Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

GMM Estimation of Non-Gaussian Structural Vector Autoregression

Markku Lanne & Jani Luoto

To cite this article: Markku Lanne & Jani Luoto (2021) GMM Estimation of Non-Gaussian Structural Vector Autoregression, Journal of Business & Economic Statistics, 39:1, 69-81, DOI: 10.1080/07350015.2019.1629940

To link to this article: https://doi.org/10.1080/07350015.2019.1629940

GMM Estimation of Non-Gaussian Structural Vector Autoregression

MARKKU LANNE
Faculty of Social Sciences, University of Helsinki, Helsinki, Finland (*markku.lanne@helsinki.fi*)

JANI LUOTO
Faculty of Social Sciences, University of Helsinki, Helsinki, Finland (*jani.luoto@helsinki.fi*)

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, the GMM estimator is shown to achieve local identification of the structural shocks. The optimal set of moment conditions can be found by well-known moment selection criteria. Compared to recent alternatives, our approach has the advantage that the structural shocks need not be mutually independent, but only orthogonal, provided they satisfy a number of co-kurtosis conditions that prevail under independence. 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 United States data illustrates the use of the methods.

KEY WORDS: Generalized method of moments; Non-Gaussian time series; Structural vector autoregression.

1. INTRODUCTION

In the recent literature, a number of approaches to statistically identifying the structural vector autoregressive (SVAR) model have been introduced. Typically, they make use of non-Gaussianity of the errors that may show up as structural breaks in their covariance matrix, their conditional heteroscedasticity, or their following a parametric non-Gaussian distribution (for a survey of the relevant literature, see Kilian and Lütkepohl 2017, chap. 14). Because of non-Gaussianity, the parameters of the SVAR model are statistically identified, but identification rarely provides any economic interpretation. However, in the identified model, testing and contrasting alternative identification schemes 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 article, 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. It bears resemblance to the estimation procedures in the independent component analysis (ICA) based on the use of fourth moments. However, this literature is confined to independent and identically distributed data. Moreover, the procedures are typically seen as algorithms rather than estimators, and their statistical properties are not necessarily known (see Miettinen et al. 2015) for a statistical analysis of several of these procedures). The closest counterpart of our GMM estimator in the econometric literature 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. It is also related to Herwartz's (2015) estimator based on finding the rotation of orthogonalized errors maximizing the p-value of a test of independence.

Our estimator has a number of advantages compared to its close counterparts in the previous literature. First and foremost, unlike the ICA literature, Lanne, Meitz, and Saikkonen (2017), Gouriéroux, Monfort, and Renne (2017), and Herwartz (2015), we do not assume the structural errors to be mutually independent. As pointed out by Kilian and Lütkepohl (2017, chap. 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. Instead, we only assume the shocks to be mutually orthogonal with a number of additional co-kurtosis restrictions derived from the independence assumption (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]). In other words, we do not impose independence but only its implications on selected moments. Due to relaxing independence of the structural

Journal of Business & Economic Statistics
January 2021, Vol. 39, No. 1
DOI: 10.1080/07350015.2019.1629940
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jbes*.

errors, various forms of joint conditional heteroscedasticity often found in economic data are allowed for. These include stochastic volatility, exemplified in Section 2.2.

In addition to relaxing independence, there is also no need to specify explicit non-Gaussian error distributions, in contrast to the ML estimator of Lanne, Meitz, and Saikkonen (2017). The PML estimators of Gouriéroux, Monfort, and Renne (2017) are, to some extent, robust with respect to misspecification of the error distributions, but our approach is more generic in that the selection of optimal moment conditions can always be based on conventional moment selection criteria combined with a test of over-identifying restrictions. Moreover, as an additional limitation, to show consistency and asymptotic normality Gouriéroux, Monfort, and Renne (2017) need to assume that each of the structural errors follows a different asymmetric distribution (see their Proposition 3). Nevertheless, according to our simulation results, the performance of the GMM estimator seems comparable to their PML estimator and superior to their recursive PML estimator.

Instrumental-variables methods have also previously been employed in estimating time series models, but, to the best of our knowledge, non-Gaussianity of the errors of the SVAR model has not been used to facilitate identification in the GMM framework before. However, during the revision process, we learned about two closely related studies that had appeared after the first version of this article. First, Guay and Normandin (2018) derive conditions for identification through the third and fourth unconditional moments. Second, Lewis (2018) considers identification based on the autocovariance structure of the second moments of the errors implied by an arbitrary stochastic process for the shock variances without parametric assumptions. Both papers contain an empirical application, and Lewis also compares a range of estimators, including the GMM estimator, by simulations experiments, but they concentrate on identification instead of statistical inference.

In the earlier related literature, Bernanke and Mihov (1995) showed consistency and asymptotic normality of the GMM estimator of the SVAR model over-identified by short-run restrictions, while Shapiro and Watson (1988) considered estimation of a SVAR model by instrumental-variable methods under long-run identification restrictions. Pagan and Robertson (1998) showed how the SVAR model can be estimated by instrumental-variable methods under both short-run and long-run identification restrictions (see also Watson 1994). More recently, the GMM has been employed in the literature on identification of SVAR models by external instruments (see, e.g., Montiel Olea, Stock, and Watson 2015). Wright's (2017) rank-based estimator can also be implemented as a GMM estimator; while he showed that the estimator brings efficiency gains under non-Gaussian errors, in his article, identification is based on recursive ordering or external instruments. Finally, non-Gaussianity has been utilized in GMM estimation of noninvertible reduced-form moving average models by Gospodinov and Ng (2015).

The rest of the article is organized as follows. In Section 2, we introduce the SVAR model along with the central assumptions and discuss the specification of moment conditions to be used in GMM estimation. In Section 3.1, we discuss the implementation of the GMM estimator in the SVAR model, while in Section 3.2, we introduce the regularity conditions under which the GMM estimator is consistent and asymptotically normal. In Section 3.3, we describe our moment selection procedure. Section 3.4 contains some finite-sample simulation results. In Section 4, we illustrate the use of the GMM estimator in an empirical application to a small United States macroeconomic model. Finally, Section 5 concludes. The detailed discussion on the conditions for local identification as well as the proof of the related proposition are deferred to the Appendix.

2. MODEL SETUP

2.1. Structural Vector Autoregression

We consider the structural VAR model of order p,

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

where y_t is the n-dimensional time series of interest, ν $(n \times 1)$ is an intercept term, $A_1, \ldots, A_p$ 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 stable, and hence weakly stationary, that is,

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

In the literature, model (1) is often referred to as the B-model (see, e.g., Lütkepohl 2005, chap. 9). An alternative SVAR formulation is obtained by left-multiplying (1) by the inverse of B

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

where $A_0 = B^{-1}$, $\nu^* = B^{-1}\nu$, and $A_j^* = B^{-1}A_j$ $(j = 1, \ldots, p)$. Typically, in this so-called A-model (Lütkepohl 2005, chap. 9), the diagonal elements of A_0 are normalized to unity, and the covariance matrix of ε_t is a diagonal matrix. Model (3) is useful when the main interest is in quantifying the contemporaneous 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, Meitz, and Saikkonen (2017) showed that identification of B (up top 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, Meitz, and Saikkonen (2017). In contrast to that article, 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. However, it is important to notice that orthogonality of the errors is not sufficient for identification, but, in addition, a number of fourth-moment restrictions implied by independence must hold. Specifically, we make the following assumption on model (1), under which B is shown to be locally identified up to permutation and multiplication by -1 of its columns (see Proposition 1 in Section 2.2):

Assumption 1.

