Numerical methods for estimating linear econometric models

Paolo Foschi

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy

Institut d’informatique
Université de Neuchâtel
Switzerland

IMPRIMATUR POUR LA THESE

Numerical Methods for estimating linear econometric models

M. Paolo FOSCHI

UNIVERSITE DE NEUCHATEL

FACULTE DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel, sur le rapport des membres du jury

MM. H.-H. Naegeli (directeur de thèse),
E. Kontoghiorghes (co-directeur de thèse), P.-J. Erard,
B. Rustem (Londres UK), B. Philippe (Rennes F)
et M. Gilli (Genève)

autorise l'impression de la présente thèse.

Neuchâtel, le 22 avril 2004

La doyenne:

Martine Rahier

Abstract

The estimation of the Seemingly Unrelated Regressions (SUR) model and its variants is a core area of econometrics. The purpose of this thesis is the investigation and development of efficient numerical and computational methods for solving large scale SUR models. Specifically, its aim is twofold. Firstly to continue past successful research into the design of numerically efficient methods for estimating the basic SUR model. Secondly to extend these methods to variants and special cases of that model.

The basic computational formulae for deriving the estimators of SUR models involve Kronecker products and direct sum of matrices that make the solution of the models computationally expensive even for modest sized models. Alternative numerical methods which substantially reduce the computational burden of the estimation procedures are proposed. Such methods successfully tackle the estimation of the basic SUR model, and that of SUR models derived from VAR(p) processes, SUR models with VAR disturbances, SUR models with unequal size observations and SUR models with orthogonal regressors. The proposed methods, are based on orthogonal transformations, and thus, results to be numerically stable. Furthermore, they do not require the common assumption which is usually made in most theoretical analyses, that the disturbance covariance matrix be non-singular.

To Elisa

Acknowledgments

Firstly, I would like to thank my supervisor, Prof. Erricos J. Kontoghiorghes. He has engaged me for Ph.D. studies in Neuchâtel which gave me the opportunity to work in the interface of Mathematics, Statistics, Economics and Computer Science. He gave me the flexibility to accomplish my research studies without disregarding my family needs. Further, he has always supported me and accustomed to my not always impeccable timing. Most important he has shown to be a real friend.

Then my thanks, goes to Prof. Charles Broyden who, with his peculiar personality, introduced me to the beauty of numerical linear algebra and taught the importance of elegance in mathematics.

I would also like to thank Prof. David A. Belsley (Boston College) who has always encouraged our research and frequently made constructive suggestions and corrections to our work.

I thank Prof. Aristide Mingozzi, Prof. Vittorio Maniezzo and Prof. Sergio Polidoro of the University of Bologna who have supported and provided me with the necessary infrastructure in Italy in order to continue my research.

I thank my friends and colleagues Cristian Gatu and Petko Yanev. Their help during my Ph.D. studies is very much appreciated. They have revealed to be splendid office mates and never refused to join me and Erricos for a coffee. I hope they will extend the research in this thesis in their specific domains.

I wish also to acknowledge the people who assisted me in accomplish the administration duties that are required by the University and State of Neuchâtel. Specifically, I'm referring to Gianfranca Cerrito, Jaroslava Messerli, Prof. Hans-Heinrich Naegeli and Prof. Pierre-Jean Erard.

I could not avoid to acknowledge and thank the Swiss National Science Foundation which financially supported this research through the projects 21-54109.98 and 2000-061875.00.

Contents

Chapter 1

Introduction

The estimation of Seemingly Unrelated Regressions (SUR) models have broad applicability in the analysis and estimation of econometrics models. The SUR model arise in the estimation Simultaneous Equations (SE), Time Series and Panel Data models, to name just a few [4, 15, 58, 79]. Procedures that provide theoretically efficient estimators to SUR models with special properties and the theory of inference for SUR models have been an active research area in econometrics for more than forty years [80, 86, 89]. The computational and numerical aspects of the various proposed estimation procedures have been investigated only recently [16, 45, 46, 48, 50, 51, 53]. Eventhough, the most commonly used estimation procedures are based on the direct implementation of theoretical formulae, which are computationally expensive even for modest sized problems, and gives meaningless results for models with ill-conditioned matrices [6, 76]. For example, in the case of a SUR model of 10 equations with an average of 8 regressors and 100 observations each, the estimation problem can be seen as equivalent to a General Linear Model (GLM) of 1000 observations and 80 variables.

When the disturbance covariance matrix is known, the most commonly used estimator for the SUR model is the Generalized Least Squares (GLS) estimator. This estimator gives a Best Linear Unbiased Estimator (BLUE). Otherwise, when the covariance matrix is unknown, the iterative Feasible GLS (FGLS) and Maximum Likelihood (ML) procedures are used. The FGLS and the ML estimators come from the solution of normal equations that involve Kronecker products, direct sums and with the unknown disturbance covariance matrix replaced at each iteration by an estimator until convergence has been achieved [3, 66, 69, 80, 83].

The equally important development of numerical and computational tools for solving SUR models lags behind the theoretical advances made in econometrics. Algorithms for computing the BLUE of the SUR model usually require the disturbance covariance matrix to be non-singular, eventhough this is not the case in many economic applications [38, 84].

1.1 Linear Models

A common problem in statistics is that of estimating parameters of some assumed relationship between one or more variables. A linear model is one relationship in which a dependent (endogenous, explained) variable y can be expressed as a linear function of independent (exogenous, explanatory) variables $x_1, \ldots, x_n$. When there are m samples observations this relationship can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}.$$

where ε_i is an error term for which specific values cannot be predicted. In compact form the latter can be written as

$$y = X\beta + \varepsilon, \tag{1.1}$$

where $y, \varepsilon \in \mathbb{R}^m$, $X \in \mathbb{R}^{m \times n}$ and $\beta \in \mathbb{R}^n$. Additional assumptions should be specified to complete the linear model (1.1). The first assumption is that the expected value of ε is zero, that is, $\mathrm{E}(\varepsilon) = 0$. The second assumption is that X is a non stochastic matrix, and thus $\mathrm{E}(X^T \varepsilon) = 0$. The final assumption is that the variance-covariance matrix of ε is $\sigma^2 \Omega$, where Ω is a symmetric non negative definite matrix and σ is an unknown scalar. In summary the complete mathematical specification of the (general) Linear Model (GLM) is given by

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim (0, \sigma^2 \Omega), \tag{1.2}$$

where the notation $\varepsilon \sim (0, \sigma^2 \Omega)$ means that the disturbance vector ε comes from a distribution having zero mean and variance-covariance matrix $\sigma^2 \Omega$.

The notation used in this treatment is consistent with that used in [28, 46] and is here briefly resumed. An $m \times n$ matrix with elements $a_{i,j}$ ($i = 1, \ldots, m$ and $j = 1, \ldots, n$) will be denoted by $A \in \mathbb{R}^{m \times n}$. Similarly, $v \in \mathbb{R}^m$ will denote a vector with elements v_i, ($i = 1, \ldots, m$). Standard *colon-notation* is used to denotes submatrices and subvectors [28]. The kth column and row of A are denoted by $A_{:,k}$ and $A_{k,:}$, respectively. The submatrix $A_{i:k,j:l}$ has dimension $(k-i+1)\times(l-j+1)$ and its first elements is given by $a_{i,j}$. Similarly, $v_{i:k}$ is the subvector of v having $(k-i+1)$ elements and starting with v_i. When the lower or upper index in the subscript is omitted, then the default values will be one or an the upper bound of this subscript, respectively. A zero dimension denotes a null matrix or vector. For example, $A_{i:,:l}$ is equivalent to $A_{i:m,1:l}$. The transpose of A will be denoted by A^T and if $A^{m\times m}$ is non singular then its transpose inverse will be written as A^{-T}. The $m\times m$ identity matrix and its ith column will be denoted by I_m and e_i, respectively. Thus, $I_m = \begin{pmatrix} e_1 & e_2 & \cdots & e_m \end{pmatrix}$. Furthermore, $\|\cdot\|$, $\|\cdot\|_\Omega$ and $\|\cdot\|_F$ will denote the Euclidean, energy and Frobenious norms, respectively. That is, $\|v\|^2 = v^Tv$, $\|v\|^2_\Omega = v^T\Omega v$ and $\|A\|^2_F = \sum_{i=1,j=1}^{m,n} a^2_{i,j}$, where Ω is positive definite.

1.2 QR Decomposition and the Ordinary Linear Model

The QR Decomposition (QRD) is one of the main computational tools in regression [8, 9, 11, 19, 28, 31, 57, 75]. It is mainly used in the solution of linear systems. It provides more accurate solutions than the LU and other similar decompositions, and involves fewer computations than the Singular Value Decomposition. The QRD is often associated with the solution of Least-Squares (LS) problems which arise in various applications, such as statistics, econometrics, optimization and signal processing to name but a few. The development of numerically stable and computationally efficient methods for solving LS problems has been an active research area for more than fifty years [29, 41, 57]. The QRD can be used efficiently to compute various diagnostic measures in regression [7].

The matrices might have special structures which need to be exploited. Often in econometrics and signal processing, the matrices have Toeplitz, Kronecker products and block-diagonal structures [18, 58, 67]. Computationally efficient methods to solve the LS problems should exploit the non-dense structure of the information matrices and enacts non-literal computation on the Kronecker products.

Consider the Ordinary Linear Model (OLM)

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim (0, \sigma^2 I_m), \tag{1.3}$$

where $y, \varepsilon \in \mathbb{R}^m$ is the response vector, $X \in \mathbb{R}^{m\times n}$ is the full rank exogenous matrix, $\beta \in \mathbb{R}^n$ are the coefficients to be estimated and $\varepsilon \in \mathbb{R}^m$ is the vector of disturbances. The Ordinary Least

Squares (OLS) estimator of (1.3) is given by $\hat{\beta} = (X^T X)^{-1} X^T y$, that is, it is the solution of the system of normal equations $X^T X\beta = X^T y$. Furthermore, if $\rho = X\hat{\beta} - y$ is the residual vector of ρ, then the scalar σ is estimated by $\hat{\sigma}^2 = \rho^T\rho/(m-n)$. The OLS estimator is linear and provides an unbiased estimator, that is $\mathrm{E}(\hat{\beta}) = \beta$. Furthermore, if $\tilde{\beta}$ is another linear unbiased estimator for β, then $\mathrm{E}((\tilde{\beta}-\beta)(\tilde{\beta}-\beta)^T) - \mathrm{E}((\hat{\beta}-\beta)(\hat{\beta}-\beta)^T)$ is non negative definite. Thus $\hat{\beta}$ is a Best Linear Unbiased Estimator (BLUE) for the linear model (1.3) [68].

Alternatively, let the QR Decomposition (QRD) of X be given by

$$Q^T X = \begin{pmatrix} R \\ 0 \end{pmatrix} \begin{matrix} {\scriptstyle n} \\ {\scriptstyle m-n} \end{matrix} \qquad \text{and} \qquad Q^T y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \begin{matrix} {\scriptstyle n} \\ {\scriptstyle m-n} \end{matrix}, \tag{1.4}$$

where $Q \in \mathbb{R}^{m\times m}$ is orthogonal, that is $Q^T Q = I_m$, and $R \in \mathbb{R}^{n\times n}$ is upper triangular. The OLS estimator of β comes from the solution of the triangular system $R\beta = y_1$ and σ is estimated by $\hat{\sigma}^2 = y_2^T y_2/(m-n)$.

1.2.1 Forming the QR decomposition

The $m \times m$ Householder matrix (or Householder reflector or Householder transformation) is of the form

$$H = I_m - 2\frac{hh^T}{\|h\|^2},$$

where $h \in \mathbb{R}^m$ is a non-zero vector. Householder matrices are orthogonal and symmetric. They can be used to annihilate specified elements of a vector or a matrix [9, 28]. Let $x \in \mathbb{R}^m$ be non zero, a Householder matrix H can be chosen such that $y = Hx$ has zero elements in positions 2 to m by setting $h = x \pm \alpha e_1$, where $\alpha = x^T x$ and e_1 denotes the first column of the $m \times m$ identity matrix.

Consider now the computation of the QRD (1.4) using Householder transformations. The orthogonal matrix Q is defined as the product of the n Householder transformations

$$Q = H_1 H_2 \cdots H_n,$$

where $H_i = I_m - 2h_i h_i^T/(h_i^T h_i)$ and

$$h_i = \begin{pmatrix} 0 \\ \tilde{h}_i \end{pmatrix} \begin{matrix} {\scriptstyle i-1} \\ {\scriptstyle m-i+1} \end{matrix}.$$

Let $A^{(i)} \equiv X$ and

$$A^{(i)} = H_i A^{(i-1)} \equiv \begin{array}{cc} \begin{array}{cc} i & n-i \end{array} & \\ \begin{pmatrix} R_{11}^{(i)} & R_{12}^{(i)} \\ 0 & \widetilde{A}^{(i)} \end{pmatrix} & \begin{array}{l} i \\ m-i \end{array} \end{array}, \qquad i = 1, \ldots, n-1, \tag{1.5}$$

where $R_{11}^{(i)}$ is upper triangular. The application of H_{i+1} from the left of $A^{(i)}$ annihilates the last $m-i$ elements of the first column of $\widetilde{A}^{(i)}$. The transformation $H_{i+1}A^{(i)}$ affects only $\widetilde{A}^{(i)}$ and it follows that

$$A^{(n)} \equiv \begin{pmatrix} R \\ 0 \end{pmatrix}.$$

An $m \times m$ Givens rotation is a rank-two correction of the identity matrix and has the form

$$G_{i,j} = \begin{array}{cc} \begin{array}{ccccc} & \overset{i}{\downarrow} & & \overset{j}{\downarrow} & \end{array} & \\ \begin{pmatrix} I_{i-1} & 0 & 0 & 0 & 0 \\ 0 & c & 0 & s & 0 \\ 0 & 0 & I_{j-i-1} & 0 & 0 \\ 0 & -s & 0 & c & 0 \\ 0 & 0 & 0 & 0 & I_{m-j-1} \end{pmatrix} & \begin{array}{l} \\ \leftarrow i \\ \\ \leftarrow j \\ \\ \end{array} \end{array},$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$ for some θ, that is $c^2 + s^2 = 1$ [28]. It follows that $G_{i,j}$ is orthogonal. The transformation $G_{i,j}$ when applied to the left of a matrix can annihilate a specific element in the jth row of the matrix. While the Householder reflections are useful for introducing zero elements on the grand scale, Givens rotations are important because they can annihilate the elements of a matrix more selectively.

If $A \in \mathbb{R}^{m \times n}$ and $\widetilde{A} = G_{i,j}A$, then the pth row of $\widetilde{A}$ is given by

$$\widetilde{A}_{p,:} = \begin{cases} cA_{i,:} + sA_{j,:}, & \text{if } p = i, \\ -sA_{i,:} + cA_{j,:}, & \text{if } p = j, \\ A_{p,:}, & \text{otherwise.} \end{cases}$$

Thus, if the (j,k)th element of A, i.e. $a_{j,k}$, is non zero, it can be annihilated using the Givens transformation $G_{i,j}$ by setting $c = a_{i,k}/t$, $s = a_{j,k}/t$ and $t^2 = a_{i,k}^2 + a_{j,k}^2$.

A sequence of Givens rotations can be applied to compute the QRD (1.4). One of such sequences, called *column-based*, is shown in Figure 1.1 where $m = 4$ and $n = 3$. The elements of A are annihilated column by column and bottom to the top, starting from the first column. Furthermore, the Givens rotations are applied between adjacent planes, that is, $j = i + 1$.

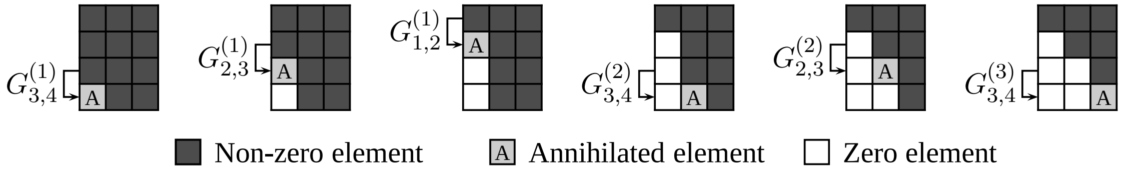


Figure 1.1: Illustration of the *column-based* Givens sequence to compute the QRD of A, where $m = 4$ and $n = 3$.

A block generalization of the Givens matrix can be used to annihilate a submatrix instead of a single element. Let $A \in \mathbb{R}^{m\times n}$ be partitioned as

$$A = \begin{array}{c@{}l} \begin{array}{cccc} n_1 & n_2 & \cdots & n_\eta \end{array} & \\ \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1\eta} \\ A_{21} & A_{22} & \cdots & A_{2\eta} \\ \vdots & \vdots & & \vdots \\ A_{\mu 1} & A_{\mu 2} & \cdots & A_{\mu\eta} \end{pmatrix} & \begin{array}{l} m_1 \\ m_2 \\ \vdots \\ m_\mu \end{array} \end{array} ,$$

where $n = \sum_{i=1}^{\eta} n_i$ and $m = \sum_{i=1}^{\mu} m_i$. Consider the QRD

$$Q^T \begin{pmatrix} A_{ij} \\ A_{kj} \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} \begin{array}{l} n_j \\ m_i+m_k-n_j \end{array} \qquad \text{with} \qquad Q \equiv \begin{array}{c@{}l} \begin{array}{cc} m_i & m_k \end{array} & \\ \begin{pmatrix} Q_{1,1} & Q_{1,2} \\ Q_{2,1} & Q_{2,2} \end{pmatrix} & \begin{array}{l} m_i \\ m_k \end{array} \end{array} ,$$

where it is assumed that $m_i + m_k > n_j$. Thus, if $m_i > n_i$, then the orthogonal matrix which annihilates the submatrix A_{kj} when applied from the left of A is given by

$$\overline{Q}^T = \begin{pmatrix} I_{m^{(1)}} & 0 & 0 & 0 & 0 \\ 0 & Q^T_{1,1} & 0 & Q^T_{2,1} & 0 \\ 0 & 0 & I_{m^{(2)}} & 0 & 0 \\ 0 & Q^T_{1,2} & 0 & Q^T_{2,2} & 0 \\ 0 & 0 & 0 & 0 & I_{m^{(3)}} \end{pmatrix} , \tag{1.6}$$

where $m^{(1)} = m_1 + m_2 + \cdots + m_{i-1}$, $m^{(2)} = m_{i+1} + m_{i+2} + \cdots + m_{k-1}$ and $m^{(3)} = m_{k+1} + m_{k+2} + \cdots + m_\eta$. Notice that in general $\overline{Q}$ is not a rotation matrix. The orthogonal matrices having the form of $\overline{Q}$ in (1.6) are extensively used in this treatment to develop strategies which exploit the sparse block-structure of the various matrices which arise in the estimation of econometric linear models [20, 24, 25, 22, 46].

1.3 Generalized QR Decomposition and the General Linear Model

The variance-covariance matrix of the disturbances Ω of the GLM (1.2) is often assumed to be positive definite. In such cases the BLUE of β in (1.2) comes from solving the Generalized Least Squares (GLS) problem

$$\operatorname*{argmin}_{\beta} \left\| y - X\beta \right\|_{\Omega^{-1}}$$

which is equivalent to the normal equations

$$X^T \Omega^{-1} X \beta = X^T \Omega^{-1} y. \tag{1.7}$$

This solution, however, can be unstable when the matrices are ill-conditioned and explicit matrix inversion are used [9, 57]. Furthermore, if Ω is singular, then the GLS estimator cannot be computed using (1.7) and the replacement of Ω^{-1} by the Moore-Penrose generalized inverse would not always give the BLUE of β [55].

To avoid problems associated with the singularity or ill-conditioning of Ω, the GLM (1.2) can be formulated as the Generalized Linear Least Squares Problem (GLLSP)

$$\operatorname*{argmin}_{\upsilon,\beta} \upsilon^T \upsilon \qquad \text{subject to} \qquad y = X\beta + C\upsilon, \tag{1.8}$$

where $\Omega \in \mathbb{R}^{m \times m}$ is non-negative definite with rank g, $\Omega = CC^T$, $C \in \mathbb{R}^{m \times g}$ has full column rank, the random g-element vector υ is defined as $C\upsilon = \varepsilon$. That is, $\upsilon \sim (0, \sigma^2 I_g)$ [55]. The Generalized QRD (GQRD) can be employed to solve GLLSP [2, 64]. Although the above formulation allows for singular Ω, without loss of generality consider the case where Ω is non-singular. The GQRD of X and C is given by the QRD (1.4) and the RQD of $Q^T C$ which can be written as

$$(Q^T C)P = U \equiv \begin{array}{c} \\ \\ \end{array}\!\!\begin{array}{cc} \scriptstyle n & \scriptstyle m-n \\ \end{array} \qquad \left(\begin{array}{cc} U_{11} & U_{12} \\ 0 & U_{22} \end{array}\right) \begin{array}{l} \scriptstyle n \\ \scriptstyle m-n \end{array}, \tag{1.9}$$

where $P \in \mathbb{R}^{m\times m}$ is orthogonal and $U \in \mathbb{R}^{m\times m}$ is upper triangular and non-singular. The GLLSP (1.8) can be equivalently written as

$$\underset{\upsilon,\beta}{\operatorname{argmin}} \left\|P^T\upsilon\right\|^2 \qquad \text{subject to} \qquad Q^Ty = Q^TX\beta + Q^TCPP^T\upsilon$$

or

$$\underset{\upsilon_1,\upsilon_2,\beta}{\operatorname{argmin}}\left(\left\|\upsilon_1\right\|^2 + \left\|\upsilon_2\right\|^2\right) \quad \text{subject to} \quad \begin{cases} y_1 = R\beta + U_{11}\upsilon_1 + U_{12}\upsilon_2, \\ y_2 = U_{22}\upsilon_2, \end{cases} \tag{1.10}$$

where υ^TP is conformably partitioned as $(\upsilon_1^T \quad \upsilon_2^T)$. In the second constraint of (1.10) υ_2 comes from solving the upper triangular system $U_{22}\upsilon_2 = y_2$, and in the first constraint the arbitrary subvector υ_1 is set to zero in order to minimize the objective function. Thus, the estimator of β derives from the solution of the upper triangular system $R\beta = y_1 - U_{12}\upsilon_2$. The variance-covariance of the coefficients estimator is given by $\hat{\sigma}^2R^{-1}U_{11}U_{11}^TR^{-T}$, where $\hat{\sigma}^2 = \upsilon_2^T\upsilon_2/(m-n)$ is an estimator of σ^2.

1.4 The SUR model

The SUR model is a special case of the GLM. It is defined by the set of regressions

$$y_i = X_i\beta_i + u_i, \qquad\qquad i = 1,\dots,G,$$

where $X_i \in \mathbb{R}^{T\times k_i}$ has full column rank, $y_i \in \mathbb{R}^T$ and the T-element disturbance vector u_i has zero mean, variance-covariance matrix $\sigma_{i,i}I_T$ and is contemporaneously correlated across the equations, so $\mathrm{E}(u_iu_j^T) = \sigma_{i,j}I_T$. In compact form the SUR model is written

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_G \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_G \end{pmatrix}$$

or

$$\mathrm{Vec}(Y) = \left(\oplus_{i=1}^G X_i\right)\mathrm{Vec}(\{\beta_i\}_G) + \mathrm{Vec}(U), \tag{1.11}$$

where $Y = \begin{pmatrix} y_1 \dots y_G \end{pmatrix} \in \mathbb{R}^{T\times G}$, $\oplus_{i=1}^G X_i = \mathrm{diag}(X_1,\dots,X_G) \in \mathbb{R}^{T\times K}$ denotes the direct sum of matrices, $\{\beta_i\}_G$ denotes the set of vectors $\beta_1 \dots \beta_G$, and $\mathrm{Vec}(\cdot)$ the column stack operator.

The disturbance term $\mathrm{Vec}(U)$ has zero mean and variance-covariance matrix $\Sigma \otimes I_T$, where $\Sigma = [\sigma_{i,j}] \in \mathbb{R}^{G\times G}$ is symmetric and non-negative definite and $\otimes$ denotes the Kronecker product [42, 50, 51, 70, 73, 78, 79, 86, 87]. That is, $\mathrm{Vec}(U) \sim (0, \Sigma \otimes I_T)$ and

$$\Sigma \otimes I_T = \begin{pmatrix} \sigma_{1,1}I_T & \sigma_{1,2}I_T & \cdots & \sigma_{1,G}I_T \\ \sigma_{2,1}I_T & \sigma_{2,2}I_T & \cdots & \sigma_{2,G}I_T \\ \vdots & \vdots & & \vdots \\ \sigma_{G,1}I_T & \sigma_{G,2}I_T & \cdots & \sigma_{G,G}I_T \end{pmatrix}$$

Notice that $(A \otimes B)(C \otimes D) = AC \otimes BD$ and $\mathrm{Vec}(ABC) = (C^T \otimes A)\,\mathrm{Vec}(B)$.

The solution of the GLLSP (1.8) has also been discussed within the context of estimating SUR models and their variants [6, 14, 34, 52, 56, 71, 89]. This formulation and the use of the GQRD allows to desing computationally efficient methods by exploiting the special structure of the regressor and covariance matrices.

Often in econometric special cases or extensions of the SUR model are considered. The most common cases are here briefly reviewed:

- **Heteroschedastic disturbances.** In this extension the assumption of a spherical distributed u_i is relaxed. The covariance matrix of u_i and u_j is assumed to be $\mathrm{E}(u_i u_j^T) = \sigma_{i,j}D$, where D is a diagonal matrix of positive elements. Thus, the variance-covariance matrix of $\mathrm{Vec}(U)$ is given by $\Sigma \otimes D$.

- **Correlation Constraints.** SUR models where the disturbances covariances are constrained includes the SUR model with Correlation Constrains (SUR-CC) [48, 79, 88]. In this models the elements of the covariance matrix is constrained so that the variance of one disturbance is smaller than that of the disturbance in the successive equation. Furthermore, the correlation between the disturbances are between zero and one. That is,

  $$\sigma_{1,1} \leq \sigma_{2,2} \leq \cdots \leq \sigma_{G,G}$$

  and

  $$0 < \rho_{i,j} < 1, \qquad i, j = 1, \ldots, G \text{ and } i \neq j,$$

  where $\rho_{i,j} = \sigma_{i,j}/\sqrt{\sigma_{i,i}\sigma_{j,j}}$ is the correlation between $u_{i,t}$ and $u_{j,t}$. This specification has applications in the context of *error components* models, where the disturbance term is defined as $u_i = \sum_{j=1}^{i} \varepsilon_j$, $\varepsilon_i \sim (0, \upsilon_i I_T)$ and $\mathrm{E}(\varepsilon_i \varepsilon_j^T) = 0$ for $i \neq j$.

- **Autoregressive disturbances.** In the SUR model (1.11) the disturbances are assumed to be uncorrelated across time. However, in several applications this is to much restrictive and correlation on time should be considered. In the SUR model with autoregressive (AR) disturbances, hereafter SUR-AR model, the errors are generated by the AR process

  $$u_{i,t} = \alpha_i u_{i,t-1} + \varepsilon_{i,t}, \qquad i = 1, \ldots, G, t = 2, \ldots, T,$$

  and where α_i are the AR parameters, $\varepsilon_{i,t} \sim (0, \sigma_{i,i})$, $\mathrm{E}(\varepsilon_{i,t}\varepsilon_{j,t}) = \sigma_{i,j}$ and $\mathrm{E}(\varepsilon_{i,s}\varepsilon_{j,t}) = 0$ for $s \neq t$ and $i, j = 1, \ldots, G$ [26, 43, 65, 85]. This can be written in compact form as $u_i = \alpha_i Z u_i + \varepsilon_i$ ($i = 1, \ldots, G$), or

  $$\mathrm{Vec}(U) = (\oplus_i \alpha_i Z)\,\mathrm{Vec}(U) + \mathrm{Vec}(E), \qquad \mathrm{Vec}(E) \sim (0, \Sigma \otimes I_T),$$

  where $\varepsilon_i = \begin{pmatrix} \varepsilon_{i,1} & \cdots & \varepsilon_{i,T} \end{pmatrix}^T$, $E = \begin{pmatrix} \varepsilon_1 & \cdots & \varepsilon_G \end{pmatrix}$ and Z denotes the $T \times T$ shift matrix

  $$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

  In this model the covariance matrix of $\mathrm{Vec}(U)$ is given by

  $$\big(\oplus_{i=1}^{G} (I_T - \alpha_i Z)\big)^{-1} \big(\Sigma \otimes I_T\big) \big(\oplus_{i=1}^{G} (I_T - \alpha_i Z)\big)^{-T}$$

- **Vector Autoregressive disturbances.** A more general form of autocorrelation is given by the Vector Autoregressive (VAR) process:

  $$U = ZUA^T + E, \qquad \mathrm{Vec}(E) \sim (0, \Sigma \otimes I_T),$$

  where $A \in \mathbb{R}^{G \times G}$ is the matrix of the AR parameters. In this case the covariance matrix of the SUR disturbances is given by

  $$\big(I_{GT} - A \otimes Z\big)^{-1} \big(\Sigma \otimes I_T\big) \big(I_{GT} - A \otimes Z\big)^{-T}.$$

  Additional assumption regarding the disturbances of the first observation, that is $u_{i,1}$ ($i = 1, \ldots, G$) should or may be specified. Furthermore, in some contexts it could be necessary to include more that one lags in the autoregressive specification, that is the disturbances are follow the VAR(p) process

$$U = ZUA_1^T + Z^2UA_2^T + \cdots + Z^pUA_p^T + E, \qquad \mathrm{Vec}(E) \sim (0, \Sigma \otimes I_T),$$

where $A_1, \ldots, A_p \in \mathbb{R}^{G \times G}$ are the matrices of the AR parameters.

- **Unequal Size Observations.** The formulation of the SUR model, as given in (1.11), assumes that each regression has the same number of observations, however this might not always be the case [72, 74, 77, 79]. The SUR model with Unequal Size Observation (SUR-USO) assumes that the ith equation has t_i observations and that the first t_i observations for the $(i+1)$th regression match in time with those for the ith regression. In this case the covariance matrices of the disturbances u_i and u_j, for $j > i$ are given by

$$\mathrm{E}(u_i u_j^T) = \sigma_{i,j} \begin{pmatrix} I_{t_i} & 0_{t_i \times (t_j - t_i)} \end{pmatrix}.$$

- **Missing Observations.** A generalization of the SUR-USO model, is given by the SUR model with Missing Observations. There, the pattern of the observations which are missing is not fixed and the t_i observations of the ith equation does not necessarily match in time with the first of $(i-1)$th regression. The covariances in this case are given by

$$\mathrm{E}(u_i u_j^T) = \sigma_{i,j} S_{i,j},$$

where $S_{i,j}$ is $t_i \times t_j$ zero-one matrix with its (k,l) element being non zero if the kth and lth observations of the ith and jth equations, respectively, match in time.

- **Common Regressors.** In the SUR model with Common Regressors, the exogenous matrix $X_i = X^d S_i$ ($i = 1, \ldots, G$), where $X^d \in \mathbb{R}^{T \times k^d}$ denotes the matrix consisting of the K^d distint regressors and $S_i \in \mathbb{R}^{K^d \times k_i}$ is a selection matrix that comprises the relevant columns of the $K^d \times K^d$ identity matrix. This is often the case because an exogenous factor can appear in more than one regressor matrix. For example such occurs when estimating the multivariate linear regression model

$$Y = X^d B + U, \qquad \mathrm{Vec}(U) \sim (0, \Sigma \otimes I_T),$$

with the constraints

$$b_{i_s, j_s} = 0, \qquad s = 1, \ldots, n.$$

That is, n elements of the parameter matrix $B \in \mathbb{R}^{K^d \times G}$ are restricted to zero, or, equivalently, the i_sth column of X^d is not included in the model to predict the j_sth column of Y.

- **Triangular SUR models.** Triangular SUR models are special cases of SUR models with common regressor. In these models, the equations and the columns of each exogenous matrix can be reordered such that the ith regressor consists of a subset of the columns of the following one, that is, $X_i = \begin{pmatrix} X_{i-1} & \widetilde{X}_i \end{pmatrix}$, where $\widetilde{X}_i \in \mathbb{R}^{T \times k_i - k_{i-1}}$.

1.5 Overview of the thesis

Each chapter of the thesis is self contained[1]. Chapter 2 considers the estimation of Vector Autoregressive (VAR) processes with zero coefficient constraints. This estimation problem reduces to that of estimating a SUR model with common regressors, where the matrix of distinct regressors X^d has block Toeplitz structure. The analysis presented is extended to the VARX model, where exogenous factors, such as linear or polynomial trends in the data generation process, are included into the model. A procedure is detailed for reducing the size of model by efficiently computing the triangular R-factor in the QRD of the exogenous matrix X^d. Then the estimation of SUR models with common regressors and that of the triangular SUR models are considered. These model derive from the estimation of VAR models with zero coefficient constraints or VAR model with Granger (non-causality) restrictions, respectively. This analysis extends the one that have been provided in [51] where a specific ordering on the equations was imposed and applies to situations where different Granger causality restrictions need to be imposed and tested.

In Chapter 3 methods for estimating the SUR model with AR and VAR disturbances (SUR-VAR) are presented. In that model the covariance matrix of the disturbances is dense and structured. When the number of observations is large the SUR-VAR model can be reduced to a GLM of smaller dimensions. Efficient strategies to solve the resulting GLLSP which exploit the structure of the Cholesky factor of the covariance matrix are derived.

The estimation of the SUR model Unequal Size Observations (SUR-USO) is considered in Chapter 4. Two algorithms which solve the GLLSP derived from the SUR-USO model are proposed. The first computes the GQRD of the regressor and Cholesky factor of the dispersion matrix. While, the second use a recursive approach to solve the GLLSP, where at each step a set of obser-

---

[1] Each chapter has been published, or is in press in a referred interantional journal.

vations are added to the model. Furthermore, an estimator for the covariance matrix is proposed for the case of normally distributed disturbances. With respect to most of the existing estimators, this has the advantage of being always non-negative definite.

Chapter 5 presents existing direct methods for estimating the basic SUR model and proposes two new methods. The first is based on a recursive approach, while the second on the transformation of the SUR model to a smaller SUR-USO model. The derivation of the algorthms is presented and their comparison based on their theoretical complexity and on computational experiments is given. Finally, the last chapter concludes and provides directions for future research.

Chapter 2

Estimation of VAR models

Abstract:

The Vector Autoregressive (VAR) model with zero coefficient restrictions can be formulated as a Seemingly Unrelated Regression Equations (SURE) model. Both the response vectors and the coefficient matrix of the regression equations comprise columns from a Toeplitz matrix. Efficient numerical and computational methods which exploit the Toeplitz and Kronecker product structure of the matrices are proposed. The methods are also adapted to provide numerically stable algorithms for the estimation of VAR(p) models with Granger–caused variables.

2.1 Introduction

The vector time series $z_t \in \mathbb{R}^n$ is a Vector Autoregressive (VAR) process of order p when its data generation process has the form

$$z_t = \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \cdots + \Phi_p z_{t-p} + u_t, \tag{2.1}$$

where $\Phi_i \in \mathbb{R}^{n\times n}$ are the coefficient matrices and the vectors $u_t \in \mathbb{R}^n$ are serially uncorrelated and identically distributed with zero mean and variance–covariance matrix Σ. That is, $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t u_t^T) = \Sigma$ and $\mathrm{E}(u_t u_\tau^T) = 0$, for $t \neq \tau$.

---

[1] This chapter is a reprint of the paper: P. Foschi, E.J. Kontoghiorghes. Estimation of VAR models: computational aspects. *Computational Economics*, 21(1):3-22, 2003.

Given a sample $z_1, \ldots, z_M$ and a presample $z_{1-p}, \ldots, z_0$ the VAR model (2.1) is efficiently estimated by Ordinary Least Squares (OLS) estimation of the model

$$\begin{pmatrix} z_M^T \\ z_{M-1}^T \\ \vdots \\ z_1^T \end{pmatrix} = \begin{pmatrix} z_{M-p}^T & z_{M+1-p}^T & \cdots & z_{M-1}^T \\ z_{M-1-p}^T & z_{M-p}^T & \cdots & z_{M-2}^T \\ \vdots & \vdots & & \vdots \\ z_{1-p}^T & z_{2-p}^T & \cdots & z_0^T \end{pmatrix} \begin{pmatrix} \Phi_p^T \\ \Phi_{p-1}^T \\ \vdots \\ \Phi_1^T \end{pmatrix} + \begin{pmatrix} u_M^T \\ u_{M-1}^T \\ \vdots \\ u_1^T \end{pmatrix},$$

which in compact form it can be written as

$$Y = XB + U, \tag{2.2}$$

where Y, X, B and U are defined by the context. The variance–covariance matrix of $\mathrm{Vec}(U)$ is $\Sigma \otimes I_M$, where $\mathrm{Vec}(\cdot)$ denotes the column stacking operator and $\otimes$ is the Kronecker product operator. The OLS and Generalized Least Squares (GLS) estimators of (2.2) are the same [58, 86].

Let $T = \begin{pmatrix} X & Y \end{pmatrix}$ and its QR decomposition (QRD) be given by

$$T = Q\begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{array}{ccc} \scriptstyle np & \scriptstyle n & \scriptstyle M-(p+1)n \\ \left(Q_T \right. & Q_Y & \left. Q_N \right) \end{array} \begin{array}{ccl} \scriptstyle np & \scriptstyle p & \\ \left(R_T \right. & \left. R_{TY} \right) & \scriptstyle np \\ \left(0 \right. & \left. R_Y \right) & \scriptstyle p \\ \left(0 \right. & \left. 0 \right) & \scriptstyle M-(p+1)n \end{array}, \tag{2.3}$$

where $Q \in \mathbb{R}^{M \times M}$ is orthogonal and $R \in \mathbb{R}^{(n+1)p \times (n+1)p}$ is upper triangular. The OLS estimator of B in (2.2) is computed by

$$\hat{B} = R_T^{-1} R_{TY}.$$

The residuals are given by

$$\hat{U} = Q_Y R_Y,$$

and the covariance matrix Σ is estimated by

$$\hat{\Sigma} = \alpha \hat{U}^T \hat{U} = \alpha R_Y^T R_Y,$$

where $\alpha = 1/M$ or $\alpha = 1/(M - np)$.

