_σ results in reduced local smoothing and increased resolution in volatility estimation. However, it is crucial to acknowledge that employing a smaller bandwidth necessitates the presence of higher-order moments. Specifically, in Assumption B(i), p must be much larger than 1. As a benchmark bandwidth, we set $h_\sigma = \sqrt{T}$ and explore the impact of varying its value in both our simulation experiments and empirical applications, aiming to assess the sensitivity of our estimation outcomes to different choices of h_σ. While more advanced methods exist for bandwidth selection, an extensive discussion of such techniques is beyond the scope of this paper. Detailed discussions can be found in Foster and Nelson (1996) and Kristensen (2010). Nonetheless, it is important to bear in mind that the HC-KML estimator may benefit from further refinements through the adoption of more sophisticated bandwidth selection approaches.

2.5 Asymptotic properties

Denote $\sigma_{it}(\theta)$ the standard deviation of the (non-structural) shock $\xi_{it}(\theta)$ and $\eta_{it}(\theta) = \xi_{it}(\theta)/\sigma_{it}(\theta)$ the corresponding independent component. Let $f_{i,\theta}$ and f_θ be the true marginal and joint density functions of $\eta_{it}(\theta)$ and $\eta_t(\theta)$, respectively. For a given $\theta \in \boldsymbol{\Theta}$, the log-likelihood function based on the true variance and joint density is

$$\log L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(\log f_\theta(\eta_t(\theta)) - \sum_{i=1}^{N}\log \sigma_{it}(\theta)\right),$$

whereas the pseudo log-likelihood function based on the true marginal densities is

$$l_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\log f_{i,\theta}(\eta_{it}(\theta)) - \log \sigma_{it}(\theta)\right). \tag{8}$$

It is important to note that at the true parameter θ_0, components in $\eta_t(\theta_0)$ are independent and thus $\log L_T(\theta_0) = l_T(\theta_0)$. To demonstrate the consistency of the HC-KML estimator, we make the following assumptions:

Assumption C.

(i) The true parameter value $\theta_0 \in \boldsymbol{\Theta}$ lies in the parameter space $\boldsymbol{\Theta}$, a compact subspace of $\mathbb{R}^{N(N-1)/2}$, on which the local identification condition holds.
(ii) For all $i \in \{1, \dots, N\}$, f_i is uniformly continuous with bounded derivatives on its support, and $\inf_x f_i(x) \geq \epsilon_1 > 0$.
(iii) The kernel function in the density estimator satisfies: $K_f(x) \geq 0$, $\int K_f(x)dx = 1$, $\int |K_f(x)|dx < \infty$, $|x|K_f(x) \to 0$ as $|x| \to \infty$, $\sup_x K_f(x) < \infty$, $\sup_x |dK_f(x)/dx| < \infty$

and $K_f(0) \geq \epsilon_2 > 0$. The bandwidth parameter satisfies: $h_f \to 0$ and $Th_f/\log(T) \to \infty$ as $T \to \infty$.

(iv) The kernel function in the volatility smoothing satisfies $\int K_\sigma(x)dx > 0$. The bandwidth parameter satisfies: $h_\sigma = \alpha T^\beta$ for some $0 < \alpha < \infty$ and $(4p)^{-1} < \beta < 1$, where p is determined in Assumption B(i).

Let $l(\theta) := \sum_{i=1}^N \mathbb{E}[\log f_{i,\theta}(\eta_{it}(\theta)) - \log \sigma_{it}(\theta)]$ denote the population objective function, which has a unique maximum at the true parameter value θ_0:

Lemma 1. *For any $\epsilon > 0$, there exists an open ball $\mathcal{B}(\theta_0, \epsilon)$ with radius ϵ s.t.*

$$\sup_{\theta \in \Theta \cap \mathcal{B}(\theta_0,\epsilon)^c} l(\theta) < l(\theta_0).$$

Theorem 4. *Under Assumptions A, B and C,*

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

The estimator remains consistent, if some or all shocks are homoskedastic.

Remark 4. *If for some $i \in \{1, \dots, N\}$, $\sigma_{it}^2 = \mathbb{E}[\sigma_{it}^2] = 1$, a.s. $\forall t$, $\hat{\theta} \xrightarrow{p} \theta_0$.*

Studying the asymptotic distribution of the estimator $\hat{\theta}$ is known to be challenging due to the random denominator in the score estimation (see, e.g., Chen and Bickel 2006, Robinson and Taylor 2017). However, convergence of $\hat{\theta}$ to the normal distribution at the usual rate can be derived under additional assumptions involving sampling-splitting and trimming conditions. For notational convenience, we drop the index θ. Define the density score as $\psi_{it}(\eta_{it}) := \frac{f_i'(\eta_{it})}{f_i(\eta_{it})}$ and its estimator as $\tilde{\psi}_{it}(\hat{\eta}_{it}) := \frac{\hat{f}_i'(\hat{\eta}_{it})}{\hat{f}_i(\hat{\eta}_{it})}$, where $\hat{f}_i'(\hat{\eta}_{it}(\theta))$ is the kernel estimate of the first derivative of the density.[7] To ensure that the sequence of score estimates is bounded, we define the trimmed estimator

$$\hat{\psi}_{it}(\hat{\eta}_{it}) := \begin{cases} \tilde{\psi}_{it}(\hat{\eta}_{it}), & \text{if } \hat{\eta}_{it} \leq M_T^{\hat{\eta}},\ \hat{f}_i(\hat{\eta}_{it}) \geq d_T,\ \hat{f}_i'(\hat{\eta}_{it}) \leq M_T^{\psi} \hat{f}_i(\hat{\eta}_{it}) \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

where we assume for constant sequences $d_T \to 0$, $M_T^{\hat{\eta}} \to \infty$ and $M_T^{\psi} \to \infty$, as $T \to \infty$. This also ensures that the estimated variance sequence is bounded away from zero.

[7] For instance, if K_f is chosen to be a Gaussian kernel, the r-th derivative of the density can be consistently estimated by

$$\hat{f}_i^{(r)}(\hat{\eta}_{it}) = \frac{1}{Th_f^{r+1}} \sum_{k=1}^T (-1)^r \mathcal{H}_r \left(\frac{\hat{\eta}_{it} - \hat{\eta}_{ik}}{h_f} \right) K_f \left(\frac{\hat{\eta}_{it} - \hat{\eta}_{ik}}{h_f} \right),$$

where $\mathcal{H}_r(x)$ is the r-th Hermite polynomial (see, e.g., Bhattacharya 1967).

Alternative trimming conditions for $\hat{\sigma}_{it}^2$ have been discussed in Hansen (1995). For simplicity, we will write $\psi_{it} = \psi_{it}(\eta_{it}(\theta_0))$. To establish asymptotic normality, we make further assumptions.

Assumption D.
(i) For all $i \in \{1, \ldots, N\}$, f_i is symmetric about zero.
(ii) The bandwidth parameter h_f satisfies $h_f M_T^{\psi} \to 0$, and $T^{-1} M_T^{\hat{\eta}} / h_f^3 \to 0$.
(iii) The true parameter value $\theta_0 \in$ interior($\mathbf{\Theta}$).

Restricting the space of distributions, as in Assumptions D(i), has been widely adopted in previous studies on adaptive estimation (see, e.g., Bickel 1982). This assumption facilitates the proof of convergence of score estimates by enabling the establishment of mean-squared convergence. Moreover, it implies that $\psi'(X)X$ is a zero-mean random variable and makes $\hat{\psi}_{it}(\hat{\eta}_{it})$ asymptotically close to $\hat{\psi}_{it}(\eta_{it})$. A proof of asymptotic normality without the symmetry assumption may be achieved by relying on additional smoothness conditions and more stringent constraints on the bandwidth sequence (see Andrews 1994, Hafner et al. 2024). In Section 3, we demonstrate that the proposed estimator performs well under skewed source distributions. Additionally, we employ sample splitting, where ι_T is a sequence of natural numbers such that $\iota_T / T \to \iota \in (0, 1)$ as $T \to \infty$. The score estimator $\hat{\psi}_{it}$ is constructed based on the sub-sample $\mathrm{z}_1, \ldots, \mathrm{z}_{\iota_T}$ if $t \in \{\iota_{T+1}, \ldots, T\}$, and on the subsample $\mathrm{z}_{\iota_{T+1}}, \ldots, \mathrm{z}_T$ if $t \in \{1, \ldots, \iota_T\}$. This effectively makes $\hat{\psi}_{it}$ i.i.d. and independent of η_{it}, thereby simplifying the proof.

Theorem 5. *Under Assumptions A, B, C and D*

$$\sqrt{T}(\hat{\theta} - \theta_0) \Rightarrow \mathcal{J}(\theta_0)^{-1} \left(\sum_{i=1}^{N} \int_0^1 \sigma_i^{-1}(s) d\boldsymbol{W}_i(s) \right),$$

where $\boldsymbol{W}_i(s)$ is a vector Brownian motion on $\mathcal{C}^{N(N-1)/2}[0, 1]$ with covariance matrix $\mathcal{I}_i(\theta_0)$ for all $i = 1, \ldots, N$ and

$$\begin{aligned} \mathcal{I}_i(\theta_0) &= \mathbb{E}\left[\psi_{it}^2 \frac{\partial \xi_{it}(\theta_0)}{\partial \theta} \frac{\partial \xi_{it}(\theta_0)}{\partial \theta^\top} \right], \\ \mathcal{J}(\theta_0) &= -\sum_{i=1}^{N} \mathbb{E}\left[\psi_{it} \frac{\partial^2 \xi_{it}(\theta_0)}{\partial \theta \partial \theta^\top} + \psi_{it}' \frac{\partial \xi_{it}(\theta_0)}{\partial \theta} \frac{\partial \xi_{it}(\theta_0)}{\partial \theta^\top} \right] \left(\int_0^1 \sigma_i^{-2}(s) ds \right), \end{aligned}$$

with $\frac{\partial \xi_{it}(\theta_0)}{\partial \theta} := \frac{\partial \xi_{it}(\theta)}{\partial \theta}|_{\theta_0}$ and $\frac{\partial^2 \xi_{it}(\theta_0)}{\partial \theta \partial \theta^\top} := \frac{\partial^2 \xi_{it}(\theta)}{\partial \theta \partial \theta^\top}|_{\theta_0}$.

The asymptotic covariance matrix can be consistently estimated by replacing expectations with sample averages and the score with its nonparametric estimate. Detailed expressions and some examples are provided in the proof. Notably, $\hat{\theta}$ has the same asymptotic distribution as the KML estimator, if the latent shocks are homoskedastic.

Remark 5. *If* $\sigma_{it}^2 = \mathbb{E}[\sigma_{it}^2] = 1$, *a.s.* $\forall t, \forall i$, $\sqrt{T}(\hat{\theta} - \theta_0) \Rightarrow \mathcal{N}\left(0, \mathcal{J}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{J}(\theta_0)^{-1}\right)$, *where* $\mathcal{I}(\theta_0) = \sum_{i=1}^{N} \mathcal{I}_i(\theta_0)$.

It is important to note that, when applying to identification of macroeconomic models such as factor models, consideration must be given to the sampling uncertainty of z_t, which depends on the reduced-form covariance matrix Ω. If x_t is not observed but also estimated as in the case of SVARs, the estimated z_t will depend on VAR slope parameters. As long as there exists consistent estimators for these reduced-form parameters, the suggested estimator for the structural parameter remains consistent under our assumption due to a continuous mapping theorem. However, the asymptotic variance of $\hat{\theta}$ will have a different form than the one presented in Theorem 5. It is feasible to devise a one-step estimator that concurrently estimates both the reduced-form and structural parameters, and attains the semiparametric lower bound, as exemplified in the work of Fiorentini and Sentana (2023). Such an estimator could exhibit improved efficiency than our chosen approach. We focus on a two-step estimator as it offers the advantage of separating the estimation of structural parameters from that of the reduced-form parameters, facilitating comparisons between alternative identification schemes. To gauge the effect of estimating x_t on the finite-sample performance, we conduct simulation exercises in the next section. To jointly account for the estimation uncertainty of both the reduced-form and structural parameters, we adopt a residual-based moving block bootstrap approach in the empirical application (Brüggemann et al. 2016). The block bootstrap can correctly replicate the fourth-order moments of the structural shocks that are present in the asymptotic covariance of Ω. While the asymptotic properties of the bootstrap approach are intriguing topics in their own right, they leave the scope of our study.

3 Monte Carlo experiments

In this section, we conduct two Monte Carlo experiments to assess the finite-sample performance of the suggested identification approach under co-heteroskedasticity with observed mixed shocks and in an estimated VAR process.