(i) The error process $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})'$ is a sequence of (strictly) stationary random vectors with each component ε_{it}, $i = 1, \ldots, n$, having mean zero and variance unity.
(ii) The components $\varepsilon_{it}, \ldots, \varepsilon_{nt}$ are uncorrelated in time, that is, $\text{cov}\left(\varepsilon_{it}, \varepsilon_{i,t+k}\right) = 0$ for all $k \neq 0$.
(iii) The components $\varepsilon_{1t}, \ldots, \varepsilon_{nt}$ are (mutually) orthogonal and at most one of them has a Gaussian marginal distribution.
(iv) $E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0$, for at least $n(n-1)/2$ combinations of i and j $(i \neq j)$.

Because matrix A_0 in Equation (3) is obtained by inverting matrix B in Equation (1), it is identified up to permutation and multiplication by -1 of its rows without further restrictions under Assumption 1. If the diagonal elements of A_0 are normalized to unity, the variances of the components of ε_t are left unrestricted. For ease of exposition, we will henceforth, for the most part, explicitly only consider the B-model formulation Equation (1), but it is to be understood that everything applies to the A-model (3) as well, with obvious modifications.

2.2. Moment Conditions

The model can be consistently estimated by the GMM as discussed in Section 3. In this section, we discuss the moment conditions to be used in GMM estimation, and show that the impact matrix B (its inverse A_0) is identified up to permutation and multiplication by -1 of its columns (rows) under Assumption 1. 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 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. However, non-Gaussianity of the reduced-form VAR residuals does not guarantee non-Gaussianity of $n-1$ of the structural shocks, which has to be checked after the SVAR model has been estimated.

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

$$y_t = \Psi \mathbf{x}_{t-1} + B\varepsilon_t, \tag{4}$$

where the $((np+1) \times 1)$ vector $\mathbf{x}_{t-1} = (1, y_{t-1}', \ldots, y_{t-p}')'$ and $\Psi = (\nu, A_1, \ldots, A_p)$. From Assumption 1(i) and the lags of y_t being predetermined, we obtain the following $2n + pn^2$ moment conditions:

$$E\left(\varepsilon_t \otimes \mathbf{x}_{t-1}\right) = 0_{n(np+1)\times 1} \tag{5a}$$

$$E\left(\varepsilon_{it}^2\right) - 1 = 0, \quad i = 1, \ldots, n \tag{5b}$$

where $\otimes$ denotes the Kronecker product. In the A-model (3) with unrestricted error variances, condition (5b) is replaced by restricting the diagonal elements of A_0 equal to unity. 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(ii). Furthermore, mutual orthogonality of the components of ε_t in Assumption 1(iii) results in $n(n-1)/2$ orthogonality conditions of the form

$$E(\varepsilon_{it}\varepsilon_{jt}) = 0, \quad i \neq j. \tag{5c}$$

The $2n + pn^2 + n(n-1)/2$ moment conditions in Equations (5a)–(5c) 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 independence, that are informative when (at least $n-1$ of) the components of the error term ε_t are non-Gaussian in accordance with Assumption 1(iii). It is well known that the 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 (5a)–(5c).

Our idea is to base estimation on some co-kurtoses 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 heteroscedasticity, as exemplified below, among other things. Our asymmetric and symmetric co-kurtosis conditions are of the form

$$E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0, \quad i \neq j, \tag{6}$$

and

$$E(\varepsilon_{it}^2 \varepsilon_{jt}^2) - 1 = 0, \quad i \neq j, \tag{7}$$

respectively. It is easy to see that both kinds of conditions are indeed implied by the mutual orthogonality conditions in (5c) and are thus redundant under Gaussianity. In conditions (6), the asymmetry manifests itself in that $E(\varepsilon_{jt}^3 \varepsilon_{it})$ need not equal zero even if $E(\varepsilon_{it}^3 \varepsilon_{jt}) = 0$ is satisfied. While at least $n(n-1)/2$ asymmetric co-kurtosis conditions are required for identification (see Proposition 1 below), symmetric co-kurtosis conditions only provide additional information that may increase the accuracy of the GMM estimator.

As an example of a SVAR model that can be estimated using the moment conditions discussed above, suppose the data are generated by model (1), with the error term ε_t following a one-factor stochastic volatility process $\varepsilon_t = \sigma_t z_t$, where $E(\sigma_t^2) = 1$, z_t is an n-dimensional normally independently distributed vector with mean zero and identity covariance matrix, and σ_t and z_t are contemporaneously independent (we would like to thank an anonymous referee for pointing out this example). The components of the error term ε_t are obviously orthogonal, but dependent, and non-Gaussian, and the asymmetric conditions (6) hold as $E(\varepsilon_{it}^3 \varepsilon_{jt}) = E(\sigma_t^4) E(z_{it}^3) E(z_{jt}) = 0$ when $i \neq j$. This example also illustrates that one also has to be careful with symmetric co-kurtosis conditions because they do not hold under any type of non-Gaussianity. In the case of this particular stochastic volatility process, the Cauchy-Schwarz inequality implies that $E(\varepsilon_{it}^2 \varepsilon_{jt}^2) = E(\sigma_t^4) E(z_{it}^2) E(z_{jt}^2) = E(\sigma_t^4)$ equals unity only if σ_t is constant (almost surely), that is, there is no stochastic volatility and the shocks are independent.

As a matter of fact, it may be hard to come by economically relevant examples of error processes where the symmetric co-kurtosis conditions hold, when the errors are mutually orthogonal but not independent. Nevertheless, in case of independence

of (some of) the components of the error term, they may provide useful additional information for inference. In any case, as part of the moment selection procedure, we recommend testing for over-identifying restrictions, to ensure that the selected moment conditions agree with the data (see Section 3.3).

If all the components of ε_t are non-Gaussian, the local identification condition (10) is satisfied whenever the set of moment conditions contains any $n(n-1)/2$ asymmetric co-kurtosis conditions in addition to conditions (5a)–(5c), as stated in Proposition 1. However, if one of the components of ε_t is 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. However, in case there are more moment conditions than parameters to estimate, it is possible to achieve local identification by a suitable selection of asymmetric co-kurtosis conditions even if any one of the shocks is Gaussian (for instance, by including all $n(n-1)$ asymmetric co-kurtosis conditions).

