CMS 2: 253–278 (2005)

DOI: 10.1007/s10287-004-0021-x

Efficient strategies for deriving the subset VAR models*

Cristian Gatu[1] **and Erricos John Kontoghiorghes**[2,3]

[1] Institut d'informatique, Université de Neuchâtel, Switzerland.
[2] Department of Public and Business Administration, University of Cyprus, Cyprus.
[3] Computer Science and Information Systems, Birkbeck College, Univ. of London, UK.

Abstract. Algorithms for computing the subset Vector Autoregressive (VAR) models are proposed. These algorithms can be used to choose a subset of the most statistically-significant variables of a VAR model. In such cases, the selection criteria are based on the residual sum of squares or the estimated residual covariance matrix. The VAR model with zero coefficient restrictions is formulated as a Seemingly Unrelated Regressions (SUR) model. Furthermore, the SUR model is transformed into one of smaller size, where the exogenous matrices comprise columns of a triangular matrix. Efficient algorithms which exploit the common columns of the exogenous matrices, sparse structure of the variance-covariance of the disturbances and special properties of the SUR models are investigated. The main computational tool of the selection strategies is the generalized QR decomposition and its modification.

Keywords: VAR models, SUR models, Subset regression, Least squares, QR decomposition

1 Introduction

A common problem in the Vector Autoregressive (VAR) process modeling is the lag structure identification or, equivalently, the specification of the subset VAR models

---

Send offprint requests to: cristian.gatu@unine.ch

* This work is in part supported by the Swiss National Foundation for Research Grants 101312-100757/1, 200020-10016/1 and 101412-105978.

Correspondence to: Cristian Gatu, Institut d'informatique, Université de Neuchâtel, Emile-Argand 11, Case Postale 2, CH-2007 Neuchâtel, Switzerland. e-mail: cristian.gatu@unine.ch, erricos@ucy.ac.cy.

CMS Computational Management Science

[3, 7, 8, 18, 22, 27]. The vector time series $z_t \in \mathbb{R}^G$ is a VAR process of order p when its data generating process has the form

$$z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \cdots + \Phi_p z_{t-p} + \epsilon_t, \tag{1}$$

where $\Phi_i \in \mathbb{R}^{G \times G}$ are the coefficient matrices and $\epsilon_t \in \mathbb{R}^G$ is the noise vector. Given a set of realizations of the process in (1), $z_1, \ldots, z_M$ and a pre-sample $z_0, \ldots, z_{1-p}$ the parameter matrices are estimated from the linear model

$$\begin{pmatrix} z_1^T \\ z_2^T \\ \vdots \\ z_M^T \end{pmatrix} = \begin{pmatrix} z_0^T & z_{-1}^T & \cdots & z_{1-p}^T \\ z_1^T & z_0^T & \cdots & z_{2-p}^T \\ \vdots & \vdots & \ddots & \vdots \\ z_{M-1}^T & z_{M-2}^T & \cdots & z_{M-p}^T \end{pmatrix} \begin{pmatrix} \Phi_1^T \\ \Phi_2^T \\ \vdots \\ \Phi_p^T \end{pmatrix} + \begin{pmatrix} \epsilon_1^T \\ \epsilon_2^T \\ \vdots \\ \epsilon_M^T \end{pmatrix}. \tag{2}$$

In the compact form the model in (2) can be written as

$$Y = XB + U, \tag{3}$$

where $Y = (y_1 \ldots y_G) \in \mathbb{R}^{M \times G}$ are the response vectors, $X \in \mathbb{R}^{M \times K}$ is the lagged exogenous data matrix having full-column rank and block-Toeplitz structure, $B \in \mathbb{R}^{K \times G}$ is the coefficient matrix, $U = (u_1 \ldots u_G) \in \mathbb{R}^{M \times G}$ are the disturbances and $K = Gp$. The expectation of U is zero, i.e. $\mathrm{E}(u_i) = 0$, and $\mathrm{E}(u_i u_j^T) = \sigma_{ij} I_M$ $(i, j = 1, \ldots, G)$ [15, 17, 18, 19, 21, 25]. The VAR model (3) can be written as

$$\operatorname{vec}(Y) = (I_G \otimes X) \operatorname{vec}(B) + \operatorname{vec}(U), \quad \operatorname{vec}(U) \sim (0, \Sigma \otimes I_M), \tag{4}$$

where vec is the vector operator and $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{G \times G}$ has full rank [6, 11]. The Ordinary and Generalized Least Squares estimators of (4) are the same and given by

$$\hat{B} = (X^T X)^{-1} X^T Y.$$

Often zero-coefficient constraints are imposed on the VAR models. This might be due to the fact that the data generating process in (1) contains only a few non-zero coefficients. Also, over-fitting the model might yield in loss of efficiency when it is used for further testing, such as forecasting [18, 28]. A zero-restricted VAR model (ZR–VAR) is called subset VAR model. When prior knowledge about zero-coefficient constraints are not available, several subset VAR models have to be compared with respect to some specified criterion. If the purpose is the identification of a model as close as possible to the data generating process, then the use of an information criterion for evaluating the subset models is appropriate. The selection criteria such as Akaike Information Criterion (AIC), Hannan-Quinn (HQ) and Schwarz Criterion (SC) are based on the residual sum of squares or the estimated residual covariance matrix [1, 2, 9, 23]. There is a trade-off between a good fit, i.e. small value of the residual sum of squares, and the number of non-zero coefficients. That is, there is a penalty related to the number of included non-zero coefficients.

Finding good models can be seen as an optimization problem, i.e. minimize or maximize a selection criterion over a set of sub-models derived from a finite realization of the process in (1) by applying a selection rule [28]. Let $B = (b_1 \dots b_G)$ and $S_i \in \mathbb{R}^{K \times k_i}$ $(i = 1, \dots, G)$ denote a selection matrix such that $\beta_i = S_i^T b_i$ corresponds to the non-zero coefficients of b_i – the ith column of B. Furthermore, let $X_i = XS_i$ which are the columns of X that correspond to the non-zero coefficients of b_i. Thus, the ZR–VAR model is equivalent to the Seemingly Unrelated Regressions (SUR) model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{pmatrix} = \begin{pmatrix} XS_1 & & & \\ & XS_2 & & \\ & & \ddots & \\ & & & XS_G \end{pmatrix} \begin{pmatrix} S_1^T b_1 \\ S_2^T b_2 \\ \vdots \\ S_G^T b_G \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_G \end{pmatrix},$$

or

$$\operatorname{vec}(Y) = (\oplus_{i=1}^G X_i) \operatorname{vec}(\{\beta_i\}_G) + \operatorname{vec}(U), \quad \operatorname{vec}(U) \sim (0, \Sigma \otimes I_M), \tag{5}$$

where $\oplus_{i=1}^G X_i = \operatorname{diag}(X_1, \dots, X_G)$, $\{\beta_i\}_G$ denotes the set $\{\beta_1, \dots, \beta_G\}$ and $\operatorname{vec}(\{\beta_i\}_G) = (\beta_1^T \dots \beta_G^T)^T$. For notational convenience the direct sum $\oplus_{i=1}^G$ and the set $\{\cdot\}_G$ are abbreviated to $\oplus_i$ and $\{\cdot\}$, respectively.

One possible approach to search for the optimal models is to enumerate all $2^{pG^2} - 1$ possible subset VAR models. However this approach is infeasible even for modest values of G and p. Thus, existing methods search in a smaller given subspace. One selection method is to enforce a whole coefficient matrix $\Phi_i (1 \leq i \leq p)$, or a combination of coefficient matrices to be zero. In this case, the number of the subset VAR models to be evaluated is $2^p - 1$. Polynomial top-down and bottom-up strategies based on the deletion and, respectively, inclusion of the coefficients in each equation separately have been also previously proposed [18]. Alternative methods use optimization heuristics such as Threshold Accepting [28].

Several algorithms for computing the subset VAR models are presented. The ZR–VAR model which is formulated as a SUR model, is transformed into one of smaller size, where the exogenous matrices comprise columns of a triangular matrix [6]. The common columns of the exogenous matrices and the Kronecker structure of the variance-covariance of the disturbances are exploited in order to derive efficient estimation algorithms.

In the next section the numerical solution of the ZR–VAR model is given. Section 3 presents an efficient variable-downdating strategy of the subset VAR model. Section 4 describes an algorithm for deriving all subset VAR models by moving efficiently from one model to another. Special cases which take advantage of the common-columns property of the data matrix and the Kronecker structure of the variance-covariance matrix are described in Section 5. Conclusion and future work are presented and discussed in Section 6.

2 Numerical solution of the ZR–VAR model

Orthogonal transformations can be employed to reduce to zero $M-K$ observations of the VAR model (3). This results in an equivalent transformed model with less observations, and thus, to a smaller-size estimation problem. Consider the QR decomposition (QRD) of the exogenous matrix X

$$\bar{Q}^T X = \begin{array}{c} K \\ \begin{pmatrix} R \\ 0 \end{pmatrix} \end{array}\begin{array}{l} \\ K \\ M-K \end{array}, \quad \text{with } \bar{Q} = \begin{array}{cc} K & M-K \\ \big(\bar{Q}_A & \bar{Q}_B\big) \end{array}, \tag{6}$$

where $\bar{Q} \in \mathbb{R}^{M\times M}$ is orthogonal and $R \in \mathbb{R}^{K\times K}$ is upper-triangular. Notice that the matrix R is a compact form of the information contained in the original data matrix X. Let

$$\bar{Q}^T Y = \begin{array}{c} G \\ \begin{pmatrix} \tilde{Y} \\ \hat{Y} \end{pmatrix} \end{array}\begin{array}{l} \\ K \\ M-K \end{array} \text{ and } \quad \bar{Q}^T U = \begin{array}{c} G \\ \begin{pmatrix} \tilde{U} \\ \hat{U} \end{pmatrix} \end{array}\begin{array}{l} \\ K \\ M-K \end{array}, \tag{7}$$

where

$$\begin{pmatrix} \operatorname{vec}(\tilde{U}) \\ \operatorname{vec}(\hat{U}) \end{pmatrix} \sim \begin{pmatrix} \Sigma \otimes I_K & 0 \\ 0 & \Sigma \otimes I_{M-K} \end{pmatrix}.$$

Premultiplying (4) by $(I_G \otimes \bar{Q}_A \quad I_G \otimes \bar{Q}_B)^T$ gives

$$\begin{pmatrix} \operatorname{vec}(\bar{Q}_A^T Y) \\ \operatorname{vec}(\bar{Q}_B^T Y) \end{pmatrix} = \begin{pmatrix} I_G \otimes \bar{Q}_A^T X \\ I_G \otimes \bar{Q}_B^T X \end{pmatrix} \operatorname{vec}(B) + \begin{pmatrix} \operatorname{vec}(\bar{Q}_A^T U) \\ \operatorname{vec}(\bar{Q}_B^T U) \end{pmatrix}.$$

From (6) and (7) it follows that the latter can be written as

$$\begin{pmatrix} \operatorname{vec}(\tilde{Y}) \\ \operatorname{vec}(\hat{Y}) \end{pmatrix} = \begin{pmatrix} I_G \otimes R \\ 0 \end{pmatrix} \operatorname{vec}(B) + \begin{pmatrix} \operatorname{vec}(\tilde{U}) \\ \operatorname{vec}(\hat{U}) \end{pmatrix}$$

which is equivalent to the reduced-size model

$$\operatorname{vec}(\tilde{Y}) = (I_G \otimes R)\operatorname{vec}(B) + \operatorname{vec}(\tilde{U}), \quad \operatorname{vec}(\tilde{U}) \sim (0, \Sigma \otimes I_K). \tag{8}$$

From the latter, it follows that (5) is equivalent to the smaller in size SUR model

$$\operatorname{vec}(\tilde{Y}) = (\oplus_i R^{(i)})\operatorname{vec}(\{\beta_i\}) + \operatorname{vec}(\tilde{U}), \quad \operatorname{vec}(\tilde{U}) \sim (0, \Sigma \otimes I_K), \tag{9}$$

where $R^{(i)} = RS_i \in \mathbb{R}^{K\times k_i}$ [5, 11, 12].

