

Estimation of structural vector autoregressive models

Helmut Lütkepohl^{1,a}

^aDIW Berlin and Department of Economics, Freie Universität Berlin, Germany

Abstract

In this survey, estimation methods for structural vector autoregressive models are presented in a systematic way. Both frequentist and Bayesian methods are considered. Depending on the model setup and type of restrictions, least squares estimation, instrumental variables estimation, method-of-moments estimation and generalized method-of-moments are considered. The methods are presented in a unified framework that enables a practitioner to find the most suitable estimation method for a given model setup and set of restrictions. It is emphasized that specifying the identifying restrictions such that they are linear restrictions on the structural parameters is helpful. Examples are provided to illustrate alternative model setups, types of restrictions and the most suitable corresponding estimation methods.

Keywords: Bayesian estimation, maximum likelihood estimation, generalized method-of-moments, method-of-moments, instrumental variables, structural vector autoregression

1. Introduction

In a seminal paper Sims (1980) criticized traditional simultaneous equations systems and proposed using vector autoregressive (VAR) models as alternatives. Since then structural VAR models have become a standard tool for macroeconomic analysis. Structural VAR models are estimated with a variety of methods that depend on the model setup and the type of structural (identifying) restrictions. The estimation techniques that have been used in this context are least-squares (LS), method-of-moments (MM), instrumental variables (IV), generalized method-of-moments (GMM), maximum likelihood (ML) and Bayesian methods. All these methods are described in the related literature (Breitung *et al.*, 2004; Kilian and Lütkepohl, 2017; Lütkepohl, 2005). In this survey a systematic account of the available methods is provided that directs researchers to the appropriate method for a given model setup and set of identifying restrictions. This paper serves researchers who look for a suitable estimation method for a specific application of structural VAR models.

The study focusses explicitly on models where the identifying restrictions are placed on the impact effects of the shocks or on the structural relations of the variables. It sets up the model as a system of simultaneous equations and directs the reader to the most suitable estimation methods for a given set of restrictions. The presentation of the estimation methods partly follows Kilian and Lütkepohl (2017). The methods are not only presented but also practical advice is given which one to use in a specific situation. The emphasis is on exposing the methods rather than specific estimation algorithms because suitable software for most of the methods is available in the internet. (For example, Matlab code for many of the methods mentioned in this review is available at <http://www-personal.umich.edu/~lkilian/book.html>. JMulTi is a menu-driven software that includes

¹ DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany. E-mail: hluetkepohl@diw.de

estimation methods for a range of structural VAR models (Krätzig, 2004). It can be downloaded from <http://www.jmulti.de>. Moreover, EViews is a commercial software with a structural VAR estimation part (EViews, 2000, <http://www.eviews.com/home.html>.) Also statistical inference derived from the estimates is not the focus of this survey because inference methods in structural VAR analysis are more important for functions of the parameters such as impulse responses than for the structural parameters. For impulse responses and related quantities frequentist inference is typically based on bootstrap methods that require the computation of a large number of point estimates. Hence, being able to obtain such estimates is important. This survey can help finding suitable estimation techniques that can be used as basis for bootstrap methods.

The study is structured as follows. In Section 2 the model setup is presented and Section 3 discusses the types of restrictions considered. Sections 4–7 describe a range of estimation methods and illustrations are presented in Section 8. Section 9 concludes and mentions some extensions.

The following general notation is used throughout. The natural logarithm is abbreviated as \log and Δ is the differencing operator defined such that $\Delta y_t = y_t - y_{t-1}$ for a time series variable y_t . The trace and the determinant of a square matrix A are denoted as $\text{tr}(A)$ and $\det(A)$, respectively, whereas $\text{chol}(A)$ denotes the Cholesky decomposition of A . In other words, $\text{chol}(A)$ is a lower-triangular square matrix with nonnegative diagonal elements such that $A = \text{chol}(A)\text{chol}(A)'$. The column stacking operator is signified by vec and the operator which stacks all columns of a square matrix from the main diagonal downward is signified by vech . For a vector x , $\text{diag}(x)$ stands for the diagonal matrix with the components of x on its main diagonal. I_K denotes a $K \times K$ identity matrix. For $n > m$, the orthogonal complement of a $n \times m$ matrix A is denoted by A_\perp . In other words, A_\perp is a $n \times (n - m)$ matrix such that the $n \times n$ matrix $[A, A_\perp]$ is nonsingular and $A'A_\perp = 0$. A normal distribution with mean μ and variance (covariance matrix) Σ is signified as $\mathcal{N}(\mu, \Sigma)$.

2. Model setup

The basic model of interest in this survey is a VAR model of order p , i.e., a VAR(p) model. It is set up in levels VAR form or, if some of the variables are integrated and possibly cointegrated, in vector error correction form. In this section both setups are considered in turn.

2.1. Vector autoregressive models in levels

We consider the K -dimensional reduced-form VAR(p) model

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t = AY_{t-1} + u_t, \quad (2.1)$$

where $Y'_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})$ is $(Kp + 1)$ -dimensional, $A \equiv [\nu, A_1, \dots, A_p]$ is $K \times (Kp + 1)$, and $u_t = (u_{1t}, \dots, u_{Kt})' \sim (0, \Sigma_u)$ is a K -dimensional white noise residual process, i.e., u_t is serially uncorrelated, has zero mean and covariance matrix Σ_u . The $K \times 1$ vector ν is a fixed, nonstochastic intercept term. Additional deterministic terms can be handled straightforwardly but are neglected because they are not important for the following discussion.

The VAR(p) process y_t is stable and stationary if the polynomial

$$\det(I_K - A_1 z - \dots - A_p z^p) \quad (2.2)$$

has all its roots outside the complex unit circle. If the polynomial (2.2) has unit roots ($z = 1$), some or all of the components of y_t are integrated and possibly cointegrated. It is assumed that all variables are at most integrated of order one such that they are stationary in first differences.

To represent structural relations between the variables, the equations (2.1) are multiplied by an invertible $K \times K$ matrix \mathbf{A} such that

$$\mathbf{A}y_t = \mathbf{A}v + \mathbf{A}A_1y_{t-1} + \cdots + \mathbf{A}A_p y_{t-p} + \mathbf{A}u_t = \mathbf{A}AY_{t-1} + \mathbf{A}u_t.$$

In this structural-form VAR representation, the structural innovations or shocks are $w_t = (w_{1t}, \dots, w_{Kt})' = \mathbf{A}u_t$. It is assumed that \mathbf{A} has a unit main diagonal and $w_t \sim (0, \Sigma_w)$ has a diagonal covariance matrix $\Sigma_w = \mathbf{A}\Sigma_u\mathbf{A}' = \text{diag}(\sigma_1^2, \dots, \sigma_K^2)$. Thereby the k^{th} structural equation can be written with y_{kt} as dependent variable on the left-hand side and instantaneous values of other variables and lags of all variables on the right-hand side. In other words, denoting the ij^{th} element of \mathbf{A} by a_{ij} , the k^{th} structural equation is

$$y_{kt} = - \sum_{\substack{i=1 \\ i \neq k}}^K a_{ki}y_{it} + a_k^* Y_{t-1} + w_{kt},$$

where a_k^* is the k^{th} row of $\mathbf{A}\mathbf{A}$. Restrictions have to be imposed on \mathbf{A} to ensure that the structural matrix \mathbf{A} is identified and can be estimated consistently. Using the terminology of Lütkepohl (2005) this model is called the A-model.

Alternatively, the structure can be imposed by defining $w_t = \mathbf{B}^{-1}u_t$ or $u_t = \mathbf{B}w_t$ and considering the structural form

$$y_t = v + A_1y_{t-1} + \cdots + A_p y_{t-p} + \mathbf{B}w_t = AY_{t-1} + \mathbf{B}w_t. \quad (2.3)$$

In other words, the reduced-form innovations u_t are replaced by $\mathbf{B}w_t$. In this setting, w_t is assumed to have unit covariance matrix such that $w_t \sim (0, I_K)$ and, hence, $\Sigma_u = \mathbf{B}\mathbf{B}'$. This model is termed B-model. In this model the identifying restrictions are imposed on the structural matrix \mathbf{B} which represents the impact or short-run effects of the structural shocks w_t .