Proposition 1. (Local identification) Suppose all n components of ε_t are non-Gaussian with $E(\varepsilon_{it}^3) \neq 0$ and/or $E(\varepsilon_{it}^4) \neq 3$, $i = 1, \ldots, n$. Then moment conditions (5a)–(5c), and $n(n-1)/2$ asymmetric co-kurtosis conditions of the form (6) exactly locally identify the parameters of SVAR model (1) (SVAR model (3)) characterized by a given permutation of the columns of B and signs of its elements. 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. □

It is important to realize that Proposition 1 only applies to a given SVAR model characterized by a given ordering of the columns of B and signs of its elements. In particular, asymptotically all the permutations of the columns of B potentially satisfy the moment conditions. Thus, to facilitate asymptotic inference, additional restrictions are needed to pinpoint a particular permutation of the columns of B and signs of its elements. These restrictions are not restrictive, however, because any permutation of the columns of B (asymptotically) produces the same shocks (reordered) and impulse responses. To this end, there are many alternative restrictions entertained in the previous literature on statistical identification that could be employed (see Lanne, Meitz, and Saikkonen 2017, and the references therein). In this article, we use the permutation convention of Pham and Garat (1997) which entails picking the permutation that maximizes the absolute value of the product of the diagonal elements of B, and restrict the diagonal elements of B positive.

3. STATISTICAL INFERENCE

3.1. GMM Estimator

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

$$Q_T(\theta) = T^{-1}\sum_{t=1}^{T} f(v_t,\theta)' W_T T^{-1}\sum_{t=1}^{T} f(v_t,\theta), \tag{8}$$

where $\theta = \left(\nu', \mathrm{vec}(A_1)', \ldots, \mathrm{vec}(A_p)', \mathrm{vec}(B)'\right)'$ is a row vector of $k \equiv n + (p+1)n^2$ parameters to be estimated, v_t, $t = 1, 2, \ldots, T$, consists 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

$$E\left[f(v_t,\theta_0)\right] = 0, \tag{9}$$

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 3.2). Finding a convenient condition for global identification is, in general, difficult in the context of a nonlinear model such as the SVAR model. However, for global identification to hold, the parameters must be identified in a neighborhood of θ_0, which is the case if the rank of the matrix of the expected partial derivatives of $f(v_t,\theta)$ with respect to the parameters evaluated at the true parameter values θ_0 equals the number of parameters,

$$\mathrm{rank}\{E[\partial f(v_t,\theta_0)/\partial\theta']\} = k. \tag{10}$$

This condition of local identification can be guaranteed by selecting a suitable set of moment conditions (see Proposition 1), of which there must be at least k. 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 3.3.

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 use 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 is estimated consistently (under regularity conditions, see Newey and West [1994]) as the following heteroscedasticity and autocorrelation consistent covariance (HAC) matrix estimator

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

where $\hat{\Gamma}_i$ is a consistent estimator of Γ_i, the ith autocovariance matrix of $f(v_t,\theta_0)$. The HAC estimator allows for heteroscedasticity 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. 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 3.4 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 Equation (8) with W_T suboptimal (such as the identity matrix), and then reestimating θ based on $\hat{S}_{\text{HAC}}$ 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 estimator (CUE) of Hansen, Heaton, and Yaron (1996) acknowledges the dependence of the efficient weighting matrix on the parameters. It is based on 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

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

is of the same form as the HAC estimator discussed above. All three estimation methods are implemented in the R package gmm (Chaussé 2010) that we use to produce the empirical and simulation results in this article. As discussed in Section 3.2, 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. Unfortunately, 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 3.4 pertaining to the estimation of the SVAR model suggests that the two-step estimator is superior to the other GMM estimators.

3.2. Asymptotic Inference

To be able to apply the asymptotic results related to the GMM estimator derived in the literature, we make a number of standard assumptions, including strict stationarity and ergodicity (see, e.g., Hall 2005, chaps 3 and 5.3), in addition to identification. Under these assumptions, the GMM estimator $\hat{\theta}_T$ is a consistent estimator of θ_0. This holds for all two-step, iterated and continuous updating GMM estimators that are asymptotically equivalent although they may behave differently in finite samples. Moreover, the efficient GMM estimator is asymptotically normally distributed with covariance matrix $\left(G_0' S^{-1} G_0\right)^{-1}$, where $G_0 = E\left[\partial f(v_t,\theta_0)/\partial\theta'\right]$ and $S = \lim_{T\to\infty}\text{var}\left[T^{1/2}\left(T^{-1}\sum_{t=1}^{T} f(v_t,\theta_0)\right)\right]$. Because the SVAR model is statistically identified, additional restrictions on the parameters can be tested, once G_0 and S are replaced by their consistent estimators, $G_T(\hat{\theta}_T)$ and $\hat{S}_{\text{HAC}}$, respectively. However, it must be borne in mind that any test on the parameters of the impact matrix only pertains to the particular ordering of its columns. Therefore, any hypothesis on the parameters is, in general, meaningful only once the shocks pertaining to those parameters have been economically identified by, say, sign constraints (see the empirical example in Section 4).

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

$$H_0 : r(\theta_0) = 0 \qquad \text{versus} \qquad H_A : r(\theta_0) \neq 0$$

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 nonredundant restrictions as there are parameters in θ. The tests considered by Newey and West are extensions of asymptotic tests related to ML 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

$$Tr(\hat{\theta}_T)'\left[R(\hat{\theta}_T)[G(\hat{\theta}_T)'\hat{S}_T(\hat{\theta}_T)^{-1}G_T(\hat{\theta}_T)]^{-1}R(\hat{\theta}_T)'\right]^{-1} r(\hat{\theta}_T). \tag{11}$$

While Equation (11) depends only on the unrestricted estimate, the likelihood ratio (LR) type test statistic

$$T[Q_T(\tilde{\theta}_T) - Q_T(\hat{\theta}_T)] \tag{12}$$

is based on the change in the minimum of the objective function between the restricted and unrestricted models. Under standard assumptions, both Equations (11) and (12) 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 reparameterization of the model or the restrictions. As shown by Hall and Peixe (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, misspecification testing should always precede (see Section 3.3) inference on the parameters.

It is also important to notice that the co-kurtosis conditions may provide only weak identification if the true errors only slightly deviate from Gaussianity. This is akin to the problem of weak instruments provided by short-run and long-run restrictions in identifying SVAR models considered by Pagan and Robertson (1998) and Gospodinov (2010). It is well known that under weak identification, standard asymptotic inference in the GMM framework is inappropriate (see, e.g., Hall 2005, sec. 8.2), and the deterioration of the asymptotic approximation when the error distributions approach the Gaussian distribution, is also clearly demonstrated by the simulation results in Section 3.4. Therefore, it is important to check for the non-Gaussianity of the structural errors.

3.3. Over-Identifying Restrictions Test and Moment Selection

As discussed in Section 2.2, the SVAR model (1) (or (3)) is locally exactly identified if estimation is based on conditions (5a)–(5c) and $n(n-1)/2$ asymmetric co-kurtosis conditions of the form (6). 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 of over-identifying restrictions is available whenever there are more moment conditions than parameters to estimate ($q > k$). When the model is exactly identified, that is, $q = k$, 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, $J_T = TQ_T(\hat{\theta}_T)$, follows asymptotically the χ^2 distribution with $q - k$ degrees of freedom under the null hypothesis of correct specification, and it is convenient in that it is obtained as a by-product of estimation.

Typically, several alternative over-identifying sets of moment conditions agree with the data, and a number of methods of selecting the optimal set among them have been put forth in the literature. In this article, we successively employ Andrews's (1999) information criterion based approach, and the relevant moment selection criterion proposed by Hall et al. (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).

