cemmap

centre for microdata methods and practice

Identification based on higher moments

Daniel Lewis

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP03/24

Economic
and Social
Research Council

Identification Based on Higher Moments

Daniel Lewis
University College London and CeMMAP*

February 15, 2024

Abstract

Identification based on higher moments has drawn increasing theoretical attention and been widely adopted in empirical practice in macroeconometrics in the last two decades. This article reviews two parallel strands of the literature: identification strategies based on heteroskedasticity and strategkes based on non-Gaussianity more generally. I outline the seminal identification results and discuss recent extensions, parametric and non-parametric implementations, and prominent empirical applications. I additionally describe key issues for the adoption of such strategies, including weak identification and interpretability of statistically identified structural shocks. I further outline key areas of ongoing research.

*daniel.j.lewis@ucl.ac.uk. I am grateful to Sascha Keweloh for helpful comments and to Brian Amorim Cabaço and Peter Jiang for research assistance.

1 Introduction

Following Sims (1980), structural vector-autoregressions (SVARs) have become the workhorse of causal inference in macroeconometrics. To recover dynamic causal effects from these models, the econometrician must identify a matrix of contemporaneous causal effects. Most traditional approaches to doing so are economic in nature. In particular, they either impose economically-motivated restrictions on these contemporaneous impacts (e.g., short-run restrictions, Sims (1980); long-run restrictions, Blanchard and Quah (1989); sign restrictions, Uhlig (2005)), or supply external economic information to help pin down a unique set of structural parameters (e.g., external instruments, Mertens and Ravn (2013)). A separate strand of the literature that has gained popularity in recent years employs statistical identification, in particular exploiting the information from higher moments. It is well-known that the second moments of VAR innovations – the reduced form covariance matrix – contains inadequate information to uniquely decompose the innovations into structural shocks and their contemporaneous impacts. However, moments beyond the covariance can contain sufficient information to achieve identification under certain conditions. This insight dates to as early as Wright (1928).

Two distinct, although closely connected, threads of this literature have developed. The first is based on heteroskedasticity, beginning with Sentana and Fiorentini (2001). If the variances of the structural shocks change through time, then there is not just a single reduced form covariance matrix to exploit. Thus, an SVAR model may no longer be underidentified using the expanded set of second moments. Over the past two decades, this idea has been applied, extended, and generalised in multiple directions, most popularly to a small number of discrete variance regimes (Rigobon (2003)), but also to smooth transition models (Lütkepohl and Netšunajev (2017a)) and parametric processes (Sentana and Fiorentini (2001); Lanne et al. (2010)). More recently, Lewis (2021) synthesised previous approaches in a non-parametric identification argument that accommodates all previously proposed processes and indeed essentially arbitrary persistent variance processes, including those with a state-space representation.

The second strand of the literature is based on the information offered by higher moments more generally, and is often referred to by the shorthand of identification via non-Gaussianity. Since Gaussian distributions are fully characterised by their first two moments, any identification scheme successfully exploiting higher moments must leverage some deviation of the structural shocks' distribution from Gaussianity. Due to the Darmois-Skitovich Theorem (e.g., Darmois (1953); Skitovich (1953)), if at least $n-1$ of the structural shocks exhibit non-Gaussianity and the structural shocks are mutually independent, then the decomposition of innovations into structural shocks is unique (e.g., Comon (1994)). This well-known result has been exploited in several different ways. These include non-parametric estimators based on independent component analysis (ICA) algorithms (Hyvärinen et al. (2010)), parametric maximum likelihood estimators (Lanne et al. (2017)), pseudo-maximum likelihood estimators (Gouriéroux et al. (2017)), and moment-based estimators (Guay (2021); Lanne and Luoto (2021); Keweloh (2021)). Of course, heteroskedasticity of the structural shocks in general implies non-Gaussianity; for example, shocks coming randomly from two different variance regimes are unconditionally non-Gaussian. Thus, the two strands of the literature are intrinsically connected.

Statistical identification approaches are attractive because they avoid making potentially controversial economic assumptions about the underlying structural parameters. Indeed, the structural parameters are typically the objects of interest to the econometrician, so imposing assumptions on them may be unappealing. Statistical identification, which can remain largely agnostic to these parameters is thus attractive when the econometrician is interested in testing hypotheses about structural parameters.

However, these identification approaches do make assumptions of their own – on the statistical properties of the data. While the presence of heteroskedasticity in macroeconomic or financial data may be uncontroversial, to avoid weak identification (and thus for standard inference techniques to be reliable) the departures from homoskedasticity may need to be substantial in typical applications; the mere existence of heteroskedasticity is not enough in finite samples. If adequate departures from homoskedasticity or Gaussianity are not present, such models will only be weakly identified, as recognized by Magnusson and Mavroeidis (2014) and Nakamura and Steinsson (2018) and recently studied by Lewis (2022), Montiel Olea et al. (2022), and Lee et al. (2022), for example. While it is in general hard to test identification conditions related to higher moments, there has been recent progress in this direction (e.g., Lewis (2022), Lütkepohl et al. (2020), Guay (2021)). Another key assumption, at least in the non-Gaussian family of approaches, is that of mutual independence of the structural shocks. Mutual independence is much stronger than the usual assumption that the structural shocks are mutually and serially uncorrelated, since it rules out that their variances follow a common or factor process, for instance – a condition that seems tenuous in light of episodes like the Great Moderation. Several recent papers have aimed to relax the independence assumption to varying degrees, (e.g., Guay (2021); Lanne and Luoto (2021); Mesters and Zwiernik (2022)). A final challenge of statistical identification is labeling. Based on the statistical information contained in higher moments alone, models are only ever identified up to sign/scale and column permutations, or "locally". There is no way to place an interpretable "label" on a structural shock, or a column of impact coefficients. Some papers have sought to provide a labeling approach based on a purely statistical preference over column order which guarantees global uniqueness, but no economic interpretation (e.g., Lanne et al. (2017)). More typically though, authors resort to economic information – meaning that these identification approaches are not in fact completely devoid of economic assumptions.

This review develops the identification strategies above in detail and outlines ongoing areas of research; the focus is macroeconometric applications, and SVARs in particular. Section 2 reviews the SVAR setting. Section 3 discusses identification approaches based on heteroskedasticity. Section 4 presents approaches based on non-Gaussianity. Section 5 describes the issues posed by weak identification as well as methods for robust inference and detection. Section 6 collects key open questions in the literature and active research areas. Section 7 concludes by identifying several appealing avenues for ongoing work.

2 Setting

Consider an $n \times 1$ vector of observed variables, Y_t, assumed mean zero for simplicity.[1] A standard SVAR has the form

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + u_t, u_t = B\epsilon_t, \tag{1}$$

or, more compactly,

$$A(L)Y_t = B\epsilon_t. \tag{2}$$

ϵ_t are mean-zero mutually orthogonal and serially uncorrelated structural shocks, or

$$E\left[\epsilon_t\right] = 0, E\left[\epsilon_{it}\epsilon_{js}\right] = 0, \forall i \neq j, s \neq t, E\left[\epsilon_t \epsilon_t'\right] = \Lambda, \tag{3}$$

where Λ is diagonal. Crucially, B is the $n \times n$ matrix of contemporaneous causal effects, assumed to be invertible.[2] The coefficients of the lag polynomial, $A(L)$, are consistently estimable via OLS. The SVAR identification problem pertains to B. In particular, the

[1] For a more general treatment of SVARs, see Kilian and Lütkepohl (2017).

[2] Note that the invertibility assumption is not innocuous, and is studied by a growing literature, see Fernández-Villaverde et al. (2007) for a discussion of such issues. Gouriéroux et al. (2019) consider the interaction of non-Gaussianity and non-invertibility.

covariance of the reduced form innovations, u_t,

$$\Sigma = E\left[u_t u_t'\right], \tag{4}$$

provides only $n\left(n+1\right)/2$ unique equations, but there are n^2 unknown parameters to be identified in the $n \times n$ matrix B and the n-entry diagonal covariance of ϵ_t, even after a suitable scale normalisation, giving n restrictions, is imposed (via either a unit diagonal for B or identity covariance for ϵ_t). Indeed, the impact coefficients of B are identified only up to orthogonal rotations. To see this, impose the unit variance normalisation, so $\Lambda = I_n$. Let $Q \in \mathcal{O}^n$, where $\mathcal{O}^n$ is the space of all $n \times n$ orthonormal matrices. Then

$$\Sigma = BI_nB' = BB' = \left(BQ\right)Q'I_nQ\left(Q'B'\right) = B^*B^{*\prime}, \tag{5}$$

where $B^* = BQ$. In other words, B and B^* are observationally equivalent up to second moments of u_t without further restrictions.

As discussed in the introduction, there are then two options to identify B uniquely. Restrictions must either be imposed on B, to reduce the number of free parameters from n^2 towards the $n\left(n+1\right)/2$ equations available in Σ (or in the case of sign restrictions, reduce the space of permissible rotations from $\mathcal{O}^n$ to some subset of $\mathcal{O}^n$), or additional information must be furnished – either through external variables or further moments of u_t. This review focuses on the final option, exploiting higher moments of the data. Two notions of identification are considered: local identification means that there is some neighbourhood around B in which no other parameters are observationally equivalent, while global identification means there exist no other observationally equivalent parameters. In what follows, higher moments alone will only ever identify B up to scale and column permutations – so are at most locally identifying without economic or other restrictions.

3 Heteroskedasticity

The intuition for how identification via heteroskedasticity can solve the SVAR identification problem is straightforward. Observational equivalence is based on the single set of equations – Σ – available under homoskedasticity. However, if there are multiple values for the variances of the structural shocks through time, the number of available equations scales up linearly. The key assumption permitting identification is that B remains constant, even as the variances change. Then, the number of new free parameters added for each variance regime is only n, fewer than than the $n(n+1)/2$ equations added.