Occasionally a combination of A- and B-models is of interest to ease the imposition of the identifying restrictions. It is set up as

$$\mathbf{A}y_t = \mathbf{A}v + \mathbf{A}A_1y_{t-1} + \cdots + \mathbf{A}A_p y_{t-p} + \mathbf{A}\mathbf{B}w_t = \mathbf{A}AY_{t-1} + \mathbf{A}\mathbf{B}w_t$$

and is referred to as AB-model (Amisano and Giannini, 1997). In this model, again $w_t \sim (0, I_K)$ and hence, $\Sigma_u = \mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}^{-1'}$. Which model type is considered in a specific situation depends on the nature of the identifying assumptions.

For the purposes of estimation it is useful to note that the exposition of the estimation methods for the structural parameters is sometimes simplified by concentrating out the reduced-form slope parameters A by replacing them with their equation-wise LS estimates, that is, A is replaced by

$$\hat{A} = \sum_{t=1}^T y_t Y_t' \left(\sum_{t=1}^T Y_{t-1} Y_{t-1}' \right)^{-1},$$

if there are no restrictions on A . In other words, defining $\hat{u}_t = y_t - \hat{A}Y_{t-1}$, we can focus on structural equations $\mathbf{A}\hat{u}_t = \tilde{w}_t$, $\hat{u}_t = \mathbf{B}\tilde{w}_t$ or $\mathbf{A}\hat{u}_t = \mathbf{B}\tilde{w}_t$ for estimating the structural parameters \mathbf{A} or/and \mathbf{B} .

2.2. The vector error correction model

Notice that the standard levels form of the VAR(p) model (2.1) can also be used if there are integrated and cointegrated variables. However, for cointegrated VAR models it is often useful to consider a model setup which involves the cointegrating relations explicitly. If there are r linearly independent cointegrating relations, the VAR(p) model can be set up in vector error correction model (VECM) reduced-form as

$$\Delta y_t = \nu + \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

where α and β are $K \times r$ matrices of rank r . If there are no over-identifying restrictions on the VECM parameters, the Johansen reduced rank regression/Gaussian ML approach can be used to estimate the parameters of this reduced form (Johansen, 1995; Lütkepohl, 2005, Chapter 7).

Denoting the corresponding estimated residuals by \hat{u}_t , the structural parameters can again be estimated from $\mathbf{A}\hat{u}_t = \tilde{w}_t$ in the A-model, $\hat{u}_t = \mathbf{B}\tilde{w}_t$ in the B-model, and $\mathbf{A}\hat{u}_t = \mathbf{B}\tilde{w}_t$ in the AB-model. Thus, if there are no restrictions on the reduced-form slope parameters A or the VECM parameters, the structural parameters can often be estimated using the relevant candidate of the three foregoing systems of equations. Although estimating the structural parameters in this way opens up useful algorithms for computing point estimates, it may not be equally useful for evaluating asymptotic variances or covariance matrices. As mentioned earlier, in structural VAR analysis having simple computational methods for getting point estimates is important because the structural parameters are often only of interest for computing impulse responses and other quantities for inference.

The preferred estimation method depends not only on the model setup but also on the identifying restrictions. In the next section alternative sets of restrictions are presented which are important in this context. The corresponding estimation methods are discussed in Sections 4–7.

3. Types of restrictions

It is recommended to use a model setup that facilitates imposing the identifying assumptions in the form of linear restrictions. These are the most popular restrictions in practice. Important examples are exclusion restrictions on the impact effects of the structural shocks or on their long-run effects.

3.1. Restrictions on the impact effects of shocks

Recursive models are perhaps the most common structural VAR models identified by short-run restrictions on the impact effects of the structural shocks. They amount to restricting \mathbf{A} or \mathbf{B} to be upper or lower triangular. Recursive models are typically just-identified although there can also be additional over-identifying restrictions. Whether they are set up as \mathbf{A} or \mathbf{B} models does not matter for just-identified recursive models because the inverse of a triangular matrix is also triangular. If a triangular matrix \mathbf{A} is estimated, an estimate for \mathbf{B} can be constructed from the inverse of the estimate of \mathbf{A} and vice versa.

A recursive setup is also often used in partially identified models where, for example, only one economic relation is identified or only one structural shock is of interest. In that case only one row of \mathbf{A} or one column of \mathbf{B} is identified. Since just-identifying restrictions on other equations do not affect the identified equation, it is often possible to impose recursive restrictions on the model and thereby fully identify \mathbf{A} or \mathbf{B} from a statistical point of view, meaning that the parameters are not necessarily meaningful for economic interpretation but can be estimated consistently. For example, in a three-variable system consisting of the log of gross national product (gnp_t), an inflation rate (π_t)

and a short-term interest rate (r_t), one may draw on a Taylor rule relation to postulate that gnp_t and π_t may have an instantaneous effect on r_t , while interest rate changes have only a delayed effect on the former two variables. Ignoring lags, a structural system of the form

$$\begin{bmatrix} 1 & a_{12} & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{pmatrix} \text{gnp}_t \\ \pi_t \\ r_t \end{pmatrix} = w_t \quad (3.1)$$

can be set up in this case. The first two equations in this system are clearly not identified but can be identified by setting $a_{12} = 0$ which results in a recursive model. Such a restriction is arbitrary if no economic reasoning suggests it. Still, if the first two equations are not of economic interest, the restriction can be used to ensure statistical identification of the model. Given that recursive models are particularly easy to estimate, this type of identification restrictions is quite common in practice (e.g., Christiano *et al.*, 1999; Kilian and Lütkepohl, 2017, Chapter 8).

For non-recursive models or over-identified structural VAR models it is important whether linear restrictions are formulated for **A** or **B**. Clearly, linear restrictions for **A** may translate into nonlinear restrictions for its inverse and linear restrictions for **B** may correspond to nonlinear restrictions for **A**. Working with linear restrictions is recommended from a computational point of view. Therefore it is useful to choose a model setup that accommodates the restrictions in linear form.

Generally linear restrictions on **A** can be written as

$$\text{vec}(\mathbf{A}) = R\gamma + \mathbf{r},$$

where R is a known $K^2 \times J$ matrix, γ is the $J \times 1$ vector of unrestricted elements of **A** and \mathbf{r} is a known $K^2 \times 1$ vector. In the example (3.1) with $a_{12} = 0$, the elements on the main diagonal of **A** are restricted to be one such that

$$\begin{aligned} \mathbf{r} &= (1, 0, 0, 0, 1, 0, 0, 0, 1)', \\ \gamma &= (a_{21}, a_{31}, a_{32})' \end{aligned}$$

and

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}'.$$

Alternatively, there may be linear restrictions on **B** that can be written as

$$\text{vec}(\mathbf{B}) = R\gamma + \mathbf{r}, \quad (3.2)$$

where $\mathbf{r} = 0$, if there are just exclusion restrictions. Sometimes it is preferable to write the restrictions in the form

$$Q \text{vec}(\mathbf{B}) = \mathbf{q}, \quad (3.3)$$

where Q is a known $J \times K^2$ restriction matrix and \mathbf{q} is a known $J \times 1$ vector. Likewise linear restrictions for **A** could be represented in this form. Whether the restrictions are written as in (3.2) or in (3.3) is often a matter of convenience and depends to some extent on the estimation method used. Of course, there may be analogous restrictions on both **A** and **B** in the AB-model.

3.2. Restrictions on the long-run effects of shocks

Structural identification in VAR models is often achieved by imposing restrictions on the long-run effects of the shocks. Since the long-run effects of the shocks in a stable, stationary VAR model are zero, the long-run effects considered in the context of stable models are the accumulated effects of the shocks. Defining $A(1) \equiv I_K - A_1 - \dots - A_p$, the relevant long-run effects or long-run multiplier matrix is

$$\Xi = A(1)^{-1} \mathbf{A}^{-1} \mathbf{B} \quad (3.4)$$

for the AB-model and analogously for the A-model and the B-model. It is more common in this context to use either the A-model or the B-model. In particular, the B-model with long-run multiplier matrix $\Xi = A(1)^{-1} \mathbf{B}$ is popular in practice (see Kilian and Lütkepohl, 2017, Chapter 10, for examples and references).

