



Helsinki
Center
of
Economic
Research

Discussion Papers

GMM Estimation of Non-Gaussian Structural Vector Autoregression

Markku Lanne
University of Helsinki and HECER

and

Jani Luoto
University of Helsinki and HECER

Discussion Paper No. 423
January 2018

ISSN 1795-0562

GMM Estimation of Non-Gaussian Structural Vector Autoregression*

Abstract

We consider estimation of the structural vector autoregression (SVAR) by the generalized method of moments (GMM). Given non-Gaussian errors and a suitable set of moment conditions, containing a sufficient number of relevant co-kurtosis conditions, the GMM estimator is shown to achieve global identification of the parameters of the SVAR model up to changing the signs of the structural shocks. We also propose a procedure, based on well-known moment selection criteria, to find the optimal set of moment conditions among the sets that guarantee identification. According to simulation results, the finite-sample performance of our estimation method is comparable, or even superior to that of the recently proposed pseudo maximum likelihood estimators. The two-step estimator is found to outperform the alternative GMM estimators. An empirical application to a small macroeconomic model estimated on postwar U.S. data illustrates the use of the methods.

JEL Classification: C32

Keywords: structural VAR model, non-Gaussian time series, generalized method of moments

Markku Lanne

Faculty of Social Sciences
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: markku.lanne@helsinki.fi

Jani Luoto

Faculty of Social Sciences
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: jani.luoto@helsinki.fi

* Financial support from the Academy of Finland (grant 308628) is gratefully acknowledged.

1 Introduction

The structural vector autoregressive (SVAR) model is one of the most popular tools in empirical macroeconomics and finance. It is obtained by imposing identifying restrictions on a vector autoregression (VAR), which is a purely statistical model summarizing the joint dynamics of a number of time series. In order for the SVAR model to yield meaningful economic interpretations, such restrictions must, in general, be motivated by information from various outside sources, such as economic theory, or institutional knowledge. Finding credible identifying restrictions can be challenging, and in case they are only sufficient to exactly identify the parameters of the SVAR model, they are not testable. Moreover, different identification schemes may yield quite different results, and the comparison of alternative identification strategies is typically not possible.

In the recent literature, a number of approaches to statistically identifying the SVAR model have been introduced. Typically, they make use of non-Gaussianity of the errors of the SVAR model that may show up as structural breaks in their covariance matrix, their conditional heteroskedasticity, or their following a parametric non-Gaussian distribution (for a survey of the relevant literature, see Kilian and Lütkepohl (2017, Chapter 14)). Because of non-Gaussianity, the parameters of the SVAR model are statistically identified, but, in contrast to other kinds of identifying restrictions, statistical identification rarely provides any economic interpretation. However, in the identified model, testing and contrasting alternative identification schemes with an economic motivation becomes possible. The economic restrictions that are not rejected, can then convincingly be used in the empirical analysis. Statistical identification may also be combined with economic information such as the signs of the impact effects of economic shocks implied by an economic model to facilitate interpretation (see, e.g., Lanne and Luoto (2016) and the references therein).

In this paper, we propose a generalized method of moments (GMM) estimator of the parameters of the SVAR model, with moment conditions that are informative when the error term of the model is non-Gaussian. Its closest counterpart is the maximum likelihood (ML) estimator of Lanne, Meitz, and Saikkonen (2017) that Gouriéroux, Monfort, and Renne (2017) have recently extended to pseudo ML (PML) estimators. In addition, it bears a resemblance to Herwartz's (2015) estimator, which is based on finding the rotation of orthogonalized errors that maximizes the p-value of a test of independence.

Our estimator has at least three advantages compared to its close counterparts previ-

ously put forth in the statistical identification literature. First, in contrast to the maximum likelihood (ML) estimator of Lanne et al. (2017), there is no need to specify an explicit non-Gaussian distribution. While the PML estimators of Gouriéroux et al. (2017) are, to some extent, robust with respect to misspecification of the error distributions, our approach is simpler, yet according to simulation results, its performance seems comparable to their PML estimator and superior to their recursive PML estimator. Second, unlike Lanne et al., Gouriéroux et al. (2017), and Herwartz (2015), we do not assume the structural errors to be independent, but only mutually orthogonal with, a number of additional co-kurtosis restrictions. As pointed out by Kilian and Lütkepohl (2017, Chapter 14.5), the independence assumption may be problematic because there is not necessarily any linear transformation that makes the errors of the reduced-form VAR model independent. In particular, our assumptions allow for various forms of joint conditional heteroskedasticity often found in economic data. Finally, both Lanne et al. and Gouriéroux et al. have to impose a number of technical restrictions to uniquely identify the parameters in addition to assuming non-Gaussianity and independence. In our setup, in turn, the corresponding additional restrictions are dictated by the moment conditions that estimation is based on, and they can be determined by well-known moment selection criteria. Hence, our estimator is completely driven by the data.

The GMM has previously been employed in estimating SVAR models in at least two contexts. First, Bernanke and Mihov (1995) showed consistency and asymptotic normality of the GMM estimator of SVAR models over-identified by short-run restrictions. Second, the GMM has been employed in the recent literature on identification of SVAR models by external instruments (see, e.g., Montiel Olea, Stock and Watson (2016)). However, to the best of our knowledge, this paper is the first that makes use of non-Gaussianity of the errors of the SVAR model in the GMM framework to facilitate identification of its parameters.

The rest of the paper is organized as follows. In Section 2, we introduce the SVAR model along with the central assumptions. Section 3 is concerned with statistical inference in the GMM framework. In particular, in Subsection 3.1, we discuss the implementation of the GMM estimator in the SVAR model, while Subsection 3.2 is devoted to the specification of moment conditions. In Subsection 3.3, we introduce regularity conditions under which the GMM estimator is consistent and asymptotically normal, and in Subsection 3.4, we describe our moment selection procedure. Subsection 3.5 contains some finite-sample simulation results. In Section 4, we illustrate the use of the GMM estimator in an empirical application

to a small U.S. macroeconomic model. Finally, Section 5 concludes. The detailed discussion on the conditions for local and global identification as well as the proofs of the related propositions are deferred to the Appendix.

2 Model

We consider the structural VAR (SVAR) model of order p ,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B \varepsilon_t, \quad (1)$$

where y_t is the n -dimensional time series of interest, ν ($n \times 1$) is an intercept term, A_1, \dots, A_p and B ($n \times n$) are parameter matrices with B nonsingular, and ε_t ($n \times 1$) is a serially uncorrelated strictly stationary error term with zero mean and identity covariance matrix.

We further assume y_t to be stationary, i.e.,

$$\det A(z) \stackrel{def}{=} \det (I_n - A_1 z - \cdots - A_p z^p) \neq 0, \quad |z| \leq 1. \quad (2)$$

In the literature, model (1) is often referred to as the B-model (see, e.g., Lütkepohl 2005, Chapter 9), and it is the most convenient formulation when the main emphasis is on impulse response analysis. An alternative SVAR formulation is the so-called A-model (Lütkepohl 2005, Chapter 9), obtained by left-multiplying (1) by the inverse of B :

$$A_0 y_t = \nu^* + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t, \quad (3)$$

where ε_t is as in (1), $A_0 = B^{-1}$, $\nu^* = B^{-1}\nu$, and $A_j^* = B^{-1}A_j$ ($j = 1, \dots, p$). Model (3) is useful when the main interest is on quantifying the instantaneous relations between the variables included in y_t .

Irrespective of the formulation, the central problem in SVAR analysis is the identification of the matrix B (or its inverse A_0) embodying the contemporaneous simultaneities. Recently, Lanne et al. (2017) showed that identification of B (up to permutation and scaling of its columns) can be reached when the error term ε_t is serially uncorrelated, and its components are contemporaneously independent and at most one of them is Gaussian. Similar results have been put forth in the related literature by Hyvärinen et al. (2010), and Moneta et al. (2013), *inter alia*, but they all assume ε_t to be an independent and identically distributed process (instead of being just serially uncorrelated).

The moment conditions that we impose in GMM estimation are inspired by the assumptions of Lanne et al. (2017). In particular, we make use of non-Gaussianity of the errors of the SVAR model, which implies different co-kurtosis conditions. However, we do not assume the components of the error term to be independent, but only contemporaneously uncorrelated as is typically the case in SVAR analysis. Specifically, we make the following assumption:

Assumption 1.

- (i) *The error process $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt})'$ is a sequence of (strictly) stationary random vectors with each component ε_{it} , $i = 1, \dots, n$, having mean zero and variance unity.*
- (ii) *The components $\varepsilon_{1t}, \dots, \varepsilon_{nt}$ are (mutually) orthogonal and at most one of them has a Gaussian marginal distribution.*
- (iii) *The components $\varepsilon_{it}, \dots, \varepsilon_{nt}$ are uncorrelated in time, i.e., $Cov(\varepsilon_{it}, \varepsilon_{i,t+k}) = 0$ for all $k \neq 0$.*

As pointed out above, Lanne et al. (2017) prove identification of matrix B in (1) (its inverse A_0 in (3)) only up to permutation and scaling of its columns (rows). In other words, they show that there is a class of observationally equivalent SVAR models, each with different signs and ordering of the structural shocks in the vector ε_t . Lack of unique identification hampers statistical inference: the derivation of the asymptotic properties of the maximum likelihood estimator requires additional restrictions to pinpoint one particular member of the class of SVAR models. As we will discuss in Section 3.2, in the GMM framework, uniqueness with respect to permutations can be achieved, provided the set of moment conditions contains certain asymmetric co-kurtosis conditions that are informative about the parameters in the presence of non-Gaussian errors. However, it is still necessary to introduce restrictions to fix the signs of the shocks. To that end, it suffices to set one element in each column of B positive (or negative); in the empirical application of Section 4, we will set its diagonal elements positive in estimation. In impulse response analysis, the columns may be rescaled to obtain impulse responses of shocks with desired sign and size.

Because matrix A_0 in the A-model (3) is obtained by inverting matrix B in the B-model, it is identified up to permutation and multiplication by -1 of its rows without

further restrictions under the same assumptions as the matrix B . Analogously to the B-model, by appropriately selecting the moment conditions and restricting one element on each row positive (or negative), ambiguity concerning the ordering of the equations and signs of the elements of the A_0 matrix can be resolved. However, the equations of the model cannot be labeled or provided with economic interpretation without additional (non-sample) information. For instance, the i th equation cannot necessarily be interpreted as the equation of the i th variable in y_t .