Sentana and Fiorentini (2001) provide the first formal result exploiting heteroskedasticity directly for identification.[3] Let Λ_t denote the diagonal covariance matrix of the structural shocks at time t, which may follow an arbitrary stochastic process. Then, their Proposition 3 shows that if the paths of the n diagonal elements, λ_t, through time have full rank, then B is uniquely determined up to permutations (and scale/sign normalisation). The condition allows at most one shock to be homoskedastic. Indeed, the requirement of $n-1$ dimensions of linearly independent time-varying volatility will be shared by all schemes exploiting heteroskedasticity below, and mirrored by the requirement of at least $n-1$ non-Gaussian shocks in the next section. In general, this result provides substantial overidentifcation, as is discussed below. However, a major limitation of this result for practical use is that identification is based on the time path of Λ_t, and thus the time path of Σ_t, denoting the time-specific covariance of the reduced form innovations. These time-specific reduced form covariances are not, in general, consistently estimable without further parametric assumptions on the variance process for the structural shocks: as T increases, the sample size informative for time t without any further restrictions stays fixed. Sentana and Fiorentini's (2001) solution is to impose a GARCH functional form for the structural variance process, under which the variances are deterministic functions of past data and consistently estimable parameters. A relatively small empirical literature has employed the Sentana and Fiorentini (2001) approach directly, based on estimating a GARCH process for the structural shocks by (pseudo-)maximum likelihood, and thus estimating the structural parameters of interest, see for example King et al. (1994), Normandin and Phaneuf (2004), Bouakez and Normandin (2010), and Lütkepohl and Milunovich (2016).

3.1 Variance regimes

Rigobon (2003) makes perhaps the best-known, and most widely applied, contribution to this literature. In a special case of Sentana and Fiorentini (2001), he argues that if there are two discrete regimes for the structural variances, then B is uniquely determined provided that Λ_1 and Λ_2 are not scalar multiples of each other. Simple equation-counting shows why: $2 \times n((n-1)/2 = n^2 + n$, which is the number of structural parameters in B and two diagonal variance matrices, after n elements are normalised for scale. If there are more than two regimes, or values for the structural variances, then the model is overidentified. His Proposition 1 gives the now well-known

[3] However, the result appears earlier in an unpublished working paper, Sentana (1992).

identification condition in the two-regime case: identification holds as long as the variances do not change proportionally across regimes,

$$\Lambda_2 \neq \alpha\Lambda_1, \tag{6}$$

for any scalar α. There are extensions allowing for additional "common shocks" (requiring further regimes), see his Proposition 2. Crucially, it is not required that the econometrician knows the precise dates of the two variance regimes, or indeed that the variances follow a discrete process at all. Provided that the econometrician specifies the regimes such that the associated covariance estimators are consistent for distinct (non-proportional) variance regimes, the identification condition is met, see his Proposition 3. However, the better specified, and thus more distinct the regimes are, the stronger identification is likely to be. Additional regimes that provide overidentifying information allow the econometrician to test the modeling assumptions via the overidentifying restrictions, for example the assumption that B stays fixed over time.

In the special case of two regimes, where the model is just identified, a convenient closed form solution exists. As noted by Lanne et al. (2010), for example, B is identified as the left eigenvectors of the matrix $\Sigma_2\Sigma_1^{-1}$:

$$\Sigma_2\Sigma_1^{-1} = B\Lambda_2\Lambda_1^{-1}B^{-1}, \tag{7}$$

where $\Lambda_2\Lambda_1^{-1}$ is diagonal, containing the eigenvalues, and thus B contains the eigenvectors. These eigenvectors are unique (up to normalisation and order) provided that there are no repeated eigenvalues – no variances change proportionally across regimes. This result also clarifies how partial identification can occur: if there are repeated eigenvalues, the columns of B (eigenvectors) corresponding to the non-repeated entries are still identified, but only the column space associated with repeated values is identified. The case of partial identification has received little formal attention, although Bacchiocchi and Kitagawa (2023) derive identified sets resulting from partially identifying heteroskedasticity and sign or zero restrictions. In general, estimation of these regime-based models proceeds via GMM. Moments take the form

$$\phi\left(B, \Lambda_1, \ldots, \Lambda_N, \eta_t\right) = \begin{pmatrix} 1[t \in R_1]\left(E\left[\eta_t\eta_t'\right] - B\Lambda_1 B'\right) \\ \vdots \\ 1[t \in R_N]\left(E\left[\eta_t\eta_t'\right] - B\Lambda_N B'\right) \end{pmatrix}, \tag{8}$$

where $1[t \in R_j]$ is an indicator for whether observation t belongs to the j^{th} regime. However, maximum likelihood approaches are also available, and Brunnermeier et al.

(2021), Bacchiocchi and Kitagawa (2023), and Bacchiocchi et al. (2023) adopt fully Bayesian frameworks, for example.

Under additional assumptions, such that the variance of only one shock changes through time, the other(s) remaining homoskedastic, Rigobon and Sack (2004) show that an instrumental variables type estimator is available for a typical parameter of interest. In particular, let B_{ij} be the coefficient of interest, and assume that it measures the effect of the shock whose variance does change, say j, on some variable, say i. In this case, under the unit diagonal normalisation,

$$\frac{\Sigma_{2,ij} - \Sigma_{1,ij}}{\Sigma_{2,jj} - \Sigma_{1,jj}} = \frac{B_{ij}\lambda_{2j} + B_{ji}\lambda_i - B_{ij}\lambda_{1j} - B_{ji}\lambda_i}{\lambda_{2j} + B_{ji}^2\lambda_i - \lambda_{1j} - B_{ji}^2\lambda_i} = B_{ij}\frac{\lambda_{2j} - \lambda_{1j}}{\lambda_{2j} - \lambda_{1j}} = B_{ij}. \tag{9}$$

The left-hand side is equivalent to

$$\frac{E\left[u_{it}Z_t\right]}{E\left[u_{jt}Z_t\right]}, \tag{10}$$

where the instrument Z_t is given by

$$Z_t = \frac{T}{T_2} 1\left[t \in R_2\right] u_{jt} - \frac{T}{T_1} 1\left[t \in R_1\right] u_{jt} \tag{11}$$

and T_r denotes the number of observations in regime r.

For the estimators described above, provided that identification is strong – that is, (6) is satisfied, inference can proceed using the standard asymptotic results associated with each estimation strategy. However, the final formulation of the identification strategy (under stronger assumptions), (10), presents an analogy to IV estimation, which should make clear that weak identification is possible – in this case when the variance of shock j changes little between the two regimes; see Section 5 for further discussion.

One key distinction between the Sentana and Fiorentini (2001) and Rigobon (2003) results is the type of heteroskedasticity accommodated. The former in principle applies to both unconditional and conditional heteroskedasticity, although the paper focuses on conditional heteroskedasticity. The latter exploits unconditional heteroskedasticity, and it remains to determine the variance regimes. In some cases, external information proves helpful, with perhaps the most popular example being, in daily financial data, to use dates corresponding to announcements, especially monetary policy announcements, as a “high variance” regime, and dates far away from such announcements as a ”control” regime (see, e.g., Rigobon and Sack (2004), Nakamura and Steinsson

(2018)). Applying the same logic to continuous time periods like the Great Moderation, for instance, is more problematic, since doing so may mask considerable variation (periods of larger shocks within an otherwise low-variance interval). Rigobon (2003) uses narrative information on the dates of tranquil and crisis periods in Latin American debt markets to define regimes for identification. Alternatively, Rigobon and Sack (2003) propose to estimate the regime dates based on realised volatility, in particular comparing rolling averages of squared reduced form residuals to the average levels of the squared residuals to determine high- and low-variance periods. In population, the precise regime breaks do not matter so much, but in finite samples, estimation error can lead to more muted variance changes, and thus weak identification. One further drawback of estimating regimes, however, that has not been explored in the literature is that the realisations of squared residuals are driven by realised values of the structural shocks. Certain combinations of values of the shocks will be more conducive to "high" or "low" realisations of squared reduced form residuals, depending on B. Thus, for regimes determined based on finite-length windows, it need not be the case that the covariance matrix of the structural shocks is diagonal *conditional* on estimated regime membership, possibly introducing bias.

When regimes are unknown, a perhaps more natural option is to estimate them parametrically, using a Markov switching model. This is precisely the innovation of Lanne et al. (2010). In a straightforward extension, they propose a maximum likelihood estimator exploiting the Rigobon (2003) identification result; the identification conditions remain unchanged. Herwartz and Lütkepohl (2014) combine a Markov switching model with more conventional short- and long-run restrictions. Lütkepohl and Woźniak (2020) develop a Bayesian implementation of this approach and use it to test overidentifying restrictions on B.

Another important extension of the regime based model is the so-called "smooth transition" approach, in which the covariances are assumed to move as a convex combination between two matrices in a continuous manner. Lütkepohl and Netšunajev (2017b) propose such a model. Concretely, the reduced form covariance, Ω_t, follows the law of motion

$$\Omega_t = (1 - F(s_t))\,\Sigma_1 + F(s_t)\Sigma_2, \tag{12}$$

where $F(\cdot)$ is a parametric function. In particular, they propose

$$F(\gamma, b, s_t) = \left(1 + \exp\left[-\exp(\gamma)(s_t - b)\right)\right])^{-1}, \tag{13}$$

where γ is a slope parameter determining the speed of transition and b is a location parameter. s_t is the "transition variable", which governs the state at time t. It can either be a random variable (lagged inflation in Lütkepohl and Netšunajev (2017b)), or a deterministic variable, like t. If the former, it must be exogenous in order for standard inference results to hold. As $\gamma \to \infty$, the model approaches a threshold model, where $\Omega_t = \Sigma_1$ for $s_t < b$ and $\Omega_t = \Sigma_2$ for $s_t > b$. As $\gamma \to 0$, the model becomes unidentified (variances are constant), so weak identification is a concern for small values of γ. Conditional on a set of parameters in the transition equation, (γ, b), identification follows by the Rigobon (2003) argument, applied to Σ_1, Σ_2. In practice, the authors note that (γ, b) must be identified and that the choice of s_t may be important. For example, if s_t does not evolve with the underlying structural variances, then identification will likely be weak or non-existent. The authors propose to estimate the model via Gaussian maximum likelihood across a grid of parameters for (γ, b) using a two step procedure, alternating between estimating the reduced form and structural parameters. Then, the final log-likelihoods can be compared across the grid for (γ, b).

3.2 Unconditional moments

All of the preceding papers adopt different forms for the variance process, but rely on the same key insight – multiple values for the structural variances offer additional covariance matrices from which B may be identified. These approaches can all be thought of as relying on the *path* of the structural variances for identification. A much smaller literature argues for identification based on the parameters governing the evolution of the variances through time. Milunovich and Yang (2013) revisit identification based on ARCH-type functional forms. They formulate the mapping between parameters of the reduced form GARCH process for the residuals and those of the structural GARCH process for the shocks. They assume that structural ARCH coefficient matrix is diagonal. Based on the Jacobian of these equations, they show that B and the structural GARCH parameters are jointly locally identified from the reduced form GARCH parameters provided that there is at most one zero on the diagonal of the structural ARCH coefficient matrix. This condition amounts to a requirement that at most one shock is homoskedastic, mirroring the arguments above.