Alternatively R in (2.3) may be computed using the Cholesky factorization of $T^T T$, but this is neither computationally efficient nor numerically stable due to the poor numerical properties of the matrix $T^T T$. Efficient methods avoid this problem by computing the Cholesky factorization without forming that matrix explicitly.

The matrix T is Block-Toeplitz with blocks of size $1 \times n$ that are constant along the diagonals. Exploiting the structure of T, a fast algorithm is possible for computing the upper triangular matrices in (2.3) and, thus, a fast estimation algorithm can be designed.

When other, possibly endogenous, factors such as deterministic trends are added, (2.1) becomes

$$z_t = \Theta w_t + \Phi_1 z_{t-1} + \Phi_2 z_{t-2} + \cdots + \Phi_p z_{t-p} + u_t, \tag{2.4}$$

where $w_t \in \mathbb{R}^\eta$ and $\Theta \in \mathbb{R}^{n \times \eta}$. The OLS estimates of (2.4) can be computed the same way as those of (2.1). However, care must be taken in arranging the matrix of regressors (endogenous and exogenous) in (2.3) to minimize the loss of structure derived from the introduction of the endogenous variables w_t. A good choice is $M = \big(W \ X \ Y\big)$, where $W^T = \big(w_M \ w_{M-1} \ \cdots \ w_1\big)$.

The complexity of the algorithms will be given in flops (floating point operations per second), where flop denotes a single scalar multiplication or addition. Throughout the paper, the following notation is used: the vector e_i denotes the ith column of the $n \times n$ identity matrix I_n and the $n \times n$ shift matrix is denoted by $Z = (e_2 \ e_3 \ \cdots \ e_n \ 0)$, that is

$$Z = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

A set of vectors $v_1, v_2, \ldots, v_n$ is denoted by $\{v_i\}_n$ and the direct sums of two or more matrices $A \oplus B$ and $\oplus_{i=1}^n A_i$ are equivalent to the block diagonal matrices

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{pmatrix},$$

respectively [3, 30, 69]. For notational convenience the subscript n in the set operator $\{\cdot\}_n$ is dropped and $\oplus_{i=1}^n$ will be abbreviated to $\oplus_i$.

The purpose of this work is twofold. First, to exploit the structure of the Toeplitz matrix in (2.2) and provide a fast algorithm to compute the upper triangular matrix R and, consequently, an efficient estimation procedure for the VAR(p) models. Second, to design computationally efficient methods to estimate VAR models with coefficient constraints by exploiting the Kronecker product structure of the Seemingly Unrelated Regression Equations (SURE) models.

In section 2.2 the Generalized Schur Algorithm (GSA) and its block version are presented. In section 2.3 the adaption of the algorithm to estimate the VAR models (2.1) and (2.4) is considered. In section 2.4 the estimation of the VAR model with Zero Coefficient Constraints (VAR-ZCC) and the resulting SURE model are investigated. Finally in section 2.5 the estimation of the VAR model with Granger–causality restrictions is presented. This model is considered as a SURE model with Proper Subset Regressors (SURE-PSS).

2.2 Structured matrices and the Generalized Schur Algorithm

Let $A \in \mathbb{R}^{n \times n}$ be a positive definite matrix and $F \in \mathbb{R}^{n \times n}$ be strictly lower triangular; that is, $F = [f_{ij}]$ with $f_{ij} = 0$ for $i \leq j$. The displacement of A with respect to (w.r.t.) F is

$$\nabla_F A = A - FAF^T \tag{2.5}$$

and its rank $d = \mathrm{rank}(\nabla_F A)$ is called the displacement rank of A. The matrix A is said to have a displacement structure or, more simply, be structured w.r.t. F if it has a small displacement rank. In this case

$$\nabla_F A = G^T J G, \tag{2.6}$$

where $G \in \mathbb{R}^{d \times n}$, $J = I_k \oplus (-I_l)$ and $k + l = d$. The rows of G are called the generators of A [39, 40].

Given F, the matrix A is uniquely defined by k, l and G. In fact, since F is strictly lower triangular, $F^n = 0$, and from (2.5) and (2.6) it follows that

$$A = \sum_{i=0}^{n-1} F^i (A - FAF^T)(F^i)^T = \sum_{i=0}^{n-1} F^i G^T J G (F^i)^T, \tag{2.7}$$

where it has been assumed $F^0 = I_n$. Consider for example the symmetric Toeplitz matrix

$$T = \begin{pmatrix} t_1 & t_2 & t_3 & \cdots & t_n \\ t_2 & t_1 & t_2 & \cdots & t_{n-1} \\ t_3 & t_2 & t_1 & \cdots & t_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_n & t_{n-1} & t_{n-2} & \cdots & t_1 \end{pmatrix}.$$

This is a structured matrix and its displacement rank w.r.t. the shift matrix $F = \begin{pmatrix} e_2 \; e_3 \; \cdots \; e_n \; 0 \end{pmatrix}$ – the matrix with ones on the first sub-diagonal and zero elsewhere – is 2, $k = l = 1$, and the generators are

$$G = \frac{1}{\sqrt{t_1}} \begin{pmatrix} t_1 & t_2 & t_3 & \cdots & t_n \\ 0 & t_2 & t_3 & \cdots & t_n \end{pmatrix} .$$

A Generalized Schur Algorithm (GSA) can be used to compute the Cholesky factorization $A = R^T R$ when A has displacement structure (2.6). At each iteration a row of R is computed. Since the first column of $F^i G$ is zero for $i \geq 1$, it follows from (2.7) that the first row of A is given by

$$\begin{aligned} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix} &= r_{11} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \end{pmatrix} \\ &= \begin{pmatrix} g_{11} & g_{21} & \cdots & g_{d1} \end{pmatrix} JG. \end{aligned} \tag{2.8}$$

If Q is a J-invariant matrix (a hyperbolic transformation), that is $QJQ^T = J$, then the rows of $\tilde{G} = Q^T G$ are again generators for A. If Q is chosen to annihilate the first column of G except for the $(1, 1)$–element, then (2.8) is given by

$$\tilde{g}_{11} \begin{pmatrix} \tilde{g}_{11} & \tilde{g}_{12} & \cdots & \tilde{g}_{1n} \end{pmatrix} ;$$

that is, $\tilde{g}_1^T = \begin{pmatrix} \tilde{g}_{11} \; \tilde{g}_{1n} \; \cdots \; \tilde{g}_{1n} \end{pmatrix}$ is the first row of R.

Consider now the partitioning

$$R = \begin{array}{c} \begin{array}{cc} \scriptstyle i-1 & \scriptstyle n-i+1 \end{array} \\ \begin{pmatrix} R_1 & R_{12} \\ 0 & R_2 \end{pmatrix} \end{array} \begin{array}{l} \scriptstyle i-1 \\ \scriptstyle n-i+1 \end{array} , \qquad A = \begin{array}{c} \begin{array}{cc} \scriptstyle i-1 & \scriptstyle n-i+1 \end{array} \\ \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \end{array} \begin{array}{l} \scriptstyle i-1 \\ \scriptstyle n-i+1 \end{array}$$

and define

$$A^{(i)} = A - \begin{pmatrix} R_1^T \\ R_{12}^T \end{pmatrix} \begin{pmatrix} R_1 & R_{12} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix} ,$$

where $S = R_2^T R_2$ is the Schur complement of A_{11}. If $A^{(i)}$ has displacement structure

$$\nabla_F A^{(i)} = G_i^T J G_i \ , \tag{2.9}$$

with

$$G_i = \begin{array}{c} \begin{array}{ccc} \scriptstyle i-1 & \scriptstyle 1 & \scriptstyle n-i \end{array} \\ \begin{pmatrix} 0 & u_1 & u_2^T \\ 0 & v_1 & V_2 \end{pmatrix} \end{array} \begin{array}{l} \scriptstyle 1 \\ \scriptstyle d-1 \end{array} , \tag{2.10}$$

and Q_i is a J-invariant matrix that annihilates the elements of v_1, that is,

$$\tilde{G}_i = Q_i^T G_i = \begin{array}{c} \begin{array}{ccc} i-1 & 1 & n-i \end{array} \\ \left(\begin{array}{ccc} 0 & \rho & \tilde{u}^T \\ 0 & 0 & \tilde{V} \end{array}\right) \end{array} \begin{array}{l} \\ 1 \\ d-1 \end{array} , \tag{2.11}$$

then the first row of R_2 is given by $\begin{pmatrix} \rho \ \tilde{u}^T \end{pmatrix}$. Now, if $r_i^T = \begin{pmatrix} 0^T \ \rho \ \tilde{u}^T \end{pmatrix}$, then $A^{(i+1)} = A^{(i)} - r_i r_i^T$, which has displacement given by

$$\begin{aligned} \nabla_F A^{(i+1)} &= G_i^T J G_i - r_i r_i^T + F r_i r_i^T F^T \\ &= \tilde{G}_i^T J \tilde{G}_i - r_i r_i^T + F r_i r_i^T F^T \\ &= G_{i+1}^T J G_{i+1}, \end{aligned}$$

where

$$G_{i+1} = \begin{pmatrix} \multicolumn{2}{c}{r_i^T F^T} \\ 0_{(r-1)\times i} & \tilde{V} \end{pmatrix} \tag{2.12}$$

has the same structure as (2.10), i.e., the first i elements of Fr_i are zero. Thus, given the generators of A in (2.6), the rows of the Cholesky factor R can be computed by iterating equations (2.11) and (2.12). The algorithm may breakdown if

$$\begin{pmatrix} u_1^T & v_1^T \end{pmatrix} J \begin{pmatrix} u_1 \\ v_1 \end{pmatrix} < 0.$$

At each step of the algorithm, a J-invariant matrix Q_i should be computed to annihilate the elements of v_1 in (2.10). The computation of this matrix plays a key role in the numerical stability of the whole algorithm [81]. In particular the number of hyperbolic transformations should be minimized. Here only one hyperbolic Givens rotation is used and this is in factored form [81].

Let $(x^T \ y^T) = (u_1^T \ v_1^T)$, where $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$. If $Q_x \in \mathbb{R}^{k\times k}$ and $Q_y \in \mathbb{R}^{l\times l}$ are two orthogonal matrices, then $Q_x \oplus Q_y$ is J-invariant. If Q_x and Q_y are the Householder transformations such that $Q_x^T x = \alpha e_1$ and $Q_y^T y = \beta e_1$, and $H \in \mathbb{R}^{2\times 2}$ is such that

$$H \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \rho \\ 0 \end{pmatrix}$$

and

$$H^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

then the matrix

$$Q = \begin{pmatrix} Q_x & 0 \\ 0 & Q_y \end{pmatrix} \begin{pmatrix} h_{11} & 0 & h_{12} & 0 \\ 0 & I_{k-1} & 0 & 0 \\ h_{21} & 0 & h_{22} & 0 \\ 0 & 0 & 0 & I_{l-1} \end{pmatrix}$$

is J-invariant and satisfies

$$Q^T \begin{pmatrix} x \\ y \end{pmatrix} = \rho e_1.$$

Notice that if no breakdown occurs, then the matrix H can be always computed as the hyperbolic Givens rotation

$$H = \frac{1}{c} \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix},$$

where $s = -\beta/\alpha$ and $c^2 = 1 - s^2$. In factored form this is

$$H = \begin{pmatrix} 1 & 0 \\ s & c \end{pmatrix} \begin{pmatrix} 1/c & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix},$$

which represents a stable implementation of the transformation H [81].

Algorithm 1 summarizes the Generalized Schur Algorithm. It needs to store only the generators G_i and the matrix R; the matrix $A^{(i)}$ is never computed explicitly. Supposing that the matrix–vector multiplication involving F is negligible (when F is a shift matrix) and using Householder matrices for Q_x and Q_y, the computational cost of the algorithm is dominated by the steps 5 and 7, which require $4d(n-i)$ and $6(n-i)$ flops, respectively [28]. Therefore, the complexity of the algorithm is $(2d+3)n^2$.

For some applications a block version of the algorithm is more appropriate. That is, if $A, F \in \mathbb{R}^{Nn \times Nn}$ are matrices with block–size $n \times n$, then F is block strictly lower triangular and A has displacement rank d w.r.t. F. A possible implementation of the Block Generalized Schur Algorithm (BGSA) is outlined in Algorithm 2. Steps 4–7 compute a J-invariant transformation Q_i such that

$$Q_i^T \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} R_i \\ 0 \end{pmatrix}. \tag{2.13}$$

Notice that other implementations exist for computing (2.13) each of which has different numerical properties. In Algorithm 2 the only critical part, concerning numerical stability, is the computation

Algorithm 1 Generalized Schur Algorithm

```
Input: The matrix of generators $G$ and the shift matrix $F$ as in (2.5) and (2.6)
Output: $R$, the upper triangular Cholesky factor of $A$
 1: Set $G_{(1)} = G$
 2: for $i = 1, 2, \ldots, n-1$ do
 3:     Let
            $$G_{(i)} = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccc} {\scriptstyle i-1} & {\scriptstyle 1} & {\scriptstyle n-i} & \\ \end{array}$$
            $$G_{(i)} = \left(\begin{array}{ccc|c} 0 & x & X & k \\ 0 & y & Y & l \end{array}\right).$$
 4:     Compute the Householder reflections $Q_x$ and $Q_y$ such that $Q_x^T x = \alpha e_1$ and $Q_y^T y = \beta e_1$, respectively.
 5:     Apply $Q_x$ and $Q_y$ to $X$ and $Y$, respectively, such that:
            $$\tilde{X} = Q_x^T X, \qquad \tilde{Y} = Q_y^T Y.$$
 6:     Compute a hyperbolic Givens transformation $H$ such that
            $$H^T \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho e_1.$$
 7:     Apply $H$ to the first rows of $\tilde{X}$ and $\tilde{Y}$ to obtain $\hat{X}$, $\hat{Y}$ and
            $$\hat{G}_{(i)} = \begin{pmatrix} 0 & \rho e_1 & \hat{X} \\ 0 & 0 & \hat{Y} \end{pmatrix}.$$
 8:     Store $(0 \ \cdots \ 0 \quad \rho \quad \hat{X} e_1)$ in the $i$th row of $R$.
 9:     $G_{(i+1)}$ is defined by $\hat{G}_{(i)}$ after its first row has been multiplied by $F$.
10: end for
```

of the hyperbolic transformation H in Step 6. The breakdown of the block algorithm [27] is related to the non-existence or the singularity of R_i in (2.13).

If Householder transformations are used, then the computational complexity of Steps 4 and 5 is $4n^2(N-i)(d-n)$ flops. The number of flops for Steps 6 and 7 using hyperbolic Givens rotations is given by

$$\begin{aligned} \sum_{j=1}^{n} \sum_{h=1}^{j} 6((N-i)n + h) &\simeq \sum_{j=1}^{n} 6(j(N-i)n + \frac{1}{2}j^2) \\ &\simeq 4(N-i)n^3. \end{aligned}$$

Algorithm 2 Generalized Schur Algorithm – Block Version

```
Input: The matrix of generators $G$ and the shift matrix $F$ as in (2.5) and (2.6)
Output: $R$, the upper triangular Cholesky factor of $A$
Let $G_{(1)} = G$ and $J = I_k \oplus -I_l$.
for $i = 1, 2, \ldots, N$ do
    Let
        $$G_i = \begin{array}{cccc} (i-1)n & n & (N-i)n & \\ \left(\begin{matrix} 0 \\ 0 \end{matrix}\right. & \begin{matrix} X_1 \\ Y_1 \end{matrix} & \left.\begin{matrix} X_2 \\ Y_2 \end{matrix}\right) & \begin{matrix} k \\ l \end{matrix} \end{array} .$$
    Compute the QRDs
        $$X_1 = Q_X \begin{pmatrix} R_X \\ 0 \end{pmatrix} \quad \text{and} \quad Y_1 = Q_Y \begin{pmatrix} R_Y \\ 0 \end{pmatrix}.$$
    Compute
        $$\begin{array}{c} n \\ k-n \end{array} \begin{pmatrix} \hat{X}_{2A} \\ \hat{X}_{2B} \end{pmatrix} = Q_X^T X_2 \quad \text{and} \quad \begin{array}{c} n \\ l-n \end{array} \begin{pmatrix} \hat{Y}_{2A} \\ \hat{Y}_{2B} \end{pmatrix} = Q_Y^T Y_2.$$
    Compute a J-invariant matrix $H$ such that
        $$H^T \begin{pmatrix} R_X \\ R_Y \end{pmatrix} = \begin{pmatrix} R_i \\ 0 \end{pmatrix}.$$
    Compute
        $$\begin{pmatrix} \check{X}_{2A} \\ \check{Y}_{2B} \end{pmatrix} = H^T \begin{pmatrix} \hat{X}_{2A} \\ \hat{Y}_{2B} \end{pmatrix}.$$
    Store $\begin{pmatrix} 0 & R_i & \check{X}_{2A} \end{pmatrix}$ in the $i$th block–row of $R$.
    Compute
        $$\begin{array}{cc} in & (N-i)n \\ \left(0\right. & \left.\hat{X}_A\right) \end{array} = \begin{array}{ccc} (i-1)n & n & (N-i)n \\ \left(0\right. & R_i & \left.\check{X}_{2A}\right) \end{array} F^T$$
    and form $G_{i+1}$ as:
        $$G_{i+1} = \begin{array}{cccc} (i-1)n & n & (N-i)n & \\ \left(\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix} & \left.\begin{matrix} \hat{X}_A \\ \tilde{X}_{2B} \\ \check{Y}_{2A} \\ \tilde{Y}_{2B} \end{matrix}\right) & \begin{matrix} n \\ k-n \\ n \\ l-n \end{matrix} \end{array} .$$
end for
```

Thus, the computational complexity of the whole procedure is given by

$$\begin{aligned} \sum_{i=1}^{N} (N-i)n^2 \left(4n + 4(d-n)\right) &\simeq \sum_{i=1}^{N} 4d(N-i)n^2 \\ &\simeq 2dN^2n^2. \end{aligned}$$

2.3 A fast algorithm for the OLS estimation of the VAR model

The BGSA shown by Algorithm 2 is designed to compute the Cholesky factorization of structured matrices. To compute the matrix R in (2.3) the algorithm should be applied to the matrix T^TT. Consider the block Toeplitz matrix $T \in \mathbb{R}^{Mm \times Nn}$ defined by

$$T = [T_{i-j}]_{i=1,\ldots,M}^{j=1,\ldots,N},$$

that is,

$$T = \begin{pmatrix} T_0 & T_{-1} & \cdots & T_{1-N} \\ T_1 & T_0 & \cdots & T_{2-N} \\ \vdots & \vdots & & \vdots \\ T_{M-1} & T_{M-2} & \cdots & T_{M-N} \end{pmatrix}, \tag{2.14}$$

where $T_k \in \mathbb{R}^{m \times n}$. Let $A = T^TT$, where the (i,j)th block $A_{ij} \in \mathbb{R}^{m \times n}$ is given by

$$A_{ij} = \sum_{k=1}^{M} T_{k-i}^T T_{k-j} = Y_{i-1}^T Y_{j-1}$$

and $Y_i^T = \left(T_{-i}^T \; T_{1-i}^T \; \cdots \; T_{M-i-1}^T\right)$ is the $(i+1)$th block column of T. If Y_0 has full column rank, then the displacement rank of A w.r.t. $Z_m = Z \otimes I_m$ is $2(m+n)$ and the generators (see Appendix 1) are given by $J = I_{m+n} \oplus -I_{m+n}$ and

$$G = \begin{array}{c} \begin{array}{cccc} n & n & \cdots & n \end{array} \\ \begin{pmatrix} R_0 & Q_0^T Y_1 & \cdots & Q_0^T Y_{N-1} \\ 0 & T_{-1} & \cdots & T_{1-N} \\ 0 & Q_0^T Y_1 & \cdots & Q_0^T Y_{N-1} \\ 0 & T_{M-1} & \cdots & T_{M+1-N} \end{pmatrix} \begin{array}{c} n \\ m \\ n \\ m \end{array} \end{array}, \tag{2.15}$$

where $Y_0 = Q_0 R_0$, $R_0 \in \mathbb{R}^{n \times n}$ is upper triangular and $Q_0 \in \mathbb{R}^{Mm \times n}$ is orthogonal.

If $M = \left(W \; T\right)$, where $W \in \mathbb{R}^{Mm \times \eta}$, then the displacement equation for $\bar{A} = M^TM$ w.r.t. the shift matrix $\bar{Z} = 0_{\eta \times \eta} \oplus Z_m$ is

$$\nabla_{\bar{Z}} \bar{A} = \begin{pmatrix} 0 & 0 \\ 0 & \nabla_{Z_m} T^T T \end{pmatrix} + \begin{pmatrix} W^T W & W^T T \\ T^T W & 0 \end{pmatrix},$$

so that the generators are given by $J = I_{\eta+m+n} \oplus -I_{\eta+m+n}$ and

$$G = \begin{array}{cc} \begin{array}{ccccc} \eta & n & n & \cdots & n \end{array} & \\ \begin{pmatrix} R_W & Q_W^T Y_0 & Q_W^T Y_1 & \cdots & Q_W^T Y_{N-1} \\ 0 & R_0 & Q_0^T Y_1 & \cdots & Q_0^T Y_{N-1} \\ 0 & 0 & T_{-1} & \cdots & T_{1-N} \\ 0 & Q_W^T Y_0 & Q_W^T Y_1 & \cdots & Q_W^T Y_{N-1} \\ 0 & 0 & Q_0^T Y_1 & \cdots & Q_0^T Y_{N-1} \\ 0 & 0 & T_{M-1} & \cdots & T_{M+1-n} \end{pmatrix} & \begin{array}{c} \eta \\ n \\ m \\ \eta \\ n \\ m \end{array} \end{array} . \tag{2.16}$$

The computational complexity of (2.16) can be approximated by $4n^2((n + mM/2)N + \eta)$. Notice that if the generators are given by (2.15), or by (2.16), then Steps 4–7 of Algorithm 2 are not needed for the first iteration.

In the specific case of the estimation of VAR models $m = 1$, $N = p + 1$, $T_k = y_{M-N-k}^T$, $Y_i^T = \begin{pmatrix} y_{M+i-p} & \cdots & y_{2+i-p} & y_{1+i-p} \end{pmatrix}$ and the displacement rank of A and $\bar{A}$ are $2(n + 1)$ and $2(n + \eta + 1)$, respectively. Thus, the computation of the Cholesky decomposition of $\bar{A}^T \bar{A}$ using Algorithm 2 requires $2(n + \eta)n^2p^2$ flops, and the computation of the generators requires a further $4n^2((n + M/2)p + \eta)$ flops.

2.4 VAR models with Zero Coefficient Constraints

The VAR model (2.2) can be written as the SURE model

$$\mathrm{Vec}(Y) = (I_n \otimes X)\,\mathrm{Vec}(B) + \mathrm{Vec}(U),$$

where $\mathrm{Vec}(U)$ has zero mean and variance–covariance matrix given by $\Sigma \otimes I_M$. Often zero coefficient constraints (ZCC) are imposed to VAR models, hereafter called VAR-ZCC model [58]. Thus, some elements of B are zero. Let $\beta_i \in \mathbb{R}^{k_i}$ the vector of nonzero elements in the ith column of B. The VAR-ZCC model can be written as the SURE model

$$\mathrm{Vec}(Y) = (\oplus_{i=1}^n X_i)\,\mathrm{Vec}(\{\beta_i\}_n) + \mathrm{Vec}(U), \tag{2.17}$$

where $X_i = XS_i$ and $S_i \in \mathbb{R}^{np \times k_i}$ is a selection matrix. Notice that the SURE model has common regressors.

The Best Linear Unbiased Estimator (BLUE) of $\operatorname{Vec}(\{\beta_i\})$ is obtained from the solution of the General Least Squares (GLS) problem

$$\underset{\beta_1,\ldots,\beta_n}{\operatorname{argmin}} \left\|\operatorname{Vec}(Y) - \operatorname{Vec}(\{X_i\beta_i\})\right\|_{\Sigma^{-1}\otimes I_M} \tag{2.18}$$

which is given by

$$\operatorname{Vec}(\{\hat{\beta}_i\}) = \left((\oplus_i X_i^T)(\Sigma^{-1}\otimes I_M)(\oplus_i X_i)\right)^{-1}(\oplus_i X_i^T)\operatorname{Vec}(Y\Sigma^{-1}). \tag{2.19}$$

The computation of (2.19) has poor numerical properties and does not allow for a singular dispersion matrix Σ. Therefore, it is preferable to formulate (2.18) as the Generalized Linear Least Squares Problem (GLLSP)

$$\underset{V,\{\beta_i\}}{\operatorname{argmin}} \left\|V\right\|_F \text{ subject to } \operatorname{Vec}(Y) = (\oplus_i X_i)\operatorname{Vec}(\{\beta_i\}) + \operatorname{Vec}(VC^T), \tag{2.20}$$

where $\|\cdot\|_F$ denotes the Frobenious norm, $\Sigma = CC^T$, the upper triangular $C \in \mathbb{R}^{n\times n}$ has full rank and the random matrix V is defined as $(C\otimes I_M)\operatorname{Vec}(V) = \operatorname{Vec}(U)$. That is, $VC^T = U$, which implies that $\operatorname{Vec}(V)$ has zero mean and variance-covariance matrix I_{nM} [50, 55, 61, 63]. Without loss of generality it will be assumed that Σ is non-singular.

Consider the Generalized QR decomposition (GQRD)

$$\mathcal{Q}^T(\oplus_i X_i) = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \tag{2.21a}$$

and

$$\mathcal{Q}^T(C\otimes I_M)\mathcal{P} = \begin{array}{c} \begin{array}{cc} K & nM-K \end{array} \\ \begin{pmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{pmatrix} \end{array} \begin{array}{l} \\ K \\ nM-K \end{array}, \tag{2.21b}$$

where $K = \sum_{i=1}^n k_i$, $R_i \in \Re^{k_i\times k_i}$ and W_{22} are upper triangular, and $\mathcal{Q},\mathcal{P} \in \mathbb{R}^{nM\times nM}$ are orthogonal [2, 64]. Using (2.21) the GLLSP (2.20) can be written as

$$\underset{\substack{\{v_{iA}\},\\ \{v_{iB}\},\{\beta_i\}}}{\operatorname{argmin}} \sum_{i=1}^{G} \left(\ \|v_{iA}\|_2 + \|v_{iB}\|_2\right) \quad \text{subject to}$$

$$\begin{pmatrix} \operatorname{Vec}(\{y_{iA}\}) \\ \operatorname{Vec}(\{y_{iB}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \operatorname{Vec}(\{\beta_i\}) + \begin{pmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{pmatrix} \begin{pmatrix} \operatorname{Vec}(\{v_{iA}\}) \\ \operatorname{Vec}(\{v_{iB}\}) \end{pmatrix}, \tag{2.22}$$

where

$$\mathcal{Q}^T \operatorname{Vec}(Y) = \begin{array}{c@{}l} \begin{pmatrix} \operatorname{Vec}(\{y_{iA}\}) \\ \operatorname{Vec}(\{y_{iB}\}) \end{pmatrix} & \begin{array}{l} K \\ nM-K \end{array} \end{array}$$

and

$$\mathcal{P}^T \operatorname{Vec}(V) = \begin{array}{c@{}l} \begin{pmatrix} \operatorname{Vec}(\{v_{iA}\}) \\ \operatorname{Vec}(\{v_{iB}\}) \end{pmatrix} & \begin{array}{l} K \\ nM-K \end{array} \end{array} .$$

From (2.22) it follows that $\operatorname{Vec}(\{v_{iB}\}) = W_{22}^{-1} \operatorname{Vec}(\{\widehat{y}_i\})$ and $\operatorname{Vec}(\{v_{iA}\}) = 0$. Thus, the solution for the SURE model comes from solving the triangular system

$$\begin{pmatrix} \operatorname{Vec}(\{y_{iA}\}) \\ \operatorname{Vec}(\{y_{iB}\}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i & W_{12} \\ 0 & W_{22} \end{pmatrix} \begin{pmatrix} \operatorname{Vec}(\{\beta_i\}) \\ \operatorname{Vec}(\{v_{iB}\}) \end{pmatrix} . \tag{2.23}$$

The matrix $\mathcal{Q}$ in (2.21) is defined as

$$\mathcal{Q} = \begin{pmatrix} \oplus_i Q_{iA} & \oplus_i Q_{iB} \end{pmatrix} ,$$

where

$$Q_i^T X_i = \begin{pmatrix} R_i \\ 0 \end{pmatrix} , \quad \text{with} \quad Q_i = \begin{array}{c} \begin{array}{cc} \;\;k_i & \;\;M-k_i \end{array} \\ \begin{pmatrix} Q_{iA} & \quad Q_{iB} \end{pmatrix} \end{array} ,$$

is the QRD of X_i ($i = 1, \ldots, G$).

The computation of (2.21b) occurs in two stages. The first stage computes

$$\mathcal{Q}^T (C \otimes I) \mathcal{Q} = \begin{array}{c@{}l} \begin{array}{cc} K & nM-K \end{array} & \\ \begin{pmatrix} \widetilde{W}_{11} & \widetilde{W}_{12} \\ \widetilde{W}_{21} & \widetilde{W}_{22} \end{pmatrix} & \begin{array}{l} K \\ nM-K \end{array} \end{array} , \tag{2.24}$$

where $\tilde{W}_{ij}$ ($i, j = 1, 2$) is block upper triangular. Furthermore, the main block–diagonals of $\tilde{W}_{12}$ and $\tilde{W}_{21}$ are zero, and the ith ($i = 1, \ldots, G$) blocks of the main diagonal of $\tilde{W}_{11}$ and $\tilde{W}_{22}$ are given by $C_{ii} I_{k_i}$ and $C_{ii} I_{M-k_i}$, respectively. The second stage computes the RQD

$$\begin{pmatrix} \widetilde{W}_{21} & \widetilde{W}_{22} \end{pmatrix} \widetilde{\mathcal{P}} = \begin{pmatrix} 0 & W_{22} \end{pmatrix} \tag{2.25a}$$

and

$$\begin{pmatrix} \widetilde{W}_{11} & \widetilde{W}_{12} \end{pmatrix} \widetilde{\mathcal{P}} = \begin{pmatrix} W_{11} & W_{12} \end{pmatrix} . \tag{2.25b}$$

Thus, in (2.21b) $\mathcal{P} = \mathcal{Q}\tilde{\mathcal{P}}$. Sequential and parallel strategies for computing the RQD (2.25) have been described in [44, 46].

The computations of (2.24) and (2.25) are the most expensive operations – their computational cost is in the order of n^3M^3 – and they become quite important when computing the feasible GLS (FGLS) estimators, that is, when computed at each iteration for different C. However, the common regressors that exist in the SURE model can be used to reduce the computational complexity of the iterative estimation procedure [54].

Consider the QRD given in (2.3). Premultiplying (2.17) from the left by $\bar{Q}^T = \begin{pmatrix} I_n \otimes Q_T & I_n \otimes Q_Y & I_n \otimes Q_N \end{pmatrix}^T$ gives

$$\begin{pmatrix} \mathrm{Vec}(R_{TY}) \\ \mathrm{Vec}(R_Y) \\ 0 \end{pmatrix} = \begin{pmatrix} \oplus_i RS_i \\ 0 \\ 0 \end{pmatrix} \mathrm{Vec}(\{\beta_i\}) + \begin{pmatrix} \mathrm{Vec}(U_T) \\ \mathrm{Vec}(U_Y) \\ \mathrm{Vec}(U_N) \end{pmatrix},$$

where

$$Q^T U = \begin{array}{c} n \\ \begin{pmatrix} U_T \\ U_Y \\ U_N \end{pmatrix} \end{array} \begin{array}{l} \\ np \\ n \\ M-(p+1)n \end{array} .$$

The covariance matrix of $\mathrm{Vec}(\begin{pmatrix} U_T & U_Y & U_N \end{pmatrix})$ is given by

$$\begin{pmatrix} \Sigma \otimes I_{np} & 0 & 0 \\ 0 & \Sigma \otimes I_n & 0 \\ 0 & 0 & \Sigma \otimes I_{M-n(p+1)} \end{pmatrix}.$$

Thus the estimator of the SURE model (2.17) arises from the solution of the reduced size model

$$\mathrm{Vec}(R_{TY}) = (\oplus_i R_T S_i)\,\mathrm{Vec}(\{\beta_i\}) + \mathrm{Vec}(U_T), \tag{2.26}$$

where the covariance matrix of $\mathrm{Vec}(U_M)$ is given by $\Sigma \otimes I_{np}$. The estimator of Σ is now given by

$$\hat{\Sigma} = \frac{1}{M}\left(\hat{U}_T^T \hat{U}_T + R_Y^T R_Y\right),$$

where $\hat{U}$ is the residual matrix estimated from (2.26). Notice that $R_Y^T R_Y$ does not depend on the covariance matrix Σ.

2.5 VAR models with Granger causality restrictions

Consider partitioning the time series z_t and the coefficient matrices Φ_l in (2.1) as

$$z_t = \begin{pmatrix} z_{At} \\ z_{Bt} \end{pmatrix} \qquad \Phi_l = \begin{pmatrix} \Phi_{AAl} & \Phi_{ABl} \\ \Phi_{BAl} & \Phi_{BBl} \end{pmatrix}.$$

If $\Phi_{BAl} = 0$ for $l = 1, 2, \ldots, p$, then the series z_{At} does not Granger–cause z_{Bt}; that is, z_{At} is not linearly informative about the future of the time series z_{Bt} [32, 58]. This concept can be generalized. Define the permutation $(\pi_1, \ldots, \pi_n)$ of the indices $(1, \ldots, n)$, and let $\Pi = \begin{pmatrix} e_{\pi_1} & e_{\pi_2} & \cdots & e_{\pi_n} \end{pmatrix}$ be the associated permutation matrix. Now, if there exists a permutation $(\pi_1, \ldots, \pi_n)$ such that

$$\Pi^T \Phi_l \Pi \sim \begin{array}{cc} \begin{array}{cccc} h_1 & h_2 & \cdots & h_s \end{array} & \\ \begin{pmatrix} \times & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \times \end{pmatrix} & \begin{array}{c} k_1 \\ k_2 \\ \vdots \\ k_s \end{array} \end{array}, \tag{2.27}$$

for $l = 1, 2, \ldots, p$, then the series z_{π_j} does not Granger–cause z_{π_l} when the (i, j)th element of the matrix in (2.27) belongs to the block lower triangular part.

When the Granger causality restrictions are known a priori, the variables can be ordered so that the matrices Φ_l have the structure given in (2.27). In this case the model is equivalent to a triangular SURE model [79, 51]. Conversely, if different models with different causality restriction have to be estimated, then a reordering is not convenient since the Toeplitz structure of X is destroyed. In this case the VAR model is equivalent to a SURE model with proper subset regressors [79].

Multiplying (2.2) on the right by Π gives

$$\tilde{Y} = X\tilde{B} + \tilde{U},$$

where $\tilde{Y} = Y\Pi$, $\tilde{U} = U\Pi$ and $\tilde{B} = B\Pi$. Figure 2.1 shows an example of the structure of $\tilde{B}^T$. The matrix $\tilde{B}$ is characterized by the property that if the (j, i)th element is zero, then also the $(j, i+1)$th element has to be zero. That is $\tilde{B} = \begin{pmatrix} S_1\beta_1 & S_2\beta_2 & \cdots & S_n\beta_n \end{pmatrix}$, where $S_i = \prod_{k=1}^{i} \hat{S}_k = S_{i-1}\hat{S}_i$ and $\hat{S}_k$ are selection matrices. The regressor matrices of the SURE model are defined as $X_i = XS_i = X_{i-1}\hat{S}_i$.