The best linear unbiased estimator (BLUE) of the SUR model in (9) comes from the solution of the Generalized Linear Least Squares Problem (GLLSP)

$$\operatorname*{argmin}_{V,\{\beta_i\}} \|V\|_F^2 \quad \text{subject to } \operatorname{vec}(\tilde{Y}) = (\oplus_i R^{(i)})\operatorname{vec}(\{\beta_i\}) + \operatorname{vec}(VC^T). \tag{10}$$

Here $\Sigma = CC^T$, the random matrix $V \in \mathbb{R}^{K\times G}$ is defined as $VC^T = \tilde{U}$ which implies $\text{vec}(V) \sim (0, I_{GK})$, and $\|\cdot\|_F$ denotes the Frobenius norm i.e. $\|V\|_F^2 = \sum_{i=1}^{K}\sum_{j=1}^{G} V_{i,j}^2$ [15, 17, 19, 20, 21]. The upper-triangular $C \in \mathbb{R}^{G\times G}$ is the Cholesky factor of Σ. For the solution of (10) consider the Generalized QR Decomposition (GQRD) of the matrices $\oplus_i R^{(i)}$ and $C \otimes I_K$:

$$Q^T(\oplus_i R^{(i)}) = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \begin{matrix} K^* \\ GK-K^* \end{matrix} \tag{11a}$$

and

$$\begin{aligned} Q^T(C\otimes I_K)\Pi P &= \begin{pmatrix} W^{(0,1)} & W^{(0,2)} \\ W^{(0,3)} & W^{(0,4)} \end{pmatrix} P \\ &= W \equiv \begin{array}{c} \begin{matrix} K^* & GK-K^* \end{matrix} \\ \begin{pmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{pmatrix} \end{array} \begin{matrix} K^* \\ GK-K^* \end{matrix}, \end{aligned} \tag{11b}$$

where $K^* = \sum_{i=1}^{G} k_i$, $\oplus R_i$ and W are upper triangular of order K^* and GK, respectively, and Π is a $GK \times GK$ permutation matrix defined as $\Pi = (\oplus_i(I_{k_i}\ 0)^T \ \oplus_i(0\ I_{K-k_i})^T)$. Furthermore, $W^{(0,j)}$ $(j = 1, 2, 3, 4)$ are block-triangular and $R_i \in \mathbb{R}^{k_i\times k_i}$ is the upper-triangular factor in the QRD of $R^{(i)}$. That is,

$$Q_i^T R^{(i)} = \begin{pmatrix} R_i \\ 0 \end{pmatrix} \begin{matrix} k_i \\ K-k_i \end{matrix}, \quad \text{with} \quad Q_i^T = \begin{pmatrix} Q_{Ai}^T \\ Q_{Bi}^T \end{pmatrix} \begin{matrix} k_i \\ K-k_i \end{matrix} \quad (i = 1, \ldots, G), \tag{12}$$

where $Q_i \in \mathbb{R}^{K\times K}$ is orthogonal and Q in (11a) is defined by

$$Q = (\oplus_i Q_{Ai} \quad \oplus_i Q_{Bi}) = \begin{pmatrix} Q_{A1} & & & Q_{B1} & & \\ & \ddots & & & \ddots & \\ & & Q_{AG} & & & Q_{BG} \end{pmatrix}.$$

Now, since $\|V\|_F^2 = \|P^T\Pi^T \text{vec}(V)\|^2$, the GLLSP (10) is equivalent to

$$\begin{aligned} &\underset{V,\{\beta_i\}}{\text{argmin}} \|P^T\Pi^T \text{vec}(V)\|^2 \quad \text{subject to} \\ &\quad Q^T \text{vec}(\tilde{Y}) = Q^T(\oplus_i R^{(i)})\,\text{vec}(\{\beta_i\}) + Q^T(C\otimes I_K)\Pi P P^T \Pi^T \text{vec}(V), \end{aligned} \tag{13}$$

where $\|\cdot\|$ denotes the Euclidian norm. Using (11a) and (11b) the latter can be re-written as

$$\underset{\{\tilde{v}_{Ai}\},\{\tilde{v}_{Bi}\},\{\beta_i\}}{\text{argmin}} \sum_{i=1}^{G}(\|\tilde{v}_{Ai}\|^2 + \|\tilde{v}_{Bi}\|^2) \quad \text{subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\tilde{y}_{Ai}\}) \\ \operatorname{vec}(\{\tilde{y}_{Bi}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \operatorname{vec}(\{\beta_i\}) + \begin{pmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\tilde{v}_{Ai}\}) \\ \operatorname{vec}(\{\tilde{v}_{Bi}\}) \end{pmatrix}, \tag{14}$$

where $\tilde{y}_{Ai}, \tilde{v}_{Ai} \in \mathbb{R}^{k_i}$, $\tilde{y}_{Bi}, \tilde{v}_{Bi} \in \mathbb{R}^{K-k_i}$,

$$\tilde{y}_{Ai} = Q_{Ai}^T \tilde{y}_i, \quad \tilde{y}_{Bi} = Q_{Bi}^T \tilde{y}_i \tag{15a}$$

and

$$P^T \Pi^T \operatorname{vec}(V) = \begin{pmatrix} \operatorname{vec}(\{\tilde{v}_{Ai}\}) \\ \operatorname{vec}(\{\tilde{v}_{Bi}\}) \end{pmatrix} \begin{matrix} K^* \\ GK-K^* \end{matrix}. \tag{15b}$$

From the constraint in (14) it follows that

$$\operatorname{vec}(\{\tilde{v}_{Bi}\}) = W_{22}^{-1} \operatorname{vec}(\{\tilde{y}_{Bi}\}), \quad i = 1, \ldots, G \tag{16}$$

and the GLLSP is reduced to

$$\underset{\{\tilde{v}_{Ai}\},\{\beta_i\}}{\operatorname{argmin}} \sum_{i=1}^{G} \|\tilde{v}_{Ai}\|^2 \quad \text{subject to}$$

$$\operatorname{vec}(\{\tilde{\tilde{y}}_i\}) = (\oplus_i R_i) \operatorname{vec}(\{\beta_i\}) + W_{11} \operatorname{vec}(\{\tilde{v}_{Ai}\}), \tag{17}$$

where

$$\operatorname{vec}(\{\tilde{\tilde{y}}_i\}) = \operatorname{vec}(\{\tilde{y}_{Ai}\}) - W_{12} \operatorname{vec}(\{\tilde{v}_{Bi}\}). \tag{18}$$

The solution of (17), and thus, the BLUE of (9), is obtained by setting $\tilde{v}_{Ai} = 0$ $(i = 1, \ldots, G)$ and solving the linear system

$$(\oplus_i R_i) \operatorname{vec}(\{\hat{\beta}_i\}) = \operatorname{vec}(\{\tilde{\tilde{y}}_i\}),$$

or, equivalently, by solving the set of triangular systems

$$R_i \hat{\beta}_i = \tilde{\tilde{y}}_i, \quad i = 1, \ldots, G.$$

3 Variable-downdating of the ZR–VAR model

Consider the re-estimation of the SUR model (9) when new zero constraints are imposed to the coefficients β_i $(i = 1, \ldots, G)$. That is, after estimating (9) the new SUR model to be estimated is given by

$$\operatorname{vec}(\tilde{Y}) = (\oplus_i \tilde{R}^{(i)}) \operatorname{vec}(\{\tilde{\beta}_i\}) + \operatorname{vec}(\tilde{U}), \quad \operatorname{vec}(\tilde{U}) \sim (0, \Sigma \otimes I_K), \tag{19}$$

where $\tilde{R}^{(i)} = R^{(i)}\tilde{S}_i$, $\tilde{\beta}_i = \tilde{S}_i^T\beta_i$ and $\tilde{S}_i \in \mathbb{R}^{k_i\times\tilde{k}_i}$ is a selection matrix ($0 \le \tilde{k}_i \le k_i$). This is equivalent to solving the GLLSP

$$\operatorname*{argmin}_{V,\{\tilde{\beta}_i\}} \|V\|_F^2 \quad \text{subject to} \quad \operatorname{vec}(\tilde{Y}) = (\oplus_i \tilde{R}^{(i)}) \operatorname{vec}(\{\tilde{\beta}_i\}) + \operatorname{vec}(VC^T). \tag{20}$$

From (11) and (15) it follows that (20) can be written as

$$\operatorname*{argmin}_{\{\tilde{v}_{Ai}\},\{\tilde{v}_{Bi}\},\{\tilde{\beta}_i\}} \sum_{i=1}^{G}(\|\tilde{v}_{Ai}\|^2 + \|\tilde{v}_{Bi}\|^2) \quad \text{subject to}$$
$$\begin{pmatrix}\operatorname{vec}(\{\tilde{y}_{Ai}\})\\ \operatorname{vec}(\{\tilde{y}_{Bi}\})\end{pmatrix} = \begin{pmatrix}\oplus_i R_i\tilde{S}_i\\ 0\end{pmatrix}\operatorname{vec}(\{\tilde{\beta}_i\}) + \begin{pmatrix}W_{11} & W_{12}\\ 0 & W_{22}\end{pmatrix}\begin{pmatrix}\operatorname{vec}(\{\tilde{v}_{Ai}\})\\ \operatorname{vec}(\{\tilde{v}_{Bi}\})\end{pmatrix}.$$

Using (16) and (18) the latter becomes

$$\operatorname*{argmin}_{\{\tilde{v}_{Ai}\},\{\tilde{\beta}_i\}} \|\operatorname{vec}(\{\tilde{v}_{Ai}\})\|^2 \quad \text{subject to}$$
$$\operatorname{vec}(\{\tilde{\tilde{y}}_i\}) = (\oplus_i R_i\tilde{S}_i)\operatorname{vec}(\{\tilde{\beta}_i\}) + W_{11}\operatorname{vec}(\{\tilde{v}_{Ai}\}). \tag{21}$$