Andrews's (1999) moment selection criterion

$$\mathrm{MSC}(c) = J_T(c) - (q - k)\ln(T) \tag{13}$$

is computed for several sets of moment conditions, indexed by selection vector c, and the over-identifying set minimizing its value is selected. The selection vector c, with elements equal to either 0 or 1, corresponds to the $q_{\max}$ dimensional vector of all potential moment conditions $f_{\max}(\cdot)$, whose jth element is picked if $c_j = 1$. The first term of Equation (13) is just the value of the J statistic of over-identifying restrictions, whose small values lend support to the moment conditions, while the latter term increases with the degree of over-identification $(q - k)$. 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

$$\mathrm{RMSC}(c) = \ln[|\hat{V}_{\theta,T}(c)|] + (q - k)\ln[(T/b_T)^{1/2}](T/b_T)^{-1/2} \tag{14}$$

is, in turn, concerned with the efficiency and nonredundancy of the moment conditions. The smaller is first term, dependent on the covariance matrix $\hat{V}_{\theta,T}(c)$ of the estimator, the more accurately the parameters have been estimated. The bandwidth parameter b_T of the $\hat{S}_{\mathrm{HAC}}$ estimator accounts 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 nondecreasing in the number of moment conditions, whereas the second term penalizes for additional conditions, attempting to avoid redundancy.

For practical moment selection, we recommend a version of the combined strategy of Hall (2005, sec. 7.3.3), where the MSC and RMSC are employed in succession. In the second step, the optimal set of moment conditions is obtained by minimizing the RMSC over all admissible subsets of the set selected by the MSC in the first step. In all cases, the optimal set contains conditions (5a)–(5c), and the two-step procedure is used to augment them with the optimal overidentifying combination of at least $n(n-1)/2$ asymmetric co-kurtosis conditions, and potentially some symmetric co-kurtosis conditions.

We first estimate the model with all over-identifying combinations of at least $n(n-1)/2$ asymmetric and $0, \ldots, \binom{n}{2}$ symmetric co-kurtosis conditions, and select among them the combination of conditions that minimizes the MSC. The set of moment conditions selected by the MSC contains the maximal number of conditions supported by the data. Then, we estimate the model with all combinations of the moment conditions included in this set, and select the optimal 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 include in each set a sufficient number of moment conditions such that the model remains over-identified ($q > k$).

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. 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 (5a)–(5c)) not rejected by the J-test, and then dropping 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.

3.4. Finite-Sample Properties

To gauge the properties of the GMM estimator in small samples, we conduct a number of Monte Carlo simulation experiments. To facilitate comparison to the PML and recursive PML estimators of Gouriéroux, Monfort, and Renne (2017), we first consider the bivariate SVAR(0) model that they used in their simulations:

$$y_t = B\varepsilon_t,$$

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 12, 24, or 48 degrees of freedom, and standardized to have variance unity. With increasing degrees of freedom, the t-distribution approaches the normal distribution, which is expected to show up as deteriorating performance of both the GMM and PML estimators, as identification is based on non-Gaussianity. We base GMM estimation on the following set of 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.

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

	$T = 250$			$T = 500$			$T = 1000$		
DF	12	24	48	12	24	48	12	24	48
					Bias				
Two-step	0.043	0.061	0.065	0.030	0.052	0.061	0.013	0.042	0.060
Iterated	0.061	0.077	0.080	0.037	0.062	0.072	0.018	0.047	0.066
CUE	0.061	0.076	0.078	0.039	0.061	0.070	0.018	0.048	0.065
Recursive PML	−0.103	−0.116	−0.124	−0.098	−0.125	−0.122	−0.091	−0.117	−0.124
PML	0.033	0.056	0.074	0.015	0.046	0.068	0.005	0.031	0.058
					Standard deviation				
Two-step	0.097	0.099	0.099	0.088	0.091	0.092	0.075	0.087	0.090
Iterated	0.103	0.102	0.102	0.091	0.094	0.094	0.078	0.088	0.092
CUE	0.101	0.101	0.101	0.090	0.093	0.093	0.077	0.088	0.092
Recursive PML	0.307	0.324	0.333	0.297	0.328	0.327	0.285	0.319	0.330
PML	0.082	0.090	0.090	0.072	0.086	0.090	0.061	0.083	0.088
					Rejection rate of LR (type) test				
Two-step	0.076	0.093	0.098	0.066	0.081	0.076	0.066	0.057	0.074
Iterated	0.087	0.109	0.119	0.065	0.100	0.099	0.058	0.057	0.079
CUE	0.083	0.109	0.114	0.062	0.089	0.088	0.058	0.057	0.080
PML	0.036	0.013	0.001	0.036	0.030	0.005	0.036	0.040	0.016
					Rejection rate of J-test				
Two-step	0.059	0.041	0.037	0.047	0.038	0.032	0.053	0.032	0.025
Iterated	0.043	0.029	0.022	0.041	0.026	0.022	0.053	0.032	0.019
CUE	0.035	0.023	0.022	0.037	0.025	0.019	0.051	0.029	0.019

NOTES: The results for the two-step, iterated and continuous updating (CUE) GMM estimators as well as the PML and recursive PML estimators of Gouriéroux, Monfort, and Renne (2017) are based on 5000 simulated samples of $T = 250$, 500, and 1000 observations. The components of the error term $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, are first generated from independent t-distributions with 12, 24, or 48 degrees of freedom (DF). 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 four panels contain the bias, standard deviation and the rejections rates of the LR (type) test of B_{11} being equal to its true value and the J test of over-identifying restrictions at the 5% nominal significance level. GMM estimation is based on 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 $E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = 1$. The PML estimators assume the true error distribution.

The two topmost panels of Table 1 contain the bias and standard deviation of the estimates of B_{11}. Recall that B is identified only up to permutation and multiplication by -1 of its columns. Therefore, the estimator of B_{11} may estimate either B_{11}, $-B_{11}$, B_{12} or $-B_{12}$, and the measures of bias and standard deviation are based on a transformation of the estimate of B_{11}, that is, closest to the true value of B_{11} (minimizing the squared deviation; cf. Gouriéroux, Monfort, and Renne 2017, sec. 2.7).

With given degrees of freedom of the t-distribution assumed for the error term, the bias and standard deviation of all estimators decreases with increasing sample size, as might be expected. With greater degrees of freedom, identification is weaker, and the estimators tend to have greater bias and standard deviation. Perhaps somewhat surprisingly, the two-step GMM estimator is more accurate than the iterated and continuous updating estimators in all cases. The PML estimator outperforms both the GMM and recursive PML estimators when based on the true error distribution. In practice, the error distribution is, of course, unknown, and while the PML estimator seems quite robust with respect to misspecification of the degrees of freedom, with an incorrect distribution, its performance can deteriorate to some extent.

