Far East Journal of Theoretical Statistics
Volume 40, Number 1, 2012, Pages 31-44
Published Online: August 2012
Available online at http://pphmj.com/journals/fjts.htm
Published by Pushpa Publishing House, Allahabad, INDIA

IDENTIFYING STRUCTURAL VAR MODEL WITH LATENT VARIABLES USING OVERCOMPLETE ICA

Wei Gao and Haizhong Yang

School of Statistics
Xi'an University of Finance and Economics
Xi'an, Shaanxi 710061, P. R. China
e-mail: gaoweiufe@yahoo.cn

Abstract

In this paper, we propose a method based on independent component analysis to represent the dependence structure of structural VAR model affected by latent variables in the case of non-Gaussian. The parameters of the model are estimated by a synthesis of least-squares method and independent component analysis. Based on the results of parameters estimation, the directed acyclic graph of structural VAR model with latent variables is constructed. Finally, simulation results demonstrate that the proposed method can correctly identify structural VAR model with latent variables.

I. Introduction

Since the 1980s, structural vector autoregressive (VAR) models have become a prevalent tool to empirically analyze dynamic general equilibrium models [1]. Recently, the graphical models have been introduced to model dependence structures among multivariate time series [2, 3]. The recursive

2010 Mathematics Subject Classification: 62-09.

Keywords and phrases: independent component analysis, structural vector autoregression, directed acyclic graph, latent variable.

Received April 13, 2012

structure of structural VAR can be represented by a directed acyclic graph (DAG) with each variable at a specific time represented by a separate vertex in the graph. Standard graph-theoretic techniques permit the researcher to infer the causal relations sufficient to identify the structural VAR [4-6]. One major problem in the application of graphical models is the possible presence of latent variables that affect the observed variables and thus lead to false dependence structure when we establish model for observed variables. In many cases, the data generating process might involve unobserved confounders. It is well known that the omission of important variables can lead to spurious correlations which are falsely detected as causal relationships among the observed variables. In recent years, several authors have proposed methods to model the causal dependence of observed variables when there exist the affections of latent variables. For multivariate time series, the study mainly focuses on identifying the dependence structure of observed variables that are affected by latent variables such as the researches of Eichler [7], Chu [8], and Gao and Tian [9].

Such methods require the user to make various assumptions about the data generating process. In much empirical work, it is additionally assumed that disturbances have a normal distribution. As has been extensively discussed elsewhere, these assumptions are quite strong and can often be violated in real data. Independent component analysis (ICA) is a statistical method used to find a linear representation of non-Gaussian data so that the components are as independent as possible. Such a representation would then capture only the essential structure of the data in these independent components [10, 11]. Shimizu et al. [12] proposed linear non-Gaussian acyclic model (LiNGAM) for the estimation of causal effects between the observed variables in the linear, non-Gaussian domain based on ICA. Hyvärinen et al. [13] applied the LiNGAM analysis to estimate a structural vector autoregressive (VAR) model based on non-Gaussianity. Hoyer et al. [14] generalized LiNGAM to the case with latent variable and obtained latent variable linear non-Gaussian acyclic model (lvLiNGAM).

In this paper, we defined a class of structural VAR models affected by latent variables and given the directed acyclic graph representation for the model. A computational method based on independent component analysis is proposed to estimate the parameters in the model. The directed acyclic graph of structural VAR model with latent variables is constructed based on the results of parameters estimation and conditional independent test.

This paper is organized as follows: In Section II, we define structural VAR model with latent variable and the foundation assumptions of the model to be identified. Section III presents the parameters estimation method for the model. The validity of the methods is demonstrated by simulations in Section IV. Section V concludes the paper.

II. Model Definition and Assumptions

Suppose that $X(t) = (X_1(t), X_2(t), ..., X_k(t))'$ is a k-dimensional stationary time series with zero-mean, for any $t \in \mathbb{Z}$, the basic VAR model has the reduced form representation

$$X(t) = A_1 X(t-1) + \cdots + A_p X(t-p) + u(t). \tag{1}$$

Here A_j $(j = 1, ..., p)$ are $k \times k$ coefficient matrices, $u(t)$ is the $k \times 1$ vector of random disturbances, which is assumed to be a zero-mean white noise process (i.e., no correlations across time) with contemporaneous covariance matrix $E(u_t u_t') = \Sigma_u$. From the diagonalization of covariance matrix, there exists matrix P which satisfies that $PP' = \Sigma_u$. We get the structural VAR representation