Now, consider the GQRD of the matrices $\oplus_i R_i\tilde{S}_i$ and W_{11}, that is

$$\tilde{Q}^T(\oplus_i R_i\tilde{S}_i) = \begin{pmatrix}\oplus_i\tilde{R}_i\\ 0\end{pmatrix}\begin{array}{l}\tilde{K}^*\\ K^*-\tilde{K}^*\end{array} \tag{22a}$$

and

$$\tilde{Q}^T W_{11}\tilde{\Pi}\tilde{P} = \tilde{W} = \begin{array}{cc} \begin{array}{cc}\tilde{K}^* & K^*-\tilde{K}^*\end{array} & \\ \begin{pmatrix}\tilde{W}_{11} & \tilde{W}_{12}\\ 0 & \tilde{W}_{22}\end{pmatrix} & \begin{array}{l}\tilde{K}^*\\ K^*-\tilde{K}^*\end{array}\end{array}, \tag{22b}$$

where $\tilde{K}^* = \sum_{i=1}^{G}\tilde{k}_i$, $\oplus_i\tilde{R}_i$ and $\tilde{W}$ are upper triangular of order $\tilde{K}^*$ and K^*, respectively, and $\tilde{\Pi} = (\oplus_i(I_{\tilde{k}_i}\ \ 0)^T \oplus_i(0\ \ I_{k_i-\tilde{k}_i})^T)$. Notice that, the upper-triangular $\tilde{R}_i \in \mathbb{R}^{\tilde{k}_i\times\tilde{k}_i}$ comes from the QRD

$$\tilde{Q}_i^T R_i\tilde{S}_i = \begin{pmatrix}\tilde{R}_i\\ 0\end{pmatrix}\begin{array}{l}\tilde{k}_i\\ k_i-\tilde{k}_i\end{array}, \quad \text{with} \quad \tilde{Q}_i^T = \begin{pmatrix}\tilde{Q}_{Ai}^T\\ \tilde{Q}_{Bi}^T\end{pmatrix}\begin{array}{l}\tilde{k}_i\\ k_i-\tilde{k}_i\end{array} \quad (i = 1,\ldots,G), \tag{23}$$

where $\tilde{Q}_i \in \mathbb{R}^{k_i\times k_i}$ is orthogonal. The latter factorization is a re-triangularization of a triangular factor after deleting columns [10, 11, 14, 29]. Furthermore, $\tilde{Q}$ in (22) is defined by $\tilde{Q} = (\oplus_i\tilde{Q}_{Ai}\quad \oplus_i\tilde{Q}_{Bi})$. If $\tilde{y}_{Ai}^* = \tilde{Q}_{Ai}^T\tilde{\tilde{y}}_i$, $\tilde{y}_{Bi}^* = \tilde{Q}_{Bi}^T\tilde{\tilde{y}}_i$,

$$\tilde{P}^T\tilde{\Pi}^T\operatorname{vec}(\{\tilde{v}_{Ai}\}) = \begin{pmatrix}\operatorname{vec}(\{\tilde{v}_{Ai}^*\})\\ \operatorname{vec}(\{\tilde{v}_{Bi}^*\})\end{pmatrix}\begin{array}{l}\tilde{K}^*\\ K^*-\tilde{K}^*\end{array}$$

and

$$\operatorname{vec}(\{\tilde{y}_i^*\}) = \operatorname{vec}(\{\tilde{y}_{Ai}^*\}) - \tilde{W}_{12} \operatorname{vec}(\{\tilde{v}_{Bi}^*\}),$$

then the solution of the GLLSP (21) is obtained by solving

$$(\oplus_i \tilde{R}_i) \operatorname{vec}(\{\hat{\tilde{\beta}}_i\}) = \operatorname{vec}(\{\tilde{y}_i^*\}),$$

or $\tilde{R}_i \hat{\tilde{\beta}}_i = \tilde{y}_i^*$ $(i = 1, \ldots, G)$. Notice that, the GQRD (22) is the most expensive computation required for deriving the BLUE of the SUR model (19) after the factorization (11) has been computed.

An efficient strategy for computing the orthogonal factorization (22b) has been proposed within the context of updating SUR models [13]. Notice that $\tilde{Q}^T W_{11} \tilde{\Pi}$ in (22b) can be written as

$$\tilde{Q}^T W_{11} \tilde{\Pi} = W^{(0)} = \begin{array}{c} \begin{array}{cc} \tilde{K}^* & K^*-\tilde{K}^* \end{array} \\ \begin{pmatrix} W^{(0,1)} & W^{(0,2)} \\ W^{(0,3)} & W^{(0,4)} \end{pmatrix} \end{array} \begin{array}{l} \\ \tilde{K}^* \\ K^*-\tilde{K}^* \end{array}, \tag{24}$$

where $W^{(0,i)}$ $(i = 1, \ldots, 4)$ has a block-triangular structure. That is, $W^{(0)}$ has the structural form

$$W^{(0)} = \begin{array}{c} \begin{array}{cccccccc} \tilde{k}_1 & \tilde{k}_2 & \ldots & \tilde{k}_G & k_1-\tilde{k}_1 & k_2-\tilde{k}_2 & \ldots & k_G-\tilde{k}_G \end{array} \\ \left(\begin{array}{cccc|cccc} W_{11}^{(0,1)} & W_{12}^{(0,1)} & \cdots & W_{1G}^{(0,1)} & W_{11}^{(0,2)} & W_{12}^{(0,2)} & \cdots & W_{1G}^{(0,2)} \\ 0 & W_{22}^{(0,1)} & \cdots & W_{2G}^{(0,1)} & 0 & W_{22}^{(0,2)} & \cdots & W_{2G}^{(0,2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{GG}^{(0,1)} & 0 & 0 & \cdots & W_{GG}^{(0,2)} \\ \hline W_{11}^{(0,3)} & W_{12}^{(0,3)} & \cdots & W_{1G}^{(0,3)} & W_{11}^{(0,4)} & W_{12}^{(0,4)} & \cdots & W_{1G}^{(0,4)} \\ 0 & W_{22}^{(0,3)} & \cdots & W_{1G}^{(0,3)} & 0 & W_{22}^{(0,4)} & \cdots & W_{1G}^{(0,4)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{GG}^{(0,3)} & 0 & 0 & \cdots & W_{GG}^{(0,4)} \end{array}\right) \end{array} \begin{array}{l} \\ \tilde{k}_1 \\ \tilde{k}_2 \\ \vdots \\ \tilde{k}_G \\ k_1-\tilde{k}_1 \\ k_2-\tilde{k}_2 \\ \vdots \\ k_G-\tilde{k}_G \end{array}. \tag{25}$$

The orthogonal matrix $\tilde{P}$ in (22b) computes the RQ decomposition of (25) using a sequence of $(G+1)$ orthogonal factorizations. That is, $\tilde{P} = \tilde{P}^{(0)} \tilde{P}^{(1)} \ldots \tilde{P}^{(G)}$, where $\tilde{P}^{(i)} \in \mathbb{R}^{K^* \times K^*}$ $(i = 0, \ldots, G)$ is orthogonal. Initially, $\tilde{P}^{(0)}$ triangularizes the blocks of the main block-diagonal of $W^{(0)}$. I.e. $\tilde{P}^{(0)} = \operatorname{diag}(\tilde{P}_{11}^{(0)}, \ldots, \tilde{P}_{1G}^{(0)}, \tilde{P}_{41}^{(0)}, \ldots, \tilde{P}_{4G}^{(0)})$, where the RQ decomposition of $W_{jj}^{(0,i)}$ is given by $W_{jj}^{(1,i)} = W_{jj}^{(0,i)} \tilde{P}_{ij}^{(0)}$ $(i = 1, 4$ and $j = 1, \ldots, G)$. Here, $W_{jj}^{(1,i)}$ is triangular. The matrix

$$W^{(0)} \tilde{P}^{(0)} = W^{(1)} \equiv \begin{pmatrix} W^{(1,1)} & W^{(1,2)} \\ W^{(1,3)} & W^{(1,4)} \end{pmatrix}$$

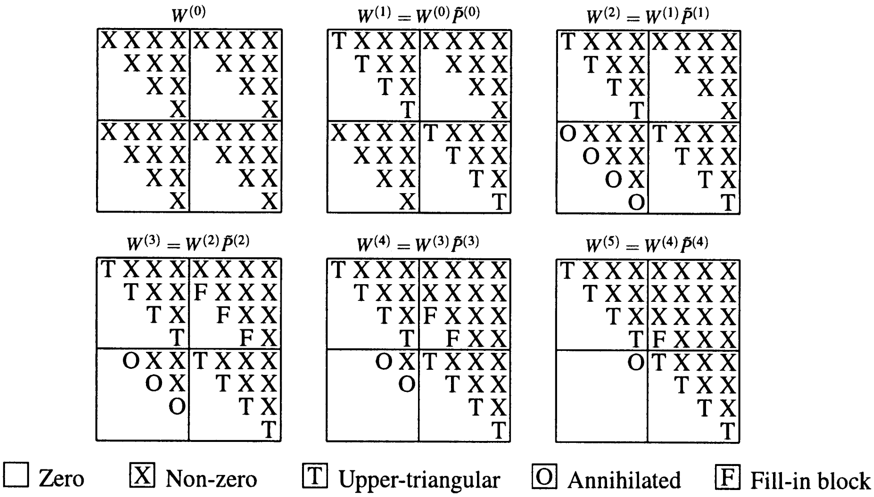
$W^{(0)}$
$W^{(1)} = W^{(0)}\tilde P^{(0)}$
$W^{(2)} = W^{(1)}\tilde P^{(1)}$
$W^{(3)} = W^{(2)}\tilde P^{(2)}$
$W^{(4)} = W^{(3)}\tilde P^{(3)}$
$W^{(5)} = W^{(4)}\tilde P^{(4)}$
Zero
X Non-zero
T Upper-triangular
O Annihilated
F Fill-in block

Fig. 1. Re-triangularization of $W^{(0)}$ in (25)

has the same structure as in (25), but with $W^{(1,1)}$ and $W^{(1,4)}$ being triangular. The transformation $W^{(i+1)} = W^{(i)}\tilde P^{(i)}$ annihilates the ith super block-diagonal of $W^{(i,3)}$, i.e. the block $W^{(i,3)}_{j,j+i-1}$ $(j = 1, \ldots, G-i+1)$, and preserves the triangular structure of $W^{(1,1)}$ and $W^{(1,4)}$. Specifically, $\tilde P^{(i)}$ is defined by

$$\tilde P^{(i)} = \left(\begin{array}{cccc|cccc}
I_{J_i} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & \tilde P_1^{(i,1)} & \cdots & 0 & \tilde P_1^{(i,2)} & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & \tilde P_{G-i+1}^{(i,1)} & 0 & \cdots & \tilde P_{G-i+1}^{(i,2)} & 0 \\
\hline
0 & \tilde P_1^{(i,3)} & \cdots & 0 & \tilde P_1^{(i,4)} & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & \tilde P_{G-i+1}^{(i,3)} & 0 & \cdots & \tilde P_{G-i+1}^{(i,4)} & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 & I_{\rho_i}
\end{array}\right),$$

where $J_i = \sum_{j=1}^{i-1}\tilde k_i$ and $\rho_i = \sum_{j=1}^{i-1}(k_i - \tilde k_i)$. Figure 1 shows the process of re-triangularizing $W^{(0)}$, where $G = 4$.