Consider the QRDs

$$X_i = \begin{pmatrix} Q_{iA} & Q_{iB} \end{pmatrix} \begin{pmatrix} R_i \\ 0 \end{pmatrix} \tag{2.28}$$

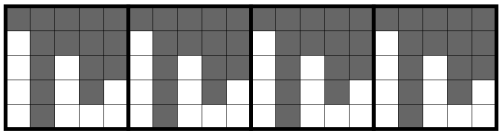


Figure 2.1: Proper Subset Structure derived from Granger causality restrictions.

and

$$\begin{pmatrix} R_i \\ 0 \end{pmatrix} \hat{S}_{i+1} = \begin{pmatrix} \bar{Q}_{i+1,A} & \bar{Q}_{i+1,B} \end{pmatrix} \begin{pmatrix} R_{i+1} \\ 0 \end{pmatrix}, \tag{2.29}$$

where $Q_i = \begin{pmatrix} Q_{iA} \; Q_{iB} \end{pmatrix}$ and $\bar{Q}_{i+1} = \begin{pmatrix} \bar{Q}_{i+1,A} \; \bar{Q}_{i+1,B} \end{pmatrix}$ are orthogonal matrices. The QRDs (2.28) and (2.29) imply

$$X_{i+1} = Q_i \begin{pmatrix} R_i \\ 0 \end{pmatrix} S_{i+1} = Q_i \bar{Q}_{i+1} \begin{pmatrix} R_{i+1} \\ 0 \end{pmatrix},$$

and $Q_{i+1} = Q_i \bar{Q}_{i+1}$. Notice that

$$\bar{Q}_{i+1} = \begin{pmatrix} \hat{Q}_{i+1} & 0 \\ 0 & I_{M-k_i} \end{pmatrix}.$$

In (2.24) for $x, y \in \{A, B\}$

$$\widetilde{W}_{xy} = \begin{pmatrix} c_{11} Q_{1x}^T Q_{1y} & c_{12} Q_{1x}^T Q_{2y} & \cdots & c_{1n} Q_{1x}^T Q_{ny} \\ 0 & c_{22} Q_{2x}^T Q_{2y} & \cdots & c_{2n} Q_{2x}^T Q_{ny} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{nn} Q_{nx}^T Q_{ny} \end{pmatrix}.$$

Thus, for $i < j$

$$\begin{aligned} Q_{ix}^T Q_{jy} &= \bar{Q}_{ix}^T \bar{Q}_{i-1}^T \cdots \bar{Q}_1^T \bar{Q}_1 \cdots \bar{Q}_{i-1} \bar{Q}_i \cdots \bar{Q}_{j-1} \bar{Q}_{jy} \\ &= \bar{Q}_{ix}^T \bar{Q}_i \cdots \bar{Q}_{j-1} \bar{Q}_{jy} \\ &= \begin{cases} \begin{pmatrix} I & 0 \end{pmatrix} \bar{Q}_{i+1} \cdots \bar{Q}_{j-1} \bar{Q}_{jy}, & \text{if } x = A, \\ \begin{pmatrix} 0 & I \end{pmatrix} \bar{Q}_{i+1} \cdots \bar{Q}_{j-1} \bar{Q}_{jy}, & \text{if } x = B, \end{cases} \end{aligned} \tag{2.30}$$

and for $i = j$

$$Q_{ix}^T Q_{jy} = \begin{cases} I_{k_i} & \text{if } x = y = A, \\ I_{M-k_i} & \text{if } x = y = B \text{ and} \\ 0 & \text{if } x \neq y. \end{cases} \tag{2.31}$$

From (2.30) and (2.31) it follows that $\widetilde{W}_{BA} = 0$, since

$$\begin{aligned} Q_{iB}^T Q_{jA} &= \begin{pmatrix} 0 & I_{M-k_{i-1}} \end{pmatrix} \left(\bar{Q}_{i+1} \cdots \bar{Q}_{j-1} \bar{Q}_{jA} \right) \\ &\sim \begin{array}{c} \begin{array}{cc} k_{i-1} & M-k_{i-1} \end{array} \\ \begin{pmatrix} 0 & \times \end{pmatrix} \end{array} \begin{array}{c} k_{j-1} \\ \begin{pmatrix} \times \\ 0 \end{pmatrix} \end{array} \begin{array}{l} k_{j-1} \\ M-k_{j-1} \end{array} \end{aligned} \tag{2.32}$$

and $k_{i-1} > k_{j-1}$. The matrix in (2.24) has an block upper triangular form and thus the RQD (2.25) is not needed, i.e., $\widetilde{W} = W$ and $\mathcal{P} = \mathcal{Q}$.

Let y_i denote the ith column of Y and note that

$$\begin{pmatrix} y_{iA} \\ y_{iB} \end{pmatrix} = Q_i^T y_i = \bar{Q}_i^T \bar{Q}_{i-1}^T \cdots \bar{Q}_1^T y_i.$$

The vectors y_{iA} and y_{iB} can be computed from the recursion

$$\left(\begin{matrix} y_{jA} \\ y_{jB} \end{matrix} \;\middle|\; \begin{matrix} y_{j+1}^{(j)} & \cdots & y_n^{(j)} \end{matrix} \right) = \bar{Q}_j^T \left(y_j^{(j-1)} \;\middle|\; y_{j+1}^{(j-1)} \quad \cdots \quad y_n^{(j-1)} \right),$$

or, in a more compact form,

$$\left(\begin{matrix} y_{jA} \\ y_{jB} \end{matrix} \;\middle|\; Y^{(j)} \right) = \bar{Q}_j^T Y^{(j-1)}, \tag{2.33}$$

where $Y^{(0)} = \begin{pmatrix} y_1^{(0)} & y_2^{(0)} & \cdots & y_n^{(0)} \end{pmatrix} = Y$. Notice that the multiplication by $\bar{Q}_j^T$ in (2.33) affects only the first k_{j-1} rows of $Y^{(j-1)}$.

Consider the case of $n = 4$. The triangular matrix in (2.23) is

$$\begin{pmatrix} R_1 & 0 & 0 & 0 & 0 & c_{12}(I\ 0)\bar{Q}_{2B} & c_{13}(I\ 0)\bar{Q}_2\bar{Q}_{3B} & c_{14}(I\ 0)\bar{Q}_2\bar{Q}_3\bar{Q}_{4B} \\ 0 & R_2 & 0 & 0 & 0 & 0 & c_{23}(I\ 0)\bar{Q}_{3B} & c_{24}(I\ 0)\bar{Q}_3\bar{Q}_{4B} \\ 0 & 0 & R_3 & 0 & 0 & 0 & 0 & c_{34}(I\ 0)\bar{Q}_{4B} \\ 0 & 0 & 0 & R_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{11}I & c_{12}(0\ I)\bar{Q}_{2B} & c_{13}(0\ I)\bar{Q}_2\bar{Q}_{3B} & c_{14}(0\ I)\bar{Q}_2\bar{Q}_3\bar{Q}_{4B} \\ 0 & 0 & 0 & 0 & 0 & c_{22}I & c_{23}(0\ I)\bar{Q}_{3B} & c_{24}(0\ I)\bar{Q}_3\bar{Q}_{4B} \\ 0 & 0 & 0 & 0 & 0 & 0 & c_{33}I & c_{34}(0\ I)\bar{Q}_{4B} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_{44}I \end{pmatrix}$$

and rearranging the terms of (2.23) gives,

$$\begin{pmatrix} \begin{matrix} y_{1A} \\ y_{1B} \end{matrix} \\ \begin{matrix} y_{2A} \\ y_{2B} \end{matrix} \\ \begin{matrix} y_{3A} \\ y_{3B} \end{matrix} \\ \begin{matrix} y_{4A} \\ y_{4B} \end{matrix} \end{pmatrix} = \begin{pmatrix} \begin{matrix} R_1 \\ 0 \end{matrix} & \begin{matrix} 0 \\ c_{11}I \end{matrix} & 0 & c_{12}\bar{Q}_{2B} & 0 & c_{13}\bar{Q}_2\bar{Q}_{3B} & 0 & c_{14}\bar{Q}_2\bar{Q}_3\bar{Q}_{4B} \\ 0 & 0 & \begin{matrix} R_2 \\ 0 \end{matrix} & \begin{matrix} 0 \\ c_{22}I \end{matrix} & 0 & c_{23}\bar{Q}_{3B} & 0 & c_{24}\bar{Q}_3\bar{Q}_{4B} \\ 0 & 0 & 0 & 0 & \begin{matrix} R_3 \\ 0 \end{matrix} & \begin{matrix} 0 \\ c_{33}I \end{matrix} & 0 & c_{34}\bar{Q}_{4B} \\ 0 & 0 & 0 & 0 & 0 & 0 & \begin{matrix} R_4 \\ 0 \end{matrix} & \begin{matrix} 0 \\ c_{44}I \end{matrix} \end{pmatrix} \begin{pmatrix} \begin{matrix} \beta_1 \\ v_{1B} \end{matrix} \\ \begin{matrix} \beta_2 \\ v_{2B} \end{matrix} \\ \begin{matrix} \beta_3 \\ v_{3B} \end{matrix} \\ \begin{matrix} \beta_4 \\ v_{4B} \end{matrix} \end{pmatrix}.$$

Algorithm 3 Recursive solution of the SURE model with proper subset regressors

```
Input: The matrix $R_{TY}$, the upper triangular matrix $R_T$ and the selection matrices $S_2, \ldots, S_n$
Output: The vectors of parameters $\beta_1, \ldots, \beta_n$.
 1: Compute the QRD (2.3).
 2: Set $R_1 = R_T$ and $\tilde{Y} = R_{TY}$.
 3: for $i = 2, \ldots, n$ do
 4:     Compute the QRD: $R_{i-1} S_i = \hat{Q}_i \begin{pmatrix} R_i \\ 0 \end{pmatrix}$.
 5:     Compute $\tilde{Y}_{1:k_{i-1}, i:n} = \hat{Q}_i^T \tilde{Y}_{1:k_{i-1}, i:n}$.
 6: end for
 7: for $i = n, n-1, \ldots, 1$ do
 8:     Solve $R_i \beta_i = \tilde{Y}_{1:k_i, i}$.
 9:     Compute $v_{iB} = \tilde{Y}_{k_i+1:, i} / c_{ii}$.
10:     for $j = i-1, i-2, \ldots, 1$ do
11:         if $j = i-1$ then
12:             Compute $\tilde{v}_i = \bar{Q}_{iB} v_{iB}$.
13:         else
14:             Compute $\tilde{v}_i = \bar{Q}_j \tilde{v}_i$.
15:         end if
16:         Compute $\tilde{y}_j = \tilde{y}_j - c_{ji} \tilde{v}_i$.
17:     end for
18: end for
```

Conversely, for $i = 1, \ldots, n$

$$\begin{pmatrix} y_{iA} \\ y_{iB} \end{pmatrix} = \begin{pmatrix} R_i & 0 \\ 0 & c_{ii} I \end{pmatrix} \begin{pmatrix} \beta_i \\ v_{iB} \end{pmatrix} + \sum_{j=i+1}^{n} c_{ij} \left(\bar{Q}_{i+1} \bar{Q}_{i+2} \cdots \bar{Q}_{j-1} \right) \bar{Q}_{jB} v_{jB}.$$

This system is solved by a back-substitution without forming the matrices W_{AB} and W_{BB} as shown by Algorithm 3.

In order to optimize the memory access and exploit cache effects, the access to the matrices $\bar{Q}_i$ of Steps 12 and 14 of Algorithm 3 should be reorganized [82]. Updating the vectors y_j is done in one step for each i and the vectors $\tilde{v}_i$ are computed recursively by the formulae

$$\tilde{v}_i^i = \begin{pmatrix} 0 \\ v_{iB} \end{pmatrix} \quad \text{and} \quad \tilde{v}_j^{i-1} = \bar{Q}_i^T \tilde{v}_j^i.$$

2.6 Conclusions

Algorithms for solving VAR models have been proposed and analyzed. The VAR models with zero coefficient constraints or Granger–caused variables have been considered as SURE models with common or proper subset regressors, respectively. The numerically stable algorithms have efficiently exploited the Toeplitz, Kronecker, and other structures of the matrices in these models. Furthermore, the proposed algorithms can handle ill-conditioned problems.

The implementation of the algorithms needs to be investigated. Block versions of the algorithms which are suitable for conventional hight performance computers need to be designed [17]. The adaptation of the numerical methods to tackle other models that have similar matrix structures as those proposed here needs to be considered. Currently the Vector Error Correction Model (VECM) and the Johansen procedure for estimating cointegrated systems are investigated. The VECM has a structure similar to that of (2.4), while the Johansen procedure requires the OLS estimation of a linear system having a block Toeplitz structure [36, 37, 58].

2.A Displacement structures derived from LS problems involving block Toeplitz matrices.

Consider the block Toeplitz matrix $T \in \mathbb{R}^{Mm \times Nn}$ defined by

$$T = [T_{i-j}]_{i=1,\dots,M}^{j=1,\dots,N},$$

that is

$$T = \begin{pmatrix} T_0 & T_{-1} & \cdots & T_{1-N} \\ T_1 & T_0 & \cdots & T_{2-N} \\ \vdots & \vdots & & \vdots \\ T_{M-1} & T_{M-2} & \cdots & T_{M-N} \end{pmatrix},$$

where $T_k \in \mathbb{R}^{m \times n}$. Let define $A = T^T T = [A_{ij}]_{ij} \in \mathbb{R}^{Mm \times Mm}$, having blocks $A_{ij} \in \mathbb{R}^{m \times m}$ and given by

$$A_{ij} = \sum_{k=1}^{M} T_{k-i}^T T_{k-j} = Y_{i-1}^T Y_{j-1}, \tag{2.34}$$

where $Y_i^T = \begin{pmatrix} T_{-i}^T \; T_{1-i}^T \; \cdots \; T_{M-i-1}^T \end{pmatrix}$ is the $(i+1)$th block column of T.

The displacement of A w.r.t. $Z_m = Z \otimes I_m$ is given by $\nabla_{Z_m} A = M_1 + M_2$, where

$$M_1 = \left(\begin{array}{c|ccc} A_{11} & A_{12} & \cdots & A_{1n} \\ \hline A_{12}^T & & & \\ \vdots & & 0 & \\ A_{1n}^T & & & \end{array}\right)$$

and

$$M_2 = \left(\begin{array}{c|c} 0 & 0 \\ \hline & \\ 0 & A_{ij} - A_{i-1,j-1} \\ & \end{array}\right).$$

From (2.34) it follows that

$$\begin{aligned} A_{ij} - A_{i-1,j-1} &= \sum_{k=1}^{M} T_{k-i}^T T_{k-j} - \sum_{k=1}^{M} T_{k-i+1}^T T_{k-j+1} \\ &= T_{1-i}^T T_{1-j} - T_{M+1-i}^T T_{M+1-j}, \end{aligned}$$

for $i, j = 2, 3, \ldots, n$, and

$$M_2 = G_2^T G_2 - G_4^T G_4, \tag{2.35}$$

where $G_2 = \begin{pmatrix} 0 \; T_{-1} & \cdots & T_{1-N} \end{pmatrix}$ and $G_4 = \begin{pmatrix} 0 \; T_{M-1} & \cdots & T_{M+1-N} \end{pmatrix}$.

If Y_0 has full column rank and its QRD is $Q_0 R_0$, then

$$M_1 = G_1^T G_1 - G_3^T G_3, \tag{2.36}$$

where $G_1 = \begin{pmatrix} R_0 & Q_0^T Y_1 & \cdots & Q_0^T Y_{N-1} \end{pmatrix}$ and $G_3 = \begin{pmatrix} 0 & Q_0^T Y_1 & \cdots & Q_0^T Y_{N-1} \end{pmatrix}$. From (2.35) and (2.36) it follows that the displacement of A is given by

$$\nabla_{Z_p} A = \begin{pmatrix} G_1^T & G_2^T & G_3^T & G_4^T \end{pmatrix} \begin{pmatrix} I_n & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 \\ 0 & 0 & -I_n & 0 \\ 0 & 0 & 0 & -I_m \end{pmatrix} \begin{pmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \end{pmatrix}.$$

Chapter 3

Estimating seemingly unrelated regression models with vector autoregressive disturbances

Abstract:

The numerical solution of seemingly unrelated regression (SUR) models with vector autoregressive disturbances is considered. Initially an orthogonal transformation is applied to reduce the model to one with smaller dimensions. The transformed model is expressed as a reduced-size SUR model with stochastic constraints. The generalized QR decomposition is used as the main computational tool to solve this model. An iterative estimation algorithm is proposed when the variance-covariance matrix of the disturbances and the matrix of autoregressive coefficients are unknown. Strategies to compute the orthogonal factorizations of the non-dense structured matrices which arise in the estimation procedure are presented. Experimental results demonstrate the computational efficiency of the proposed algorithm.

---

[1] This chapter is a reprint of the paper: P. Foschi, E.J. Kontoghiorghes. Estimating seemingly unrelated regression models with vector autoregressive disturbances. *Journal of Economics Dynamics and Control*, 2003 (In press).

3.1 Introduction

The seemingly unrelated regression (SUR) model is given by

$$y_i = X_i\beta_i + u_i, \qquad i = 1, 2, \ldots, G, \tag{3.1}$$

where $y_i \in \mathbb{R}^M$ is the endogenous vector, $X_i \in \mathbb{R}^{M\times k_i}$ is the exogenous matrix with full column rank, $\beta_i \in \mathbb{R}^{k_i}$ are the coefficients and $u_i \in \mathbb{R}^M$ is the disturbance vector having zero mean. The covariance matrix of u_i and u_j is given by $\sigma_{ij}I_M$ ($i, j = 1, \ldots, G$). In compact form the SUR model can be written as

$$\operatorname{Vec}(Y) = \left(\oplus_{i=1}^{G} X_i\right) \operatorname{Vec}(\{\beta_i\}_G) + \operatorname{Vec}(U), \tag{3.2}$$

where $Y = \begin{pmatrix} y_1 & \cdots & y_G \end{pmatrix}$, $U = \begin{pmatrix} u_1 & \cdots & u_G \end{pmatrix}$, $\oplus_{i=1}^{G} X_i = \operatorname{diag}(X_1, \ldots, X_G)$, $\{\beta_i\}_G$ denotes the set of vectors $\beta_1, \ldots, \beta_G$ and $\operatorname{Vec}(\cdot)$ is the vector operator which stacks one column under the other of its matrix or set of vectors argument. The disturbance term $\operatorname{Vec}(U) \sim (0, \Sigma \otimes I_M)$, i.e., it has zero mean and dispersion matrix $\Sigma \otimes I_M$, where $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{G\times G}$ is symmetric and positive definite [3, 30, 69]. The subscript G in the set operator $\{\cdot\}$ is dropped and $\oplus_{i=1}^{G}$ is abbreviated to $\oplus_i$.

Often the SUR model has vector (VAR) or scalar (AR) autoregressive disturbances [26, 38, 43, 58, 65, 67, 79, 85]. In such cases the disturbance matrix U in (3.2) satisfies

$$U - ZUA^T = E \tag{3.3a}$$

or, equivalently,

$$(I_{GM} - A \otimes Z) \operatorname{Vec}(U) = \operatorname{Vec}(E), \tag{3.3b}$$

where $A \in \mathbb{R}^{G\times G}$ is the matrix of the AR coefficients, the $M \times M$ shift matrix Z is defined as

$$Z = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

and $E \in \mathbb{R}^{M\times G}$. The dispersions of $\operatorname{Vec}(E)$ and $\operatorname{Vec}(U)$ are given, respectively, by $\Sigma \otimes I_M$ and

$$(I_{GM} - A \otimes Z)^{-1} (\Sigma \otimes I_M) (I_{GM} - A \otimes Z)^{-T}, \tag{3.4}$$

where $-T$ denotes the transpose of the inverse.

Now, premultiplying the SUR with VAR disturbances (hereafter SUR-VAR) model (3.2) by $(I_{GM} - A \otimes Z)$ it gives

$$\mathrm{Vec}(Y - ZYA^T) = (I_G \otimes X - A \otimes ZX)\,(\oplus_i S_i)\,\mathrm{Vec}\,(\{\beta_i\}) + \mathrm{Vec}(E), \tag{3.5a}$$

or the general linear model (GLM)

$$\mathrm{Vec}(\Psi) = F\,\mathrm{Vec}\,(\{\beta_i\}) + \mathrm{Vec}(E), \tag{3.5b}$$

where $X \in \mathbb{R}^{M \times K^d}$ denotes the matrix comprising the K^d distinct regressors of the SUR model and S_i is a $K^d \times k_i$ selection matrix such that $X_i = XS_i$ ($i = 1, \ldots, G$). The matrices Ψ and F are defined by context. Notice that in the case where there are no common regressors, X and S_i are given by $X = \begin{pmatrix} X_1 & X_2 & \cdots & X_G \end{pmatrix}$ and $S_i = \begin{pmatrix} 0 & I_{k_i} & 0 \end{pmatrix}$, respectively. Furthermore, the matrix F in (3.5b) is full rather than block-diagonal as in the case of the conventional SUR model. The estimator of the SUR-VAR model (3.2) derives from computing the generalized least squares (GLS) estimator

$$\mathrm{Vec}\,(\{\hat{\beta}_i\}) = \big(F^T(\Sigma^{-1} \otimes I_M)F\big)^{-1} F^T \big(\Sigma^{-1} \otimes I_M\big)\,\mathrm{Vec}(\Psi). \tag{3.6}$$

In the case where Σ and A are unknown, an iterative procedure is used to derive the feasible GLS estimator. Initially, $\mathrm{Vec}\,(\{\hat{\beta}_i\})$ is computed from (3.6) based on some initial estimates of Σ and A. The residuals provide new estimates for Σ and A, which are then used in (3.6) to derive a new GLS estimator. This procedure is repeated until convergence.

The existing methods for computing the GLS, or the feasible GLS, estimator of the SUR-VAR model solve the normal equation (3.6) explicitly by computing Kronecker products and inverses of matrices [43, 79]. This results in computationally expensive and numerically unstable estimation procedures [76]. The purpose of this work is to provide computationally efficient algorithms for computing the GLS estimator for the SUR-VAR model. These algorithms use non-literal Kronecker operations and for numerical stability use orthogonal factorizations [28].

In the next section the numerical solution to the SUR-VAR model is considered and an iterative algorithm to compute the feasible GLS estimator is proposed. Section 3.3 considers various strategies for computing the matrix factorizations arising in the estimation procedures of the model. Computational results are shown in section 3.4. Finally conclusions and future research are presented.

3.2 Numerical solution of SUR-VAR models

Consider the QRD

$$\begin{array}{c} \begin{array}{cccc} K^d & G & K^d & G \end{array} \\ \begin{pmatrix} ZX & ZY & X & Y \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{ccc} K^* & K^* & M^* \end{array} \\ \begin{pmatrix} Q_A^* & Q_B^* & Q_C^* \end{pmatrix} \end{array} \begin{array}{cc} \begin{array}{cc} K^* & K^* \end{array} & \\ \begin{pmatrix} \widetilde{R}_A^* & \widetilde{R}_A \\ 0 & \widetilde{R}_B \\ 0 & 0 \end{pmatrix} & \begin{array}{c} K^* \\ K^* \\ M^* \end{array} \end{array} , \tag{3.7}$$

where $K^* = K^d + G$, $M^* = M - 2K^*$, $\widetilde{R}_A^*$ and $\widetilde{R}_B$ are upper-triangular defined by

$$\widetilde{R}_A^* = \begin{array}{c} \begin{array}{cc} K^d & G \end{array} \\ \begin{pmatrix} R_A^* & Y_A^* \end{pmatrix} \end{array}, \quad \widetilde{R}_A = \begin{array}{c} \begin{array}{cc} K^d & G \end{array} \\ \begin{pmatrix} R_A & Y_A \end{pmatrix} \end{array}, \quad \widetilde{R}_B = \begin{array}{c} \begin{array}{cc} K^d & G \end{array} \\ \begin{pmatrix} R_B & Y_B \end{pmatrix} \end{array}$$

and the orthogonal matrix $Q^* \in \mathbb{R}^{M \times M}$ is partitioned as

$$Q^* = \begin{pmatrix} Q_A^* & Q_B^* & Q_C^* \end{pmatrix} . \tag{3.8}$$

Pre-multiplying (3.5) by the orthogonal matrix $\begin{pmatrix} I_G \otimes Q_A^* & I_G \otimes Q_B^* & I_G \otimes Q_C^* \end{pmatrix}^T$ gives

$$\begin{pmatrix} \text{Vec}(\Psi_A) \\ \text{Vec}(Y_B) \\ 0 \end{pmatrix} = \begin{pmatrix} I_G \otimes R_A - A \otimes R_A^* \\ I_G \otimes R_B \\ 0 \end{pmatrix} (\oplus_i S_i) \, \text{Vec}\left(\{\beta_i\}\right) + \begin{pmatrix} \text{Vec}(E_A) \\ \text{Vec}(E_B) \\ \text{Vec}(E_C) \end{pmatrix} , \tag{3.9}$$

where

$$\Psi_A = Y_A - Y_A^* A^T \tag{3.10}$$

and

$$\begin{pmatrix} \text{Vec}(E_A) \\ \text{Vec}(E_B) \\ \text{Vec}(E_C) \end{pmatrix} \sim \left(0, \quad \begin{pmatrix} \Sigma \otimes I_{K^*} & 0 & 0 \\ 0 & \Sigma \otimes I_{K^*} & 0 \\ 0 & 0 & \Sigma \otimes I_{M^*} \end{pmatrix} \right) . \tag{3.11}$$

From (3.11) it follows that (3.9), and consequently the SUR-VAR model, can be written as the SUR model

$$\text{Vec}(Y_B) = (\oplus_i R_{Bi}) \, \text{Vec}\left(\{\beta_i\}\right) + \text{Vec}(E_B) \tag{3.12a}$$

with the stochastic constraints

$$\mathrm{Vec}(\Psi_A) = F_A\,\mathrm{Vec}\left(\{\beta_i\}\right) + \mathrm{Vec}(E_A), \tag{3.12b}$$

where $R^*_{Ai} = R^*_A S_i$, $R_{Ai} = R_A S_i$, $R_{Bi} = R_B S_i$ and

$$F_A = (\oplus_i R_{Ai}) - (A \otimes I_{K^*})(\oplus_i R^*_{Ai}). \tag{3.13}$$

Notice that this orthogonal transformation has reduced the size of the SUR-VAR model by GM^*. Computationally this is very significant for large–scale models where M is much bigger than the number of distinct regressors, i.e., $M \gg K^d$ [48].

The model (3.12) can be reformulated as the generalized linear least-squares problem (GLLSP):

$$\underset{\{\beta_i\},V_A,V_B}{\mathrm{argmin}}\ \left\|V_A\right\|_F^2 + \left\|V_B\right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix}\mathrm{Vec}(Y_B)\\ \mathrm{Vec}(\Psi_A)\end{pmatrix} = \begin{pmatrix}\oplus_i R_{Bi}\\ F_A\end{pmatrix}\mathrm{Vec}\left(\{\beta_i\}\right) + \begin{pmatrix}C\otimes I_{K^*} & 0\\ 0 & C\otimes I_{K^*}\end{pmatrix}\begin{pmatrix}\mathrm{Vec}(V_B)\\ \mathrm{Vec}(V_A)\end{pmatrix}, \tag{3.14}$$

where $\|\cdot\|_F$ denotes the *Frobenius* norm, $\Sigma = CC^T$, $C \in \mathbb{R}^{G\times G}$ is upper triangular, $V_A C^T = E_A$, $V_B C^T = E_B$ and $\mathrm{Vec}(\begin{pmatrix}V_B & V_A\end{pmatrix}) \sim (0, I_{2K^*})$ [55, 61, 62]. The solution to (3.14) is computed in two stages and provides the GLS estimator of (3.12). First, as in the case of the conventional SUR model, the GQRD of $\oplus_i R_{Bi}$ and $C \otimes I_{K^*}$ is computed:

$$Q^T\left(\oplus_i R_{Bi}\right) = \begin{array}{c@{}l} \scriptstyle K & \\ \begin{pmatrix}\oplus_i R_i\\ 0\end{pmatrix} & \begin{array}{l}\scriptstyle K\\ \scriptstyle GK^*-K\end{array}\end{array} \tag{3.15a}$$

and

$$Q^T\left(C\otimes I_{K^*}\right)P = W \equiv \begin{array}{c@{}l} \begin{array}{cc}\scriptstyle K & \scriptstyle GK^*-K\end{array} & \\ \begin{pmatrix}W_{11} & W_{13}\\ 0 & W_{33}\end{pmatrix} & \begin{array}{l}\scriptstyle K\\ \scriptstyle GK^*-K\end{array}\end{array}. \tag{3.15b}$$

Here, $K = \sum_i k_i$, Q and P are orthogonal, $\oplus_i R_i$ and W are upper triangular, and R_i is the triangular factor in the QRD of R_{Bi} [46]. The second stage computes the updated GQRD:

$$Q_u^T\begin{pmatrix}\oplus_i R_i\\ F_A\end{pmatrix} = \begin{array}{c@{}l} \scriptstyle K & \\ \begin{pmatrix}R_u\\ 0\end{pmatrix} & \begin{array}{l}\scriptstyle K\\ \scriptstyle GK^*\end{array}\end{array} \tag{3.16a}$$

and

$$Q_u^T \begin{pmatrix} W_{11} & 0 \\ 0 & C \otimes I_{K^*} \end{pmatrix} P_u = W_u \equiv \begin{array}{c@{}l} \begin{array}{cc} \scriptstyle K & \scriptstyle GK^* \end{array} & \\ \begin{pmatrix} W_{11}^u & W_{12}^u \\ 0 & W_{22}^u \end{pmatrix} & \begin{array}{l} \scriptstyle K \\ \scriptstyle GK^* \end{array} \end{array} , \tag{3.16b}$$

where R_u and W_u are upper triangular, and the orthogonal matrices Q_u and P_u are of order $(K + GK^*)$.

Now, let

$$Q^T \operatorname{Vec}(Y_B) = \begin{pmatrix} \operatorname{Vec}(\{\check{y}_i\}) \\ \operatorname{Vec}(\{\hat{y}_i\}) \end{pmatrix} \begin{array}{l} \scriptstyle K \\ \scriptstyle GK^*-K \end{array} , \tag{3.17a}$$

$$Q_u^T \begin{pmatrix} \operatorname{Vec}(\{\check{y}_i\}) \\ \operatorname{Vec}(\Psi_A) \end{pmatrix} = \begin{pmatrix} \operatorname{Vec}(\{\tilde{y}_i\}) \\ \operatorname{Vec}(\widetilde{Y}_A) \end{pmatrix} \begin{array}{l} \scriptstyle K \\ \scriptstyle GK^* \end{array} , \tag{3.17b}$$

$$Q_u^T \begin{pmatrix} W_{13} \\ 0 \end{pmatrix} = \begin{array}{c@{}l} \scriptstyle GK^*-K & \\ \begin{pmatrix} W_{13}^u \\ W_{23}^u \end{pmatrix} & \begin{array}{l} \scriptstyle K \\ \scriptstyle GK^* \end{array} \end{array} , \tag{3.17c}$$

$$P^T \operatorname{Vec}(V_B) = \begin{pmatrix} \operatorname{Vec}(\{\check{v}_i\}) \\ \operatorname{Vec}(\{\hat{v}_i\}) \end{pmatrix} \begin{array}{l} \scriptstyle K \\ \scriptstyle GK^*-K \end{array}$$

and

$$P_u^T \begin{pmatrix} \operatorname{Vec}(\{\check{v}_i\}) \\ \operatorname{Vec}(V_A) \end{pmatrix} = \begin{pmatrix} \operatorname{Vec}(\{\tilde{v}_i\}) \\ \operatorname{Vec}(\widetilde{V}_A) \end{pmatrix} \begin{array}{l} \scriptstyle K \\ \scriptstyle GK^* \end{array} .$$

Thus, the GLLSP (3.14) can be reformulated as

$$\underset{\substack{\{\beta_i\},\{\tilde{v}_i\}, \\ \{\hat{v}_i\},\widetilde{V}_A}}{\operatorname{argmin}} \sum_{i=1}^{G} \left(\left\| \tilde{v}_i \right\|^2 + \left\| \hat{v}_i \right\|^2 \right) + \left\| \widetilde{V}_A \right\|_F^2 \quad \text{subject to} \tag{3.18a}$$

$$\begin{pmatrix} \operatorname{Vec}(\{\tilde{y}_i\}) \\ \operatorname{Vec}(\widetilde{Y}_A) \\ \operatorname{Vec}(\{\hat{y}_i\}) \end{pmatrix} = \begin{pmatrix} R_u \\ 0 \\ 0 \end{pmatrix} \operatorname{Vec}(\{\beta_i\}) + \begin{pmatrix} W_{11}^u & W_{12}^u & W_{13}^u \\ 0 & W_{22}^u & W_{23}^u \\ 0 & 0 & W_{33} \end{pmatrix} \begin{pmatrix} \operatorname{Vec}(\{\tilde{v}_i\}) \\ \operatorname{Vec}(\widetilde{V}_A) \\ \operatorname{Vec}(\{\hat{v}_i\}) \end{pmatrix} , \tag{3.18b}$$

with solution given by

$$\begin{pmatrix} R_u & W_{12}^u & W_{13}^u \\ 0 & W_{22}^u & W_{23}^u \\ 0 & 0 & W_{33} \end{pmatrix} \begin{pmatrix} \operatorname{Vec}(\{\beta_i\}) \\ \operatorname{Vec}(\widetilde{V}_A) \\ \operatorname{Vec}(\{\hat{v}_i\}) \end{pmatrix} = \begin{pmatrix} \operatorname{Vec}(\{\tilde{y}_i\}) \\ \operatorname{Vec}(\widetilde{Y}_A) \\ \operatorname{Vec}(\{\hat{y}_i\}) \end{pmatrix} . \tag{3.19}$$

The covariance matrix of the coefficient estimator $\operatorname{Vec}(\{\beta_i\})$ is given by $R_u^{-1}W_{11}W_{11}^T R_u^{-T}$ [46, 50, 52, 55, 61].

In the case where there are no common regressors, it follows from (3.7) that R_{Ai}^* and R_{Bi} have the same structure. That is,

$$R_{Ai}^* = \begin{array}{c@{}l} \left(\begin{matrix} \widetilde{R}_{Ai}^* \\ 0 \end{matrix}\right) & \begin{matrix} {\scriptstyle k^{(i)}} \\ {\scriptstyle K^*-k^{(i)}} \end{matrix} \end{array} \qquad \text{and} \qquad R_{Bi} = \begin{array}{c@{}l} \left(\begin{matrix} \widetilde{R}_{Bi} \\ 0 \end{matrix}\right) & \begin{matrix} {\scriptstyle k^{(i)}} \\ {\scriptstyle K^*-k^{(i)}} \end{matrix} \end{array} ,$$

where $k^{(i)} = \sum_{j=1}^{i} k_j$ and the last $k_i \times k_i$ blocks of $\widetilde{R}_{Ai}^*$ and $\widetilde{R}_{Bi}$ are upper triangular. Notice that if $a \in \mathbb{R}^G$, then $(a^T \otimes I_{K^*})(\oplus_i R_{Ai}^*) = R_A^*(\oplus_i a_i I_{k_i}) \in \mathbb{R}^{K^* \times K^d}$ is upper triangular. Thus, $(A \otimes I_{K^*})(\oplus_i R_{Ai}^*)$ consists of G upper-triangular blocks stacked one atop the other. Figure 3.1 shows this structure and that of F_A for the case $G = 5$.

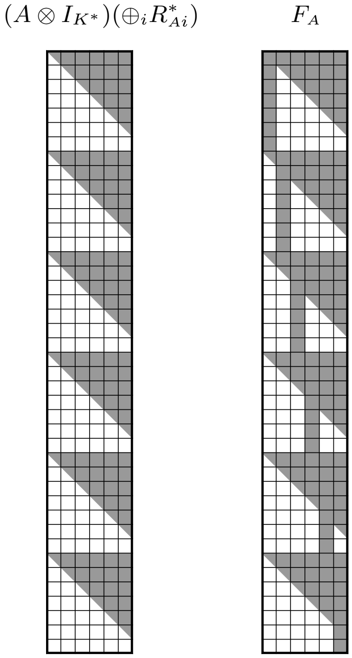


Figure 3.1: Structure of $(A \otimes I)(\oplus_i R_{Ai}^*)$ and F_A, where there are no common regressors and $G = 5$.

Notice that the QRD (3.7) does not depend on the value of either Σ or A. Furthermore, the value of the autoregressive parameter matrix A does not affect the SUR model (3.12a), but only the constraints (3.12b). Thus, for the derivation of the iterative feasible GLS estimator, initially, the GQRD (3.15) is computed to solve the SUR model (3.12a), and then the updated GQRD (3.16) is obtained. This updating is performed whenever A changes, while (3.15) is re-computed when the covariance matrix Σ is modified. Algorithm 4 summarizes the steps of this iterative procedure. The precision and condition that determines the equality, and thus inequality, between matrices and the estimation methods for Σ and A are not shown.

Algorithm 4 Iterative estimation of the SUR-VAR model.