3.1 Data generation and volatility scenarios

Our baseline analysis focuses on a bivariate model with observable variables z_t given by $\mathrm{z}_t = Q\xi_t$. Here, the elements in the vector ξ_t share a common volatility process σ_t, where $\xi_{it} = \sigma_t \eta_{it}$ for $i = 1, 2$, and σ_t is independent of η_t. To investigate estimator performance under various types of non-Gaussianity, we use six alternative distributions for the vector of independent components η_t, encompassing leptokurtic (Student's

t and exponential mixture), platykurtic (uniform and symmetric Gaussian mixture), and skewed (asymmetric Gaussian mixture and exponential) distributions. These distributions are visualized in the left-hand-side panel of Figure 3. To ensure comparability, we adjust the independent components η_t simulated from all data-generating distributions to have zero mean and unit variance across populations. Additionally, we consider two volatility-generating processes, which allow us to generate both stochastic autoregressive volatility and level shifts, i.e.,

$$\mathcal{V}_t^{(a)}(c_1, c_2) = \exp\left(c_1 T^{-1/2} h_t\right) \text{ and } \mathcal{V}_t^{(b)}(\kappa, \tau) = (1 - \mathcal{G}_t(\kappa, \tau))\varsigma_2^2 + \mathcal{G}_t(\kappa, \tau)\varsigma_3^2,$$

where $h_t = (1-c_2T^{-1})h_{t-1}+\zeta_t$, $\zeta_t \overset{iid}{\sim} \mathcal{N}(0, \varsigma_1^2)$, and $\mathcal{G}_t(\kappa, \tau) = (1 + \exp(-\kappa(t - \lfloor T\tau \rfloor)))^{-1}$, for finite constants $c_1, c_2, \varsigma_1, \varsigma_2, \varsigma_3, \kappa \in \mathbb{R}$ and $\tau \in (0, 1)$. The volatility process $\mathcal{V}_t^{(a)}$ has been considered in previous studies (see, e.g., Hansen 1995) as a discrete version of a continuous stochastic process. It is well-known that $T^{-1/2}h_{\lfloor sT \rfloor} \Rightarrow W_{c_2}(s)$ with $W_{c_2}(s)$ being a diffusion process, such that $dW_{c_2}(s) = -c_2 W_{c_2}(s) + dW(s)$, where $W(s)$ is a standard Brownian motion. The local-to-unity specification $(1-c_2T^{-1})$ in the representation of h_t is motivated by empirical estimates of stochastic autoregressive volatility processes (see, e.g., Harvey et al. 1994) and allows for unit roots ($c_2 = 0$) as well as roots close to unity. The volatility process $\mathcal{V}_t^{(b)}$ has been considered in Lütkepohl and Netšunajev (2017) as a smooth transition in volatility modeled by a logistic transition function $\mathcal{G}_t(\kappa, \tau)$. Coefficients κ and τ capture the smoothness and location of the transition, respectively.

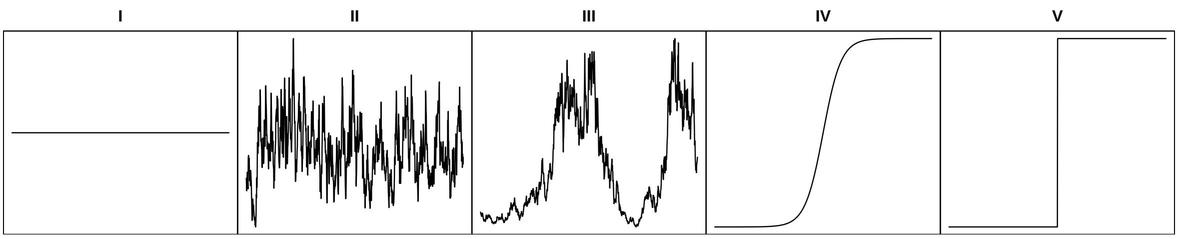

Figure 2: A random realization of σ_t^2 under the volatility scenarios (I) – (V) ($T = 1{,}000$).

For our analysis, we set $\varsigma_2 = 1$, $\varsigma_1 = \varsigma_3 = 3$ and consider five alternative volatility scenarios: (I) homoskedasticity, (II) – (III) autoregressive processes with distinct degrees of persistence, and (IV) – (V) volatility shifts occurring at the center of the sample

($\tau = 0.5$) with varying degrees of smoothness. Formally, we have

$$\begin{aligned}
\sigma_t^{2,(I)} &= \mathcal{V}_t^{(a)}(0, c_2) = 1 \\
\sigma_t^{2,(II)} &= \mathcal{V}_t^{(a)}(1, T/10) \\
\sigma_t^{2,(III)} &= \mathcal{V}_t^{(a)}(1, 0) \\
\sigma_t^{2,(IV)} &= \mathcal{V}_t^{(b)}((T/24)^{-1}, 0.5) \\
\sigma_t^{2,(V)} &= \lim_{\kappa \to \infty} \mathcal{V}_t^{(b)}(\kappa, 0.5) = 1 \cdot (1 - \delta_{t>T/2}) + 9\delta_{t>T/2}.
\end{aligned}$$

In scenario I, the shocks ξ_t are generated under homoskedasticity. For scenarios II and III, the stochastic component is given by $h_t = 0.9h_{t-1} + \zeta_t$ and $h_t = h_{t-1} + \zeta_t$, representing stationary autoregressive and unit root processes, respectively. Scenarios IV and V capture volatility shifts, with scenario IV exhibiting a smooth transition and scenario V involving an instant change at $t = T/2$. We consider four sample sizes: $T = 50, 100, 200,$ and $1,000$. Figure 2 displays exemplary realizations of each of the five volatility scenarios for a sample size of $T = 1,000$.

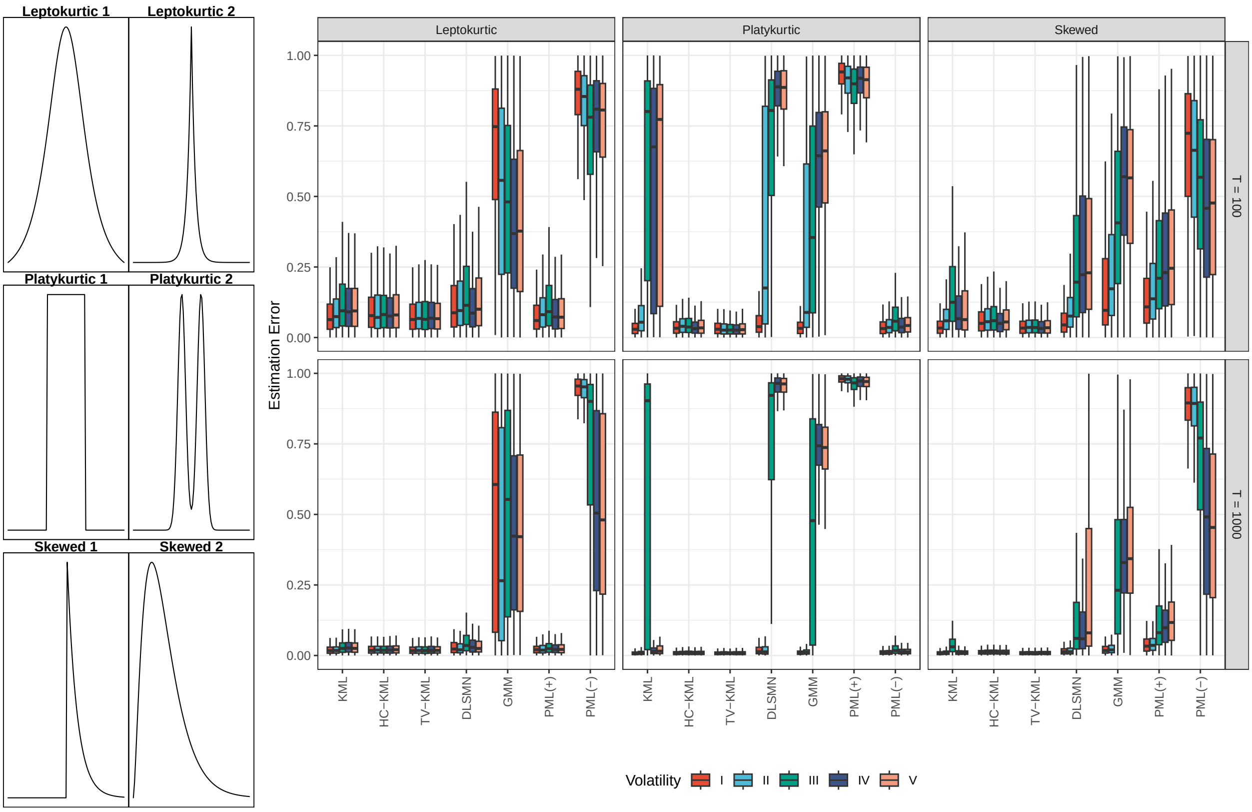

Figure 3: The left panel displays six data-generating-distributions, including leptokurtic (Student's t and exponential mixture), platykurtic (uniform and symmetric Gaussian mixture), and skewed (asymmetric Gaussian mixture and exponential) distributions. The right panel presents boxplots of the estimation error for alternative approaches with sample sizes $T = 100$ and $1,000$. Thick lines indicate the medians.

3.2 Implementation of alternative identification approaches

Following the generation of shocks under alternative volatility scenarios, the structural shocks in ξ_t are mixed by an orthogonal matrix $Q(\theta)$, where θ is uniformly sampled from $[-\pi;\pi]^{N(N-1)/2}$. The mixed shocks z_t are observed, and the structural parameters in Q are then identified using various approaches.

First, we employ the unadjusted KML approach that relies on assumption (i) in Theorem 1. For kernel density estimation, we follow Hafner et al. (2024) and adopt a general and agnostic setting with a Gaussian kernel and a rule-of-thumb bandwidth. Second, we utilize the novel HC-KML identification approach by adjusting the components in $Q(\theta)^\top \mathrm{z}_t$ with their estimated standard deviations. The Epanechnikov kernel is employed for variance smoothing with the bandwidth chosen as $T^{1/2}$. It is worth noting that alternative kernels and bandwidth choices, such as $T^{1/3}$, $T^{2/3}$, $T^{1/4}$, and $T^{3/4}$, yield similar performance. Third, we report estimates obtained by maximizing the function (5) assuming Σ_t is known. While this estimator, denoted as TV-KML, is infeasible in practice, it helps to gauge the efficiency loss caused by variance smoothing. Fourth, we consider ML estimation based on unrestricted two-component DLSMN densities, as also studied in the Monte Carlo analysis of Fiorentini and Sentana (2023) and Hafner et al. (2024). Furthermore, we employ a GMM estimator by imposing a set of asymmetric third- and fourth-order moment conditions, given by:

$$\mathbb{E}[\xi_{it}^2\xi_{jt}] = 0, \quad \text{and} \quad \mathbb{E}[\xi_{it}^3\xi_{jt}] = 0, \quad \text{for } i \neq j \text{ and } i,j \in \{1,\dots,N\}. \tag{10}$$

It is worth highlighting that these asymmetric moment conditions remain valid under (co-)heteroskedasticity.[8] Finally, we identify the matrix Q by means of PML based on a Student-t and a Gaussian mixture pseudo density, resulting in estimators PML(+) and PML(−), respectively. For a comprehensive comparison between the KML approach and a broader variety of ICA procedures (such as dCovICA, FastICA and JADE etc.) under a wider range of distribution scenarios, we refer the reader to the Section 3 of Hafner et al. (2024). To evaluate the estimates obtained from the alternative procedures, we use the Amari distance as a measure of estimation error. The Amari distance

[8]Let $\xi_{it} = \sigma_{it}\eta_{it}$, then $\mathbb{E}[\xi_{it}^2\xi_{jt}] = \mathbb{E}[\sigma_{it}^2\eta_{it}^2\sigma_{jt}\eta_{jt}] = \mathbb{E}[\sigma_{it}^2\sigma_{jt}]\mathbb{E}[\eta_{it}^2]\mathbb{E}[\eta_{jt}] = 0$ and $\mathbb{E}[\xi_{it}^3\xi_{jt}] = \mathbb{E}[\sigma_{it}^3\eta_{it}^3\sigma_{jt}\eta_{jt}] = \mathbb{E}[\sigma_{it}^3\sigma_{jt}]\mathbb{E}[\eta_{it}^3]\mathbb{E}[\eta_{jt}] = 0$, where the second and fourth equalities hold due to the independence between σ_{it} and η_{it} for all i,j and the independence between η_{it} and η_{jt} for all $i \neq j$. The argument for co-heteroskedasticity is similar. However, it is worth mentioning that under co-heteroskedasticity, these conditions alone (i.e., without mutual independence) may not be sufficient for identification in the lack of skewness (see Keweloh 2021).

(see, e.g., Bach and Jordan 2003) is given by:

$$d_{\text{Amari}}(Q,\hat{Q}) = \sum_{i=1}^{N}\left(\frac{\sum_{j=1}^{N}|r_{ij}|}{\max_j|r_{ij}|} - 1\right) + \sum_{j=1}^{N}\left(\frac{\sum_{i=1}^{N}|r_{ij}|}{\max_i|r_{ij}|} - 1\right), \text{ with } r_{ij} = (Q^{-1}\hat{Q})_{ij}. \tag{11}$$

The Amari distance is designed to be invariant to sign slips and column permutations, making it a suitable measure for comparing estimated matrices. Additionally, we compute the standard deviation of the Amari distance to provide a measure of uncertainty associated with the estimation error.