Let $\tilde k_1 = \ldots = \tilde k_G \equiv K/2$. The complexities of computing the GQRD (22) using this variable-downdating method and that which does not exploit the structure of the matrices are given, respectively, by:

$$GK^2\big(8G^2(K+1) + G(31K+12) + 7(15K+4)\big)/24 \approx O(G^3K^3/3),$$

and

$$2G^2K^2(11GK/3 + 1) \approx O(22G^3K^3/3).$$

Thus, the proposed variable-downdating method is approximately 22 times faster. The number of flops required for the estimation of all subset VAR models using a simple enumeration strategy is of $O((G^3K^3/3)2^{GK})$.

4 Deriving the subset VAR models

All possible subset VAR models can be generated by moving from one model to another. Consider deleting only the μth variable from jth block of the reduced ZR–VAR model (9). This is equivalent to the SUR model in (19), where $\tilde{S}_j = (e_1 \ldots e_{\mu-1}\ e_{\mu+1} \ldots e_{k_j})$, e_l is the lth column of the identity matrix I_{k_j}, $\tilde{S}_i = I_{k_i}$, $\tilde{\beta}_i = \beta_i$, $\tilde{k}_i = k_i$, for $i = 1, \ldots, G$ and $i \neq j$. Thus, the ZR–VAR model to be estimated is equivalent to the SUR model

$$\begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_j \\ \vdots \\ \tilde{y}_G \end{pmatrix} = \begin{pmatrix} R^{(1)} & & & & \\ & \ddots & & & \\ & & R^{(j)}\tilde{S}_j & & \\ & & & \ddots & \\ & & & & R^{(G)} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \tilde{S}_j^T\beta_j \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} \tilde{u}_1 \\ \vdots \\ \tilde{u}_j \\ \vdots \\ \tilde{u}_G \end{pmatrix}. \tag{26}$$

The BLUE of (26) comes from the solution of the GLLSP in (21). Now, let W_{11} be partitioned as

$$W_{11} = \begin{array}{c} \begin{array}{cccccc} k_1 & k_2 & \cdots & k_j & \cdots & k_G \end{array} \\ \left(\begin{array}{cccccc} \Gamma_{1,1} & \Gamma_{1,2} & \cdots & \Gamma_{1,j} & \cdots & \Gamma_{1,G} \\ 0 & \Gamma_{2,2} & \cdots & \Gamma_{2,j} & \cdots & \Gamma_{2,G} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Gamma_{j,j} & \cdots & \Gamma_{j,G} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \Gamma_{G,G} \end{array}\right) \begin{array}{c} k_1 \\ k_2 \\ \vdots \\ k_j \\ \vdots \\ k_G \end{array} \end{array}.$$

The computation of the GQRD (22) can be efficiently derived in two stages. The first stage, initially computes the QRD

$$\check{Q}^T\tilde{R}^{(j)} = \begin{pmatrix} \tilde{R}_j \\ 0 \end{pmatrix}$$

and the product $\check{Q}^T(\Gamma_{j,j} \cdots \Gamma_{j,G}) = (\Gamma^*_{j,j} \cdots \Gamma^*_{j,G})$. Then, it computes the RQD $\Gamma^*_{j,j}\check{P} = \tilde{\Gamma}^*_{j,j}$ and the product $(\Gamma^T_{1,j} \cdots \Gamma^T_{j-1,j})^T\check{P} = (\tilde{\Gamma}^T_{1,j} \cdots \tilde{\Gamma}^T_{j-1,j})^T$. Here, the

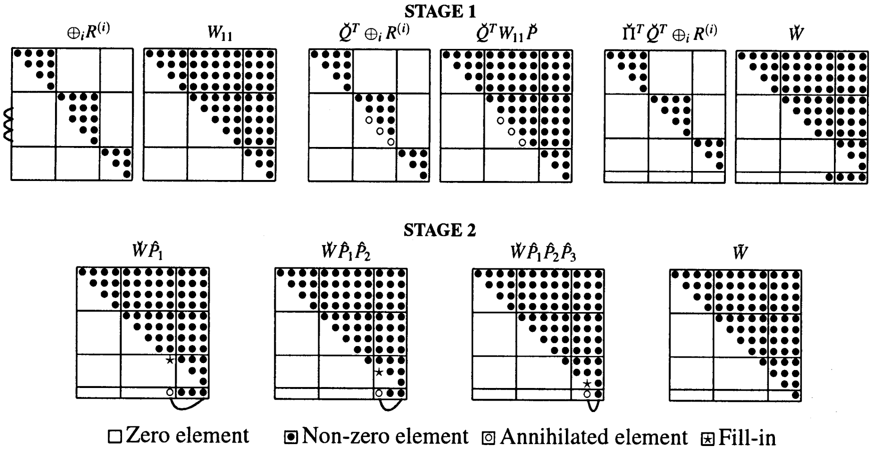
STAGE 1
$\oplus_i R^{(i)}$
W_{11}
$\check{Q}^T \oplus_i R^{(i)}$
$\check{Q}^T W_{11} \check{P}$
$\check{\Pi}^T \check{Q}^T \oplus_i R^{(i)}$
$\check{W}$
STAGE 2
$\check{W}\hat{P}_1$
$\check{W}\hat{P}_1\hat{P}_2$
$\check{W}\hat{P}_1\hat{P}_2\hat{P}_3$
$\tilde{W}$
☐ Zero element
▣ Non-zero element
▣ Annihilated element
⊞ Fill-in

Fig. 2. The two stages of estimating a ZR–VAR model after deleting one variable

orthogonal $\check{Q}^T$ and $\check{P}$ are the products of $k_i - \mu$ left and right Givens rotations which re-triangularize $\tilde{R}^{(j)}$ and $\Gamma^*_{j,j}$, respectively. Furthermore, let

$$\check{\Pi}^T = \begin{pmatrix} I_{K^*_j - 1} & 0 & 0 \\ 0 & 0 & I_{K^*_G - K^*_j} \\ 0 & 1 & 0 \end{pmatrix}, \quad \text{with} \quad K^*_j = \sum_{i=1}^{j} k_i .$$

Thus, in (22a) $\tilde{Q}^T = \check{\Pi}^T \check{Q}^T_*$ and

$$\tilde{Q}^T \oplus_i (R_i \tilde{S}_i) = \begin{pmatrix} \oplus_i \tilde{R}_i \\ 0 \end{pmatrix} \begin{matrix} \tilde{K}^* \\ 1 \end{matrix},$$

where $\check{Q}^T_* = \text{diag}(I^*_{K_{j-1}}, \check{Q}^T, I_{K^* - K^*_j})$ and $K^* \equiv K^*_G$.

The second stage computes the RQD $\tilde{W} = (\check{\Pi}^T \check{W})\hat{P}$, where $\tilde{W}$ and $\check{W}$ are upper-triangular, $\check{W} = \check{Q}^T_* W_{11} \check{P}_*$, $\check{P}_* = \text{diag}(I_{K^*_{j-1}}, \check{P}, I_{K^* - K^*_j})$ and $\hat{P}$ is the product of $(K^* - K^*_j)$ Givens rotations. The ρth rotation ($\rho = 1, \ldots, K^* - K^*_j$), say $\hat{P}_\rho$, annihilates the $(K^*_j + \rho - 1)$th element of the last row of $\check{\Pi}^T \check{W}$ by rotating adjacent planes. Figure 2 illustrates the computation of the two stages, where $G = 3$, $k_1 = 4$, $k_2 = 5$, $k_3 = 3$, $j = 2$ and $\mu = 2$. Notice that in (22b), $\tilde{\Pi}$ is the identity matrix and $\tilde{P} \equiv \check{P}^*$.

Now, let $V = [v_1, v_2, \ldots, v_n]$ denote the set of $n = |V|$ indices of the selected columns (variables) included in the sub-matrices $R^{(i)}$ ($i = 1, \ldots, G$). The sub-models corresponding to the sub-sets $[v_1]$, $[v_1, v_2]$, $\cdots$, $[v_1, v_2, \ldots, v_n]$ are immediately available. A function Drop will be used to derive the remaining sub-models [8]. This function downdates the ZR–VAR by one variable. That is,

$$\text{Drop}(V, i) = [v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n], \quad \text{where } i = 1, \ldots, n-1.$$

An efficient algorithm, called Dropping Columns Algorithm (DCA) has been previously introduced within the context of generating all subset models of the ordinary and general linear models [4, 8, 24]. The DCA generates a regression tree. It moves from one node to another by applying a Drop operation, that is, by deleting a single variable. A formal and detailed description of the regression tree which generates all subset models can be found in [8]. Here the basic concepts using a more convenient notation are introduced.

Let V denote the set of indices and $0 \leq \gamma < |V|$. A node of the regression tree is a tuple (V, γ), where γ indicates that the children of this node will include the first γ variables. If $V = [v_1, v_2, \ldots, v_n]$, then the regression tree $T(V, \gamma)$ is a $(n-1)$–tree having as root the node (V, γ), where $\gamma = 0, \ldots, n-1$. The children are defined by the tuples $(\text{Drop}(V, i), i-1)$ for $i = \gamma+1, \ldots, n-1$. Formally this can be expressed recursively as

$$T(V,\gamma) = \begin{cases} (V,\gamma) & \text{if} \quad \gamma = n-1, \\ \big((V,\gamma), T(\text{Drop}(V,\gamma+1),\gamma), \cdots & \\ \qquad \cdots, T(\text{Drop}(V,n-1),n-2)\big) & \text{if} \quad \gamma < n-2. \end{cases}$$

The number of nodes in the sub-tree $T(\text{Drop}(V, i), i-1)$ is given by $\delta_i = 2^{n-i-1}$ and $\delta_i = 2\delta_{i+1}$, where $i = 1, \ldots, n-1$ [8].

Computing all possible subset regressions of a model having n variables is equivalent to generating $T(V, 0)$, where $V = [1, 2, \ldots, n]$. Generally, the complexity –in terms of flops– of generating $T(V, \gamma)$ in the General Linear model case is of $O((|V| + \gamma)2^{|V|-\gamma})$. Thus the complexity of generating all subset VAR models using the DCA is of $O((GK)2^{GK})$. Figure 3 shows $T(V, \gamma)$ together with the sub-models generated from each node, where $V = [1, 2, 3, 4, 5]$ and $\gamma = 0$. A sub-model is denoted by a sequence of numerals which correspond to the indices of variables.

The DCA will generate all the subset VAR models by deleting one variable from the upper-triangular regressors of the reduced-size model (8). It avoids estimating each ZR–VAR model afresh, i.e. it derives efficiently the estimation of one ZR–VAR model from another after deleting a single variable. Algorithm 1 summarizes this procedure.