Input: The regressors X, Y and the selection matrices $S_1, \ldots, S_G$

Output: The vectors of parameters $\beta_1, \ldots, \beta_G$, the autoregressive coefficient matrix A and the covariance matrix Σ

```
Compute the QRD (3.7).
Compute the QRD (3.15a) and (3.17a).
Let $\Sigma^{(0)} = 0$, $A^{(0)} = 0$ and $\text{Vec}\,(\{\beta_i^{(0)}\}) = 0$.
Estimate $\Sigma^{(1)}$ and $A^{(1)}$.
for $j = 1, 2, \ldots$ do
    if ($\Sigma^{(j)} \neq \Sigma^{(j-1)}$) then
        Compute the Cholesky factorization $\Sigma^{(j)} = CC^T$.
        Compute the RQD (3.15b).
    end if
    if ($A^{(j)} \neq A^{(j-1)}$) then
        Compute (3.13) with $A^{(j)}$ in place of $A$.
        Compute the updated QRD (3.16a) and (3.17b).
    end if
    Compute the RQD (3.16b).
    Compute (3.17c).
    Solve the upper triangular system (3.19) and let $\text{Vec}\,(\{\beta_i^{(j)}\}) = \text{Vec}\,(\{\beta_i\})$.
    Estimate $\Sigma^{(j+1)}$ and $A^{(j+1)}$.
until $\Sigma^{(j+1)} = \Sigma^{(j)}$, $A^{(j+1)} = A^{(j)}$ and $\text{Vec}\,(\{\beta_i^{(j)}\}) = \text{Vec}\,(\{\beta_i^{(j-1)}\})$
```

3.2.1 Estimation of the AR parameters and disturbance Covariance matrix

Commonly used formulae for estimating the covariance matrix and the matrix of autoregressive parameters for the SUR-VAR model (3.2) are given, respectively, by

$$\hat{\Sigma} = \frac{\hat{U}^T\hat{U}}{M} \tag{3.20}$$

and

$$\hat{A} = (\hat{U}^T\hat{U})^{-1}(\hat{U}^T Z\hat{U}), \tag{3.21}$$

where $\hat{U}$ is the residuals of the GLS estimator. That is

$$\hat{U} = Y - X\hat{B}, \tag{3.22}$$

where $\hat{B} = \begin{pmatrix} S_1\hat{\beta}_1 & \cdots & S_G\hat{\beta}_G \end{pmatrix} \in \mathbb{R}^{K^d \times G}$ and $\{\hat{\beta}_i\}$ denotes the GLS estimator of (3.2). The latter is equivalent to

$$Z\hat{U} = ZY - ZX \begin{pmatrix} S_1\hat{\beta}_1 & \cdots & S_G\hat{\beta}_G \end{pmatrix}. \tag{3.23}$$

Premultiplying (3.22) and (3.23) by the orthogonal matrix $(Q^*)^T$ from (3.8) gives

$$\begin{pmatrix} \hat{U}_A \\ \hat{U}_B \\ \hat{U}_C \end{pmatrix} = \begin{pmatrix} Y_A \\ Y_B \\ 0 \end{pmatrix} + \begin{pmatrix} R_A \\ R_B \\ 0 \end{pmatrix} \hat{B} \tag{3.24a}$$

and

$$\begin{pmatrix} \hat{U}_A^* \\ \hat{U}_B^* \\ \hat{U}_C^* \end{pmatrix} = \begin{pmatrix} Y_A^* \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} R_A^* \\ 0 \\ 0 \end{pmatrix} \hat{B}. \tag{3.24b}$$

Now, using (3.24), the estimators for Σ and A in (3.20) and (3.21) can be equivalently expressed, respectively, as:

$$\hat{\Sigma} = \frac{1}{M} \left(\hat{U}_A^T \hat{U}_A + \hat{U}_B^T \hat{U}_B \right) \tag{3.25}$$

and

$$\hat{A} = \left(\hat{U}_A^T \hat{U}_A + \hat{U}_B^T \hat{U}_B \right)^{-1} \left(\hat{U}_A^T \hat{U}_A^* \right). \tag{3.26}$$

For the efficient computation of (3.25) and (3.26), consider the QL decomposition

$$\check{Q}^T \begin{pmatrix} \hat{U}_A \\ \hat{U}_B \end{pmatrix} = \begin{pmatrix} 0 \\ L \end{pmatrix} \begin{matrix} {\scriptstyle 2K^*-G} \\ {\scriptstyle G} \end{matrix}, \tag{3.27}$$

where $\check{Q} = \begin{pmatrix} \check{Q}_A & \check{Q}_B \end{pmatrix}$, $\check{Q}_A \in \mathbb{R}^{2K^* \times (2K^*-G)}$, $\check{Q}_B \in \mathbb{R}^{2K^* \times G}$ and $L \in \mathbb{R}^{G \times G}$ is lower triangular [52]. The upper triangular Cholesky factor of $\widehat{\Sigma} = \widehat{C}\widehat{C}^T$, which is required by Algorithm 4, is given by $\widehat{C} = L^T$. Furthermore, $\widehat{A}$ in (3.26) derives from the solution of the triangular system

$$L\hat{A} = \frac{1}{M} \check{Q}_A^T \begin{pmatrix} \hat{U}_A^* \\ 0 \end{pmatrix}. \tag{3.28}$$

3.2.2 Considerations regarding the first observation

Different treatments of the first observation provide different estimators for the SUR-VAR model. Particularly when the first observation of the sample is dropped, the disturbance matrix U is assumed to satisfy

$$\bar{Z}^T U = \bar{Z} U A^T + E, \tag{3.29}$$

where $\mathrm{Vec}(E) \sim (0, \Sigma \otimes I_{M-1})$ and the $(M-1) \times M$ shift matrix $\bar{Z}$ is defined as

$$\bar{Z} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Under this assumption the SUR-VAR model (3.5a) becomes

$$\mathrm{Vec}(\bar{Z}^T Y - \bar{Z} Y A^T) = (I_G \otimes \bar{Z}^T X - A \otimes \bar{Z} X)(\oplus_i S_i)\,\mathrm{Vec}\left(\{\beta_i\}\right) + \mathrm{Vec}(E)$$

and the proposed solution method is still valid if the QRD (3.7) is replaced by the QRD of

$$\begin{pmatrix} \bar{Z}X & \bar{Z}Y & \bar{Z}^T X & \bar{Z}^T Y \end{pmatrix}.$$

An alternative means for handling the first observation is to enforce the stationarity of the VAR process U [58]. The disturbance matrix U then satisfies

$$U = Z U A^T + e_1 e_1^T U (I_G - A_1)^T + E, \tag{3.30}$$

where e_1 denotes the first column of I_M and $A_1 \in \mathbb{R}^{G \times G}$ is computed as a function of A and Σ. In this case, the QRD (3.7) should be replaced by the QRD

$$\begin{aligned} &\begin{pmatrix} e_1 e_1^T X & e_1 e_1^T Y & ZX & ZY & X & Y \end{pmatrix} \\ &= \begin{array}{c} \begin{array}{cccc} {\scriptstyle 1} & {\scriptstyle K^*} & {\scriptstyle K^*} & {\scriptstyle M^*} \end{array} \\ \begin{pmatrix} Q_1^* & Q_2^* & Q_3^* & Q_4^* \end{pmatrix} \end{array} \begin{array}{cccc} & {\scriptstyle K^*} & {\scriptstyle K^*} & {\scriptstyle M^*} \\ & \left(\begin{matrix} R_{11} \\ 0 \\ 0 \\ 0 \end{matrix}\right. & \begin{matrix} R_{12} \\ R_{22} \\ 0 \\ 0 \end{matrix} & \left.\begin{matrix} R_{13} \\ R_{23} \\ R_{33} \\ 0 \end{matrix}\right) \end{array} \begin{array}{c} \\ {\scriptstyle 1} \\ {\scriptstyle K^*} \\ {\scriptstyle K^*} \\ {\scriptstyle M^*} \end{array}, \end{aligned}$$

where $M^* = M - 2K^* - 1$. Premultiplying the SUR-VAR model by the orthogonal matrix $\begin{pmatrix} I_G \otimes Q_1^* & I_G \otimes Q_2^* & I_G \otimes Q_3^* & I_G \otimes Q_4^* \end{pmatrix}^T$ results in a reduced-size model of $2K^* + 1$ observations which can be solved by employing a method similar to the one that solves (3.12).

3.3 Computing the orthogonal factorizations

Strategies for computing the GQRD (3.15) have been previously proposed [46, 50]. Here the computation of the updated GQRD (3.16) will be considered for the case of distinct regressors. Let $R^{(0)} = \oplus_i R_i \in \mathbb{R}^{K \times K}$, and $R^{(p)}$ denote the pth $K^* \times K$ block of F_A; that is,

$$R^{(p)} = \left(e_p^T \otimes R_{Al}\right) + \left(\begin{pmatrix} A_{p1} & A_{p2} & \cdots & A_{pG} \end{pmatrix} \otimes I_{K^*}\right)\left(\oplus_i R^*_{Ai}\right), \tag{3.31}$$

for $p = 1, \ldots, G$. Furthermore let $R_{i,j}^{(p,G+1)}$ ($i = 1, \ldots, G+1$, $j = 1, \ldots, G$ and $p = 0, \ldots, G$) denote the (i,j)th $k_i \times k_j$ block of $R^{(p)}$, with $k_{G+1} \equiv G$; that is,

$$R^{(p)} = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{ccc|c} k_1 & \cdots & k_G & \\ R_{1,1}^{(p,G+1)} & \cdots & R_{1,G}^{(p,G+1)} & k_1 \\ \vdots & & \vdots & \vdots \\ R_{G,1}^{(p,G+1)} & \cdots & R_{G,G}^{(p,G+1)} & k_G \\ R_{G+1,1}^{(p,G+1)} & \cdots & R_{G+1,G}^{(p,G+1)} & k_{G+1} \end{array}.$$

Similarly, let

$$\begin{pmatrix} W_{11} & 0 \\ 0 & C \otimes I_{K^*} \end{pmatrix} \equiv \begin{array}{cccc|c} K & K^* & \cdots & K^* & \\ W^{(0,0)} & W^{(0,1)} & \cdots & W^{(0,G)} & K \\ W^{(1,0)} & W^{(1,1)} & \cdots & W^{(1,G)} & K^* \\ \vdots & \vdots & & \vdots & \vdots \\ W^{(G,0)} & W^{(G,1)} & \cdots & W^{(G,G)} & K^* \end{array} \tag{3.32}$$

and $W_{p,s}^{(i,j)} \in \mathbb{R}^{k_p \times k_s}$ denote the (p,s)th sub-block of $W^{(i,j)}$, where $i,j = 0, 1, \ldots, G$, $p = 1, \ldots, G$ if $i = 0$, while $p = 1, \ldots, G+1$ if $i > 0$ and analogously for s and j. Thus, for $i,j = 1, \ldots, G$,

$$W^{(i,j)} \equiv \begin{array}{cccc|c} k_1 & k_2 & & k_{G+1} & \\ W_{1,1}^{(i,j)} & W_{1,2}^{(i,j)} & \cdots & W_{1,G+1}^{(i,j)} & k_1 \\ \vdots & \vdots & & \vdots & \vdots \\ W_{G+1,1}^{(i,j)} & W_{G+1,2}^{(i,j)} & \cdots & W_{G+1,G+1}^{(i,j)} & k_{G+1} \end{array}.$$

The updated GQRD (3.16) is computed in two stages. The first stage applies a series of orthogonal transformations to reduce $R^{(1)}, \ldots, R^{(G)}$ to block-upper triangular. The orthogonal transformations are also applied from the left of (3.32), which is then re-triangularized from the right. Specifically, the $R_{i,i}^{(0,G+1)}$ ($i = 1, \ldots, G$) is used as pivot in order to annihilate from bottom to top the non-zero blocks $R_{i+1,i}^{(i,G+1)}, \ldots, R_{G+1,i}^{(i,G+1)}$ in the lower triangular part of $R^{(i)}$. That is, it computes the updated QRDs

$$\widehat{Q}_{i,j}^T \begin{pmatrix} R_{i,i}^{(0,j)} \\ R_{j,i}^{(i,j)} \end{pmatrix} = \begin{pmatrix} R_{i,i}^{(0,j-1)} \\ 0 \end{pmatrix} \tag{3.33a}$$

and

$$\widehat{Q}_{i,j}^T \begin{pmatrix} 0 & R_{i,j+1}^{(0,j)} & \cdots & R_{i,G}^{(0,j)} \\ R_{j,j}^{(i,j)} & R_{j,j+1}^{(i,j)} & \cdots & R_{j,G}^{(i,j)} \end{pmatrix} = \begin{pmatrix} R_{i,j}^{(0,j-1)} & R_{i,j+1}^{(0,j-1)} & \cdots & R_{i,G}^{(0,j-1)} \\ R_{j,j}^{(i,j-1)} & R_{j,j+1}^{(i,j-1)} & \cdots & R_{j,G}^{(i,j-1)} \end{pmatrix}, \tag{3.33b}$$

where $\widehat{Q}_{i,j}^T$ is a $(k_i + k_j) \times (k_i + k_j)$ orthogonal matrix, $R_{i,i}^{(0,j)}$ is $k_i \times k_i$ upper-triangular, $i = 1, \ldots, G$ and $j = G+1, G, \ldots, i+1$. Notice that the top i block-rows of $R^{(i,G+1)} \equiv R^{(i)}$ remain unchanged during the factorization (3.33), that is, $R_{p,q}^{(i,i)} \equiv R_{p,q}^{(i,G+1)}$ for $p = 1, \ldots, i$ and $q = 1, \ldots, G$. The factorization (3.33) can be computed simultaneously for $i = 1, \ldots, G$ using Householder transformation or Givens rotations [28, 46]. Figure 3.2 illustrates the triangularization process of $R^{(i)}$ using (3.33), where $i = 3$ and $G = 5$. An arc indicates the block-rows affected by the orthogonal factorization.

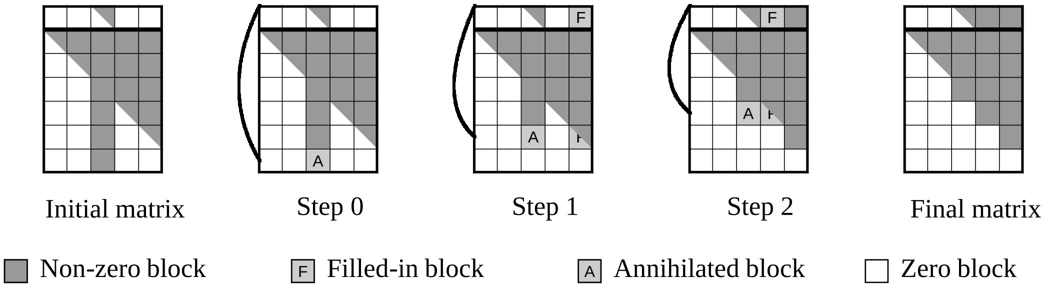


Figure 3.2: Computation of (3.33), where $i = 3$ and $G = 5$.

The modified structure of (3.32), say $\widehat{W}$, following the application of the orthogonal matrices $\widehat{Q}_{i,j}^T$ is shown in Figure 3.3. Notice that for $i, j = 1, \ldots, G$, $\widehat{W}^{(i,j)} = 0$ if $i > j$, otherwise

$$\widehat{W}^{(i,j)} = \begin{pmatrix} C_{i,j} I_{k^{(i)}} & 0 \\ 0 & \widehat{W}_*^{(i,j)} \end{pmatrix},$$

where $k^{(j)} = \sum_{p=1}^{j} k_p$ and $\widehat{W}_*^{(i,j)}$ is block-upper triangular of order $K^* - k^{(i)}$. Furthermore, $\widehat{W}^{(0,0)}$ is block-upper triangular, the first $k^{(i)}$ rows and $k^{(i-1)}$ columns of $\widehat{W}^{(i,0)}$ are zero, and $\widehat{W}_{p,q}^{(0,i)} = 0$ if $p > i$ or $p \geq q$, where $i, p = 1, \ldots, G$ and $q = 1, \ldots, G + 1$.

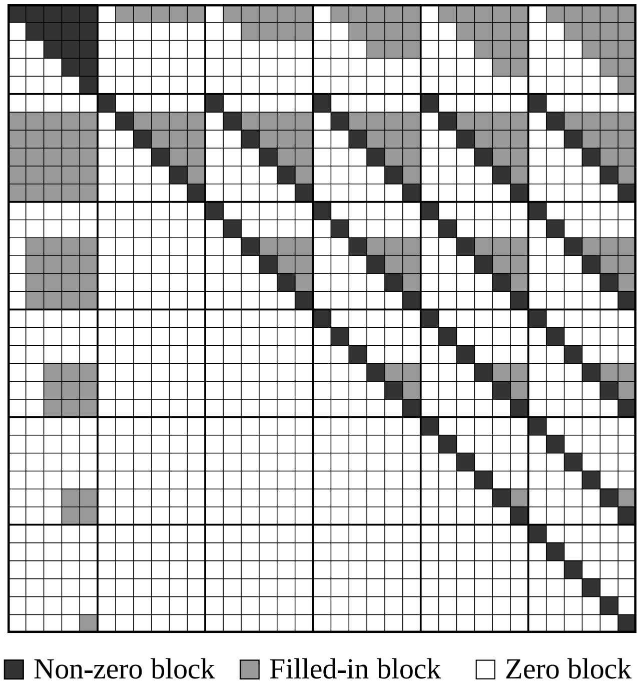


Figure 3.3: The structure of $\widehat{W}$ (fill-in induced in the Cholesky factor of the covariance matrix when applying the orthogonal matrices in (3.33)), where $G = 5$.

The matrix $\widehat{W}$ is retriangularized from the right by computing the RQ decomposition

$$\widehat{W}\widehat{P} = \widetilde{W}, \tag{3.34}$$

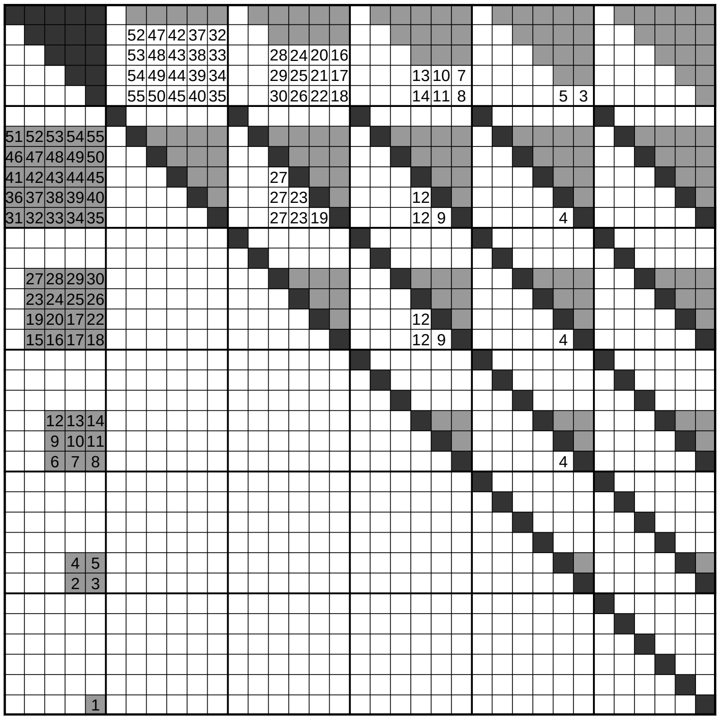
where $\widetilde{W}$ is upper triangular and $\widehat{P}$ is orthogonal. This is computed by using the diagonal blocks $\widehat{W}^{(i,i)}$ ($i = G, G - 1, \ldots, 1$) to annihilate the non-zero blocks in the first block-column. That is, the right transformation of the ith ($i = G, G - 1, \ldots, 1$) step computes the RQDs

$$\begin{pmatrix} \widehat{W}_{p,q}^{(i,0)} & \widehat{W}_{p,p}^{(i,i)} \end{pmatrix} \widehat{P}_{p,q}^{(i)} = \begin{pmatrix} 0 & \widetilde{W}_{p,p}^{(i,i)} \end{pmatrix}, \tag{3.35}$$

where $\widehat{P}_{p,q}^{(i)}$ is orthogonal, $\widetilde{W}_{p,p}^{(i,i)}$ is upper-triangular, $p = i + 1, \ldots, G + 1$ and $q = i, \ldots, G$. The orthogonal matrix $\widehat{P}_{p,q}^{(i)}$ is also applied from the right of the non-zero sub-blocks

$$\begin{pmatrix} \widehat{W}_{1,q}^{(0,0)} & \widehat{W}_{1,p}^{(0,i)} \\ \vdots & \vdots \\ \widehat{W}_{q,q}^{(0,0)} & \widehat{W}_{q,p}^{(0,i)} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \widehat{W}_{j,q}^{(j,0)} & \widehat{W}_{j,p}^{(j,i)} \\ \vdots & \vdots \\ \widehat{W}_{G+1,q}^{(j,0)} & \widehat{W}_{G+1,p}^{(j,i)} \end{pmatrix}, \quad \text{for} \quad j = 1, \ldots, i-1.$$

Notice that, this process does not produce fill-in in the lower-triangular part of $\widehat{W}$ and does not involve the first j block-columns of $\widehat{W}^{(i,j)}$. Figure 3.4 shows this process when $G = 5$. The sequence of numerals indicate the order of annihilating the sub-blocks in the first block-column and the corresponding fill-in in the first block-row of $\widehat{W}$.



■ Non-zero block ▨ Annihilated block at step i ☐ Filled-in block at step i ☐ Zero block

Figure 3.4: Re-triangularization of $\widehat{W}$, where $G = 5$.

The second stage of computing the updated GQRD (3.16) gives the GQRD

$$\widetilde{Q}^T \begin{pmatrix} \widehat{R}^{(0)} \\ \widehat{R}^{(1)} \\ \vdots \\ \widehat{R}^{(G)} \end{pmatrix} = \begin{pmatrix} R_u \\ 0 \end{pmatrix}, \tag{3.36a}$$

and

$$\widetilde{Q}^T \widetilde{W}\, \widetilde{P} = W_u, \tag{3.36b}$$

where R_u and W_u are the triangular factors defined in (3.16a) and (3.16b), respectively, and $\widehat{R}^{(i)}$ denotes the modified $R^{(i)}$ following the factorization (3.33). The QRD in (3.36a) results from computing the updated QRDs

$$\widetilde{Q}_i^T \begin{pmatrix} \widehat{R}^{(i-1)} \\ \widetilde{R}^{(i)} \end{pmatrix} = \begin{pmatrix} \widetilde{R}^{(i-1)} \\ 0 \end{pmatrix} \tag{3.37a}$$

and

$$\widetilde{Q}_i^T \begin{pmatrix} \widetilde{W}^{(i-1,i-1)} & \widetilde{W}^{(i-1,i)} & \cdots & \widetilde{W}^{(i-1,G)} \\ 0 & \widetilde{W}^{(i,i)}_{(i)} & \cdots & \widetilde{W}^{(i,G)}_{(i)} \end{pmatrix} = \\ \begin{pmatrix} \widetilde{W}^{(i-1,i-1)}_{(i-1)} & \widetilde{W}^{(i-1,i)}_{(i-1)} & \cdots & \widetilde{W}^{(i-1,G)}_{(i-1)} \\ \widetilde{W}^{(i,i-1)}_{(i-1)} & \widetilde{W}^{(i,i)}_{(i-1)} & \cdots & \widetilde{W}^{(i,G)}_{(i-1)} \end{pmatrix}, \tag{3.37b}$$

for $i = G, G-1, \ldots, 1$ and where $\widetilde{Q}_i$ is orthogonal, $\widetilde{R}^{(G)} \equiv \widehat{R}^{(G)}$, $\widetilde{R}^{(0)} \equiv R_u$ and $\widetilde{W}^{(G,G)}_{(G)} \equiv \widetilde{W}^{(G,G)}$. Let $\widetilde{W}_{(0)}$ denote the matrix resulting from the computations (3.37b), that is, the matrix comprising the sub-blocks $\widetilde{W}^{(i,j)}_{(i-1)}$, where $i, j = 0, 1, \ldots, G$. It follows that $\widetilde{W}_{(0)}$ is block upper-Hessenberg, that is $\widetilde{W}^{(i,j)}_{(0)} = 0$, for $i > j + 1$.

Within the context of rank-k updating and block downdating the QRD, strategies for retriangularizing block upper Hessenberg matrices have been proposed [46, 47]. Here the block subdiagonal of $\widetilde{W}_{(0)}$ is annihilated from the right by computing the RQDs

$$\begin{pmatrix} \widetilde{W}^{(i,i-1)}_{(0)} & \widehat{W}^{(i,i)}_{(i)} \end{pmatrix} \widetilde{P}^{(i)} = \begin{pmatrix} 0 & \widehat{W}^{(i,i)}_{(i-1)} \end{pmatrix} \tag{3.38a}$$

and

$$\begin{pmatrix} \widetilde{W}^{(0,i-1)}_{(0)} & \widehat{W}^{(0,i)}_{(i)} \\ \vdots & \vdots \\ \widetilde{W}^{(i-1,i-1)}_{(0)} & \widehat{W}^{(i-1,i)}_{(i)} \end{pmatrix} \widetilde{P}^{(i)} = \begin{pmatrix} \widehat{W}^{(0,i-1)}_{(i-1)} & \widehat{W}^{(0,i)}_{(i-1)} \\ \vdots & \vdots \\ \widehat{W}^{(i-1,i-1)}_{(i-1)} & \widehat{W}^{(i-1,i)}_{(i-1)} \end{pmatrix}, \tag{3.38b}$$

where $\widetilde{P}^{(i)}$ is orthogonal, $\widehat{W}^{(i,i)}_{(i-1)}$ is upper triangular and $i = G, G-1, \ldots, 1$.

3.4 Computational results

The proposed Algorithm 4 has been implemented using the LAPACK subroutines DGEQRF and DGERQF to perform the various matrix factorizations. The performance of the algorithm has been compared with that of computing the GLS estimator of the GLM (3.5b) using the LAPACK subroutine DGGGLM [1, 2]. Double precision has been used to implement the algorithm on a PC of a single Intel Pentium IV processor with clock of 1.7GHz and 512Mb of RAM.

Table 3.1: Execution times of solving the SUR-VAR model, where $k_1 = \cdots = k_G = 15$.

		Execution time		Ratio
M	G	DGGGLM	Algor. 4	DGGGLM / Algor. 4
250	5	5.8100	0.3519	16.51
500	5	48.0114	0.4022	119.39
750	5	152.5446	0.4745	321.51
1000	5	338.8241	0.5696	594.89
500	4	24.3317	0.2062	118.00
500	6	77.9108	0.8035	96.97
500	8	189.3951	3.2238	58.75
500	10	349.8498	10.6832	32.75

Table 3.2: Execution times of solving the SUR-VAR model, where $k_1 = \cdots = k_G = 20$.

		Execution time		Ratio
M	G	DGGGLM	Algor. 4	DGGGLM / Algor. 4
250	5	6.1658	0.7740	7.97
500	5	49.0770	0.8546	57.43
750	5	155.2297	0.8858	175.25
1000	5	342.0672	0.9659	354.14
500	4	24.9131	0.3692	67.47
500	6	80.6672	1.8915	42.65
500	8	193.0501	9.9281	19.44
500	10	357.6688	33.3050	10.74

Tables 3.1 and 3.2 show the execution times in seconds of the two estimation procedures. The first two columns indicate the dimension of the SUR-VAR model for fixed number of regressors per

equations. The execution times of the LAPACK DGGGLM and Algorithm 4 for a single iteration are shown in the third and fourth column, respectively. The ratio of the two execution times is shown in the last column. In Table 3.1 $k_1 = \cdots = k_G = 15$, while Table 3.2 considers 20 regressors per equation. Table 3.3 shows the time required to solve the problem afresh and that of recomputing the estimator when the AR parameters matrix A, and, or, the covariance matrix Σ have changed.

Table 3.3: Execution time of Algorithm 4 to re-estimate the SUR-VAR model, when A, and, or, Σ have changed, where $T = 1000$, $G = 5$ and $k_1 = \cdots = k_G = 20$.

	Execution time	% of time saved
Initial estimation	0.96590	
Re-estimation after A has changed	0.61503	36.33%
Re-estimation after Σ has changed	0.66587	31.06%
Re-estimation after A and Σ has changed	0.70812	26.69%

The computational results confirm the computational efficiency of Algorithm 4, which outperforms the existing LAPACK strategy. The performance of Algorithm 4 become better as the number of observations increases. This is mainly for two reasons. First, the LAPACK algorithm, which solves the general linear model, does not exploit the structure of the matrices of the SUR-VAR model. Second, the new estimation procedure initially transforms the model to one of smaller dimension. This allows the complexity of the algorithm to be linear with the number of observations. Furthermore, Algorithm 4 utilizes part of the computations performed in previous iterations. For the particular case shown in Table 3.3, this produces approximately 30% reduction in execution time compared to that required to solve the model afresh.

3.5 Conclusions

A computationally efficient method to estimate the SUR model with Vector Autoregressive disturbances (SUR-VAR model) has been proposed. The SUR-VAR model is transformed to a smaller model having dimension $2GK^*$ using the QRD (3.12). Using the GQRD as the main computational tool and exploiting the structure of the matrices results in an efficient procedure to estimate the reduced-size model. The computational savings are significant for large samples, that is for large values of M. Computational results confirm the efficiency of the proposed method when compared to solving the general linear model.

The estimator of the model is computed iteratively when the covariance matrix Σ or the AR parameters are unknown and need to be estimated. Some of the computations are unnecessary if only Σ or A changes. An iterative estimation procedure that efficiently utilizes computations from previous steps has been developed.

The proposed method can be extended when the SUR model has VAR(p) disturbances, that is, when the disturbance matrix satisfies $U - \sum_{i=1}^{p} Z^i U A_i^T = E$. Now, if

$$M > (p+1)K^*,$$

then the QRD

$$\begin{array}{c} \begin{array}{ccccccc} K^d & G & \cdots & K^d & G & K^d & G \end{array} \\ \begin{pmatrix} Z^pX & Z^pY & \cdots & ZX & ZY & X & Y \end{pmatrix} = \end{array}$$

$$\begin{array}{c} \begin{array}{ccccc} K^* & K^* & \cdots & K^* & M^* \end{array} \\ \begin{pmatrix} Q_1^* & Q_2^* & \cdots & Q_{p+1}^* & Q_{p+2}^* \end{pmatrix} \end{array} \begin{array}{cc} \begin{array}{cccc} K^* & K^* & \cdots & K^* \end{array} & \\ \begin{pmatrix} R^*_{1,1} & R^*_{1,2} & \cdots & R^*_{1,p+1} \\ 0 & R^*_{2,2} & \cdots & R^*_{2,p+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R^*_{p+1,p+1} \\ 0 & 0 & \cdots & 0 \end{pmatrix} & \begin{array}{c} K^* \\ K^* \\ \vdots \\ K^* \\ M^* \end{array} \end{array},$$

is computed. Here $Q^* = \begin{pmatrix} Q_1^* & Q_2^* & \cdots & Q_{p+2}^* \end{pmatrix}$ is orthogonal, $R^*_{i,i}$ ($i = 1, \ldots, p+1$) is upper triangular and $M^* = M - (p+1)K^*$. Premultiplying the SUR-VAR(p) model by the orthogonal matrix

$$\begin{pmatrix} I_G \otimes Q_1^* & I_G \otimes Q_2^* & \cdots & I_G \otimes Q_{p+2}^* \end{pmatrix}^T$$

results in a reduced-size SUR model with pK^* stochastic constraints compared to the K^* constraints of (3.12b).

In the case of autoregressive disturbances the proposed estimation method is simplified. The coefficient matrix A is diagonal, and consequently F_A in (3.13) is block-diagonal, given by

$$F_A = \oplus_i (R_{Ai} - A_{ii} R^*_{Ai}).$$

The estimation of SUR model with stochastic constraints (3.12) can be computed by estimating the SUR model (3.12a) and then adding the observations in (3.12b). Methods to re-estimate the standard SUR model after it has been updated with new observations can be adapted to the computation of the updated GQRD (3.16). Currently, strategies for updating the GQRD are considered within the context of recursive estimation of SUR models [20, 23, 49].

Chapter 4

Seemingly unrelated regression model with unequal size observations

Abstract:

The computational solution of the seemingly unrelated regression (SUR) model with unequal size observations is considered. Two algorithms to solve the model when treated as a generalized linear least squares problem are proposed. The algorithms have as a basic tool the generalized QR decomposition (GQRD) and efficiently exploit the block–sparse structure of the matrices. One of the algorithms reduces the computational burden of the estimation procedure by not computing explicitly the RQ factorization of the GQRD. The maximum likelihood estimation of the model when the covariance matrix is unknown is also considered.

4.1 Seemingly unrelated regression with unequal size observations

The seemingly unrelated regression (SUR) model is defined by the set of regressions

$$y_i = X_i\beta_i + u_i, \qquad i = 1, \ldots, G,$$

---

[1] This chapter is a reprint of the paper: P. Foschi, E.J. Kontoghiorghes. Seemingly unrelated regression model with unequal size observations: computational aspects. *Computational Statistics and Data Analysis,* 41(1):211-229, 2002.

where $X_i \in \mathbb{R}^{t \times k_i}$, $y_i \in \mathbb{R}^t$ and the disturbance vector $u_i \in \mathbb{R}^t$ has zero mean and variance–covariance matrix $\sigma_{i,i} I_t$. Furthermore the disturbances are contemporaneously correlated across the equations, i.e. $E(u_i u_j^T) = \sigma_{i,j} I_t$. In the compact form the SUR model can be written as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_G \end{pmatrix} = \begin{pmatrix} X_1 & & \\ & \ddots & \\ & & X_G \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_G \end{pmatrix}$$

or

$$\mathrm{Vec}(Y) = \left(\oplus_{i=1}^{G} X_i\right) \mathrm{Vec}(\{\beta_i\}_G) + \mathrm{Vec}(U),$$

where $Y = \begin{pmatrix} y_1 & \ldots & y_G \end{pmatrix}$, $U = \begin{pmatrix} u_1 & \ldots & u_G \end{pmatrix}$, the direct sum of matrices $\oplus_{i=1}^{G} X_i \equiv \oplus_i X_i \equiv \mathrm{diag}(X_1, \ldots, X_G)$, $\{\beta_i\}_G$ – abbreviated to $\{\beta_i\}$ – denotes the set of vectors $\beta_1, \ldots, \beta_G$ and $\mathrm{Vec}(\cdot)$ is the column stack operator with $\mathrm{Vec}(\{\beta_i\}) = (\beta_1^T, \ldots, \beta_G^T)^T$. The disturbance term $\mathrm{Vec}(U)$ has zero mean and dispersion matrix $\Sigma \otimes I_t$, where, $\Sigma = [\sigma_{i,j}] \in \mathbb{R}^{G \times G}$ is symmetric and positive semidefinite [79].

Computationally efficient methods for solving SUR models have been proposed [20, 26, 46, 48]. These methods formulate the SUR model as a Generalized Linear Least Squares Problem (GLLSP) and use the Generalized QR decomposition (GQRD) to solve it [55, 61]. Often it is assumed that each regression equation has the same number of observations, but this might not always be the case [79]. The solution of SUR models with unequal size observations (abbreviated to SUR-USO) has been previously considered [72, 74, 77]. Emphasis was given in the statistical properties of the estimators. The SUR-USO model assumes that the observations for the ith ($i > 1$) regression match in time with those for the $(i-1)$th regression. Here computational strategies for solving SUR-USO models are provided.

Firstly, recent methods for solving SUR models are extended to the numerical solution of the SUR-USO model when this is considered as a GLLSP. A method based on the GQRD is proposed for solving the GLLSP by exploiting the block-sparse and recursive structures of the exogenous matrix and Cholesky factor, respectively. A recursive strategy to reduce the computational burden of this method is presented. Finally, Maximum Likelihood expressions that can be used in the iterative solution of the SUR-USO model are derived.

4.2 Numerical solution of the SUR-USO model

In the SUR-USO model each regression has different number of observations. That is, $y_i, u_i \in \mathbb{R}^{t_i}$, $X_i \in \mathbb{R}^{t_i \times k_i}$ and the covariance matrices, for $j > i$, are given by

$$E(u_i u_j^T) = \sigma_{ij} \begin{pmatrix} I_{t_i} & 0_{t_i \times (t_j - t_i)} \end{pmatrix}, \tag{4.1}$$

where it has been assumed that $t_i \leq t_{i+1}$. The compact form of the SUR-USO model is given by

$$\text{Vec}(\{y_i\}) = (\oplus_i X_i) \, \text{Vec}(\{\beta_i\}) + \text{Vec}(\{u_i\}). \tag{4.2}$$

The dispersion of $\text{Vec}(\{u_i\})$ has a block matrix structure, where the (i,j)th block is given by (4.1).

Consider partitioning and reordering the observations of each regression by

$$y_i = \begin{array}{c} \begin{pmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{i,i} \end{pmatrix} \end{array} \begin{array}{l} h_1 \\ h_2 \\ \vdots \\ h_i \end{array}, \quad X_i = \begin{pmatrix} X_{1,i} \\ X_{2,i} \\ \vdots \\ X_{i,i} \end{pmatrix} \begin{array}{l} h_1 \\ h_2 \\ \vdots \\ h_i \end{array} \quad \text{and} \quad u_i = \begin{pmatrix} u_{1,i} \\ u_{2,i} \\ \vdots \\ u_{i,i} \end{pmatrix} \begin{array}{l} h_1 \\ h_2 \\ \vdots \\ h_i \end{array}, \tag{4.3}$$

where $h_1 = t_1$ and $h_i = t_i - t_{i-1}$ for $i = 2, 3, \ldots, G$. The SUR-USO model can be formulated as the set of regression equations

$$y_{i,j} = X_{i,j} \beta_j + u_{i,j}, \qquad \text{for } i, j = 1, 2, \ldots, G \; \text{ and } \; i \leq j, \tag{4.4}$$

where $u_{i,j}$ has zero mean and dispersion matrix given by $\sigma_{j,j} I_{h_i}$. Furthermore, the cross equation covariances are given by

$$E\big(u_{k,i} u_{l,j}^T\big) = \begin{cases} \sigma_{i,j} I_{h_k}, & \text{for } l = k, \\ 0_{h_k \times h_l}, & \text{for } l \neq k, \end{cases}$$

where $k \leq i$ and $l \leq j$. The regressions (4.4) are also equivalent to the general line model (GLM)

$$\begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_G \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_G \end{pmatrix} \text{Vec}\left(\{\beta_i\}\right) + \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \vdots \\ \bar{u}_G \end{pmatrix}, \tag{4.5}$$

which, after appropriate substitutions, can be written as

$$\bar{y} = \bar{X} \beta + \bar{u},$$

where $\bar{y}_i^T = \begin{pmatrix} y_{i,i}^T & y_{i,i+1}^T & \cdots & y_{i,G}^T \end{pmatrix} \in \mathbb{R}^{\mu_i}$, $\bar{u}_i^T = \begin{pmatrix} u_{i,i}^T & u_{i,i+1}^T & \cdots & u_{i,G}^T \end{pmatrix} \in \mathbb{R}^{\mu_i}$, $\beta = \operatorname{Vec}(\{\beta_i\})$,

$$\bar{X}_i = \begin{array}{c} \begin{array}{cccc} k_1 & \cdots & k_{i-1} & k_i+\cdots+k_G \end{array} \\ \begin{pmatrix} 0 & \cdots & 0 & \bigoplus_{j=i}^{G} X_{i,j} \end{pmatrix} \mu_i \end{array} , \tag{4.6}$$

and $\mu_i = (G-i+1)h_i$. The disturbance vector $\bar{u}$ has zero mean and covariance matrix $\Sigma_{(i)} \otimes I_{h_i}$, where $\Sigma_{(i)} \equiv \Sigma_{i:,i:}$ denotes the $(G-i+1) \times (G-i+1)$ submatrix of Σ starting at position (i,i) [28]. Furthermore the vectors $\bar{u}_i$ and $\bar{u}_j$ are uncorrelated for $i \neq j$. Thus, the covariance matrix of $\bar{u}$ is given by $\bar{\Sigma} = \oplus_i \left(\Sigma_{(i)} \otimes I_{h_i}\right) \in \mathbb{R}^{T \times T}$, where $T = \sum_i t_i = \sum_i \mu_i$. Without loss of generality it is assumed that $\Sigma_{(i)}$ is non-singular and $t_1 \geq k_i$ for $i = 1, \ldots, G$.