Typical restrictions in this setup are exclusion restrictions. The most common set of restrictions constrains Ξ to be a triangular matrix (e.g., Blanchard and Quah, 1989). More generally, for the B-model, linear restrictions on Ξ can be written as

$$Q' \text{vec}(\Xi) = Q' (A(1)^{-1'} \otimes I_K) \text{vec}(\mathbf{B}) = \mathbf{q}',$$

where Q' is a suitable known restriction matrix and \mathbf{q}' is a suitable vector. Thus, given the reduced-form parameters, the implied restrictions for the structural parameters \mathbf{B} are linear restrictions with restriction matrix

$$\mathbf{Q} = Q' (A(1)^{-1'} \otimes I_K). \quad (3.5)$$

It is easy to combine these restrictions with further linear restrictions on the impact effects.

In a model with integrated and cointegrated variables, a shock can have permanent effects on the variables. Hence, considering such effects for identification purposes makes sense. The long-run effects matrix for the B-model in this case is

$$\Xi = \beta_{\perp} \left[\alpha'_{\perp} \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_{\perp} \right]^{-1} \alpha'_{\perp} \mathbf{B},$$

where β_{\perp} and α_{\perp} are orthogonal complements of β and α , respectively (e.g., King *et al.*, 1991; Lütkepohl, 2005, Chapter 9). For the A- and AB-model, \mathbf{B} has to be replaced by \mathbf{A}^{-1} or by $\mathbf{A}^{-1} \mathbf{B}$, respectively. It is noteworthy that the long-run effects matrix has rank $K - r$ and is hence a reduced rank matrix if there are $r > 0$ linearly independent cointegration relations. The k^{th} column of Ξ represents the long-run effects of the k^{th} structural shock, w_{kt} , on the components of y_t .

Again linear restrictions can be imposed easily on Ξ . Since α_{\perp} , β_{\perp} , and the Γ_i are determined by the reduced form, such restrictions amount to linear restrictions on \mathbf{B} if the B-model is used. Typical restrictions on Ξ are exclusion restrictions which amount to imposing that certain shocks do not have a long-run or permanent effect on some of the variables. If a whole column of Ξ is restricted to zero this means that the corresponding shock does not have any permanent effects on any of the variables and is hence transitory. A shock is called persistent if it has permanent effects on at least one of the variables.

Typically such long-run restrictions are complemented with linear restrictions directly on the impact effects. Such restrictions are in fact necessary for identifying all shocks properly if there is more than one purely transitory shock in which case such shocks can obviously not be distinguished by their long-run effects which are all zero.

4. Estimation of the A-model

All the frequentist estimators considered in the following are consistent and have asymptotic normal distributions under standard conditions. Many of them are even equally efficient asymptotically. However, as mentioned in the introduction, the structural parameters as such are often not of main interest but derived quantities such as the implied impulse responses and forecast error variance decompositions are of interest. Inference for these quantities is typically based on bootstrap methods for which a large number of estimates has to be computed. Thus, it is essential that a computationally efficient and robust estimation method is used for the structural parameters. This survey is meant to direct the reader to suitable methods for a given setup. Asymptotic properties of the estimators and conditions for their validity can be found, for example, in Kilian and Lütkepohl (2017, Chapters 9 and 11).

4.1. Short-run restrictions

4.1.1. Just-identified models

If the model is just-identified and there are just exclusion restrictions, estimation of the A-model is particularly simple. A popular example is a recursive model where the structural matrix \mathbf{A} is lower triangular. In that case, \mathbf{A} and Σ_w can be estimated by MM using the relation $\Sigma_u = \mathbf{A}^{-1}\Sigma_w\mathbf{A}^{-1'}$. The MM estimator can be computed easily with the help of a Cholesky decomposition of the reduced-form error covariance matrix Σ_u . Typically Σ_u is estimated as

$$\tilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t' \quad \text{or} \quad \hat{\Sigma}_u = (T - Kp - 1)^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'.$$

The latter estimator uses a degrees-of-freedom adjustment and may be preferable in small samples. The estimator $\tilde{\Sigma}_u$ is the Gaussian ML estimator and is used repeatedly in the following. It should be understood that the asymptotic properties of the MM estimators remain unaffected if $\tilde{\Sigma}_u$ is replaced by $\hat{\Sigma}_u$.

The MM estimator of \mathbf{A} is

$$\hat{\mathbf{A}} = \text{diag}(\hat{d}_1, \dots, \hat{d}_K)^{-1} \text{chol}(\tilde{\Sigma}_u)^{-1},$$

where the \hat{d}_k , $k = 1, \dots, K$, are the diagonal elements of $\text{chol}(\tilde{\Sigma}_u)^{-1}$. The MM estimator for the matrix Σ_w is

$$\tilde{\Sigma}_w = \hat{\mathbf{A}} \tilde{\Sigma}_u \hat{\mathbf{A}}'.$$

Generally, if the A-model is just-identified and there are just exclusion restrictions for the off-diagonal elements of \mathbf{A} , simple estimation methods are available even if \mathbf{A} is not triangular. The k^{th} equation of the A-model can be written as

$$\hat{u}_{kt} = \hat{\mathbf{u}}_{kt}' a_k + \tilde{w}_{kt},$$

where $\hat{\mathbf{u}}_{kt}$ is the column vector of the elements of \hat{u}_t that appear on the right-hand side of the k^{th} equation of the model $\mathbf{A}\hat{u}_t = \tilde{w}_t$ and where a_k is the vector of associated structural parameters. For example, for a lower-triangular matrix \mathbf{A} , no parameters in the first equation have to be estimated and the ‘regressors’ for $k = 2, \dots, K$ are of the form $\hat{\mathbf{u}}_{kt}' = (\hat{u}_{1t}, \dots, \hat{u}_{k-1,t})$ with parameters $a_k = (-a_{k1}, \dots, -a_{k,k-1})'$.

Summarizing the model for $t = 1, \dots, T$ in matrix notation yields

$$\mathbf{x}_k = X_k a_k + \mathbf{w}_k, \quad (4.1)$$

where $\mathbf{x}_k \equiv (\hat{u}_{k1}, \dots, \hat{u}_{kT})'$, $X_k \equiv [\hat{\mathbf{u}}_{k1}, \dots, \hat{\mathbf{u}}_{kT}]'$ and $\mathbf{w}_k \equiv (\tilde{w}_{k1}, \dots, \tilde{w}_{kT})'$. The LS estimator is

$$\hat{a}_k^{\text{LS}} = (X_k' X_k)^{-1} X_k' \mathbf{x}_k. \quad (4.2)$$

This estimator is equivalent to the MM estimator based on solving the system of equations

$$\tilde{\Sigma}_u = \mathbf{A}^{-1} \text{diag}(\sigma_1^2, \dots, \sigma_K^2) \mathbf{A}^{-1'}$$

for the unrestricted elements of \mathbf{A} and $\sigma_1^2, \dots, \sigma_K^2$. It can also be interpreted as a GMM estimator based on the moment conditions $\mathbb{E}(\mathbf{u}'_{kt} w_{kt}) = 0$. Thus,

$$\hat{a}_k^{\text{LS}} = \hat{a}_k^{\text{MM}} = \hat{a}_k^{\text{GMM}}.$$

4.1.2. Over-identified models

LS estimation can also be used if there are over-identifying restrictions of the form

$$a_k = R_k \gamma_k, \quad (4.3)$$

where R_k is a matrix that relates the unrestricted parameters γ_k to the restricted parameter vector a_k . The vector γ_k can be estimated by LS from the regression

$$\mathbf{x}_k = X_k^\dagger \gamma_k + \mathbf{w}_k,$$

where $X_k^\dagger = X_k R_k$. Now there may be more instruments in X_k than parameters in the k^{th} equation. These instruments can be used to estimate the parameters γ_k possibly more efficiently by two-stage LS (2SLS) using the instruments $\hat{X}_k^\dagger \equiv X_k (X_k' X_k)^{-1} X_k' X_k^\dagger$,

$$\hat{\gamma}_k^{\text{IV}} = \hat{\gamma}_k^{\text{2SLS}} = (\hat{X}_k^{\dagger'} \hat{X}_k^\dagger)^{-1} \hat{X}_k^{\dagger'} \mathbf{x}_k. \quad (4.4)$$