3 Statistical Inference

3.1 GMM Estimator

Models (1) and (3) can be estimated by minimizing

$$Q_T(\theta) = T^{-1} \sum_{t=1}^T f(v_t, \theta)' W_T T^{-1} \sum_{t=1}^T f(v_t, \theta), \quad (4)$$

where $\theta = (\nu', \text{vec}(A_1)', \dots, \text{vec}(A_p)', \text{vec}(B)')$ is a $(k \times 1)$ vector of $k \equiv n + (p + 1)n^2$ parameters to be estimated, v_t , $t = 1, 2, \dots, T$, is a vector of random variables consisting of y_t , its lags and deterministic terms. W_T is a $(q \times q)$ positive semi-definite matrix, potentially dependent on data, that converges to a positive definite weighting matrix of constants, W , containing the weights of the sample counterparts of the $(q \times 1)$ vector of population moment conditions

$$E[f(v_t, \theta_0)] = 0, \quad (5)$$

where θ_0 denotes the true value of θ . For the consistency of the GMM estimator, the moment conditions should only hold at one value (θ_0) in the entire parameter space, (see Section 3.3). Finding a convenient condition for global identification is, in general, difficult in the context of a nonlinear model such as the SVAR model, but in Section 3.2, we argue that, in our setup, by a suitable selection of moment conditions, global identification in a given SVAR model is achieved (see Proposition 2 and the discussion following it).

In order for the weaker condition of local identification to be satisfied, certain combinations of co-kurtosis conditions are ruled out, as shown in Proposition 1 in Section 3.2. This condition states that

$$\text{rank}\{E[\partial f(v_t, \theta_0)/\partial \theta']\} = k.$$

In other words, the matrix of expected partial derivatives of $f(v_t, \theta)$ with respect to the parameters evaluated at the true parameter values θ_0 is of full column rank. It follows, that for local identification, there must necessarily be at least k moment conditions. If $q > k$, it may be possible to run a test of over-identifying restrictions as a general specification test, as discussed below in Section 3.4.

In case of over-identification ($q > k$), inference may be sensitive to the choice of the weighting matrix W . Therefore it is, in general, desirable to base inference on the most accurate estimator, and as shown by Hansen (1982), the efficient estimator with minimum asymptotic variance is obtained by setting $W = S^{-1}$, the inverse of the long-run covariance matrix of the moment conditions, S . The latter can be estimated consistently (under regularity conditions, see Newey and West 1994) as the following heteroskedasticity and autocorrelation covariance (HAC) matrix:

$$\hat{S}_{HAC} = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega_{i,T} \left(\hat{\Gamma}_i + \hat{\Gamma}'_i \right),$$

where $\hat{\Gamma}_i$ is a consistent estimator of Γ_i , the i th autocovariance matrix of $f(v_t, \theta_0)$. The HAC estimator allows for heteroskedasticity and autocorrelation in the moment conditions, and the *bandwidth* parameter b_T embedded in the weights $\omega_{i,T}$ (or kernel) controls for the number of autocovariances included in the HAC estimator. A number of different kernels have been put forth in the GMM literature, including the Bartlett, Parzen and Quadratic Spectral kernels, but according to the simulation evidence of Newey and West (1994), the bandwidth is far more important for the finite-sample performance of the HAC estimator than the choice of the kernel, and they propose an automatic bandwidth selection procedure, which, coupled with the Bartlett kernel, we also employ in Sections 3.5 and 4.

In practice, estimation can be carried out in at least three different ways using numerical optimization methods. First, Hansen's (1982) two-step estimator is obtained by first minimizing (4) with W_T suboptimal (such as the identity matrix), and then re-estimating θ based on \hat{S}_{HAC} computed using the first-step estimator of θ . Second, this procedure can be continued iteratively until the estimate of θ converges to obtain the iterated GMM estimator. Finally, the continuous updating GMM estimator of Hansen, Heaton and Yaron (1996) acknowledges the dependence of the efficient weighting matrix on the parameters, and obtains the GMM estimator of θ by minimizing with respect to θ ,

$$T^{-1} \sum_{t=1}^T f(v_t, \theta)' S_T(\theta)^{-1} T^{-1} \sum_{t=1}^T f(v_t, \theta),$$

where

$$S_T(\theta) = \Gamma_{0,T}(\theta) + \sum_{i=1}^{T-1} \omega_{i,T} [\Gamma_{i,T}(\theta) + \Gamma_{i,T}(\theta)']$$

is of the same form as the HAC estimator discussed above. All three estimation methods are implemented in the R package `gmm` (Chaussé 2015) that we have used to produce the empirical and simulation results in this paper. As discussed in Section 3.3 below, all three estimators are consistent under regularity conditions. However, they may have different finite-sample properties, and the simulation results in the previous literature tend to favor the iterated and continuous updating estimators. However, such results may not be very helpful as they seem to depend considerably on the particular model. As a matter of fact, our limited simulation study in Section 3.5 pertaining to the estimation of the SVAR model suggests that the two-step estimator is superior to the other GMM estimators.

3.2 Moment Conditions

As discussed in Section 2, matrix B (its inverse A_0) is identified up to permutation and multiplication by -1 of its columns (rows) if the components of the error term ε_t are mutually independent and at most one of them is Gaussian. This result, recently shown by Lanne et al. (2017), adapted to the properties of the SVAR model incorporated in our Assumption 1, suggests a number of moment conditions that we next introduce, and then discuss them from the viewpoint of identification. In particular, we show that the components of the error term ε_t need not be mutually independent for identification, but, in addition to Assumption 1, a number of co-kurtosis conditions are sufficient. It should be borne in mind that identification of the parameters indeed depends on non-Gaussianity of (at least $n - 1$ of) the components of the error term. Therefore, it is always important to start the empirical analysis by checking whether the residuals of the reduced-form VAR model exhibit normality. If they turn out to be Gaussian, the moment conditions discussed below are not going to be sufficiently informative for identification. For ease of exposition, below we for the most part explicitly refer only to the B-model formulation (1), but it is to be understood that everything applies to the A-model (3) as well, with obvious modifications.

Let us, for notational convenience, rewrite model (1) as

$$y_t = \Gamma \mathbf{x}_{t-1} + B\varepsilon_t, \tag{6}$$

where the $((np + 1) \times 1)$ vector $\mathbf{x}_{t-1} = (1, y'_{t-1}, \dots, y'_{t-p})'$. From Assumption 1(*i*) and the lags of y_t being predetermined, we obtain the following $2n + pn^2$ moment conditions:

$$E(\varepsilon_t \otimes \mathbf{x}_{t-1}) = 0_{n(np+1) \times 1} \quad (7a)$$

$$E(\varepsilon_{it}^2) - 1 = 0, \quad i = 1, \dots, n \quad (7b)$$

where \otimes denotes the Kronecker product. It is implicitly assumed that the lag length p is sufficient to make the components of the error term ε_t serially uncorrelated as stated in Assumption 1(*iii*). Furthermore, mutual orthogonality of the components of ε_t in Assumption 1(*ii*) implies $n(n - 1)/2$ orthogonality conditions of the form

$$E(\varepsilon_{it}\varepsilon_{jt}) = 0, \quad i \neq j. \quad (7c)$$

The $2n + pn^2 + n(n - 1)/2$ moment conditions in (7a)–(7c) are not yet sufficient to identify the $n + (p + 1)n^2$ parameters of the SVAR model, but at least $n(n - 1)/2$ additional conditions are necessarily needed. To that end, we invoke co-kurtosis conditions implied by non-Gaussianity of (at least $n - 1$ of) the components of the error term ε_t in Assumption 1(*ii*). It is well known that co-kurtosis of two Gaussian random variables is a function of their variances and the correlation coefficient between them (see, e.g., Kendall and Stuart 1977, p. 94), whereas this need not be the case if either (or both) of the variables is non-Gaussian. Hence, co-kurtosis conditions can be informative in the presence of non-Gaussianity, while in the Gaussian case they provide no information over and above conditions (7a)–(7c). The co-kurtosis properties of economic shocks have recently been utilized in examining the effects of macro risks in a different econometric setup by Bekaert, Engstrom, and Ermolov (2017). Our idea is to base estimation on imposing asymmetric and symmetric co-kurtosis to take values that would prevail if the structural errors were independent. Hence, we obtain shocks that are close to being independent without actually imposing independence, and thus allowing for various forms of conditional heteroskedasticity, among other things.

Let us first consider asymmetric co-kurtosis conditions of the form

$$E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0, \quad i \neq j, \quad (8)$$

which are particularly informative if any of the errors follows a skewed distribution. However, also in the absence of skewness (and presence of non-Gaussianity), they may provide useful additional information for estimation because then, even if conditions (7a)–(7c) hold, $E(\varepsilon_{it}^3 \varepsilon_{jt}) = Cov(\varepsilon_{it}^3, \varepsilon_{jt}) + E(\varepsilon_{it}^3)E(\varepsilon_{jt}) = Cov(\varepsilon_{it}^3, \varepsilon_{jt})$ may deviate from zero if $\theta \neq \theta_0$. In

contrast, under Gaussianity, condition (7c) implies independence of ε_{it} and ε_{jt} , and hence of ε_{it}^3 and ε_{jt} such that $E(\varepsilon_{it}^3 \varepsilon_{jt}) = Cov(\varepsilon_{it}^3, \varepsilon_{jt}) + E(\varepsilon_{it}^3)E(\varepsilon_{jt}) = E(\varepsilon_{it}^3)E(\varepsilon_{jt}) = 0$.

If all the components of ε_t are non-Gaussian, the local identification condition (6) is satisfied whenever the set of moment conditions contains any $n(n-1)/2$ asymmetric co-kurtosis conditions in addition to conditions (7a)–(7c), as stated in Proposition 1 below. However, if one of the components of ε_t is suspected to be Gaussian, the exactly locally identifying asymmetric co-kurtosis conditions must be such that they do not involve its third power. For instance, if ε_{1t} is Gaussian in a trivariate model, the set of moment conditions containing the asymmetric co-kurtosis conditions $E(\varepsilon_{1t}^3 \varepsilon_{2t})$, $E(\varepsilon_{2t}^3 \varepsilon_{1t})$ and $E(\varepsilon_{2t}^3 \varepsilon_{3t})$ does not yield local identification, whereas the set where $E(\varepsilon_{1t}^3 \varepsilon_{2t})$ is replaced by, say, $E(\varepsilon_{3t}^3 \varepsilon_{2t})$, does. This is not very restrictive, however, as it concerns only the case where the number of moment conditions q equals the number of parameters k , while for $q > k$, the set of asymmetric co-kurtosis conditions can contain any of the components of ε_t in their third power.