$$X(t) = \Phi_0 X(t) + \Phi_1 X(t-1) + \cdots + \Phi_p X(t-p) + \varepsilon(t), \tag{2}$$

where $\Phi_0 = I - P^{-1}$, $\Phi_j = P^{-1} A_j$, $j = 1, ..., p$, $\varepsilon(t) = P^{-1} U(t)$ and the variance matrix $D = P^{-1} \Sigma_u P^{-1'}$ of $\varepsilon(t)$ is diagonal. Let $\Phi_q(ij)$ denote the element with i row and j column in matrix Φ_q, for any indices sequence $(j_1, j_2, ..., j_m)$, there is a zero in the m coefficient $\Phi_0(j_2 j_1)$, $\Phi_0(j_3 j_2), ..., \Phi_0(j_m j_{m-1}), \Phi_0(j_1 j_m)$. Matrix Φ_0 represents a recursive (causal) dependence of each component of $X(t)$ on other contemporaneous components.

The model is recursive and can be represented by directed acyclic graphs.

Definition 2.1. Let $X(t)$ be a k-dimensional stationary Gaussian process with structural VAR representation of (2). In a graph $G = (V, E)$, the vertex set $V = \{X_1(t), \ldots, X_k(t), \ldots, X_1(t-p), \ldots, X_k(t-p)\} := (a_t, a_{t-1}, \ldots, a_{t-p})$, $a \in \{1, 2, \ldots, k\}$ denotes the variables of the components in X_t at different times and edge set E satisfies that two vertices being without an edge if and only if they are conditional independent given all the remainder variables, the directions of the edges denoting the directions of causal dependence. Then G is called the *directed acyclic graph* (DAG) for $X(t)$.

Therefore, each possible ordering of the components in $X(t)$ gives a potentially distinct form of (1), but all these forms are statistically equivalent and have the same DAG representation. Since $X_a(t-u)$ and $X_b(t)$ are conditional independence if and only if the corresponding entries $\Phi_u(ba)$ vanish, we have the next theorem.

Theorem 2.2. *Let* $X(t)$ *be a* k*-dimensional stationary Gaussian process with structural VAR representation* (1). *Then the DAG with* $X(t)$ *is the graph* $G = (V, E)$ *with vertex set* $V = (a_t, a_{t-1}, \ldots, a_{t-p})$, $a \in \{1, 2, \ldots, k\}$ *and edge set E such that*

(1) *for distinct* $a, b \in \{1, 2, \ldots, k\}$, $a_{t-u} \to b_t \notin E \Leftrightarrow \Phi_u(ba) = 0$, $u \in \{0, 1, 2, \ldots, p\}$;

(2) $a_{t-u} \to a_t \notin E \Leftrightarrow \Phi_u(aa) = 0$, $u \in \{1, 2, \ldots, p\}$.

Then we define the structural VAR models with latent variables.

Definition 2.3. Suppose that $X(t) = (X_1(t), X_2(t), \ldots, X_k(t))'$, $t \in \mathbb{Z}$ is a k-dimensional observed time series, $U(t) = (U_1(t), \ldots, U_q(t))$ is a q-dimensional unobserved time series, and $\varepsilon(t)$ is a k-dimensional white noise. The structural VAR model with latent variables can be defined as

$$X(t) = \sum_{\tau=0}^{\tau=p} \Phi_\tau X(t-\tau) + HU(t) + \varepsilon(t), \tag{3}$$

where the process $X(t)$ satisfies the following conditions:

(A1) $\varepsilon_i(t),\ i = 1, ..., k,\ t \in \mathbb{Z}$ and $U_j(t),\ j = 1, ..., q,\ t \in \mathbb{Z}$ are jointly independent and non-Gaussian.

(A2) The matrix Φ_0 modelling instantaneous effects corresponds to an acyclic graph, as is typical in causal analysis.

(A3) Each latent variable $U_j(t),\ \ j = 1, ..., q$ is a root node (i.e., has no parents) and has at least two children (direct descendants). Furthermore, although different $U_j(t)$ may have the same sets of children, no two latent variables exhibit exactly proportional sets of connection strengths to the observed variables.

(A4) Each latent variable is restricted to have zero-mean and unit variance.

The acyclicity condition (A2) is equivalent to the existence of a permutation matrix P, which corresponds to an ordering of the variables $X_i(t),\ \ i = 1, ..., k$ such that the matrix $P\Phi_0 P^T$ is lower-triangular (i.e., entries above the diagonal are zero).