As in the case of the SUR model, the Best Linear Unbiased Estimator (BLUE) of the SUR-USO model derives from the solution of the Generalized Linear Least Squares problem (GLLSP)

$$\underset{\bar{v},\beta}{\operatorname{argmin}} \|\bar{v}\| \quad \text{subject to} \quad \bar{y} = \bar{X}\beta + \bar{C}\bar{v}, \tag{4.7}$$

where $\bar{u} = \bar{C}\bar{v}$, $\bar{\Sigma} = \bar{C}\bar{C}^T$ and the upper triangular matrix $\bar{C}$ has full rank. Thus, the random vector $\bar{v}$ has zero mean and dispersion matrix given by I_T. Notice that the matrix $\bar{C}$ is block diagonal with the ith ($i = 1, \ldots, G$) block given by $\bar{C}_{i,i} = C_{i:,i:} \otimes I_{h_i}$, where $\Sigma = CC^T$ and C is upper triangular. Figure 4.1 shows the structure of the SUR-USO model (4.2), GLM (4.5) and GLLSP (4.7) for $G = 3$.

For the solution of (4.7) consider the GQRD:

$$\bar{Q}^T \bar{X} = \begin{array}{cc} K & \\ \begin{pmatrix} \bar{R} \\ 0 \end{pmatrix} & \begin{array}{c} K \\ T-K \end{array} \end{array} \tag{4.8a}$$

and

$$\bar{Q}^T \bar{C} \bar{P} = \begin{array}{cc} \begin{array}{cc} K & T-K \end{array} & \\ \begin{pmatrix} W_{1,1} & W_{1,2} \\ 0 & W_{2,2} \end{pmatrix} & \begin{array}{c} K \\ T-K \end{array} \end{array} , \tag{4.8b}$$

The SUR-USO model (4.2): $\mathrm{Vec}\left(\{y_i\}\right) = (\oplus_i X_i)\,\mathrm{Vec}\left(\{\beta_i\}\right) + \mathrm{Vec}\left(\{u_i\}\right)$.

$$\begin{pmatrix} y_{11} \\ \hline y_{12} \\ y_{22} \\ \hline y_{13} \\ y_{23} \\ y_{33} \end{pmatrix} = \begin{pmatrix} X_{11} & 0 & 0 \\ \hline 0 & X_{12} & 0 \\ 0 & X_{22} & 0 \\ \hline 0 & 0 & X_{13} \\ 0 & 0 & X_{23} \\ 0 & 0 & X_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} u_{11} \\ \hline u_{12} \\ u_{22} \\ \hline u_{13} \\ u_{23} \\ u_{33} \end{pmatrix}$$

The covariance matrix of $\mathrm{Vec}\left(\{u_i\}\right)$.

$$\left(\begin{array}{c|cc|ccc} \sigma_{1,1}I & \sigma_{1,2}I & 0 & \sigma_{1,3}I & 0 & 0 \\ \hline \sigma_{2,1}I & \sigma_{2,2}I & 0 & \sigma_{2,3}I & 0 & 0 \\ 0 & 0 & \sigma_{2,2}I & 0 & \sigma_{2,3}I & 0 \\ \hline \sigma_{3,1}I & \sigma_{3,2}I & 0 & \sigma_{3,3}I & 0 & 0 \\ 0 & 0 & \sigma_{3,2}I & 0 & \sigma_{3,3}I & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{3,3}I \end{array}\right)$$

The GLM (4.5): $\bar{y} = \bar{X}\beta + \bar{u}$.

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ \hline y_{22} \\ y_{23} \\ \hline y_{33} \end{pmatrix} = \begin{pmatrix} X_{11} & 0 & 0 \\ 0 & X_{12} & 0 \\ 0 & 0 & X_{13} \\ \hline 0 & X_{22} & 0 \\ 0 & 0 & X_{23} \\ \hline 0 & 0 & X_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} u_{11} \\ u_{12} \\ u_{13} \\ \hline u_{22} \\ u_{23} \\ \hline u_{33} \end{pmatrix}$$

The covariance matrix of $\bar{u}$.

$$\left(\begin{array}{ccc|cc|c} \sigma_{1,1}I & \sigma_{1,2}I & \sigma_{1,3}I & 0 & 0 & 0 \\ \sigma_{2,1}I & \sigma_{2,2}I & \sigma_{2,3}I & 0 & 0 & 0 \\ \sigma_{3,1}I & \sigma_{3,2}I & \sigma_{3,3}I & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \sigma_{2,2}I & \sigma_{2,3}I & 0 \\ 0 & 0 & 0 & \sigma_{3,2}I & \sigma_{3,3}I & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \sigma_{3,3}I \end{array}\right),$$

The constraints of the GLLSP (4.7): $\bar{y} = \bar{X}\beta + \bar{C}\bar{v}$.

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ \hline y_{22} \\ y_{23} \\ \hline y_{33} \end{pmatrix} = \begin{pmatrix} X_{11} & 0 & 0 \\ 0 & X_{12} & 0 \\ 0 & 0 & X_{13} \\ \hline 0 & X_{22} & 0 \\ 0 & 0 & X_{23} \\ \hline 0 & 0 & X_{33} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \left(\begin{array}{ccc|cc|c} C_{1,1}I & C_{1,2}I & C_{1,3}I & 0 & 0 & 0 \\ 0 & C_{2,2}I & C_{2,3}I & 0 & 0 & 0 \\ 0 & 0 & C_{3,3}I & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & C_{2,2}I & C_{2,3}I & 0 \\ 0 & 0 & 0 & 0 & C_{3,3}I & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & C_{3,3}I \end{array}\right) \begin{pmatrix} v_{11} \\ v_{12} \\ v_{13} \\ \hline v_{22} \\ v_{23} \\ \hline v_{33} \end{pmatrix}.$$

Figure 4.1: Examples of models (4.2), (4.5) and (4.7) for $G = 3$.

where $K = \sum_i k_i$, $\bar{R}$ and W_{22} are upper triangular, and $Q, P \in \mathbb{R}^{T \times T}$ are orthogonal. Using (4.8) the GLLSP (4.7) can be written as

$$\underset{\bar{v}_A, \bar{v}_B, \beta}{\operatorname{argmin}} \left\|\bar{v}_A\right\|^2 + \left\|\bar{v}_B\right\|^2 \quad \text{subject to}$$

$$\begin{pmatrix} \bar{y}_A \\ \bar{y}_B \end{pmatrix} = \begin{pmatrix} \bar{R} \\ 0 \end{pmatrix} \beta + \begin{pmatrix} W_{1,1} & W_{1,2} \\ 0 & W_{2,2} \end{pmatrix} \begin{pmatrix} \bar{v}_A \\ \bar{v}_B \end{pmatrix},$$

where

$$Q^T \bar{y} = \begin{pmatrix} \bar{y}_A \\ \bar{y}_B \end{pmatrix} \quad \text{and} \quad P^T \bar{v} = \begin{pmatrix} \bar{v}_A \\ \bar{v}_B \end{pmatrix}.$$

It follows that $\bar{v}_B = W_{22}^{-1}\bar{y}_B$ and $\bar{v}_A = 0$. Thus, the solution of the SUR-USO model comes from solving the triangular system

$$\begin{pmatrix} \bar{y}_A \\ \bar{y}_B \end{pmatrix} = \begin{pmatrix} \bar{R} & W_{1,2} \\ 0 & W_{2,2} \end{pmatrix} \begin{pmatrix} \beta \\ \bar{v}_B \end{pmatrix}. \tag{4.9}$$

The main operations for solving the SUR-USO is the computation of the GQRD (4.8) and in some extent the solution of the triangular system (4.9). Clearly the computational burden of solving the SUR-USO will be reduced if the GQRD (4.8) is computed efficiently. Furthermore, the efficient computation of (4.8) will have a greater impact in the overall computational complexity if the iterative feasible estimator of the SUR-USO is required [79]. In such a case, at each iteration an estimator in the place of the unknown Σ is used. Thus, the QRD (4.8a) is computed once, while (4.8b), and consequently (4.9), need to be solved at each iteration for different $\bar{C}$.

4.3 Efficient solution of the GLLSP

For the efficient solution of the GLLSP (4.7) using the GQRD (4.8) the block-sparse structure of the matrices needs to be exploited. Consider first the GQRD

$$\begin{pmatrix} \bar{R}^{(0)} \\ 0 \end{pmatrix} = Q_0^T \bar{X}_1 \tag{4.10a}$$

and

$$Q_0^T \bar{C}_{1,1} P_0 = \begin{array}{c} \begin{array}{cc} \scriptstyle K & \scriptstyle \mu_1 - K \end{array} \\ \begin{pmatrix} \bar{C}_{1,1}^{(0)} & \widehat{W}_{1,1} \\ 0 & \widetilde{W}_{1,1} \end{pmatrix} \end{array} \begin{array}{l} \\ \scriptstyle K \\ \scriptstyle \mu_1 - K \end{array}, \tag{4.10b}$$

where $\bar{C}_{1,1}^{(0)}$ and $\widetilde{W}_{1,1}$ are upper triangular and P_0 is orthogonal. Furthermore, $\bar{R}^{(0)} = \oplus_i R_i^{(0)}$ and

$$Q_0 = \begin{pmatrix} \oplus_i \widehat{Q}_{0,i} & \oplus_i \widetilde{Q}_{0,i} \end{pmatrix} \equiv \begin{pmatrix} \widehat{Q}_{0,1} & & & \widetilde{Q}_{0,1} & & \\ & \ddots & & & \ddots & \\ & & \widehat{Q}_{0,G} & & & \widetilde{Q}_{0,G} \end{pmatrix},$$

where

$$X_{1,i} = \begin{pmatrix} \widehat{Q}_{0,i} & \widetilde{Q}_{0,i} \end{pmatrix} \begin{pmatrix} R_i^{(0)} \\ 0 \end{pmatrix} = \widehat{Q}_{0,i} R_i^{(0)}$$

is the QRD of $X_{1,i}$ for $i = 1, 2, \ldots, G$. Using (4.10), the GLLSP (4.7) can be equivalently written as

$$\underset{\substack{\beta,\hat{v}_1,\tilde{v}_1,\\ \bar{v}_2,\ldots,\bar{v}_G}}{\operatorname{argmin}} \ \|\widehat{v}_1\|^2 + \|\widetilde{v}_1\|^2 + \sum_{j=2}^{G} \|\bar{v}_j\|^2 \quad \text{subject to}$$

$$\begin{pmatrix} \widehat{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_G \\ \widetilde{y}_1 \end{pmatrix} = \begin{pmatrix} \bar{R}^{(0)} \\ \bar{X}_2 \\ \vdots \\ \bar{X}_G \\ 0 \end{pmatrix} \beta + \begin{pmatrix} \bar{C}_{1,1}^{(0)} & 0 & \cdots & 0 & \widehat{W}_{1,1} \\ 0 & \bar{C}_{2,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \bar{C}_{G,G} & 0 \\ 0 & 0 & \cdots & 0 & \widetilde{W}_{1,1} \end{pmatrix} \begin{pmatrix} \widehat{v}_1 \\ \bar{v}_2 \\ \vdots \\ \bar{v}_G \\ \widetilde{v}_1 \end{pmatrix}, \tag{4.11}$$

where $\widehat{y}_1 = \widehat{Q}_0^T \bar{y}_1$, $\widetilde{y}_1 = \widetilde{Q}_0^T \bar{y}_1$ and $P_0^T \bar{v}_1$ is conformably partitioned as $P_0^T \bar{v}_1 = \begin{pmatrix} \widehat{v}_1^T & \widetilde{v}_1^T \end{pmatrix}^T$. Here $\widetilde{v}_1$ can be computed by $\widetilde{v}_1 = \widetilde{W}_{11}^{-1} \widetilde{y}_1$ and thus, (4.11) can be reduced to

$$\underset{\substack{\beta,\hat{v}_1,\\ \bar{v}_2,\ldots,\bar{v}_G}}{\operatorname{argmin}} \ \|\widehat{v}_1\|^2 + \sum_{j=2}^{G} \|\bar{v}_j\|^2 \quad \text{subject to}$$

$$\begin{pmatrix} \bar{y}_1^{(0)} \\ \bar{y}_2 \\ \vdots \\ \bar{y}_G \end{pmatrix} = \begin{pmatrix} \bar{R}^{(0)} \\ \bar{X}_2 \\ \vdots \\ \bar{X}_G \end{pmatrix} \beta + \begin{pmatrix} \bar{C}_{1,1}^{(0)} & 0 & \cdots & 0 \\ 0 & \bar{C}_{2,2}^{(0)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{C}_{G,G}^{(0)} \end{pmatrix} \begin{pmatrix} \widehat{v}_1 \\ \bar{v}_2 \\ \vdots \\ \bar{v}_G \end{pmatrix}, \tag{4.12}$$

where

$$\bar{y}_1^{(0)} = \widehat{y}_1 - \widehat{W}_{1,1} \widetilde{v}_1 = \widehat{y}_1 - \widehat{W}_{1,1} \widetilde{W}_{1,1}^{-1} \widetilde{y}_1$$

and $\bar{C}_{i,i}^{(0)} \equiv \bar{C}_{i,i}$ for $i = 2, \ldots, G$.

The blocks $\bar{X}_i$ ($i = 2, \ldots, G$) can be annihilated by using a block generalization of a Givens sequence. Starting from the bottom to the top $\bar{R}^{(0)}$ is used as a pivot in order to annihilate $\bar{X}_G, \ldots, \bar{X}_2$ one at a time. That is, for $i = G, G-1, \ldots, 2$, the QRD

$$Q_i^T \begin{pmatrix} \bar{R}^{(G-i)} \\ \bar{X}_i \end{pmatrix} = \begin{array}{c} K \\ \begin{pmatrix} \bar{R}^{(G-i+1)} \\ 0 \end{pmatrix} \end{array} \begin{array}{l} \\ K \\ \mu_i \end{array}, \tag{4.13a}$$

$$Q_i^T \begin{pmatrix} \bar{y}_1^{(G-i)} \\ \bar{y}_i \end{pmatrix} = \begin{pmatrix} \bar{y}_1^{(G-i+1)} \\ \bar{y}_i^{(G-i+1)} \end{pmatrix} \begin{array}{l} K \\ \mu_i \end{array} \tag{4.13b}$$

and

$$Q_i^T \begin{array}{c} \\ \\ \end{array}\begin{array}{ccccc} K & \mu_i & \mu_{i+1} & \cdots & \mu_G \\ \end{array}$$

$$Q_i^T\begin{pmatrix} \bar{C}_{1,1}^{(G-i)} & 0 & \bar{C}_{1,i+1}^{(G-i)} & \cdots & \bar{C}_{1,G}^{(G-i)} \\ 0 & \bar{C}_{i,i}^{(0)} & 0 & \cdots & 0 \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} K & \mu_i & \cdots & \mu_G \end{array} \\ \begin{pmatrix} \bar{C}_{1,1}^{(G-i+1)} & \bar{C}_{1,i}^{(G-i+1)} & \cdots & \bar{C}_{1,G}^{(G-i+1)} \\ \bar{C}_{i,1}^{(G-i+1)} & \bar{C}_{i,i}^{(G-i+1)} & \cdots & \bar{C}_{i,G}^{(G-i+1)} \end{pmatrix} \begin{array}{c} K \\ \mu_i \end{array} \end{array} \tag{4.13c}$$

are computed, where $Q_i \in \mathbb{R}^{(K+\mu_i)\times(K+\mu_i)}$ is orthogonal, $\bar{R}^{(i)} = \oplus_{j=1}^{G} R_j^{(i)}$ and $R_j^{(i)} \in \mathbb{R}^{k_j \times k_j}$ is upper triangular. Notice that at each step $\bar{C}_{1,i}^{(G-i+1)}$, $\bar{C}_{i,1}^{(G-i+1)}$ and $\bar{C}_{i,i+1}^{(G-i+1)}, \ldots, \bar{C}_{i,G}^{(G-i+1)}$ are filled-in. This results in filling the block superdiagonals and first block-column of $\oplus_i \bar{C}_{i,i}^{(0)}$.

Let W denote the modified $\oplus_i \bar{C}_{i,i}^{(0)}$, that is,

$$W \equiv \begin{array}{c} \begin{array}{ccccc} K & \mu_2 & \mu_3 & \cdots & \mu_G \end{array} \\ \begin{pmatrix} W_{1,1}^{(0)} & W_{1,2}^{(0)} & W_{1,3}^{(0)} & \cdots & W_{1,G}^{(0)} \\ W_{2,1}^{(0)} & W_{2,2}^{(0)} & W_{2,3}^{(0)} & \cdots & W_{2,G}^{(0)} \\ W_{3,1}^{(0)} & 0 & W_{3,3}^{(0)} & \cdots & W_{3,G}^{(0)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_{G,1}^{(0)} & 0 & 0 & \cdots & W_{G,G}^{(0)} \end{pmatrix} \end{array}$$

$$= \begin{array}{c} \begin{array}{ccccc} K & \mu_2 & \mu_3 & \cdots & \mu_G \end{array} \\ \begin{pmatrix} \bar{C}_{1,1}^{(G-1)} & \bar{C}_{1,2}^{(G-1)} & \bar{C}_{1,3}^{(G-1)} & \cdots & \bar{C}_{1,G}^{(G-1)} \\ \bar{C}_{2,1}^{(G-1)} & \bar{C}_{2,2}^{(G-1)} & \bar{C}_{2,3}^{(G-1)} & \cdots & \bar{C}_{2,G}^{(G-1)} \\ \bar{C}_{3,1}^{(G-2)} & 0 & \bar{C}_{3,3}^{(G-2)} & \cdots & \bar{C}_{3,G}^{(G-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \bar{C}_{G,1}^{(1)} & 0 & 0 & \cdots & \bar{C}_{G,G}^{(1)} \end{pmatrix} \begin{array}{c} K \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_G \end{array} \end{array}. \tag{4.14}$$

The RQD of W can derive by a sequence of $G-1$ orthogonal factorizations which annihilate from the bottom to the top the submatrices $W_{2,1}^{(0)}, \ldots, W_{G,1}^{(0)}$. The ith ($i = 1, \ldots, G-1$) factorization computes

$$\begin{pmatrix} W_{1,1}^{(i-1)} & W_{1,G-i+1}^{(0)} \\ \vdots & \vdots \\ W_{G-i,1}^{(i-1)} & W_{G-i,G-i+1}^{(0)} \\ W_{G-i+1,1}^{(i-1)} & W_{G-i+1,G-i+1}^{(0)} \end{pmatrix} P_i = \begin{array}{c c} \begin{array}{cc} K & \mu_{G-i+1} \end{array} & \\ \begin{pmatrix} W_{1,1}^{(i)} & W_{1,G-i+1}^{(i)} \\ \vdots & \vdots \\ W_{G-i,1}^{(i)} & W_{G-i,G-i+1}^{(i)} \\ 0 & W_{G-i+1,G-i+1}^{(i)} \end{pmatrix} & \begin{array}{l} K \\ \vdots \\ \mu_{G-i} \\ \mu_{G-i+1} \end{array} \end{array}, \tag{4.15}$$

where $P_i \in \mathbb{R}^{(K+\mu_{G-i+1})\times(K+\mu_{G-i+1})}$ is orthogonal and $W_{G-i+1,G-i+1}^{(i)}$ is upper triangular. Thus, the upper triangular factor in the RQD of W is given by

$$\begin{array}{c c} & \begin{array}{cc} K & l^* \end{array} \\ \begin{array}{c} K \\ l^* \end{array} & \begin{pmatrix} W_{1,1}^* & W_{1,2}^* \\ 0 & W_{2,2}^* \end{pmatrix} \end{array} = \begin{array}{c c} \begin{array}{cccc} K & \mu_2 & \cdots & \mu_G \end{array} & \\ \begin{pmatrix} W_{1,1}^{(G-1)} & W_{1,2}^{(G-1)} & \cdots & W_{1,G}^{(1)} \\ 0 & W_{2,2}^{(G-1)} & \cdots & W_{2,G}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{G,G}^{(1)} \end{pmatrix} & \begin{array}{l} K \\ \mu_2 \\ \vdots \\ \mu_G \end{array} \end{array}, \tag{4.16}$$

where $l^* = \sum_{i=2}^{G} \mu_i$. Figure 4.2 shows at each step of the procedure the annihilation and filled-in of $\bar{X}_2, \ldots, \bar{X}_G$ and $\oplus_i \bar{C}_{i,i}^{(0)}$, respectively, and the retriangularization of W. From (4.13a), (4.13b) and (4.16), it follows that the solution of GLLSP (4.12) is given by the solution of the triangular system

$$\begin{pmatrix} \bar{y}_1^{(G-1)} \\ \bar{y}^* \end{pmatrix} = \begin{pmatrix} \bar{R}^{(G-1)} & W_{1,2}^* \\ 0 & W_{2,2}^* \end{pmatrix} \begin{pmatrix} \beta \\ \bar{v}_B^* \end{pmatrix},$$

where

$$\bar{y}^* = \begin{pmatrix} \bar{y}_2^{(G-1)} \\ \vdots \\ \bar{y}_G^{(1)} \end{pmatrix}.$$

The matrices in (4.13a), (4.13c) and (4.15) have block-sparse structures which can facilitate the development of fast factorization algorithms. From (4.6) and the block-diagonal structure of $\bar{R}^{(G-i)}$ it follows that the QRD (4.13a) can be derived by computing the $(G-i+1)$ updating

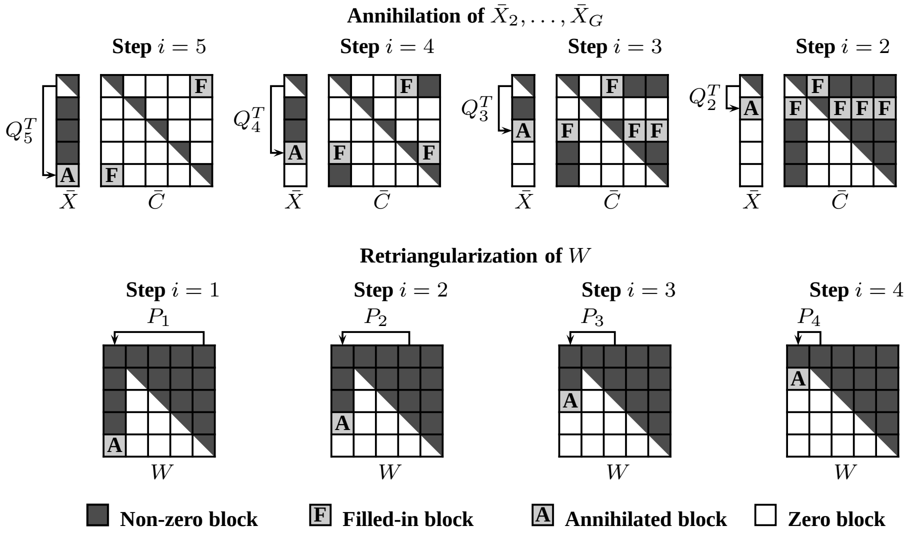


Figure 4.2: Annihilation of $\bar{X}_i$, fill-in of $\oplus_i \bar{C}_{i,i}^{(0)}$ and retriangularization of W, where $G = 5$.

QRDs (UQRDs)

$$Q_{i,j}^T \begin{pmatrix} R_j^{(G-i)} \\ X_{i,j} \end{pmatrix} = \begin{array}{cc} k_j & \\ \begin{pmatrix} R_j^{(G-i+1)} \\ 0 \end{pmatrix} & \begin{array}{c} k_j \\ h_i \end{array} \end{array} \qquad (j = 1, \ldots, G - i + 1). \tag{4.17}$$

Thus, in (4.13a) $R_s^{(G-i)} \equiv R_s^{(G-i+1)}$ for $s = 1, \ldots, i-1$ and

$$Q_i = \begin{pmatrix} I_{\lambda_i} & 0 & 0 \\ 0 & \oplus_{j=i}^G Q_{i,j}^{(1,1)} & \oplus_{j=i}^G Q_{i,j}^{(1,2)} \\ 0 & \oplus_{j=i}^G Q_{i,j}^{(2,1)} & \oplus_{j=i}^G Q_{i,j}^{(2,2)} \end{pmatrix}, \tag{4.18}$$

where $\lambda_i = \sum_{j=1}^{i-1} k_j$ and

$$Q_{i,j} = \begin{array}{cc} \begin{array}{cc} k_j & h_i \end{array} & \\ \begin{pmatrix} Q_{i,j}^{(1,1)} & Q_{i,j}^{(1,2)} \\ Q_{i,j}^{(2,1)} & Q_{i,j}^{(2,2)} \end{pmatrix} & \begin{array}{c} k_j \\ h_i \end{array} \end{array} .$$

Notice that when Q_i^T in (4.18) is used to compute (4.13c), then in (4.14)

$$W^{(0)}_{i,1} = \begin{array}{cc} \begin{array}{cc} \lambda_i & K-\lambda_i \end{array} & \\ \begin{pmatrix} 0 & \widetilde{W}^{(0)}_{i,1} \end{pmatrix} & \mu_i \end{array} , \tag{4.19a}$$

$$W^{(0)}_{1,j} = \begin{array}{cc} \mu_j & \\ \begin{pmatrix} 0 \\ \widetilde{W}^{(0)}_{1,j} \end{pmatrix} & \begin{array}{c} \lambda_j \\ K-\lambda_j \end{array} \end{array} , \tag{4.19b}$$

and

$$W^{(0)}_{i,j} = \begin{array}{cc} \mu_j & \\ \begin{pmatrix} 0 \\ \widetilde{W}^{(0)}_{i,j} \end{pmatrix} & \begin{array}{c} (j-i)h_i \\ (G-j+1)h_i \end{array} \end{array} \qquad \text{if } i < j, \tag{4.19c}$$

where $W^{(0)}_{i,i}$, $\widetilde{W}^{(0)}_{i,1}$, $\widetilde{W}^{(0)}_{1,i}$ and $\widetilde{W}^{(0)}_{i,j}$ are block upper triangular ($i, j = 2, \ldots, G$). Figure 4.3 shows the structure of W after the UQRDs (4.17) have been computed, where $G = 5$. A numeral i in $\bar{X}_{i,j}$ and W denotes, respectively, the annihilated and filled-in submatrices which resulted from the UQRDs at step i ($i = 1, \ldots, G$).

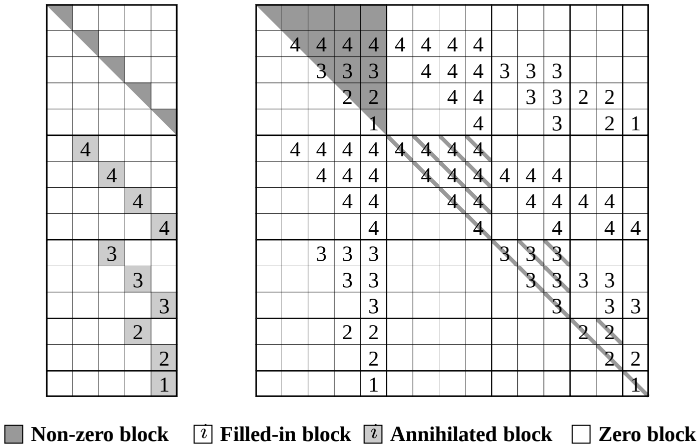


Figure 4.3: Annihilation of $\bar{X}_i$ and fill-in of $\oplus_i \bar{C}^{(0)}_{i,i}$, where $G = 5$ and $i = 1, \ldots, G$.

The RQD of W using (4.15) needs to take into account the sparse structure of the submatrices. Sequential and parallel strategies for computing similar factorizations have been proposed [44, 46]. The block-diagonals of $\widetilde{W}^{(i-1)}_{G-i+1,1}$ in (4.15) can be annihilated one at a time with a series of factorizations which preserve the sparse and triangular structure of $\widetilde{W}^{(i-1)}_{1,1}, \ldots, \widetilde{W}^{(i-1)}_{G-i,1}$. The orthogonal matrix P_i is defined as $P_i = \widetilde{P}_{i,1} \cdots \widetilde{P}_{i,i}$, where $\widetilde{P}_{i,j} = \widehat{P}^{(1)}_{i,j} \cdots \widehat{P}^{(i-j+1)}_{i,j}$ and $\widehat{P}^{(s)}_{i,j}$ annihilates the sth block of the $(G - i + j)$th block-diagonal of $\widetilde{W}^{(i-1)}_{G-i+1,1}$ ($i = 1, \ldots, G - 1$, $j = 1, \ldots, i$ and $s = 1, \ldots, i - j + 1$). Figure 4.4 illustrates this strategy for computing (4.15), where $i = 3$ and $G = 5$. Arcs connecting the blocks and block-columns indicate those affected by the orthogonal factorization $\widehat{P}^{(s)}_{i,j}$.

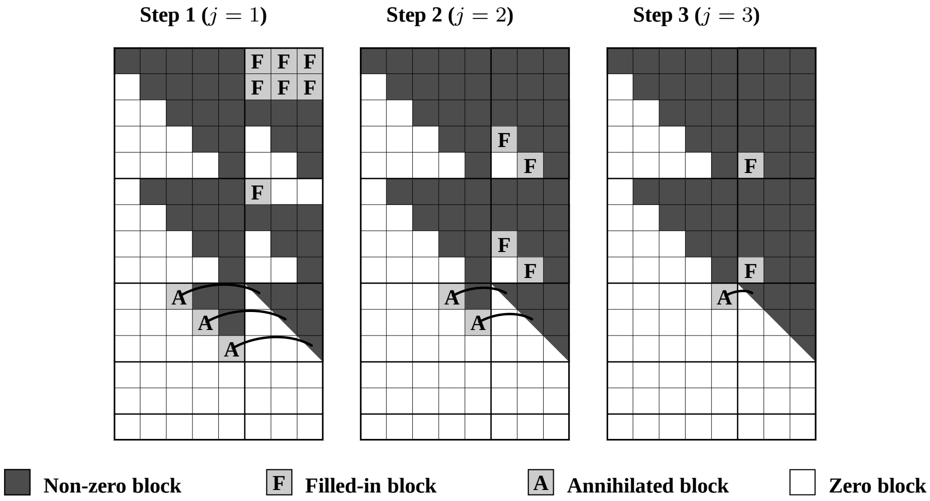


Figure 4.4: Computing (4.15), where $i = 3$ and $G = 5$.

4.4 A recursive strategy for solving the SUR-USO model

In the GQRD (4.8) the computations of the QRD (4.8a) and RQD (4.8b) can be interleaved. The orthogonal matrix Q_i^T in (4.13a) when applied from the left of $\begin{pmatrix} \bar{X} & \bar{C} \end{pmatrix}$ to annihilate $\bar{X}_i$ will fill-in a block in the lower part of $\bar{C}$. This fill-in is eliminated by the application of an orthogonal transformation from the right of the modified $\bar{C}$. That is, following (4.13a) and (4.13b),

$$Q_i^T\begin{array}{c}\begin{array}{ccccc} K & \mu_i & \mu_{i+1} & \cdots & \mu_G\end{array}\\ \begin{pmatrix} \bar{C}_{1,1}^{(G-i)} & 0 & \bar{C}_{1,i+1}^{(G-i)} & \cdots & \bar{C}_{1,G}^{(G-i)} \\ 0 & \bar{C}_{i,i}^{(0)} & 0 & \cdots & 0 \end{pmatrix}\end{array}$$

$$=\begin{array}{cc}\begin{array}{ccccc} K & \mu_i & \mu_{i+1} & \cdots & \mu_G\end{array} & \\ \begin{pmatrix} \widehat{C}_{1,1}^{(G-i)} & \widehat{C}_{1,i}^{(G-i)} & \bar{C}_{1,i+1}^{(G-i+1)} & \cdots & \bar{C}_{1,G}^{(G-i+1)} \\ \widehat{C}_{i,1}^{(G-i)} & \widehat{C}_{i,i}^{(G-i)} & \bar{C}_{i,i+1}^{(G-i+1)} & \cdots & \bar{C}_{i,G}^{(G-i+1)} \end{pmatrix} & \begin{array}{c} K \\ \mu_i \end{array}\end{array} \tag{4.20}$$

and the RQD

$$\begin{array}{cc} & \begin{array}{cc} K & \mu_i \end{array} \\ \begin{array}{c} K \\ \mu_i \end{array} & \begin{pmatrix} \widehat{C}_{1,1}^{(G-i)} & \widehat{C}_{1,i}^{(G-i)} \\ \widehat{C}_{i,1}^{(G-i)} & \widehat{C}_{i,i}^{(G-i)} \end{pmatrix}\end{array} P_i = \begin{array}{cc}\begin{array}{cc} K & \mu_i \end{array} & \\ \begin{pmatrix} \bar{C}_{1,1}^{(G-i+1)} & \bar{C}_{1,i}^{(G-i+1)} \\ 0 & \bar{C}_{i,i}^{(G-i+1)} \end{pmatrix} & \begin{array}{c} K \\ \mu_i \end{array}\end{array} \tag{4.21}$$

are computed, where $\bar{C}_{1,1}^{(G-i+1)}$ and $\bar{C}_{i,i}^{(G-i+1)}$ are upper triangular, $P_i \in \mathbb{R}^{(K+\mu_i)\times(K+\mu_i)}$ is orthogonal, $\widehat{C}_{i,1}^{(G-i)}$ and $\widehat{C}_{1,i}^{(G-i)}$ have, respectively, the same structure as $W_{i,1}^{(0)}$ and $W_{1,i}^{(0)}$ in (4.19) and $i = G, G-1, \ldots, 2$. The orthogonal matrices $\bar{Q}^T$ and $\bar{P}$ in (4.8) are defined as the products of the left and right transformations, respectively. Furthermore, notice that (4.21) involves only 4 blocks of $\bar{C}$, instead of $2i$ blocks of its corresponding (4.15). This results in an algorithm which has less computational complexity and lower memory usage. The annihilations and fill-ins occuring at each step of this procedure are shown in Fig. 4.5, where $G = 5$.