In the two bottom panels of Table 1, we report the rejection rates of the LR (type) test of B_{11} being equal to its true value and the J-test of over-identifying restrictions. In both cases, the nominal significance level of the test is 5%. The LR test based on PML estimation is under-sized in all cases, with rejection rates close to zero in the case of 48 degrees of freedom, when the error distribution is very close to being Gaussian. When the number of degrees of freedom is misspecified, severe size distortions may occur (these results are not shown to save space, but they are available upon request). The LR type test related to GMM estimation is somewhat over-sized, and severest over-rejection takes place when the sample size is small and the error distribution is close to a Gaussian distribution. Of the three GMM estimators, the two-step estimator tends to be the most accurate in terms of size. The Wald test, in contrast, exhibits severe size distortions, with the rejection rate at least 14% in all cases (the results are not shown to save space, but are available upon request). Hence, based on this limited simulation evidence, it is advisable to rely on the LR type test instead of the Wald test. The J-test follows a similar pattern. When the error terms follow the t distribution with 12 degrees of freedom, the rejection rates are relatively close to the nominal size of the test even in small samples, but as identification weakens with increasing degrees of freedom, the rejection rates fall.

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

Table 2. Simulation results of the two-step GMM estimator of the SVAR(1) model

		$T = 250$			$T = 500$			$T = 1000$		
C_{11}	DF	Bias	Std.	J-test	Bias	Std.	J-test	Bias	Std.	J-test
0.0	12	0.036	0.096	0.056	0.025	0.085	0.050	0.012	0.075	0.063
	24	0.050	0.094	0.047	0.047	0.091	0.037	0.040	0.086	0.034
	48	0.057	0.097	0.040	0.061	0.091	0.033	0.056	0.090	0.024
0.5	12	0.036	0.096	0.057	0.025	0.085	0.049	0.012	0.074	0.063
	24	0.050	0.094	0.046	0.047	0.092	0.037	0.040	0.086	0.033
	48	0.056	0.096	0.041	0.061	0.091	0.033	0.056	0.090	0.024
0.9	12	0.037	0.096	0.050	0.025	0.085	0.050	0.012	0.075	0.066
	24	0.051	0.095	0.040	0.046	0.091	0.037	0.041	0.086	0.034
	48	0.058	0.096	0.036	0.061	0.092	0.030	0.056	0.090	0.024
0.97	12	0.041	0.096	0.033	0.026	0.085	0.052	0.012	0.074	0.049
	24	0.054	0.096	0.022	0.047	0.091	0.038	0.041	0.086	0.025
	48	0.061	0.096	0.019	0.063	0.092	0.029	0.057	0.090	0.019

NOTES: See the notes to Table 1. The data are generated from the DGP in Equation (15).

previous data-generating process

$$y_t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} C_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix} y_{t-1} + B\varepsilon_t, \tag{15}$$

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 reported in Table 2 are comparable to those in Table 1, with the bias and standard deviation in most cases even smaller. While the persistence seems to have little effect on the properties of the estimator, the rejection rate of the J test seems to fall more with weak identification when persistence is higher.

To gauge the performance of the J-test in a larger system, we consider a trivariate SVAR model, where the B matrix in the data generating process is taken to be the estimated impact matrix $\hat{B}$ in our empirical application in Section 4. Each of the independent components of ε_t is generated from a Student's t-distribution with 12, 24, or 48 degrees of freedom, and standardized to have variance unity. For simplicity, the model contains no lags. We use four asymmetric and one symmetric co-kurtosis conditions in estimation (the ones included in the optimal set in the empirical application in Section 4), and because only three asymmetric conditions are necessarily needed for identification, the model is over-identified, with two over-identifying conditions. In all cases, the rejection rates are lower than 5%, and this example demonstrates that when the errors are close to being Gaussian, the J-test can be severely undersized. With errors following the t distribution with only 12 degrees of freedom, the rejection rate of the J-test slowly moves closer to the nominal size as the sample size increases, but with as few as 250 observations, there are great differences among the three GMM estimators such that the two-step estimator seems to be the most reliable in small samples.

All in all, our simulation results warrant three conclusions. First, the GMM estimators are slightly inferior to the PML estimator of Gouriéroux, Monfort, and Renne (2017) in accuracy (assuming the correct error distributions are known), but outperform it in terms of size of the LR (type) test. Moreover, both the GMM and PML estimators are superior to the recursive PML estimator. Second, the performance of both the GMM and PML estimators deteriorates as the error distributions get closer to normal distributions. Third, the two-step GMM estimator seems superior to the iterated and CUE estimators in terms of accuracy and size of the LR type and J-tests. The falling rejection rates of the J-test with weakening identification as well as the superiority of the two-step estimator are also in line with the findings of Gospodinov, Kan, and Robotti (2014) and Gospodinov, Kan, and Robotti (2017). However, our simulation study is quite limited, and should be extended in future research. Among other things, our conclusions may depend on a number of things, including the data-generating process and the moment conditions employed.

4. EMPIRICAL ILLUSTRATION

We demonstrate SVAR analysis based on GMM estimation by means of an empirical application to quarterly United States macroeconomic data covering the period from 1960:I to 2017:II (230 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. In the period 2009–2015, when the federal funds rate was virtually constant at the zero lower bound, it is replaced by the "shadow rate" of Wu and Xia (2016). Otherwise, 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 2 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

Table 3. Rejection rates of the J-test at the nominal 5% significance level in the trivariate SVAR model

	$T = 250$			$T = 500$			$T = 1000$		
DF	12	24	48	12	24	48	12	24	48
Two-Step	0.021	0.008	0.004	0.027	0.005	0.004	0.032	0.007	0.003
Iterated	0.009	0.002	0.001	0.017	0.003	0.001	0.029	0.004	0.002
CUE	0.006	0.003	0.002	0.021	0.004	0.002	0.034	0.007	0.003

NOTES: The results for the two-step, iterated, and continuous updating (CUE) GMM estimators are based on 5000 simulated samples of T = 250, 500, and 1000 observations. The components of the error term $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$, are first generated from independent t-distributions with 12, 24, or 48 degrees of freedom (DF). Then the data y_t are computed from $y_t = B\varepsilon_t$, where B equals the estimated matrix of impact effects $\hat{B}$ in Section 4. The co-kurtosis conditions used in GMM estimation are the following: $E(\varepsilon_{1t}^3\varepsilon_{2t}) = 0, E(\varepsilon_{2t}^3\varepsilon_{3t}) = 0, E(\varepsilon_{3t}^3\varepsilon_{1t}) = 0, E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$, and $E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = 1$.

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 all of the latter would imply normality of the former and, hence, violation of identification. The quantile–quantile (Q–Q) plots of the reduced form residuals in the upper panel of Figure 1 indicate heavy tails and, hence, clear deviation from normality, which suggests that the necessary conditions for identification might be satisfied.