Vola. Scenario	I	II	III	IV	V	I	II	III	IV	V
			$T = 50$					$T = 100$		
KML	**.094 (.115)**	.228 (.242)	.304 (.277)	.252 (.248)	.281 (.268)	**.056 (.063)**	.107 (.143)	.248 (.262)	.185 (.229)	.190 (.230)
HC-KML	.141 (.167)	**.171 (.192)**	**.163 (.183)**	**.146 (.170)**	**.156 (.183)**	.077 (.102)	**.090 (.110)**	**.090 (.112)**	**.075 (.088)**	**.082 (.109)**
TV-KML	.091 (.106)	.103 (.132)	.100 (.130)	.096 (.119)	.098 (.119)	.056 (.062)	.054 (.068)	.052 (.058)	.059 (.072)	.057 (.066)
DLSMN	.139 (.164)	.280 (.264)	.361 (.294)	.396 (.303)	.417 (.306)	.107 (.148)	.162 (.203)	.346 (.299)	.403 (.322)	.406 (.304)
GMM	.338 (.300)	.412 (.288)	.445 (.292)	.442 (.279)	.455 (.284)	.277 (.296)	.337 (.281)	.429 (.287)	.436 (.269)	.432 (.270)
PML(+)	.360 (.308)	.404 (.299)	.416 (.306)	.426 (.307)	.434 (.306)	.345 (.313)	.364 (.306)	.398 (.312)	.408 (.322)	.422 (.320)
PML(−)	.491 (.322)	.458 (.310)	.461 (.316)	.443 (.314)	.432 (.314)	.489 (.340)	.488 (.339)	.454 (.316)	.440 (.329)	.447 (.329)
			$T = 200$					$T = 1,000$		
KML	**.034 (.040)**	.048 (.060)	.198 (.255)	.101 (.175)	.123 (.208)	**.012 (.011)**	**.014 (.013)**	.143 (.269)	.029 (.096)	.033 (.112)
HC-KML	.041 (.040)	**.046 (.046)**	**.048 (.049)**	**.044 (.043)**	**.044 (.042)**	.014 (.012)	.015 (.013)	**.015 (.012)**	**.014 (.012)**	**.015 (.012)**
TV-KML	.034 (.032)	.032 (.029)	.033 (.030)	.032 (.034)	.033 (.030)	.012 (.011)	.013 (.012)	.012 (.010)	.012 (.010)	.012 (.010)
DLSMN	.074 (.121)	.088 (.130)	.304 (.293)	.389 (.333)	.376 (.323)	.048 (.106)	.053 (.125)	.252 (.316)	.346 (.368)	.377 (.367)
GMM	.216 (.278)	.225 (.255)	.402 (.293)	.444 (.261)	.436 (.263)	.093 (.207)	.092 (.200)	.335 (.291)	.389 (.255)	.382 (.255)
PML(+)	.315 (.321)	.345 (.327)	.377 (.325)	.405 (.342)	.426 (.339)	.307 (.351)	.320 (.356)	.346 (.364)	.410 (.393)	.404 (.388)
PML(−)	.518 (.371)	.490 (.364)	.476 (.334)	.435 (.341)	.420 (.335)	.493 (.414)	.496 (.414)	.494 (.369)	.389 (.337)	.394 (.333)

Table 1: Mean and standard deviation (in parentheses) of the estimation error for alternative methods under different volatility scenarios labeled from I to V, and various sample sizes T, with independent components being randomly sampled from six data-generating distributions. Boldface numbers indicate the lowest mean and standard deviation of estimation error among the feasible methods (i.e., excluding TV-KML).

3.3 Baseline simulation results

Figure 3 displays the simulation performance of alternative estimators in terms of estimation errors, conditioned on specific shock distributions (leptokurtic, platykurtic, and skewed), for sample sizes $T = 100$ and $T = 1,000$. Beyond showcasing the consistent performance of the practically infeasible TV-KML estimator and its feasible HC-KML counterpart, both of which exhibit remarkable robustness across various source distributions and volatility scenarios, the presented simulation results offer three pivotal insights into the performance of KML, DLSMN, PML, and GMM estimation.

First, both the KML and DLSMN estimators exhibit vulnerabilities when encountering platykurtic distributions of independent components with non-stationary volatility (scenarios III, IV, and V). Specifically, in scenario III, characterized by platykurtic distributions alongside nonstationary volatility, KML estimation shows inconsistency, while the DLSMN approach falters in the presence of a smooth or instantaneous level shift in shock variance. Second and unsurprisingly, the performance of PML hinges on the choice of an appropriate pseudo density. Specifically, when underlying independent components are drawn from platykurtic distributions, adopting a Student-t pseudo density (PML(+)) leads to significant estimation biases that persist even in asymptotic settings. Similarly, the use of a Gaussian mixture pseudo density (PML(−)) lacks consistency in detecting leptokurtic independent components. Interestingly, the PML(+) estimator outperforms PML(−) when latent independent components follow a skewed distribution. Third, the GMM approach displays substantial distortions when dealing with heavy-tailed (i.e., leptokurtic) source distributions, even in situations of homoskedasticity or stationary volatility. In larger samples ($T = 1,000$), GMM estimation suffers from pronounced nonstationary heteroskedasticity (volatility scenarios III, IV, and V), which may lead to the emergence of markedly outlying observations. These outliers make it especially difficult to obtain accurate estimate of the asymptotically efficient GMM weighting matrix, which involves estimating up to eighth-order moments (such as $\mathbb{E}[\xi_{it}^3\xi_{jt}\xi_{kt}^3\xi_{lt}]$) (see Keweloh 2023 for discussion).

Table 1 documents performance statistics, including means and standard deviations, under alternative sample sizes, where shocks are randomly sampled with equal probability from six data-generating distributions. In the homoskedastic baseline scenario, KML, DLSMN and GMM approaches demonstrate consistent performance while HC-KML estimator presents a marginal inefficiency when contrasted with the benchmark KML estimator in small samples. When considering co-heteroskedasticity and among the KML variations, the infeasible TV-KML estimator, which employs true standard deviations for shock rescaling, (unsurprisingly) emerges as the most effective regardless of the specific volatility process. Intriguingly, the feasible HC-KML estimator yields parameter estimates that are only marginally less precise. Finally, the baseline KML estimator demonstrates accurate performance across all distributional scenarios and various heteroskedasticity specifications for underlying independent components. However, in cases of non-stationary volatility, the KML estimator's average large sample performance ($T = 1,000$) appears to slightly trail behind that of TV-KML and HC-KML. Further simulation results are documented in OA A.

3.4 Performance in estimated VAR processes

In a second simulation exercise, we consider a three-dimensional VAR(1) model, wherein the independent components η_t are mapped to observable variables in y_t through the

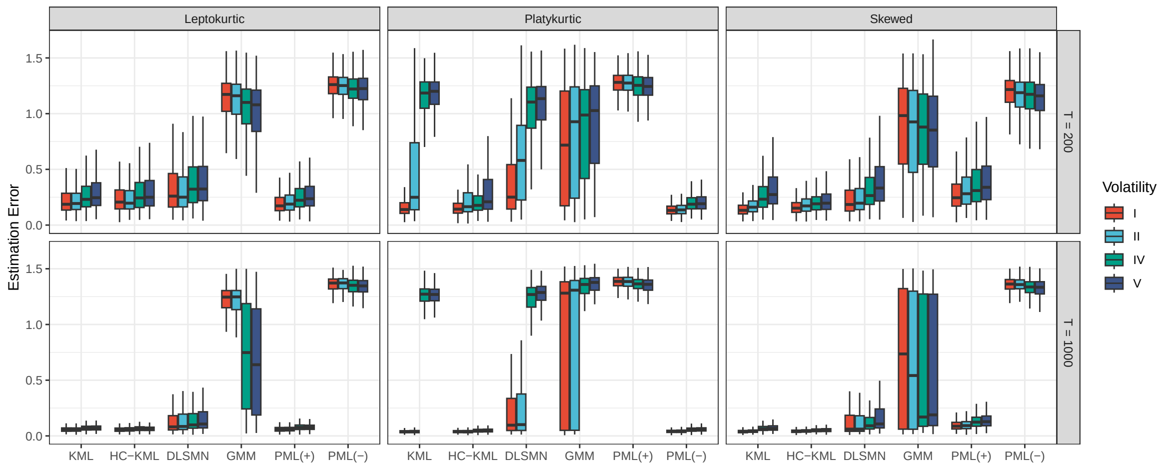

Figure 4: The estimation error for alternative approaches with sample sizes $T = 200$ and $1,000$ as boxplots, where x_t are estimated from a VAR process.

following DGP:

$$y_t - \begin{bmatrix} 0.79 & 0.00 & 0.25 \\ 0.19 & 0.95 & -0.46 \\ 0.12 & 0.00 & 0.62 \end{bmatrix} y_{t-1} = B\Sigma_t^{1/2}\eta_t,$$

where the autoregressive slope parameter is calibrated based on an economic model studied by An and Schorfheide (2007) and was previously employed in related simulation experiments (Lütkepohl and Schlaak 2022).[9] Similar to the baseline experiment, η_t are generated from six alternative distributions, and structural shocks in ξ_t share a common volatility process σ_t. We consider four volatility scenarios (I, II, IV and V) which are compatible with the stable VAR setting. To simulate the structural mixing matrix B, we first draw N^2 standard Gaussian variables and store them in a $N \times N$ matrix denoted by M in each experiment. Then, a covariance matrix is generated as $\widetilde{\Omega} = M^\top \mathrm{diag}(N, \dots, 1)M$ and standardized $\Omega = \widetilde{\Omega}\ \mathrm{diag}(1/\omega_1, \dots, 1/\omega_N)$, where ω_i^2 is the i-th main diagonal element of $\widetilde{\Omega}$. Finally, $B = \Omega^{1/2}Q(\theta)$ where rotation angles θ are uniformly drawn from $[-\pi; \pi]^{N(N-1)/2}$.

In each experiment, the analyst observes values of $\{y_t\}_{t=1}^T$, fitted a VAR(1) model, estimates both the slope parameter and covariance matrix Ω with least squares, from which the estimated residual vector $\hat{\mathrm{x}}_t$ and whitening matrix $\hat{\Omega}^{1/2}$ are extracted. Then, based on whitened residuals $\hat{\mathrm{z}}_t$, the analyst identifies the orthogonal mixing matrix $\hat{Q}$ using alternative ICA approaches and obtains estimates $\hat{B} = \hat{\Omega}^{1/2}\hat{Q}$. We document $d_{\mathrm{Amari}}(B, \hat{B})$ for six feasible ICA approaches, in four volatility scenarios under two al-

[9]The roots of the reverse characteristic polynomial implied by the process are 1.05, 1.11 and 1.95. Thus, y_t is stable with a certain degree of persistence given the two roots close to unity.

ternative sample sizes $T = 200$ and $T = 1,000$. The simulation results are displayed in Figure 4. Notably, the conclusions drawn for the baseline experiment remain unchanged. In particular, HC-KML still exhibits very favorable finite-sample properties when the analysis is conditional on an estimated VAR process.

4 A small scale US monetary policy model

In this section, we present the outcomes obtained by applying the proposed identification scheme and the newly developed HC-KML estimator to a small-scale monetary policy model for the US economy. First, we introduce the VAR model under scrutiny, outline the implementation details, and subsequently discuss the implications of the structural model, as demonstrated by a benchmark non-parametric variance estimator. Furthermore, we rationalize the model choice within the context of the implied variance estimators and their relationship with information-rich indicators of macroeconomic uncertainty (Jurado et al. 2015). As the yielded structural findings hinge on the implementation of variance smoothing, we conclude by scrutinizing the robustness of core structural model features.

4.1 Model representation and structural shocks

As an empirical illustration, we apply the novel identificaion approach to analyze the impact of US monetary policy shocks on real economic activity. For this purpose, we examine a three dimensional system of jointly endogenous variables conditional on changes in the global price for crude oil (i.e., refiner acquisition cost of imported crude oil) that we consider as an exogenous variable. In the vein of stylized trinity models, the set of endogenous variables comprises a measure of real economic activity, inflation and monetary policy indicator. More specifically, our VAR consists of the output gap as the deviation of the natural logarithm of real GDP from a measure of log potential GDP, annualized inflation as 400× the first difference of the natural logarithm of the quarterly GDP deflator, and a treasury bond rate with constant maturity of one year.[10] The deterministic components of the VAR include a constant term and a linear trend. A similar model with exogenous crude oil prices has been used in Gouriéroux et al. (2017) to illustrate the empirical performance of their pseudo ML estimator. Notably, our model choice distinguishes itself from Gouriéroux et al. (2017) by utilizing a one-

[10] All endogeneous variables are constructed utilizing data downloaded from the FRED database (FRED mnemonic in parentheses): real GDP (`GDPC1`), real potential GDP (`GDPPOT`), GDP deflator (`GDPDEF`), and yields of 1 year treasury bill (`GDPDEF`). The real oil price series is obtained on the basis of the refiner acquisition cost of imported crude oil provided by the US Department of Energy, and deflated by the US consumer price index (`CPIAUCSL`).

year treasury rate as an indicator of monetary policy conduct, instead of the federal funds rate. This choice takes into account the shift in monetary policy communication strategies over recent years, which places emphasis on steering expectations about future interest rate paths. This is particularly important given that in comparison with the policy analysis for the period 1959Qu4 until 2015Qu1 in Gouriéroux et al. (2017), the quarterly data used for our analysis spans a slightly shorter period from 1975Qu4 until 2019Qu4 but includes more recent pre-Covid observations. It has been shown that medium-maturity bond rates can serve as valuable indicators of the Federal Reserve's intentions, potentially revealing insights about surprises in forward guidance strategies (Gertler and Karadi 2015, Jarociński and Karadi 2020). This conditioning of the jointly endogenous variables on commodity market information offers an alternative to utilizing larger-dimensional SVAR models, as considered by Uhlig (2005) and Arias et al. (2019). Based on the model selection criteria AIC and BIC, we opt for a lag order of $p = 4$, which results in a dataset comprising $T = 173$ observations (see Tables B.1 to B.3 in OA B for diagnostic tests on remaining serial correlation, fundamentalness, and non-Gaussianity).