Proposition 1. *(Local identification) Suppose all n components of ε_t are non-Gaussian. Then moment conditions (7a)–(7c), and $n(n-1)/2$ asymmetric co-kurtosis conditions of the form (8) exactly locally identify the parameters of SVAR model (1) (SVAR model (3)) up to permutation and multiplication of the columns of B by -1 . If one of the components of ε_t is Gaussian, the asymmetric co-kurtosis conditions must not involve its third power.*

Proof. See the Appendix. □

The asymmetric co-kurtosis conditions are particularly useful because by including a suitable collection of $n(n-1)/2$ of them, local and global (exact) identification can be reached (up to multiplying any of the columns of B by -1). Intuitively this follows from the fact that, for a given SVAR model (i.e., a model with a fixed permutation of the columns of B), $E(\varepsilon_{jt}^3 \varepsilon_{it})$ need not equal zero at θ_0 even if $E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0$. In other words, (5) may hold when the latter is included in the set of moment conditions, but not when it is replaced by the former. However, care must be taken in selecting the asymmetric co-kurtosis conditions to include because not all of their combinations yield global identification, but some combinations are satisfied by multiple permutations of the columns of B . Proposition 2 below states that the $n(n-1)/2$ asymmetric co-kurtosis conditions obtained by setting $i > j$, or $i < j$ in (8) are sufficient for global identification. However, as discussed in

the Appendix, these are not the only possibilities, but exact global identification (or over-identification) can be achieved also with other sets of at least $n(n - 1)/2$ asymmetric co-kurtosis conditions. In practice, it is important to check for global identification in each case, and we have incorporated an algorithm for this purpose in our estimation software.

To guarantee global identification, it is finally necessary to rule out multiplication by -1 of the columns of B by restricting the sign of one element in each column, which boils down to multiplying ε_t by a diagonal matrix with ± 1 (the sign depending on the element to be fixed) on the main diagonal, or, in other words, fixing the signs of the shocks. In the simulation experiments in Section 3.5 and the empirical application in Section 4, we will restrict the diagonal elements of B positive to accomplish this. While such a restriction is required for global identification, it is not restrictive for empirical analysis, as once the model has been estimated, the shocks can be freely rescaled for impulse response analysis if desired.

Proposition 2. (*Global identification*) *Suppose the set of $n + (p + 1)n^2$ moment conditions contains only the $n(n - 1)/2$ asymmetric co-kurtosis conditions of the form (8) such that $i > j$ (or $i < j$), and the parameters θ of SVAR model (1) are locally exactly identified (B up to permutation and multiplication of its columns by -1). Then, if the diagonal elements of B are restricted positive, θ is also globally identified in the sense that $E[f(v_t, \theta_0)] = 0$ for only one permutation of the components of ε_t (or, equivalently, of the columns of B), and $E[f(v_t, \theta_0)] \neq 0$ for any other permutation of the components of ε_t .*

Proof. See the Appendix. □

While there are multiple sets of $n(n - 1)/2$ asymmetric co-kurtosis conditions that, in addition to conditions (7a)–(7c), uniquely identify the SVAR model (with a given ordering of the columns of B), introducing over-identifying conditions facilitates selecting the SVAR model (i.e., the permutation of the columns of B) that the data lend the strongest support to. The selection of moment conditions can be based on the standard moment selection criteria discussed in Section 3.4, and demonstrated in Section 4. The data-driven procedure for selecting the moment conditions, and hence the particular SVAR model, can be seen an advantage of the GMM estimator over the ML estimator of Lanne et al. (2017) and the PML estimator of Gouriéroux et al. (2017), which require somewhat complicated additional restrictions to be imposed to guarantee uniqueness (that fix an arbitrary SVAR model).

The over-identifying conditions can be additional asymmetric co-kurtosis conditions of the form (8), provided global identification is preserved. As discussed in the Appendix, global identification may fail if the set of asymmetric co-kurtosis conditions is inappropriately augmented. In particular, including all $n!$ asymmetric co-kurtosis conditions renders the model globally unidentified because in that case the columns of B can be permuted without affecting the value of the objective function $Q_T(\theta)$ in (4). This follows from the fact that, in that case, the moment conditions would include both $E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0$ and $E(\varepsilon_{jt}^3 \varepsilon_{it}) = 0$ for all $i \neq j$, so that interchanging the i th and j th columns of B results in the same value of $Q_T(\theta)$.

In addition, symmetric co-kurtosis conditions of the form

$$E(\varepsilon_{it}^2 \varepsilon_{jt}^2) - 1 = 0 \quad i \neq j. \quad (9)$$

may be included. Also these conditions are redundant in the Gaussian case, where zero covariance and independence coincide, i.e. by conditions (7c) and (7b), ε_{it} and ε_{jt} are orthogonal and $E(\varepsilon_{it}^2) = E(\varepsilon_{jt}^2) = 1$, respectively, so that under Gaussianity $E(\varepsilon_{it}^2 \varepsilon_{jt}^2) = E(\varepsilon_{it}^2)E(\varepsilon_{jt}^2) = 1$. However, if at most one of the errors is Gaussian, their orthogonality no longer implies independence, and hence symmetric co-kurtosis conditions may be informative if $\theta \neq \theta_0$. It is important to notice that symmetric co-kurtosis conditions are not alone sufficient for global identification because $E(\varepsilon_{it}^2 \varepsilon_{jt}^2) = E(\varepsilon_{jt}^2 \varepsilon_{it}^2)$ for any permutation of the columns of B . In other words, the contribution of these conditions to the value of $Q_T(\theta)$ is independent of the permutation of the columns of B (i.e., the particular SVAR model).

3.3 Asymptotic Properties

To be able to apply standard asymptotic results related to the GMM estimator derived in the literature, we make a number of assumptions on which they are based and that can be easily checked or assumed in empirical applications. Our presentation in this section draws heavily upon Hall (2005, Chapters 3 and 5.3). First, to show consistency of the GMM estimator of the parameter vector θ , we make the following assumption:

Assumption 2.

- (i) *The random vectors $\{v_t : -\infty < t < \infty\}$ form a strictly stationary process.*
- (ii) *The function $f(\cdot, \cdot)$ is continuous on the parameter space Θ for each v_t , $E[f(v_t, \theta)]$ exists and is finite for every $\theta \in \Theta$, and is continuous on Θ .*

- (iii) The random vector v_t and the parameter vector θ_0 satisfy the population moment conditions $E[f(v_t, \theta_0)] = 0$.
- (iv) $E[f(v_t, \bar{\theta})] \neq 0$ for all $\bar{\theta} \in \Theta$ such that $\bar{\theta} \neq \theta_0$.
- (v) W_T is a positive semi-definite matrix which converges in probability to the positive definite matrix of constants W .
- (vi) The random process v_t is ergodic.
- (vii) Θ is a compact set.
- (viii) $E[\sup_{\theta \in \Theta} \|f(v_t, \theta)\|] < \infty$.

The requirement of stationarity in part (i) is in concert with assumption (2), and it entails including unit root processes often encountered in macroeconomic applications as differences in the SVAR model. The population moment conditions implied by non-Gaussian independent errors, such as those discussed in Section 3.1, obviously satisfy the regularity conditions in part (ii). Global identification (part (iv)) can be guaranteed by fixing the signs of the shocks and including suitable asymmetric co-kurtosis conditions, as discussed in Section 3.2. Parts (vi)–(viii) are technical assumptions that establish uniform convergence in probability of the objective function $Q_T(\theta)$. They can typically be safely assumed even if no knowledge of the bounds of the parameter space Θ is available.

Under Assumption 2, the GMM estimator $\hat{\theta}_T$ converges in probability to θ_0 (see, e.g. Hall 2005, Theorem 3.1). This consistency result holds for all two-step, iterated and continuous updating GMM estimators that are asymptotically equivalent although they may behave differently in finite samples.

In order to derive the asymptotic distribution of the GMM estimator $\hat{\theta}_T$, the following additional assumption is needed:

Assumption 3.

- (i) The derivative matrix $\partial f(v_t, \theta)/\partial \theta'$ exists and is continuous on Θ for each v_t , θ_0 is an interior point of Θ , and $E[\partial f(v_t, \theta_0)/\partial \theta']$ exists and is finite.
- (ii) $E[f(v_t, \theta_0)f(v_t, \theta_0)']$ exists and is finite, and $\lim_{T \rightarrow \infty} \text{Var} \left[T^{1/2} \left(T^{-1} \sum_{t=1}^T f(v_t, \theta_0) \right) \right] = S$ exists and is a finite valued positive definite matrix.
- (iii) $E[\partial f(v_t, \theta)/\partial \theta']$ is continuous on some neighborhood N_ϵ of θ_0 .

(iv) $\sup_{\theta \in N_\varepsilon} \|G_T(\theta) - E[\partial f(v_t, \theta)/\partial \theta']\| \xrightarrow{p} 0$, where $G_T(\theta) = T^{-1} \sum_{t=1}^T \partial f(v_t, \theta)/\partial \theta'$.

The existence of the derivative matrix of the moment conditions in part (i) is obviously satisfied by conditions such as those discussed in Section 3.1 above, and the rest of the assumption can reasonably be expected to hold as well. Assumptions 2 and 3 together imply asymptotic normality of the GMM estimator, summarized in the following result (adapted from Hall, Theorem 3.2):

Theorem 1. *If Assumptions 2 and 3 hold, then $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, MSM')$, where $M = (G_0'WG_0)^{-1}G_0'W$ with $G_0 = E[\partial f(v_t, \theta_0)/\partial \theta']$.*

For the efficient GMM estimator with $W = S^{-1}$, the asymptotic covariance matrix of $\hat{\theta}_T$ reduces to $(G_0'S^{-1}G_0)^{-1}$. This result facilitates testing hypotheses on the parameters θ once G_0 and S are replaced by their consistent estimators, $G_T(\hat{\theta}_T)$ and \hat{S}_{HAC} , respectively. As discussed in the Introduction, while we are able to statistically identify the SVAR model, its economic interpretation may call for the imposition of additional identifying restrictions, such as those imposed in similar models in the previous literature. However, to be useful such restrictions should be supported in the data, which we can actually test. In addition to labeling equations or economic shocks, economic theory may imply hypotheses restricting the parameters.