This estimator can also be interpreted as a GMM estimator based on the moment conditions $\mathbb{E}(\mathbf{u}'_{kt} w_{kt}) = 0$ and the GMM objective function

$$\begin{aligned} J(\gamma_k) &= \left(\frac{1}{T} \mathbf{w}'_k X_k \right) \left(\frac{1}{T} X_k' X_k \right)^{-1} \left(\frac{1}{T} X_k' \mathbf{w}_k \right) \\ &= \left(\frac{1}{T} (\mathbf{x}_k - X_k^\dagger \gamma_k)' X_k \right) \left(\frac{1}{T} X_k' X_k \right)^{-1} \left(\frac{1}{T} X_k' (\mathbf{x}_k - X_k^\dagger \gamma_k) \right). \end{aligned} \quad (4.5)$$

In other words, if this GMM objective function is used for estimation,

$$\hat{\gamma}_k^{\text{GMM}} = \hat{\gamma}_k^{\text{2SLS}}.$$

So far only the estimation of the k^{th} equation of the structural VAR model has been discussed. Since the components of w_t are instantaneously uncorrelated, single-equation GMM is identical to estimating the full system

$$\begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{pmatrix} = \begin{bmatrix} X_1^\dagger & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_K^\dagger \end{bmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_K \end{pmatrix} + \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{pmatrix}$$

by GMM. Note that there may be equations without any parameters to estimate like in the recursive model. Such equations are dropped from the system. The full system has to be estimated jointly, if there are cross-equations restrictions.

It is worth emphasizing again that concentrating out the parameters of the lagged variables and working with the estimated reduced-form residuals instead of the observations is not the same as the standard GMM framework for estimating the structural model including the lagged values. If standard GMM software is used, it is preferable to work with the original observations and include the lagged values in the equations. The same formulas can be used in that case but the symbols have to be adjusted properly. For example, \mathbf{x}_k has to include the original observations, i.e., $\mathbf{x}_k = (y_{k1}, \dots, y_{kT})'$ instead of LS residuals, X_k and X_k^\dagger are matrices that include the lagged observations in addition to unlagged observations, and the parameter vector γ_k has to include also all parameters associated with the lags of the observed variables etc. If such modifications are used, the standard output of GMM estimation for asymptotic covariance matrices, t -ratios etc. can be used.

Estimates of the elements of \mathbf{A} are obtained from the relation $a_k = R_k \gamma_k$ by replacing γ_k with an estimator. The diagonal elements of Σ_w can be estimated in the usual way from the estimated structural residuals as

$$\hat{\sigma}_k^2 = \frac{1}{T} \sum_{t=1}^T \hat{w}_{kt}^2.$$

4.2. Other restrictions and estimation approaches

4.2.1. Just-identifying restrictions

Suppose there are just-identifying restrictions of the form

$$\mathbf{Q} \text{vec}(\mathbf{A}) = \mathbf{q},$$

where this set of equations may contain short-run and long-run restrictions. In other words, \mathbf{Q} may contain reduced-form parameters as in (3.5). In that case, we denote by $\tilde{\mathbf{Q}}$ the matrix obtained by replacing all reduced-form parameters by LS estimators for the levels VAR model or by Gaussian ML estimators for VECMs.

For this type of restrictions, MM estimation is the preferred method for just-identified models. Since

$$\mathbf{A} \Sigma_u \mathbf{A}' = \text{diag}(\sigma_1^2, \dots, \sigma_K^2),$$

Σ_u is replaced by $\tilde{\Sigma}_u$ and the system of equations

$$\begin{bmatrix} \text{vech}(\mathbf{A} \tilde{\Sigma}_u \mathbf{A}') \\ \tilde{\mathbf{Q}} \text{vec}(\mathbf{A}) \end{bmatrix} = \begin{bmatrix} \text{vech}(\text{diag}(\sigma_1^2, \dots, \sigma_K^2)) \\ \mathbf{q} \end{bmatrix}$$

is solved by a nonlinear equations solver. It can be shown that the resulting estimates $\tilde{\mathbf{A}}$ and $\tilde{\Sigma}_w = \text{diag}(\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_K^2)$ are identical to the Gaussian ML estimates.

4.2.2. Gaussian maximum likelihood estimation

If there are over-identifying restrictions, Gaussian ML estimates of the structural parameters are obtained by maximizing the concentrated log-likelihood as a function of the structural parameters,

$$\log l_c(\mathbf{A}, \Sigma_w) = \text{constant} + \frac{T}{2} \log(\det \mathbf{A})^2 - \frac{T}{2} \log \det(\Sigma_w) - \frac{T}{2} \text{tr}(\mathbf{A}' \Sigma_w^{-1} \mathbf{A} \tilde{\Sigma}_u), \quad (4.6)$$

where $\tilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$ is the usual ML estimator of the reduced-form residual covariance matrix (Lütkepohl, 2005, Chapter 9). The log-likelihood function has to be maximized with respect to \mathbf{A} and $\Sigma_w = \text{diag}(\sigma_1^2, \dots, \sigma_K^2)$ subject to the identifying restrictions which may require numerical optimization methods. If the restrictions are just-identifying, the estimator is the same as the MM estimator discussed earlier.

4.2.3. Instrumental variable estimation of structural vector error correction models

Pagan and Pesaran (2008) observe that in a structural VECM the cointegration relations can serve as instruments under certain conditions. They show that a VECM with r ($0 < r < K$) linearly independent cointegration relations, r structural shocks with transitory effects only and $K-r$ persistent shocks with permanent effects, can be arranged such that

$$\mathbf{A}^p \Delta y_t = \mathbf{\Gamma}_1^{\dagger p} \Delta y_{t-1} + \dots + \mathbf{\Gamma}_{p-1}^{\dagger p} \Delta y_{t-p+1} + w_t^p, \quad (4.7)$$

$$\mathbf{A}^{tr} \Delta y_t = \alpha_{(r \times r)}^{\dagger} \beta' y_{t-1} + \mathbf{\Gamma}_1^{\dagger tr} \Delta y_{t-1} + \dots + \mathbf{\Gamma}_{p-1}^{\dagger tr} \Delta y_{t-p+1} + w_t^{tr}, \quad (4.8)$$

where the intercept has been deleted for simplicity, w_t^p is a $(K-r)$ -dimensional vector of persistent shocks and w_t^{tr} is an r -dimensional vector of transitory shocks. In this model setup, the cointegration relations, $\beta' y_{t-1}$, do not appear in the first $K-r$ structural equations of the system. If the r cointegration relations $\beta' y_{t-1}$ are known, they can be used as instruments in the first set of equations and w_t^p can be used as a vector of instruments for \mathbf{A}^{tr} in the second set of equations because w_t^p is uncorrelated with w_t^{tr} .

Since \mathbf{A}^p has up to $(K-1)(K-r)$ unknown elements to be estimated, the r cointegration relations may not provide enough instruments for consistent estimation of all the structural parameters. In that case the instruments have to be complemented with other identifying information such as exclusion restrictions for the impact effects of the structural shocks.

Kilian and Lütkepohl (2017, Section 11.2) emphasize that this IV method for estimating the structural parameters in the presence of long-run restrictions requires that (1) the number of transitory shocks is equal to the cointegrating rank, r , of the system and the number of persistent shocks is equal to the number of common trends, $K-r$; (2) the cointegration relations are known. In any case, if the instruments are insufficient for estimating all structural parameters, identifying information from other sources is needed.