Notice that after the $(i+1)$th ($i = G-1, G-2, \ldots, 1$) step of the above strategy the GLLSP (4.12) can be written as:

$$\underset{\substack{\bar{v}_1^{(G-i)}, \bar{v}_*^{(G-i)}, \\ \bar{v}_2, \ldots, \bar{v}_i, \beta}}{\operatorname{argmin}} \quad \left\|\bar{v}_1^{(G-i)}\right\|^2 + \left\|\bar{v}_*^{(G-i)}\right\|^2 + \sum_{j=2}^{i} \left\|\bar{v}_j\right\|^2 \quad \text{subject to}$$

$$\begin{pmatrix} \bar{y}_1^{(G-i)} \\ \bar{y}_2 \\ \vdots \\ \bar{y}_i \\ \bar{y}_*^{(G-i)} \end{pmatrix} = \begin{pmatrix} \bar{R}^{(G-i)} \\ \bar{X}_2 \\ \vdots \\ \bar{X}_i \\ 0 \end{pmatrix} \beta + \begin{pmatrix} \bar{C}_{1,1}^{(G-i)} & 0 & \cdots & 0 & \bar{C}_{1,i:G}^{(G-i)} \\ 0 & \bar{C}_{2,2}^{(0)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \bar{C}_{i,i}^{(0)} & 0 \\ 0 & 0 & \cdots & 0 & \bar{C}_*^{(G-i)} \end{pmatrix} \begin{pmatrix} \hat{v}_1^{(G-i)} \\ \bar{v}_2 \\ \vdots \\ \bar{v}_i \\ \bar{v}_*^{(G-i)} \end{pmatrix}, \tag{4.22}$$

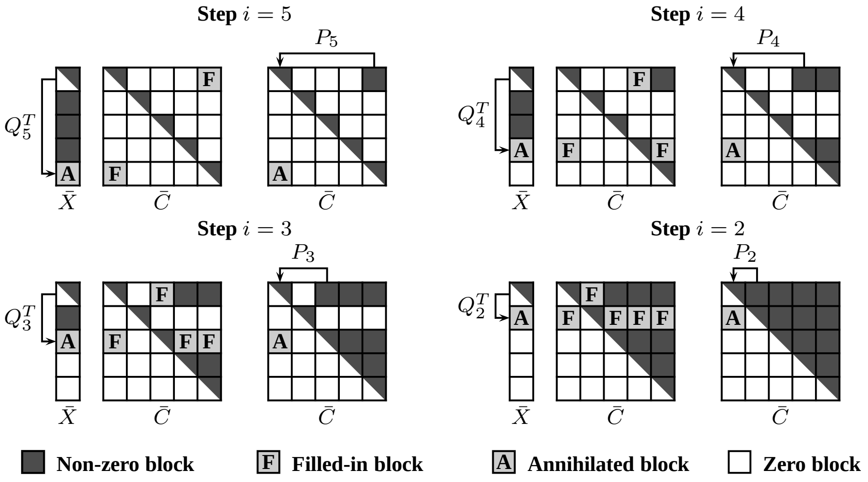


Figure 4.5: Annihilation of $\bar X_2, \ldots, \bar X_G$ and retriangularization of $\oplus_i \bar C_{i,i}^{(0)}$, where $G = 5$.

where $\bar C_*^{(G-i)}$ is $(\sum_{j=i+1}^{G} \mu_j) \times (\sum_{j=i+1}^{G} \mu_j)$ upper triangular and non-singular, $\bar y_*^{(G-i)} = \operatorname{Vec}(\{\bar y_{i+1}^{(G-i)}, \bar y_{i+2}^{(G-i-1)}, \ldots, \bar y_G^{(1)}\})$ and $\bar v_*^{(G-i)} = \operatorname{Vec}(\{\bar v_{i+1}, \ldots, \bar v_G\})$. Thus, (4.22) is equivalent to

$$\underset{\substack{\beta,\hat v_1^{(G-i)},\\ \bar v_2,\ldots,\bar v_i}}{\operatorname{argmin}} \left\|\hat v_1^{(G-i)}\right\|^2 + \sum_{j=2}^{i} \left\|\bar v_j\right\|^2 \quad \text{subject to}$$

$$\begin{pmatrix}\hat y_1^{(G-i)}\\ \bar y_2\\ \vdots\\ \bar y_i\end{pmatrix} = \begin{pmatrix}\bar R^{(G-i)}\\ \bar X_2\\ \vdots\\ \bar X_i\end{pmatrix}\beta + \begin{pmatrix}\bar C_{1,1}^{(G-i)} & 0 & \cdots & 0\\ 0 & \bar C_{2,2}^{(0)} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \bar C_{i,i}^{(0)}\end{pmatrix}\begin{pmatrix}\hat v_1^{(G-i)}\\ \bar v_2\\ \vdots\\ \bar v_i\end{pmatrix}, \tag{4.23}$$

where

$$\hat y_1^{(G-i)} = \bar y_1^{(G-i)} - \bar C_{1,i+1:G}^{(G-i)} \bar v_*^{(G-i)} \quad \text{and} \quad \bar v_*^{(G-i)} = (\bar C_*^{(G-i)})^{-1} \bar y_*^{(G-i)}.$$

The latter suggest a recursive strategy which solves a sequence of smaller in size GLLSP and requires less computational effort of computing the RQD. At the ith ($i = G, G-1, \ldots, 2$) step the recursive algorithm solves the GLLSP (4.23) by computing the QRD in (4.13a),

$$Q_i^T \begin{pmatrix} \widehat{y}_1^{(G-i)} \\ \bar{y}_i \end{pmatrix} = \begin{pmatrix} \widetilde{y}_1^{(G-i)} \\ \widehat{y}_i \end{pmatrix}, \tag{4.24}$$

$$Q_i^T \begin{pmatrix} \bar{C}_{1,1}^{(G-i)} & 0 \\ 0 & \bar{C}_{i,i}^{(0)} \end{pmatrix} = \begin{pmatrix} \widehat{C}_{1,1}^{(G-i)} & \widehat{C}_{1,i}^{(G-i)} \\ \widehat{C}_{i,1}^{(G-i)} & \widehat{C}_{i,i}^{(G-i)} \end{pmatrix} \tag{4.25}$$

and the RQD in (4.21). As in the case of (4.12), the GLLSP (4.23) is reduced to

$$\underset{\substack{\beta,\hat{v}_1^{(G-i+1)}, \\ \bar{v}_2,\ldots,\bar{v}_{i-1}}}{\operatorname{argmin}} \quad \|\hat{v}_1^{(G-i)}\|^2 + \sum_{j=2}^{i-1} \|\bar{v}_j\|^2 \quad \text{subject to}$$

$$\begin{pmatrix} \widehat{y}_1^{(G-i+1)} \\ \bar{y}_2 \\ \vdots \\ \bar{y}_{i-1} \end{pmatrix} = \begin{pmatrix} \bar{R}^{(G-i+1)} \\ \bar{X}_2 \\ \vdots \\ \bar{X}_{i-1} \end{pmatrix} \beta + \begin{pmatrix} \bar{C}_{1,1}^{(G-i+1)} & 0 & \cdots & 0 \\ 0 & \bar{C}_{2,2}^{(0)} \cdots & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{C}_{i-1,i-1}^{(0)} \end{pmatrix} \begin{pmatrix} \widehat{v}_1^{(G-i+1)} \\ \bar{v}_2 \\ \vdots \\ \bar{v}_{i-1} \end{pmatrix}, \tag{4.26}$$

where

$$\widehat{y}_1^{(G-i+1)} = \widetilde{y}_1^{(G-i)} - \bar{C}_{1,i}^{(G-i+1)} \left(\bar{C}_{i,i}^{(G-i+1)} \right)^{-1} \widehat{y}_i. \tag{4.27}$$

The structure of (4.26) is the same as that of (4.23), but smaller in size. The GLLSP (4.12) is equivalent to (4.23) when $i = G$, and $\widehat{y}_1^{(G-i)} \equiv \bar{y}_1^{(0)}$. Thus, this process can be applied iteratively to solve (4.12) and derive the BLUE of the SUR-USO model. Algorithm 5 summarizes the steps of this recursive procedure and Fig. 4.6 illustrates the factorization steps for $G = 5$.

Algorithm 5 Iterative estimation of the SUR-USO model.

```
Input: The regressors $X_1, \ldots, X_G$ and $Y$
Output: The vector $\beta$ of estimated parameters
Compute the GQRD (4.10), $\hat{y}_1 = \widehat{Q}_0^T \bar{y}_1$ and $\tilde{y}_1 = \widetilde{Q}_0^T \bar{y}_1$.
Solve the upper triangular system $\widetilde{W}_{11} \tilde{v}_1 = \tilde{y}_1$.
Compute $\bar{y}_1^{(0)} = \hat{y}_1 - \widehat{W}_{11} \tilde{v}_1$.
for $i = G, G-1, \ldots, 2$ do
    Compute the UQRD (4.13a).
    Compute (4.24) and (4.25).
    Compute the RQD (4.21).
    Compute (4.27).
end for
Solve the upper triangular system $\bar{R}^{(G-1)} \beta = \hat{y}_1^{(G-1)}$.
```

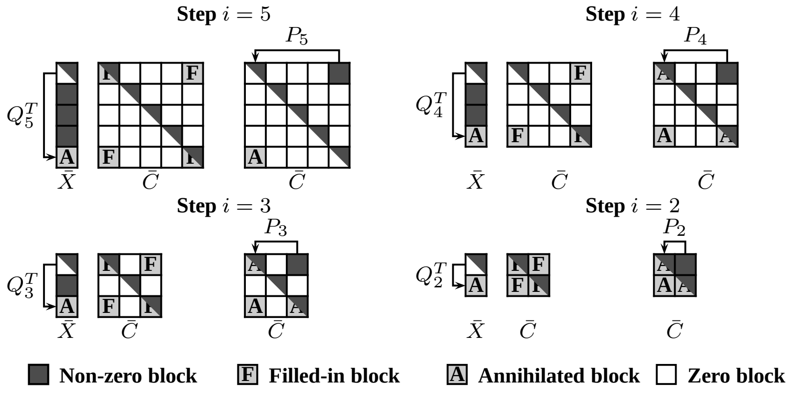


Figure 4.6: Annihilation of $\bar{X}_2, \ldots, \bar{X}_G$ and retriangularization of $\oplus_i \bar{C}^{(0)}_{i,i}$, where $G = 5$.

4.5 Maximum Likelihood Estimation

Under normality assumptions, the Maximum Likelihood (ML) estimators for β_i and Σ derive from the solution of the nonlinear equations

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \tag{4.28a}$$

and

$$\frac{\partial \mathcal{L}}{\partial \Sigma} = 0, \tag{4.28b}$$

where $\mathcal{L}$ is the log-likelihood function for the SUR-USO model (4.2). The nonlinear equations (4.28) are solved by using the EM algorithm. An initial estimator for Σ is choosen in order to obtain an estimator for β_i from (4.28a), which in term is used to provide a new estimator for Σ. This process is repeated until convergence [13].

The solution of (4.28a) is equivalent to the GLS estimator (4.2) and can be computed using the previously derived methods. Thus, only the numerical solution of (4.28b) will be considered. Notice that, when the disturbances are not normally distributed this approach can be considered as quasi-maximum likelihood estimation procedure.

The SUR-USO model (4.2) is equivalent to the set of equations,

$$\left.\begin{aligned} Y_1 &= \begin{pmatrix} X_{1,1}\beta_1 & X_{1,2}\beta_2 & \cdots & X_{1,G}\beta_G \end{pmatrix} + U_1, \\ Y_2 &= \begin{pmatrix} X_{2,2}\beta_2 & \cdots & X_{2,G}\beta_G \end{pmatrix} + U_2, \\ &\vdots \\ Y_G &= X_{G,G}\beta_G + U_2, \end{aligned}\right\} \tag{4.29}$$

where $Y_i = \begin{pmatrix} y_{i,i} & \dots & y_{i,G} \end{pmatrix} \in \mathbb{R}^{h_i \times (G-i+1)}$ and $U_i = \begin{pmatrix} u_{i,i} & \dots & u_{i,G} \end{pmatrix} \in \mathbb{R}^{h_i \times (G-i+1)}$ has a multivariate distribution with zero mean and covariance matrix given by $\Sigma_{(i)}$. That is, $\mathrm{Vec}(U_i)$ has zero mean and covariance matrix given by $\Sigma_{(i)} \otimes I_{G-i+1}$. Furthermore, the elements of U_i and U_j are uncorrelated for $i \neq j$.

The log-likelihood function of the ith equation in (4.29) and of the whole set are given by

$$\mathcal{L}_i = -\frac{1}{2}\left(\mu_i + h_i \log\left(\det(\Sigma_{(i)})\right) + \mathrm{tr}(U_i^T U_i \Sigma_{(i)}^{-1})\right)$$

and $\mathcal{L} = \sum_{i=1}^{G} \mathcal{L}_i$, respectively. Now, from $\Sigma_{(i)} = C_{(i)} C_{(i)}^T$ and $C_{(i)} = C_{i:,i:}$, it follows that

$$\frac{\partial \mathcal{L}_i}{\partial C_{(i)}^{-1}} = h_i C_{(i)}^T - C_{(i)}^{-1} U_i^T U_i. \tag{4.30}$$

Furthermore, since $C_{(i)}^{-1}$ is a submatrix of C^{-1}, the derivative for the log-likelihood function of SUR-USO model (4.29) with respect to C^{-1} is given by

$$\frac{\partial \mathcal{L}}{\partial C^{-1}} = \sum_{i=1}^{G} \begin{pmatrix} 0_{(i-1)\times(i-1)} & 0_{(i-1)\times(G-i+1)} \\ 0_{(G-i+1)\times(i-1)} & \partial \mathcal{L}_i / \partial C_{(i)}^{-1} \end{pmatrix}. \tag{4.31}$$

Substituting (4.30) in (4.31) and considering only the nonzero elements of C^{-1} – the elements in its upper triangular – gives

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial\,\mathrm{Vech}(C^{-T})} &= \mathrm{Vech}\left(D_T D_C - \sum_{i=1}^{G} \left(\begin{smallmatrix} 0 & 0 \\ 0 & U_i^T U_i C_{(i)}^{-T} \end{smallmatrix}\right)\right) \\ &= \mathrm{Vech}\left(\left(D_T - \sum_{i=1}^{G} \left(\begin{smallmatrix} 0 & 0 \\ 0 & U_i^T U_i C_{(i)}^{-T} \end{smallmatrix}\right) D_C^{-1}\right) D_C\right), \end{aligned} \tag{4.32}$$

where $D_T = \mathrm{diag}(t_1, t_2, \dots, t_G)$, $D_C = \mathrm{diag}(C_{1,1}, C_{2,2}, \dots, C_{G,G})$ and Vech is the half-vectorization operator which stacks the columns of its matrix argument from the principal diagonal downwards [59]. That is, if $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$, then $\mathrm{Vech}(A) = \begin{pmatrix} a_{1:n,1}^T & a_{2:n,2}^T & \cdots & a_{n,n} \end{pmatrix}^T$.

From

$$\begin{array}{cc} & \begin{array}{cc} i-1 & G-i+1 \end{array} \\ \begin{array}{c} i-1 \\ G-i+1 \end{array} & \begin{pmatrix} 0 & 0 \\ 0 & I_{G-i+1} \end{pmatrix} \end{array} C^{-1} = \begin{array}{cc} \begin{array}{cc} i-1 & G-i+1 \end{array} & \\ \begin{pmatrix} 0 & 0 \\ 0 & C_{(i)}^{-1} \end{pmatrix} & \begin{array}{c} i-1 \\ G-i+1 \end{array} \end{array}$$

it follows that

$$D_C^{-1} \left(\begin{smallmatrix} 0 & 0 \\ 0 & C_{(i)}^{-1} U_i^T U_i \end{smallmatrix} \right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & I_{G-i+1} \end{smallmatrix} \right) \widetilde{C}^{-1} \left(\begin{smallmatrix} 0 & 0 \\ 0 & U_i^T U_i \end{smallmatrix} \right),$$

where $\widetilde{C} = CD_C$. Thus,

$$\operatorname{Vech} \left(\sum_{i=1}^{G} \begin{pmatrix} 0 & 0 \\ 0 & U_i^T U_i C_{(i)}^{-T} \end{pmatrix} D_C^{-1} \right) = \bar{A} \operatorname{Vech}(\widetilde{C}^{-T}), \tag{4.33}$$

where

$$\bar{A} = L_G \left(\sum_{i=1}^{G} \left(\begin{smallmatrix} 0 & 0 \\ 0 & I_{G-i+1} \end{smallmatrix} \right) \otimes \left(\begin{smallmatrix} 0 & 0 \\ 0 & U_i^T U_i \end{smallmatrix} \right) \right) L_G^T \tag{4.34}$$

and the $G(G+1)/2 \times G^2$ elimination matrix L_G is defined by $\operatorname{Vech}(X) = L_G \operatorname{Vec}(X)$, for any matrix $X \in \mathbb{R}^{G \times G}$ [59].

Now, if $\bar{U}_i = \begin{pmatrix} 0_{h_i \times (i-1)} & U_i \end{pmatrix} \in \mathbb{R}^{h_i \times G}$, then $\bar{A}$ in (4.34) can be written as

$$\begin{aligned} \bar{A} = \sum_{i=1}^{G} & L_G \left(\left(\begin{smallmatrix} 0 & 0 \\ 0 & I_{G-i+1} \end{smallmatrix} \right) \otimes \bar{U}_i^T \bar{U}_i \right) L_G^T \\ &= \sum_{i=1}^{G} \bigoplus_{j=1}^{G} \left(\begin{smallmatrix} 0_{(G-j+1)\times(j-1)} & I_{(G-j+1)} \end{smallmatrix} \right) \bar{U}_i^T \bar{U}_i \left(\begin{smallmatrix} 0_{(G-j+1)\times(j-1)} \\ I_{(G-j+1)} \end{smallmatrix} \right) \\ &= \bigoplus_{j=1}^{G} \bar{A}_j, \end{aligned}$$

where $\bar{A}_j \in \mathbb{R}^{(G-j+1)\times(G-j+1)}$ is given by

$$\begin{aligned} \bar{A}_j &= \left(\begin{smallmatrix} 0_{(G-j+1)\times(j-1)} & I_{(G-j+1)} \end{smallmatrix} \right) \left(\sum_{i=j}^{G} \bar{U}_i^T \bar{U}_i \right) \left(\begin{smallmatrix} 0_{(G-j+1)\times(j-1)} \\ I_{(G-j+1)} \end{smallmatrix} \right) \\ &= \begin{pmatrix} u_{1,i} & \cdots & u_{1,G} \\ \vdots & & \vdots \\ u_{i,i} & \cdots & u_{i,G} \end{pmatrix}^T \begin{pmatrix} u_{1,i} & \cdots & u_{1G} \\ \vdots & & \vdots \\ u_{i,i} & \cdots & u_{i,G} \end{pmatrix}. \end{aligned}$$

Notice that if $G > t_1$, then $\bar{A}_1$, and thus $\bar{A}$, is semidefinite.

From (4.32) and (4.33) it follows that the solution of the nonlinear equation in (4.28b) derives from the solution of the symmetric linear system

$$\bar{A}\,\mathrm{Vech}(M) = \mathrm{Vech}(D_T),$$

or, equivalently, from solving the set of symmetric linear systems

$$\bar{A}_i M_{i,i:G} = t_i e_1 \qquad (i = i = 1, 2, \cdots, G), \tag{4.35}$$

where $M \equiv \widetilde{C}^{-1}$ and e_1 denotes the first column of the identity matrix. Once $\widetilde{C} = M^{-1}$ is computed, from the definition of $\widetilde{C}$ it follows that the elements of C are given by

$$C_{i,j} = \begin{cases} \sqrt{\widetilde{C}_{i,i}}, & \text{for } i = j, \\ \widetilde{C}_{i,j}/C_{i,i} & \text{for } j = 1, 2, \ldots, i-1, \end{cases}$$

where it has been assumed that $\bar{A}$ is positive definite and thus, $\widetilde{C}_{i,i} > 0$. Notice that when $t_1 < G$, $\bar{A}_1$, and thus $\bar{A}$, is positive semidefinite, which implies that (4.35) may not have solutions.

4.6 Conclusions

Computationally efficient methods to solve the SUR model with unequal size of observations (SUR-USO) which is treated as a GLLSP have been proposed. The algorithms use the GQRD to solve the GLLSP by exploiting the block-sparse structure of the matrices. The first algorithm initially computes the QRD of the exogenous matrix by annihilating from bottom to the top blocks of observations which consist of a non-zero block-superdiagonal. The annihilation of the blocks is obtained by orthogonal transformations which do not create any fill-in. These transformations are also applied from the left of the Cholesky factor and then, a sequence of orthogonal factorizations is applied to retriangularize it from the right. The second recursive algorithm interleaves the QRD and RQD of the exogenous and modified Cholesky factors, respectively. This avoids the explicit computation of the RQD and thus, reduces the computational burden of the estimation procedure.

The algorithms presented here assumed for simplicity that $t_1 \geq k_i$, $(i = 1, \ldots, G)$. This implies that $\bar{R}^{(G-i)}$ in (4.13a) is upper triangular and not trapezoidal. Generally this assumption should be relaxed and the algorithms modified to deal with cases where the QRD (4.13a) yields a trapezoidal. This generalization will allow the investigation of alternative block-generalizations of Givens sequences to compute the QRD (4.8a) without imposing additional assumptions so that $\bar{R}^{(G-i)}$ is triangular.

For the case of normally distributed disturbances the maximum likelihood estimation has been considered. A closed-form solution of the Cholesky factor of the covariance matrix has been derived by solving the first order conditions (4.28). This resulted an iterative procedure to estimate the SUR-USO model when the variance–covariance matrix Σ is unknown. Furthermore, this procedure never yields a non-definite estimator for Σ.

The extension of the proposed methods to solve SUR models with missing observations will be investigated [33]. Currently, the adaptation and (parallel) implementation of the recursive algorithm to solve the standard SUR model – with equal size observations – is considered.

Chapter 5

A comparative study of algorithms for solving seemingly unrelated regressions models

---

Abstract:

The computational efficiency of various algorithms for solving Seemingly Unrelated Regressions (SUR) models is investigated. Some of the algorithms adapt known methods; others are new. The first transforms the SUR model to an ordinary linear model and uses the QR decomposition to solve it. Three others employ the generalized QR decomposition to solve the SUR model formulated as a generalized linear least squares problem. Strategies to exploit the structure of the matrices involved are developed. The algorithms are reconsidered for solving the SUR model after it has been transformed to one of smaller dimensions.

---

[1] This chapter is a reprint of the paper: D.A. Belsley, P. Foschi, E.J. Kontoghiorghes. A comparative study of algorithms for solving seemingly unrelated regressions models. *Computaitonal Statistics and Data Analysis,* 2003 (In press).

Appendix 5.C has been published as: P. Foschi and E. J. Kontoghiorghes. Estimating SUR models with orthogonal regressors: computational aspects. *Linear Algebra and Applications*, 2003 (In press).

5.1 Introduction

The Seemingly Unrelated Regressions (SUR) model is defined by the set of regressions

$$y_i = X_i\beta_i + u_i, \qquad i = 1, \ldots, G,$$

where $X_i \in \mathbb{R}^{M\times k_i}$ has full column rank, $y_i \in \mathbb{R}^M$, and the M–element disturbance vector $u_i \sim (0, \sigma_{i,i}I_M)$ and is contemporaneously correlated across the equations so $E(u_iu_j^T) = \sigma_{i,j}I_M$ [78, 79, 83, 86, 87]. Compactly, the SUR model is written

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_G \end{pmatrix} = \begin{pmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_G \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_G \end{pmatrix}$$

or

$$\operatorname{Vec}(Y) = \left(\oplus_{i=1}^G X_i\right)\operatorname{Vec}(\{\beta_i\}_G) + \operatorname{Vec}(U), \tag{5.1}$$

where $Y = (y_1 \ldots y_G)$, $U = (u_1 \ldots u_G)$, $\oplus_{i=1}^G X_i \equiv \oplus_i X_i \equiv \operatorname{diag}(X_1, \ldots, X_G)$, denotes the direct sum of matrices, $\{\beta_i\}_G$ denotes the set of vectors $\beta_1, \ldots, \beta_G$, and $\operatorname{Vec}(\cdot)$ is the column stack operator. The disturbance term $\operatorname{Vec}(U) \sim (0, \Sigma \otimes I_M)$, where $\Sigma = [\sigma_{i,j}] \in \mathbb{R}^{G\times G}$ is symmetric and non-negative definite and $\otimes$ denotes the Kronecker product. In this treatment the following properties of the Kronecker product will be used: $(A \otimes B)(C \otimes D) = AC \otimes BD$, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ and $\operatorname{Vec}(ABC) = (C^T \otimes A)\operatorname{Vec}(B)$. For notational convenience, the subscript G in the set operator $\{\cdot\}$ is omitted, and $\oplus_{i=1}^G$ is abbreviated as $\oplus_i$. Also, $\operatorname{Vec}(\{\beta_i\})$ is denoted simply β, so $\beta^T \equiv \begin{pmatrix}\beta_1^T & \cdots & \beta_G^T\end{pmatrix}$ [69]. The notation is consistent with that employed in [22]. The standard colon notation will be used to denote subvectors and submatrices [28].

When Σ is non-singular, a Best Linear Unbiased Estimator (BLUE) of β results from solving the Generalized (linear) Least Squares (GLS) problem

$$\operatorname*{argmin}_{\beta_1,\ldots,\beta_G} \left\| \operatorname{Vec}(Y) - \operatorname{Vec}(\{X_i\beta_i\})\right\|_{\Sigma^{-1}\otimes I_M}, \tag{5.2}$$

which can be obtained from the normal equations

$$\left((\oplus_i X_i^T)(\Sigma^{-1} \otimes I_M)(\oplus_i X_i)\right)\hat{\beta} = (\oplus_i X_i^T)\operatorname{Vec}(Y\Sigma^{-1}). \tag{5.3}$$

This solution, however, can be unstable when the matrices are ill-conditioned and explicit matrix inversions are used [9, 57]. Alternatively, multiplying (5.1) by $(C^{-1} \otimes I_M)$ gives the ordinary linear model (OLM)

$$\mathrm{Vec}(YC^{-T}) = (C^{-1} \otimes I_M)(\oplus_i X_i)\beta + \mathrm{Vec}(UC^{-T}), \tag{5.4}$$

where $C \in \mathbb{R}^{G \times G}$ is a Cholesky decomposition of $\Sigma \equiv CC^T$ and is upper triangular. Computing the least squares estimator of (5.4) derives the BLUE of the SUR model (5.1) [66].

The SUR model can also be formulated as a generalized linear least squares problem (GLLSP)

$$\mathrm{argmin} \left\| V \right\|_F^2 \quad \text{subject to} \quad \mathrm{Vec}(Y) = (\oplus_i X_i)\beta + (C \otimes I_M)\,\mathrm{Vec}(V), \tag{5.5}$$

where $VC^T = U$, $\mathrm{Vec}(V) \sim (0, I_{GM})$ and $\| \cdot \|_F$ denotes the *Frobenius* norm [46, 50]. This approach allows the derivation of algorithms that are numerically more stable than those based on (5.4). Furthermore, the GLLSP allows solution of the BLUE of (5.1) even when C is not full rank, that is, when Σ is singular [55, 61, 63, 76].

Often, Σ is unknown and an iterative procedure is used to obtain the feasible GLS estimator [83]. Given a consistent estimator of Σ the solution of the model (5.2) derives an estimator for β. From the residual of the estimated coefficients, another estimator of Σ is obtained. This procedure is repeated until convergence. Thus, the GLS problem (5.2), or the corresponding GLLSP (5.5), is solved a number of times for different Σ. Here, the computational cost of deriving the estimator β during a single iteration is considered. The particular properties of the SUR model that affect the convergence of the iterative estimation procedure are not investigated.

In this work, the computational efficiencies of various methods for computing the BLUE of the SUR model are considered. Some of the algorithms are well known while others are new. All of the algorithms are based on an orthogonal factorization obtained through the QR decomposition. In the next section the solutions of the SUR model using the QR and generalized QR decompositions are considered. Recursive estimation algorithms are presented in section 5.3. Size reduction of large-scale SUR models is shown in section 5.4. The computational results are discussed in section 5.5. Section 5.6 provides summary comments.

5.2 Numerical estimation of the SUR model

5.2.1 Estimating the OLM using the QR decomposition

The OLM (5.4) can be written as

$$\bar{y} = \overline{X}\beta + \bar{\varepsilon}, \tag{5.6}$$

where $\bar{y} = \text{Vec}(YC^{-T})$, $\overline{X} = (C^{-1} \otimes I_M)(\oplus_i X_i)$, and $\bar{\varepsilon} = \text{Vec}(UC^{-T})$. Let the QR decomposition (QRD) of $\overline{X}$ be given by

$$\overline{Q}^T\overline{X} = \begin{pmatrix} \overline{R} \\ 0 \end{pmatrix} \begin{matrix} {\scriptstyle K} \\ {\scriptstyle GM-K} \end{matrix} \qquad \text{and} \qquad \overline{Q}^T\bar{y} = \begin{pmatrix} \overline{y}_A \\ \overline{y}_B \end{pmatrix} \begin{matrix} {\scriptstyle K} \\ {\scriptstyle GM-K} \end{matrix} \quad , \tag{5.7}$$

where $\overline{R}$ is upper triangular, $\overline{Q} \in \mathbb{R}^{GM \times GM}$ is orthogonal, and $K = \sum_i k_i$ [9, 28]. The least-squares estimator of β is given by solving the triangular system

$$\overline{R}\beta = \bar{y}\ . \tag{5.8}$$

This straightforward solution of (5.4) is computationally inefficient since it computes $\overline{X}$ explicitly and ignores its sparsity.

To solve (5.4) efficiently, consider the QRD of X_i:

$$Q_i^T X_i = \begin{pmatrix} R_i \\ 0 \end{pmatrix}, \qquad \text{with} \qquad Q_i^T = \begin{pmatrix} \widetilde{Q}_i^T \\ \widehat{Q}_i^T \end{pmatrix} \begin{matrix} {\scriptstyle k_i} \\ {\scriptstyle M-k_i} \end{matrix} \quad , \tag{5.9}$$

where $Q_i \in \mathbb{R}^{M \times M}$ is orthogonal and $R_i \in \mathbb{R}^{k_i \times k_i}$ is upper triangular. From (5.9) it follows that the QRD of $\oplus_i X_i$ is given by

$$Q^T(\oplus_i X_i) = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \begin{matrix} {\scriptstyle K} \\ {\scriptstyle GM-K} \end{matrix} \quad , \tag{5.10}$$

where $Q = \begin{pmatrix} \oplus_i \widetilde{Q}_i & \oplus_i \widehat{Q}_i \end{pmatrix}$. Premultiplying (5.4) by Q^T gives

$$Q^T \, \text{Vec}(YC^{-T}) = Q^T(C^{-1} \otimes I_M)(\oplus_i X_i)\beta + Q^T \, \text{Vec}(UC^{-T}) \ \ ,$$

or

$$\begin{pmatrix} \text{Vec}\left(\{\tilde{\bar{y}}_i\}\right) \\ \text{Vec}\left(\{\hat{\bar{y}}_i\}\right) \end{pmatrix} = \begin{pmatrix} \widetilde{\overline{W}} \\ \widehat{\overline{W}} \end{pmatrix} \beta + \begin{pmatrix} \text{Vec}(\widetilde{\overline{V}}) \\ \text{Vec}(\widehat{\overline{V}}) \end{pmatrix}, \tag{5.11}$$

where

$$\widetilde{\widetilde{W}} = \begin{array}{c} \begin{array}{ccc} k_1 & \cdots & k_G \end{array} \\ \begin{pmatrix} \widetilde{\widetilde{W}}_{1,1} & \cdots & \widetilde{\widetilde{W}}_{1,G} \\ \vdots & & \vdots \\ \widetilde{\widetilde{W}}_{G,1} & \cdots & \widetilde{\widetilde{W}}_{G,G} \end{pmatrix} \end{array} \begin{array}{l} \\ k_1 \\ \vdots \\ k_G \end{array}, \qquad \widehat{\widehat{W}} = \begin{array}{c} \begin{array}{ccc} k_1 & \cdots & k_G \end{array} \\ \begin{pmatrix} \widehat{\widehat{W}}_{1,1} & \cdots & \widehat{\widehat{W}}_{1,G} \\ \vdots & & \vdots \\ \widehat{\widehat{W}}_{G,1} & \cdots & \widehat{\widehat{W}}_{G,G} \end{pmatrix} \end{array} \begin{array}{l} \\ M-k_1 \\ \vdots \\ M-k_G \end{array}, \tag{5.12a}$$

$$\widetilde{\widetilde{W}}_{i,j} = \begin{cases} \gamma_{i,j}\widetilde{Q}_i^T X_j, & \text{if } i < j \\ \gamma_{i,i} R_i, & \text{if } i = j \\ 0, & \text{if } i > j \end{cases}, \tag{5.12b}$$

$$\widehat{\widehat{W}}_{i,j} = \begin{cases} \gamma_{i,j}\widehat{Q}_i^T X_j, & \text{if } i < j \\ 0, & \text{if } i \geq j, \end{cases} \tag{5.12c}$$

and $\gamma_{i,j}$ is the (i,j)-th element of C^{-1}. Notice that $\widetilde{\widetilde{W}}$ and $\widehat{\widehat{W}}$ are block upper-triangular and strictly block upper-triangular matrices, respectively.

Now, compute a row-updating QRD (hereafter abbreviated to UQRD)

$$\bar{Q}^T \begin{pmatrix} \widetilde{\widetilde{W}} \\ \widehat{\widehat{W}} \end{pmatrix} = \begin{pmatrix} \overline{R} \\ 0 \end{pmatrix} \begin{array}{l} K \\ GM-K \end{array} \tag{5.13a}$$

and

$$\bar{Q}^T \begin{pmatrix} \operatorname{Vec}(\{\tilde{\tilde{y}}_i\}) \\ \operatorname{Vec}(\{\hat{\hat{y}}_i\}) \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{y}_* \end{pmatrix} \begin{array}{l} K \\ GM-K \end{array}. \tag{5.13b}$$

It follows that the least-squares solution of (5.11), and thus the BLUE of β, is given by (5.8). Algorithm 6 summarizes these steps for solving (5.4). Two block strategies, the column- and diagonally-based methods, that can be used to compute (5.13) – step 8 – are described in Appendix 5.A.

5.2.2 The GLLSP and generalized QRD

The solution of the GLLSP (5.5) can be obtained by computing the generalized QRD (GQRD) of $\oplus_i X_i$ and $(C \otimes I_M)$ [9, 46, 52, 55, 64]; that is, by computing the QRD (5.10) and the RQ

Algorithm 6 Ordinary Least Squares estimation of the OLM (5.4).

```
Input: The regressors $X_1, \ldots, X_G$ and $Y$ and the disturbances covariance matrix $\Sigma$
Output: The vector of parameters $\beta$
Compute $\Sigma = CC^T$
Compute $C^{-1} = [\gamma_{i,j}]$ and $\overline{Y} = \begin{pmatrix} \bar{y}_1 & \cdots & \bar{y}_G \end{pmatrix} = YC^{-T}$
for $i = 1, \ldots, G$ do
    Compute the QRD (5.9)
    Compute $\tilde{\tilde{y}}_i = \widetilde{Q}_i^T \bar{y}_i$ and $\hat{y}_i = \widehat{Q}_i^T \bar{y}_i$
end for
Compute $\widetilde{\widetilde{W}}$ and $\widehat{\widehat{W}}$ as in (5.12a)
Compute the UQRD (5.13a) and (5.13b)
Solve the triangular system (5.8) for $\beta$
```

decomposition

$$Q^T(C \otimes I_M)P = W \equiv \begin{array}{cc} & \begin{array}{cc} K & GM-K \end{array} \\ & \begin{pmatrix} W_{AA} & W_{AB} \\ 0 & W_{BB} \end{pmatrix} \begin{array}{l} K \\ GM-K \end{array} \end{array} , \tag{5.14}$$

where $K = \sum_i k_i$, $P \in \mathbb{R}^{GM \times GM}$ is orthogonal, and W_{BB} is upper triangular. Premultiplying the constraints in (5.5) by Q^T and using $\text{Vec}(V) \equiv PP^T \,\text{Vec}(V)$, the GLLSP can be written as

$$\underset{\beta,\{\tilde{v}_i\},\{\hat{v}_i\}}{\text{argmin}} \sum_{i=1}^{G} \left(\left\| \tilde{v}_i \right\|^2 + \left\| \hat{v}_i \right\|^2 \right) \quad \text{subject to}$$

$$\begin{pmatrix} \text{Vec}\left(\{\tilde{y}_i\}\right) \\ \text{Vec}\left(\{\hat{y}_i\}\right) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \beta + \begin{pmatrix} W_{AA} & W_{AB} \\ 0 & W_{BB} \end{pmatrix} \begin{pmatrix} \text{Vec}\left(\{\tilde{v}_i\}\right) \\ \text{Vec}\left(\{\hat{v}_i\}\right) \end{pmatrix} , \tag{5.15}$$

where $\widetilde{Q}_i^T y_i = \tilde{y}_i$, $\widehat{Q}_i^T y_i = \hat{y}_i$, $P^T \,\text{Vec}(V) = \begin{pmatrix} \text{Vec}\left(\{\tilde{v}_i\}\right)^T & \text{Vec}\left(\{\hat{v}_i\}\right)^T \end{pmatrix}^T$, $\tilde{y}_i, \tilde{v}_i \in \mathbb{R}^{k_i}$ and $\hat{y}_i, \hat{v}_i \in \mathbb{R}^{M-k_i}$. The solution of (5.15) is given by $\text{Vec}\left(\{\tilde{v}_i\}\right) = 0$ and

$$\begin{pmatrix} \oplus_i R_i & W_{AB} \\ 0 & W_{BB} \end{pmatrix} \begin{pmatrix} \beta \\ \text{Vec}\left(\{\hat{v}_i\}\right) \end{pmatrix} = \begin{pmatrix} \text{Vec}\left(\{\tilde{y}_i\}\right) \\ \text{Vec}\left(\{\hat{y}_i\}\right) \end{pmatrix} . \tag{5.16}$$

The RQD (5.14) derives in two stages. The first computes the permutation

$$Q^T(C \otimes I_T)\Pi = \begin{array}{cc} & \begin{array}{cc} K & GM-K \end{array} \\ & \begin{pmatrix} \widetilde{W}_{AA} & \widetilde{W}_{AB} \\ \widetilde{W}_{BA} & \widetilde{W}_{BB} \end{pmatrix} \begin{array}{l} K \\ GM-K \end{array} \end{array} , \tag{5.17}$$

where $\Pi = \begin{pmatrix} \oplus_i \, (I_{k_i} \;\, 0)^T & \oplus_i \, (0 \;\, I_{M-k_i})^T \end{pmatrix}$. This results in $\widetilde{W}_{AA}$, $\widetilde{W}_{AB}$, $\widetilde{W}_{BA}$ and $\widetilde{W}_{BB}$ being block upper-triangular. The second stage computes the RQD

$$\begin{pmatrix} \widetilde{W}_{BA} & \widetilde{W}_{BB} \end{pmatrix} \widetilde{P} = \begin{pmatrix} 0 & W_{BB} \end{pmatrix} \tag{5.18a}$$

and

$$\begin{pmatrix} \widetilde{W}_{AA} & \widetilde{W}_{AB} \end{pmatrix} \widetilde{P} = \begin{pmatrix} W_{AA} & W_{AB} \end{pmatrix}, \tag{5.18b}$$

where $\widetilde{P} \in \mathbb{R}^{GM \times GM}$ is orthogonal. Notice that P in (5.14) is given by $\Pi\widetilde{P}$ and that (5.18) does not compute the RQD of the whole matrix in (5.17). The leading submatrix W_{AA}, which is not used in the solution of the GLLSP, is not triangularized. Furthermore, the RQD (5.18a) is equivalent to the QL decomposition

$$\widetilde{P}^T \begin{pmatrix} \widetilde{W}_{BA}^T \\ \widetilde{W}_{BB}^T \end{pmatrix} = \begin{pmatrix} 0 \\ W_{BB}^T \end{pmatrix}.$$

This indicates that (5.18a) can be computed using adaptations of the diagonally-based and column-based strategies (see Appendix 5.A) that are used for the computation of (5.13a) [44, 46]. Furthermore, these strategies produce W_{AA} in (5.18b) that are block upper-triangular.

The first step of the diagonally-based strategy annihilates the main block-diagonal of $\widetilde{W}_{BA}$. However, the permutation in (5.17) along with this step is equivalent to applying Q to the the right of $Q^T(C \otimes I_M)$; that is

$$Q^T(C \otimes I_M)Q = \widetilde{W}^{(0)} \equiv \begin{array}{c} \begin{array}{cc} \scriptstyle K & \scriptstyle GM-K \end{array} \\ \begin{pmatrix} \widetilde{W}_{AA}^{(0)} & \widetilde{W}_{AB}^{(0)} \\ \widetilde{W}_{BA}^{(0)} & \widetilde{W}_{BB}^{(0)} \end{pmatrix} \end{array} \begin{array}{l} \scriptstyle K \\ \scriptstyle GM-K \end{array}, \tag{5.19}$$

where $\widetilde{W}_{AA}^{(0)}$ and $\widetilde{W}_{BB}^{(0)}$ are block upper-triangular with the ith block of their main diagonals given by $C_{i,i}I_{k_i}$ and $C_{i,i}I_{M-k_i}$, respectively. Furthermore the matrices $\widetilde{W}_{AB}^{(0)}$ and $\widetilde{W}_{BA}^{(0)}$ are strictly block upper-triangular. Figure 5.1 shows the structure of the matrices $\widetilde{W}$ and $\widetilde{W}^{(0)}$, where $G = 5$.