Conditions (A3) and (A4) mean that the model for $X_i(t), U_j(t)$ is a canonical model defined by Hoyer et al. [14].

Under assumptions (A1) to (A4), the DAG represented the causal dependence structure of structural VAR models with latent variables defined in Definition 2.4 is as follows.

Definition 2.4. Let $X(t) = (X_1(t), ..., X_k(t))$ and

$$U(t) = (U_1(t), ..., U_q(t)), \quad t \in \mathbb{Z}$$

be the time series with structural VAR representation (1). Then the DAG with $X(t)$ and $U(t)$ is the graph $G = (V, E)$ with vertex set $V = (a_t, a_{t-1}, ..., a_{t-p}, l_t)$, $a \in \{1, 2, ..., k\}$, $l \in \{1, 2, ..., q\}$ and edge set E such that

(1) for distinct $a, b \in \{1, 2, ..., k\}$, $a_{t-u} \to b_t \notin E \Leftrightarrow \Phi_u(ba) = 0$, $u \in \{0, 1, 2, ..., p\}$;

(2) $a_{t-u} \to a_t \notin E \Leftrightarrow \Phi_u(aa) = 0$, $u \in \{0, 1, 2, ..., p\}$;

(3) $l_t \to a_t \notin E \Leftrightarrow H(al) = 0$.

III. Parameter Estimation Method Based on Overcomplete ICA

In this section, a method combining classic least-squares estimation of an autoregressive model with latent variable overcomplete ICA estimation [15] is proposed to estimate the parameters of the structural VAR model defined in Definition 2.3.

Independent component analysis solves the non-identifiability of factor analytic models using the assumption of non-Gaussianity of the factors [10, 11]. We begin by considering the full data vector $\tilde{Y} = \{y_1, ..., y_m\}$ which includes the latent variables. If we first subtract out the means of the variables, then the full data satisfies

$$\tilde{Y} = \tilde{B}\tilde{Y} + e, \tag{4}$$

where $\tilde{B}$ is a matrix that could be permuted (by simultaneous equal row and column permutations) to strict lower triangularity if one knew a causal ordering $k(i)$ of the variables. Solving for $\tilde{Y}$, we obtain

$$\tilde{Y} = \tilde{A}e, \tag{5}$$

where $\tilde{A} = (I - \tilde{B})^{-1}$ contains the influence of the disturbance variables onto the observed variables (the total effects). Again, $\tilde{A}$ could be permuted to lower triangularity (although not strict lower triangularity) with an appropriate permutation $k(i)$. Taken together, the linear relationship between

e and $\tilde{Y}$ and the independence and non-Gaussianity of the components of e define the standard linear independent component analysis model [10, 11].

Now consider the effect of hiding some of the variables. This yields $Y = Ae$; where A contains just the rows of $\tilde{A}$ corresponding to the observed variables. When the number of observed variables is less than the number of disturbance variables, A is non-square with more columns than rows. This is known as an overcomplete basis in the ICA literature [15].

Thus, Hoyer et al. [14] formulated the algorithm for calculating all observationally equivalent canonical models compatible with any given ICA basis matrix A (containing exact zeros). For more details on the algorithm, please see the reference.

The VAR-lvLiNGAM algorithm for identification of the structural VAR model with latent variables is provided in Algorithm 1.

Algorithm 1: Structural VAR-lvLiNGAM

Step 1. Using least-squares method, estimate a classic autoregressive model for the data $X(t) = \sum_{\tau=1}^{\tau=p} M_\tau X(t-\tau) + n(t)$. Denote the estimates of the autoregressive matrices by $\hat{M}_\tau,\ \tau = 1, \ldots, p$.

Step 2. Compute the residuals, i.e., estimates of innovations $n(t)$ as $\hat{n}(t) = X(t) - \sum_{\tau=1}^{\tau=p} \hat{M}_\tau X(t-\tau)$.

Step 3. Check whether the $\hat{n}(t)$ indeed are non-Gaussian, and proceed only if this is so.

Step 4. $n(t) = \Phi_0 n(t) + \tilde{\varepsilon}(t) + HU(t)$ can be rewritten as $(I - \Phi_0)n(t) = (I \quad H)\begin{pmatrix} \tilde{\varepsilon}(t) \\ H(t) \end{pmatrix}$. This yields $n(t) = Ae(t)$, where $A = (I - \Phi_0)^{-1}(I \quad H)$ contains just the rows of A corresponding to the observed variables. This is an overcomplete basis [16]. From the assumptions in Definition 2.3, the model $n(t) = Ae(t)$ is a canonical model.