Algorithm 1 Generating the regression tree $T(V, \gamma)$ given the root node (V, γ).

```
procedure SubTree($V, \gamma$)
    From $(V, \gamma)$ obtain the the sub–models $(v_1 \cdots v_{\gamma+1}), \ldots, (v_1 \cdots v_{|V|})$
    for $i = \gamma + 1, \ldots, |V| - 1$ do
        $V^{(i)} \leftarrow \text{Drop}(V, i)$
        SubTree($V^{(i)}, i - 1$)
    end for
end procedure
```

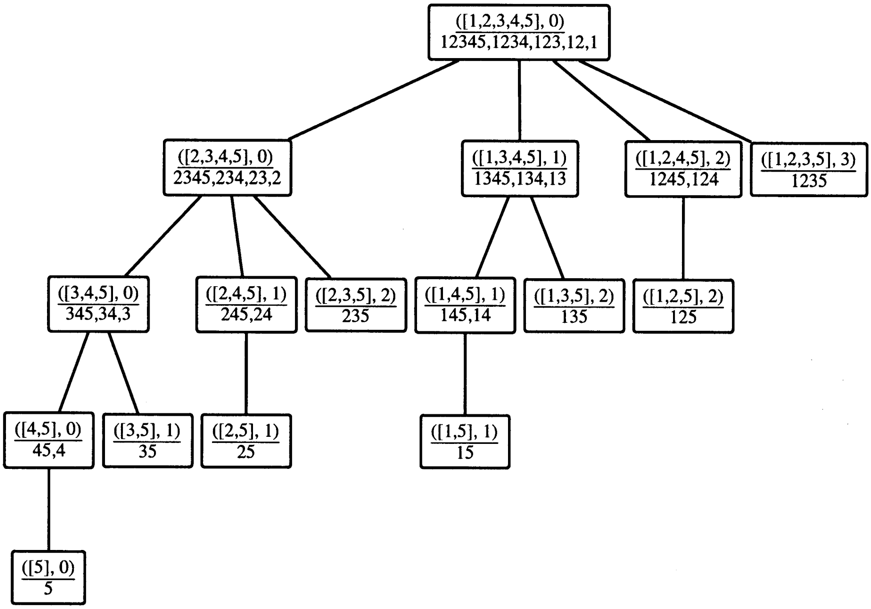

Fig. 3. The regression tree $T(V, \gamma)$, where $V = [1, 2, 3, 4, 5]$ and $\gamma = 0$

5 Special cases

The method of generating all subset VAR models becomes rapidly infeasible when the dimensions of the generating process (1), i.e. G and p increase. Thus, two approaches can be envisaged. The first is to compare models from a smaller given search space. The second is the use of heuristic optimization techniques [28]. Here, the former approach is considered.

A simplified approach is to consider a block-version of Algorithm 1, i.e. a ZR–VAR model is derived by deleting a block rather than a single variable. Within this context, in Figure 3 the numerals will represent indices of blocks of variables. This approach will generate $2^G - 1$ subset VAR models and can be implemented using fast block-downdating algorithms [11]. Notice that the deletion of the entire jth block is equivalent in deleting the jth row from all $\Phi_1, \ldots, \Phi_p$. This is different than the method in [18, pp. 180] where a whole coefficient matrix Φ_i $(i = 1, \ldots, p)$ is deleted at one time.

5.1 Deleting identical variables

Deleting the same variables from all the G blocks of the ZR–VAR model corresponds to deletion of whole columns from some of the coefficient matrices $\Phi_1, \ldots, \Phi_p$. This is equivalent to the SUR model in (9), where $S_i \equiv S \in \mathbb{R}^{K \times \tilde{k}}$

for $i = 1, \ldots, G$ and $0 \leq \tilde{k} < K$. Thus, (9) can be written as

$$\begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_G \end{pmatrix} = \begin{pmatrix} RS & & \\ & \ddots & \\ & & RS \end{pmatrix} \begin{pmatrix} S^T b_1 \\ \vdots \\ S^T b_G \end{pmatrix} + \begin{pmatrix} \tilde{u}_1 \\ \vdots \\ \tilde{u}_G \end{pmatrix}. \qquad (27)$$

The estimation of (27) comes from the solution of GLLSP (10), where now, $R^{(i)} = RS$, $\beta_i = S^T b_i$ and $k_i = \tilde{k}$ for $i = 1, \ldots, G$. The orthogonal matrices in (12) are identical, i.e. $Q_i^T = \check{Q}^T$ for $i = 1, \ldots, G$ and have the structure

$$\check{Q}^T = \begin{pmatrix} \check{Q}_A^T \\ \check{Q}_B^T \end{pmatrix} \begin{matrix} \tilde{k} \\ K - \tilde{k} \end{matrix} .$$

Multiplying respectively, Q^T and Q from the left and right of $(C \otimes I_K)$ it results

$$Q^T(C \otimes I_K)Q = \begin{pmatrix} C \otimes W_A \\ C \otimes W_B \end{pmatrix} (\oplus_i \check{Q}_A \quad \oplus_i \check{Q}_B) = \begin{pmatrix} C \otimes I_{\tilde{k}} & 0 \\ 0 & C \otimes I_{K-\tilde{k}} \end{pmatrix}.$$

The latter is upper-triangular. Figure 4 shows the computation of $Q^T(C \otimes I_K)Q$, where $G = 3$, $K = 5$ and $\tilde{k} = K - 1$. In this case, the permutation matrix Π in (11) is not required, i.e. $\Pi = I_{KG}$ and the matrix $P \equiv Q$. Notice that for the construction of Q in (11) only a $K \times K$ orthogonal matrix $\check{Q}$ needs to be computed rather than the $K \times K$ matrices $Q_1, \ldots, Q_G$.

The DCA can be modified to generate the subset VAR models derived by deleting identical variables from each block. Given a node (V, γ), the set V denotes the indices of the non-deleted (selected) variables. The parameter γ has the same definition as in section 4. The model (27) is estimated. This provides $G|V| - \gamma$ sub-leading VAR models. Then, one variable is deleted, specifically $V^{(i)} \equiv [v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{|V|}]$ for $i = \gamma + 1, \ldots, |V|$. The procedure is recursively repeated for $V^{(\gamma+1)}, \ldots, V^{(|V|)}$.

This method is summarized by Algorithm 2 which generates a regression tree of $2^K - 1$ nodes. Each node corresponds to one of the possible combination of selecting variables out of K. In general, the regression tree with the root node (V, γ) has $2^{|V|-\gamma} - 1$ nodes and provides $2^{|V|-\gamma-1}(|V| + \gamma + 2) - 1$ subset VAR models. Figure 5 shows the regression tree for the case $K = 4$ and $G = 2$. Each node shows (V, γ) and the indices of the corresponding subset VAR model in (27) together with its sub-leading models.

Notice that Algorithm 2 generates all the subset VAR models by deleting the same variables from each block when initially $V \equiv [1, \ldots, K]$ and $\gamma = 0$. Compared to the standard variable-deleting strategy of Algorithm 1, it requires $O(G)$ less computational complexity in order to generate $(K + 2)2^{K-1} - 1$ out of the $2^{KG} - 1$ possible subset VAR models.

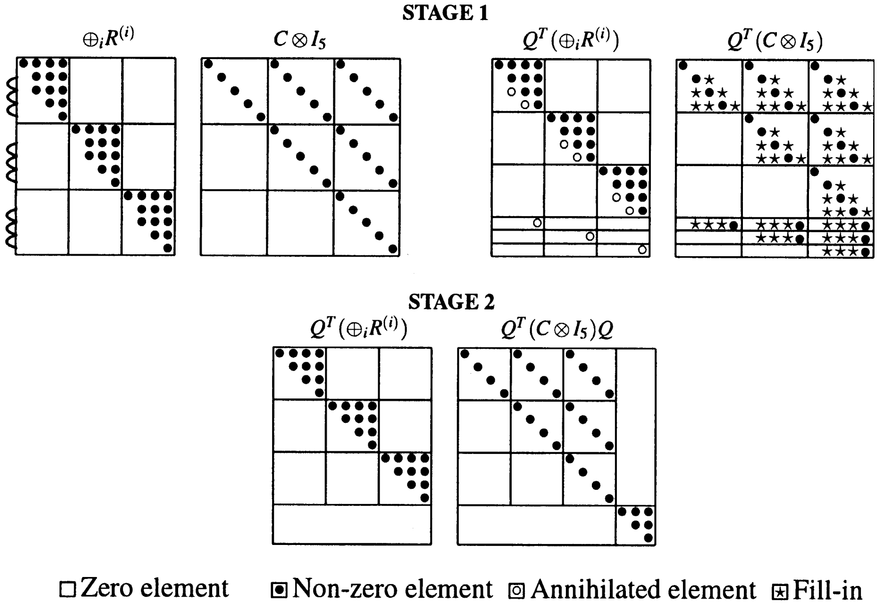

Fig. 4. The two stages of estimating a SUR model after deleting the same variable from each block

Algorithm 2 Generating the subset VAR models by Deleting Identical Variables (DIV).

```
procedure SubTree_DIV($V, \gamma$)
    Let the selection matrix $S \equiv [e_{v_1} \cdots e_{v_{|V|}}]$
    Estimate the subset VAR model $\mathrm{vec}(\tilde{Y}) = (I_K \otimes RS)\,\mathrm{vec}(\{S^T b_i\}) + \mathrm{vec}(\tilde{U})$
    for $i = \gamma + 1, \ldots, |V|$ do
        $V^{(i)} \equiv [v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{|V|}]$
        if ($|V^{(i)}| > 0$) then SubTree_DIV($V^{(i)}, i - 1$) end if
    end for
end procedure
```

5.2 Deleting subsets of variables

The computational burden is reduced when subsets of variables are deleted from the blocks of the ZR–VAR model (9). Consider the case of proper subsets and specifically when

$$S_i = \begin{array}{c} \begin{array}{cc} k_{i+1} & k_i - k_{i+1} \end{array} \\ \begin{pmatrix} S_{i+1} & S_i^* \end{pmatrix} \end{array}, \quad i = 1, \ldots, G-1. \tag{28}$$

From the QRD of $R^{(1)}$

$$Q_1^T R^{(1)} = \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \begin{array}{l} k_1 \\ K-k_1 \end{array}, \tag{29}$$

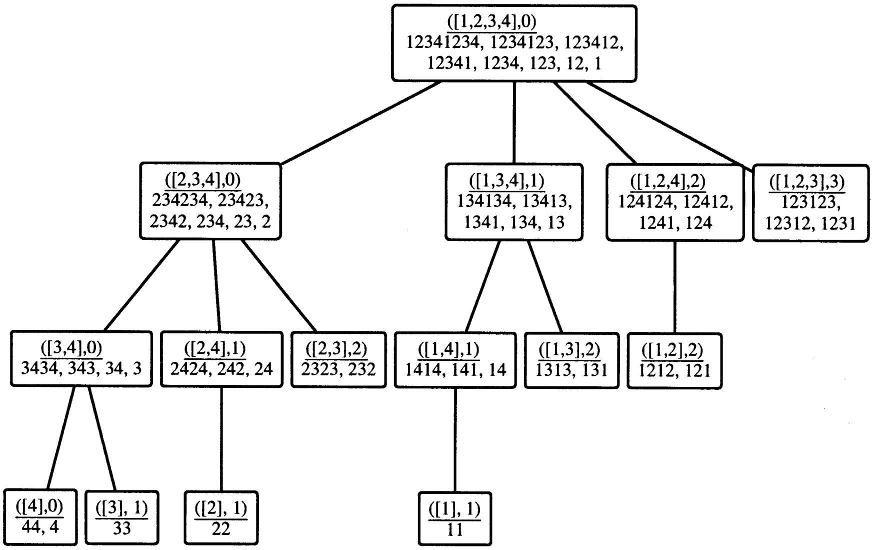

Fig. 5. The regression tree $T(V, \gamma)$, where $V = [1, 2, 3, 4]$, $K = 2$, $G = 2$ and $\gamma = 0$