Newey and West (1987) show how hypotheses of the form

$$H_0 : r(\theta_0) = 0 \quad \text{vs.} \quad H_A : r(\theta_0) \neq 0$$

can be tested in the GMM framework. Here $r(\cdot)$ is an $(s \times 1)$ vector of real-valued, continuous and differentiable functions, and the $(s \times k)$ matrix $R(\theta) = \partial r(\theta)/\partial \theta'$ has rank s , so that there are at most as many non-redundant restrictions as there are parameters in θ . The tests considered by Newey and West are extensions of asymptotic tests related to the method of maximum likelihood estimation. Let $\hat{\theta}_T$ and $\tilde{\theta}_T$ denote the unrestricted and restricted (by $r(\theta) = 0$) efficient GMM estimators, respectively. Then the Wald test statistic can be written as

$$Tr(\hat{\theta}_T)' \left[R(\hat{\theta}_T)[G(\hat{\theta}_T)' \hat{S}_T(\hat{\theta})^{-1} G_T(\hat{\theta}_T)]^{-1} R(\hat{\theta}_T)' \right]^{-1} r(\hat{\theta}_T). \quad (10)$$

While (10) depends only on the unrestricted estimate, the likelihood ratio (LR) type test statistic

$$T[Q_T(\tilde{\theta}_T) - Q_T(\hat{\theta}_T)] \quad (11)$$

is based on the change in the minimum of the objective function between the restricted and unrestricted models. Under Assumptions 2 and 3, both (10) and (11) follow asymptotically the χ^2 distribution with s degrees of freedom when H_0 is true. Compared to the LR type test, the Wald test has the advantage that only the unrestricted model needs to be estimated, but it is not invariant to reparametrization of the model or the restrictions. As shown by Hall and Inoue (2003), these tests have also power against misspecification, indicating that they may reject because the moment conditions are violated even if the restriction $r(\theta_0) = 0$ holds. Therefore, it is important to test for misspecification (see Section 3.4) before conducting inference on the parameters.

A distinguishing feature of our approach is that tests on the parameters of the matrix B can be given a general interpretation. This follows from the fact that each set of moment conditions containing a sufficient number of admissible asymmetric co-kurtosis conditions yields a different SVAR model, and the data-driven moment selection procedure, outlined in Section 3.4, yields a unique set of moment conditions. Hence, although such tests concern only one of the SVAR models, it is the optimal model based on the data. In contrast, the ML and PML approaches of Lanne et al. (2017) and Gouriéroux et al. (2017), respectively, call for additional assumptions to pinpoint a particular SVAR model that need not be strongly supported by the data, and therefore, tests in those setups can be only interpreted as tests on the parameters of the particular model (which is only one model in a set of observationally equivalent models). For instance, the test of the null hypothesis that B is lower triangular, can be interpreted as a tests of the existence of a recursive SVAR model. In the ML and PML approaches, such a test would only concern the particular model picked by the pre-specified restrictions.

3.4 Over-identifying Restrictions Test and Moment Selection

As discussed in Section 3.2, the SVAR model (1) (or (3)) is globally exactly identified if estimation is based on conditions (7a)–(7c) and $n(n-1)/2$ appropriately selected asymmetric co-kurtosis conditions of the form (8). Over-identification can be achieved by introducing additional co-kurtosis conditions. Once the model has been estimated, it is important to ensure that the moment conditions agree with the data. To that end, Hansen’s (1982) well-known J test for over-identifying restrictions is available whenever there are more moment conditions than parameters to estimate ($q > k$). When the model is exactly identified, i.e.,

$q = k$, the moment conditions are automatically satisfied, while in the over-identified case, the additional moment conditions are informative about the correctness of the specification. The test statistic, $J_T = TQ_T(\hat{\theta}_T)$, is convenient in that it is obtained as a by-product of estimation, and asymptotically it follows the χ^2 distribution with $q - k$ degrees of freedom under the null hypothesis of correct specification.

Typically, several alternative sets of moment conditions agree with the data, and a number of methods of selecting the optimal moment conditions among them have been put forth in the literature. In this paper, we employ Andrews's (1999) information criterion based approach backed up by previous simulation evidence, and the relevant moment selection criterion proposed by Hall, Inoue, Jana and Shin (2007), which concentrate on different aspects of the moment conditions. The former attempts to find the largest set that is supported by the data, while the latter tries to find the most relevant moment conditions, yielding maximal estimation efficiency and avoiding redundancy. Finding a relevant set of moment conditions is important because introducing too many conditions might adversely affect the finite-sample properties of the GMM estimator (see, e.g., Hall and Peixe (2003) in the context of linear regression). In practice, we recommend a moment selection strategy based on a combination of these criteria along the lines of Hall's (2005, Section 7.3.3) suggestion.

The SVAR framework is special in the sense that, as discussed in Section 3.2, each admissible combination of asymmetric co-kurtosis conditions corresponds to a different SVAR model (involving a different permutation of the columns of the matrix B). Hence, moment selection also entails selecting a particular SVAR model. While Lanne et al. (2017) and Gouriéroux et al. (2017) impose pre-specified additional restrictions to pinpoint a particular SVAR model, our approach is purely data-driven. In other words, by selecting the optimal set of moment conditions by means of moment selection criteria, we also select a unique SVAR model, which emphasizes the importance of moment selection in the analysis of SVAR models.

Andrews's (1999) moment selection criterion

$$MSC(c) = J_T(c) - (q - k)\ln(T) \quad (12)$$

is computed for several sets of moment conditions, indexed by c , and the set minimizing its value is selected. The first term is just the value of the J statistic of overidentifying restrictions, whose small values lend support to the moment conditions, while the latter

term increases with the degrees of freedom $(q - k)$. Hence, this criterion tends to favor a large set of valid moment conditions, without paying attention to efficiency or redundancy. The relevant moment selection criterion

$$RMSC(c) = \ln[|\hat{V}_{\theta,T}(c)|] + (q - k)\ln[(T/b_T)^{1/2}](T/b_T)^{-1/2} \quad (13)$$

is, in turn, concerned with the efficiency and non-redundancy of the moment conditions. The first term indicates estimation accuracy, and it is the smaller the more accurately the parameters have been estimated, i.e., the smaller is the determinant of their estimated covariance matrix $\hat{V}_{\theta,T}(c)$. The penalty term involves the bandwidth parameter b_T of the \hat{S}_{HAC} estimator to account for its rate of convergence. Also this criterion is computed for several sets of moment conditions, and the set yielding the minimum value is selected. The first term is obviously non-decreasing in the number of moment conditions, whereas the second term penalizes for additional conditions, attempting to avoid redundant conditions.

For practical moment selection, we recommend a version of the combined strategy of Hall (2005, Section 7.3.3), where the MSC and RMSC are employed in succession. In all cases, the moment conditions (7a)–(7c) are included, and the procedure is used to augment them with the optimal combination of (at least $n(n - 1)/2$) asymmetric, and symmetric co-kurtosis conditions.

As the first step, we estimate the model with all combinations of the maximal number of asymmetric and $0, 1, \dots, n$ symmetric co-kurtosis conditions such that identification is preserved, and select among them the combination of conditions that minimizes the MSC. For instance, in our empirical illustration with three variables in Section 4, estimation can be based on at most five asymmetric co-kurtosis conditions. The set of moment conditions selected by the MSC should include the maximal number of moment conditions supported by the data. Then, we estimate the model with all $q > k$ combinations of the moment conditions included in this set, such that the model remains over-identified, and select the set of moment conditions that minimizes the RMSC. We should, thus, end up with the most informative set of moment conditions among those that the data lend strongest support to.

At both steps, it is important to ensure that each set of moment conditions satisfies global identification, and discard those that do not. This feature can be easily built into the estimation procedure. The combined strategy outlined above, and illustrated in Section 4, should efficiently make use of all information in the data such that the set of moment conditions selected is the largest possible set with as few redundancies as possible.

In high-dimensional SVAR models, the moment selection procedure outlined above may become computationally burdensome because the number of subsets of admissible asymmetric co-kurtosis conditions increases rapidly with the number of variables included. For instance, a five-dimensional SVAR model involves 600 such admissible subsets only in the first step of the procedure. To keep the moment selection problem tractable, it may, therefore, be necessary to devise some kind of a sequential procedure based on only the RMSC, starting out with the largest admissible set of asymmetric co-kurtosis conditions (in addition to conditions (7a)–(7c)) not rejected by the J test, and then drop co-kurtosis conditions, one at a time, until the RMSC cannot be made smaller. Alternatively, in the combined procedure, it may be required that the sets of moment conditions to be compared differ by at least $r > 1$ asymmetric co-kurtosis conditions.

3.5 Finite-sample Properties

In order to gauge the properties of the GMM estimator in small samples, we conducted a number of Monte Carlo simulation experiments. For a comparison to the results of Gouriéroux et al. (2017) related to the PML estimator, we first considered the same bivariate SVAR(0) model (with no lagged terms):

$$y_t = B\varepsilon_t,$$

where B is an orthogonal matrix dependent on a single parameter, i.e.,

$$B = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

with $\theta = -\pi/5$. Because all elements of B depend only on θ , it suffices to concentrate on the estimates of just one element, say $B_{11} = \cos(-\pi/5) \approx 0.809$. Each of the independent components of ε_t is assumed to follow Student's t distribution with 5 degrees of freedom, and standardized to have variance unity. We base estimation on two alternative sets of moment conditions. First, the following five moment conditions: $E(\varepsilon_{1t}^2) = E(\varepsilon_{2t}^2) = 1$, $E(\varepsilon_{1t}\varepsilon_{2t}) = 0$, $E(\varepsilon_{1t}^3\varepsilon_{2t}) = 0$, and $E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = 1$. As discussed in Section 3.1, one asymmetric co-kurtosis condition is necessarily required for identification, and with the symmetric co-kurtosis condition, over-identification is reached. Second, in order to examine the effect of including a redundant moment condition, we augment this set by the conditions that $E(\varepsilon_{1t}\varepsilon_{1,t-1}) = E(\varepsilon_{2t}\varepsilon_{2,t-1}) = 0$, which provides no new information when the model is

correctly specified in that it captures all autocorrelation in the data, resulting in serially uncorrelated errors.

Table 1: Simulation results of the SVAR(0) model.