Step 5. Use overcomplete ICA algorithm [17] to obtain overcomplete basis A and estimate the means of the observed variables, calculate all observationally equivalent canonical latent variable LiNGAM models compatible with the basis.

Step 6. The matrix Φ_0 and H can be estimated as the solution of the instantaneous causal model with latent variables $\hat{n}(t) = \hat{\Phi}_0\hat{n}(t) + \tilde{\varepsilon}(t) + \hat{H}U(t)$.

Step 7. Finally, compute the estimates of the causal effect matrices Φ_τ, $\tau = 1, \ldots, p$ as $\hat{\Phi}_\tau = (I - \hat{\Phi}_0)\hat{M}_\tau$.

The next theorem proves the efficiency of the algorithm.

Theorem 3.1. (a) *The estimate of matrix* Φ_0 *and H can be computed as the solution of the instantaneous causal model with latent variables*,

$$\hat{n}(t) = \hat{\Phi}_0\hat{n}(t) + \tilde{\varepsilon}(t) + \hat{H}U(t). \tag{6}$$

(b) *The estimates of the causal effect matrices* Φ_τ, $\tau = 1, \ldots, p$ *can be computed as*

$$\hat{\Phi}_\tau = (I - \hat{\Phi}_0)\hat{M}_\tau. \tag{7}$$

Proof. From (3), we have

$$X(t) = \Phi_0 X(t) + \sum_{\tau=1}^{\tau=p} \Phi_\tau X(t-\tau) + HU(t) + \varepsilon(t), \tag{8}$$

then

$$(I - \Phi_0)X(t) = \sum_{\tau=1}^{\tau=p} \Phi_\tau X(t-\tau) + HU(t) + \varepsilon(t), \tag{9}$$

and thus

$$X(t) = \sum_{\tau=1}^{\tau=p} (I - \Phi_0)^{-1}\Phi_\tau X(t-\tau) + (I - \Phi_0)^{-1}HU(t) + (I - \Phi_0)^{-1}\varepsilon(t). \tag{10}$$

Comparing this with (3), we can equate the autoregressive matrices, which give

$$(I - \Phi_0)^{-1}\Phi_\tau = M_\tau \tag{11}$$

and

$$n(t) = (I - \Phi_0)^{-1}HU(t) + (I - \Phi_0)^{-1}\varepsilon(t). \tag{12}$$

Equation (12) multiplied by $(I - \Phi_0)$,

$$(I - \Phi_0)n(t) = HU(t) + \varepsilon(t). \tag{13}$$

The equation above can be written as

$$n(t) = \Phi_0 n(t) + HU(t) + \varepsilon(t) \tag{14}$$

which is just an lvLiNGAM model [14] on the residuals $\hat{n}(t)$. □

This estimation method is consistent, since the least-squares estimation is consistent and the consistency of the estimator of Φ_0 and H follow from the consistency of lvLiNGAM estimation [14].

In the problem of identification, a more reassuring result is the following: if the data follows the same causal ordering for all time lags, then ordering is not contradicted by neglecting instantaneous effect. The existence of latent variable makes the problem more complex. A rigorous definition of this property is the following.

Theorem 3.2. *Assume that there is an ordering* $i(j),\ j = 1, \ldots, n$ *of the variables such that no effect goes backward*, *i.e.*,

$$\Phi_\tau(i(j - \delta), i(j)) = 0 \ \ for\ \ \delta > 0, \ \ \tau \geq 0, \ \ 1 \leq j \leq n. \tag{15}$$

Then the same property applies to the $M_\tau,\ \tau \geq 1$ *as well*. *Conversely*, *if there is an ordering such that* (18) *applies to* $M_\tau,\ \tau \geq 1$ *and* Φ_0, *then it applies to* $\Phi_\tau,\ \tau \geq 1$ *as well*.