Lewis (2021) takes a different approach to argue that B is non-parametrically identified from the unconditional autocovariances of the squares of the reduced form residuals. Under the slightly stronger assumption that the shocks are a martingale difference sequence with respect to past shocks and current and past volatilities, and finite fourth moments, the following equations hold

$$\mathrm{cov}(\mathrm{vec}(u_t u_t'), \mathrm{vec}(u_s u_s')') = (B \otimes B) G \mathrm{cov}(\lambda_t, \mathrm{vec}(\epsilon_{t-s}\epsilon_{t-s}')')(B \otimes B)', s > 0 \tag{14}$$

$$E\left[u_t u_t'\right] = BE[\Lambda_t]B', \tag{15}$$

where G is a selection matrix of zeros and ones. The main result is that this system of equations has a unique solution for $B, \mathrm{cov}(\lambda_t, \mathrm{vec}(\epsilon_{t-s}\epsilon_{t-s}')'), E[\Lambda_t]$, up to normalisation and column order, provided that, for some lag s, $\left[\mathrm{cov}\left(\lambda_t, \mathrm{vec}(\epsilon_{t-s}\epsilon_{t-s}')'\right) \quad E[\lambda_t]\right]$ has rank of at least two and no proportional rows. If there are no ARCH effects, $\mathrm{cov}(\lambda_t, \mathrm{vec}(\epsilon_{t-s}\epsilon_{t-s}')') = \mathrm{cov}(\lambda_t, \lambda_{t-s}')G'$, the autocovariance matrix of λ_t at lag s. Applied to $\mathrm{cov}(\lambda_t, \mathrm{vec}(\epsilon_{t-s}\epsilon_{t-s}')')$, the rank condition essentially means that the autocovariance structures of the shock volatilities are not proportional. This will be satisfied if the volatility processes have persistence coming from a source other than a lower-dimensional factor structure, for example. Augmented with $E[\lambda_t]$, the condition states that even if the autocovariance structures are proportional, identification will still hold so long as the constants of proportionality are not equal to the ratios between the mean variances. Notably, this condition can be interpreted as allowing at most 1 homoskedastic shock, much like all of the previous schemes. If the identification conditions for any of the preceding schemes hold, then the rank condition above is satisfied. In this sense, the Lewis (2021) argument nests all previous identification schemes based on heteroskedasticity, without relying on parametric features for identification or consistent estimability of identifying moments.

The intuition behind the argument is that if the shocks are serially uncorrelated (and in fact are martingale difference sequences), then the only persistence in the squared residuals is that of the variance process (and/or ARCH effects). Then B, or rather $(B \otimes B)$ contains the coefficients relating the autocovariance of the squared reduced form residuals to the autocovariance of the squared structural shocks.[4]

There are several important features of the Lewis (2021) argument. First, it is nonparametric, so can be implemented without assuming any form for the identifying heteroskedasticity, for instance simply estimating the equations (14) by GMM. However, given standard macroeconomic sample sizes, it will often be challenging to estimate the required higher moments precisely, so in practice it may be more appealing to view it as a general-purpose argument that shows that a very wide range of parametric models will identify the structural parameters of interest. Provided that a parametric model

[4]The structure of the argument, particularly in a simplified motivating example presented in the paper, is not dissimilar to the seminal argument of Blundell et al. (2008) identifying coefficients of an income process in panel data.

satisfies the rank condition, then B will be identified from the associated likelihood. Finally, distinct from any previous arguments based on heteroskedasticity, it permits identification based on volatility models including state variables. For example, the autoregressive log stochastic volatility model is very popular in empirical work (e.g., Cogley and Sargent (2005); Primiceri (2005)), but previously was not compatible with any available identification arguments, since all require consistent estimation of the *path* of (reduced form) volatilities; see for example the implementation in Bertsche and Braun (2022), developed concurrently. Since the Lewis (2021) argument instead requires consistent estimation of unconditional *moments* of the volatilities, it is compatible with state-space volatility models.

Although this review focuses on identification as opposed to estimation and inference, one common estimation issue across these identification scheme warrants further discussion. Most commonly, SVARs are estimated in a two-step process: in the first step, the reduced form VAR is estimated via OLS, and the implied estimated residuals are then treated as data to estimate the structural parameters in the second step. For estimation purposes, these two steps are often treated as entirely separable (although inference may adjust for estimation error in $\hat{u}_t$). However, identification via heteroskedasticity motivates GLS-type estimators for the entire model, since they may offer efficiency gains if heteroskedasticity is present. In practice, maximum likelihood estimation alternates between estimating the reduced form parameters and structural parameters, computing time-varying variances and thus weights, then updating, until convergence, as described for Lütkepohl and Netšunajev (2017a) for example above. Moreover, without additional assumptions, inference on the structural parameters (and in particular IRFs) is complicated. Indeed, the estimation error in $\hat{u}_t$ cannot be ignored for the purpose of inference on $\hat{B}$ and IRFs in general. However, if the shocks, ϵ_t, are assumed to follow a symmetric distribution, it can be shown that the estimators of the reduced form covariances and those of the reduced form VAR parameters are asymptotically independent, which means that estimation error in $\hat{u}_t$ is asymptotically negligible in estimating $\hat{B}$. For further discussion, see Brüggemann et al. (2016), and in particular their Theorem 2.1, and the summary in Lewis (2021) (in particular footnotes 15 and 20). For a detailed discussion of estimation for many of the schemes described above, see Kilian and Lütkepohl (2017).

4 Non-Gaussianity

While the previous section considered exploiting a particular type of higher moment for identification, identification based on non-Gaussianity (and some form of Independence assumption) exploits moments beyond the second generically. Some approaches take a stand on what types of deviation from Gaussianity are likely to be informative in macroeconomic data, while others are more flexible. Note that in general, heteroskedasticity generates non-Gaussianity in the structural shocks, even if the underlying (standardised) disturbances are themselves Gaussian, so identification via heteroskedasticity can be thought of as a special case of identification based on non-Gaussianity.

This approach can be motivated from the idea that structural shocks should be independent (or at least more than uncorrelated), as discussed in Keweloh (2024). For example, the shocks $\epsilon_{1t} \sim \mathcal{N}(0,1)$ and $\epsilon_{2t} = \epsilon_{1t}^2 - 1$ are uncorrelated, but exhibit dependence that is inconsistent with how must macroeconomists think about structural shocks. Independence assumptions rule out these types of dependence. Given such an assumption, non-Gaussianity can identify the shocks. Consider the moment $E\left[\epsilon_{1t}^2 \epsilon_{2t}\right]$, a coskewness condition measuring the dependence of the two shocks. If $E\left[\epsilon_{1t}^2 \epsilon_{2t}\right] \neq 0$, then the first shock's size is informative for the sign of the second shock, and independence is violated; the key is to find an orthogonal rotation of the structural shocks such that independence holds. If the shocks are Gaussian, the moment is mechanically zero, but if they are non-Gaussian, moments like this are informative for B.

4.1 The main result

The idea that non-Gaussianity in general suffices to identify B follows from the Darmois (1953)-Skitovich (1953) Theorem. This result states that if the elements of $\zeta = (\zeta_1, \ldots, \zeta_n)'$ are independent random variables and $\mu'\zeta$ and $\beta'\zeta$ are independent (for non-zero μ, β), then all ζ_i are Gaussian. As a consequence, Comon (1994) (Theorem 11 and Corollary 13) shows that a decomposition of the form $u_t = B\epsilon_t$ is unique up to column order and scale provided at most 1 component of ϵ_t is Gaussian and the components of ϵ_t are independent. The result is based on a "contrast function", $\Psi(\cdot)$ that measures the (negative) deviation of a random vector from independence; intuitively, any candidate shocks cannot cannot be any "more independent" than the original shocks, ϵ_t. By Comon's Definition 5, for some random vector e_t, $\Psi(f_{e_t}) = \Psi(f_{\epsilon_t})$, where f_{e_t} denotes the density of e_t, if and only if e_t is a scaled permutation of ϵ_t. Further, if ϵ_t are independent, then $\Psi(f_{M\epsilon_t}) \leq \Psi(f_{\epsilon_t})$, with M invertible, so maximising the contrast function will return (a scaled permutation of) ϵ_t. It remains to propose candidate functions with these properties. Earlier results on identifiability under non-Gaussianity can be found in Geary (1941) and Reiersøl (1950). A notable feature of all of these results is the independence assumption on ϵ_t, which is notably stronger than the mutually orthogonal and serial uncorrelatedness assumption typically employed elsewhere in the SVAR literature. We will return to this below.

4.2 Independent Components Analysis

The Comon (1994) result is central to the "Independent Components Analysis" (ICA) literature, which originated the idea of identification via non-Gaussianity in modern macroeconometrics. ICA is focused on deconvolutions of the the form $u_t = B\epsilon_t$, where ϵ_t are independent, and is a key technique in the signal processing and neural networks literatures. While there are variants of ICA, the basic idea is, for an intial set of candidate standardised shocks, $\tilde{e}_t$, say, (for example, those arising from a Cholesky factorisation) to find the rotation Q for which the shocks $e_t = \tilde{e}_t Q$ are as independent as possible – for which some contrast function is maximised. For example, Comon (1994) proposes to use a feasible approximation to the negative of the Kullback-Leibler divergence of the joint density of e_t from the density under the independence of $e_{1t}, \ldots, e_{nt}$ ("mutual information"). This approximation can be expressed in terms of the cumulants of u_t and Q.