Estimator	T	No redundant conditions			Two redundant conditions		
		Bias	St.dev.	J test	Bias	St.dev.	J test
Two Step	200	-0.0211	0.1415	0.0576	-0.0038	0.1485	0.1046
	500	-0.0125	0.0867	0.0510	0.0026	0.0876	0.0902
	1000	-0.0081	0.0619	0.0504	-0.0077	0.0628	0.0577
Iterated	200	-0.0492	0.1709	0.0381	-0.0144	0.1967	0.0859
	500	-0.0224	0.1020	0.0333	0.0027	0.1080	0.0707
	1000	-0.0134	0.0734	0.0364	-0.0113	0.0745	0.0416
CUE	200	-0.0521	0.1520	0.0207	0.0238	0.1760	0.0412
	500	-0.0260	0.0960	0.0234	0.0158	0.0996	0.0477
	1000	-0.0145	0.0673	0.0316	-0.0167	0.0695	0.0316

The results for the two-step, iterated and continuous updating (CUE) GMM estimators are based on $N = 10,000$ simulated samples of $T = 200, 500,$ and $1,000$ observations. The components of the error term $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, are first generated from independent t distributions with 5 degrees of freedom. Then the data y_t are computed from $y_t = B\varepsilon_t$, where the entries of B are $B_{11} = \cos(\theta)$, $B_{12} = \sin(\theta)$, $B_{21} = -\sin(\theta)$, and $B_{22} = \cos(\theta)$ with $\theta = -\pi/5$. The errors are centered and standardized to have variance unity. The left panel contains the results in the case of no redundant moment conditions, while in the right panel there are two redundant moment conditions. The bias and standard deviation of the GMM estimates of B_{11} are reported, and the columns entitled “ J test” report the rejection rates of the 5% nominal level J test of over-identifying restrictions.

Table 1 contains the results (based on 10,000 replications) related to all three GMM estimators discussed in Section 3.1. The GMM estimator is consistent, but not necessarily unbiased, and, following Gouriéroux et al. (2017), we report the bias and standard deviation of the estimates of B_{11} . In addition, we examine the rejection rates of the 5% nominal level J test of over-identifying restrictions. In all cases, the two-step estimator is the winner among the three GMM estimators with the smallest bias and standard deviation. It is noteworthy that iteration actually seems to be detrimental to estimation accuracy. In terms of bias and standard deviation, the performance of the two-step GMM estimator is, in general, comparable to that of the PML estimator of Gouriéroux et al. (2017), and

it outperforms their recursive PML estimator, also based on the correctly specified error distribution. In the absence of redundant moment conditions, the two-step estimator also seems to lead to the best size control in the J test, whereas the other two estimators tend to under-reject. The addition of redundant moment conditions makes all GMM estimators less accurate, with the two-step estimator being, in general, the most robust of them. Interestingly, however, the size of the J test related to the two-step estimator becomes more distorted, while the under-rejection problem of the other two estimators is alleviated. Following Gouriéroux et al., we also considered generating ε_{1t} and ε_{2t} from t distributions with 7 and 12 degrees of freedom, respectively, with little change in the results.

Table 2: Simulation results of the SVAR(1) model.

C_{11}	T	No redundant conditions			Two redundant conditions		
		Bias	St.dev.	J test	Bias	St.dev.	J test
0.0	200	0.1059	0.3225	0.2260	0.0887	0.3058	0.2368
	500	0.0546	0.2622	0.1871	0.0467	0.2467	0.1875
	1000	0.0241	0.1924	0.1332	0.0219	0.1986	0.1343
0.5	200	0.1075	0.3830	0.2316	0.0974	0.3392	0.2474
	500	0.0546	0.2624	0.1786	0.0497	0.2665	0.1909
	1000	0.0255	0.2034	0.1313	0.0239	0.1979	0.1389
0.9	200	0.0710	0.2730	0.2034	0.0742	0.2751	0.2156
	500	0.0326	0.2105	0.1636	0.0316	0.2090	0.1792
	1000	0.0154	0.1741	0.1246	0.0108	0.1619	0.1336
0.97	200	0.0499	0.2518	0.1754	0.0587	0.2950	0.1577
	500	0.0223	0.2466	0.1396	0.0196	0.2417	0.1519
	1000	0.0077	0.1502	0.1111	0.0085	0.1579	0.1334

See the notes to Table 1. The data are generated from the DGP in (14).

Next, we introduce some autocorrelation, and examine the performance of the GMM estimator in a SVAR(1) model, concentrating on the two-step GMM estimator found superior above. Specifically, we consider the following extension of the previous data-generating process:

$$y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} C_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix} y_{t-1} + B\varepsilon_t, \quad (14)$$

where the components of $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ are generated in the same way as above, and

$C_{11} \in \{0, 0.5, 0.9, 0.97\}$, with persistence increasing in the value of C_{11} . The same sets of moment conditions as above are entertained.

The results are reported in Table 2. In all cases, the bias and standard deviation of the GMM estimator are larger than in the SVAR(0) case, and they tend to diminish with increasing persistence. The J test tends to strongly over-reject in all cases. Comparison of the left and right panels reveals that, somewhat surprisingly, estimation accuracy tends to improve due to the introduction of redundant moment conditions although the differences are not great. It may be that the redundant conditions marginally help to guard against remaining autocorrelation in the errors.

4 Empirical Illustration

We demonstrate SVAR analysis based on GMM estimation by means of an empirical application to quarterly U.S. macroeconomic data covering the period from 1960:I to 2017:II (229 observations). In particular, we consider a stylized three-variable VAR model for $y_t = (\pi_t, u_t, r_t)'$, where π_t is inflation, u_t is the unemployment gap, and r_t is the federal funds rate. All data are extracted from the Federal Reserve Economic Database (FRED). Inflation is computed as the logarithmic difference, multiplied by 400, of the seasonally adjusted GDP deflator (mnemonic GDPDEF) and the unemployment gap as the difference between the observed unemployment rate (mnemonic UNRATE) and the natural rate of unemployment (mnemonic NROU).

To obtain initial estimates of the autoregressive parameters, we start out by estimating a reduced-form VAR model with an intercept term. The Akaike and Schwartz information criteria pick models with 6 and 3 lags, respectively. The latter exhibits remaining autocorrelation in the residuals of all equations, while in the former, it is clearly a problem only in the equation of the federal funds rate. Qualitatively, the fit of the model with four lags is similar to the former, so in the interest of parsimony, we proceed with the VAR(4) model. For identification, non-Gaussianity of at least two of the structural shocks is crucial, and, therefore, we check the residuals of the estimated VAR model for normality. Because the structural errors are linear combinations of the reduced-form errors, normality of any of the latter might imply normality of multiple structural errors and, hence, violation of identification. The results of the Jarque-Bera test of normality as well as estimated skewness and kurtosis of the residuals are reported in Table 3. Normality is clearly rejected at

conventional significance levels for all residual series, showing up as kurtosis in excess of value 3 implied by normality, and as skewness. Hence, the non-Gaussianity condition for identification of the parameters of the SVAR model seems to be satisfied.

Table 3: Normality diagnostics of the residuals of the VAR(4) model.

Equation	Skewness	Kurtosis	Jarque-Bera
Inflation	0.3399	3.7891	0.0115
Unemployment Gap	0.6799	4.1976	0.0010
Federal Funds Rate	0.5766	13.0857	2.2e-16

The entries are skewness and kurtosis of the residuals of the equations of the VAR(4) model, and the p values of the Jarque-Bera test for their normality.

We next estimate a three-variable SVAR(4) model using different combinations of moment conditions. In order to ensure unique identification, the diagonal elements of the B matrix of instantaneous effects are constrained positive. In all cases, conditions (7a)–(7c) are included, while the rest of the moment conditions are selected by the sequential procedure outlined in Section 3.4. In Table 4, we first report the results for the combinations of five asymmetric moment conditions (8) that minimize the MSC with with 0, 1, 2 and 3 symmetric co-kurtosis conditions. With any given number of symmetric co-kurtosis conditions, the set of asymmetric co-kurtosis conditions contains terms involving the third power of each component of the error term. This guarantees identification even if one of the components is Gaussian albeit, in view of the results in Table 3, this is unlikely to be the case. The MSC seems to improve with the inclusion of additional symmetric moment conditions, and the J test of over-identifying restrictions does not reject at conventional significance levels even when all three symmetric co-kurtosis conditions are included. The set of moment conditions selected by the MSC containing all three symmetric co-kurtosis conditions excludes only the asymmetric co-kurtosis condition $E(\varepsilon_{2t}^3 \varepsilon_{1t}) = 0$. Hence, in order to select the most informative conditions, we proceed with the RMSC criterion among all 118 subsets of these eight co-kurtosis conditions containing at least four conditions (of which at least three asymmetric) such that the model remains over-identified.

In Table 5, we report the results for the over-identifying sets of moment conditions containing combinations of 3, 4, and 5 asymmetric and 0, 1, 2, and 3 symmetric co-

Table 4: Moment selection based on the MSC.

Co-Kurtosis Conditions		J_T	MSC
Asymmetric	Symmetric		
$E(\varepsilon_{1t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$	-	0.989	-10.809
$E(\varepsilon_{1t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = 1$	0.999	-16.219
$E(\varepsilon_{1t}^3\varepsilon_{2t}) = E(\varepsilon_{1t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = E(\varepsilon_{1t}^2\varepsilon_{3t}^2) = 1$	0.799	-20.007
$E(\varepsilon_{1t}^3\varepsilon_{2t}) = E(\varepsilon_{1t}^3\varepsilon_{3t}) = E(\varepsilon_{2t}^3\varepsilon_{3t}) = E(\varepsilon_{3t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = E(\varepsilon_{1t}^2\varepsilon_{3t}^2) = E(\varepsilon_{2t}^2\varepsilon_{3t}^2) = 1$	0.659	-23.815

The entries in the last two columns are the p values of the J test of over-identifying restrictions and the values of the MSC criterion for the different sets of moment conditions, respectively. In addition to the asymmetric and symmetric co-kurtosis conditions, conditions (7a)–(7c) are included in each set.

Table 5: Moment selection based on the RMSC.

Co-Kurtosis Conditions		J_T	RMSC
Asymmetric	Symmetric		
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = 0$	$E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.083	-375.094
$E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.188	-380.565
$E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{1t}^2 \varepsilon_{3t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.445	-367.430
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	-	0.891	-370.119
$E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{2t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.160	-383.683
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.306	-385.032
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{1t}^2 \varepsilon_{3t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.570	-364.456
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	-	0.845	-370.442
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.294	-384.945
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.451	-387.352
$E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$	$E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{1t}^2 \varepsilon_{3t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$	0.667	-342.588

The entries in the last two columns are the p values of the J test of over-identifying restrictions and the values of the RMSC criterion for

different sets of moment conditions, respectively. In addition to the asymmetric and symmetric co-kurtosis conditions, conditions (7a)–(7c)

are included in each set.

kurtosis conditions selected by the RMSC. For any given number of asymmetric co-kurtosis conditions, the RMSC is minimized by additionally including the conditions $E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = 1$ and $E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$. With these two moment conditions included, the RMSC decreases with the number of asymmetric co-kurtosis conditions, and the overall minimum is reached with all five asymmetric co-kurtosis conditions selected by the MSC. We thus proceed with the set of moment conditions selected by the sequential procedure, including the following co-kurtosis conditions $E(\varepsilon_{1t}^3 \varepsilon_{2t}) = E(\varepsilon_{1t}^3 \varepsilon_{3t}) = E(\varepsilon_{2t}^3 \varepsilon_{3t}) = E(\varepsilon_{3t}^3 \varepsilon_{1t}) = E(\varepsilon_{3t}^3 \varepsilon_{2t}) = 0$ and $E(\varepsilon_{1t}^2 \varepsilon_{2t}^2) = E(\varepsilon_{2t}^2 \varepsilon_{3t}^2) = 1$.