The small-scale model facilitates the recovering of three stylized macroeconomic shocks, each characterized by well-established effect patterns. In terms of the role played by 'positive' shocks, we observe that (i) a demand shock is anticipated to lead to an upswing across all macroeconomic aggregates (including output, prices, and interest rates), (ii) a supply shock brings about higher prices while tempering economic activity, and (iii) a contractionary monetary policy shock leads to elevated interest rates coupled with subdued inflation. It's noteworthy that while we achieve, akin to the approach of Gouriéroux et al. (2017), a focus on a lower-dimensional system by conditioning on crude oil prices, the distinct property of monetary policy shocks, causing interest rates and prices to move in opposing directions, remains a distinctive feature even in more complex monetary policy SVARs with higher dimensions (as discussed, for example, by Montiel Olea et al. 2022). Consequently, one could contend that any potential adverse impacts of excluding commodity prices from the set of endogenous variables are likely to be minimal, particularly concerning the identification of the monetary policy shock.[11]

[11] Montiel Olea et al. (2022) illustrate, through simulations using data from a medium-scale DSGE model of Smets and Wouters (2007) with seven shocks, that the monetary policy shock retains near-invertibility within a three-dimensional model. Moreover, Herbst and Schorfheide (2012) emphasize that the predictive performance of a larger-scale model, akin to Smets and Wouters (2007), doesn't necessarily surpass that of a related smaller-scale model.

4.2 Implementing the identification approach

To jointly identify all three shocks using the proposed estimator, a selection of kernel functions and bandwidth parameters is necessary, both for non-parametric density estimation and variance smoothing. As elaborated in Section 2 and supported by the evidence gathered from the Monte Carlo analysis in Section 3, we adopt the Gaussian kernel and a rule-of-thumb bandwidth for estimating the joint multiplicative density of standardized orthogonalized model residuals $\xi_{it}/\hat{\sigma}_{it}$. Regarding the non-parametric variance estimation, we opt for the Epanechnikov kernel. With regard to bandwidth selection for variance estimation, however, the common practice of choosing global smoothing parameters through leave-one-out cross-validation criteria is unfeasible here. This stems from the fact that the empirical counterparts of leave-one-out variance estimators, namely scaled squared shocks (ξ_{it}^2), remain latent, precluding the determination of cross-validation criteria. Recognizing the essential role of the bandwidth parameter in non-parametric variance estimation, we first evaluate the HC-KML estimator's performance concerning a benchmark bandwidth choice.

Given the quarterly nature of our sample observations, we consider a discrete range of alternative bandwidth choices: $h_\sigma = 4, 8, 12$, and 16, approximately corresponding to $0.3T^{1/2}, 0.6T^{1/2}, 0.9T^{1/2}$, and $1.2T^{1/2}$, and ultimately select $h_\sigma = 8$ for benchmark results. Subsequently, we provide explicit details on the robustness of structural estimates implied by these benchmark outcomes, supporting their plausibility. Supplementary IRFs derived from these alternative bandwidth parameters are available in OA C.

4.3 Structural impact multiplier and impulse responses

The estimated structural impact multiplier is given by

$$\widehat{B} = \begin{bmatrix} \underset{(7.182)}{0.571} & \underset{(-0.607)}{-0.107} & \underset{(0.261)}{0.037} \\ \underset{(0.879)}{0.206} & \underset{(6.549)}{0.765} & \underset{(-0.017)}{-0.003} \\ \underset{(1.073)}{0.16} & \underset{(0.175)}{0.019} & \underset{(4.517)}{0.592} \end{bmatrix},$$

where the values in parentheses represent the corresponding t-ratios obtained through a residual-based moving block bootstrap with 1,000 replications.[12] Importantly, the sign pattern of the estimated impact effects of the structural shocks in $\widehat{B}$ facilitates a distinctive and economically plausible labeling of ξ_{1t}, ξ_{2t}, and ξ_{3t} as the demand, supply, and monetary policy shocks, respectively. Particularly noteworthy is that the

[12]As evident from test results documented in Tables B.3 and B.4 in OA B, the model implied structural shocks align with the identifying Assumptions A.

monetary policy shock is the only one that induces inflation and interest rates to move in opposite directions upon impact. The structural IRFs are depicted in Figure 5. While demand shocks exhibit a positive impact on all variables (as seen in the first row of Figure 5), their effects are transitory. A supply shock, presented in the second row of Figure 5, elicits opposite movements in output and prices upon impact, and is the only shock in the system having a significant long-run effects on the economic activity. Turning to the effects of monetary policy shocks, the third row of Figure 5 indicates transient and statistically insignificant output and price puzzles. Notably, the U-shaped reaction of economic activity attains its peak magnitude after six quarters and eventually levels off. In quantifiable terms, an unanticipated monetary tightening leading to a median impact increase of 55 basis points in one-year treasury rates induces a 0.09% reduction in the output gap after six quarters. Regarding inflation control, a discernible downward pressure on inflation becomes evident around two quarters following a monetary contraction, with significance manifesting after five quarters. This impact remains noteworthy in the medium to long run, spanning up to four years. To further support the economic content of the monetary policy shocks identified by the HC-KML approach, a comparative assessment with alternative monetary policy measures, suggested by the literature based on information-rich models or plausible narrative data, is intriguing. In this context, we consider monetary policy shocks extracted by Smets and Wouters (2007) from a medium-scale DSGE model, the narrative shocks constructed by Romer and Romer (2004) based on Federal Open Market Committee minutes, and the SVAR shocks identified by Herwartz and Wang (2023) using a six-dimensional model using sign restrictions and independence criteria. Analyzing overlapping sample periods, we find substantial and statistically significant correlations between the HC-KML monetary policy shocks and their counterparts from the literature, with correlation coefficients of 0.48 (113 observations), 0.61 (125), and 0.67 (141), respectively.

In summary, our analysis underscores the effectiveness of US monetary policy in shaping medium to long-term inflation and reveals that its impact on economic activity is transitory within intermediate horizons. Next, we turn to the estimated variance patterns implied by the model. Considering the strides made by recent research in macroeconomic uncertainty, starting with Bloom (2009), the time-varying variance patterns we estimate might offer valuable information to corroborate the identified model. If these variance estimates align with the fundamental findings of the uncertainty literature (see, Bloom 2014 for a literature review), it would provide cross-confirmation. In case that the estimated variance patterns $\hat{\sigma}it^2$ significantly diverge from established results in this literature, it raises the possibility of criticism towards the adopted HC-KML model and prompts consideration for potential adjustments in bandwidth choice.

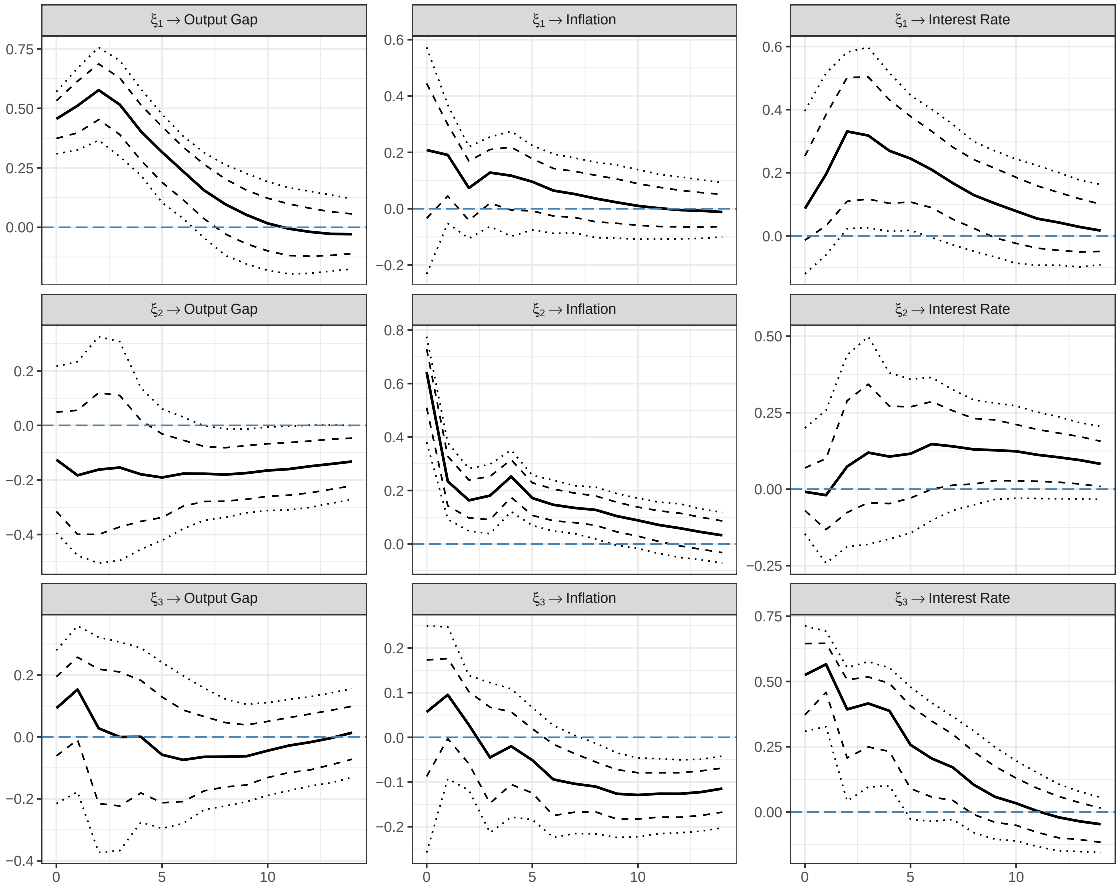

Figure 5: Structural impulse responses to demand shock (ξ_1), supply shock (ξ_2) and monetary policy shock (ξ_3) based on a $h_\sigma = 8$ bandwidth. The solid, dashed and dotted lines indicate the median, 68% and 90% confidence intervals from a moving block bootstrap (recursive design) with 1,000 replications, respectively. See Figures C.1 and C.2 in OA C for very similar IRFs estimates resulting from $h_\sigma = 4$ or $h_\sigma = 12$.

4.4 Implied variance estimates

The first three panels of Figure 6 illustrate the model-implied estimated variances of the structural shocks, demonstrating the diverse curves that result from varying bandwidth choices: $h_\sigma = 4, 8, 12$. To facilitate the detection of variance heterogeneity, horizontal lines indicate the unconditional variance of unity. Overall, these alternative smoothing parameters offer comparable insights into the secondary-order characteristics of the shocks. Several noteworthy observations pertain to the variance estimates, particularly those derived from the benchmark bandwidth $h_\sigma = 8$. First, aligning well with the narrative of the *Great Moderation* era, the shock variances tend to be moderate during the mid-1980s until the mid-2000s, while significant variances characterize

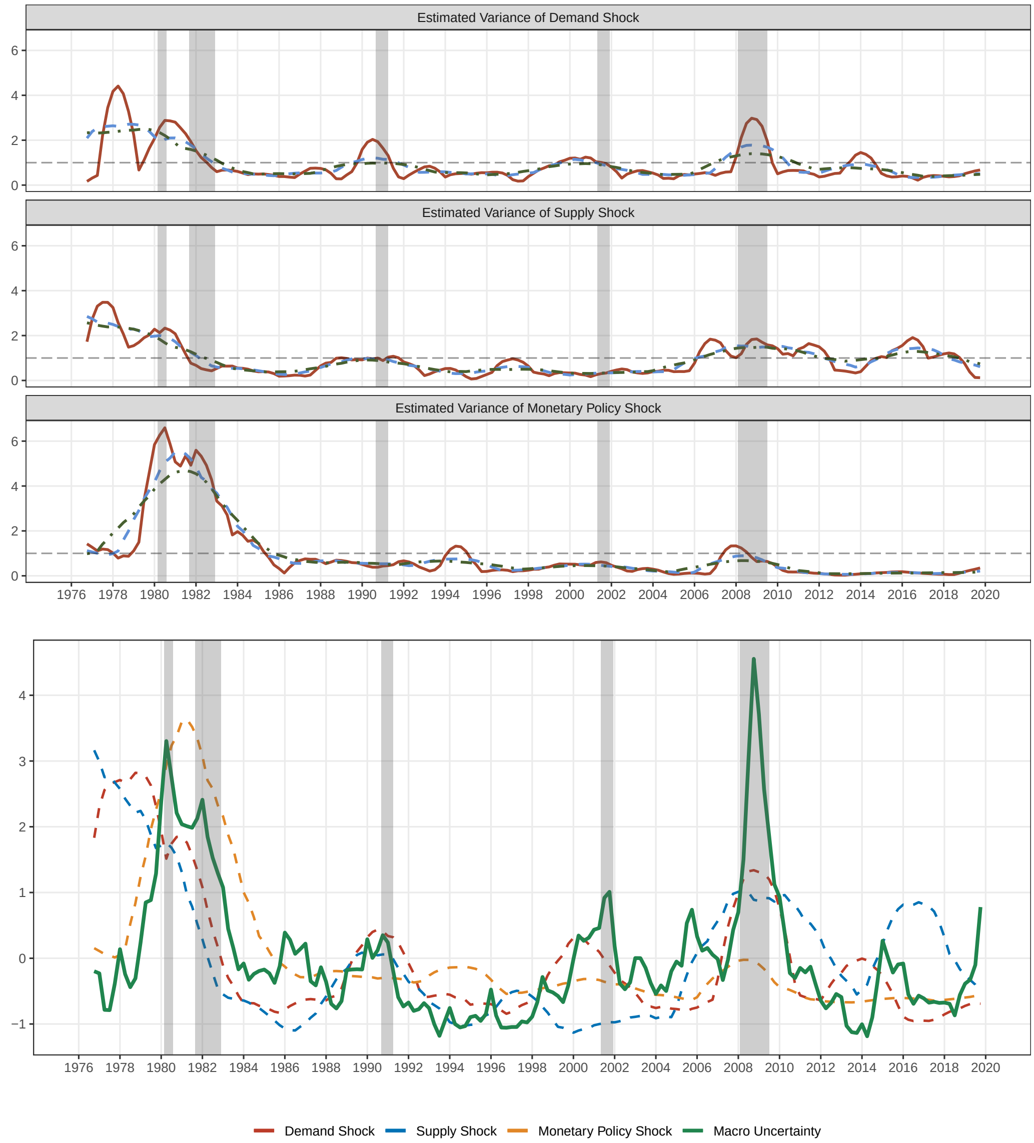

Figure 6: Upper panel: Variance estimates of demand, supply and monetary policy shocks using bandwidths $h_\sigma = 4$ (red solid), $h_\sigma = 8$ (blue dashed) and $h_\sigma = 12$ (green dot-dashed). Dashed horizontal lines indicate the unconditional unit variance. Lower panel: Variance estimates based on $h_\sigma = 8$ (dashed) compared to the macro uncertainty estimates of Jurado et al. (2015) (solid), expressed in standardized units.