This subsection focuses on non-parametric ICA. This ICA literature is vast, and a comprehensive discussion is outside of the scope of this review. For a review of this literature, see Hyvärinen (2013), for example, or the recent literature review in Mesters and Zwiernik (2022). We focus on contributions that have directly impacted macroeconometrics. One such variant is the FastICA algorithm (Hyvarinen (1999)), which is in wide use, and available in many statistical software packages (many others exist, see for example the JADE algorithm of Cardoso and Souloumiac (1993)). FastICA exploits results showing that the negentropy (which measures the distance between some distribution and the normal distribution) can be better approximated by the maximum entropy principle than by cumulants (e.g., Hyvärinen (1997)) and further that minimising the mutual information is roughly equivalent to maximising the negentropy, or the degree of non-Gaussianity, of each shock. In practice, the algorithm uses an approximation to the negentropy of $e_{it} = c_i' u_t$, $J_H(\cdot)$

$$J_H(c_i) = \left(E\left[H(c_i' u_t)\right] - E\left[H(z)\right]\right)^2, \tag{16}$$

for some vector c_i and non-quadratic function H, where z is standard normal. For symmetric variables, this is a generalisation of the cumulant-based approximation in Comon (1994), where $H(e_{it}) = e_{it}^4$. Then the contrast function is $\sum_{i=1}^{n} J_H(c_i)$, which is maximized over $C = B^{-1}$. Hyvarinen (1999) proposes a number of different choices for H, implemented in the algorithm, including $1/k_1 \log\cosh(k_1 u)$, $-1/k_2 \exp\left(-k_2 u^2/2\right)$, and $1/4\ u^4$, motivated by efficiency and robustness considerations. The FastICA algorithm then proceeds to maximise the negentropy given $H(\cdot)$ using a computationally efficient fixed point algorithm. While most of the ICA literature assumes i.i.d. shocks, a concern that will be discussed in detail below, some recent algorithms have been extended to accommodate heteroskedasticity, for example, gJADE, Matilainen et al. (2015).

Results on the statistical properties of estimators are relatively rare in the ICA literature. In particular, consistency results for the estimators resulting from many algorithms are hard to find, and some are in fact inconsistent, see the discussion in Gouriéroux et al. (2017). Consistency and asymptotic normality of the FastICA algorithm have been established by Wei (2015), although expressions for the asymptotic variance (assuming consistency) date to Shimizu et al. (2006), for example. Bonhomme and Robin (2009) provide earlier consistency and asymptotic normality results for a modification of the JADE algorithm.

Hyvärinen et al. (2010) is likely the first paper to exploit modern non-Gaussianity results to identify an SVAR, using financial data. They consider a SVAR of four global stock indices and estimate the reduced form before applying the FastICA-based LiNGAM algorithm of Shimizu et al. (2006). They impose a sparsity penalty in the estimation, and find that more than half of the impact coefficients are zero, and that $\hat{B}$ can be permuted to a lower-triangular matrix, with strong spillovers from the Dow Jones to both Nikkei and Hang Seng indices. Moneta et al. (2013) introduce the same methodology, but in an economics publication, with applications to both firm growth and the effects of monetary policy.

4.3 Likelihood approaches

The next generation of results identifying SVARs using non-Gaussianity rely on likelihood approaches instead of non-parametric ICA. Lanne et al. (2017) assume that the shocks are sequences of independent and identically distributed and mutually independent disturbances with variances λ_i, with at most one shock Gaussian. Their identification result is based on those above. They propose a maximum likelihood estimator in terms of the density $f_{i,\lambda_i}(x, \theta_i) = \lambda_i^{-\frac{1}{2}} f_i(x/\lambda_i^{\frac{1}{2}}, \theta_i)$, but in practice assume that each shock follows a Student-t distribution. Standard asymptotic properties for maximum likelihood estimation hold, provided the choice of densities is correct, a rather heroic assumption given that there are infinitely many ways to model non-Gaussianity, and presenting a challenge to empirical users. They propose a computationally simpler three-step estimator that is efficient if the shocks follow symmetric distributions. In an extension, they are able to relax the assumption of temporal independence to no serial correlation, which admits time-conditional heteroskedasticity for the shocks, provided that the volatility processes remain independent. They also provide an approach for refining local identification to global identification, which is discussed in detail in Section 6.4. As an empirical application they study the relationship between the macroeconomy and financial conditions, and are able to marginally reject a conventional recursive structure in their SVAR.

Contemporaneously, Gouriéroux et al. (2017) instead consider pseudo-maximum likelihood (PML) estimation. They assume that the shocks ϵ_t are independent, with at most one Gaussian. Importantly, they establish consistency for the PML estimator even when the likelihood is misspecified, provided the misspecified model is identified. Additional results offer testable implications of Gaussianity of 2 or more shocks for the observed data, Y_t. They further provide expressions for the asymptotic variance. In an SVAR-X application in real activity, inflation, and the Fed Funds rate, with oil prices exogenous, they are able to reject two recursive schemes, depending on the chosen real activity variable. Further PML results can be found in HKO (2001).

More recently, Jarociński (2021) proposes to estimate the effects of four different dimensions of monetary policy based on non-Gaussianity using maximum likelihood estimators. His baseline approach uses the Student-t distribution, as in Lanne et al. (2017), but he also proposes an alternative estimator allowing for dependence. In particular, he allows for endogenously determined dependence in the tails of the shock distributions by designing a new partially dependent multivariate t-distribution that nests both independent and multivariate t-distributions as extreme cases. In the data, the level of dependence is found to be small, leading to minimal changes in results, but this contribution may be of interest in many applications.

Elsewhere, Chen and Bickel (2006) avoid the choice of likelihood by proposing a semi-parametric estimator based on the efficient score function, which they show is asymptotically efficient. Hafner et al. (2023) propose to maximise a kernel estimate of the likelihood in another non-parametric approach. Fiorentini and Sentana (2023) propose mixture of normals PML estimators. They prove that while for other likelihood approaches estimates of impact coefficients (and autoregressive parameters) will in general be consistent under misspecification, they are inconsistent for VAR intercepts and shock moments. Their proposed estimators provide consistent estimates of all parameters. Maxand (2020) considers maximum likelihood estimation of models with possibly more than one Gaussian shock, and provides tests for the number of such shocks.

Bayesian implementations of identification based on non-Gaussianity can be found in Lanne and Luoto (2020) (t-distributed shocks to compute the probability of sign restrictions holding), Anttonen et al. (2023) (generalised skewed t-distributed shocks, with MCMC methods proposed), Braun (2021) (Dirichlet process mixture model), Keweloh et al. (2023b) (skewed t-distributed shocks and potentially invalid proxy variables), and Lanne et al. (2023b) (t-distributed errors and a GARCH process for the shock volatilities).

4.4 Moment-based approaches

More recently, the literature has turned to moment-based estimators exploiting non-Gaussian-ity. Typically, these results relax the independence assumption and instead require uncorrelated shocks that satisfy a number of zero restrictions on co-skewness and/or co-kurtosis. These co-skewness and co-kurtosis conditions are implied by the stronger (and previously maintained) independence assumption. Unsurprisingly, this exercise (like similar efforts to allow dependence in the likelihoods above) presents a trade-off, since additional restrictions implied by independence can improve identification when valid. By virtue of selecting specific moments, all such approaches rely on specific deviations from Gaussianity, but in practice, the identification condition is heuristically thought of as "at least $n-1$ non-Gaussian shocks".

A first set of identification results focuses on using fourth moments, and in particular excess co-kurtosis restrictions. These date to Bonhomme and Robin (2009), who show that B is identified from the covariance and co-kurtosis (e.g., the collection equations $E[u_{it}u_{jt}u_{lt}u_{mt}]$) of reduced form errors u_t, provided the co-kurtosis tensor of ϵ_t is diagonal, as implied by independence, and at most one shock has zero excess kurtosis. Their proof is based on spectral arguments, in a factor model setting, and identification is global. They establish consistency and asymptotic normality of GMM estimators, as do the following papers. Guay (2021) considers the same moment conditions as Bonhomme and Robin (2009), but establishes local identification under the assumption of zero excess co-kurtosis by studying the rank of the Jacobian, which facilitates additonal results discussed below. Keweloh (2021) works from slightly different moments, written instead in terms of the underlying shocks, which, under a mutual independence assumption and normalisation, have the form

$$E[\epsilon_{it}^2] = 1 \tag{17}$$

$$E[\epsilon_{it}\epsilon_{jt}] = 0,\ i \neq j \tag{18}$$

$$E[\epsilon_{it}^2\epsilon_{jt}^2] = 1,\ i \neq j \tag{19}$$

$$E[\epsilon_{it}^3\epsilon_{jt}] = 0,\ i \neq j. \tag{20}$$

After restricting the set of permissible B matrices to those matching the permutation/scaling rule in Lanne et al. (2017) and assuming that $n-1$ shocks exhibit excess kurtosis, the covariance and co-kurtosis conditions globally identify B, following an argument similar to Comon (1994). Note that while each paper may write the assumptions and results somewhat differently, the identification conditions are indeed the same. The only difference is that writing the moments in terms of u_t versus ϵ_t entails fewer moments but requires the inversion of B, both of which may have finite sample consequences. Lanne and Luoto (2021) argue that B is identified if $E\left[\epsilon_{it}^3\epsilon_{jt}\right] = 0$ for at least $n(n-1)/2$ combinations of $i \neq j$. This condition appears weaker than those in the preceding three papers. However, this result does not in fact hold under the stated assumptions; rather, it requires all co-kurtosis restrictions (symmetric and asymmetric) implied by independence to be satisfied – the same as the previous papers. The authors rely on these conditions in their proof, which establishes local identification via the Jacobian of the identifying moments. Keweloh (2021) shows by counterexample that under the stated conditions, the model is only locally and not globally identified, and in particular not identified up to sign and permutation. Ultimately, all of the above studies require exactly the same restrictions on the dependence of the shocks. Lanne et al. (2023a) subsequently introduce the additional assumption that $n-1$ shocks have excess kurtosis of the same sign, under which they achieve global identification using only $n(n-1)/2$ higher moments – all of the symmetric co-kurtosis conditions implied by independence. However, asymptotic normality additionally requires $n(n-1)/2$ asymmetric co-kurtoisis restrictions. They argue that this is potentially far fewer than the moments required by Keweloh (2021). Nevertheless, they still require all fourth moments implied by independence to hold, even if they are not exploited for identification. They propose a moment selection procedure, although it is unlikely to perform well in the larger SVARs where it would be most useful.