The GMM estimate of the matrix of impact effects is

$$\hat{B} = \begin{bmatrix} 0.933 & -0.118 & 0.114 \\ (0.081) & (0.096) & (0.087) \\ 0.046 & 0.287 & -0.086 \\ (0.031) & (0.015) & (0.018) \\ 0.010 & -0.137 & 1.162 \\ (0.119) & (0.076) & (0.176) \end{bmatrix}$$

where the figures in parentheses are asymptotic standard errors. Only the diagonal elements and B_{23} are significant at the 5% level. Hence, the first and second shocks have no significant effect (at the 5% level) on impact on the federal funds rate, suggesting that the third shock is the likeliest candidate for the monetary policy shock. If, based on this, we label the third shock as the monetary policy shock, we can test the identification scheme considered in the previous literature that B is a lower triangular matrix, i.e., the null hypothesis $B_{12} = B_{13} = B_{23} = 0$ (see, e.g., Castelnuovo (2016) and the references therein). This hypothesis is clearly rejected with a p value 6.75e-12. The weaker restriction that the third shock has no effect on impact on inflation and unemployment gap, i.e., the hypothesis $B_{13} = B_{23} = 0$ is also rejected with a p value 5.27e-11. There is thus little support for the recursive identification scheme popular in the previous literature.

The impulse responses of the three shocks along with their 95% confidence bands are depicted in Figure 1. Only the third shock was found to have a statistically significant effect on the federal funds rate on impact at the 5% level in the asymptotic Wald test, and this conclusion is reconfirmed by the bootstrapped confidence bands on the bottom row of Figure 1. Thus, only the third shock can indeed be labeled as the monetary policy shock. The effect of a contractionary monetary policy shock on inflation is seen to be initially positive, and turning negative only after a relatively long time. This may reflect the so called price puzzle effect due to the model being very simple. Its effect on the unemployment gap is negative on impact, but turns positive after a few quarters, and eventually converges to

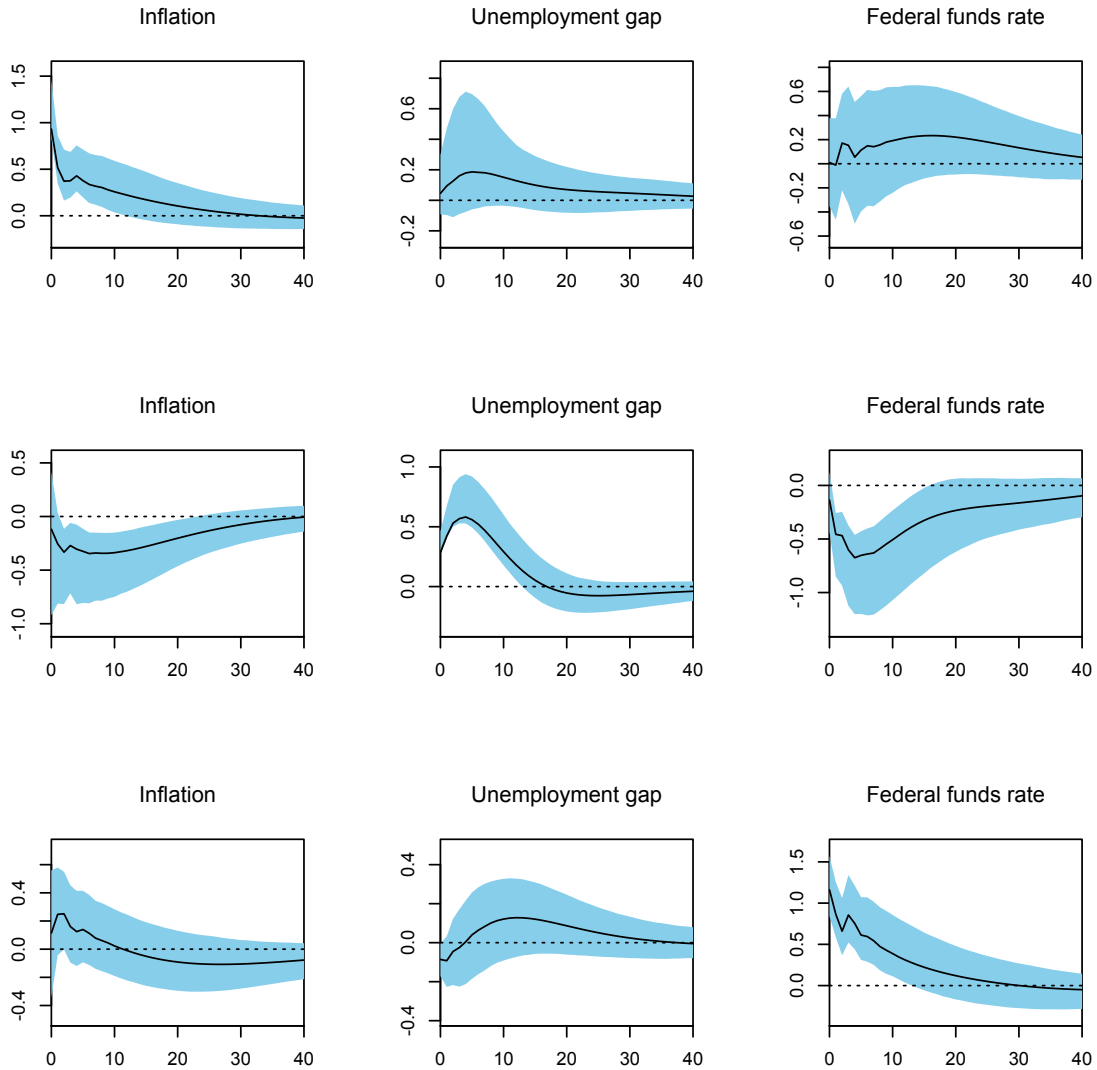


Figure 1: Impulse responses of the three shocks in the SVAR model estimated using the full set of moment conditions. Each row contains the impulse responses of one shock on all variables. The shaded areas are the pointwise 95% Hall's percentile confidence bands obtained by bootstrap with 10,000 replications.

zero. Hence, the effect of the contractionary monetary policy shock is negative on inflation and output in the medium term, albeit these effects are not significant at the 5% level.

The first shock (top row of Figure 1) has a positive effect on both inflation and output gap, and can thus be labelled a positive supply (or cost-push) shock. The second shock (the middle row), having a negative effect on inflation and a positive effect on the unemployment gap, in turn, can be labelled as a contractionary demand shock. Visual inspection of

the confidence bands suggests that asymptotic theory does not provide quite accurate approximation, as there seem to be more insignificant (at the 5% level) impact effects than the asymptotic significance tests imply. In particular, based on the bootstrapped confidence bands, the evidence is weaker against the lower-triangularity restriction on B that was strongly rejected by the Wald test. It must, of course, be kept in mind that the Wald test involving joint restrictions and confidence intervals on single parameters are not directly comparable.

5 Conclusion

In this paper, we have considered GMM estimation of structural SVAR models whose errors are non-Gaussian. In particular, we have shown that by suitable selection of moment conditions, non-Gaussianity can be exploited to uniquely identify the parameters of the SVAR model. Our approach deviates from the related statistical identification literature in that no particular distributional assumptions are required, and the structural shocks do not have to be independent. Although pseudo maximum likelihood estimators, to some extent robust with respect to distributional misspecification, have recently been proposed by Gouriéroux et al. (2017), our approach appears simpler and more general, and according to our simulation study its performance seems comparable to that of the PML estimators. Also the independence assumption may be problematic because there need not be a linear transformation that makes the errors of the reduced-form VAR model independent.

Our approach is completely data driven in that we use well-known moment selection criteria to select the optimal set of moment conditions under only the constraint that they provide local and global identification. It is the moment conditions that determine the SVAR model, and our procedure can thus be seen as selecting the model that the data lend the strongest support to. This is in contrast to Lanne et al. (2017) and Gouriéroux et al. (2017), where rather technical additional restrictions are required for unique identification. It follows that tests on the parameters of the SVAR model can be given a general interpretation: they are not only tests on the parameters of one of the observationally equivalent SVAR models, but on those of the unique SVAR model (that is the most in accordance with the data).

Statistical identification rarely produces structural shocks with economic interpretation, and our approach is no exception. In our empirical application to a trivariate U.S. macroe-

conomic model, in order to label the shocks, we made use of the shapes of the impulse responses on which there is a relatively wide agreement in the literature. Labeling could also be based on short-run or long-run restrictions that, due to unique identification, are testable in our setup. As a matter of fact, we tested recursive identification restrictions entertained in some of the related empirical literature, but they were strongly rejected, and could thus not be used to interpret the shocks. Yet another possibility would be to combine the moment conditions arising from non-Gaussianity with those related to external instruments. The instruments typically used in the literature are known to be potentially weak, and the larger set of moment conditions might enhance inference. We leave the latter issue to future research.

Appendix: Local and Global Identification

This appendix contains the proofs of Propositions 1 and 2, and further discussion on local and global identification.

Proof of Proposition 1. The necessary condition for θ to be locally identified is that the expectation of the Jacobian matrix $E[\partial f(v_t, \theta_0)/\partial \theta']$, evaluated at θ_0 , the true value of θ , has full column rank k . Because the row rank equals the column rank, it suffices to show that k rows of the Jacobian matrix are linearly independent. The Jacobian matrix corresponding to conditions (7a)–(7c) and $n(n-1)/2$ conditions of the form (8) is obtained by stacking (A.1), (A.2), the $n(n-1)/2$ components of the form (A.3), and $n(n-1)/2$ components of the form (A.4) below:

$$E \left[\frac{\partial(\varepsilon_t \otimes \mathbf{x}_{t-1})}{\partial \theta'} \right] = - \left[E \left[(I_n \otimes \mathbf{x}_{t-1})(\mathbf{x}'_{t-1} \otimes A) \right], 0_{n(np+1) \times n^2} \right] \quad (\text{A.1})$$

$$E \left[\frac{\partial(\varepsilon_{it}^2)}{\partial \theta'} \right] = \left[0_{1 \times n(np+1)}, -2(e'_i \otimes a_i) \right], \quad (\text{A.2})$$

$$E \left[\frac{\partial(\varepsilon_{it}\varepsilon_{jt})}{\partial \theta'} \right] = \left[0_{1 \times n(np+1)}, -(e'_j \otimes a_i) - (e'_i \otimes a_j) \right], i \neq j, \quad (\text{A.3})$$

$$E \left[\frac{\partial(\varepsilon_{it}^3\varepsilon_{jt})}{\partial \theta'} \right] = - \left[E(\varepsilon_{it}^3)(E(\mathbf{x}'_{t-1}) \otimes a_j), 3(e'_j \otimes a_i) + E(\varepsilon_{it}^4)(e'_i \otimes a_j) \right], i \neq j. \quad (\text{A.4})$$

Here $i, j \in \{1, \dots, n\}$, e_i is the i th column of the $n \times n$ identity matrix, and a_i is the i th row of $A \equiv B^{-1}$.