it follows that the QRD of $R^{(i)}$ can be written as

$$Q_1^T R^{(i)} = \begin{pmatrix} R_i \\ 0 \end{pmatrix} \quad \text{with} \quad Q_1^T = \begin{pmatrix} Q_{Ai}^T \\ Q_{Bi}^T \end{pmatrix} \begin{matrix} k_i \\ K-k_i \end{matrix}, \quad i = 1, \ldots, G, \tag{30}$$

where R_i is the leading triangular $k_i \times k_i$ sub-matrix of R_1. Now, the GLLSP (10) can be written as

$$\begin{aligned} &\underset{V,\{\beta_i\}}{\operatorname{argmin}} \| Q^T \operatorname{vec}(V) \| \quad \text{subject to} \\ &\qquad Q^T \operatorname{vec}(Y) = Q^T \oplus_i (R^{(i)} \beta_i) + Q^T (C \otimes I_K) Q Q^T \operatorname{vec}(V), \end{aligned} \tag{31}$$

where $Q^T = (\oplus_i Q_{Ai}^T \quad \oplus_i Q_{Bi}^T)$. The latter is equivalent to (14), but with $\tilde{v}_{Ai} = Q_{Ai}^T v_i$, $\tilde{v}_{Bi} = Q_{Bi}^T v_i$. Furthermore, the triangular W_{pq} $(p, q = 1, 2)$ can be partitioned as

$$W_{11} = \left(\begin{array}{c|c|c|c} c_{11} I_{k_1} & \begin{matrix} c_{12} I_{k_2} \\ 0 \end{matrix} & \cdots & \begin{matrix} c_{1G} I_{k_G} \\ 0 \end{matrix} \\ \hline 0 & c_{22} I_{k_2} & \cdots & \begin{matrix} c_{2G} I_{k_G} \\ 0 \end{matrix} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & c_{GG} I_{k_G} \end{array}\right), \quad W_{12} = \left(\begin{array}{c|c|c|c} 0 & \begin{matrix} 0 & 0 \\ c_{12} I_{k_1-k_2} & 0 \end{matrix} & \cdots & \begin{matrix} 0 & 0 \\ c_{1G} I_{k_1-k_G} & 0 \end{matrix} \\ \hline 0 & 0 & \cdots & \begin{matrix} 0 & 0 \\ c_{2G} I_{k_2-k_G} & 0 \end{matrix} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & 0 \end{array}\right),$$

$$W_{21} = 0 \qquad \text{and} \qquad W_{22} = \left(\begin{array}{c|c|c|c} c_{11} I_{K-k_1} & 0 \;\; c_{12} I_{K-k_1} & \cdots & 0 \;\; c_{1G} I_{K-k_1} \\ \hline 0 & c_{22} I_{K-k_2} & \cdots & 0 \;\; c_{2G} I_{K-k_2} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & c_{GG} I_{K-k_G} \end{array}\right).$$

This simplifies the computation of the estimation of the ZR–VAR model (9) [15, 16]. Expression (16) becomes

$$\tilde{v}_{Bi} = (\tilde{y}_{Bi} - \sum_{j=i+1}^{G} c_{ij}(0 \quad I_{K-k_i})\tilde{v}_{Bj})/c_{ii}, \qquad (i = G, \ldots, 1) \tag{32a}$$

and the estimation $\hat{\beta}_i$ $(i = 1, \ldots, G)$ is computed by solving the triangular system

$$R_i \hat{\beta}_i = \big(\tilde{y}_{Ai} - \sum_{j=i+1}^{G} c_{ij} \begin{pmatrix} 0 & 0 \\ I_{k_i - k_j} & 0 \end{pmatrix} \tilde{v}_{Bj}\big). \tag{32b}$$

The proper subsets VAR models are generated by enumerating all the possible selection matrices in (28) and estimating the corresponding models using (32). This enumeration can be obtained by considering all the possibilities of deleting variables on the first block, i.e. generating S_1 and then constructing the remaining selection matrices $S_2, \ldots, S_G$ conformly with (28).

This method is summarized by Algorithm 3 and consists of two procedures. The first, SubTreeM, is the modified SubTree procedure of Algorithm 1. It generates the regression tree as shown in Figure 3. In addition, for each node (V, γ), the ProperSubsets procedure is executed. The latter performs no factorization, but computes the estimated coefficients using (32). Specifically, it derives all possible proper subsets $(S_1, \ldots, S_G)$ in (28), for $S_1 = [e_{v_1}, \ldots, e_{v_{\gamma+1}}], \ldots, [e_{v_1}, \ldots, e_{v_{\gamma+1}}, \ldots, e_{v_{|V|}}]$. The ProperSubsets procedure is based on a backtracking scheme. That is, given $S_1, \ldots, S_{i-1}$ $(i = 1, \ldots, G)$, it generates a new S_i and increments i. If this is not possible, then it performs a backtracking step, i.e. it decrements i and repeats the procedure. As shown in the Appendix, the number of proper subsets VAR models generated by Algorithm 3 is given by

$$f(K, G) = \begin{cases} 2^K - 1 & \text{if} \quad G = 1; \\ \sum_{i=1}^{\min(K, G-1)} C_{G-2}^{i-1} C_K^i 2^{K-i} & \text{if} \quad G \geq 2, \end{cases}$$

where $C_n^k = n!/(k!(n-k)!)$.

The order of generating the models is illustrated in Figure 6 for the case $K = 4$ and $G = 3$. The highlighted models are the common models which are generated when the proper subsets are in increasing order, i.e.

$$S_{i+1} = \begin{array}{c} \\ \end{array}\!\!\begin{array}{cc} k_i & k_{i+1} - k_i \\ \left(S_i \right. & \left. S_{i+1}^* \right) \end{array}, \quad i = 2, \ldots, G.$$

In this case the ZR–VAR model is permuted so that (28) holds, and thus, Algorithm 3 can be employed.

The computational burden of the ProperSubsets procedure can be further reduced by utilizing previous computations. Assume the proper subsets VAR model

Algorithm 3 Generating the subset VAR models by deleting proper subsets of variables.

```
procedure SubTreeM($V, \gamma$)
   Compute the QRD (29) for $S_1 = [e_{v_1}, \dots, e_{v_{|V|}}]$
   ProperSubsets($V, \gamma$)
   for $i = \gamma + 1, \dots, |V| - 1$ do
      $V^{(i)} \leftarrow$ Drop($V, i$)
      SubTreeM($V^{(i)}, i - 1$)
   end for
end procedure

procedure ProperSubsets($V, \gamma$)
   Let $S_1 \leftarrow [e_{v_1}, \dots, e_{v_\gamma}]$;   $k_1 \leftarrow \gamma$;   $i \leftarrow 1$
   while ($i \geq 1$) do
      if ($i = 1$ and $k_1 < |V|$) or ($i > 1$ and $k_i < k_{i-1}$) then
         $k_i \leftarrow k_i + 1$;   $S_i(k_i) \leftarrow e_{v_{k_i}}$
         if ($i = G$) then
            Extract $R_1, \dots R_G$ in (30) from $R_1$ in (29) corresponding to $(S_1, \dots S_G)$
            Solve the GLLSP (31) using (32)
         else
            $i \leftarrow i + 1$;   $k_i \leftarrow 0$
         end if
      else
         $i \leftarrow i - 1$
      end if
   end while
end procedure
```

corresponding to $(S_1, \dots, S_{G-1}, S_G)$ has been estimated. Consider now the estimation of the proper subsets VAR model corresponding to $(S_1, \dots, S_{G-1}, \tilde{S}_G)$, where $\tilde{S}_G = (S_G \quad e_{v_{k_G+1}})$. For example, in Figure 6, this is the case when moving from step 15 to step 16. Let $\tilde{y}_{BG} = (\psi_{BG} \quad \tilde{\tilde{y}}_{BG})^T$ and $\tilde{v}_{BG} = (\upsilon_{BG} \quad \tilde{v}^*_{BG})^T$, where $\tilde{\tilde{y}}_{BG}, \tilde{v}^*_{BG} \in \mathbb{R}^{K-\tilde{k}_G}$ and $\tilde{k}_G = k_G + 1$. That is ψ_{BG} and υ_{BG} are the first elements of $\tilde{y}_{BG}$ and $\tilde{v}_{BG}$, respectively. Notice that from $\tilde{k}_G = k_G + 1$, it implies that $K - k_i < K - k_G$ for $i = 1, \dots, G - 1$. Thus, in (32a), υ_{BG} corresponds to a zero entry and therefore,

$$\tilde{v}_{Bi} = \Big(\tilde{y}_{Bi} - \sum_{j=i+1}^{G-1} c_{ij}(0 \ I_{K-k_i})\tilde{v}_{Bj} - c_{iG}(0 \ I_{K-k_i})\tilde{v}^*_{BG}\Big)/c_{ii} \quad \text{for} \quad i = G-1, \dots, 1.$$

The recursive updating formulae (32a) become

$$\tilde{\tilde{v}}_{BG} = \tilde{\tilde{y}}_{BG}/c_{GG} \equiv \tilde{v}^*_{BG} \quad \text{and} \quad \tilde{\tilde{v}}_{Bi} = \tilde{v}_{Bi} \quad \text{for} \quad i = G-1, \dots, 1.$$

Now,

$$\tilde{\tilde{y}}_i = \tilde{y}_{Ai} - \sum_{j=i+1}^{G-1} c_{ij} \begin{pmatrix} 0 & 0 \\ I_{k_i-k_j} & 0 \end{pmatrix} \tilde{v}_{Bj} - c_{iG} \begin{pmatrix} 0 & 0 \\ I_{k_i-k_G-1} & 0 \end{pmatrix} \tilde{v}^*_{BG}$$

Step	Selected variables			Step	Selected variables			Step	Selected variables		
1	1	1	1	**21**	2	2	2	**39**	1 3	1	1
2	1 2	1	1	**22**	2 3	2	2	**40**	1 3	1 3	1
3	1 2	1 2	1	**23**	2 3	2 3	2	**41**	1 3	1 3	1 3
4	1 2	1 2	1 2	**24**	2 3	2 3	2 3	**42**	1 3 4	1	1
5	1 2 3	1	1	**25**	2 3 4	2	2	**43**	1 3 4	1 3	1
6	1 2 3	1 2	1	**26**	2 3 4	2 3	2	**44**	1 3 4	1 3	1 3
7	1 2 3	1 2	1 2	**27**	2 3 4	2 3	2 3	**45**	1 3 4	1 3 4	1
8	1 2 3	1 2 3	1	**28**	2 3 4	2 3 4	2	**46**	1 3 4	1 3 4	1 3
9	1 2 3	1 2 3	1 2	**29**	2 3 4	2 3 4	2 3	**47**	1 3 4	1 3 4	1 3 4
10	1 2 3	1 2 3	1 2 3	**30**	2 3 4	2 3 4	2 3 4	**48**	1 4	1	1
11	1 2 3 4	1	1	**31**	3	3	3	**49**	1 4	1 4	1
12	1 2 3 4	1 2	1	**32**	3 4	3	3	**50**	1 4	1 4	1 4
13	1 2 3 4	1 2	1 2	**33**	3 4	3 4	3	**51**	1 2 4	1	1
14	1 2 3 4	1 2 3	1	**34**	3 4	3 4	3 4	**52**	1 2 4	1 2	1
15	1 2 3 4	1 2 3	1 2	**35**	4	4	4	**53**	1 2 4	1 2	1 2
16	1 2 3 4	1 2 3	1 2 3	**36**	2 4	2	2	**54**	1 2 4	1 2 4	1
17	1 2 3 4	1 2 3 4	1	**37**	2 4	2 4	2	**55**	1 2 4	1 2 4	1 2
18	1 2 3 4	1 2 3 4	1 2	**38**	2 4	2 4	2 4	**56**	1 2 4	1 2 4	1 2 4
19	1 2 3 4	1 2 3 4	1 2 3								
20	1 2 3 4	1 2 3 4	1 2 3 4								