5. Estimation of the B-model

5.1. Just-identified models

If the B-model is just-identified, a MM approach to estimating \mathbf{B} is recommended. That approach is based on the covariance matrix of the reduced-form VAR residuals, Σ_u , which may be expressed in terms of the structural model parameters as

$$\Sigma_u = \mathbf{B}\mathbf{B}'. \quad (5.1)$$

Replacing Σ_u by a consistent estimate $\tilde{\Sigma}_u$ or $\hat{\Sigma}_u$ and solving the system of equations (5.1) subject to the identifying restrictions gives an estimate of \mathbf{B} . For the A-model, MM estimation is typically considered only if the model is recursive or if numerical methods are needed for estimation and no simple closed form estimator exists. In contrast, for the B-model, in some cases computing the MM estimator is straightforward. In the following the focus is on such cases but the general case which involves nonlinear numerical techniques is also presented.

5.1.1. Recursive models

If \mathbf{B} is lower-triangular the easiest way of computing the MM estimator of the structural parameters involves estimating the reduced-form VAR parameters, computing the residual variance-covariance matrix $\hat{\Sigma}_u$ or $\tilde{\Sigma}_u$ and applying the Cholesky decomposition. In other words,

$$\hat{\mathbf{B}} = \text{chol}(\hat{\Sigma}_u) \quad \text{or} \quad \hat{\mathbf{B}} = \text{chol}(\tilde{\Sigma}_u).$$

If \mathbf{B} is upper triangular, the same approach can be used, but it is necessary to multiply the estimate of \mathbf{B} from the left and the right by the orthogonal matrix

$$G = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{bmatrix}$$

such that, for example,

$$\hat{\mathbf{B}} = G \text{chol}(\tilde{\Sigma}_u) G$$

(see, Lütkepohl, 2005, A.9.3). This MM estimator is also the Gaussian ML estimator if Σ_u is estimated by $\tilde{\Sigma}_u$.

Estimation by Cholesky decomposition is also possible if there are recursive long-run restrictions, i.e., if $\Xi = A(1)^{-1}\mathbf{B}$ is triangular. This case is quite common in practice (e.g., Blanchard and Quah, 1989). For simplicity we assume that Ξ is lower triangular. In that case, $\text{chol}(\hat{A}(1)^{-1}\tilde{\Sigma}_u\hat{A}(1)^{-1'})$ is an estimator of $A(1)^{-1}\mathbf{B}$ and

$$\hat{\mathbf{B}} = \hat{A}(1)\text{chol}(\hat{A}(1)^{-1}\tilde{\Sigma}_u\hat{A}(1)^{-1'})$$

is the MM estimator of the structural matrix \mathbf{B} .

It is important to note that this MM estimator involves the inverse of $\hat{A}(1) = I_K - \hat{A}_1 - \cdots - \hat{A}_p$ which may be ill-conditioned if some of the components of y_t are integrated and possibly cointegrated because in that case $A(1)$ is singular. Thus, the MM estimator should only be used if the underlying VAR process is stable. In fact, even in that case the estimates may not be accurate if there are very persistent variables with autoregressive roots near unity because in that case $A(1)$ will be near singular which may translate into large variances of the MM estimator. This problem has led some researchers to question the usefulness of long-run identifying restrictions (see Kilian and Lütkepohl, 2017, Section 11.4, for a discussion). Lütkepohl *et al.* (2017) show, however, that long-run restrictions can lead to more precise inference for impulse responses derived from the structural parameters than short-run restrictions, provided the autoregressive roots are not close to the unit circle.

5.1.2. Non-recursive models

If the restrictions are just-identifying but not recursive so that \mathbf{B} is not triangular, a nonlinear equations solver can be used to solve the system of equations

$$\hat{\Sigma}_u = \mathbf{B}\mathbf{B}' \quad \text{or} \quad \tilde{\Sigma}_u = \mathbf{B}\mathbf{B}'$$

for \mathbf{B} , subject to the identifying restrictions. For small systems with exclusion restrictions it may be possible to obtain a recursive system by rearranging the variables. Moreover, sometimes the solution

is easy to find without a nonlinear equations solver. In that case, such solutions may be preferable in terms of computing time. Rubio-Ramírez *et al.* (2010) present a non-iterative algorithm for just-identified models which can be used in this context and which may have computational advantages.

5.2. Over-identified models

5.2.1. Generalized method-of-moments estimation

If over-identifying restrictions are available for \mathbf{B} , GMM estimation can be used. In that case, no \mathbf{B} matrix exists that satisfies all identifying restrictions and also the relations $\tilde{\Sigma}_u = \mathbf{B}\mathbf{B}'$. Therefore the objective is to find the structural parameter values that minimize a weighted average of the moment conditions where the asymptotically optimal set of weights corresponds to the inverse of the variance-covariance matrix of the sample moment conditions.

A set of moment conditions for estimating \mathbf{B} is

$$\mathbb{E}(\text{vech}(u_t u_t') - \text{vech}(\mathbf{B}\mathbf{B}')) = 0$$

for which the empirical counterpart is

$$\text{vech}\left(\frac{1}{T} \sum_{t=1}^T \hat{u}_t \hat{u}_t' - \mathbf{B}\mathbf{B}'\right) = \text{vech}(\tilde{\Sigma}_u) - \text{vech}(\mathbf{B}\mathbf{B}').$$

Defining

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T (\text{vech}(\hat{u}_t \hat{u}_t') - \overline{\text{vech}(\hat{u}\hat{u}')})(\text{vech}(\hat{u}_t \hat{u}_t') - \overline{\text{vech}(\hat{u}\hat{u}')}')$$

with

$$\overline{\text{vech}(\hat{u}\hat{u}')} = \frac{1}{T} \sum_{t=1}^T \text{vech}(\hat{u}_t \hat{u}_t'),$$

the GMM estimator for \mathbf{B} chooses the unknown elements of \mathbf{B} so as to minimize

$$J = T (\text{vech}(\tilde{\Sigma}_u) - \text{vech}(\mathbf{B}\mathbf{B}'))' \hat{\Omega}^{-1} (\text{vech}(\tilde{\Sigma}_u) - \text{vech}(\mathbf{B}\mathbf{B}'))$$

subject to all identifying and over-identifying restrictions. Typically numerical algorithms have to be used for minimizing J subject to over-identifying restrictions.

5.2.2. Gaussian maximum likelihood estimation

As mentioned earlier, the MM estimator is equivalent to Gaussian ML estimation if $\tilde{\Sigma}_u$ is used as estimator for Σ_u and the model is just-identified. For an over-identified B-model, concentrating on the structural parameters, the concentrated log-likelihood function is

$$\log l_c(\mathbf{B}) = \text{constant} - \frac{T}{2} \log(\det \mathbf{B})^2 - \frac{T}{2} \text{tr}(\mathbf{B}^{-1'} \mathbf{B}^{-1} \tilde{\Sigma}_u). \quad (5.2)$$

In general the concentrated log-likelihood function can be maximized by a numerical optimization algorithm with respect to \mathbf{B} subject to the identifying restrictions.

If y_t is a stationary Gaussian VAR(p) process, the ML estimation framework also facilitates the derivation of the asymptotic properties of the estimator. General ML theory implies that the unrestricted parameters are consistent and asymptotically normal. If there are over-identifying restrictions on \mathbf{B} , the restricted ML estimator of Σ_u is

$$\tilde{\Sigma}_u^r \equiv \tilde{\mathbf{B}}\tilde{\mathbf{B}}', \quad (5.3)$$

where $\tilde{\mathbf{B}}$ is the restricted ML estimator of \mathbf{B} satisfying all identifying restrictions.