The submatrix consisting of the first $n(np + 1)$ rows in (A.1) and the $n + n(n - 1)/2$ rows in (A.2)–(A.3) below them is block diagonal, and thus the first $n(np + 1)$ rows are linearly independent of the following $n + n(n - 1)/2$ rows. Moreover, (A.1) is of full row rank, which can be easily seen by writing it as

$$E [P((I_n \otimes \mathbf{x}_{t-1})(\mathbf{x}'_{t-1} \otimes A))] = E(\mathbf{x}_{t-1}\mathbf{x}'_{t-1}) \otimes A \quad (\text{A.5})$$

for a particular permutation matrix P . Because permuting the rows of a matrix does not change its rank, we can focus on $E(\mathbf{x}_{t-1}\mathbf{x}'_{t-1}) \otimes A$. Now, as a positive definite matrix, the $(np + 1) \times (np + 1)$ square matrix $E(\mathbf{x}_{t-1}\mathbf{x}'_{t-1})$ is of full rank, and the $n \times n$ matrix B (and hence $A = B^{-1}$) is assumed to be of full rank. Because $\text{rank}(E(\mathbf{x}_{t-1}\mathbf{x}'_{t-1}) \otimes A) = \text{rank}(E(\mathbf{x}_{t-1}\mathbf{x}'_{t-1})) \times \text{rank}(A) = (np + 1)n$, the first $n(np + 1)$ rows of $E[\partial f(v_t, \theta_0)/\partial \theta']$ must be linearly independent.

It is clear that the $n + n(n - 1)/2$ rows given by (A.2) and (A.3) must be mutually linearly independent because the rows of A are linearly independent. All sets of moment conditions that we consider, contain moment conditions (7a)–(7c), and thus all the Jacobian matrices considered contain the $n(np + 1) + n + n(n - 1)/2$ rows that were just shown to be linearly independent.

The remaining rows of $E[\partial f(v_t, \theta_0)/\partial \theta']$, obtained from (A.4) can be readily seen to be mutually linearly independent, and independent of the first $n(np + 1)$ rows of the Jacobian matrix. However, they can be linearly independent of the rows given by (A.2) and (A.3) if at most one of the components of ε_t is Gaussian and suitable asymmetric co-kurtosis moment conditions are selected. Suppose first that all n components of ε_t are non-Gaussian. In this case generally $E(\varepsilon_{it}^3) \neq 0$ and $E(\varepsilon_{it}^4) \neq 3$ for all i , and it is not possible to express any of the rows of the form (A.4) as a linear combination of rows given by (A.3), and thus the Jacobian matrix is of full row (and column) rank. If the i th component of ε_t is Gaussian, $E(\varepsilon_{it}^4) = 3$ and $E(\varepsilon_{it}^3) = 0$, and one of the rows given by (A.4) equals 4 times one of the rows given by (A.3), and the Jacobian matrix is of reduced rank. However, by inspecting (A.4), it is easy to see that if the asymmetric co-kurtosis conditions selected do not involve the third power of the Gaussian element of ε_t , the rows given by (A.3) and (A.4) are linearly independent, and the Jacobian matrix is of full rank.

Finally, multiplication of the columns of B by -1 is equivalent to multiplication of

the rows of A by -1 , which is readily seen to have no effect on the rank of the Jacobian matrix. \square

Let us illustrate Proposition 1 by an example of a trivariate SVAR model ($n = 3$), estimated by imposing $q = k$ moment conditions such that $i < j$ in (8). We thus have three asymmetric co-curtosis conditions: $E(\varepsilon_{1t}^3 \varepsilon_{2t}) = 0$, $E(\varepsilon_{1t}^3 \varepsilon_{3t}) = 0$, and $E(\varepsilon_{2t}^3 \varepsilon_{3t}) = 0$. The expectation of the Jacobian matrix becomes

$$\begin{pmatrix} -E[(I_3 \otimes \mathbf{x}_{t-1})(\mathbf{x}'_{t-1} \otimes A)] & 0_{3k \times 3} & 0_{3k \times 3} & 0_{3k \times 3} \\ \hline 0_{1 \times 3(3p+1)} & -2a_1 & 0_{1 \times 3} & 0_{1 \times 3} \\ 0_{1 \times 3(3p+1)} & 0_{1 \times 3} & -2a_2 & 0_{1 \times 3} \\ 0_{1 \times 3(3p+1)} & 0_{1 \times 3} & 0_{1 \times 3} & -2a_3 \\ \hline 0_{1 \times 3(3p+1)} & -a_2 & -a_1 & 0_{1 \times 3} \\ 0_{1 \times 3(3p+1)} & -a_3 & 0_{1 \times 3} & -a_1 \\ 0_{1 \times 3(3p+1)} & 0_{1 \times 3} & -a_3 & -a_2 \\ \hline -E(\varepsilon_{1t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_2) & -E(\varepsilon_{1t}^4)a_2 & -3a_1 & 0_{1 \times 3} \\ -E(\varepsilon_{1t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_3) & -E(\varepsilon_{1t}^4)a_3 & 0_{1 \times 3} & -3a_1 \\ -E(\varepsilon_{2t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_3) & 0_{1 \times 3} & -E(\varepsilon_{2t}^4)a_3 & -3a_2 \end{pmatrix}.$$

It is seen that the last $n(n-1)/2 = 3$ rows of the Jacobian matrix can be linearly independent of the rows of its middle row block, only if ε_{1t} and ε_{2t} are non-Gaussian because then $E(\varepsilon_{1t}^4)$ and $E(\varepsilon_{2t}^4)$ are generally different from each other and from 3. Also, the quantities $E(\varepsilon_{1t}^3)$ and $E(\varepsilon_{2t}^3)$ are generally different from zero, if the distributions of these errors are asymmetric. However, even if their distributions are symmetric, their being leptokurtic suffices to guarantee that the Jacobian matrix is of full rank. For Gaussian ε_{1t} , the quantities $E(\varepsilon_{1t}^4)$ and $E(\varepsilon_{1t}^3)$ equal 3 and 0, respectively, and the second row in bottom block is 3 times the second row in the third block, so that they are linearly dependent. Likewise, if ε_{2t} is Gaussian, the third rows in the middle and bottom blocks are linearly dependent. The third component ε_{3t} can be Gaussian because it does not enter the moment conditions in its third power, and thus its moments do not appear in the Jacobian matrix.

When the number of orthogonality conditions exceeds the number of parameters ($q > k$), all shocks can enter the orthogonality conditions in their third power, even if any one of them is Gaussian. Let us demonstrate this by expanding the set of the asymmetric moment conditions in our previous example by $E(\varepsilon_{3t}^3 \varepsilon_{1t}) = 0$. The expanded set hence contains the following asymmetric co-kurtosis conditions: $E(\varepsilon_{1t}^3 \varepsilon_{2t}) = 0$, $E(\varepsilon_{1t}^3 \varepsilon_{3t}) = 0$, $E(\varepsilon_{2t}^3 \varepsilon_{3t}) = 0$,

$E(\varepsilon_{3t}^3 \varepsilon_{1t}) = 0$. The corresponding Jacobian matrix has $k + 1$ rows, and its bottom row is obtained from (A.4) by setting $i = 3$ and $j = 1$. Its expectation is

$$\begin{pmatrix} -E[(I_3 \otimes \mathbf{x}_{t-1})(\mathbf{x}'_{t-1} \otimes A)] & 0_{3k \times 3} & 0_{3k \times 3} & 0_{3k \times 3} \\ \hline 0_{1 \times 3(3p+1)} & -2a_1 & 0_{1 \times 3} & 0_{1 \times 3} \\ 0_{1 \times 3(3p+1)} & 0_{1 \times 3} & -2a_2 & 0_{1 \times 3} \\ 0_{1 \times 3(3p+1)} & 0_{1 \times 3} & 0_{1 \times 3} & -2a_3 \\ \hline 0_{1 \times 3(3p+1)} & -a_2 & -a_1 & 0_{1 \times 3} \\ 0_{1 \times 3(3p+1)} & -a_3 & 0_{1 \times 3} & -a_1 \\ 0_{1 \times 3(3p+1)} & 0_{1 \times 3} & -a_3 & -a_2 \\ \hline -E(\varepsilon_{1t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_2) & -E(\varepsilon_{1t}^4)a_2 & -3a_1 & 0_{1 \times 3} \\ -E(\varepsilon_{1t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_3) & -E(\varepsilon_{1t}^4)a_3 & 0_{1 \times 3} & -3a_1 \\ -E(\varepsilon_{2t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_3) & 0_{1 \times 3} & -E(\varepsilon_{2t}^4)a_3 & -3a_2 \\ -E(\varepsilon_{3t}^3)(E(\mathbf{x}'_{t-1}) \otimes a_1) & -3a_3 & 0_{1 \times 3} & -E(\varepsilon_{3t}^4)a_1 \end{pmatrix}.$$

Now, based on the previous discussions, it is obvious that this matrix is of full column rank if at most one of the structural errors is Gaussian. Indeed, if the i th shock ε_{it} is Gaussian, then $E(\varepsilon_{it}^4) = 3$, and $E(\varepsilon_{it}^3) = 0$, and the matrix has two linearly dependent rows. Because it has in total $k + 1$ rows, its rank is k , i.e., it is of full column rank. Inspection of (A.1)–(A.4) reveals that this is also the case for general $n > 3$. Thus, the necessary condition for θ to be locally identified is satisfied for the considered set of orthogonality conditions, when at most one of the structural shocks is Gaussian. The degree of over-identification may, of course, exceed one, but then care must be taken to avoid including a combination of asymmetric co-kurtosis conditions that renders the model globally unidentified (see Proposition 2).