Fig. 6. The sequence of proper subset models generated by Algorithm 3, for $G = 3$ and $K = 4$

$$= \tilde{y}_i + (c_{iG} \upsilon_{BG}) e_{k_G+1},$$

where $\tilde{y}_i = R_i \hat{\beta}_i$, i.e. the righthand-side of (32b). The estimation of the proper subsets VAR model comes from the solution of the triangular systems

$$\tilde{R}_G \hat{\beta}_G^* = \begin{pmatrix} \tilde{y}_{AG} \\ \psi_{BG} \end{pmatrix} \quad \text{and} \quad R_i \hat{\beta}_i^* = \tilde{y}_i, \quad \text{for} \quad i = G-1, \ldots, 1.$$

The computational cost of the QRDs (12) is also reduced in the general subsets case where $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_G$. The QRD of $R^{(i)} = RS_i$ is equivalent to retriangularizing the smaller in size matrix R_{i+1} after deleting columns. Notice that $RS_i = RS_{i+1}S_{i+1}^T S_i$ and

$$\begin{aligned} Q_{i+1}^T RS_i &= Q_{i+1}^T RS_{i+1} S_{i+1}^T S_i \\ &= \begin{pmatrix} R_{i+1} \\ 0 \end{pmatrix} S_{i+1}^T S_i \\ &= \begin{pmatrix} R_{i+1} S_i^* \\ 0 \end{pmatrix}. \end{aligned}$$

Here $S_i^* = S_{i+1}^T S_i$ is of order $k_{i+1} \times k_i$ and selects the subset of S_{i+1} and in turn the selected columns from R_{i+1}. Now, computing the QRD

$$\hat{Q}_i^T (R_{i+1} S_i^*) = \begin{pmatrix} R_i \\ 0 \end{pmatrix} \begin{matrix} k_i \\ k_{i+1} - k_i \end{matrix} \tag{33}$$

it follows that the orthogonal Q_i^T of the QRD (12) is given by $Q_i^T = \check{Q}_i^T Q_{i+1}^T$, where

$$\check{Q}_i^T = \begin{pmatrix} \hat{Q}_i^T & 0 \\ 0 & I_{K-k_{i+1}} \end{pmatrix}.$$

Thus, following the initial QRD of $R^{(G)} = RS_G$, the remaining QRDs of $R^{(i)}$ are computed by (33) for $i = G-1, \ldots, 1$.

Consider the case where $S_G \subseteq \cdots \subseteq S_2 \subseteq S_1$. Computations are simplified if the GLLSP (10) is expressed as

$$\underset{V,\{\beta_i\}}{\operatorname{argmin}} \|V\|_F^2 \quad \text{subject to } \operatorname{vec}(\tilde{Y}) = (\oplus_i L^{(i)}) \operatorname{vec}(\{\beta_i\}) + \operatorname{vec}(VC^T), \tag{34}$$

where $L^{(i)} = LS_i$ and now, L and C are lower triangular [15]. Thus, instead of (6), the QL decomposition of X needs to be computed:

$$\bar{Q}^T X = \begin{pmatrix} 0 \\ L \end{pmatrix} \begin{matrix} M-K \\ K \end{matrix},$$

with

$$\bar{Q}^T (Y \quad U) = \begin{pmatrix} \hat{Y} & \hat{U} \\ \tilde{Y} & \tilde{U} \end{pmatrix} \begin{matrix} M-K \\ K \end{matrix}.$$

Furthermore, the QL decomposition

$$Q_i^T L^{(i+1)} = \begin{pmatrix} 0 \\ L_{i+1} \end{pmatrix} \begin{matrix} K-k_i \\ k_i \end{matrix}$$

can be seen as the re-triangularization of L_i after deleting columns [14]. If S_{i+1} is a proper subset of S_i, i.e. $S_i = (S_i^* \quad S_{i+1})$, then L_{i+1} is the trailing lower $k_{i+1} \times k_{i+1}$ sub-matrix of L_i [15]. Notice that if (9) rather than (34) is used, then R_{i+1} derives from the more computational expensive (updating) QRD

$$Q_{i+1}^T \begin{pmatrix} R_{i1} \\ R_{i2} \end{pmatrix} = \begin{pmatrix} R_{i+1} \\ 0 \end{pmatrix},$$

where

$$R_i = \begin{array}{c} \begin{matrix} k_i - k_{i+1} & k_{i+1} \end{matrix} \\ \begin{pmatrix} R_i^* & R_{i1} \\ 0 & R_{i2} \end{pmatrix} \begin{matrix} k_i - k_{i+1} \\ k_{i+1} \end{matrix} \end{array}.$$

6 Conclusion and future work

Efficient numerical and computational strategies for deriving the subset VAR models have been proposed. The VAR model with zero-coefficient restriction, i.e. ZR–VAR model, has been formulated as a SUR model. Initially, the QR decomposition is employed to reduce to zero $M - K$ observations of the VAR, and consequently, ZR–VAR model. The numerical estimation of the ZR–VAR model has been derived. Within this context an efficient variable-downdating strategy has been presented. The main computational tool of the estimation procedures is the Generalized QR decomposition. During the proposed selection procedures only the quantities required by the the selection criteria, i.e. the residual sum of squares or the estimated residual covariance matrix, should be computed. The explicit computation of the estimated coefficients is performed only for the final selected models.

An algorithm which generates all subset VAR models by efficiently moving from one model to another has been described. The algorithm generates a regression tree and avoids estimating each ZR–VAR model afresh. However, this strategy is computational infeasible even for modest size VAR models due to the exponential number $(2^{pG^2} - 1)$ of sub-models that derives. An alternative block-version of the algorithm generates $(2^G - 1)$ sub-models. At each step of the block-strategy a whole block of observations is deleted from the VAR model. The deletion of the ith block is equivalent in deleting the ith row from each coefficient matrix $\Phi_1, \ldots, \Phi_p$ in (1).

Two special cases of subset VAR models which are derived by taking advantage of the common-columns property of the data matrix and the Kronecker structure of the variance-covariance matrix have been presented. Both of them require $O(G)$ less computational complexity than generating the models afresh. The first special case derives $(pG+2)2^{(pG-1)} - 1$ subset VAR models by deleting the same variable from each block of the reduced ZR–VAR model. The second case is based on deleting subsets of variables from each block of the regressors in the ZR–VAR model (10). An algorithm that derives all proper subsets models given the initial VAR model has been designed. This algorithm generates $\sum_{i=1}^{\min(K,G-1)} C_{G-2}^{i-1} C_K^i 2^{K-i}$ models, when G is greater than one. In both cases the computational burden of deriving the generalized QR decomposition (11), and thus, estimating the sub-models, is significantly reduced. In the former case only a single column-downdating of a triangular matrix is required. This is done efficiently using Givens rotations. The second case performs no factorizations, but efficiently computes the coefficients using (32).

The new algorithms allow the investigation of more subset VAR models when trying to identify the lag structure of the process in (1). The implementation and application of the proposed algorithms need to be pursued. These methods are based on a regression tree structure. This suggest that a branch and bound strategy which derives the best models without generating the whole regression tree should be considered [7]. Within this context the use of parallel computing to allow the tackling of large scale models merits investigation.

The permutations of the exogenous matrices $X_1, \ldots, X_G$ in the SUR model (5) can provide $G!$ new subset VAR models. If the ZR–VAR model (9) has been already estimated, then the computational cost of estimating these new subset models will be significantly lower since the exogenous matrices $X_1, \ldots, X_G$ have been already factorized ($i = 1, \ldots, G$). Furthermore, the efficient computation of the RQ factorization in (11b) should be investigated for some permutations, e.g. when two adjacent exogenous matrices are permuted. Strategies that generate efficiently all $G!$ sub-models and the best ones using the branch and bound method are currently under investigation.

References

[1] Akaike H (1969) Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21: 243–247

[2] Akaike H (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723

[3] Chao JC, Phillips PCB (1999) Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* 91(2): 227–271

[4] Clarke MRB (1981) Algorithm AS163. A Givens algorithm for moving from one linear model to another without going back to the data. *Applied Statistics* 30(2): 198–203

[5] Foschi P, Kontoghiorghes EJ (2002) Estimation of seemingly unrelated regression models with unequal size of observations: computational aspects. *Computational Statistics and Data Analysis* 41(1): 211–229

[6] Foschi P, Kontoghiorghes EJ (2003) Estimation of VAR models: computational aspects. *Computational Economics* 21(1–2): 3–22

[7] Gatu C, Kontoghiorghes EJ (2002) A branch and bound algorithm for computing the best subset regression models. Technical Report RT-2002/08-1, Institut d'informatique, Université de Neuchâtel, Switzerland

[8] Gatu C, Kontoghiorghes EJ (2003) Parallel algorithms for computing all possible subset regression models using the QR decomposition. *Parallel Computing* 29: 505–521

[9] Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B* 41(2): 190–195

[10] Kontoghiorghes EJ (1999) Parallel strategies for computing the orthogonal factorizations used in the estimation of econometric models. *Algorithmica* 25: 58–74

[11] Kontoghiorghes EJ (2000) *Parallel Algorithms for Linear Models: Numerical Methods and Estimation Problems*, vol 15. *Advances in Computational Economics*. Kluwer Academic Publishers, Boston, MA

[12] Kontoghiorghes EJ (2000) Parallel strategies for solving SURE models with variance inequalities and positivity of correlations constraints. *Computational Economics* 15(1+2): 89–106

[13] Kontoghiorghes EJ (2004) Computational methods for modifying seemingly unrelated regressions models. *Journal of Computational and Applied Mathematics* 162: 247–261

[14] Kontoghiorghes EJ, Clarke MRB (1993) Parallel reorthogonalization of the QR decomposition after deleting columns. *Parallel Computing* 19(6): 703–707

[15] Kontoghiorghes EJ, Clarke MRB (1995) An alternative approach for the numerical solution of seemingly unrelated regression equations models. *Computational Statistics & Data Analysis* 19(4): 369–377

[16] Kontoghiorghes EJ, Dinenis E (1996) Solving triangular seemingly unrelated regression equations models on massively parallel systems. In: Gilli M (ed) *Computational Economic Systems: Models, Methods & Econometrics*, vol 5. *Advances in Computational Economics*, pp 191–201. Dordrecht: Kluwer Academic Publishers