6. Estimation of the AB-model

6.1. Just-identified models

MM estimation can also be used for just-identified AB-models. The moment conditions are

$$\Sigma_u = \mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}^{-1'}$$

Given identifying restrictions

$$\mathbf{Q} \begin{bmatrix} \text{vec}(\mathbf{A}) \\ \text{vec}(\mathbf{B}) \end{bmatrix} = \mathbf{q},$$

the system of equations to be solved for \mathbf{A} and \mathbf{B} is

$$\begin{bmatrix} \text{vech}(\mathbf{A}\tilde{\Sigma}_u\mathbf{A}') \\ \tilde{\mathbf{Q}} \begin{bmatrix} \text{vec}(\mathbf{A}) \\ \text{vec}(\mathbf{B}) \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \text{vech}(\mathbf{B}\mathbf{B}') \\ \mathbf{q} \end{bmatrix},$$

where $\tilde{\mathbf{Q}}$ is obtained by replacing all reduced-form parameters in \mathbf{Q} by their ML/LS estimates. Solving this system of equations for \mathbf{A} and \mathbf{B} typically requires the use of a nonlinear equations solver. Under normality assumptions and if reduced-form estimation is done by Gaussian ML, then the resulting estimates are also ML estimates. This result shows the good asymptotic properties of MM estimators under standard assumptions.

6.2. Over-identified models

For over-identified models, GMM or Gaussian ML estimation of the AB-model may be used.

6.2.1. Generalized method-of-moments estimation

GMM estimation of the AB-model is a straightforward extension of GMM estimation of the B-model. If there are identifying restrictions on both \mathbf{A} and \mathbf{B} in an over-identified structural VAR model, the GMM estimators are obtained by minimizing the objective function

$$J = T \text{vech}(\tilde{\Sigma}_u - \mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}^{-1'})' \hat{\Omega}^{-1} \text{vech}(\tilde{\Sigma}_u - \mathbf{A}^{-1}\mathbf{B}\mathbf{B}'\mathbf{A}^{-1'})$$

subject to all restrictions on both \mathbf{A} and \mathbf{B} .

6.2.2. Gaussian maximum likelihood estimation

If there are no restrictions on the reduced-form parameters, Gaussian ML estimation is again executed by concentrating on the structural parameters. The relevant concentrated log-likelihood function is

$$\log l_c(\mathbf{A}, \mathbf{B}) = \text{constant} + \frac{T}{2} \log(\det \mathbf{A})^2 - \frac{T}{2} \log(\det \mathbf{B})^2 - \frac{T}{2} \text{tr}(\mathbf{A}' \mathbf{B}^{-1'} \mathbf{B}^{-1} \mathbf{A} \tilde{\Sigma}_u) \quad (6.1)$$

(see Lütkepohl, 2005, Chapter 9).

In general the concentrated log-likelihood function can be maximized by a numerical optimization algorithm with respect to \mathbf{A} and \mathbf{B} , subject to the identifying restrictions. For just-identified models it can be shown that the ML estimates satisfy

$$\tilde{\mathbf{A}}^{-1} \tilde{\mathbf{B}} \tilde{\mathbf{B}}' \tilde{\mathbf{A}}^{-1'} = \tilde{\Sigma}_u.$$

In other words, ML estimation and MM estimation provide identical results if the same estimator for Σ_u is used. If there are over-identifying restrictions on the structural parameters, the reduced-form error covariance matrix Σ_u may be estimated as

$$\tilde{\Sigma}_u' \equiv \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{B}} \tilde{\mathbf{B}}' \tilde{\mathbf{A}}^{-1'}. \quad (6.2)$$

7. Bayesian estimation

In practice, structural VAR analysis is often based on Bayesian estimation methods. Bayesian estimators are obtained by evaluating the posterior distribution of the parameters of interest or functions of the structural parameters. The first step in constructing the posterior distribution of the structural parameters is the specification of a prior. In the framework of structural VAR analysis the prior is either imposed on the reduced-form parameters or on the structural parameters. The latter approach is more plausible if the structural parameters are of main interest and prior beliefs are available for the structural parameters. Priors imposed on the reduced-form parameters often reflect the limited structural knowledge of the investigator. They are formulated with the objective to obtain posterior distributions which are easy to sample from and to induce little distortion for the structural analysis.

Gaussian-inverse Wishart priors are conventional priors for reduced-form VAR models. The mean of the prior for the VAR coefficients is typically specified as for a so-called Minnesota prior which shrinks the VAR slope parameters to zero or a random walk process depending on the persistence properties of the data (Kilian and Lütkepohl, 2017, Chapter 5).

Suppose that the VAR(p) model (2.1) has Gaussian errors, $u_t \sim \mathcal{N}(0, \Sigma_u)$, and the priors for A and Σ_u are

$$\text{vec}(A) | \Sigma_u \sim \mathcal{N}(\text{vec}(A^*), V_A = V \otimes \Sigma_u) \quad (7.1)$$

and

$$\Sigma_u \sim \mathcal{IW}_K(S_*, n), \quad (7.2)$$

where $\mathcal{IW}_K(S_*, n)$ denotes an inverse Wishart distribution with $K \times K$ matrix parameter S_* and degrees-of-freedom parameter n . Using the covariance matrix Kronecker product $V \otimes \Sigma_u$ in the conditional prior for A simplifies the posterior distribution which turns out to be also a Gaussian-inverse Wishart distribution,

$$\text{vec}(A) | \Sigma_u, \mathbf{y} \sim \mathcal{N}(\text{vec}(\bar{A}), \bar{\Sigma}_A), \quad \Sigma_u | \mathbf{y} \sim \mathcal{IW}_K(S, \tau), \quad (7.3)$$

where $\mathbf{y} = \text{vec}(y_1, \dots, y_T)$ denotes the data,

$$\bar{A} = (A^*V^{-1} + YZ')(V^{-1} + ZZ')^{-1} \quad (7.4)$$

and

$$\bar{\Sigma}_A = (V^{-1} + ZZ')^{-1} \otimes \Sigma_u.$$

Here $Z \equiv [Y_0, \dots, Y_{T-1}]$ and $Y \equiv [y_1, \dots, y_T]$. The parameters of the inverse Wishart distribution in (7.3) are

$$S = T\bar{\Sigma}_u + S_* + \hat{A}ZZ'\hat{A}' + A^*V^{-1}A^{*'} - \bar{A}(V^{-1} + ZZ')\bar{A}' \quad (7.5)$$

and $\tau = T + n$ (Koop and Korobilis, 2010; Kilian and Lütkepohl, 2017, Section 5.2.4; Uhlig, 1994, 2005). Here A^* and \bar{A} are $K \times (Kp + 1)$ matrices and $\hat{A} = YZ'(ZZ')^{-1}$.

Since the posterior is from the same distributional family as the prior, the prior (7.1)/(7.2) is a conjugate prior. Given that the prior is also from the same distributional family as the likelihood, it is even a natural conjugate prior. The question is, of course, how to choose the prior parameters A^* , V , S_* and n . As mentioned before, the prior mean A^* is often chosen as in the Minnesota prior, i.e., A^* is chosen such that the corresponding model for persistent variables is a random walk and for non-persistent variables a white noise process (see Kilian and Lütkepohl, 2017, Section 5.2.3, for the Minnesota and other priors). The variance related prior parameters V , S_* and n are chosen such that the prior has a limited impact on the posterior. In other words, very large or even infinite variances are chosen. If infinite variances are chosen, V^{-1} is replaced by a zero matrix in the posterior distributions and the prior is improper.

Using a prior that gives rise to a known form of the posterior distribution of the reduced-form parameters is convenient because it makes it easy to draw samples from the reduced-form posterior. For just-identified models, reduced-form posterior draws can then be transformed to draws of structural parameters (e.g., Canova, 1991; Gordon and Leeper, 1994). For example, for a recursive identification scheme, a Cholesky decomposition of the i^{th} draw of Σ_u yields the i^{th} posterior draw for \mathbf{B} .

Given that the prior in this approach is not specified for the structural parameters which are supposedly the parameters of interest, this approach is unsatisfactory. It is even more problematic if there are over-identifying restrictions for \mathbf{A} or \mathbf{B} because they would be ignored in a general purpose prior for the reduced form. Sims and Zha (1998, 1999) have developed an approach for imposing priors directly on the structural parameters. The approach is specifically designed for structural VAR models with linear restrictions on \mathbf{A} . There are no obvious extensions for \mathbf{B} -models. A generalization for certain nonlinear restrictions on \mathbf{A} was recently proposed by Canova and Pérez Forero (2015).