Proof of Proposition 2. Consider SVAR model (1). All $n!$ observationally equivalent SVAR models, each corresponding to a different permutation of the components of ε_t can be written as

$$y_t = \nu^* + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + B^* \varepsilon_t^*, \quad (\text{A.6})$$

where $\nu^* = \nu P$, $A_j^* = A_j P$, $j = 1, \dots, p$, $B^* = BP$, $\varepsilon_t^* = P' \varepsilon_t$, and P denotes any one of the $n!$ permutation matrices (with model (1) obtained by setting P equal to an identity matrix). Because moment conditions other than asymmetric co-kurtosis conditions are invariant to permutations of the components of ε_t , it suffices to concentrate on the $n(n-1)/2$ asymmetric

co-kurtosis conditions of the form (8). With $i > j$, they are given by lower-triangular elements of

$$E[\varepsilon_t^* \varepsilon_t^{*'} dg(\varepsilon_t^* \varepsilon_t^{*'})] = 0, \quad (\text{A.7})$$

where $dg(\varepsilon_t^* \varepsilon_t^{*'}) = \text{diag}(\varepsilon_{1t}^{*2}, \dots, \varepsilon_{nt}^{*2})$, and the elements of the $n \times n$ matrix $\varepsilon_t^* \varepsilon_t^{*'} dg(\varepsilon_t^* \varepsilon_t^{*'})$ are of the form $E(\varepsilon_{it}^3 \varepsilon_{jt})$ for $i, j \in \{1, \dots, n\}$. Substituting ε_t^* for $P' \varepsilon_t$, (A.7) can equivalently be written as

$$E[P' \varepsilon_t \varepsilon_t' P P' dg(\varepsilon_t \varepsilon_t') P] = 0,$$

and because P is an orthogonal matrix, $PP' = I$ we obtain

$$E[P' \varepsilon_t \varepsilon_t' dg(\varepsilon_t \varepsilon_t') P] = 0. \quad (\text{A.8})$$

Now, recall that even if $E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0$ ($i \neq j$), $E(\varepsilon_{jt}^3 \varepsilon_{it})$ need not equal zero. Hence, it suffices to show the $n!$ sets of the asymmetric moment conditions obtained from the lower triangle of (A.8) with different permutation matrices P are distinct (i.e., they do not contain precisely the same elements). Notice that pre- and postmultiplying $\varepsilon_t \varepsilon_t' dg(\varepsilon_t \varepsilon_t')$ by P' and P , respectively, interchanges at least two rows and columns of $\varepsilon_t \varepsilon_t' dg(\varepsilon_t \varepsilon_t')$. Hence, the matrices $P' \varepsilon_t \varepsilon_t' dg(\varepsilon_t \varepsilon_t') P$ with different P must be different in the sense that at least two of their rows and columns are in different positions (although these matrices have exactly the same elements). This implies that the $n!$ sets of the asymmetric moment conditions given by the lower- triangular elements of (A.8) never share the same conditions. The same logic applies to the upper triangle of (A.8) that contains the asymmetric moment conditions with $i < j$. \square

Proposition 2 gives a sufficient but not necessary condition for global identification. Hence, there are other sets of asymmetric co-kurtosis conditions of the form (8) besides those obtained by setting $i < j$ or $i > j$ that exactly globally identify the parameters of the SVAR model (1). Moreover, the set of such conditions can be augmented to include more than $n(n-1)/2$ conditions to reach over-identification. However, care must be taken in selecting the conditions to preserve global identification.

Let us take a closer look at a couple of special cases. Consider first a trivariate SVAR model, corresponding to our empirical illustration. In that case, the matrix $\varepsilon_t \varepsilon_t' dg(\varepsilon_t \varepsilon_t')$ in

the proof of Proposition 2 is

$$\begin{pmatrix} \varepsilon_{1t}^4 & \varepsilon_{2t}^3\varepsilon_{1t} & \varepsilon_{3t}^3\varepsilon_{1t} \\ \varepsilon_{1t}^3\varepsilon_{2t} & \varepsilon_{2t}^4 & \varepsilon_{3t}^3\varepsilon_{2t} \\ \varepsilon_{1t}^3\varepsilon_{3t} & \varepsilon_{2t}^3\varepsilon_{3t} & \varepsilon_{3t}^4 \end{pmatrix},$$

and clearly the $n(n-1)/2 = 3$ asymmetric co-kurtosis conditions corresponding to the lower- or upper-triangular elements yield global identification because any permutation of the elements of ε_t (i.e., permutation of rows and columns of this matrix), would change at least one of the lower- or upper-triangular elements. As a matter of fact, any combination of three, four or five conditions corresponding to the off-diagonal elements of this matrix would yield global identification because these sets are all different for different permutations of the components of ε_t . Any three conditions involve all three components of ε_t , and therefore, it is easy to see that any of their permutations always yields a different set of conditions. For instance, if the three conditions are $E(\varepsilon_{it}^3\varepsilon_{jt}) = 0$, $E(\varepsilon_{jt}^3\varepsilon_{it}) = 0$, and $E(\varepsilon_{kt}^3\varepsilon_{it}) = 0$, by permuting ε_{it} and ε_{jt} , we obtain a set where the first two conditions are the same, but the third is different. In contrast, interchanging ε_{it} and ε_{kt} yields a completely different set of asymmetric co-kurtosis conditions. In the same vein, all sets with four or five conditions must be different across the different permutations of the components of ε_t . It is only the combination of all six asymmetric co-kurtosis conditions that breaks global identification because then the conditions are the same irrespective of the permutation of the components of ε_t .

In SVAR models with more than three variables, situations may arise where a set of at least $n(n-1)/2$ but less than n asymmetric co-kurtosis conditions fails to globally identify the parameters. For instance, in a five-dimensional SVAR model, the following $n(n-1)/2 = 10$ asymmetric co-kurtosis conditions do not yield global identification because by interchanging ε_{1t} and ε_{2t} we obtain the same set of conditions: $E(\varepsilon_{2t}^3\varepsilon_{1t}) = 0$, $E(\varepsilon_{3t}^3\varepsilon_{1t}) = 0$, $E(\varepsilon_{4t}^3\varepsilon_{1t}) = 0$, $E(\varepsilon_{5t}^3\varepsilon_{1t}) = 0$, $E(\varepsilon_{1t}^3\varepsilon_{2t}) = 0$, $E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$, $E(\varepsilon_{4t}^3\varepsilon_{2t}) = 0$, $E(\varepsilon_{5t}^3\varepsilon_{2t}) = 0$, $E(\varepsilon_{4t}^3\varepsilon_{3t}) = 0$, and $E(\varepsilon_{5t}^3\varepsilon_{3t}) = 0$. It is in practice straightforward to check whether a given set of asymmetric co-kurtosis conditions yields global identification, and we have incorporated this feature in our estimation software.

References

- Andrews, D. W. K. (1999), “Consistent Moment Selection Procedures for Generalized Method of Moments Estimation,” *Econometrica* 67, 543–564.
- Baumeister, C., and Hamilton, J. D. (2015), “Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information,” *Econometrica*, 83, 1963–1999.
- Bekaert, G., Engstrom, E., and Ermolov, A. (2017) “Macro Risks and the Term Structure of Interest Rates,” *Finance and Economics Discussion Series* 2017-058. Washington, D.C.: Board of Governors of the Federal Reserve System.
- Bernanke, B. S., and Mihov, I. (1995) “Measuring Monetary Policy,” *NBER Working Paper* No. 5145.
- Castelnuovo, E. (2016), “Monetary Policy Shocks and Cholesky VARs: An Assessment for the Euro Area,” *Empirical Economics*, 50, 383–414.
- Chaussé, P. (2015). gmm: Generalized Method of Moments and Generalized Empirical Likelihood. R package version 1.5.2.
- Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017), “Statistical Inference for Independent Component Analysis: Application to Structural VAR Models,” *Journal of Econometrics*, 196, 111–126.
- Eichenbaum, M., Hansen, L. P., and Singleton, K. J. (1988), “A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty,” *Quarterly Journal of Economics*, 103, 51–78.
- Hall, A. R. (2005), *Generalized Method of Moments*. New York: Oxford University Press.
- Hall, A. R., and Inoue, A. (2003), “The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models,” *Journal of Econometrics*, 114, 361–394.
- Hall, A. R., Inoue, A., Jana, K., and Sin, C. (2007), “Information in Generalized Method of Moments Estimation and Entropy-Based Moment Selection,” *Journal of Econometrics*, 138, 488–512.

- Hall, A.R., and Peixe, F. P. M. (2003), “A Consistent Method for the Selection of Relevant Instruments,” *Econometric Reviews* 7: 269–288.
- Hansen, L. P. (1982), “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054
- Hansen, L.P., Heaton, J., and Yaron, A. (1996), “Finite Sample Properties of Some Alternative GMM Estimators Obtained from Financial Market Data,” *Journal of Business and Economic Statistics*, 14, 262–280.
- Herwartz, H. (2015), “Structural VAR Modelling with Independent Innovations—An Analysis of Macroeconomic Dynamics in the Euro Area Based on a Novel Identification Scheme,” working paper, University of Göttingen.
- Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. (2010), “Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity,” *Journal of Machine Learning Research*, 11, 1709–1731.
- Kendall, M.G., and Stuart, A. (1977), *The Advanced Theory of Statistics*, Vol. 1, 4th ed. London: Charles Griffin and Co.
- Kilian, L., and Lütkepohl, H. (2017), *Structural Vector Autoregressive Analysis*, Cambridge, U.K.: Cambridge University Press.
- Lanne, M., and Luoto, J. (2016), “Data-Driven Inference on Sign Restrictions in Bayesian Structural Vector Autoregression,” *CREATES Research Paper* 2016-14.
- Lanne, M., Meitz, M., and Saikkonen, P. (2017), “Identification and Estimation of Non-Gaussian Structural Vector Autoregressions,” *Journal of Econometrics*, 196, 288–304.
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- Moneta, A., Entner, D., Hoyer, P. O., and Coad, A. (2013), “Causal Inference by Independent Component Analysis: Theory and Applications,” *Oxford Bulletin of Economics and Statistics*, 75, 705–730.
- Montiel Olea, J. L., Stock, J. H., and Watson, M. W. (2015), “Inference in Structural VARs with an External Instrument,” working paper, Columbia University.

Newey, W.K., and West, K. D. (1987), “Hypothesis Testing with Efficient Method of Moments Estimation,” *International Economic Review*, 61, 631–653.

Newey, W. K., and West, K. D. (1994), “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, 61, 631–653

Neyman, J., and Pearson, E. S. (1928), “On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part II,” *Biometrika*, 20A, 263–294.