the late 1970s and early 1980s. It is vital to note that our identification approach does not enforce any pre-existing knowledge or consensus on moment structure. Second, the estimated variances of the shocks labeled as demand and supply shocks exhibit a remarkable degree of comovement across significant portions of the sample, indicating a dependence in their second-order moment. When considering bandwidths of $h_\sigma = 4, 8,$ and $h_\sigma = 12$, the variance processes of both shocks show correlations of 0.51, 0.77, and 0.82, respectively. Such pronounced comovement profiles hold potential implications for ICA-based identification schemes that overlook such dependence patterns. Notably, parametric ML, pseudo ML, or kernel ML estimators may fail to capture this co-heteroskedasticity, treating the relevant components as jointly independent. Third, from a broader perspective, the data-driven non-parametric variance estimators do not seem to conform to stylized representations like structural breaks, smooth transitions, or conditional autoregressive heteroskedasticity. Consequently, identification schemes relying on explicit formalizations of informative (co)variance shifts might be challenging or even infeasible to apply in this context. Intuitively, the estimated variance of the latent structural shocks contribute to prediction uncertainty in the real and financial economies. Greater variance may complicate the formation of reliable predictions for macroeconomic aggregates, and contribute to the overall level of macroeconomic uncertainty. The lower panel of Figure 6 displays the variance estimators based on the benchmark bandwidth parameter (adjusted through centering and rescaling) in conjunction with the information-rich measures of macroeconomic uncertainty proposed by Jurado et al. (2015).[13] It is noteworthy that the time series of estimated variances of structural shocks largely align with the evolution of the macroeconomic uncertainty measure. Specifically, the variances of demand and monetary policy shocks exhibit significant comovement with macroeconomic uncertainty, while the variance of supply shocks in the late 1970s and 2007 appears to anticipate peaks in macroeconomic uncertainty. Interestingly, both periods experienced substantial increases in international oil prices (see Hamilton 2009 for a comparative analysis of the causes and consequences of the oil price surge in 2007-08). This implies that while the analysis conditions against the first-order-moment effects of commodity pricing by focusing on crude oil prices, oil markets could play a pivotal role in driving variations in latent supply shocks and subsequent macroeconomic uncertainty. In terms of explicit linear correlation measures, the variances of demand, supply, and monetary policy shocks (derived from the benchmark choice $h_\sigma = 8$) exhibit linear correlations with macroeconomic uncertainty of 0.51, 0.35, and 0.62, respectively. Consequently, it can be concluded that the non-parametric variance estimators find substantial validation from the intricate and information-rich uncertainty indicators proposed by Jurado et al. (2015).

4.5 Robustness

As noted earlier, we considered a discrete set of alternative bandwidth parameters that involve 'averaging' squared residuals within time windows spanning one, two (bench-

[13]Macroeconomic and financial uncertainty indices of Jurado et al. (2015) are constructed from an extensive collection of macroeconomic and financial time series and can be accessed from Sydney C. Ludvigson's website `https://www.sydneyludvigson.com.`.

mark), and three years of observations. Among these alternatives, selecting a larger bandwidth parameter will yield smoother patterns of implied variances in comparison to benchmark results, while opting for a smaller parameter could lead to more sensitive variance estimators. This distinction is evident from the top three panels of Figure 6. With the confirmation of variance estimators through macroeconomic uncertainty, as detailed above for the benchmark choice $h_\sigma = 8$, it becomes intriguing to explore the repercussions of varying bandwidth choices for variance estimation. Three key observations are worth highlighting in this context. First, adopting a finer bandwidth of $h_\sigma = 4$ preserves the described confirmation, as the implied variances of demand, supply, and monetary policy shocks still maintain linear correlations with the macroeconomic uncertainty measure of 0.55, 0.29, and 0.64, respectively, which are of comparable magnitude to the benchmark outcomes. Second, opting for larger bandwidths encompassing three or four years of observations ($h_\sigma = 12, 16$) generates distinctive outcomes. While the transition to $h_\sigma = 12$ upholds the integrity of the correlations of interest,[14] a selection of $h_\sigma = 16$ appears to excessively smooth the data, leading to a notably less accurate representation of macroeconomic uncertainties. Specifically, the implied variances of demand, supply, and monetary policy shocks correlate linearly with macroeconomic uncertainty at 0.46, 0.35, and 0.52, respectively. Third, moving towards even larger bandwidth choices, such as $h_\sigma = \infty$, may eventually compromise the ability to identify independent components, as certain shocks are likely influenced by co-heteroskedasticity for substantial portions of the analyzed sample period.

Beyond the realm of alternative variance estimators capturing the core patterns of macroeconomic uncertainties, it is worth examining the resilience of structural model characteristics under varying bandwidth selections. In this context, the patterns of structural IRFs remain remarkably consistent with benchmark results when opting for either a more sensitive ($h_\sigma = 4$) or less sensitive ($h_\sigma = 12$) bandwidth parameter. However, expanding the bandwidth further towards $h_\sigma = 16$ leads to a combination of an output and price puzzle. Comparable findings are obtained for the KML approach, which is implicit in the choice of $h_\sigma = \infty$.[15]

In conclusion, the outcomes attributed to bandwidth selections of $h_\sigma = 16$ and $h_\sigma = \infty$ prompt us to dismiss the implied structural models. The former is susceptible to

[14] In this case, the implied variances of demand, supply, and monetary policy shocks maintain linear correlations with macroeconomic uncertainty of 0.51, 0.35, and 0.58, respectively.

[15] Similarly, utilizing the PML estimator using a Student-t distribution with five degrees of freedom as the pseudo density (i.e., PML+ in the terminology of Section 3) also results in puzzling impact effects of monetary policy signals and an inability to identify any shock with the characteristic impact effects of demand shocks.

over-smoothing second-order features of structural shocks and results in an output puzzle, while the latter is prone to estimation bias due to the documented dependence, especially involving demand and supply shocks, in the form of co-heteroskedasticity.

5 Conclusion

Statistical identification approaches that utilize higher-order moments have become essential toolkits in structural analysis for econometricians. This paper reviews the identifying assumptions underlying prevalent approaches, such as those employing time-varying volatility or assuming independence under non-Gaussianity, and demonstrates that both approaches overlook the empirically relevant scenario where shocks share common volatility processes, known as co-heteroskedasticity. We introduce a novel statistical identification scheme that accommodates latent shocks with second-order moment features of unknown form and propose a new heteroskedasticity-consistent kernel-based maximum likelihood (HC-KML) estimator. We establish the estimator's consistency under standard regularity conditions and demonstrate its asymptotic normality. Notably, the suggested estimator recovers the unique linear combinations of heteroskedastically rescaled independent components regardless whether the shocks are (co-)heteroskedastic or homoskedastic. Through extensive Monte Carlo simulations, we demonstrate its favorable finite-sample properties. Specifically, the performance of the feasible HC-KML estimator is on par with the infeasible counterpart that employs the true variances for rescaling orthogonalized residuals.

In an empirical application, we showcase the effectiveness of the new identification approach within the context of a small-scale US monetary policy model incorporating exogenous crude oil prices, akin to the empirical analysis conducted by Gouriéroux et al. (2017). Our empirical findings shed light on the significance of co-heteroskedasticity issues surrounding model-implied demand and monetary policy shocks. The structural estimates derived from the novel HC-KML estimator reveal that an unexpected monetary tightening effectively manages inflation over a medium-term horizon of one year. As a valuable complement augmenting the structural analysis, the estimated variance profiles of latent shocks exhibit a strong alignment with a benchmark information-rich measure of macroeconomic uncertainty (Jurado et al. 2015).

References

An, S. and F. Schorfheide (2007): "Bayesian analysis of DSGE models," *Econometric Reviews*, 26, 113–172.

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

Arias, J. E., D. Caldara, and J. F. Rubio-Ramírez (2019): "The systematic component of monetary policy in SVARs: An agnostic identification procedure," *Journal of Monetary Economics*, 101, 1–13.

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

Bertsche, D. and R. Braun (2022): "Identification of structural vector autoregressions by stochastic volatility," *Journal of Business & Economic Statistics*, 40, 328–341.

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

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

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

Bloom, N. (2009): "The impact of uncertainty shocks," *Econometrica*, 77, 623–685.

——— (2014): "Fluctuations in uncertainty," *Journal of Economic Perspectives*, 28, 153–76.

Boswijk, H. P. (2005): "Adaptive testing for a unit root with nonstationary volatility," *UvA-Econometrics Discussion Paper*, 2005/07.

Bouezmarni, T. and O. Scaillet (2005): "Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data," *Econometric Theory*, 21, 390–412.

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

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

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

Cramér, H. (1936): "Über eine Eigenschaft der normalen Verteilungsfunktion," *Mathematische Zeitschrift*, 41, 405–414.

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

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

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

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

Foster, D. P. and D. B. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64, 139–174.

Gertler, M. and P. Karadi (2015): "Monetary policy surprises, credit costs, and economic activity," *American Economic Journal: Macroeconomics*, 7, 44–76.

Gouriéroux, C., A. Monfort, and J. Renne (2017): "Statistical inference for independent component analysis: Application to structural VAR models," *Journal of Econometrics*, 196, 111–126.

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

Hafner, C., H. Herwartz, and S. Wang (2024): "Statistical identification of independent shocks with kernel-based maximum likelihood estimation and an application to the global crude oil market," *Journal of Business & Economic Statistics*.

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

Hamilton, J. D. (2009): "Causes and consequences of the oil Shock of 2007-08," *Brookings Papers on Economic Activity*, 40, 215–283.

Hansen, B. E. (1995): "Regression with nonstationary volatility," *Econometrica*, 63, 1113–1132.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate stochastic variance models," *The Review of Economic Studies*, 61, 247–264.

Herbst, E. and F. Schorfheide (2012): "Evaluating DSGE model forecasts of comovements," *Journal of Econometrics*, 171, 152–166.

Herwartz, H. and H. Lütkepohl (2014): "Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks," *Journal of Econometrics*, 183, 104–116.

Herwartz, H., S. Maxand, and Y. M. Walle (2023): "Forward detrending for heteroskedasticity-robust panel unit root testing," *Econometric Reviews*, 42, 28–53.

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

——— (2024): "Statistical identification in panel structural vector autoregressive models based on independence criteria," *Journal of Applied Econometrics*.

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

Jarociński, M. (2024): "Estimating the Fed's unconventional policy shocks," *Journal of Monetary Economics*, 103548.

Jarociński, M. and P. Karadi (2020): "Deconstructing monetary policy surprises—The role of information shocks," *American Economic Journal: Macroeconomics*, 12, 1–43.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring uncertainty," *American Economic Review*, 105, 1177–1216.

Kagan, A. M., Y. V. Linnik, and C. R. Rao (1973): *Characterization problems in mathematical statistics*, John Wiley & Sons.

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

——— (2023): "Structural vector autoregressions and higher moments: Challenges and solutions in small samples," .

Kristensen, D. (2010): "Nonparametric filtering of the realized spot volatility: A kernel-based approach," *Econometric Theory*, 26, 60–93.

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

Lanne, M. and H. Lütkepohl (2008): "Identifying monetary policy shocks via changes in volatility," *Journal of Money, Credit and Banking*, 40, 1131–1149.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010): "Structural vector autoregressions with Markov switching," *Journal of Economic Dynamics and Control*, 34, 121–131.

Lanne, M., M. Meitz, and P. Saikkonen (2017): "Identification and estimation of non-Gaussian structural vector autoregressions," *Journal of Econometrics*, 196, 288–304.

Lewis, D. J. (2021): "Identifying shocks via time-varying volatility," *Review of Economic Studies*, 88, 3086–3124.

——— (2024): "Identification based on higher moments," cemmap working paper CWP03/24, London.

Linnik, Y. V. (1964): *Decomposition of probability laws*, Oliver and Boyd.

Lütkepohl, H. and A. Netšunajev (2017): "Structural vector autoregressions with smooth transition in variances: The interaction between U.S. monetary policy and the stock market," *Journal of Economic Dynamics & Control*, 84, 43–57.

Lütkepohl, H. and T. Schlaak (2022): "Heteroscedastic proxy vector autoregressions," *Journal of Business & Economic Statistics*, 40, 1268–1281.