Canova (2007, Section 10.3) compares the structural priors discussed in Sims and Zha (1998) with more traditional Minnesota priors on the reduced-form parameters and points out that there are important differences. Thus, imposing the prior on the structural parameters can make a difference for the posterior of the structural parameters.

8. Illustrations

In this section some illustrative examples for estimating structural VAR models are provided. The IS-LM example data from Breitung *et al.* (2004) are used to illustrate the estimation of the structural parameters under alternative sets of restrictions. The data consists of three seasonally adjusted, quarterly

Table 1: Structural restrictions and frequentist estimation methods

| Restrictions | Estimation method |
|---|----------------------------|
| Recursive model (triangular A or B) | Cholesky decomposition, MM |
| Just-identifying linear restrictions on A | LS |
| Over-identifying linear restrictions on A | IV, 2SLS, GMM |
| Just-identifying linear restrictions on B | MM |
| Over-identifying linear restrictions on B | GMM, ML |
| Just-identifying linear restrictions on A and B | MM |
| Over-identifying linear restrictions on A and B | GMM, ML |

MM = method-of-moments; LS = least-squares; IV = instrumental variables; 2SLS = two-stage LS;
GMM = generalized method-of-moments; ML = maximum likelihood.

U.S. time series for log real GDP (q_t), the log of the real monetary base (m_t) and a three-months inter-bank interest rate (r_t) for the period 1970Q1–1997Q4. (The data is available at <http://www.jmulti.de>. Further details on the data sources are given by Breitung *et al.* (2004).) Thus, $y_t = (q_t, m_t, r_t)'$ is three-dimensional.

A VAR(4) model is fitted to the data, as in Breitung *et al.* (2004), and various alternative sets of identifying restrictions for the structural parameters are used. Some of these restrictions are not plausible from an economic point of view but are just chosen to illustrate specific identification restrictions and the related estimation methods for the structural parameters.

As discussed in the previous sections, the choice of estimation method depends on the type of model and restrictions. The frequentist estimation methods most suitable for specific model types and restrictions are summarized in Table 1. To illustrate the methods, the following three alternative A-models with **A** matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & * & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ * & 1 & * \\ * & 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & * \\ * & 0 & 1 \end{bmatrix}$$

are considered. They all have diagonal structural covariance matrix

$$\Sigma_w = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}.$$

The asterisks denote unrestricted elements while 0 and 1 stand for restricted parameters. The first model is recursive. The second identification scheme is not recursive but just-identified and the third scheme is over-identified with just two structural parameters to be estimated in the **A** matrix.

Based on Table 1, MM via Cholesky decomposition is the recommended estimation method for the recursive model. The resulting estimates are presented in Table 2. Equivalently, the unknown elements of **A** can be estimated by applying LS to the two equations

$$m_t = v_2 - a_{21}q_t + a_2^*Y_{t-1} + w_{2t}$$

and

$$r_t = v_3 - a_{31}q_t - a_{32}m_t + a_3^*Y_{t-1} + w_{3t},$$

where a_k^* denotes the k^{th} row of **AA**, as before.

Table 2: Estimated A-models

| Model | $\hat{\mathbf{A}}$ | $\hat{\Sigma}_w$ |
|-----------------|--|---|
| Recursive | $\begin{bmatrix} 1 & 0 & 0 \\ -0.0458 & 1 & 0 \\ -0.0502 & 0.8822 & 1 \end{bmatrix}$ | $\begin{bmatrix} 0.4614 & 0 & 0 \\ 0 & 0.2980 & 0 \\ 0 & 0 & 0.7738 \end{bmatrix} \times 10^{-4}$ |
| Nonrecursive | $\begin{bmatrix} 1 & 0 & 0 \\ -0.0483 & 1 & 0.2614 \\ -0.0098 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} 0.4614 & 0 & 0 \\ 0 & 0.2293 & 0 \\ 0 & 0 & 1.0058 \end{bmatrix} \times 10^{-4}$ |
| Over-identified | $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.2612 \\ -0.0098 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} 0.4614 & 0 & 0 \\ 0 & 0.2304 & 0 \\ 0 & 0 & 1.0058 \end{bmatrix} \times 10^{-4}$ |

Computations were done with suitably adjusted Matlab code provided at <http://www-personal.umich.edu/~lkilian/book.html>.

Table 3: Estimated B-models

| | |
|-----------------------|--|
| Recursive model | $\begin{bmatrix} 0.0072 & 0 & 0 \\ 0.0003 & 0.0058 & 0 \\ 0.0001 & -0.0051 & 0.0094 \end{bmatrix}$ |
| Nonrecursive model | $\begin{bmatrix} 0.0072 & 0.0004 & 0 \\ 0 & 0.0058 & 0 \\ 0.0004 & -0.0051 & 0.0094 \end{bmatrix}$ |
| Over-identified model | $\begin{bmatrix} 0.0072 & 0.0006 & 0 \\ 0 & 0.0058 & 0 \\ 0.0005 & 0 & 0.0107 \end{bmatrix}$ |

Computations were done with JMulTi (see <http://www.jmulti.de>).

LS is also the recommended method for estimating the nonrecursive just-identified model. In other words, LS is applied to the two equations

$$m_t = v_2 - a_{21}q_t - a_{23}r_t + a_2^*Y_{t-1} + w_{2t}$$

and

$$r_t = v_3 - a_{31}q_t + a_3^*Y_{t-1} + w_{3t}.$$

The resulting estimates are also presented in Table 2.

Finally, 2SLS may be considered for estimating the second equation of the over-identified A-model,

$$m_t = v_2 - a_{23}r_t + a_2^*Y_{t-1} + w_{2t}.$$

The estimates are also given in Table 2, where the third equation of the over-identified model is estimated by LS.

To illustrate the estimation of B-models, three such models with **B** matrices

$$\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}, \quad \begin{bmatrix} * & * & 0 \\ 0 & * & 0 \\ * & * & * \end{bmatrix}, \quad \begin{bmatrix} * & * & 0 \\ 0 & * & 0 \\ * & 0 & * \end{bmatrix}$$

and $\Sigma_w = I_3$ are considered. The first model has a lower-triangular **B** matrix and is, hence, recursive. The second scheme belongs to a just-identified but nonrecursive model and the third model is over-identified. The recommended estimation method for the first model is by a Cholesky decomposition

Table 4: Estimated AB-model

| $\hat{\mathbf{A}}$ | $\hat{\mathbf{B}}$ |
|--|--|
| $\begin{bmatrix} 1 & 0 & 0 \\ -0.0483 & 1 & 0.2614 \\ -0.0098 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} 0.0068 & 0 & 0 \\ 0 & 0.0048 & 0 \\ 0 & 0 & 0.0100 \end{bmatrix}$ |

Computations were done with suitably adjusted Matlab programs provided at <http://www-personal.umich.edu/~lkilian/book.html>.

of the estimated reduced-form covariance matrix. The second model is easily estimated by MM estimation, that is, by a suitable decomposition of the estimated reduced-form residual covariance matrix and the third model can be estimated by Gaussian ML. The three resulting estimates are given in Table 3. They are all obtained by the software JMulTi which is a powerful tool for structural VAR estimation (Krätzig, 2004).

To illustrate an AB-model, I note that an A-model can be cast in AB form by specifying the \mathbf{B} matrix as a diagonal matrix and standardizing the innovation covariance matrix to be an identity matrix, $\Sigma_w = I_K$. For example, the nonrecursive just-identified A-model can be parameterized as

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ * & 1 & * \\ * & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{bmatrix}.$$

MM estimation is a suitable method for this type of model (Table 1). It is equivalent to Gaussian ML. The estimates for the example model are given in Table 4.