A second strand of this literature uses third cumulants, or skewness, for identification, either separately or in conjunction with kurtosis. Again, all restrictions on co-skewness implied by independence are used for identification. In parallel to their result using kurtosis, Bonhomme and Robin (2009) show that under the restriction of zero co-skewness, i.e. $E[\epsilon_{it}^2 \epsilon_{jt}] = 0,\ i \neq j$, provided at most one shock has zero skewness, B is identified, following similar spectral arguments. Guay's (2021) identification argument also applies to sets of moments including skewness conditions alongside kurtosis conditions. The identification condition holds when all but one shock exhibits skewness and/or non-zero excess kurtosis. Moreover, he establishes partial identification results, such that the parameters corresponding to all skewed and/or non-mesokurtic shocks are identified regardless of the properties of the remaining shocks. He further shows that the rank can be decomposed into the sum of the ranks of certain blocks of the Jacobian. Importantly, this decomposition makes the identification conditions testable. The number of skewed shocks is equal to the rank of the coskewness matrix of the reduced form residuals and likewise the number of non-mesokurtic shocks is equal to the rank of the cokurtosis matrix of the reduced form residuals. He proposes a bootstrap procedure to implement this test. Keweloh's (2021) identification argument, based on Comon (1994), goes through whether skewness conditions, kurtosis conditions, or both, are used. An advantage of using skewness conditions alone is that they admit possible coheteroskedasticity in the data, an empirically relevant form of dependence that violates the diagonal coskewness assumptions discussed above.

As originally noted by Bonhomme and Robin (2009), estimation of the higher moments required can be challenging without in the absence of parametric restrictions due to relatively small sample sizes; this is especially true for inference, which requires up to eighth moments. Guay (2021) uses the identity weighting matrix due to difficulty estimating the efficient weighting matrix. For improved finite sample performance, Keweloh (2021) proposes the "fast-GMM" estimator, which uses a diagonal weighting matrix in which the weight on the covariance moments goes to infinity, so that the shocks are always whitened. The idea is similar to ICA, with the weighting matrix replacing the problem of minimising dependencies between the shocks with that of maximising non-Gaussianity, subject to satisfying the covariance moments. In larger models there is a significant computational advantage to this change, and he does not find efficiency loss relative to the efficient estimator in simulations. Moreover, he finds that the estimators of Lanne and Luoto (2021) and Gouriéroux et al. (2017) do not effectively exploit information found in skewed shocks, instead relying on excess kurtosis, and more generally that information contained in coskewness can be important. Lanne and Luoto (2021) propose 2-step, iterative, and CUE GMM estimators, as well as a test for overidentifying restrictions and a moment selection procedure based on Andrews (1999), although this procedure may struggle in larger SVARs. In simulations, they prefer the two-step procedure, which in general performs slightly worse than the baseline PML estimator of Gouriéroux et al. (2017), but better than their iterative estimator. Keweloh (2023) further studies the limitations of GMM estimators in short samples. In particular, he notes the challenge of precisely estimating the efficient weighting matrix (which requires up to eighth moments) and also that bias is typically introduced towards parameters implying shock variances below unity. He proposes to impose independence assumptions not just for identification, but also in the estimation of the efficient weighting matrix for the former and a continuously updated scaling term for the latter. Simulations show reduced bias and improved coverage. Discussed in more detail below, Keweloh (2024) provides a method to incorporate uncertain economic prior information into moment-based estimators exploiting non-Gaussianity.

In a minimum distance setting, Mesters and Zwiernik (2022) are able to relax the independence assumption in important ways. They show that in general, diagonality of any higher-order cumulant tensor is sufficient for identification up to sign and permutation, generalising away from previous work focused on third and fourth moments (e.g., Bonhomme and Robin (2009); Guay (2021); Keweloh (2021)). More importantly, they show that "reflectionally invariant restrictions", where the only non-zero cumulant tensor entries are those where each index appears an even number of times, are similarly sufficient for identification This is the first identification result exploiting non-Gaussianity and fourth moments that is able to accommodate co-heteroskedasticity in the errors, an empirically-relevant form of higher-order dependence. However, the shocks must satisfy an additional genericity condition. Working in a similar direction, Herwartz and Wang (2023a) combine the non-parametric approach of Hafner et al. (2023) with standardisation using a kernel estimate of time-varying volatility to identify B in the presence of co-heteroskedasticity.

5 Weak Identification Based on Higher Moments

It may be tempting to view identification based on higher moments as a free lunch for recovering causal effects in macroeconometrics. No "economic" assumptions are required – at least for local identification – and identification obtains provided that often uncontroversial properties of the data hold. However, the separation of statistical and economic assumptions is a false dichotomy; assuming that the variances of certain shocks has changed over time in a particular pattern, that all shocks are mutually independent, or that shocks have meaningful excess kurtosis has real economic content. More importantly, the higher moments leveraged by the identification schemes discussed above are all non-trivial to estimate in realistic macroeconomic sample sizes. At best, the schemes exploit the difference in second moments, but more generally rely on up to the fourth moments of the data, which typically will be very imprecisely estimated in at most 50 years of quarterly data – 200 observations.

Researchers are now generally familiar with the weak instruments problem, but the challenges posed by weak instruments apply to any identification scheme. As described formally by Stock and Wright (2000), when the objective function is relatively flat in the neighbourhood of the true parameters, standard inference methods will be invalid, exhibiting large size distortions and poor coverage. This occurs when deviations from homoskedasticity are small, relative to estimation error so additional values for the variances offer little information beyond the original $n(n-1)/2$ covariance equations (and deviations $\hat{B}$ from B are at best weakly rejected by the data in a non-trivial neighbourhood of B). Alternatively, when deviations from Gaussianity are small, higher moments offer little information beyond those of a Gaussian distribution – where they are completely redundant with the original covariance restrictions. The possibility that identification based on heteroskedasticity might only be "weak" – and thus that standard inference methods like Wald tests might perform poorly – dates back to at least Magnusson and Mavroeidis (2014), who consider how instability in moments can be used to sharpen identification in general, and propose identification-robust test statistics (S-statistics) tailored to such settings. They consider the Rigobon (2003) model as a leading example. More recently, Montiel Olea et al. (2022) argue that weak identification is likely present in many applications of identification based on higher moments.

Lewis (2022) studies the problem of weak identification via heteroskedasticity directly, again in the context of the Rigobon (2003) model. Rigobon and Sack (2004) show that in a bivariate model where only one variance changes that regime-based identification can be rewritten as an instrumental variables problem. In that setting, it is unsurprising that weak identification may arise; it does so when the variance that does change only does so by a small amount. In that case, the pre-test for weak instruments of Montiel Olea and Pflueger (2013) applies (with bias-based critical values), and the usual robust inference methods for IV estimators can be adopted. Lewis (2022) further characterises weak identification problems in models identified using variance regimes more generally, which arise when variance changes are close to proportional across regimes. In this general setting, the size-based Andrews (2018) procedure to detect weak identification in GMM estimators can be applied. In these models, robust inference is more complicated due to the projection problem: robust inference can have prohibitively conservative limiting distributions when the econometrician is only interested in a subset of the parameters in B. Lewis (2022) provides conditions under which limiting distributions providing exact size can be derived, which hinge on whether the remaining "nuisance" parameters can be uniquely determined from the data conditional on the null hypothesis. Empirical evidence suggests that weak identification is present in event studies based on daily financial data, a popular setting in applied practice. There are several limitations to the methods proposed. The bias-based pre-test is attractive, but only applies in a restricted bivariate model. The Andrews (2018) test requires the computation of robust confidence sets, which can be computationally demanding in larger models, and prohibitively conservative when the conditions provided for improved limiting distributions do not hold. Finally, those conditions place strong limits on the extent of weak identification, which may be hard to justify for $n > 2$.

Lütkepohl et al. (2020) provide a test of the identification condition for the Rigobon (2003) identification scheme using two regimes. They test whether equality can be rejected for each subset of eigenvalues in equation (7). This is not a test for weak identification, but rather a test of non-identification. A limitation of this test is that it presents a potentially substantial multiple testing problem when $n > 2$: to reject non-identification, as many tests as there are adjacent subsets of the integers $1, \ldots, n$ must be conducted. However, it remains the most attractive option in applied practice given the computational and performance issues of the Andrews (2018) procedure for larger models.

Similar tests for non-identification exist for other schemes. Lewis (2021) derives testable implications of that paper's non-parametric identification conditions and proposes a suitable Cragg-Donald statistic. Lanne and Saikkonen (2007) propose two LM-type tests for the order of the GARCH process driving the SVAR residuals, which can help determine whether the required $n - 1$ dimensions are present, or whether identification is only partial. Lütkepohl and Milunovich (2016) propose a further test and study all three in extensive simulations. The size and power of the tests and their ranking varies with DGPs and sample sizes, and they tend to be conservative in determining the correct model order. However, these papers only consider the case where there are r heteroskedastic shocks and $n - r$ homoskedastic shocks, rather than the models where all shocks may be heteroskedastic, but with a factor structure in the volatilities.

Turning to non-Gaussianity, serious effort has recently been applied to robust inference. Lee and Mesters (2021) and Hoesch et al. (2023) provide a robust inference approach based on a semi-parametric score statistic (with the non-parametric part corresponding on the shock distributions). The former considers identification via non-

Gaussianity from observed innovations, while the latter extends the results to SVARs and IRFs, and also shows how the score can be used to construct an efficient estimator. Inference on IRFs ultimately requires a Bonferroni step and projection methods. The confidence intervals constructed in Drautzburg and Wright (2021), based on sign restrictions and an independence assumption, are also robust to weak identification (via the independence assumption).

Recently, Amengual et al. (2022a) and Amengual et al. (2022b) provide tests of the assumptions for identification based on non-Gaussianity. Both are based on the estimated shocks. The former proposes moment-based tests for whether one (or more) shock(s) has a Gaussian distribution and for dependence in the shocks. The latter is a test for dependence amongst the shocks that compares the estimated joint cdf of the shocks to the estimated marginal cdfs. Both are based on the mixture of normals pseudo-maximum likelihood estimator of Fiorentini and Sentana (2023), for which influence functions allow adjustment for estimation error in the shocks. There is further work on testing the independence assumption alone, coming from the ICA literature, see for instance Matteson and Tsay (2017) and Davis and Ng (2023).

6 Open issues

6.1 Time-varying B

A possible tension in standard models identified via heteroskedasticity is that the structural parameters in B – the causal effects – are required to be constant over time, while the variances are allowed to change. Indeed, these two sources of variation are often included together in models where identification comes from other sources (e.g., Cogley and Sargent (2005), Primiceri (2005)). On the other hand, many DSGE models include time-varying volatility, but not changes in the deep structural parameters that dictate B in a VAR representation. In any case, arguments for identification based on higher moments hinge on the constancy of B. Nevertheless, it is possible to accommodate time-varying reduced form parameters, $A(L)$, and the failure to do so when demanded by the data can be a source of spurious heteroskedasticity (Sims (2002)). There have been several recent attempts to combine identification via heteroskedasticity with particular forms of time-variation in B. Typically, these approaches make use of additional regimes that would be overidentifying with constant coefficents in order to identify changes in a restricted number of coefficients. Note, however, that adding additional regimes cannot identify time-varying B and volatilities without such restrictions: each additional regime adds n new volatility parameters and n^2 new coefficients, but only $(n^2 + n)/2$ new equations. Typically, identification results in this literature provide local identification conditions under which the Jacobian of the reduced form covariance matrices with respect to the unrestricted parameters is full-rank.