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

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

Mesters, G. and P. Zwiernik (2022): "Non-independent components analysis," Working Papers 1358, Barcelona School of Economics.

Montiel Olea, J. L., M. Plagborg-Møller, and E. Qian (2022): "SVAR identification from higher moments: Has the simultaneous causality problem been solved?" *AEA Papers and Proceedings*, 112, 481–85.

Normadin, M. and L. Phaneuf (2004): "Monetary policy shocks: Testing identification conditions under time-varying conditional volatility," *Journal of Monetary Economics*, 1217–1243.

Rigobon, R. (2003): "Identification through heteroskedasticity," *The Review of Economics and Statistics*, 85, 777–792.

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

Romer, C. D. and D. H. Romer (2004): "A new measure of monetary shocks: Derivation and implications," *American Economic Review*, 94, 1055–1084.

Sentana, E. and G. Fiorentini (2001): "Identification, estimation and testing of conditionally heteroskedastic factor models," *Journal of Econometrics*, 102, 143–164.

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

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

Smets, F. and R. Wouters (2007): "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American Economic Review*, 97, 586–606.

Stock, J. H. and M. Watson (2012): "Disentangling the channels of the 2007-2009 recession," *Brookings Papers on Economic Activity*, 43, 81–156.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

A Proofs of theorems

Lemma A.1 (Darmois-Skitovich). *For two scalar random variables Y_1 and Y_2 of the linear form*

$$Y_1 = \sum_{j=1}^{N} \alpha_{1j} X_j \quad \text{and} \quad Y_2 = \sum_{j=1}^{N} \alpha_{2j} X_j, \tag{12}$$

where $\alpha_{1j}, \alpha_{2j} \in \mathbb{R}$ are constant coefficients and X_j are independent scalar random variables for $j = 1, \ldots, N$, which are not necessarily identically distributed. If Y_1 and Y_2 are independent, all variables X_j whose coefficients satisfy $\alpha_{1j}\alpha_{2j} \neq 0$ are Gaussian.

The proof of Lemma A.1 relies on the application of Cramér's Theorem (Cramér 1936) and a extended version of the Marcinkiewicz's Theorem. The latter states that if a characteristic function has the form $\phi(s) = \mathbb{E}[\exp P(t)]$, where $P(t)$ is a polynomial, then $P(t)$ is at most a quadratic polynomial, and $\phi(s)$ corresponds to the characteristic function of a normal law (see, e.g. Linnik 1964). A detailed proof of Lemma A.1 can be found in standard statistical textbooks that address characterization problems, such as Section 3 of Kagan et al. (1973).

Proof of Theorem 1: Local identifiability under condition (i) follows directly from Lemma A.1 and can be proved by contradiction. The proof is identical to the proof of Theorem 2, and thus omitted here. See proof of Theorem 2 for more details. We prove local identifiability under condition (ii). First, given the conditional covariance matrix of the observable variable $\mathbb{E}[\mathrm{x}_t\mathrm{x}_t^\top|\mathcal{F}_{t-1}] = \Omega_t = B\Sigma_t B^\top$, both matrices B and Σ_t are identified up to an orthogonal transformation. Let $\tilde{B} = BQ^\top$, $\tilde{\Sigma}_t = Q\Sigma_t Q^\top$ with $Q \in \mathcal{O}(N)$, then $(\tilde{B}, \tilde{\Sigma}_t)$ and (B, Σ_t) are observationally equivalent, i.e.,

$$\tilde{B}\tilde{\Sigma}_t\tilde{B}^\top = BQ^\top Q\Sigma_t Q^\top Q B^\top = B\Sigma_t B^\top.$$

The associated vector of non-structural orthogonalized shocks is given by $\tilde{\xi}_t = \tilde{B}^{-1}\mathrm{x}_t$ with conditional covariance matrix given by

$$\tilde{\Sigma}_t = Q\Sigma_t Q^\top = \begin{bmatrix} \sum_{k=1}^N \sigma_{kt}^2 q_{1k}q_{1k} & \sum_{k=1}^N \sigma_{kt}^2 q_{1k}q_{2k} & \cdots & \sum_{k=1}^N \sigma_{kt}^2 q_{1k}q_{Nk} \\ \sum_{k=1}^N \sigma_{kt}^2 q_{2k}q_{1k} & \sum_{k=1}^N \sigma_{kt}^2 q_{2k}q_{2k} & \cdots & \sum_{k=1}^N \sigma_{kt}^2 q_{2k}q_{Nk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^N \sigma_{kt}^2 q_{Nk}q_{1k} & \sum_{k=1}^N \sigma_{kt}^2 q_{Nk}q_{2k} & \cdots & \sum_{k=1}^N \sigma_{kt}^2 q_{Nk}q_{Nk} \end{bmatrix}.$$

Since components in $\tilde{\xi}_t$ are uncorrelated, $\tilde{\Sigma}_t$ is a diagonal matrix and thus $\sum_{k=1}^N \sigma_{kt}^2 q_{ik}q_{jk} = 0$ for all $i \neq j$ for all t. This condition can be expressed as

$$\underbrace{\begin{bmatrix} \sigma_{11}^2 & \sigma_{21}^2 & \cdots & \sigma_{N1}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \cdots & \sigma_{N2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1T}^2 & \sigma_{2T}^2 & \cdots & \sigma_{NT}^2 \end{bmatrix}}_{\boldsymbol{S}_T} \begin{bmatrix} q_{i1}q_{j1} \\ q_{i2}q_{j2} \\ \cdots \\ q_{iN}q_{jN} \end{bmatrix} = \mathbf{0}, \ \forall i, j \in \{1, \ldots, N\}, i \neq j.$$

Since rank $\boldsymbol{S}_T = N$ by (ii), the only solution to the system above is $q_{ik}q_{jk} = 0$ for all $i \neq j$ and all $k = 1, \ldots, N$. However, since $Q \in \mathcal{O}(N)$, $\sum_{j=1}^N q_{jk}^2 = 1$ for all k, each column in Q can not have two (or more) non-zero elements and the only non-zero element is either $+1$ or -1. This completes the proof. $\square$

Proof of Theorem 2: We prove the Theorem by proving that the orthogonal mixing matrix Q in (4) is identified up to a right-multiplication of $\mathcal{DP}$. Denote $\eta_t = \Sigma_t^{-1/2} Q^\top \mathrm{z}_t = (Q_t^*)^{-1}\mathrm{z}_t$ the vector of true independent components, where Q is the true orthogonal mixing matrix, $Q_t^* := Q\Sigma_t^{1/2}$. Under Assumption A, at most one component in η_t is Gaussian. Suppose that there exists another orthogonal mixing matrix $\tilde{Q}$ with $\tilde{Q}_t^* := \tilde{Q}\Sigma_t^{1/2}$ and $\tilde{\eta}_t$ the associated vector of independent components, i.e., $\tilde{\eta}_t = \Sigma_t^{-1/2}\tilde{Q}^\top \mathrm{z}_t = (\tilde{Q}_t^*)^{-1}\mathrm{z}_t$. We claim that $\tilde{Q}_t^*$ can be expressed as $Q_t^*\mathcal{D}_1\mathcal{P}$ for some diagonal matrix $\mathcal{D}_1$ and permutation matrix $\mathcal{P}$. We prove the claim by a contradiction.

Denote $\mathcal{A}_t := (\tilde{Q}_t^*)^{-1}Q_t^*$ and suppose that matrix $\mathcal{A}_t$ can not be expressed as $\mathcal{A}_t = \mathcal{D}_1\mathcal{P}$. Note that

$$\tilde{\eta}_t = (\tilde{Q}_t^*)^{-1}\mathrm{z}_t = (\tilde{Q}_t^*)^{-1}Q_t^*\eta_t = \mathcal{A}_t\eta_t.$$

Since the matrix $\mathcal{A}_t$ does not exhibit decomposition $\mathcal{D}_1\mathcal{P}$, it must contain two non-zero elements in at least two distinct columns. Let us denote these columns as j_1 and j_2, and without loss of generality, suppose that the non-zero elements are $\alpha_{m,j_1,t}$ and $\alpha_{n,j_2,t}$ with $\alpha_{i,j,t}$ being the $[i,j]$-th element in matrix $\mathcal{A}_t$ ($j_1, j_2, m, n \in \{1,\dots,N\}$, $j_1 \neq j_2$, $m \neq n$). Note that

$$\tilde{\eta}_{mt} = \sum_{j=1}^{N} \alpha_{m,j,t}\eta_{jt} \quad \text{and} \quad \tilde{\eta}_{nt} = \sum_{j=1}^{N} \alpha_{n,j,t}\eta_{jt},$$

where $\tilde{\eta}_{mt}$ and $\tilde{\eta}_{nt}$ are independent. Since $\alpha_{m,j_1,t}\alpha_{n,j_1,t} \neq 0$ and $\alpha_{m,j_2,t}\alpha_{n,j_2,t} \neq 0$, according to Lemma A.1, the corresponding shocks $\eta_{j_1,t}$ and $\eta_{j_2,t}$ must be both Gaussian. This contradicts our identifying Assumption A, which states that at most one of η_t is Gaussian. Therefore, $\mathcal{A}_t$ must take the form $\mathcal{D}_1\mathcal{P}$, or in other words, $\tilde{Q}_t^* = Q_t^*\mathcal{D}_1\mathcal{P}$. This implies

$$\tilde{Q}\Sigma_t^{1/2} = Q\underbrace{\Sigma_t^{1/2}\mathcal{D}_1}_{\mathcal{D}_2}\mathcal{P}$$

$$\tilde{Q} = Q\mathcal{D}_2\underbrace{\mathcal{P}\Sigma_t^{-1/2}\mathcal{P}^\top}_{\mathcal{D}_3}\mathcal{P} = Q\underbrace{\mathcal{D}_2\mathcal{D}_3}_{\mathcal{D}}\mathcal{P} = Q\mathcal{D}\mathcal{P}$$

which proves Theorem 2. $\square$

Proof of Theorem 3: We provide a proof for the bivariate case. The extension towards arbitrary $N > 2$ is straightforward by showing pairwise independence for $N(N-1)/2$ pairs of components and then applying the equivalence theorem for mutual independence in (Comon 1994 Theorem 11). Suppose σ_{it} has a marginal pdf denoted by f_{σ_i} for all i, which has support on $\mathbb{R}^+ := (0, \infty)$. Let $\boldsymbol{\sigma}_t = (\sigma_{1t}, \ldots, \sigma_{Nt})^\top$ and denote the marginal pdf of η_{it} and ξ_{it} by f_i and f_{ξ_i}, and the joint pdf of $\boldsymbol{\sigma}_t$, ξ_t and η_t by $f_{\boldsymbol{\sigma}}$, f_ξ and f, respectively. Then, for all i and for any $\xi_i \in \mathbb{R}$,

$$f_{\xi_i}(\xi_i) = \int_{\mathbb{R}^+} s_i^{-1} f_{\sigma_i}(s_i) f_i(\xi_i/s_i) ds_i,$$

while for any $\xi \in \mathbb{R}^2$,

$$f_\xi(\xi) = \int_{\mathbb{R}^+ \times \mathbb{R}^+} (s_1 s_2)^{-1} f_{\boldsymbol{\sigma}}(\boldsymbol{s}) f(\text{diag}\,(\boldsymbol{s})^{-1}\,\xi) |d\boldsymbol{s}|,$$

where $|d\boldsymbol{s}| = ds_1 \ldots ds_N$. To show $(i*)$, suppose – without loss of generality – $\sigma_{1t} = 1$ a.s. $\forall t$, then $f_{\xi_1}(\xi_1) = \int_{\mathbb{R}^+} s_1^{-1} f_{\sigma_1}(s_1) f_1(\xi_1/s_1) ds_1 = f_1(\xi_1)$. It follwos from Assumption A that for any $\eta = (\eta_1, \eta_2)^\top \in \mathbb{R}^2$, $f(\eta) = f_1(\eta_1) f_2(\eta_2)$, hence for any $\xi = (\xi_1, \xi_2)^\top \in \mathbb{R}^2$,

$$\begin{aligned} f_\xi(\xi) &= \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} f_{\boldsymbol{\sigma}}((s_1, s_2)^\top) s_1^{-1} f_1(\xi_1/s_1) s_2^{-1} f_2(\xi_2/s_2) ds_2 ds_2 \\ &= f_1(\xi_1) \int_{\mathbb{R}^+} s_2^{-1} f_{\sigma_2}(s_2) f_2(\xi_2/s_2) ds_2 = f_{\xi_1}(\xi_1) f_{\xi_2}(\xi_2), \end{aligned}$$

which proves that condition (i) in Theorem 1 is satisfied. To show $(ii*)$, if processes in $\boldsymbol{\sigma}_t^2$ are independent, so do processes in $\boldsymbol{\sigma}_t$ and thus for any $\boldsymbol{s} = (s_1, s_2)^\top \in \mathbb{R}^+$, $f_{\boldsymbol{\sigma}}(\boldsymbol{s}) = f_{\sigma_1}(s_1) f_{\sigma_2}(s_2)$. Therefore, for any $\xi = (\xi_1, \xi_2)^\top \in \mathbb{R}^2$,

$$\begin{aligned} f_\xi(\xi) &= \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} f_{\sigma_1}(s_1) f_{\sigma_2}(s_2) s_1^{-1} f_1(\xi_1/s_1) s_2^{-1} f_2(\xi_2/s_2) ds_2 ds_2 \\ &= \int_{\mathbb{R}^+} s_1^{-1} f_{\sigma_1}(s_1) f_1(\xi_1/s_1) ds_1 \int_{\mathbb{R}^+} s_2^{-1} f_{\sigma_2}(s_2) f_2(\xi_2/s_2) ds_2 \\ &= f_{\xi_1}(\xi_1) f_{\xi_2}(\xi_2), \end{aligned}$$

which proves the Theorem. □

Lemma A.2. *(i) Assume B(i), then* $\forall \theta \in \boldsymbol{\Theta}$ *and* $\forall i$, $\|\eta_{it}(\theta)\|_{4p}^{4p} \leq M_\eta M_\sigma < \infty$.
(ii) Assume A and B(ii), let $\sigma_{it}(\theta)$ *denote the standard deviation of* $\xi_{it}(\theta)$*, then* $\forall \theta \in \boldsymbol{\Theta}$ *and* $\forall i \in \{1, \ldots, N\}$, $\sigma_{i,\lfloor sT \rfloor}(\theta) \Rightarrow \sigma_i(s, \theta)$ *with* $\sigma_i(s, \theta) \in \mathcal{C}[0, 1]$*. Moreover,* $\forall i$, $\sup_{\theta \in \boldsymbol{\Theta}} \sup_{s \in [0,1]} \sigma_i^2(s, \theta) \leq \sigma_+^2 < \infty$ *and* $\inf_{\theta \in \boldsymbol{\Theta}} \inf_{s \in [0,1]} \sigma_i^2(s, \theta) \geq \sigma_-^2 > 0$.