Finally, to illustrate the estimation of a structural VAR model in the presence of long-run restrictions, consider a B-model with lower-triangular long-run multiplier matrix Ξ ,

$$\Xi = A(1)^{-1}\mathbf{B} = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}.$$

Note that, in practice, for the example model such restrictions are not recommended because the variables may be integrated in which case $A(1)$ is not invertible (Breitung *et al.*, 2004). However, for illustrative purposes, since the LS reduced-form estimate of $A(1)$ is nonsingular, I proceed as if $A(1)$ were nonsingular and use the long-run restrictions to identify \mathbf{B} . The recursive restrictions suggest using an estimator based on the Cholesky decomposition of $\hat{A}(1)^{-1}\hat{\Sigma}_u\hat{A}(1)^{-1'}$. The resulting estimate of \mathbf{B} is

$$\hat{\mathbf{B}} = \hat{A}(1)\text{chol}\left(\hat{A}(1)^{-1}\hat{\Sigma}_u\hat{A}(1)^{-1'}\right) = \begin{bmatrix} -0.0018 & -0.0044 & 0.0054 \\ -0.0037 & 0.0039 & 0.0023 \\ 0.0101 & -0.0002 & 0.0034 \end{bmatrix}.$$

This estimate was also computed with the software JMulTi.

9. Conclusions and extensions

This study reviews methods for estimating structural VAR models set up as A-models, B-models or AB-models. A range of alternative estimation methods are considered. It is stressed that the identifying restrictions should be placed such that easy estimation methods can be used. More precisely, it is

useful if the restrictions can be imposed as linear restrictions on \mathbf{A} or \mathbf{B} or both of these matrices. The choice of estimation method depends on the type of model and restrictions.

Many estimators allow to concentrate out the reduced-form parameters and estimate the structural parameters from reduced-form residuals or the residual covariance matrix estimator directly. Gaussian ML and GMM are the most general estimation methods which can be used even if there are over-identifying restrictions on \mathbf{A} or \mathbf{B} . For the A-model the GMM estimator can be computed in closed form even in many situations when there are over-identifying restrictions. For just-identified models, MM estimation is often a suitable choice and for recursive models estimates can be computed using a Cholesky decomposition of a matrix based on reduced-form parameters.

It is also possible to estimate structural VAR models by Bayesian methods. The conventional approach of formulating a prior for the reduced-form parameters and generating posterior draws for the structural parameters \mathbf{A} and/or \mathbf{B} from the draws for Σ_u has the drawback that it does not allow for over-identifying restrictions. So far Bayesian methods that impose priors directly on the structural parameters are primarily suitable for A-models.

There are a number of extensions of the basic model setup considered in this survey. First of all, linear identifying restrictions are emphasized because they facilitate estimation of the structural parameters. Some of the methods can in principle be adopted to nonlinear restrictions as well. For example, Gaussian ML and GMM estimation can be used in principle in conjunction with nonlinear restrictions. The computational challenges may become more burdensome for that case, however.

The basic model in this survey is a finite order VAR(p) model. Many of the methods can be justified even if the true VAR order is infinite and the finite order VAR model used for structural analysis is just an approximation to an infinite order process. The corresponding methods are discussed in the literature under the headline of sieve estimation (e.g., Kilian and Lütkepohl, 2017).

Another extension of the basic model setup allows for heteroskedasticity or conditional heteroskedasticity. Clearly, time-varying residual volatility is a common feature of financial data and therefore it also plays an important role in structural VAR analysis. Such features are in principle easy to deal with by using, for example, generalized LS methods rather than LS or by adjusting the likelihood function appropriately. Since the structural parameters are related to the residual covariance matrix, it is in fact possible to use a time-varying covariance structure for the identification of structural shocks. A broad literature addressing the topic of identification through heteroskedasticity has evolved following proposals by Rigobon (2003), Lanne and Lütkepohl (2008) and others. Recent surveys of the related literature are provided by Lütkepohl (2013) and Kilian and Lütkepohl (2017, Chapter 14).

This review focusses on point-identified structural VAR models. In practice set-identified models based on sign restrictions for the structural parameters or the derived impulse responses are often considered (see Uhlig (2005) for an important contribution to this literature and Kilian and Lütkepohl (2017, Chapter 13) for a recent survey). Although extensions of the setup discussed in this review to such models are not straightforward, some of the estimation algorithms presented in the current study are important building blocks of the related algorithms used in the sign restriction literature. Therefore being familiar with the methods discussed in the current study is useful.

References

- Amisano G and Giannini C (1997). *Topics in Structural VAR Econometrics* (2nd ed), Springer, Berlin.
Blanchard OJ and Quah D (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review*, **79**, 655–673.

- Breitung J, Brüggemann R, and Lütkepohl H (2004). Structural vector autoregressive modeling and impulse responses. In H. Lütkepohl and M. Krätzig (Eds), *Applied Time Series Econometrics* (pp. 159–196), Cambridge University Press, Cambridge.
- Canova F (1991). The sources of financial crisis: pre- and post-Fed evidence, *International Economic Review*, **32**, 689–713.
- Canova F (2007). *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton.
- Canova F and Pérez Forero FJ (2015). Estimating overidentified, nonrecursive, time-varying coefficients structural vector autoregressions, *Quantitative Economics*, **6**, 359–384.
- Christiano LJ, Eichenbaum M, and Evans C (1999). Monetary policy shocks: What have we learned and to what end? In JB Taylor and M Woodford (Eds), *Handbook of Macroeconomics* (pp. 65–148), Elsevier, Amsterdam.
- EViews (2000). *EViews 4.0 User's Guide*, Quantitative Micro Software, Irvine.
- Gordon DB and Leeper EM (1994). The dynamic impacts of monetary policy: an exercise in tentative identification, *Journal of Political Economy*, **102**, 1228–1247.
- Johansen S (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Kilian L and Lütkepohl H (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press, Cambridge.
- King RG, Plosser CI, Stock JH, and Watson MW (1991). Stochastic trends and economic fluctuations, *American Economic Review*, **81**, 819–840.
- Koop G and Korobilis D (2010). Bayesian multivariate time series methods for empirical macroeconomics, *Foundations and Trends in Econometrics*, **3**, 267–358.
- Krätzig M (2004). The software JMulTi. In H Lütkepohl and M Krätzig (Eds), *Applied Time Series Econometrics* (pp. 289–299), Cambridge University Press, Cambridge.
- Lanne M and Lütkepohl H (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking*, **40**, 1131–1149.
- Lütkepohl H (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Lütkepohl H (2013). Identifying structural vector autoregressions via changes in volatility. In TB Fomby, L Kilian and A Murphy (Eds), *Advances in Econometrics* (pp. 169–204), JAI Press, Greenwich.
- Lütkepohl H, Staszewska-Bystrova A, and Winker P (2017). Estimation of structural impulse responses: short-run versus long-run identifying restrictions, *Advances in Statistical Analysis*, 1–16. Retrieved from: <https://link.springer.com/article/10.1007/s10182-017-0300-9>
- Pagan AR and Pesaran MH (2008). Econometric analysis of structural systems with permanent and transitory shocks, *Journal of Economic Dynamics and Control*, **32**, 3376–3395.
- Rigobon R (2003). Identification through heteroskedasticity, *Review of Economics and Statistics*, **85**, 777–792.
- Rubio-Ramírez JF, Waggoner D, and Zha T (2010). Structural vector autoregressions: theory of identification and algorithms for inference, *Review of Economic Studies*, **77**, 665–696.
- Sims CA (1980). Macroeconomics and reality, *Econometrica*, **48**, 1–48.
- Sims CA and Zha T (1998). Bayesian methods for dynamic multivariate models, *International Economic Review*, **39**, 949–968.
- Sims CA and Zha T (1999). Error bands for impulse responses, *Econometrica*, **67**, 1113–1155.
- Uhlig H (1994). What macroeconomists should know about unit roots: a Bayesian perspective, *Econometric Theory*, **10**, 645–671.

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

Received August 1, 2017; Revised August 28, 2017; Accepted August 29, 2017