Bacchiocchi and Fanelli (2015) and Bacchiocchi et al. (2018) propose a framework that nests that of Rigobon (2003). They consider two regimes with the covariance matrix of the shocks constant, but with two matrices of coefficients, B and $\tilde{B} \equiv B + W$, modeled as functions of a vector of $n^2 + n$ unknown parameters. Note that changes in shock variances can be subsumed into $\tilde{B} = B + W$, where each column is just rescaled by the new volatility additively instead of multiplicatively. They derive the rank condition for identification; unsurprisingly, changes in *relative* effects of the shocks can only be identified if some of the variances are restricted to be unchanged (via the matrix W); these are the same equations that were just-identifying with B fixed and all volatilities allowed to change. Angelini et al. (2019) go a step further. They likewise do not explicitly model heteroskedasticity, but exploit it to identify three different regimes of impact coefficients in a study of economic uncertainty. In particular, they keep the shock variances fixed at unity and model the impact coefficients as $B, B + W_1, B + W_1 + W_2$. This structure allows for changes in the volatilities via W_1 and W_2, as well as changes in the relative effects of the shocks. Identification is achieved by imposing a suitable number of zero restrictions in W_1 and W_2.

Brenna et al. (2023) take a different perspective. Studying macro-financial linkages, they note that there are too few uncontroversial restrictions that can be imposed on time variation in B to offer point identification. Instead, for a regime $i > 1$ they let $B_i = B + W_i$, construct the full set of combinations of zero restrictions that could be sufficient to identify the model, and recover the identified set of parameters across that collection of sets of possible identifying restrictions.

Dungey et al. (2015) marry the smooth-transition and GARCH-based identification approaches. They model volatilities as evolving according to a GARCH process and the time-varying B_t as a convex combination of $N + 1$ unit diagonal matrices,

$$B_t = (1 - S_N)\left[\ldots\{(1 - S_2)\left[(1 - S_1) B_0 + S_1 B_1\right] + S_2 B_2\}\ldots\right] + S_N B_N. \tag{21}$$

with $S_j = \left(1 + e^{-\gamma(x_t - d_j)}\right)^{-1}$, where d_j is the center of the transition between regimes $j-1$ and j and $x_t = t/T$ expresses time t as a fraction of the sample. The identification argument contends that when γ, the speed of transition, is large enough so that within each regime j, B_t is essentially B_j, then identification results like that of Milunovich and Yang (2013) can be applied within each regime. However, the true parameters only appear to be identified in the limit as $\gamma \to \infty$.

6.2 Choosing the right functional form

Amongst identification approaches based on heteroskedasticity, the econometrician is generally required to take a stance on the form of heteroskedasticity taken, at least to some extent. While now behind the frontier range of options for the volatility process, Lütkepohl and Netsunajev (2015) provide an early survey of possible models, highlighting their relative strengths and weaknesses. With the exception of Lewis (2021), all identification schemes rely on estimating the paths of the reduced form variances through time, at least for feasible implementations. While using a regime approach can be thought of as unrestrictive – since variance regimes can be estimated through time regardless of the true variance process – they will only offer sharp identifying information if there are distinct differences in structural variances across regimes. If the true variance process is poorly approximated by a regime structure, it is unlikely that identification will be strong, and at the very least valuable identifying information will be left on the table. However, as discussed in Section 6.2, even though non-parameteric estimators are available under fairly weak assumptions, it will often be preferable to specify a functional form for the variance process in finite samples for efficiency reasons.

Therefore, how should the researcher go about choosing the variance process to fit to the data? This remains an open question, with three possibilities. The first option is to use application-specific knowledge to choose an appropriate model. The second is to conduct statistical tests to learn the correct functional form, as proposed by Lütkepohl and Schlaak (2018). They provide various information criteria for determining the correct model of heteroskedasticity. They then evaluate their performance in a simulation study. They consider exogenously determined regimes, Markov switching, smooth transition, and GARCH DGPs. They find that the information criteria struggle to differentiate between different models for the structural variances, in terms of making the correct binary determination. They tend to favour the exogenous regime model, and in particular struggle to detect GARCH data. However, adopting the criteria can still help to reduce the MSE of impulse response estimates. Ultimately, they conclude that these tests can be helpful in ascertaining whether adequate heteroskedasticity is present in the data for identification, but are not yet well-suited to discriminating between different volatility models.

Elsewhere, Lewis (2021) and Bertsche and Braun (2022) conduct parallel simulation studies comparing the performance of various different estimators (in terms of

MSE and other criteria) when correctly and incorrectly specified, pointing to the third possibility of choosing a demonstrably robust functional form. In particular, they consider estimators based on AR(1) log stochastic volatility, GARCH, Markov Switching, and regime-based models for the structural variances, as well as a non-parametric estimator based on GMM; Lewis (2021) adds further estimators, including two exploiting non-Gaussianity. Both papers consider DGPs corresponding to each of these models. Both studies find that the AR(1) log stochastic volatility model is remarkably robust to misspeficiation. It performs very well when correctly specified, but even when badly misspecified often performs nearly as well as correctly specified estimators. This is not true of any of the other estimators, whose performance is generally very heterogeneous. Both papers recommend the stochastic volatility model for use in practice on the basis of its robustness to misspecification. However, beyond the simple flexibility of the DGP (*ex post*, paths for the latent volatilities can approximate those coming from any of the other variance processes, and even the fat tails of homoskedastic non-Gaussian DGPs), no theoretical explanation for this performance or justification for this guidance is available.

For the case of non-Gaussianity, similar questions arise, and there is, as yet, unfortunately little clear guidance for empirical researchers. There are maximum likelihood and PML approaches, typically based on variations on student-t distributions, and non-parametric estimators. For the latter, there is an important choice of which co-skewness and/or co-kurtosis restrictions to impose or of which contrast function to use for ICA. Moneta and Pallante (2022) compare a variety of estimators in a simulation study. They include FastICA, the PML estimator of Gouriéroux et al. (2017), and two other ICA approaches based on Givens matrices; unfortunately, they do not include recent moment-based estimators. Overall, they find that FastICA has an edge in terms of bias, efficiency, size distortions, and coverage, with PML slightly behind.

6.3 Combining multiple sources of identifying information

An exciting avenue for ongoing research focuses on combining identification based on higher moments with other identification schemes. This can serve three purposes. First, additional over-identifying assumptions – particularly economic ones – can be tested. Second, combining statistical identifying information with other types of identifying information, like an external instrument, zero restrictions, or sign restrictions, can serve to sharpen identification when the heteroskedasticity or non-Gaussianity may not be pronounced enough to provide strong identification in finite samples. Finally, it can also help to resolve the labeling indeterminacy; discussion is deferred to Section 6.4.

The first purpose is one of the key advantages of statistical identification in the first place. For example, B can be estimated based on higher moments alone, and then a simple joint Wald test for all of the entries above the diagonal can test whether recursive identification assumptions are rejected by the data. Alternatively, the same can be done using likelihood ratio tests etc. There are myriad examples of this approach to testing economic identification assumptions, including the empirical applications of virtually all of the references in this review; see for example Normandin and Phaneuf (2004); Lanne and Lütkepohl (2008); Herwartz and Lütkepohl (2014); Lütkepohl and Woźniak (2020); Lewis (2021); Bertsche and Braun (2022).

Combining higher moments with additional identifying assumptions to achieve sharper identification, for example when it is feared that the higher moments may only be weakly identifying, is still an emerging literature. Carriero et al. (2023) provide a wide-ranging treatment in a Bayesian framework. They propose algorithms to estimate SVARs combining heteroskedasticity with sign and narrative restrictions as well as external instruments. They argue that heteroskedasticity, since it is potentially point-identifying, can substantially reduce the identified sets resulting from sign or narrative restrictions. On the other hand, those restrictions can resolve the labeling problem associated with statistical identification. One key challenge to such strategies is understanding how conflicting identifying information will interact. Indeed, the identified set may be empty if two sources of identifying information are at odds; for instance, if no shock exists, according to the moments arising from heteroskedasticity, that satisfies stipulated sign restrictions. In that case, determining which set of assumptions is incorrect presents a further challenge. Lütkepohl and Schlaak (2022) consider a slightly different problem: they take the presence of heteroskedasticity (in regimes, say) as potential evidence of time-varying impact coefficients. They combine heteroskedasticity and an external instrument to test whether the column of B identified by the instrument changes across variance regimes. Bacchiocchi et al. (2023) derive identified sets resulting from partially identifying heteroskedasticity and zero or sign restrictions and provide methods to compute them as well as for robust Bayesian inference. Bacchiocchi and Kitagawa (2023) provide a comprehensive treatment of identification in SVARs with breaks, which nests models identified with heteroskedasticity regimes, but also allows for breaks in B and the reduced form parameters. They allow for additional information in the form of inequalities on various SVAR objects and stability restrictions and provide Bayesian and robust Bayesian algorithms and methods for valid inference.

Drautzburg and Wright (2021) combine sign restrictions with non-Gaussianity: they propose an identified set that is the intersection of that arising from sign restrictions and the set of models for which independence of the shocks cannot be rejected. Keweloh et al. (2023a) combine a block recursive structure with non-Gaussian shocks to propose an estimator that has advantages over purely statistical information in terms of performance, shock labelling, and weaker independence requirements. Braun (2021) develops a Bayesian framework combining priors involving sign restrictions and non-Gaussianity. He models each non-Gaussian shock as a univariate Dirichlet process mixture model, and shows that, combined with weak priors on important coefficients, the non-Gaussianity sharpens inference to deliver results similar to those under much strong restrictions in an application to oil markets. Keweloh et al. (2023b) propose a Bayesian framework that allows for potentially endogenous proxy variables, and show that that such endogeneity helps to reconcile the range of empirical estimates for fiscal multipliers. Keweloh (2024) combines non-Gaussianity with potentially invalid short-run restrictions, proposing an estimator with data-dependent shrinkage towards those restrictions. Herwartz and Wang (2023b) develop a point estimator that minimises the dependence of the implied structural shocks subject to sign restrictions. Crucil et al. (2023) exploit external instruments with non-Gaussianity to sharpen identification and partially resolve the shock labelling problem.