To find the optimal set of moment conditions, we next estimate a three-variable SVAR(4) model using different combinations of these conditions. In view of the simulation results in Section 3.4, we use the two-step GMM estimator throughout. In all cases, conditions (5a)–(5c) are included, while the rest of the moment conditions are selected by the sequential procedure outlined in Section 3.3. With three variables in the SVAR, there are six asymmetric and three symmetric co-kurtosis conditions in total. In the first step, the set of moment conditions minimizing the MSC contains all six asymmetric co-kurtosis conditions and two symmetric co-kurtosis conditions ($E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = E(\varepsilon_{2t}^2\varepsilon_{3t}^2) = 1$). The p-value of the J-test for this set equals 0.571. In the second step, we consider all over-identifying subsets of these conditions not rejected by the J-test at the 5% level. The set minimizing the RMSC among these subsets contains four asymmetric co-kurtosis conditions ($E(\varepsilon_{1t}^3\varepsilon_{2t}) = E(\varepsilon_{2t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{1t}) = E(\varepsilon_{3t}^3\varepsilon_{2t}) = 0$) and one symmetric co-kurtosis condition ($E(\varepsilon_{1t}^2\varepsilon_{2t}^2) = 1$). The p-value of the J-test for the selected optimal set is 0.661. In addition to the set of moment conditions selected, we also experimented with a number of alternative over-identifying sets, and in this particular application, the results turned out to be quite robust with respect to the moment conditions entertained.

The Q–Q plots in the lower panel of Figure 1 indicate non-Gaussianity of all three structural errors, so local identification is reached. The GMM estimate of the matrix of impact effects is

$$\hat{B} = \begin{bmatrix} \underset{(0.077)}{0.901} & \underset{(0.074)}{-0.153} & \underset{(0.063)}{0.185} \\ \underset{(0.030)}{0.050} & \underset{(0.015)}{0.264} & \underset{(0.022)}{-0.074} \\ \underset{(0.093)}{-0.076} & \underset{(0.069)}{-0.129} & \underset{(0.250)}{0.959} \end{bmatrix}$$

where the figures in parentheses are asymptotic standard errors. According to the asymptotic Wald test, only the (2,1) and (3,1) elements are insignificant at the 5% level. However, in view of the simulation results in Section 3.4, the Wald test tends to over-reject, and, therefore, we also conducted the LR type test of significance of the elements of the impact matrix. These tests indicate significance of only the (2,2), (2,3), and (3,3) elements. Thus, only the third shock appears to have a significant effect (at the 5% level) on impact on the federal funds rate, suggesting that it 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 for the identification scheme considered in the previous literature that B is a lower triangular matrix, that is, 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 p-values 2.79e–12 and 0.013 in the Wald and LR type tests, respectively. The markedly smaller p-value of the Wald test may indicate its poor finite-sample properties, but in any case, there seems to be little support for the recursive identification scheme popular in the previous literature. It must be borne in mind, though, that the test is based on the assumption that the third shock is the monetary policy shock.

The impulse responses of the three shocks along with their 95% confidence bands are depicted in Figure 2. 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 LR type test, and this conclusion is reconfirmed by the bootstrapped confidence bands in the rightmost column of Figure 2. 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 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 2) has a positive effect on both inflation and unemployment 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 on impact, in turn, can be labelled as a contractionary demand shock. Visual

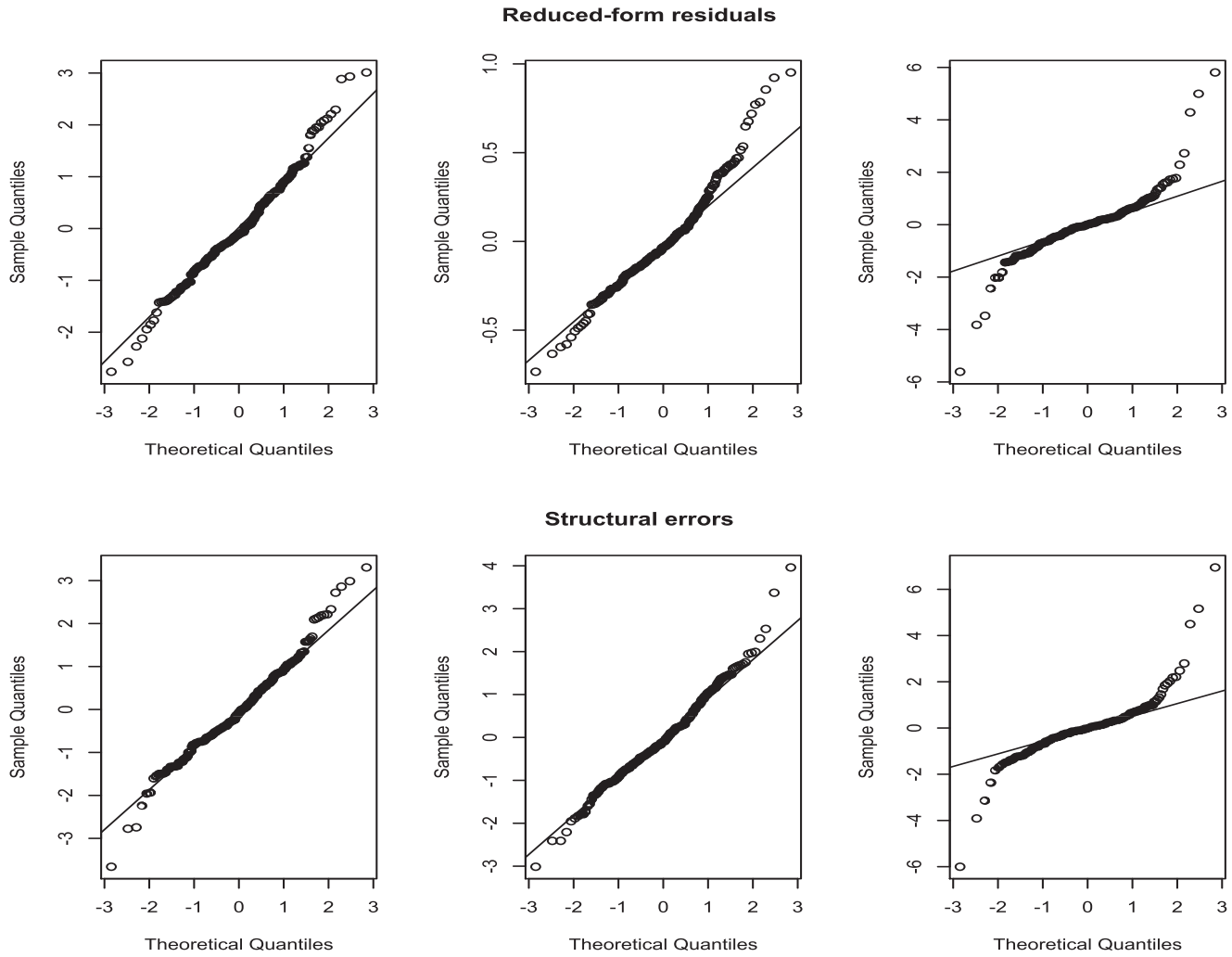

Figure 1. Quantile–quantile plots of the residuals (upper panel) of the equations of inflation, unemployment gap and the federal funds rate (from left to right) and the structural errors (lower panel) of the selected SVAR(4) model.

inspection of the confidence bands suggests that asymptotic theory does not provide a very accurate approximation. 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 asymptotic tests. It must, of course, be borne in mind that the tests involving joint restrictions and confidence intervals on single parameters are not directly comparable. Moreover, the results should be interpreted with caution because the residual-based wild bootstrap does not produce confidence intervals with asymptotically correct coverage rates in the presence of conditional heteroscedasticity, especially at short horizons. Nevertheless, according to the simulation results of Brüggemann, Jentsch, and Trenkler (2016), it performs reasonably well and is not outperformed by the asymptotically valid residual-based moving block bootstrap even in moderately large samples.