Proof. To show (i), note that $\xi_t(\theta) = Q(\theta)^\top \mathrm{z}_t = Q(\theta)^\top Q(\theta_0) \xi_t$, where $Q(\theta)^\top Q(\theta_0)$ is orthonormal and hence $\exists \tilde{\theta} \in \boldsymbol{\Theta}$, s.t. $\xi_t(\theta) = Q(\tilde{\theta})^\top \xi_t$ and $\eta_{it}(\theta) = e_i^\top \Sigma_t^{-1/2}(\theta) Q(\tilde{\theta})^\top \xi_t$. Denote the $[i, j]$ entry of $Q(\tilde{\theta})$ by $q_{ij}(\tilde{\theta})$, then

$$\sigma_{it}^2(\theta) = \mathbb{E}\left[\left(\sum_{j=1}^N q_{ij}(\tilde{\theta}) \xi_{jt}\right)^2 |\mathcal{F}_{t-1}\right] = \sum_{j=1}^N q_{ij}^2(\tilde{\theta}) \mathbb{E}[\xi_{jt}^2 | \mathcal{F}_{t-1}] = \sum_{j=1}^N q_{ij}^2(\tilde{\theta}) \sigma_{jt}^2.$$

Note that $\forall\theta \in \mathbf{\Theta}$ and $\forall i$, $\mathbb{E}[\sigma_{it}^2(\theta)] = \sum_{j=1}^N q_{ij}^2(\tilde{\theta})\mathbb{E}[\sigma_{jt}^2] = 1$, since $\sum_{j=1}^N q_{ij}^2(\tilde{\theta}) = 1$. Moreover, $\forall i \in \{1, \ldots, N\}$,

$$\mathbb{E}|\xi_{it}|^{4p} = \mathbb{E}|\sigma_{it}\eta_{it}|^{4p} = \mathbb{E}|\sigma_{it}|^{4p}\mathbb{E}|\eta_{it}|^{4p} \leq M_\eta M_\sigma < \infty$$

by independence between σ_{it} and η_{it}. Thus, for some $p > 1$,

$$\mathbb{E}|\eta_{it}(\theta)|^{4p} = \mathbb{E}\left[\frac{1}{\sigma_{it}^{4p}(\theta)}\left|\sum_{j=1}^N q_{ij}(\tilde{\theta})\xi_{jt}\right|^{4p}\right] \leq \mathbb{E}[\sigma_{it}^{-4p}(\theta)]\mathbb{E}\left[\sum_{j=1}^N |q_{ij}(\tilde{\theta})|^{4p}|\xi_{jt}|^{4p}\right]$$
$$\leq \mathbb{E}[\sigma_{it}^2(\theta)]^{-2p} M_\eta M_\sigma \sum_{j=1}^N |q_{ij}(\tilde{\theta})|^{4p} \leq M_\eta M_\sigma \left(\sum_{j=1}^N q_{ij}(\tilde{\theta})^2\right)^{2p} = M_\eta M_\sigma < \infty.$$

For (ii), since $\sigma_{it}^2(\theta) = \sum_{j=1}^N q_{ij}^2(\tilde{\theta})\sigma_{jt}^2$, $\sigma_{i,\lfloor sT \rfloor}(\theta) \Rightarrow \sigma_i(s,\theta)$, where $\sigma_i^2(s,\theta) = \sum_{j=1}^N q_{ij}(\tilde{\theta})^2\sigma_{jt}^2(s)$ is continuous a.s. on $[0,1]$. Moreover, $\sup_{\theta\in\mathbf{\Theta}}\sup_{s\in[0,1]}\sigma_i^2(s,\theta) \geq \sigma_+^2 \sum_{j=1}^N q_{ij}(\tilde{\theta})^2 = \sigma_+^2 < \infty$ and $\inf_{\theta\in\mathbf{\Theta}}\inf_{s\in[0,1]}\sigma_i^2(s,\theta) \geq \sigma_-^2 \sum_{j=1}^N q_{ij}(\tilde{\theta})^2 = \sigma_-^2 > 0$. □

Proof of Lemma 1: As shown in the proof of Lemma A.2, $\forall\theta$, $\exists\tilde{\theta} \in \mathbf{\Theta}$, s.t. $\eta_t(\theta) = \Sigma_t^{-1/2}(\theta)Q(\tilde{\theta})^\top\Sigma_t^{1/2}\eta_t = W(\theta)\eta_t$. By a change of variable theorem,

$$\log f_\theta(\eta_t(\theta)) = \log f_{\theta_0}(W(\theta)\eta_t(\theta)) - \log|\det W(\theta)|, \tag{13}$$

where $\log|\det W(\theta)| = \sum_{i=1}^N \log\frac{\sigma_{it}}{\sigma_{it}(\theta)}$, since $Q(\tilde{\theta})$ is also orthogonal s.t. $\det Q(\tilde{\theta}) = 1$. For illustration purpose, in the following we provide a proof for the bivariate case. The generalization of the proof for arbitrary $N > 2$ is straightforward by replacing the joint and marginal densities with conditional densities by conditioning on the remaining variables. First, note that

$$\begin{aligned}\sum_{i=1}^2 \mathbb{E}\left[\log f_{i,\theta}(\eta_{it}(\theta))\right] &= \int_{\mathbb{R}} f_{1,\theta}(\eta_1)\log f_{1,\theta}(\eta_1)d\eta_1 + \int_{\mathbb{R}} f_{2,\theta}(\eta_2)\log f_{2,\theta}(\eta_2)d\eta_2 \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f_\theta(\eta_1,\eta_2)d\eta_2 \log f_{1,\theta}(\eta_1)d\eta_1 + \int_{\mathbb{R}}\int_{\mathbb{R}} f_\theta(\eta_1,\eta_2)d\eta_1 \log f_{2,\theta}(\eta_2)d\eta_2 \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f_\theta(\eta_1,\eta_2)\log\left(f_{1,\theta}(\eta_1)f_{2,\theta}(\eta_2)\right)d\eta_1 d\eta_2 \\ &= \mathbb{E}\left[\log\left(f_{1,\theta}(\eta_1(\theta))f_{2,\theta}(\eta_2(\theta))\right)\right],\end{aligned}$$

and

$$\sum_{i=1}^2 \mathbb{E}\left[\log f_{i,\theta_0}(\eta_{it})\right] = \mathbb{E}\left[\log\left(f_{1,\theta_0}(\eta_1)f_{2,\theta_0}(\eta_2)\right)\right] = \mathbb{E}\left[\log f_{\theta_0}(\eta_t)\right],$$

where the last equality follows from the independence. Then, by equation (13),

$$\begin{aligned}
l(\theta) - l(\theta_0) &= \mathbb{E}\left[\log\left(f_{1,\theta}(\eta_1(\theta))f_{2,\theta}(\eta_2(\theta))\right)\right] - \mathbb{E}\left[\log f_{\theta_0}(\eta_t)\right] - \mathbb{E}[\log|\det W(\theta)|] \\
&= \mathbb{E}\left[\log\left(f_{1,\theta}(\eta_1(\theta))f_{2,\theta}(\eta_2(\theta))\right)\right] - \mathbb{E}\left[\log f_{\theta}(\eta_t(\theta))\right] \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \log\left(\frac{f_{1,\theta}(\eta_1)f_{2,\theta}(\eta_2)}{f_\theta(\eta_1,\eta_2)}\right) f_\theta(\eta_1,\eta_2)d\eta_1 d\eta_2 \\
&\leq \int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{f_{1,\theta}(\eta_1)f_{2,\theta}(\eta_2)}{f_\theta(\eta_1,\eta_2)} - 1\right) f_\theta(\eta_1,\eta_2)d\eta_1 d\eta_2 \\
&= \int_{\mathbb{R}} f_{2,\theta}(\eta_2)d\eta_2 \int_{\mathbb{R}} f_{1,\theta}(\eta_1)d\eta_1 - \int_{\mathbb{R}}\int_{\mathbb{R}} f_\theta(\eta_1,\eta_2)d\eta_1 d\eta_2 \\
&= 1 - 1 = 0.
\end{aligned}$$

The equality holds if and only if $f_{1,\theta}(\eta_1)f_{2,\theta}(\eta_2) = f_\theta(\eta_1,\eta_2)$, i.e., $\eta_1(\theta)$ and $\eta_2(\theta)$ are independent, which under the local identification condition implies that $\theta = \theta_0$. Thus, for any $\epsilon > 0$, $\theta \in \mathbf{\Theta} \cap \mathcal{B}(\theta_0, \epsilon)^c$, the inequality holds strictly. $\square$

Proof of Theorem 4: We start by showing that $\sup_\theta \max_{1\leq t\leq T}|\hat{\sigma}^2_{it}(\theta) - \sigma^2_{it}(\theta)| \xrightarrow{p} 0$ for any $i \in \{1,\dots,N\}$. Denote the weights in the volatility smoothing

$$\bar{K}_\sigma(k-t) \equiv \frac{K_\sigma((k-t)/h_\sigma)}{\sum_{k=1}^T K_\sigma((k-t)/h_\sigma)}.$$

Note that $\sum_{k=1}^T \bar{K}_\sigma(k-t) = 1$. For the case $h_\sigma < t < T - h_\sigma$, note that

$$\begin{aligned}
\hat{\sigma}^2_{it}(\theta) - \sigma^2_{it}(\theta) &= \sum_{k=1}^T \bar{K}_\sigma(k-t)[\xi^2_{ik}(\theta) - \sigma^2_{it}(\theta)] \\
&= \sum_{k=1}^T \bar{K}_\sigma(k-t)[(\xi^2_{ik}(\theta) - \sigma^2_{ik}(\theta)) + (\sigma^2_{ik}(\theta) - \sigma^2_{it}(\theta))] \\
&= \sum_{k=1}^T \bar{K}_\sigma(k-t)\sigma^2_{ik}(\theta)[\eta^2_{ik}(\theta) - 1] + \sum_{k=1}^T \bar{K}_\sigma(k-t)[\sigma^2_{ik}(\theta) - \sigma^2_{it}(\theta)] \\
&\equiv R^a_{t,T} + R^b_{t,T}.
\end{aligned}$$

By Lemma A.2,

$$R^a_{t,T} = \sum_{k=1}^T \bar{K}_\sigma(k-t)\sigma^2_{ik}(\theta)[\eta^2_{ik}(\theta) - 1] \leq \sigma^2_+ \sum_{k=1}^T \bar{K}_\sigma(k-t)[\eta^2_{ik}(\theta) - 1],$$

where $\bar{K}_\sigma(k-t)$ is summable and $\eta^2_{ik}(\theta) - 1$ is bounded in L^{4p}. Hence, it follows from Lemma 2 of Hansen (1991), $\left\|\sum_{k=1}^T \bar{K}_\sigma(k-t)[\eta^2_{ik}(\theta) - 1]\right\|_{4p} \to 0$, which implies

$$\max_{h_\sigma < t < T - h_\sigma} |R^a_{t,T}| \xrightarrow{p} 0,$$

by the generalized Chebychev's inequality. Furthermore,

$$\begin{aligned}
R^b_{t,T} &\equiv \sum_{k=1}^{T} \bar{K}_\sigma(k-t)[\sigma^2_{ik}(\theta) - \sigma^2_{it}(\theta)] \\
&\leq \max_{h_\sigma < t < T-h_\sigma} \max_{t-h_\sigma \leq k \leq t+h_\sigma} [\sigma^2_{ik}(\theta) - \sigma^2_{it}(\theta)] \sum_{k=1}^{T} \bar{K}_\sigma(k-t) \\
&= \max_{h_\sigma < t < T-h_\sigma} \max_{t-h_\sigma \leq k \leq t+h_\sigma} [\sigma^2_{ik}(\theta) - \sigma^2_{it}(\theta)].
\end{aligned}$$