Thus the remaining steps of the diagonally-based method annihilate the strictly block upper-triangular matrix $\widetilde{W}_{BA}^{(0)}$ by preserving the block-triangular structure of $\widetilde{W}_{AA}^{(0)}$ and $\widetilde{W}_{BB}^{(0)}$. This annihilation strategy is illustrated in Figure 5.2, where an arc denotes an updating RQD (URQD). Algorithm 7 summarizes the steps of this estimation procedure.

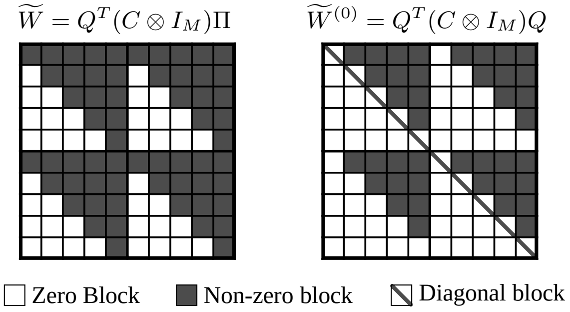


Figure 5.1: Structure of matrices $\widetilde{W}$ and $\widetilde{W}^{(0)}$, where $G = 5$.

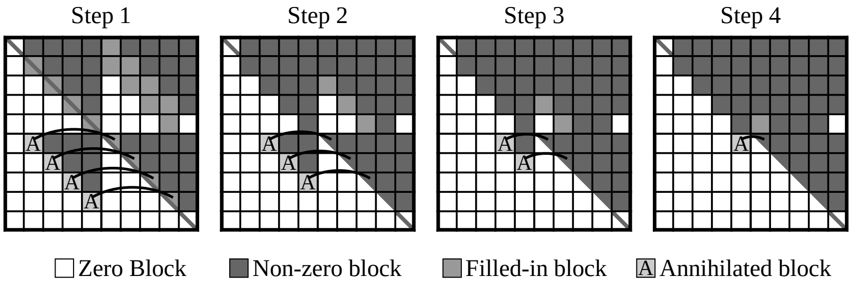


Figure 5.2: Structure of matrices $\widetilde{W}$ and $\widetilde{W}^{(0)}$, where $G = 5$.

Algorithm 7 Solution of the GLLSP (5.5) using the GQRD.

```
Input: The regressors $X_1, \ldots, X_G$ and $Y$ and the disturbances covariance matrix $\Sigma$
Output: The vector of parameters $\beta$
Compute $\Sigma = CC^T$
for $i = 1, \ldots, G$ do
    Compute the QRD (5.9)
    Compute $\tilde{y}_i = \widetilde{Q}_i^T y_i$ and $\hat{y}_i = \widehat{Q}_i^T y_i$
end for
Compute $Q^T(C \otimes I_T)Q$ as in (5.19)
Compute the URQD $\begin{pmatrix} \widetilde{W}_{BA}^{(0)} & \widetilde{W}_{BB}^{(0)} \end{pmatrix} \widetilde{P}^{(0)} = \begin{pmatrix} 0 & W_{BB} \end{pmatrix}$
Compute $\begin{pmatrix} \widetilde{W}_{AA}^{(0)} & \widetilde{W}_{AB}^{(0)} \end{pmatrix} \widetilde{P}^{(0)} = \begin{pmatrix} W_{AA} & W_{AB} \end{pmatrix}$
Solve the triangular system (5.16) for $\beta$ and $\mathrm{Vec}(\{\hat{v}_i\})$
```

5.2.3 An interleaving approach to solving the GLLSP

The RQD (5.14) is the most expensive operation in computing the GQRD of $\oplus_i X_i$ and $C \otimes I_M$ (see Appendix 5.B). An iterative procedure that does not compute (5.14) can be employed [63]. At each iteration a smaller problem is solved. Let $\check{X}^{(0)} = \oplus_{i=1}^{G} X_i$, $\check{W}^{(0)} = C \otimes I_M$, $\check{y}^{(0)} = \operatorname{Vec}(Y)$, and $\check{v}^{(0)} = \operatorname{Vec}(V)$. The sth ($s = 0, \ldots, G-1$) iteration deals with the GLLSP

$$\underset{\check{v}^{(s)},\beta}{\operatorname{argmin}} \left\|\check{v}^{(s)}\right\| \quad \text{subject to} \quad \check{y}^{(s)} = \check{X}^{(s)}\beta + \check{W}^{(s)}\check{v}^{(s)}, \tag{5.20}$$

computing the factorizations

$$\check{Q}_s^T \check{X}^{(s)} = \begin{array}{cc} K & \\ \left(\begin{array}{c} \check{X}^{(s+1)} \\ 0 \end{array}\right) & \begin{array}{l} \mu_{s+1} \\ M-k_{G-s} \end{array} \end{array} \tag{5.21a}$$

and

$$\big(\check{Q}_s^T \check{W}^{(s)} \check{Q}_s\big)\check{P}_s = \begin{array}{cc} \begin{array}{cc} \mu_{s+1} & M-k_{G-s} \end{array} & \\ \left(\begin{array}{cc} \check{W}^{(s+1)} & \widetilde{W}_{AB}^{(s)} \\ 0 & \widetilde{W}_{BB}^{(s)} \end{array}\right) & \begin{array}{l} \mu_{s+1} \\ M-k_{G-s} \end{array} \end{array}, \tag{5.21b}$$

where $\check{Q}_s, \check{P}_s \in \mathbb{R}^{\mu_s \times \mu_s}$ are orthogonal, $\mu_s = (G-s)M + \lambda_{G-s+1}$, $\lambda_i = \sum_{j=i}^{G} k_j$,

$$\check{X}^{(s+1)} = \begin{array}{cc} \begin{array}{cccccc} k_1 & \cdots & k_{G-s-1} & k_{G-s} & \cdots & k_G \end{array} & \\ \left(\begin{array}{cccccc} X_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & X_{G-s-1} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & R_{G-s} & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & R_G \end{array}\right) & \begin{array}{l} M \\ \vdots \\ M \\ k_{G-s} \\ \vdots \\ k_G \end{array} \end{array}, \tag{5.22}$$

$$\check{W}^{(s+1)} = \begin{array}{cc} \begin{array}{cc} (G-s-1)M & \lambda_{G-s} \end{array} & \\ \left(\begin{array}{cc} \check{W}_{AA}^{(s+1)} & \check{W}_{AB}^{(s+1)} \\ 0 & \check{W}_{BB}^{(s+1)} \end{array}\right) & \begin{array}{l} (G-s-1)M \\ \lambda_{G-s} \end{array} \end{array}, \tag{5.23}$$

R_{G-s+i} and $\check{W}^{(s+1)}_{BB}$ are upper-triangular, and $\check{W}^{(s+1)}_{AA} = C_{1:G-s-1,1:G-s-1} \otimes I_M$. Furthermore,

$$\check{Q}_s = \begin{pmatrix} I_{(G-s-1)M} & 0 & 0 & 0 \\ 0 & \widetilde{Q}_{G-s} & 0 & \widehat{Q}_{G-s} \\ 0 & 0 & I_{\lambda_{G-s+1}} & 0 \end{pmatrix} \tag{5.24}$$

and

$$\check{Q}_s^T \check{W}^{(s)} \check{Q}_s = \begin{array}{c@{}c} \begin{array}{cccc} \scriptstyle (G-s-1)M & \scriptstyle k_{G-s} & \scriptstyle \lambda_{G-s+1} & \scriptstyle M-k_{G-s} \end{array} & \\ \begin{pmatrix} \check{W}^{(s+1)}_{AA} & \widehat{W}^{(s)}_{AB} & \widehat{W}^{(s)}_{AC} & \widehat{W}^{(s)}_{AD} \\ 0 & C_{G-s,G-s} I_{k_{G-s}} & \widehat{W}^{(s)}_{BC} & 0 \\ 0 & 0 & \widehat{W}^{(s)}_{CC} & 0 \\ 0 & 0 & \widehat{W}^{(s)}_{DC} & C_{G-s,G-s} I_{M-k_{G-s}} \end{pmatrix} & \begin{array}{l} \scriptstyle (G-s-1)M \\ \scriptstyle k_{G-s} \\ \scriptstyle \lambda_{G-s+1} \\ \scriptstyle M-k_{G-s} \end{array} \end{array},$$

where $\widehat{W}^{(s)}_{AC} = \begin{pmatrix} I_{(G-s-1)M} & 0 \end{pmatrix} \check{W}^{(s)}_{AB}$, $\widehat{W}^{(s)}_{CC} = \check{W}^{(s)}_{BB}$ and $\check{W}^{(0)}_{BB}$, and consequently $\check{W}^{(0)}_{CC}$, has zero dimension.

Note that (5.21a) computes the QRD of X_s, while the RQD (5.21b) is equivalent to the URQD

$$\begin{pmatrix} \widehat{W}^{(s)}_{DC} & C_{G-s,G-s} I_{M-k_{G-s}} \end{pmatrix} \overline{P}_s = \begin{pmatrix} 0 & \widetilde{W}^{(s)}_{BB} \end{pmatrix} \tag{5.25a}$$

and

$$\begin{pmatrix} \widehat{W}^{(s)}_{AC} & \widehat{W}^{(s)}_{AD} \\ \widehat{W}^{(s)}_{BC} & 0 \\ \widehat{W}^{(s)}_{CC} & 0 \end{pmatrix} \overline{P}_s = \begin{pmatrix} \widetilde{W}^{(s)}_{AC} & \widetilde{W}^{(s)}_{AD} \\ \widetilde{W}^{(s)}_{BC} & \widetilde{W}^{(s)}_{BD} \\ \widetilde{W}^{(s)}_{CC} & \widetilde{W}^{(s)}_{CD} \end{pmatrix}, \tag{5.25b}$$

where

$$\check{P}_s = \begin{pmatrix} I_{(\mu_{s+1} - \lambda_{G-s+1})} & 0 \\ 0 & \overline{P}_s \end{pmatrix}, \qquad \check{W}^{(s+1)}_{AB} = \begin{pmatrix} \widetilde{W}^{(s)}_{AB} & \widetilde{W}^{(s)}_{AC} \end{pmatrix},$$

$$\check{W}^{(s+1)}_{BB} = \begin{pmatrix} C_{G-s,G-s} I_{k_{G-s}} & \widetilde{W}^{(s)}_{BC} \\ 0 & \widetilde{W}^{(s)}_{CC} \end{pmatrix} \quad \text{and} \quad \widetilde{W}^{(s)}_{AB} = \begin{pmatrix} \widetilde{W}^{(s)}_{AD} \\ \widetilde{W}^{(s)}_{BD} \\ \widetilde{W}^{(s)}_{CD} \end{pmatrix}.$$

Let

$$\check{Q}_s^T \check{y}^{(s)} = \begin{pmatrix} \check{y}^{(s)}_A \\ \check{y}^{(s)}_B \end{pmatrix} \begin{array}{l} \scriptstyle \mu_{s+1} \\ \scriptstyle M-k_{G-s} \end{array} \quad \text{and} \quad \check{P}_s^T \check{Q}_s^T \check{v}^{(s)} = \begin{pmatrix} \check{v}^{(s+1)} \\ \check{v}^{(s)}_B \end{pmatrix} \begin{array}{l} \scriptstyle \mu_{s+1} \\ \scriptstyle M-k_{G-s} \end{array}.$$

Premultiplying the constraints in (5.20) by $\check{Q}_s^T$ and using (5.21), it follows that the GLLSP is equivalent to

$$\underset{\check{v}_A^{(s)},\check{v}_B^{(s)},\beta}{\operatorname{argmin}} \ \left\|\check{v}_A^{(s)}\right\|^2 + \left\|\check{v}_B^{(s)}\right\|^2 \quad \text{subject to}$$

$$\begin{pmatrix} \check{y}_A^{(s)} \\ \check{y}_B^{(s)} \end{pmatrix} = \begin{pmatrix} \check{X}^{(s+1)} \\ 0 \end{pmatrix} \beta + \begin{pmatrix} \check{W}^{(s+1)} & \widetilde{W}_{AB}^{(s)} \\ 0 & \widetilde{W}_{BB}^{(s)} \end{pmatrix} \begin{pmatrix} \check{v}^{(s+1)} \\ \check{v}_B^{(s)} \end{pmatrix},$$

or, again, the smaller GLLSP

$$\underset{\check{v}^{(s+1)},\beta}{\operatorname{argmin}} \left\|\check{v}^{(s+1)}\right\| \quad \text{subject to} \quad \check{y}^{(s+1)} = \check{X}^{(s+1)}\beta + \check{W}^{(s+1)}\check{v}^{(s+1)}, \tag{5.26}$$

where $\check{v}_B^{(s)} = (\widetilde{W}_{BB}^{(s)})^{-1}\check{y}_B^{(s)}$ and $\check{y}^{(s+1)} = \check{y}_A^{(s)} - \widetilde{W}_{AB}^{(s)}\check{v}_B^{(s)}$.

The solution to (5.26) can be obtained iteratively by employing the method used for the GLLSP (5.20). At the end of iteration $(G-1)$ the GLLSP becomes

$$\underset{\check{v}^{(G)},\beta}{\operatorname{argmin}} \left\|\check{v}^{(G)}\right\| \quad \text{subject to} \quad \check{y}^{(G)} = (\oplus_i R_i)\beta + \check{W}^{(G)}\check{v}^{(G)},$$

which has solution $\check{v}^{(G)} = 0$ and $\beta = (\oplus_i R_i^{-1})\check{y}^{(G)}$.

Now consider the computation of (5.25a). Let

$$\begin{pmatrix} \widehat{W}_{DC}^{(s)} & C_{G-s,G-s} I_{M-k_{G-s}} \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} k_{G-s+1} & \cdots & k_G & M-k_{G-s} \end{array} \\ \begin{pmatrix} A_1 & \cdots & A_s & A_{s+1}^{(0)} \end{pmatrix} \end{array}. \tag{5.27}$$

The submatrices $A_1, \ldots, A_s$ are annihilated one at a time by computing the URQDs

$$\begin{pmatrix} A_i & A_{s+1}^{(i-1)} \end{pmatrix} P_i = \begin{pmatrix} 0 & A_{s+1}^{(i)} \end{pmatrix}, \qquad i = 1, \ldots, s, \tag{5.28}$$

where $A_{s+1}^{(i)}$ is upper triangular and P_i is orthogonal. Thus, in (5.25a) $\widetilde{W}_{BB}^{(s)} = A_{s+1}^{(s)}$. This produces $\widetilde{W}_{CC}^{(s)}$ with a block upper triangular structure. Figure 5.3 shows the steps for annihilating $\widehat{W}_{DC}^{(s)}$ and the fill-ins induced in $\widetilde{W}_{CC}^{(s)}$ and $\widetilde{W}_{CD}^{(s)}$ in (5.25), where $s = 4$. Algorithm 8 summarizes the steps of the interleaving procedure for solving the GLLSP (5.5). Figure 5.4 illustrates the computations of the second step ($s = 2$), where $G = 5$.

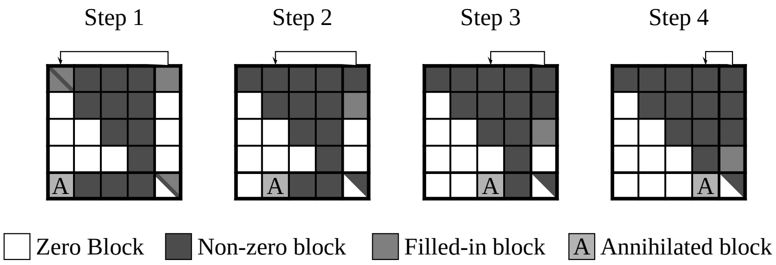


Figure 5.3: Computation of (5.25) at step $s = 4$.

Algorithm 8 Solution of the GLLSP (5.5) using the interleaving approach.

```
Input: The regressors $X_1, \ldots, X_G$ and $Y$ and the disturbances covariance matrix $\Sigma$
Output: The vector of parameters $\beta$
Compute $\Sigma = CC^T$
Let $\check{X}^{(0)} = \oplus_{i=1}^{G} X_i$, $\check{W}^{(0)} = C \otimes I_M$ and $\check{y}^{(0)} = \text{Vec}(Y)$
for $s = 0, 1, \ldots, G-1$ do
    Compute the QRD (5.9), where $i = G - s$ and let $\check{Q}_s$ be given by (5.24)
    Compute $\begin{pmatrix} \check{y}_A^{(s)} \\ \check{y}_B^{(s)} \end{pmatrix} = \check{Q}_s^T \check{y}^{(s)}$
    Compute $\check{Q}_s^T \check{W}^{(s)} \check{Q}_s$
    Compute the URQD (5.25a) and (5.25b)
    Solve the triangular system $\widetilde{W}_{BB}^{(s)} \check{v}_B^{(s)} = \check{y}_B^{(s)}$ for $\check{v}_B^{(s)}$
    Compute $\check{y}^{(s+1)} = \check{y}^{(s)} - \widetilde{W}_{AB}^{(s)} \check{v}_B^{(s)}$
end for
Solve the triangular system $(\oplus_i R_i)\beta = \check{y}^{(G)}$ for $\beta$
```

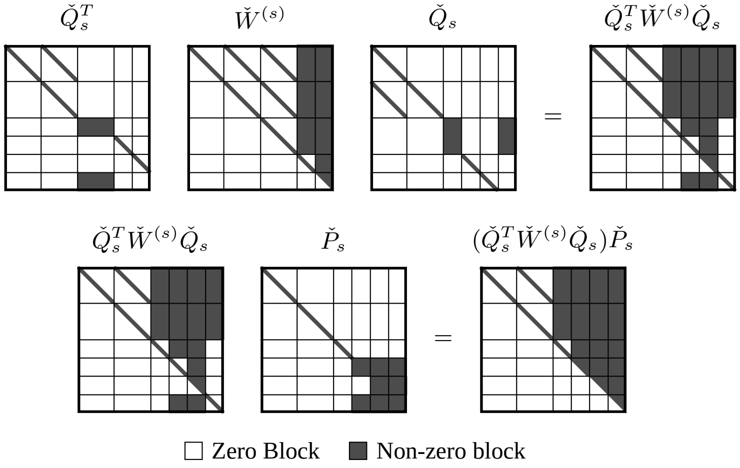


Figure 5.4: Computation of (5.21b), where $G = 5$ and $s = 2$.

5.3 A recursive algorithm for the estimation of the SUR model

The BLUE of the SUR model can be computed recursively [10, 49]. Consider the partitioning

$$X_i = \left(\begin{array}{c} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{array}\right)\begin{array}{l} M_1 \\ \vdots \\ M_p \end{array}, \quad Y = \left(\begin{array}{c} Y^{(1)} \\ \vdots \\ Y^{(p)} \end{array}\right)\begin{array}{l} M_1 \\ \vdots \\ M_p \end{array}, \quad U = \left(\begin{array}{c} U^{(1)} \\ \vdots \\ U^{(p)} \end{array}\right)\begin{array}{l} M_1 \\ \vdots \\ M_p \end{array}, \quad \text{and} \quad V = \left(\begin{array}{c} V^{(1)} \\ \vdots \\ V^{(p)} \end{array}\right)\begin{array}{l} M_1 \\ \vdots \\ M_p \end{array}, \tag{5.29}$$

for $i = 1, \ldots, G$. The SUR model (5.1) and the GLLSP (5.5) can be respectively expressed equivalently as

$$\begin{pmatrix} \mathrm{Vec}(Y^{(1)}) \\ \mathrm{Vec}(Y^{(2)}) \\ \vdots \\ \mathrm{Vec}(Y^{(p)}) \end{pmatrix} = \begin{pmatrix} \oplus_i X_i^{(1)} \\ \oplus_i X_i^{(2)} \\ \vdots \\ \oplus_i X_i^{(p)} \end{pmatrix} \beta + \begin{pmatrix} \mathrm{Vec}(U^{(1)}) \\ \mathrm{Vec}(U^{(2)}) \\ \vdots \\ \mathrm{Vec}(U^{(p)}) \end{pmatrix}$$

and

$$\underset{\beta, V^{(1)}, \ldots, V^{(p)}}{\mathrm{argmin}} \sum_{j=1}^{p} \left\| V^{(j)} \right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix} \mathrm{Vec}(Y^{(1)}) \\ \mathrm{Vec}(Y^{(2)}) \\ \vdots \\ \mathrm{Vec}(Y^{(p)}) \end{pmatrix} = \begin{pmatrix} \oplus_i X_i^{(1)} \\ \oplus_i X_i^{(2)} \\ \vdots \\ \oplus_i X_i^{(p)} \end{pmatrix} \beta + \begin{pmatrix} C \otimes I_{M_1} & 0 & \cdots & 0 \\ 0 & C \otimes I_{M_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \otimes I_{M_p} \end{pmatrix} \begin{pmatrix} \mathrm{Vec}(V^{(1)}) \\ \mathrm{Vec}(V^{(2)}) \\ \vdots \\ \mathrm{Vec}(V^{(p)}) \end{pmatrix}. \tag{5.30}$$

Assume that $M_1 \geq \max(k_1, \ldots, k_G)$, and let the GQRD of $\oplus_i X_i^{(1)}$ and $C \otimes I_{M_1}$ be given by

$$Q_{(1)}^T (\oplus_i X_i^{(1)}) = \begin{array}{c} K \\ \left(\begin{array}{c} \oplus_i R_i^{(1)} \\ 0 \end{array}\right) \end{array} \begin{array}{l} \\ K \\ GM_1 - K \end{array} \tag{5.31a}$$

and

$$Q_{(1)}^T (C \otimes I_{M_1}) P_{(1)} = W^{(1)} \equiv \begin{array}{c} \begin{array}{cc} K & GM_1 - K \end{array} \\ \left(\begin{array}{cc} W_{AA}^{(1)} & W_{AB}^{(1)} \\ 0 & W_{BB}^{(1)} \end{array}\right) \end{array} \begin{array}{l} \\ K \\ GM_1 - K \end{array}, \tag{5.31b}$$

where $R_i^{(1)}$ and $W_{(1)}$ are upper-triangular. Furthermore, let

$$Q_{(1)}^T \operatorname{Vec}(Y^{(1)}) = \begin{pmatrix} \tilde{y}^{(1)} \\ \hat{y}^{(1)} \end{pmatrix} \begin{matrix} K \\ GM_1-K \end{matrix} \quad \text{and} \quad P_{(1)}^T \operatorname{Vec}(V^{(1)}) = \begin{pmatrix} \tilde{v}^{(1)} \\ \hat{v}^{(1)} \end{pmatrix} \begin{matrix} K \\ GM_1-K \end{matrix} . \tag{5.32}$$

Using (5.31) and (5.32) it follows that the GLLSP (5.30) can be written as

$$\operatorname*{argmin}_{\substack{\beta,\tilde{v}^{(1)},\hat{v}^{(1)},\\ V^{(2)},\ldots,V^{(p)}}} \left\|\tilde{v}^{(1)}\right\|^2 + \left\|\hat{v}^{(1)}\right\|^2 + \sum_{j=2}^{p} \left\|V^{(j)}\right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix} \tilde{y}^{(1)} \\ \operatorname{Vec}(Y^{(2)}) \\ \vdots \\ \operatorname{Vec}(Y^{(p)}) \\ \hat{y}^{(1)} \end{pmatrix} = \begin{pmatrix} \oplus_i R_i^{(1)} \\ \oplus_i X_i^{(2)} \\ \vdots \\ \oplus_i X_i^{(p)} \\ 0 \end{pmatrix} \beta + \begin{pmatrix} W_{AA}^{(1)} & 0 & \cdots & 0 & W_{AB}^{(1)} \\ 0 & C \otimes I_{M_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & C \otimes I_{M_p} & 0 \\ 0 & 0 & \cdots & 0 & W_{BB}^{(1)} \end{pmatrix} \begin{pmatrix} \tilde{v}^{(1)} \\ \operatorname{Vec}(V^{(2)}) \\ \vdots \\ \operatorname{Vec}(V^{(p)}) \\ \hat{v}^{(1)} \end{pmatrix} . \tag{5.33}$$

This is equivalent to

$$\operatorname*{argmin}_{\substack{\beta,\tilde{v}^{(1)}\\ V^{(2)},\ldots,V^{(p)}}} \left\|\tilde{v}^{(1)}\right\|^2 + \sum_{j=2}^{p} \left\|V^{(j)}\right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix} y^{(1)} \\ \operatorname{Vec}(Y^{(2)}) \\ \vdots \\ \operatorname{Vec}(Y^{(p)}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i^{(1)} \\ \oplus_i X_i^{(2)} \\ \vdots \\ \oplus_i X_i^{(p)} \end{pmatrix} \beta + \begin{pmatrix} W_{AA}^{(1)} & 0 & \cdots & 0 \\ 0 & C \otimes I_{M_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \otimes I_{M_p} \end{pmatrix} \begin{pmatrix} \tilde{v}^{(1)} \\ \operatorname{Vec}(V^{(2)}) \\ \vdots \\ \operatorname{Vec}(V^{(p)}) \end{pmatrix} , \tag{5.34}$$

where $\hat{v}^{(1)} = \left(W_{BB}^{(1)}\right)^{-1}\hat{y}^{(1)}$ and $y^{(1)} = \tilde{y}^{(1)} - W_{AB}^{(1)}\hat{v}^{(1)}$.

The solution to the GLLSP (5.34) can be obtained in $(p-1)$ iterations. The sth ($s = 2, \ldots, p$) iteration solves the GLLSP

$$\operatorname*{argmin}_{\substack{\beta,\tilde{v}^{(s-1)},\\ V^{(s)},\ldots,V^{(p)}}} \left\|\tilde{v}^{(s-1)}\right\|^2 + \sum_{j=s}^{p} \left\|V^{(j)}\right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix} y^{(s-1)} \\ \operatorname{Vec}(Y^{(s)}) \\ \vdots \\ \operatorname{Vec}(Y^{(p)}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i^{(s-1)} \\ \oplus_i X_i^{(s)} \\ \vdots \\ \oplus_i X_i^{(p)} \end{pmatrix} \beta + \begin{pmatrix} W_{AA}^{(s-1)} & 0 & \cdots & 0 \\ 0 & C \otimes I_{M_s} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \otimes I_{M_p} \end{pmatrix} \begin{pmatrix} \tilde{v}^{(s-1)} \\ \operatorname{Vec}(V^{(s)}) \\ \vdots \\ \operatorname{Vec}(V^{(p)}) \end{pmatrix} , \tag{5.35}$$

where $y^{(s-1)}, \tilde{v}^{(s-1)} \in \mathbb{R}^K$, and $W_{AA}^{(s-1)} \in \mathbb{R}^{K \times K}$ and $R_i^{(s-1)} \in \mathbb{R}^{k_i \times k_i}$ are upper triangular. For the solution of (5.35) consider the update GQRD (UGQRD)

$$Q_{(s)}^T \begin{pmatrix} \oplus_i R_i^{(s-1)} \\ \oplus_i X_i^{(s)} \end{pmatrix} = \begin{array}{c c} K & \\ \begin{pmatrix} \oplus_i R_i^{(s)} \\ 0 \end{pmatrix} & \begin{array}{l} K \\ GM_s \end{array} \end{array} \tag{5.36a}$$

and

$$Q_{(s)}^T \begin{pmatrix} W_{AA}^{(s-1)} & 0 \\ 0 & C \otimes I_{M_s} \end{pmatrix} P_{(s)} = W^{(s)} \equiv \begin{array}{c c} \begin{array}{cc} K & GM_s \end{array} & \\ \begin{pmatrix} W_{AA}^{(s)} & W_{AB}^{(s)} \\ 0 & W_{BB}^{(s)} \end{pmatrix} & \begin{array}{l} K \\ GM_s \end{array} \end{array} , \tag{5.36b}$$

where $R_i^{(s)}$ and $W^{(s)}$ are upper triangular and, $Q_{(s)}$ and $P_{(s)}$ are orthogonal. Let

$$Q_{(s)}^T \begin{pmatrix} y^{(s-1)} \\ \text{Vec}(Y^{(s)}) \end{pmatrix} = \begin{pmatrix} \tilde{y}^{(s)} \\ \text{Vec}(\hat{Y}^{(s)}) \end{pmatrix} \begin{array}{l} K \\ GM_s \end{array} \tag{5.37a}$$

and

$$P_{(s)}^T \begin{pmatrix} \tilde{v}^{(s-1)} \\ \text{Vec}(V^{(s)}) \end{pmatrix} = \begin{pmatrix} \tilde{v}^{(s)} \\ \text{Vec}(\hat{V}^{(s)}) \end{pmatrix} \begin{array}{l} K \\ GM_s \end{array} . \tag{5.37b}$$

Strategies for computing the UGQRD (5.36) have been discussed in the context of updating the SUR model [49].

Using (5.36) and (5.37), the GLLSP (5.35) becomes the smaller GLLSP

$$\underset{\substack{\beta, \tilde{v}^{(s)}, \\ V^{(s+1)}, \ldots, V^{(p)}}}{\text{argmin}} \left\| \tilde{v}^{(s)} \right\|^2 + \sum_{j=s+1}^{p} \left\| V^{(j)} \right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix} y^{(s)} \\ \text{Vec}(Y^{(s+1)}) \\ \vdots \\ \text{Vec}(Y^{(p)}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i^{(s)} \\ \oplus_i X_i^{(s+1)} \\ \vdots \\ \oplus_i X_i^{(p)} \end{pmatrix} \beta + \begin{pmatrix} W_{AA}^{(s)} & 0 & \cdots & 0 \\ 0 & C \otimes I_{M_{s+1}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C \otimes I_{M_p} \end{pmatrix} \begin{pmatrix} \tilde{v}^{(s)} \\ \text{Vec}(V^{(s+1)}) \\ \vdots \\ \text{Vec}(V^{(p)}) \end{pmatrix}, \tag{5.38}$$

where

$$W_{BB}^{(s)} \, \text{Vec}(\hat{V}^{(s)}) = \text{Vec}(\hat{Y}^{(s)}) \tag{5.39}$$

and

$$y^{(s)} = \tilde{y}^{(s)} - W_{AB}^{(s)} \operatorname{Vec}(\hat{V}^{(s)}). \tag{5.40}$$

At the last iteration, when $s = p$, the GLLSP reduces to

$$\underset{\beta,\tilde{v}^{(p)}}{\operatorname{argmin}} \left\|\tilde{v}^{(p)}\right\|^2 \quad \text{subject to} \quad y^{(p)} = (\oplus_i R_i^{(p)})\beta + W_{AA}^{(p)}\tilde{v}^{(p)} \,,$$

which has solution $\tilde{v}^{(p)} = 0$ and $\oplus_i R_i^{(p)}\beta = y^{(p)}$. Algorithm 9 summarizes the steps of this recursive estimation procedure for computing the BLUE of the SUR model. Note that at the sth iteration, the matrix retriangularized in (5.36b) is of order $(K + GM_s)$. This results in less computational cost than does the RQD of the $GM \times GM$ matrix in (5.14). Algorithm 9 also requires less memory to store the smaller dimensioned matrices involved in the factorizations.

Algorithm 9 Solution of the GLLSP (5.5) using the recursive algorithm.

```
Input: The regressors $X_1, \dots, X_G$ and $Y$ and the disturbances covariance matrix $\Sigma$
Output: The vector of parameters $\beta$
Compute $\Sigma = CC^T$
Compute the GQRD (5.31) and $Y^{(1)}$ from (5.32)
Solve the triangular system $W_{BB}^{(1)}\hat{v}^{(1)} = \hat{y}^{(1)}$
Compute $y^{(1)} = \tilde{y}^{(1)} - W_{AB}^{(1)}\hat{v}^{(1)}$
for $s = 2, \dots, p$ do
    Compute the UGQRD (5.36) and (5.37a)
    Solve the triangular system (5.39) for $\operatorname{Vec}(\hat{V}^{(s)})$
    Compute (5.40)
end for
Solve the triangular system $(\oplus_i R_i)\beta = y^{(p)}$ for $\beta$
```

5.4 Size reduction of large scale SUR models

When $M > k$, the SUR model can be transformed to one of smaller dimension [21, 46, 48]. Solving the transformed model results in a computationally efficient algorithm. Let $X^* = \begin{pmatrix} X_1 & \cdots & X_G \end{pmatrix} \in \mathbb{R}^{M \times K}$, $K = \sum_{i=1}^{G} k_i$, and $M > K$. Consider the QRD

$$\begin{pmatrix} Q_R^{*T} \\ Q_N^{*T} \end{pmatrix} X^* = \begin{pmatrix} R^* \\ 0 \end{pmatrix}, \tag{5.41}$$

where $R^* = \begin{pmatrix} R_1^* & \cdots & R_G^* \end{pmatrix} \in \mathbb{R}^{M\times M}$, $R_i^* \in \mathbb{R}^{K\times k_i}$, and the matrix $\begin{pmatrix} Q_R^* & Q_N^* \end{pmatrix} \in \mathbb{R}^{M\times M}$ is orthogonal. Now, premultiplying the SUR model (5.1) by $\begin{pmatrix} I_G \otimes Q_R^* & I_G \otimes Q_N^* \end{pmatrix}^T$ results in the transformed SUR (TSUR) model

$$\begin{pmatrix} \mathrm{Vec}(Y_R^*) \\ \mathrm{Vec}(Y_N^*) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i^* \\ 0 \end{pmatrix} \beta + \begin{pmatrix} \mathrm{Vec}(U_R^*) \\ \mathrm{Vec}(U_N^*) \end{pmatrix}, \tag{5.42}$$

where $Y_R^* = Q_R^{*T} Y$, $Y_N^* = Q_N^{*T} Y$, $U_R^* = Q_R^{*T} U$, and $U_N^* = Q_N^{*T} U$. Furthermore,

$$\begin{pmatrix} \mathrm{Vec}(U_R^*) \\ \mathrm{Vec}(U_N^*) \end{pmatrix} \sim \left(0, \begin{pmatrix} \Sigma \otimes I_K & 0 \\ 0 & \Sigma \otimes I_{M-K} \end{pmatrix} \right), \tag{5.43}$$

and thus the SUR model (5.1) is equivalent to the smaller TSUR model

$$\mathrm{Vec}(Y_R^*) = (\oplus_i R_i^*)\beta + \mathrm{Vec}(U_R^*). \tag{5.44}$$

Note that

$$R_i^* = \begin{pmatrix} R_{1,i}^* \\ \vdots \\ R_{i,i}^* \\ 0 \end{pmatrix} \begin{array}{l} k_1 \\ \vdots \\ k_i \\ \lambda_{i+1} \end{array}, \qquad (i = 1, \ldots, G),$$

where $\lambda_i = \sum_{j=i}^{G} k_j$. However, the direct implemention of Algorithms 6-9 to solve this reduced model does not exploit the special structure of the (transformed) exogenous matrices $R_1^*, \ldots, R_G^*$.