5. CONCLUSION

In this article, 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 identify the parameters of the SVAR model up to the ordering and scaling of the structural shocks. Our approach deviates from the related statistical identification literature in that the structural shocks do not have to be independent, and no particular distributional assumptions are required. 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, and it precludes various forms of conditional heteroscedasticity. Although the PML estimators of Gouriéroux, Monfort, and Renne (2017) are to some extent robust with respect to distributional misspecification, our approach is generic in that the optimal set of moment conditions can in any application be selected by conventional moment selection criteria. There seems to be no corresponding procedure for specifying the error distributions in ML and PML estimation, and the selected distributions may affect the results. According to our simulation study the performance of the GMM estimator seems comparable to that of the PML estimator and superior to the recursive PML estimator.

Statistical identification rarely produces structural shocks with economic interpretation, and our approach is no excep-

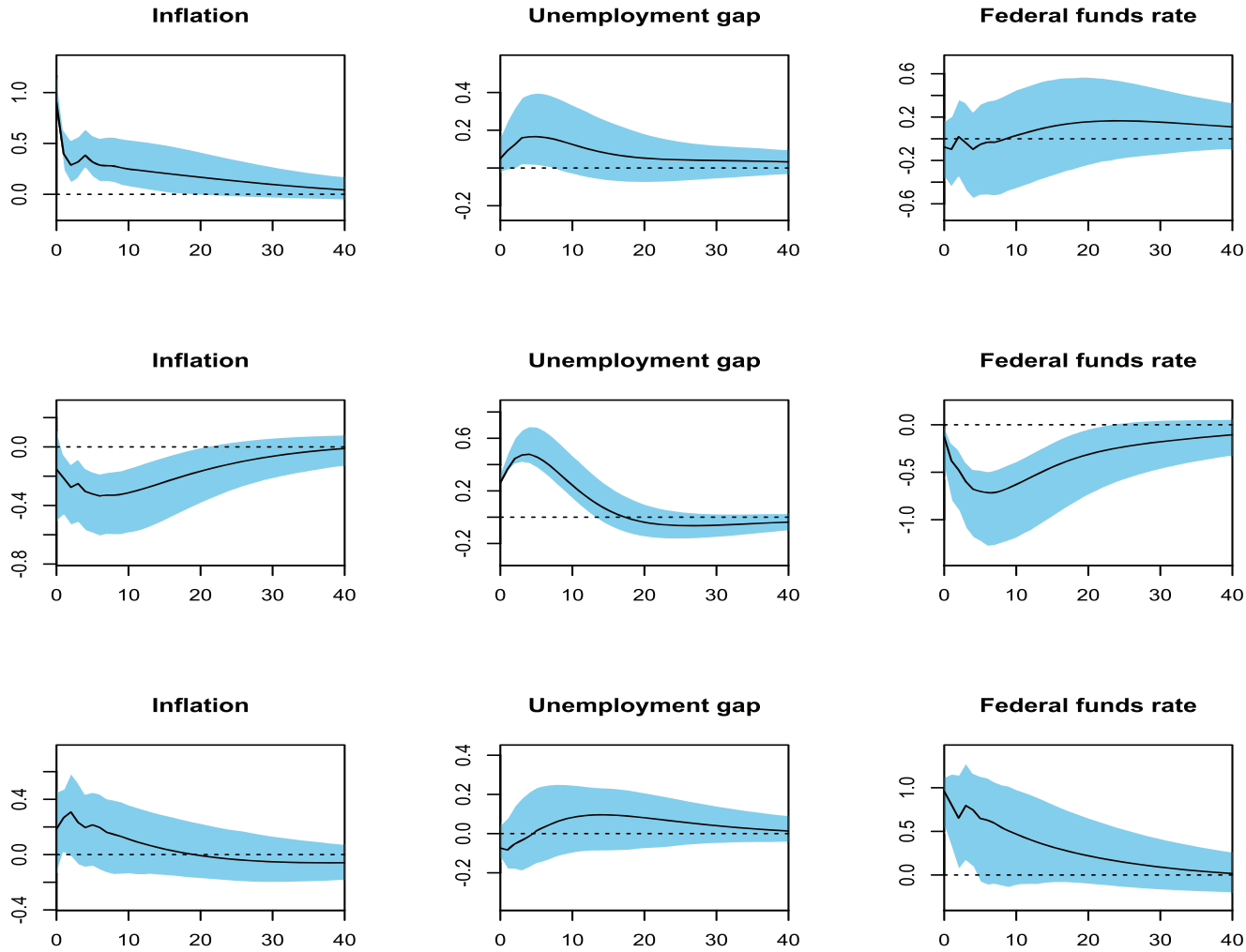

Figure 2. Impulse responses of the three shocks in the SVAR model estimated using the full set of moment conditions. Each row contains the impulse responses of one shock on all variables. The shaded areas are the pointwise 95% Hall's percentile confidence bands obtained by residual-based wild bootstrap with 10,000 replications.

tion. In our empirical application to a trivariate United States macroeconomic 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 are testable in our setup. As a matter of fact, we tested recursive identification restrictions entertained in some of the related empirical literature, and they were strongly rejected. 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.

As discussed in Section 3.2, identification may be weak, that is, the condition for local identification may be only marginally satisfied, if the distribution of the error is close to being Gaussian. In that case, standard asymptotic inference may work poorly. Alternative asymptotic theory has recently been derived by Donovon and Hall (2018). They also consider simulation-based methods, including indirect inference, simulated method of moments, and efficient method of moments that our set-up could be extended to as well. It may also be difficult to detect the presence of weak identification. In this article, we rely on checking for non-Gaussianity of the estimated structural shocks, but Guay and Normandin (2018) have recently considered testing for their normality based on the reduced-form VAR model that might be applicable in our set-up as well. Another possibility might be a test based on comparing two different GMM estimators, which have the same limiting distribution only under strong identification as in Inoue and Rossi (2011). We leave these issues for future research.

Appendix: Local Identification

Proof of Proposition 1. The necessary condition for θ to be locally identified is that the expectation of the Jacobian matrix $E\left[\partial f(\upsilon_t, \theta_0)/\partial \theta'\right]$, 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 (5a)–(5c) and $n(n-1)/2$ conditions of the form (6) 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):

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

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

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

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

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

The submatrix consisting of the first $n(np + 1)$ rows in Equation (A.1),the n rows in (A.2), and the $n(n-1)/2$ rows given by (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 its $n(np + 1)$ leftmost columns as

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

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 $E\left[\partial f(\upsilon_t, \theta_0)/\partial\theta'\right]$ must be linearly independent. Finally, it is clear that the n rows in (A.2), and the $n(n - 1)/2$ rows given by (A.3) must be mutually linearly independent because the rows of A are linearly independent.