Proof. Model

$$X(t) = \sum_{\tau=0}^{\tau=p} \Phi_\tau X(t - \tau) + HU(t) + \varepsilon(t) \tag{16}$$

can be written as

$$\overline{X}(t) = \sum_{\tau=0}^{\tau=p} \overline{\Phi}_\tau \overline{X}(t-\tau) + \overline{\varepsilon}(t), \tag{17}$$

where $\overline{X}(t) = \begin{pmatrix} U(t) \\ X(t) \end{pmatrix}$, $\overline{\Phi}_0 = \begin{pmatrix} 0 & 0 \\ H & \Phi_0 \end{pmatrix}$, $\overline{\Phi}_\tau = \begin{pmatrix} 0 & 0 \\ 0 & \Phi_\tau \end{pmatrix}$, $\tau \geq 1$, $\overline{\varepsilon}(t) = \begin{pmatrix} \varepsilon_U(t) \\ \varepsilon(t) \end{pmatrix}$.

When the variables are ordered in this way (assuming such an order exists), all the matrices Φ_τ are lower-triangular. The same applies to $(I - \Phi_0)$. Then all the matrices $\overline{\Phi}_\tau$ and $(I - \overline{\Phi}_0)$ are lower-triangular too. Now the product of two lower-triangular matrices is lower-triangular and the inverse of a lower-triangular matrix is lower-triangular; in particular,

$$\overline{M}_\tau = (I - \overline{\Phi}_0)^{-1}\overline{\Phi}_\tau = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & \Phi_\tau \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & D\Phi_\tau \end{pmatrix}, \tag{18}$$

where A, D, Φ_τ are lower-triangular, then $\overline{M}_\tau$ are also lower-triangular, which proves the first part of the theorem. The converse part follows from solving for $\overline{\Phi}_\tau$ in (18) and the fact that the inverse of a lower-triangular matrix is lower-triangular. □

IV. Simulation Example

In this section, we conduct an experiment using synthetic data to testify the validity of the methods. The simulation data is generated from the next model

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 0.5 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} + \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}\begin{pmatrix} X_1(t-1) \\ X_2(t-1) \end{pmatrix}$$

$$+ \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix} U(t) + \begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{pmatrix}.$$

Then the parameters are

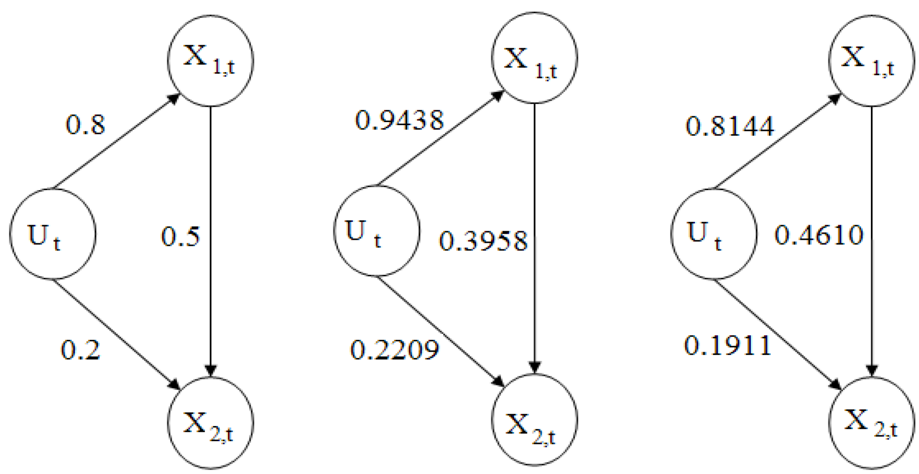
$$B_0 = \begin{pmatrix} 0 & 0.5 \\ 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}, \quad H = \begin{pmatrix} 0.8 \\ 0.2 \end{pmatrix}.$$

The influences $\varepsilon_i(t)$ and latent variable $U(t)$ are independently drawing from a two-component mixture of Gaussian density $0.9\xi_1 + 0.1\xi_2$, where $\xi_1 \sim N(0, 0.01)$, $\xi_2 \sim N(0, 1)$.

We applied our estimation method to the data with sample sizes $n = 500$ and $n = 1000$, respectively. Figure 1 gives the estimated structure of the instantaneous variables and the latent variable.

The estimate of B_1 is computed from $\hat{B}_1 = (I - \hat{B}_0)\hat{M}_1$. The results are shown in Table I.

Table I shows that the accuracy of the estimation is becoming better with the increasing of the sample size. Figure 2 gives the graph the dependence structure of the variables from Table I.



(a) (b) (c)

Figure 1. (a) Original generating model; (b) Estimated model with $n = 500$; and (c) Estimated model with $n = 1000$.