Under the continuity of the limit variance process in Lemma A.2 and our Assumption C(iv), there exists a finite real number $M^{h_\sigma} < \infty$ such that $h_\sigma \leq M^{h_\sigma} T^{\alpha'}$ for some $\alpha' \in (0,1)$ and

$$\max_{h_\sigma < t < T-h_\sigma} \max_{t-h_\sigma \leq k \leq t+h_\sigma} [\sigma^2_{ik}(\theta) - \sigma^2_{it}(\theta)] \xrightarrow{p} 0,$$

due to Lemma A.1 of Hansen (1995). Therefore,

$$\max_{h_\sigma < t < T-h_\sigma} |\hat{\sigma}^2_{it}(\theta) - \sigma^2_{it}(\theta)| \leq \max_{h_\sigma < t < T-h_\sigma} |R^a_{t,T}| + \max_{h_\sigma < t < T-h_\sigma} |R^b_{t,T}| = o_p(1).$$

The extension to cases $t < h_\sigma$ and $T - h_\sigma < t < T$ follows from Lemma A.1 of (Hansen 1995) under the continuity of the limit variance process in Assumption B(ii). Thus,

$$\max_{1 \leq t \leq T} |\hat{\sigma}^2_{it}(\theta) - \sigma^2_{it}(\theta)| \xrightarrow{p} 0,$$

for all $\theta \in \mathbf{\Theta}$. Since $\xi^2_{it}(\theta) = (e_i^\top Q(\theta)^\top z_t)^2$ and $\sigma^2_{it}(\theta) = \sum_{j=1}^{N} q^2_{ij}(\tilde{\theta})\sigma^2_{jt}$, where $q_{ij}(\tilde{\theta})$ is the $[i,j]$ entry of $Q(\tilde{\theta}) = Q(\theta)^\top Q(\theta_0)$, both $\hat{\sigma}^2_{it}(\theta)$ and $\sigma^2_{it}(\theta)$ are continuous in θ. By Assumption C(i), $\sup_\theta \max_{1\leq t\leq T} |\hat{\sigma}^2_{it}(\theta) - \sigma^2_{it}(\theta)| \xrightarrow{p} 0$. Let $\eta_{it}(\theta) = \xi_{it}(\theta)/\sigma_{it}(\theta)$ denote the structural shock standardized by the true variance. By the results above and from Lemma A.2, $\hat{\sigma}_{it}(\theta)$ is bounded away from zero, we have $\hat{\eta}_{it}(\theta) = \xi_{it}(\theta)/\hat{\sigma}_{it}(\theta) \xrightarrow{p} \eta_{it}(\theta)$. Let $\hat{f}_{i,\theta}(\eta_{it}(\theta))$ be the kernel density estimates based on $\eta_{it}(\theta)$, i.e.,

$$\hat{f}_{i,\theta}(\eta_{it}(\theta)) := \frac{1}{Th_f} \sum_{k=1}^{T} K_f\left(\frac{\eta_{it}(\theta) - \eta_{ik}(\theta)}{h_f}\right).$$

Assumptions C(ii) and C (iii) imply that (see Masry 1996)

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

Since for all $\epsilon_1, \epsilon_2 > 0$, $\inf_x f_i(x) \geq \epsilon_1 > 0$ and there exists $0 < T_{\epsilon_2} < \infty$ such that for all $T > T_{\epsilon_2}$, $K_f(0) \geq \epsilon_2 > \epsilon_2/Th_f > 0$, thus, both $f_{i,\theta}(\eta_{it}(\theta))$ and $\hat{f}_{i,\theta}(\eta_{it}(\theta))$ are bounded away from zero. Thus, for each $\theta \in \mathbf{\Theta}$, define $\Delta(\theta) := \log \hat{f}_{i,\theta}(\eta_{it}(\theta)) - \log f_{i,\theta}(\eta_{it}(\theta))$, and $\Delta(\theta) \xrightarrow{p} 0$ by the continuous mapping theorem. Then, by the mean value theorem and the Cauchy-Schwarz inequality,

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

where $M = \sup_{\theta^* \in \mathbf{\Theta}^*} ||\partial \Delta(\theta^*)/\partial\theta||$, with $\mathbf{\Theta}^*$ being the line segment joining θ and θ'. Note that M is bounded in probability since $\hat{f}_{i,\theta}(\eta_{it}(\theta))$ and $f_{i,\theta}(\eta_{it}(\theta))$ have bounded derivatives on $\mathbf{\Theta}$, and both are bounded away from zero. Therefore, $\Delta(\theta)$ is stochastically equicontinuous (see Theorem 21.10 of Davidson 1994) and thus $\sup_\theta|\log \hat{f}_{i,\theta}(\eta_{it}(\theta)) - \log f_{i,\theta}(\eta_{it}(\theta))| \xrightarrow{p} 0$ by Theorem 21.9 of Davidson (1994). Since $\sup_\theta|\log \hat{f}_{i,\theta}(\hat{\eta}_{it}(\theta)) - \log \hat{f}_{i,\theta}(\eta_{it}(\theta))| \xrightarrow{p} 0$ by the continuous mapping theorem, we have

$$\begin{aligned}
\sup_\theta|\hat{l}_T(\theta) - l_T(\theta)| &= \sup_\theta \frac{1}{T}|\sum_{t=1}^T \sum_{i=1}^N \log \hat{f}_{i,\theta}(\hat{\eta}_{it}(\theta)) - \log f_{i,\theta}(\eta_{it}(\theta)) + \log \sigma_{it}(\theta) - \log \hat{\sigma}_{it}(\theta)| \\
&\leq \sup_\theta \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N (|\log \hat{f}_{i,\theta}(\hat{\eta}_{it}(\theta)) - \log \hat{f}_{i,\theta}(\eta_{it}(\theta))| \\
&\quad + |\log \hat{f}_{i,\theta}(\eta_{it}(\theta)) - \log f_{i,\theta}(\eta_{it}(\theta))| + |\log \sigma_{it}(\theta) - \log \hat{\sigma}_{it}(\theta)|) \\
&= o_p(1).
\end{aligned}$$

Moreover, by Lemma A.2,

$$\mathbb{E}\left[\sup_\theta \left|\sum_{i=1}^N \log \sigma_{it}(\theta)\right|\right] = \mathbb{E}\left[\sup_\theta \left|\sum_{i=1}^N \log \left(\sum_{j=1}^N q_{ij}^2(\tilde{\theta})\sigma_{jt}^2\right)^{1/2}\right|\right] \leq N \log \sigma_+ < \infty,$$

and thus $\mathbb{E}[\sup_\theta|\sum_{i=1}^N \log f_{i,\theta}(\eta_{it}(\theta)) - \log \sigma_{it}(\theta)|] < \infty$. By a uniform law of large numbers, $\sup_\theta|l_T(\theta) - l(\theta)| \xrightarrow{p} 0$, and $\sup_\theta|\hat{l}_T(\theta) - l(\theta)| \leq \sup_\theta|\hat{l}_T(\theta) - l_T(\theta)| + |l_T(\theta) - l(\theta)| = o_p(1)$. Therefore, by Assumption C(i) and (ii), $\arg\max_\theta \hat{l}_T(\theta) - \arg\max l(\theta) = o_p(1)$. Let $\log L(\theta)$ denote the probability limit of $\log L_T(\theta)$, i.e., $\log L(\theta) := \mathbb{E}[\log f_\theta(\eta_t(\theta)) - \sum_{i=1}^N \log \sigma_{it}(\theta)]$, note that $l(\theta) \leq \log L(\theta)$, where the equality holds if and only if $\theta = \theta_0$ as the result of independence. By Lemma 1, we have $\arg\max_\theta \hat{l}_T(\theta) - \arg\max_\theta \log L(\theta) \xrightarrow{p} 0$ and thus $\hat{\theta} \xrightarrow{p} \theta_0$. □

Proof of Theorem 5: Define

$$\begin{aligned}
\hat{\mathcal{I}}_i(\theta_0) &:= \frac{1}{T} \sum_{t=1}^T \hat{\psi}_{it}^2(\hat{\eta}_{it}) \frac{\partial \xi_{it}(\theta_0)}{\partial\theta} \frac{\partial \xi_{it}(\theta_0)}{\partial\theta^\top}, \\
\hat{\mathcal{J}}(\theta_0) &:= -\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \left(\hat{\psi}_{it}(\hat{\eta}_{it}) \frac{\partial^2 \xi_{it}(\theta_0)}{\partial\theta\partial\theta^\top} + \hat{\psi}'_{it}(\hat{\eta}_{it}) \frac{\partial \xi_{it}(\theta_0)}{\partial\theta} \frac{\partial \xi_{it}(\theta_0)}{\partial\theta^\top}\right) \frac{1}{\hat{\sigma}_{it}^2},
\end{aligned}$$

where the estimator for density score $\hat{\psi}_{it}(\hat{\eta}_{it})$ is defined in (9). By a first-order Taylor expansion, we obtain

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

Since the estimator satisfies $\hat{l}'_T(\hat{\theta}) = 0$, we have

$$
\begin{aligned}
\hat{\mathcal{J}}(\theta_0)\sqrt{T}(\hat{\theta} - \theta_0) &= \sqrt{T}\hat{l}'_T(\theta_0) + o_p(1) \\
&= \sqrt{T}l'_T(\theta_0) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\hat{\psi}_{it}(\eta_{it}) - \psi_{it}(\eta_{it})\right)\frac{\partial\xi_{it}(\theta_0)/\partial\theta}{\sigma_{it}} \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\frac{\hat{\psi}_{it}(\hat{\eta}_{it})}{\hat{\sigma}_{it}} - \frac{\hat{\psi}_{it}(\eta_{it})}{\sigma_{it}}\right)\frac{\partial\xi_{it}(\theta_0)}{\partial\theta} + o_p(1).
\end{aligned}
$$

Since $\hat{\psi}_{it}(\eta_{it}) - \psi_{it}(\eta_{it})$ is i.i.d. due to the sample splitting construction, it suffices to show that for all $i = 1, \ldots, N$,

$$
\mathbb{E}\left[\left(\hat{\psi}_{it}(\eta_{it}) - \psi_{it}(\eta_{it})\right)^2 |\mathrm{z}_1, \ldots, \mathrm{z}_T\right] \xrightarrow{p} 0,
$$

as $T \to \infty$, which follows by Lemma 4.1 of Bickel (1982) under our conditions. A first-order Taylor expansion of $\hat{\psi}_{it}(\hat{\eta}_{it})$ at $\eta_{it} \equiv \xi_{it}/\sigma_{it}$ is

$$
\hat{\psi}_{it}(\hat{\eta}_{it}) = \hat{\psi}_{it}(\eta_{it}) - \hat{\psi}'_{it}(\eta_{it})\frac{\xi_{it}}{\sigma_{it}}\frac{(\hat{\sigma}_{it} - \sigma_{it})}{\sigma_{it}} + o_p(1),
$$

where for all $i = 1, \ldots, N$,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\psi}'_{it}(\eta_{it})\eta_{it}\frac{(\hat{\sigma}_{it} - \sigma_{it})}{\sigma_{it}} \le \max_{1\le t\le T}|\hat{\sigma}_{it} - \sigma_{it}|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\psi}'_{it}(\eta_{it})\eta_{it} = o_p(1),
$$

since $\psi'_{it}(\eta_{it})\eta_{it}$ is a bounded zero-mean innovation due to symmetry under Assumption D(i). Thus, for all $i = 1, \ldots, N$,

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\hat{\psi}_{it}(\hat{\eta}_{it})}{\hat{\sigma}_{it}} - \frac{\hat{\psi}_{it}(\eta_{it})}{\sigma_{it}}\right)\frac{\partial\xi_{it}(\theta_0)}{\partial\theta} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\psi}_{it}(\eta_{it})\frac{\sigma_{it} - \hat{\sigma}_{it}}{\sigma_{it}\hat{\sigma}_{it}}\frac{\partial\xi_{it}(\theta_0)}{\partial\theta} + o_p(1) \\
&\le \max_{1\le t\le T}|\hat{\sigma}_{it} - \sigma_{it}|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{\psi}_{it}(\eta_{it})\partial\xi_{it}(\theta_0)/\partial\theta}{\sigma_{it}\hat{\sigma}_{it}} = o_p(1),
\end{aligned}
$$

since σ_{it} and its estimates are bounded away from zero. Therefore, $\sqrt{T}\hat{l}'_T(\theta_0)$ and $\sqrt{T}l'_T(\theta_0)$ have the same limiting distribution. By Theorem 2.1 in Hansen (1992),

$$
\sqrt{T}l'_T(\theta_0) = \sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_{it}\frac{\partial\xi_{it}(\theta_0)/\partial\theta}{\sigma_{it}}\right) \Rightarrow \sum_{i=1}^{N}\left(\int_0^1 \sigma_i^{-1}(s)d\boldsymbol{W}_i(s)\right),
$$

since $\{(\sigma_{i,t+1}, \eta_{it})\}_{t\ge1}$ is adapted to $\{\mathcal{F}_t\}_{t\ge1}$ and $\{\eta_{it}\}_{t\ge1}$ is a martingale difference sequence relative $\{\mathcal{F}_t\}_{t\ge0}$ with $\sup_T T^{-1}\sum_{t=1}^{T}\mathbb{E}[\eta_{it}^2] = 1$ for all $i = 1, \ldots, N$. Since $T^{-1}\sum_{t=1}^{T}\sigma_{it}^{-2} \Rightarrow \int_0^1 \sigma^{-2}(s)ds < \infty$, by Assumption B($ii$), and by the weak law of large numbers and Assumption C(ii), we have $\hat{\mathcal{I}}_i(\theta_0) \xrightarrow{p} \mathcal{I}_i(\theta_0)$ and $\hat{\mathcal{J}}(\theta_0) \xrightarrow{p} \mathcal{J}(\theta_0)$. □