6.4 The shock labeling problem

One of the main challenges of purely statistical identification is the shock labeling problem – B is only ever identified up to column order by statistical information alone.[5] The answer, in general, is to use economic information to label the shocks, see Herwartz and Lütkepohl (2014) for an early detailed discussion. While this may seem self-defeating, having turned to statistical identification to avoid economic assumptions, the versions of economic assumptions required for labelling are generally much less restrictive than those required for point identification on their own. For example, rather than imposing a lower-triangular structure on B, the researcher can obtain the permutation of the identified columns of B that is “closest” to that structure, under the Frobenius norm, say. The economic assumptions do not determine the possible values of identified parameters – they are recovered on statistical information alone; rather, the economic information helps to choose amongst the statistically identified models, which is desirable if the restrictions are thought to be approximately but perhaps not literally true. Another alternative is to obtain a set of model-based causal effects (values for B, or

[5]This results in the challenging “label switching” problem for Bayesian inference, see e.g., Bacchiocchi and Kitagawa (2020) for a discussion.

IRFs at longer horizons) and choose the shock labeling that most closely matches the model predictions (see, e.g., Brunnermeier et al. (2021)). Augmenting the statistical identification assumptions with some economic assumption that selects a unique column order (or even selects a single column of interest) is enough to point identify the causal effects of interest. Lewis (2021) and Kilian and Lütkepohl (2017) (chapter 14) discuss this labeling problem in further detail.

There are also non-economic options for pinning down a unique value for B, beyond column order. For example, Lanne et al. (2017) propose to globally identify an SVAR based on non-Gaussianity using a series of transformations. Given any one of the $n!$ observationally equivalent identified matrices, they first normalise each column to have unit Euclidean norm. Next, they choose the unique column permutation for which the entries to the right of each diagonal element are smaller in absolute value than that diagonal element. Finally, they impose the unit diagonal normalisation. This can be thought of as an approximation to a lower-triangular matrix, since the permutation is chosen for which all of the entries above the diagonal are smaller than the corresponding diagonal entries. Another popular choice in the ICA literature, and recent papers using non-Gaussianity, is the Pham and Garat (1997) scheme, which chooses the permutation that maximises the product of the diagonal entries of B, after restricting them to be positive and making a unit variance assumption. While augmenting statistical assumptions with such column ordering rules achieves global identification, it does not directly assist in rendering identified shocks (or their effects) economically interpretable.

A further concern is the impact of the labeling problem on inference. While the asymptotic distribution of a given column permutation of B may be known (after fixing some statistical rule, like those mentioned above, for estimation), what is the distribution after subsequent permutations and re-scaling for economic labeling? Lewis (2021) applies results from the model selection literature to show that, provided a "consistent labeling criterion" is used, the error introduced by the labeling procedure is asymptotically negligible, and the asymptotic distribution of the permutation of interest is valid for the labeled estimator without modification. A consistent labeling criterion is a rule for choosing the column order of B that will choose the economically correct order with probability approaching 1 as $T \to \infty$. This meaningfully simplifies inference on the parameters of interest in practice.

7 Conclusion

To conclude, we review several areas fruitful for ongoing research.

Robust inference As discussed above, higher moments often likely only provide weakly identifying information in realistic sample sizes. While this fact has been noted in the theoretical literature for some time, it has only recently started to impact applied practice. In general, weakly identified SVARs require projection inference for a subset of the parameters, with available critical values proving sometimes prohibitively conservative, as noted by Lewis (2022) for heteroskedasticity and Lee et al. (2022) for non-Gaussianity. Lewis (2022) provides a solution for a particular class of models in the form of considerably sharper critical values, but no such results exist for the majority of applications. Robust inference methods that exhibit acceptable performance and are computationally conveninent remain elusive, and, as the literature comes to grips with the prevalence of weak identification, will be increasingly in demand.

Testing identification conditions Enlarging the range of settings in which the identification conditions are testable should be a priority. Since the identification conditions are in terms of the structural parameters, it is challenging to test them without assuming identification in the first place. To do so generally requires the derivation of testable implications in terms of only reduced form quantities. Several methods are discussed in Section 5. Testing approaches, especially those testing for weak identification as opposed to non-identification, that perform well and are computationally tractable for arbitrarily large models, remain a target for all of the identification schemes considered.

Combining identifying information As discussed in Section 6.3, several recent papers have turned to combining statistical identification with other forms of identification, aiming to combat the weakness of the identifying variation coming from higher moments, provide point identification when economic restrictions are only set-identifying, or test economic restrictions. This is a conceptually appealing strategy, since it addresses head-on the fact that many sources of identifying information fail to provide sharp identification. It has the potential to extploit the "best of both worlds" in terms of statistical and economic identifying information. The Bayesian literature is perhaps somewhat further advanced in this respect as a consequence of the fact that it is particularly appealing to combine higher moments with set identification approaches, which are most commonly implemented in a Bayesian framework. However, there is plenty of scope to pursue such developments in the frequentist framework as well, see for example Drautzburg and Wright (2021) and Keweloh (2024). Besides theoretical work, this area also presents a rich vein for empirical work, since the combination of all credible identifying information available will likely pay dividends.

Functional forms Section 6.2 outlined the challenges of choosing the best functional form for a particular application. For both identification based on heteroskedasticity and identification based on non-Gaussianity, there are meaningful decisions for the researcher to make in terms of whether to choose a non-parametric estimator or specify some particular functional form for the volatility process or distribution for the shocks. As discussed, current attempts to detect the best-fitting form of heteroskedasticity based on information criteria proved unsatisfactory in Lütkepohl and Schlaak (2018). That paper's study also predates the availability of identification results for a much richer range of functional forms, like the AR(1) log stochastic volatility model favoured by Bertsche and Braun (2022) and Lewis (2021). Thus, there is scope to both extend that paper's analysis and to investigate alternative testing methodologies that may offer better performance.

To my knowledge, there has not been a systematic comparison of the quickly multiplying options for implementation of identification via non-Gaussianity. The speed at which new implementations of identification based on non-Gaussianity are developed makes comprehensive systematic comparisons difficult; current options described in Section 4 include various parametric models, moment based approaches where the researcher must choose which higher moments to use and which restrictions to impose on cross-moments, and a plethora of non-parametric estimators from the ICA literature. Moneta and Pallante (2022) is the only simulation comparison of which I am aware. There is ample scope for further and updated simulation studies comparing the leading alternatives, as well as "pre-tests" to select suitable moments containing relevant identifying variation.

Separately, there is also scope for creative innovation under each of these identification approaches. First, Lewis (2021) justifies identification under a very wide class of persistent volatility models. This frees the researcher to exploit whatever DGP she thinks best suits the data, providing substantial ground for exploration. In the case of non-Gaussianity, while several early parametric implementations started from obvious fat-tailed distributions, the identification requirements likewise admit a continuum of possible distributions, and there is scope to propose new alternatives here as well. For example, Jarociński (2021) proposes a novel likelihood that allows for dependence in the tails of the shocks, motivated by empirical features of U.S. monetary policy.

Moreover, Mesters and Zwiernik (2022) provide exciting new results relaxing the independence assumption and opening the door for researchers to work on a much richer and more realistic range of distributions (including those with co-heteroskedasticity, for instance).

Presenting statistical identification to applied researchers While statistical identification often exploits relatively uncontroversial properties of the data, the mechanisms of identification can be much less clear. Many applied researchers want to know what *economic features* identify the parameters. Statistical properties of the data do have economic meaning. Unpacking statistical identification in economic terms can be challenging, and is an area where econometricians must do better. The original Rigobon (2003) paper is a great exemplar in this respect: it represents identification based on variance regimes graphically, and makes clear how identification obtains as the variances of two shocks change across them. Lewis (2021) attempts to reframe rather abstract conditions on the persistence and co-persistence of shock variances in terms of predictions from structural models and conceptual properties of the shocks, as well as interpreting periods of high and low volatility and particularly sizeable shocks through the lens of historical events. While it is natural to use information from the historical record to define variance regimes, there is no reason such narrative information cannot also be used to interpret the variation underlying any other approach using higher moments.

Similar arguments can be made with respect to non-Gaussianity. While somewhat more challenging, theory can also help to motivate the required independence or other restrictions on higher moments necessary to achieve identification. Guay (2021) provides a compelling graphical illustration of the identification argument, mirroring that in Rigobon (2003) for heteroskedasticity. Braun (2021) argues similarly, illustrating the power of non-Gaussianity for identifying supply and demand shocks in oil markets. Greater effort along these lines to build intuition for what drives identification, besides a collection of equations and associated rank conditions, will increase applied uptake and the appeal of future work in these areas.

References

(2001): *ICA by Maximum Likelihood Estimation*, John Wiley Sons, Ltd, chap. 9, 203–219.

Amengual, D., G. Fiorentini, and E. Sentana (2022a): "Moment tests of independent components," *SERIEs: Journal of the Spanish Economic Association*, 13, 429–474.

——— (2022b): "Specification tests for non-Gaussian structural vector autoregressions," Working Papers wp2022$_2$212, $CEMFI$.

Andrews, D. W. K. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–563.

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *The Review of Economics and Statistics*, 100, 337–348.

Angelini, G., E. Bacchiocchi, G. Caggiano, and L. Fanelli (2019): "Uncertainty across volatility regimes," *Journal of Applied Econometrics*, 34, 437–455.

Anttonen, J., M. Lanne, and J. Luoto (2023): "Bayesian Inference on Fully and Partially Identified Structural Vector Autoregressions," Tech. rep., SSRN.

Bacchiocchi, E., A. Bastianin, T. Kitagawa, and E. Mirto (2023): "Partially identified heteroskedastic SVARs," Tech. rep.

Bacchiocchi, E., E. Castelnuovo, and L. Fanelli (2018): "Gimme a break! Identification and estimation of the macroeconomic effects of monetary policy shocks in the U.S." *Macroeconomic Dynamics*, 22, 1613–1651.

Bacchiocchi, E. and L. Fanelli (2015): "Identification in Structural Vector Autoregressive Models with Structural Changes, with an Application to US Monetary Policy," *Oxford Bulletin of Economics and Statistics*, 77, 761–779.

Bacchiocchi, E. and T. Kitagawa (2020): "Locally- but not globally-identified SVARs," CeMMAP Working Papers 40/20.

——— (2023): "SVARs with breaks: Identification and inference," Tech. rep.