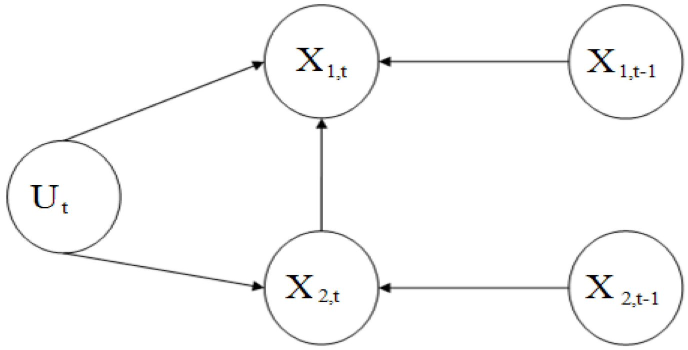


Figure 2. Graphical model from the results of Table I.

Table I. Estimations of the parameter matrix

	B_0	B_1	H
$n = 500$	$\begin{pmatrix} 0 & 0.3958 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.9319 & 0.1080 \\ 0.0427 & 0.9857 \end{pmatrix}$	$\begin{pmatrix} 0.9438 \\ 0.2209 \end{pmatrix}$
$n = 1000$	$\begin{pmatrix} 0 & 0.4610 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.9238 & 0.0495 \\ 0.0405 & 0.9845 \end{pmatrix}$	$\begin{pmatrix} 0.8144 \\ 0.1911 \end{pmatrix}$

V. Conclusion

This study extends the application of causal inference for time series to the case that the model is affected by latent variables. We present a new procedure that combines least-squares methods and overcomplete independent component analysis model to estimate the parameters of linear non-Gaussian structural VAR model with latent variables. The combination of the estimation and test procedures can give an accurate graphical model to represent the dependence structure for structural VAR model with latent variables.

Acknowledgement

The authors would like to thank the Scientific Research Program Funded by Shaanxi Provincial Education Department under Program No. 11JK0497.

References

[1] M. Lanne and H. Lütkepohl, Structural vector autoregressions with nonnormal residuals, J. Bus. Econom. Statist. 28(1) (2006), 159-168.

[2] D. R. Brillinger, Remarks concerning graphical models for time series and point processes, Revista de Econometria 16 (1996), 1-23.

[3] R. Dahlhaus, Graphical interaction models for multivariate time series, Metrika 51 (2000), 157-172.

[4] M. Reale and G. T. Wilson, Identification of vector AR models with recursive structural errors using conditional independence graphs, Stat. Meth. Appl. 10 (2001), 49-65.

[5] S. Demiralp and K. D. Hoover, Searching for the causal structure of a vector autoregression, Oxford Bull. Econom. Statist. 65 (2003), 745-767.

[6] A. Moneta, Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis, Empirical Economics 35 (2008), 275-300.

[7] M. Eichler, Graphical modeling of multivariate time series with latent variables, Technical Report, University of Maastricht, 2006.

[8] T. Chu and C. Glymour, Search for additive nonlinear time series causal models, J. Mach. Learn. Res. 9 (2008), 967-991.

[9] W. Gao and Z. Tian, Latent ancestral graph models of structure VAR, J. Systems Engineering and Electronics 21 (2010), 233-238.

[10] P. Comon, Independent component analysis: a new concept?, Signal Processing 36 (1994), 287-314.

[11] J. Eriksson and V. Koivunen, Identifiability, separability and uniqueness of linear ICA models, IEEE Signal Processing Letters 11(7) (2004), 601-604.

[12] S. Shimizu, P. O. Hoyer, A. Hyvärinen and A. J. Kerminen, A linear non-Gaussian acyclic model for causal discovery, J. Mach. Learn. Res. 7 (2006), 2003-2030.

[13] A. Hyvärinen, S. Shimizu and P. O. Hoyer, Estimation of a structural vector autoregression model using non-Gaussianity, J. Mach. Learn. Res. 11 (2010), 1709-1731.

[14] P. O. Hoyer, S. Shimizu, A. J. Kerminen and M. Palviainen, Estimation of causal effects using linear non-Gaussian causal models with hidden variables, Internat. J. Approx. Reason. 49(2) (2008), 362-378.

[15] M. S. Lewicki and T. J. Sejnowski, Learning overcomplete representations, Neural Comput. 12 (2000), 337-365.

[16] J. Eriksson and V. Koivunen, Identifiability, separability and uniqueness of linear ICA models, IEEE Signal Processing Letters 11 (2004), 601-604.

[17] H. Attias, Independent factor analysis, Neural Comput. 11 (1999), 803-851.