The remaining rows of $E\left[\partial f(\upsilon_t, \theta_0)/\partial\theta'\right]$ in (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 are linearly independent of the rows given by (A.3) only if at most one of the components of ε_t is Gaussian and suitable asymmetric co-curtosis moment conditions are selected. To see this, 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 the rows given by (A.3), and thus the Jacobian matrix is of full row (and column) rank. In contrast, if the ith component of ε_t is Gaussian, $E(\varepsilon_{it}^4) = 3$ and $E(\varepsilon_{it}^3) = 0$, then it is possible that one of the rows given by (A.4) equals 3 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 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 has no effect on the rank of the Jacobian matrix.□

Let us illustrate Proposition 1 by an example of a trivariate SVAR model ($n = 3$), estimated by imposing three asymmetric co-kurtosis conditions: $E(\varepsilon_{1t}^3\varepsilon_{2t}) = 0$, $E(\varepsilon_{1t}^3\varepsilon_{3t}) = 0$, and $E(\varepsilon_{2t}^3\varepsilon_{3t}) = 0$. Thus, $q = k$. The expectation of the Jacobian matrix becomes

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

It is seen that the last $n(n - 1)/2 = 3$ rows of the Jacobian matrix can be linearly independent of the rows in the block above it 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. If ε_{1t} is Gaussian, the quantities $E(\varepsilon_{1t}^4)$ and $E(\varepsilon_{1t}^3)$ equal 3 and 0, respectively, and the first and second rows in bottom block equal 3 times the first and second rows in the third block, respectively, so that they are linearly dependent. Likewise, if ε_{2t} is Gaussian, the third rows in the middle and bottom blocks are linearly dependent. In contrast, even if the third component ε_{3t} is Gaussian, the Jacobian matrix has full rank because this component does not enter the moment conditions in its third power, and thus its moments do not appear in the Jacobian matrix.

SUPPLEMENTARY MATERIAL

A zip file containing the data and R codes for replicating the empirical results is available online.

ACKNOWLEDGMENTS

The authors would like to thank Todd Clark (the editor) and two anonymous referees for useful comments and suggestions. Financial support from the Academy of Finland (grant 308628) is gratefully acknowledged.

[Received May 2018. Revised May 2019.]

REFERENCES

Andrews, D. W. K. (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–564. [74]

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, DC: Board of Governors of the Federal Reserve System. [69]

Bernanke, B. S., and Mihov, I. (1995) "Measuring Monetary Policy," NBER Working Paper No. 5145. [70]

Brüggemann, R., Jentsch, C., and Trenkler, C. (2016), "Inference in VARs with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 191, 69–85. [78]

Castelnuovo, E. (2016), "Monetary Policy Shocks and Cholesky VARs: An Assessment for the Euro Area," *Emprical Economics*, 50, 383–414. [77]

Chaussé, P. (2010), "Computing Generalized Method of Moments and Generalized Empirical Likelihood with R," *Journal of Statistical Software*, 34, 1–35. [73]

Donovon, P., and Hall, A. R. (2018), "The Asymptotic Properties of GMM and Indirect Inference under Second-Order Identification," *Journal of Econometrics*, 205, 76–111. [79]

Gospodinov, N. (2010), "Inference in Nearly Nonstationary SVAR Models with Long-Run Identifying Restrictions," *Journal of Business & Economic Statistics*, 28, 1–12. [73]

Gospodinov, N., Kan, R., and Robotti, C. (2014), "Spurious Inference in Unidentified Asset-Pricing Models," Federal Reserve Bank of Atlanta Working Paper Series, 2014-12. [76]

——— (2017), "Spurious Inference in Reduced-Rank Asset-Pricing Models," *Econometrica*, 85, 613–1628. [76]

Gospodinov, N., and Ng, S. (2015), "Minimum Distance Estimation of Possibly Noninvertible Moving Average Models," *Journal of Business & Economic Statistics*, 33, 403–417. [70]

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017), "Statistical Inference for Independent Component Analysis: Appliction to Sructural VAR Models," *Journal of Econometrics*, 196, 111–126. [69,70,74,75,76,78]

Guay, A., and Normandin, M. (2018), "Identification of Structural Vector Autoregresions Through Higher Unconditional Moments," Unpublished Working Paper. [70,79]

Hall, A. R. (2005), *Generalized Method of Moments*, New York: Oxford University Press. [73,74]

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. [73,74]

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. [74]

Hall, A. R., and Peixe, F. P. M. (2003), "A Consistent Method for the Selection of Relevant Instruments," *Econometric Reviews*, 7, 269–288. [73,74]

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Mometns Estimators," *Econometrica*, 50, 1029–1054 [72,73]

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. [73]

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. [69]

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. (2010), "Estimation of a Structural Vector Autoregssion Model Using Non-Gaussianity," *Journal of Machine Learning Research*, 11, 1709–1731. [70]

Inoue, A., and Rossi, B. (2011), "Testing for Weak Identification in Possibly Nonlinear Models," *Journal of Econometrics*, 161, 246–261. [79]

Kendall, M. G., and Stuart, A. (1977), *The Advanced Theory of Statistics* (Vol. 1, 4th ed.), London: Charles Griffin and Co. [71]

Kilian, L., and Lütkepohl, H. (2017), *Structural Vector Autoregressive Analysis*, Cambridge, UK: Cambridge University Press. [69]

Lanne, M., and Luoto, J. (2016), "Data-Driven Inference on Sign Restrictions in Bayesian Structural Vector Autoregression," *CREATES Research Paper* 2016–14. [69]

Lanne, M., Meitz, M., and Saikkonen, P. (2017), "Identification and Estimation of Non-Gaussian Structural Vector Autoregressions," *Journal of Econometrics*, 196, 288–304. [69,70,72]

Lewis, D. J. (2018), "Identifying Shocks via Time-Varying Volatility," Staff Report 871, Federal Reserve Bank of New York. [70]

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Berlin: Springer. [70]

Miettinen, J., Taskinen, S., Nordhausen, K., and Oja, H. (2015), "Fourth Moments and Independent Component Analysis," *Statistical Science*, 30, 372–390. [69]

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. [70]

Montiel Olea, J. L., Stock, J. H., and Watson, M. W. (2015), "Inference in Structural VARs With an Extermal Instrument," Working Paper, Columbia University. [70]

Newey, W. K., and West, K. D. (1987), "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 61, 631–653. [73]

——— (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653. [72]

Pagan, A. R., and Robertson, J. C. (1998), "Structural Models of the Liquidity Effect," *Review of Economics and Statistics*, 80, 202–217. [70,73]

Pham, D. T., and Garat, P. (1997), "Blind Separation of Mixture of Independent Sources Through a Quasi-Maximum Likelihood Approach," *IEEE Transactions on Signal Processing*, 45, 1712–1725. [72]

Shapiro, M. D., and Watson, M. W. (1988), "Sources of Business Cycle Fluctuations," in *NBER Macroeconomics Annual*, ed. S. Fischer, Cambridge, MA: MIT Press, pp. 111–148. [70]

Watson, M. W. (1994), "Vector Autoregression and Cointegration," in *Handbook of Econometrics* (Vol. IV), ed. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier, pp. 2844–2915. [70]

Wright, J. H. (2017), "Efficient Inference in Non-Gaussian Structural VARs," Unpublished Working Paper. [70]

Wu, J. C., and Xia, F. D. (2016), "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound," *Journal of Money, Credit, and Banking*, 48, 253–291. [76]