Bertsche, D. and R. Braun (2022): "Identification of Structural Vector Autoregressions by Stochastic Volatility," *Journal of Business & Economic Statistics*, 40, 328–341.

Blanchard, O. J. and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655–673.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98, 1887–1921.

Bonhomme, S. and J.-M. Robin (2009): "Consistent noisy independent component analysis," *Journal of Econometrics*, 149, 12–25.

Bouakez, H. and M. Normandin (2010): "Fluctuations in the foreign exchange market: How important are monetary policy shocks?" *Journal of International Economics*, 81, 139–153.

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

Brenna, F., de Graeve Ferre, and R. Wouters (2023): "Macro-financial feedbacks through time," Tech. rep.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021): "Feedbacks: Financial Markets and Economic Activity," *American Economic Review*, 111, 1845–79.

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

Cardoso, J. and A. Souloumiac (1993): "Blind beamforming for non-gaussian signals," *IEE Proceedings F (Radar and Signal Processing)*, 140, 362–370(8).

Carriero, A., M. Marcellino, and T. Tornese (2023): "Blended Identification in Structural VARs," Tech. rep.

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

Cogley, T. and T. J. Sargent (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302, monetary Policy and Learning.

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

Crucil, R., J. Hambuckers, and M. Simone (2023): "Do monetary policy shocks affect financial uncertainty? A non-Gaussian proxy SVAR approach," Tech. rep.

Darmois, G. (1953): "Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire," *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 21, 2–8.

Davis, R. and S. Ng (2023): "Time series estimation of the dynamic effects of disaster-type shocks," *Journal of Econometrics*, 235, 180–201.

Drautzburg, T. and J. H. Wright (2021): "Refining Set-Identification in VARs through Independence," Economics Working Paper Archive 64575, The Johns Hopkins University,Department of Economics.

Dungey, M., G. Milunovich, S. Thorp, and M. Yang (2015): "Endogenous crisis dating and contagion using smooth transition structural GARCH," *Journal of Banking Finance*, 58, 71–79.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

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

Geary, R. C. (1941): "Inherent Relations between Random Variables," *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, 47, 63–76.

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

——— (2019): "Identification and Estimation in Non-Fundamental Structural VARMA Models," *The Review of Economic Studies*, 87, 1915–1953.

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

Hafner, C. M., H. Herwartz, and S. Wang (2023): "Causal inference with (partially) independent shocks and structural signals on the global crude oil market," LIDAM Discussion Papers ISBA 2023004, Université catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA).

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, internally Consistent Modeling, Aggregation, Inference and Policy.

Herwartz, H. and S. Wang (2023a): "Consistent Statistical Identification of SVARs Under (Co-)heteroskedasticity of Unknown Form," Tech. rep.

——— (2023b): "Point estimation in sign-restricted SVARs based on independence criteria with an application to rational bubbles," *Journal of Economic Dynamics and Control*, 151, S0165188923000362.

Hoesch, L., A. Lee, and G. Mesters (2023): "Locally Robust Inference for Non-Gaussian SVAR models," Working Papers 1367, Barcelona School of Economics.

Hyvärinen, A. (1997): "New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit," in *Advances in Neural Information Processing Systems*, ed. by M. Jordan, M. Kearns, and S. Solla, MIT Press, vol. 10.

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

Hyvärinen, A., K. Zhang, S. Shimizu, and P. O. Hoyer (2010): "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity," *Journal of Machine Learning Research*, 11, 1709–1731.

Hyvärinen, A. (2013): "Independent component analysis: recent advances," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371, 20110534.

Jarociński, M. (2021): "Estimating the Fed's Unconventional Policy Shocks," Working Paper Series 20210, European Central Bank.

Keweloh, S. (2023): "Structural Vector Autoregressions and Higher Moments: Challenges and Solutions in Small Samples." .

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.

——— (2024): "Uncertain Short-Run Restrictions and Statistically Identified Structural Vector Autoregressions," .

Keweloh, S. A., S. Hetzenecker, and A. Seepe (2023a): "Monetary policy and information shocks in a block-recursive SVAR," *Journal of International Money and Finance*, 137, 102892.

Keweloh, S. A., M. Klein, and J. Prüser (2023b): "Estimating Fiscal Multipliers by Combining Statistical Identification with Potentially Endogenous Proxies," .

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, Themes in Modern Econometrics, Cambridge University Press.

King, M., E. Sentana, and S. Wadhwani (1994): "Volatility and Links between National Stock Markets," *Econometrica*, 62, 901–933.

Lanne, M., K. Liu, and J. Luoto (2023a): "Identifying Structural Vector Autoregression via Leptokurtic Economic Shocks," *Journal of Business & Economic Statistics*, 41, 1341–1351.

——— (2023b): "Identifying Structural Vector Autoregressions Via Non-Gaussianity of Potentially Dependent Structural Shocks," .

Lanne, M. and J. Luoto (2020): "Identification of Economic Shocks by Inequality Constraints in Bayesian Structural Vector Autoregression," *Oxford Bulletin of Economics and Statistics*, 82, 425–452.

——— (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.

Lanne, M. and P. Saikkonen (2007): "A Multivariate Generalized Orthogonal Factor GARCH Model," *Journal of Business & Economic Statistics*, 25, 61–75.

Lee, A., L. Hoesch, and G. Mesters (2022): "Locally Robust Inference for Non-Gaussian SVAR Models," Working Papers 1357, Barcelona School of Economics.

Lee, A. and G. Mesters (2021): "Locally Robust Inference for Non-Gaussian Linear Simultaneous Equations Models," Working Papers 1278, Barcelona School of Economics.

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

——— (2022): "Robust Inference in Models Identified via Heteroskedasticity," *The Review of Economics and Statistics*, 104, 510–524.

Lütkepohl, H., M. Meitz, A. Netšunajev, and P. Saikkonen (2020): "Testing identification via heteroskedasticity in structural vector autoregressive models," *The Econometrics Journal*, 24, 1–22.

Lütkepohl, H. and G. Milunovich (2016): "Testing for identification in SVAR-GARCH models," *Journal of Economic Dynamics and Control*, 73, 241–258.

Lütkepohl, H. and A. Netsunajev (2015): "Structural Vector Autoregressions with Heteroskedasticity: A Comparison of Different Volatility Models," Tech. rep.

Lütkepohl, H. and A. Netšunajev (2017a): "Structural vector autoregressions with smooth transition in variances," *Journal of Economic Dynamics and Control*, 84, 43–57.

——— (2017b): "Structural vector autoregressions with smooth transition in variances," *Journal of Economic Dynamics and Control*, 84, 43–57.

Lütkepohl, H. and T. Schlaak (2018): "Choosing Between Different Time-Varying Volatility Models for Structural Vector Autoregressive Analysis," *Oxford Bulletin of Economics and Statistics*, 80, 715–735.

——— (2022): "Heteroscedastic Proxy Vector Autoregressions," *Journal of Business & Economic Statistics*, 40, 1268–1281.

Lütkepohl, H. and T. Woźniak (2020): "Bayesian inference for structural vector autoregressions identified by Markov-switching heteroskedasticity," *Journal of Economic Dynamics and Control*, 113, 103862.

Magnusson, L. M. and S. Mavroeidis (2014): "Identification Using Stability Restrictions," *Econometrica*, 82, 1799–1851.

Matilainen, M., K. Nordhausen, and H. Oja (2015): "New independent component analysis tools for time series," *Statistics Probability Letters*, 105, 80–87.

Matteson, D. S. and R. S. Tsay (2017): "Independent Component Analysis via Distance Covariance," *Journal of the American Statistical Association*, 112, 623–637.

Maxand, S. (2020): "Identification of independent structural shocks in the presence of multiple Gaussian components," *Econometrics and Statistics*, 16.

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, 1212–47.

Mesters, G. and P. Zwiernik (2022): "Non-Independent Components Analysis," Working Papers 1358, Barcelona School of Economics.

Milunovich, G. and M. Yang (2013): "On Identifying Structural VAR Models via ARCH Effects," *Journal of Time Series Econometrics*, 5, 117–131.

Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013): "Causal Inference by Independent Component Analysis: Theory and Applications*," *Oxford Bulletin of Economics and Statistics*, 75, 705–730.

Moneta, A. and G. Pallante (2022): "Identification of Structural VAR Models via Independent Component Analysis: A Performance Evaluation Study," *Journal of Economic Dynamics and Control*, 144, 104530.

Montiel Olea, J. L. and C. Pflueger (2013): "A Robust Test for Weak Instruments," *Journal of Business & Economic Statistics*, 31, 358–369.

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.

Nakamura, E. and J. Steinsson (2018): "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*," *The Quarterly Journal of Economics*, 133, 1283–1330.

Normandin, M. and L. Phaneuf (2004): "Monetary policy shocks:: Testing identification conditions under time-varying conditional volatility," *Journal of Monetary Economics*, 51, 1217–1243.

Pham, D. T. and P. Garat (1997): "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Transactions on Signal Processing*, 45, 1712–1725.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821–852.

Reiersøl, O. (1950): "Identifiability of a Linear Relation between Variables Which Are Subject to Error," *Econometrica*, 18, 375–389.

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

Rigobon, R. and B. Sack (2003): "Measuring The Reaction of Monetary Policy to the Stock Market*," *The Quarterly Journal of Economics*, 118, 639–669.

——— (2004): "The impact of monetary policy on asset prices," *Journal of Monetary Economics*, 51, 1553–1575.

Sentana, E. (1992): "Identification of multivariate conditionally heteroskedastic factor models," Tech. rep.

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

Shimizu, S., P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006): "A Linear Non-Gaussian Acyclic Model for Causal Discovery," *Journal of Machine Learning Research*, 7, 2003–2030.

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

——— (2002): "Comment on Cogley and Sargentís ëEvolving Post-World War II U.S. Ináation Dynamics," in *NBER Macroeconomics Annual 2001*, ed. by B. S. Bernanke and K. Rogoff, Cambridge, Mass.: MIT Press, vol. 16, 373–379.

Skitovich, V. (1953): "On a property of the normal distribution," *Doklady Akademii Nauk SSSR*, 89, 217–219.

STOCK, J. H. AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

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.

WEI, T. (2015): "A Convergence and Asymptotic Analysis of the Generalized Symmetric FastICA Algorithm," *IEEE Transactions on Signal Processing*, 63, 6445–6458.

WRIGHT, P. (1928): *The Tariff on Animal and Vegetable Oils*, Investigations in international commercial policies, Macmillan.