[17] Kourouklis S, Paige CC (1981) A constrained least squares approach to the general Gauss–Markov linear model. *Journal of the American Statistical Association* 76(375): 620–625
[18] Lütkepohl H (1993) *Introduction to Multiple Time Series Analysis*. Berlin Heidelberg New York: Springer
[19] Paige CC (1978) Numerically stable computations for general univariate linear models. *Communications on Statistical and Simulation Computation* 7(5): 437–453
[20] Paige CC (1979) Computer Solution and Perturbation Analysis of Generalized Linear Least Squares Problems. *Mathematics of Computation* 33(145): 171–183
[21] Paige CC (1979) Fast numerically stable computations for generalized linear least squares problems. *SIAM Journal on Numerical Analysis* 16(1): 165–171
[22] Pitard A, Viel JF (1999) A model selection tool in multi-pollutant time series: The Granger-causality diagnosis. *Environmetrics* 10(1): 53–65
[23] Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6(2): 461–464
[24] Smith DM, Bremner JM (1989) All possible subset regressions using the QR decomposition. *Computational Statistics and Data Analysis* 7: 217–235
[25] Srivastava VK, Giles DEA (1987) *Seemingly Unrelated Regression Equations Models: Estimation and Inference (Statistics: Textbooks and Monographs)*, vol 80. New York: Marcel Dekker
[26] Strang G (1976) *Linear Algebra and Its Applications*. New York: Academic Press
[27] Winker P (2000) Optimized multivariate lag structure selection. *Computational Economics* 16(1–2): 87–103
[28] Winker P (2001) *Optimization Heuristics in Econometrics: Applications of Threshold Accepting*. Wiley Series in Probability and Statistics. Chichester: Wiley
[29] Yanev P, Foschi P, Kontoghiorghes EJ (2003) Algorithms for computing the QR decomposition of a set of matrices with common columns. *Algorithmica* 39(1): 83–93 (2004)

Appendix

Lemma 1 *The recurrence*

$$f(K,G)=\begin{cases}0 & \text{if}\quad K=0 \quad \text{and} \quad G\geq 1,\\ 1 & \text{if}\quad K\geq 0 \quad \text{and} \quad G=0,\\ \sum_{i=0}^{K-1}\sum_{j=0}^{G} f(i,j) & \text{if}\quad K\geq 1 \quad \text{and} \quad G\geq 1,\end{cases} \tag{35}$$

denotes the number of proper subsets models defined by (28) *with maximum K variables,* $(v_1,\ldots,v_K)$ *and G blocks.*

Proof The proof is by double induction. For $K=1$ and $G\geq 1$,

$$f(1,G)=\sum_{j=0}^{G} f(0,j)=f(0,0)=1.$$

This is the case where there is only one possible model, i.e. $S_i=[v_1]$, for $i=1,\ldots,G$. The inductive hypothesis is that Lemma 1 is true for some $K,G\geq 1$. It has to be proven that Lemma 1 is also true for $K+1$ and $G\geq 1$.

Let v_{K+1} be the new variable. First, there is a new model defined by $S_i=[v_{K+1}]$, for $i=1,\ldots,G$. Furthermore, from the inductive hypothesis there are $f(K,G)$ models which do not include v_{K+1}. Consider now all the possibilities for which S_j includes v_{K+1}, when $j=1,\ldots,G$. From the proper subsets property, it follows that $S_1,\ldots,S_{j-1}$ include also v_{K+1}. Furthermore, if two of these sets are different, then, before adding v_{K+1}, there are i $(2 \leq i \leq j)$ and $\alpha \geq 1$, so that, $S_{i-1} = [v_1, \dots, v_\rho, v_{\rho+1}, \dots, v_{\rho+\alpha}]$ and $S_i = [v_1, \dots, v_\rho]$. Now, if v_{K+1} is included, then it follows $S_{i-1} = [v_1, \dots, v_\rho, v_{\rho+1}, \dots, v_{\rho+\alpha}, v_{K+1}]$ and $S_i = [v_1, \dots, v_\rho, v_{K+1}]$. However, this contradicts the definition in (28) and therefore $S_1 \equiv \cdots \equiv S_{j-1} \equiv S_j$. Thus, the number of possibilities for which S_j includes v_{K+1} is the number of models with maximum K variables and $G - j + 1$ blocks. From the inductive hypothesis this number is given by $f(K, G - j + 1)$.

Hence, the number of proper subsets models defined by (28) with maximum $K + 1$ variables and G blocks is given by

$$\begin{aligned} 1+f(K,G)+\sum_{j=1}^{G} f(K,G-j+1) &= \sum_{i=0}^{K-1}\sum_{j=0}^{G} f(i,j)+\sum_{j=1}^{G} f(K,G-j+1)+f(K,0) \\ &= \sum_{i=0}^{K}\sum_{j=0}^{G} f(i,j) = f(K+1,G), \end{aligned}$$

which completes the proof.

Lemma 2 *The recurrence* (35) *simplifies to*

$$f(K,G) = \begin{cases} 1 & \textit{if} \quad G = 0, \\ 2^K - 1 & \textit{if} \quad G = 1, \\ \sum_{i=1}^{min(K,G-1)} C_{G-2}^{i-1} C_K^i 2^{K-i} & \textit{if} \quad G \geq 2. \end{cases}$$

Proof Consider the recurrence (35). If $K \geq 1$, then $f(K,G) = \sum_{j=0}^{G-1} f(K-1, j) + 2f(K-1, G)$. Thus, (35) can be written as $F_K^T = F_{K-1}^T \Lambda = F_0^T \Lambda^K$, where $F_K \in \mathbb{R}^{G+1}$, $F_0 = (1, 0, \dots, 0)^T$ and $\Lambda \in \mathbb{R}^{(G+1)\times(G+1)}$ is given by

$$\Lambda = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ & 2 & 1 & \cdots & 1 & 1 \\ & & 2 & \cdots & 1 & 1 \\ & & & \ddots & \vdots & \vdots \\ & & & & 2 & 1 \\ & & & & & 2 \end{pmatrix}.$$

Now, consider the computation of Λ^K which requires the Jordan form of Λ [26, pp. 335–341]. That is, $\Lambda = \Theta \bar{D} \Theta^{-1}$, where $\bar{D} = D + N$, $D \in \mathbb{R}^{(G+1)\times(G+1)}$ is diagonal, $N \in \mathbb{R}^{(G+1)\times(G+1)}$, $N^G = 0$ and $DN = ND$. Specifically

$$D = \begin{pmatrix} 1 & & & & & \\ & 2 & & & & \\ & & 2 & & & \\ & & & \ddots & & \\ & & & & 2 & \\ & & & & & 2 \end{pmatrix} \quad \text{and} \quad N = \begin{pmatrix} 0 & 0 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ & & & & & 0 \end{pmatrix}.$$

The upper-triangular matrix Λ has the two eigenvalues $\lambda = 1$ and $\mu = 2$ with the multiplicities 1 and G, respectively. The eigenvectors $v_1 = (1, 0, 0, \ldots, 0)^T$ and $v_2^{(1)} = (1, 1, 0, \ldots, 0)^T$ corresponds to the eigenvalues λ and μ, respectively. The remaining eigenvectors $v_2^{(i)}$ $(i = 2, \ldots, G)$, corresponding to the multiple eigenvalue μ, are recursively computed by deriving the solution w of $(\Lambda - 2I_{G+1})w = v_2^{(i-1)}$. Thus, $\Theta = [v_1, v_2^{(1)}, \ldots, v_2^{(G)}]$ and is given by

$$\theta_{i,j} = \begin{cases} 1, & \text{if} \quad i = 0,\ j = 0 \vee 1 \text{ or } i = 1,\ j = 0, \\ 0, & \text{if} \quad i = 0 \vee 1,\ j \geq 2 \text{ or } i = 1,\ j = 0 \text{ or } i \geq 2,\ j < i, \\ (-1)^{i+j} C_{j-2}^{i-2}, & \text{if} \quad i \geq 2,\ j \geq i, \end{cases}$$

where $C_n^k = n!/(k!(n-k)!)$. Furthermore Θ^{-1} is given by

$$\theta_{i,j}^{-1} = \begin{cases} 1 \text{ if } i = j = 0 \vee 1, \\ -1 \text{ if } i = 0,\ j = 1, \\ 0 \text{ if } i = 0 \vee 1,\ j \geq 2 \text{ or } i = 1,\ j = 0 \text{ or } i \geq 2,\ j < i, \\ C_{j-2}^{i-2} \text{ if } i \geq 2,\ j \geq i. \end{cases}$$

That is

$$\Theta^{-1} \equiv \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & C_0^0 & C1^0 & C_2^0 & \ldots & C_{G-3}^0 & C_{G-2}^0 \\ 0 & 0 & 0 & C_1^1 & C_2^1 & \ldots & C_{G-3}^1 & C_{G-2}^1 \\ 0 & 0 & 0 & 0 & C_2^2 & \ldots & C_{G-3}^2 & C_{G-2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & C_{G-3}^{G-3} & C_{G-2}^{G-3} \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & C_{G-2}^{G-2} \end{pmatrix}.$$

Now, $\Lambda^K = (\Theta\bar{D}\Theta^{-1})^K = \Theta(D+N)^K\Theta^{-1}$ and

$$(D+N)^K = \begin{cases} C_K^0 D^K + C_K^1 D^{K-1}N + \cdots + C_K^K N^K, & \\ \hfill K < G-1, & \\ C_K^0 D^K + C_K^1 D^{K-1}N + \cdots + C_K^{K-G+1} D^{K-G+1} N^{G-1}, & \\ \hfill K \geq G-1. & \end{cases}$$

For the case $K \geq G-1$, the latter can be written as

$$(D+N)^K = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 2^K C_K^0 & 2^{K-1}C_K^1 & 2^{K-2}C_K^2 & \ldots & 2^{K-G+1}C_K^{G-1} \\ 0 & 0 & 2^K C_K^0 & 2^{K-1}C_K^1 & \ldots & 2^{K-G+2}C_K^{G-2} \\ 0 & 0 & 0 & 2^K C_K^0 & \ldots & 2^{K-G+3}C_K^{G-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 2^K C_K^0 \end{pmatrix}. \tag{36}$$

In the case $K < G-1$, (36) has band-diagonal structure with band-width $K+1$.

Now $F_0 = (1, 0, 0, \ldots, 0)^T$ and $F_K = F_0\Lambda^K$. Thus, only the first row of the matrix Λ^K needs to be computed. Furthermore, the first row of Θ is $(1, 1, 0, \ldots, 0)$ which implies that the first row of the product $\Theta(D+N)^K$, say r, is given by

$$r = \begin{cases} (1, 2^K C_K^0, 2^{K-1}C_K^1, \ldots, 2^0 C_K^K, 0, \ldots, 0), & \text{if} \quad k < G-1 \\ (1, 2^K C_K^0, 2^{K-1}C_K^1, \ldots, 2^{K-G+1}C_K^{K-G+1}), & \text{if} \quad K \geq G-1. \end{cases} \tag{37}$$

From $r\Theta^{-1} = F_K$ it follows that

$$f(K,G) = \begin{cases} 1, & \text{if} \quad G = 0, \\ 2^K - 1, & \text{if} \quad G = 1, \\ \sum_{i=1}^{\min(K,G-1)} C_{G-2}^{i-1} C_K^i 2^{K-i}, & \text{if} \quad G \geq 2, \end{cases}$$

which completes the proof.