Let (5.41) be replaced with the QRD of $\widetilde{X} = \begin{pmatrix} X_G & \cdots & X_1 \end{pmatrix}$ and partition

$$Q_R^{*T} X_i \equiv \widetilde{R}_i = \begin{pmatrix} \widetilde{R}_{1,i} \\ \vdots \\ \widetilde{R}_{G-i+1,i} \\ 0 \end{pmatrix} \begin{array}{l} k_1 \\ \vdots \\ k_{G-i+1} \\ \lambda_{G-i+2} \end{array}, \quad Y_R^* = \begin{pmatrix} \widetilde{Y}_1 \\ \widetilde{Y}_2 \\ \vdots \\ \widetilde{Y}_G \end{pmatrix} \begin{array}{l} k_1 \\ k_2 \\ \vdots \\ k_G \end{array} \quad \text{and} \quad U_R^* = \begin{pmatrix} \widetilde{U}_1 \\ \widetilde{U}_2 \\ \vdots \\ \widetilde{U}_G \end{pmatrix} \begin{array}{l} k_1 \\ k_2 \\ \vdots \\ k_G \end{array}. \tag{5.45}$$

Also let $\widetilde{V}_i C^T = \widetilde{U}_i$, $\widetilde{Y}_i$, $\widetilde{V}_i$, and C $(i = 1, \ldots, G)$ be partitioned, respectively, as

$$\widetilde{Y}_i = \begin{array}{c} \begin{array}{cc} {\scriptstyle G-i+1} & {\scriptstyle i-1} \end{array} \\ \begin{pmatrix} \widetilde{Y}_{iA} & \widetilde{Y}_{iB} \end{pmatrix} \end{array}, \quad \widetilde{V}_i = \begin{array}{c} \begin{array}{cc} {\scriptstyle G-i+1} & {\scriptstyle i-1} \end{array} \\ \begin{pmatrix} \widetilde{V}_{iA} & \widetilde{V}_{iB} \end{pmatrix} \end{array}, \quad \text{and} \quad C = \begin{array}{cc} \begin{array}{cc} {\scriptstyle G-i+1} & {\scriptstyle i-1} \end{array} & \\ \begin{pmatrix} C_{AA}^{(i)} & C_{AB}^{(i)} \\ 0 & C_{BB}^{(i)} \end{pmatrix} & \begin{array}{l} {\scriptstyle G-i+1} \\ {\scriptstyle i-1} \end{array} \end{array}.$$

Then, the GLLSP formulation of the TSUR model (5.44) can be written as

$$\underset{\beta,\widetilde{V}_j}{\operatorname{argmin}}\sum_{j=1}^{G}\left\|\widetilde{V}_j\right\|_F^2 \text{ subject to } \operatorname{Vec}(\widetilde{Y}_i)=(\oplus_j\widetilde{R}_{j,i})\beta+\operatorname{Vec}(\widetilde{V}_iC^T), \quad i=1,\ldots,G,$$

or, equivalently, as

$$\underset{\beta,\widetilde{V}_{jA},\widetilde{V}_{jB}}{\operatorname{argmin}}\sum_{j=1}^{G}\left(\left\|\widetilde{V}_{jA}\right\|_F^2+\left\|\widetilde{V}_{jB}\right\|_F^2\right) \quad \text{subject to} \tag{5.46a}$$

$$\widetilde{Y}_{iA}=\left(\widetilde{R}_{i,1}\beta_1\ \cdots\ \widetilde{R}_{i,G-i+1}\beta_{G-i+1}\right)+\widetilde{V}_{iA}(C_{AA}^{(i)})^T+\widetilde{V}_{iB}(C_{AB}^{(i)})^T,\ \ i=1,\ldots,G, \tag{5.46b}$$

$$\widetilde{Y}_{iB}=\widetilde{V}_{iB}(C_{BB}^{(i)})^T,\ \ i=1,\ldots,G. \tag{5.46c}$$

From (5.46c), it follows that $\widetilde{V}_{iB}=\widetilde{Y}_{iB}\big(C_{BB}^{(i)}\big)^{-T}$, and thus, the GLLSP (5.46) can be written as

$$\underset{\beta,\widetilde{V}_{jA}}{\operatorname{argmin}}\sum_{j=1}^{G}\left\|\widetilde{V}_{jA}\right\|_F^2 \quad \text{subject to}$$

$$\overline{Y}_{iA}=\left(\widetilde{R}_{i,1}\beta_1\ \cdots\ \widetilde{R}_{i,G-i+1}\beta_{G-i+1}\right)+\widetilde{V}_{iA}(C_{AA}^{(i)})^T, \qquad i=1,\ldots,G, \tag{5.47}$$

where $\overline{Y}_{iA}=\widetilde{Y}_{iA}-\widetilde{V}_{iB}(C_{AB}^{(i)})^T$. This is equivalent to

$$\underset{\beta,\widetilde{V}_{iA}}{\operatorname{argmin}}\sum_{i=1}^{G}\left\|\widetilde{V}_{iA}\right\|_F^2 \quad \text{subject to}$$

$$\begin{pmatrix}\operatorname{Vec}(\overline{Y}_{1A})\\ \operatorname{Vec}(\overline{Y}_{2A})\\ \vdots\\ \operatorname{Vec}(\overline{Y}_{GA})\end{pmatrix}=\begin{pmatrix}\overline{R}_1\\ \overline{R}_2\\ \vdots\\ \overline{R}_G\end{pmatrix}\beta+\begin{pmatrix}C_{AA}^{(1)}\otimes I_{k_1} & 0 & \cdots & 0\\ 0 & C_{AA}^{(2)}\otimes I_{k_2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & C_{AA}^{(G)}\otimes I_{k_G}\end{pmatrix}\begin{pmatrix}\operatorname{Vec}(\widetilde{V}_{1A})\\ \operatorname{Vec}(\widetilde{V}_{2A})\\ \vdots\\ \operatorname{Vec}(\widetilde{V}_{GA})\end{pmatrix}, \tag{5.48}$$

where

$$\overline{R}_i=\begin{array}{c}\begin{array}{cc} K_{G-i+1} & \lambda_{G-i+2}\end{array}\\ \left(\begin{array}{cc}\oplus_{j=1}^{G-i+1}\widetilde{R}_{i,j} & 0\end{array}\right)(G-i+1)K_{G-i+1}\end{array} \tag{5.49}$$

and $K_i=\sum_{j=1}^{i}k_j$. Notice that the first block of the constraints

$$\operatorname{Vec}(\overline{Y}_{1A})=\overline{R}_1\beta+\left(C_{AA}^{(1)}\otimes I_{k_1}\right)\operatorname{Vec}(\widetilde{V}_{1A})$$

is analogous to the constraint of the GLLSP (5.5).

The GLLSP (5.48) corresponds to a GLLSP formulation of a SUR model with unequal number of observations [22]. Figure 5.5 shows the structure of $\big(\bar{R}_1^T \ \cdots \ \bar{R}_G^T\big)^T$ and $\oplus_i(C_{AA}^{(i)} \otimes I_{k_i})$, where $G = 4$. The recursive algorithm in [22] that solves the unequal-size-of-observations problem is similar to Algorithm 9 and can therefore be employed to compute the solution of (5.48).

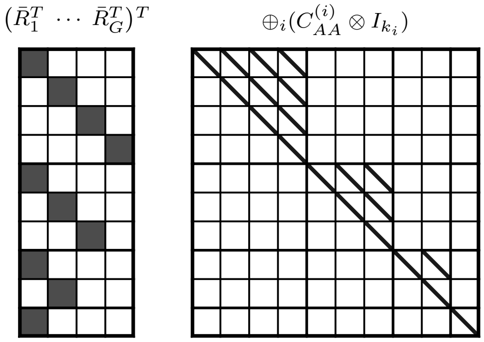


Figure 5.5: The structure of the transformed exogenous matrix $\big(\bar{R}_1^T \ \cdots \ \bar{R}_G^T\big)^T$ and Cholesky factor $\oplus_i(C_{AA}^{(i)} \otimes I_{k_i})$ of the GLLSP (5.48), where $G = 4$.

5.5 Computational comparison

The algorithms are implemented in double precision on a PC with a single 1.7GHz Intel Pentium IV processor and 512Mb of RAM. The matrix factorizations have been computed using LAPACK subroutines [1]. The diagonally-based method (see Appendix 5.A) has been used to compute the factorizations (5.13), (5.18) and (5.36b) [21]. Furthermore, in the case of the recursive Algorithm 9, the block sizes used are $M_1 = \max(k_1, \ldots, k_G)$ and $M_2 = \cdots = M_p = 10$. This is found experimentally to be the best choice for the specific architecture.

Tables 5.1 and 5.2 show the execution times (in seconds) of the algorithms. Three classes of models ($M = 51$, $M = 100$ and $M = 400$) are reported, where each regression equation is assumed to have the same number of variables and no common regressors. The elements of the exogenous matrices, the response vectors, and the Cholesky factor of the covariance matrix are randomly generated from a uniform distribution. Notice that the computational complexity of the factorization procedures does not depend on the specific values of the exogenous and covariance matrices. Thus, the performance of the algorithms is the same for matrices that have been generated using different statistical assumptions. Table 5.1 shows the performance of the algorithms when the number of equations changes, ($G = 5, 10, 15, 20$), while the number of variables in each equation remains fixed at 5. Table 5.2 shows the execution times when $G = 10$ is constant and the number of variables in each regression is k ($k = 5, 10, 15, 20, 40$); that is, $k_1 = \cdots = k_G = k$.

The execution times for solving the OLM (5.6) using the LAPACK routine DGELS and Algorithm 6 are shown in columns 3 and 4, respectively. Column 5-8 give the results for Algorithms 7-9. Specifically, the 5th column refers to the LAPACK routine DGGGLM, which solves the GLLSP (5.5) without exploiting the sparse structure of the matrices. Columns 6-8 show the execution times for Algorithms 7-9, respectively. The best times for solving the OLM (5.6) and the GLLSP (5.5) are underlined and are used to calculate the performance ratio in the last column. Computational results for the LAPACK routine DGGGLM and Algorithms 7-8 are not available (n/a) for the largest problems because the algorithms run out of memory. Analogous results for the estimation of the TSUR model (5.44) are given in Tables 5.3 and 5.4, where the execution times include the initial step of transforming the SUR model (5.1) to the TSUR model (5.44). The cost of this step is negligible compared to the overall execution time.

Table 5.5 shows the theoretical complexities in terms of floating point operations (flops) of Algorithms 6-9 and that of solving the TSUR model (5.44), where $k = k_1 = \cdots = k_G$. A detailed derivation of these complexities can be found in Appendix 5.B. The second column reports the approximate complexity of each algorithm for large values of M, G, and k. The third column shows the same complexities for large scale models, i.e., $M \gg k$. Finally, the last column gives the number of flops required by each algorithm to solve the reduced-sized model (5.44). Computations that have small marginal cost compared to the overall complexities have not been taken into account. Furthermore, the transformation (5.41) that derives the TSUR model has complexity $2G^2k^2(M - Gk/3)$ and has not been included.

From the theoretical and computational results a number of conclusions can be drawn:

- Theoretical and experimental results confirm that the OLM algorithms outperform the GLLSP algorithms. In theory the ratio between Algorithm 6 and Algorithms 7–8 is linear with k/M, while that of Algorithm 6 and 9 is constant. In practice, this performance difference decreases as the number of regressors or equations increases.

- The direct use of the standard LAPACK routine DGGGLM to solve the GLLSP is not feasible for large-scale models.

- The discrepancies between the theoretical and experimental results are due to the implementation overheads and memory usage.

Table 5.1: Execution times of solving the SUR model, where $k_1 = \cdots = k_G = 5$.

		OLM algorithms		GLLSP algorithms				Ratio
M	G	LAPACK	Alg. 6	LAPACK	Alg. 7	Alg. 8	Alg. 9	GLLSP/OLM
51	5	0.0007	0.0012	0.0968	0.0159	0.0116	0.0133	16.57
51	10	0.0070	0.0070	0.4430	0.0994	0.0776	0.0547	7.81
51	15	0.0389	0.0180	1.4195	0.2993	0.2412	0.1404	7.80
51	20	0.1274	0.0382	3.4158	0.7061	0.6279	0.2801	7.33
51	30	0.3471	0.1111	11.4602	2.4468	2.2541	0.7788	7.01
100	5	0.0014	0.0022	0.5553	0.0507	0.0409	0.0270	19.28
100	10	0.0213	0.0119	2.8841	0.3215	0.2906	0.1091	9.17
100	15	0.1135	0.0326	10.4153	1.0539	0.9871	0.2853	8.75
100	20	0.3039	0.0687	25.5424	2.6647	2.4102	0.5697	8.29
100	30	0.7402	0.2122	81.9765	9.4783	8.9455	1.5567	7.34
400	5	0.0162	0.0093	25.6388	1.2633	0.7686	0.1092	11.74
400	10	0.1675	0.0498	183.3260	8.2393	6.0076	0.4480	9.00
400	15	0.5750	0.1598	586.3929	28.0438	22.2265	1.1694	7.32
400	20	1.3203	0.4207	n/a	n/a	n/a	2.4639	5.86
400	30	n/a	1.5331	n/a	n/a	n/a	6.6609	4.34

Table 5.2: Execution times of solving the SUR model, where $G = 10$ and $k_i = 5, 10, 15, 20, 30$ for $i = 1, \ldots, G$.

		OLM algorithms		GLLSP algorithms				Ratio
M	k_i	LAPACK	Alg. 6	LAPACK	Alg. 7	Alg. 8	Alg. 9	GLLSP/OLM
51	5	0.0060	0.0069	0.4446	0.0989	0.0756	0.0556	9.26
51	10	0.0450	0.0187	0.5103	0.1383	0.1013	0.0868	4.64
51	15	0.0853	0.0334	0.5746	0.1688	0.1239	0.1327	3.71
51	20	0.0942	0.0507	0.5736	0.1887	0.1580	0.1773	3.50
51	30	0.1450	0.0898	0.7835	0.1990	0.1803	0.2610	2.91
100	5	0.0227	0.0121	2.9766	0.3384	0.2963	0.1122	9.27
100	10	0.1208	0.0327	3.1151	0.4663	0.4511	0.1897	5.80
100	15	0.1951	0.0726	3.4684	0.6214	0.5865	0.3085	4.25
100	20	0.1979	0.0989	3.6348	0.7296	0.7224	0.4416	4.46
100	30	0.3181	0.2138	4.5518	1.0656	1.0375	0.8022	3.75
400	5	0.1812	0.0499	185.3506	8.3259	6.0854	0.4527	9.07
400	10	0.6433	0.1731	191.6706	13.1758	9.2547	0.7791	4.50
400	15	1.0295	0.3869	198.3687	18.3041	13.0657	1.2969	3.35
400	20	1.1602	0.7879	204.5498	23.9272	17.4742	2.0352	2.71
400	30	1.8623	1.8808	200.3455	34.8307	25.7010	3.8604	2.07

Table 5.3: Execution times of solving the TSUR model (5.44), where $k_1 = \cdots = k_G = 5$.

		OLM algorithms		GLLSP algorithms				
M	G	LAPACK	Alg. 6	LAPACK	Alg. 7	Alg. 8	Alg. 9	GLLSP/OLM
100	5	0.0007	0.0012	0.0055	0.0049	0.0034	0.0050	4.86
100	10	0.0068	0.0079	0.3980	0.0834	0.0608	0.0463	6.81
100	15	0.0665	0.0277	4.1706	0.5426	0.5045	0.1887	6.81
100	20	0.2908	0.0729	25.1177	2.3531	2.1953	0.5194	7.21
400	5	0.0015	0.0019	0.0059	0.0056	0.0040	0.0059	3.93
400	10	0.0088	0.0110	0.3973	0.0866	0.0637	0.0496	5.64
400	15	0.0746	0.0396	4.3138	0.5676	0.5066	0.1987	5.02
400	20	0.3059	0.1010	26.2893	2.5442	2.3426	0.5566	5.51
400	30	1.1623	0.4447	79.9213	23.6311	21.8987	2.3073	5.19

Table 5.4: Execution times of solving the TSUR model (5.44), where $G = 10$ and $k_i = 5, 10, 15, 20, 40$ for $i = 1, \ldots, G$.

		OLM algorithms		GLLSP algorithms				Ratio
M	k_i	LAPACK	Alg. 6	LAPACK	Alg. 7	Alg. 8	Alg. 9	GLLSP/OLM
100	5	0.0052	0.0081	0.4015	0.0881	0.0601	0.0477	9.17
100	10	0.1088	0.0366	3.0925	0.4421	0.3805	0.1672	4.57
400	5	0.0093	0.0111	0.4249	0.0873	0.0643	0.0491	5.28
400	10	0.1236	0.0556	3.1201	0.4482	0.5142	0.1872	3.37
400	15	0.3700	0.1443	10.9318	1.2860	1.2046	0.4675	3.24
400	20	0.5482	0.3031	27.2676	3.5127	2.9855	0.9567	3.16
400	30	1.3749	1.2734	87.4721	13.9188	11.3183	2.7661	2.17

Table 5.5: Complexity of Algorithms 6-9, where $k = k_1 = \cdots = k_G$.

Algorithm	Complexity	Compl. Approx. for $M \gg k$	Compl. for solving the TSUR model
OLM Algorithms			
LAPACK	$2G^3k^2(M - k/3)$	$2G^3k^2M$	$2G^4k^3$
Alg. 1	$G^2k^2\big(M + 2G(M - k + 1)/3\big)$	$2G^3k^2M/3$	$2G^4k^3/3$
GLLSP Algorithms			
LAPACK	$G^3(4M^3/3 + 4M^2k - 2k^3/3)$	$4G^3M^3/3$	$4G^6k^3/3$
Alg. 2	$G^2kM\big(M + 2G(M - k + 1)/3\big)$	$2G^3kM^2/3$	$2G^5k^3/3$
Alg. 3	$G^2kM\big(M + G(M - k + 2)/3\big) + G^2k^2\big(M + G(M - k + 1)/3\big)$	$G^3kM^2/3$	$G^5k^3/3$
Alg. 4	$G^2k^3 + 4G^3k^2(M - k + 1/2)/3$	$4G^3k^2M/3$	$4G^4k^3/3$

- The complexity of the OLM algorithms and that of the recursive Algorithm 9 is a linear function of the sample size. It follows that, in practice, the performance of these algorithms does not deteriorate when the number of observations increases and thus they can solve large-scale problems.

- The algorithms for solving the TSUR model (5.44) outperform the corresponding algorithms for solving the initial SUR model (5.1). For the largest problems, the cost of transforming the SUR model to one of smaller dimensions and solving it is negligible compared to the cost of solving the original one.

5.6 Summary

Algorithms for solving the seemingly unrelated regressions (SUR) model have been considered. The algorithms use as a basic component the QR decomposition. Initially the SUR model is transformed to an ordinary linear model (OLM). This transformation results in a regressor matrix having a block triangular structure. The best linear unbiased estimator (BLUE) of the SUR model results from the least-squares (LS) solution of the OLM. A computationally efficient strategy (Algorithm 6) produces the LS estimator by exploiting the sparse structure of the matrices. This strategy outperforms the LAPACK DGELS subroutine, which treats the matrices as full, when the problem is not very small.

The remaining three algorithms compute the BLUE by formulating the SUR model as a generalized linear least squares problem (GLLSP). The solution of the GLLSP is obtained using the generalized QR decomposition (GQRD). The first method (Algorithm 7) computes the GQRD of the block diagonal and Kronecker structures of matrix of the exogenous variables and the Cholesky factor of the dispersion matrix, respectively. This method is computationally more efficient than the corresponding LAPACK routine (DGGGLM) that solves the general linear model. The second method (Algorithm 8) solves the GLLSP iteratively. Each iteration solves a smaller sized GLLSP. The main advantage of this method is that it avoids the formation of the computationally expensive RQ decomposition (5.14). This allows Algorithm 8 to outperform Algorithm 7. Finally, a recursive estimation strategy (Algorithm 9) is proposed. This is found to be the most efficient when the model is not very small. Furthermore, this strategy requires less memory and can thus solve larger problems. Algorithms 6, 8 and 9 are new designs, while Algorithm 7 has been originally proposed in [50].

The algorithms are reassessed after an initial orthogonal transformation is made to reduce the SUR model to one of smaller size. The matrix of exogenous variables of the transformed (TSUR) model (5.44) has dimensions $GK \times K$ compared with $GM \times K$ of the original model (5.1). This transformation is significant for large-scale models, where the number of observations in each equation is much larger than the total number of regressors, i.e., $M \gg K$. The solution for the SUR, and consequently TSUR, model when the regressions have common exogenous factors is currently under investigation. In this case, $X_i = XS_i$, where $X \in \mathbb{R}^{M \times K^d}$ is the matrix of the K^d distinct regressors, and $S_i \in \mathbb{R}^{K^d \times k_i}$ is the selection matrix comprised of the relevant columns of the $K^d \times K^d$ identity matrix. The computation of the QRD of $\widetilde{X} = X\big(S_G \ \cdots \ S_1\big)$ produces $\widetilde{R}_i$ ($i = 1, \ldots, G$) matrices in (5.45) that have a sparse structure able to be exploited by the various algorithms [26, 46, 51].

Often SUR models exhibit special properties and characteristics [25, 24, 48, 60] (see Appendix 5.C). For the efficient solution of these models the proposed algorithms need to be modified. The structures of the matrices and their properties should be exploited. Iterative algorithms for computing the estimators of models with sparse exogenous matrices merits investigation.

Although Algorithm 6 is the most computationally efficient, it is numerically less stable than Algorithms 7-9 [76]. The Algorithm 6 may provide a poor solution when C is ill-conditioned and fails when Σ is singular, i.e., when C is not of full-rank [45, 55, 61, 62]. In such cases, the GLLSP approach should be used [21, 46, 50]. The numerical stability of the algorithms needs to be investigated.

The algorithms for solving the TSUR model (5.44) can be adapted to solve simultaneous equations models (SEMs) [6, 12, 15, 52, 89]. Similarly to the SUR model (5.1), the SEM can be expressed as

$$\operatorname{Vec}(Y) = (\oplus_i W_i) \operatorname{Vec}\left(\{\beta_i\}\right) + \operatorname{Vec}(U), \tag{5.50}$$

where $W_i = \big(X_i \quad Y_i\big)$, $\beta_i \in \mathbb{R}^{k_i + g_i}$, and $Y_i \in \mathbb{R}^{M \times g_i}$ consists of g_i endogenous variables from Y, excluding y_i. The endogeneity in the SEM can be eliminated by a transformation identical to that employed to derive the TSUR model [52]. The transformed SEM can be written as

$$\operatorname{Vec}(Y_R^*) = (\oplus_i W_i^*) \operatorname{Vec}\left(\{\beta_i\}\right) + \operatorname{Vec}(U_R^*), \tag{5.51}$$

where $W_i^* = \big(R_i \quad Y_i^*\big)$, $Y_i^* = Q_R^{*T} Y_i$, and Y_R^*, R_i^* and U_R^* are defined in (5.44). Efficient algorithms for solving the SEM are currently under investigation.

5.A The column- and diagonally-based methods

Consider the updating QR decomposition (UQRD)

$$Q^T \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} \tag{5.52}$$

where $A, R \in \mathbb{R}^{k^{(G)} \times k^{(G)}}$ are upper-triangular, $B \in \mathbb{R}^{q^{(G)} \times k^{(G)}}$ is block upper-triangular and Q is orthogonal of order $k^{(G)} + q^{(G)}$. Now, let

$$A \equiv A^{(0)} = \begin{array}{cc} \begin{array}{cccc} k_1 & k_2 & \cdots & k_G \end{array} & \\ \begin{pmatrix} A_{1,1}^{(0)} & A_{1,2}^{(0)} & \cdots & A_{1,G}^{(0)} \\ 0 & A_{2,2}^{(0)} & \cdots & A_{2,G}^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{G,G}^{(0)} \end{pmatrix} & \begin{array}{c} k_1 \\ k_2 \\ \vdots \\ k_G \end{array} \end{array} ,$$

and

$$B \equiv B^{(0)} = \begin{array}{cc} \begin{array}{cccc} k_1 & k_2 & \cdots & k_G \end{array} & \\ \begin{pmatrix} B_{1,1}^{(0)} & B_{1,2}^{(0)} & \cdots & B_{1,G}^{(0)} \\ 0 & B_{2,2}^{(0)} & \cdots & B_{2,G}^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_{G,G}^{(0)} \end{pmatrix} & \begin{array}{c} q_1 \\ q_2 \\ \vdots \\ q_G \end{array} \end{array} ,$$

where $A_{i,i}$ is upper-triangular, $k^{(G)} = \sum_{i=1}^{G} k_i$ and $q^{(G)} = \sum_{i=1}^{G} q_i$. Two block strategies can be used to compute the UQRD (5.52). The first, a diagonally-based strategy, annihilates the block-superdiagonals of B one at a time. The second, a column-based strategy, annihilates the non-zero blocks of B column-by-column [21, 44].

The diagonally-based strategy computes the UQRDs

$$Q_{i,j}^T \begin{pmatrix} A_{i+j,i+j}^{(i-1)} \\ B_{j,i+j}^{(i-1)} \end{pmatrix} = \begin{pmatrix} A_{i+j,i+j}^{(i)} \\ 0 \end{pmatrix} \begin{array}{c} k_{i+j} \\ q_j \end{array} \tag{5.53}$$

and

$$Q_{i,j}^T \begin{pmatrix} A_{i+j,i+j+1:G}^{(i-1)} \\ B_{j,i+j+1:G}^{(i-1)} \end{pmatrix} = \begin{pmatrix} A_{i+j,i+j+1:G}^{(i)} \\ B_{j,i+j+1:G}^{(i)} \end{pmatrix} \begin{array}{c} k_{i+j} \\ q_j \end{array} , \tag{5.54}$$

where $i = 1, \ldots, G-1$, $j = 1, \ldots, G-i$ and

$$\begin{pmatrix} A^{(i)}_{i+j,i+j+1:G} \\ B^{(i)}_{j,i+j+1:G} \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} k_{i+j+1} & k_{i+j+2} & \cdots & k_G \end{array} \\ \begin{pmatrix} A^{(i)}_{i+j,i+j+1} & A^{(i)}_{i+j,i+j+2} & \cdots & A^{(i)}_{i+j,G} \\ B^{(i)}_{j,i+j+1} & B^{(i)}_{j,i+j+2} & \cdots & B^{(i)}_{j,G} \end{pmatrix} \end{array} \begin{array}{l} \\ k_{i+j} \\ q_j \end{array} .$$

Thus, R in (5.52) is given by

$$R = \begin{pmatrix} A^{(0)}_{1,1} & A^{(0)}_{1,2} & \cdots & A^{(0)}_{1,G} \\ 0 & A^{(1)}_{2,2} & \cdots & A^{(1)}_{2,G} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A^{(G-1)}_{G,G} \end{pmatrix} .$$

Notice that the UQRDs (5.53) and (5.54) can be computed simultaneously for $j = 1, \ldots, G-i$.

The column-based strategy computes the QRDs

$$\widetilde{Q}_i^T \begin{pmatrix} A^{(i-2)}_{i,i} \\ B^{(i-2)}_{1:i-1,i} \end{pmatrix} = \begin{pmatrix} A^{(i-1)}_{i,i} \\ 0 \end{pmatrix} \begin{array}{l} k_i \\ q(i-1) \end{array}$$

and

$$\widetilde{Q}_i^T \begin{pmatrix} A^{(i-2)}_{i,i+1:G} \\ B^{(i-2)}_{1:i-1,i+1:G} \end{pmatrix} = \begin{pmatrix} A^{(i-1)}_{i,i+1:G} \\ B^{(i-1)}_{1:i-1,i+1:G} \end{pmatrix} \begin{array}{l} k_i \\ q^{(i-1)} \end{array} ,$$

where $i = 2, \ldots, G$ and $q^{(i-1)} = \sum_{j=1}^{i-1} q_j$. Figure 5.6 shows the annihilation patterns of the two strategies in the case where $G = 4$.

Now, the computation of (5.18) and part of (5.36b) are equivalent to

$$Q^T \begin{pmatrix} C \\ D \end{pmatrix} = \begin{array}{c} p^{(G)} \\ \begin{pmatrix} \widetilde{C} \\ \widetilde{D} \end{pmatrix} \end{array} \begin{array}{l} \\ k^{(G)} \\ q^{(G)} \end{array} , \tag{5.55}$$

where

$$C \equiv C^{(0)} = \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & \cdots & p_G \end{array} \\ \begin{pmatrix} C^{(0)}_{1,1} & C^{(0)}_{1,2} & \cdots & C^{(0)}_{1,G} \\ 0 & C^{(0)}_{2,2} & \cdots & C^{(0)}_{2,G} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C^{(0)}_{G,G} \end{pmatrix} \end{array} \begin{array}{l} \\ k_1 \\ k_2 \\ \vdots \\ k_G \end{array} , \quad D \equiv D^{(0)} = \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & \cdots & p_G \end{array} \\ \begin{pmatrix} D^{(0)}_{1,1} & D^{(0)}_{1,2} & \cdots & D^{(0)}_{1,G} \\ 0 & D^{(0)}_{2,2} & \cdots & D^{(0)}_{2,G} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D^{(0)}_{G,G} \end{pmatrix} \end{array} \begin{array}{l} \\ q_1 \\ q_2 \\ \vdots \\ q_G \end{array}$$

Diagonally-based strategy

$$\begin{array}{cccc}
\text{Initial matrix} & \text{after Stage 1} & \text{after Stage 2} & \text{after Stage 3} \\
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(0)} & A_{2,3}^{(0)} & A_{2,4}^{(0)} \\
0 & 0 & A_{3,3}^{(0)} & A_{3,4}^{(0)} \\
0 & 0 & 0 & A_{4,4}^{(0)} \\ \hline
0 & B_{1,2}^{(0)} & B_{1,3}^{(0)} & B_{1,4}^{(0)} \\
0 & 0 & B_{2,3}^{(0)} & B_{2,4}^{(0)} \\
0 & 0 & 0 & B_{3,4}^{(0)} \\
0 & 0 & 0 & 0
\end{array}\right) &
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(1)} & A_{2,3}^{(1)} & A_{2,4}^{(1)} \\
0 & 0 & A_{3,3}^{(1)} & A_{3,4}^{(1)} \\
0 & 0 & 0 & A_{4,4}^{(1)} \\ \hline
0 & 0 & B_{1,3}^{(1)} & B_{1,4}^{(1)} \\
0 & 0 & 0 & B_{2,4}^{(1)} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right) &
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(1)} & A_{2,3}^{(1)} & A_{2,4}^{(1)} \\
0 & 0 & A_{3,3}^{(2)} & A_{3,4}^{(2)} \\
0 & 0 & 0 & A_{4,4}^{(2)} \\ \hline
0 & 0 & 0 & B_{1,4}^{(2)} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right) &
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(1)} & A_{2,3}^{(1)} & A_{2,4}^{(1)} \\
0 & 0 & A_{3,3}^{(2)} & A_{3,4}^{(2)} \\
0 & 0 & 0 & A_{4,4}^{(3)} \\ \hline
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right)
\end{array}$$

Column-based strategy

$$\begin{array}{cccc}
\text{Initial matrix} & \text{after Stage 1} & \text{after Stage 2} & \text{after Stage 3} \\
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(0)} & A_{2,3}^{(0)} & A_{2,4}^{(0)} \\
0 & 0 & A_{3,3}^{(0)} & A_{3,4}^{(0)} \\
0 & 0 & 0 & A_{4,4}^{(0)} \\ \hline
0 & B_{1,2}^{(0)} & B_{1,3}^{(0)} & B_{1,4}^{(0)} \\
0 & 0 & B_{2,3}^{(0)} & B_{2,4}^{(0)} \\
0 & 0 & 0 & B_{3,4}^{(0)} \\
0 & 0 & 0 & 0
\end{array}\right) &
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(1)} & A_{2,3}^{(1)} & A_{2,4}^{(1)} \\
0 & 0 & A_{3,3}^{(0)} & A_{3,4}^{(0)} \\
0 & 0 & 0 & A_{4,4}^{(0)} \\ \hline
0 & 0 & B_{1,3}^{(1)} & B_{1,4}^{(1)} \\
0 & 0 & B_{2,3}^{(0)} & B_{2,4}^{(0)} \\
0 & 0 & 0 & B_{3,4}^{(0)} \\
0 & 0 & 0 & 0
\end{array}\right) &
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(1)} & A_{2,3}^{(1)} & A_{2,4}^{(1)} \\
0 & 0 & A_{3,3}^{(2)} & A_{3,4}^{(2)} \\
0 & 0 & 0 & A_{4,4}^{(0)} \\ \hline
0 & 0 & 0 & B_{1,4}^{(2)} \\
0 & 0 & 0 & B_{2,4}^{(2)} \\
0 & 0 & 0 & B_{3,4}^{(0)} \\
0 & 0 & 0 & 0
\end{array}\right) &
\left(\begin{array}{cccc}
A_{1,1}^{(0)} & A_{1,2}^{(0)} & A_{1,3}^{(0)} & A_{1,4}^{(0)} \\
0 & A_{2,2}^{(1)} & A_{2,3}^{(1)} & A_{2,4}^{(1)} \\
0 & 0 & A_{3,3}^{(2)} & A_{3,4}^{(2)} \\
0 & 0 & 0 & A_{4,4}^{(3)} \\ \hline
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{array}\right)
\end{array}$$

Figure 5.6: Computation of the UQRD (5.13a) using the diagonally- and column-based strategies, $G = 4$.

and $p^{(G)} = \sum_{i=1}^{G} p_i$. If the diagonally-based strategy is used, then (5.55) is equivalent to computing

$$Q_{i,j}^T \begin{pmatrix} C_{i+j,j}^{(i-1)} & C_{i+j,j+1}^{(i-1)} & \cdots & C_{i+j,G}^{(i-1)} \\ D_{j,j}^{(i-1)} & D_{j,j+1}^{(i-1)} & \cdots & D_{j,G}^{(i-1)} \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} p_j & p_{j+1} & \cdots & p_G \end{array} \\ \begin{pmatrix} C_{i+j,j}^{(i)} & C_{i+j,j+1}^{(i)} & \cdots & C_{i+j,G}^{(i)} \\ D_{j,j}^{(i)} & D_{j,j+1}^{(i)} & \cdots & D_{j,G}^{(i)} \end{pmatrix} \end{array} \begin{array}{l} k_{i+j} \\ q_j \end{array} , \tag{5.56}$$

where $i = 1, \ldots, G-1$ and $j = 1, \ldots, G-i$. Notice that, the upper triangular structure of $D^{(0)}$ is preserved throughout the computation.

Now, if a column-based algorithm is used, then (5.55) is computed as

$$\widetilde{Q}_i^T \begin{pmatrix} C_{i,1}^{(i-2)} & C_{i,2}^{(i-2)} & \cdots & C_{i,2}^{(i-2)} \\ D_{1:i-1,1}^{(i-2)} & D_{1:i-1,2}^{(i-2)} & \cdots & D_{1:i-1,G}^{(i-2)} \end{pmatrix} = \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & \cdots & p_G \end{array} \\ \begin{pmatrix} C_{i,1}^{(i-1)} & C_{i,2}^{(i-1)} & \cdots & C_{i,2}^{(i-1)} \\ D_{1:i-1,1}^{(i-1)} & D_{1:i-1,2}^{(i-1)} & \cdots & D_{1:i-1,G}^{(i-1)} \end{pmatrix} \end{array} \begin{array}{l} k_i \\ q^{(i-1)} \end{array} .$$

Using this strategy, the block upper-triangular structure of $D^{(0)}$ is destroyed. This can be avoided if at the ith step the blocks of $B_{1:i-1,j}^{(i-2)}$ are annihilated one at a time and from bottom to top.

5.B Complexity analysis

The theoretical complexities of the algorithms in terms of number of flops (floating point operations) are derived in line with [28]. Initially the computational costs of the main factorizations are calculated. These are then used to determine the complexity of Algorithms 6–9. For simplicity the complexities are approximated for large values of G, M and k, where it is assumed that $k = k_1 = \cdots = k_G$.

5.B.1 Main factorizations

The number of flops required to compute the Cholesky factorization of an $n \times n$ symmetric and positive definite matrix is given by $n^3/3$ [28]. The complexities of computing the QRD of an $m \times n$ matrix using Householder transformations and that of applying the same orthogonal transformation to an m–element vector are given by $T_{QR}^1(m,n) = 2n^2(m - n/3)$ and $T_{QR}^2(m,n) = 2n(2m - n + 1)$, respectively [28, pages 224-225]. Analogously, the complexity of computing the UQRD

$$Q^T \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} \begin{array}{l} n \\ m-n \end{array} , \tag{5.57}$$

is $T^1_{UQR}(m,n) = 2n^2(m-n+1)$, where $Q \in \mathbb{R}^{m\times m}$ is orthogonal, R and A are upper triangular of order n and $B \in \mathbb{R}^{(m-n)\times n}$. The flops required to apply Q^T to a vector are $T^2_{UQR}(m,n) = 4n(m-n+1)$. Notice that the RQD and URQD have the same complexities as those of the QRD and UQRD, respectively.

Now, consider the computation of the UQRD (5.52) and (5.55) using the diagonally-based strategy. To simplify the analysis, let assume that $k_i = k$, $p_i = p$ and $q_i = q$, for $i = 1, \ldots, G$. Thus, the flops required to compute the UQRD (5.53), (5.54) and (5.56) are given, respectively, by

$$T^1_{UQR}(k,p) = 2k^2(p+1),$$
$$((G-i-j)k+1)\; T^2_{UQR}(k,p) = 4(1+(G-i-j)k)k(p+1)$$

and

$$(G-i-j+1)q\; T^2_{UQR}(k,p) = 4(G-i-j+1)k(p+1)q.$$

The complexities of the diagonally-based methods are given by

$$\begin{aligned} T_{diag}(G,k,p,q) &= \sum_{i=1}^{G-1}\sum_{j=1}^{G-i}((G-i-j)k+1+(G-i-j+1)q)T^2_{UQR}(k,p) + T^1_{UQR}(k,p) \\ &= G(G+1)k(p+1)\big(2(k+q)(G+1)-3k+6\big)/3 \\ &\simeq 2G(G+1)^2k(p+1)(k+q)/3. \end{aligned}$$

Now, if the first block diagonal of $B^{(0)}$ is already zero, then the latter becomes

$$\begin{aligned} T^*_{diag}(G,k,p,q) &= \sum_{i=2}^{G-1}\sum_{j=1}^{G-i}((G-i-j)k+1+(G-i-j+1)q)T^2_{UQR}(k,p) + T^1_{UQR}(k,p) \\ &= (G-1)(G-2)k(p+1)\big(2G(k+q)-3k+6\big)/3 \\ &\simeq 2G(G-1)(G-2)k(p+1)(k+q)/3. \end{aligned}$$

5.B.2 Algorithm 1

The complexity of Algorithm 6 is dominated by that of steps 7 and 8. Specifically, the complexity of step 7 is given by

$$\sum_{i=1}^{G-1} 2(G-i)kT^2_{QR}(M,k) = G(G-1)k^2(2M-k+1)$$

and that of step 8, i.e., computing the UQRD (5.13a) and (5.13b), by

$$\begin{aligned} T_{diag}(G-1,k,M-k,0) &= (G-1)(G-2)k(M-k+1)\big((2G-5)k+6\big)/3 \\ &\simeq 2(G-1)(G-2)(G-\frac{5}{2})k^2(M-k+1)/3. \end{aligned}$$

Thus, the number of flops required by Algorithm 6 is approximately

$$\begin{aligned} T_1(G,M,k) &\simeq (G-1)k^2\big(GM+2(M-k+1)(G^2-3G+5)/3\big) \\ &\simeq G^2k^2\big(M+2G(M-k+1)/3\big). \end{aligned}$$

Table 5.6 reports the complexity of each step of Algorithm 6.

Table 5.6: Complexity of each step of Algorithm 6

Step	Complexity	Step	Complexity
1	$G^3/3$	7	$G(G-1)k^2(2M-k+1)$
2	MG^2	8	$2(G-1)(G-2)(G-5/2)k^2(M-k+1)/3$
4	$2k^2(M-k/3)$	9	G^2k^2
5	$2k(2M-k+1)$		

5.B.3 Algorithm 2

The complexity of Algorithm 7 is approximately that of steps 6–8. The flops required by step 6 are

$$G(G-1)M\ T_{QR}^2(M,k)/2 = G(G-1)kM(2M-k+1).$$

Notice that steps 7 and 8 can be computed using an adaptation of the diagonally-based algorithm. Furthermore, since the diagonal blocks of $\widetilde{W}_{BA}^{(0)}$ are zero, the flops of these steps are

$$\begin{aligned} T_{diag}^*(G,k,M-k,M-k) &= G(G-1)k(M-k+1)(2GM-3k+6)/3 \\ &\simeq 2G^2(G-1)k(M-k+1)M/3. \end{aligned}$$

Therefore, the complexity of Algorithm 6 is

$$\begin{aligned} T_2(G,M,k) &\simeq G(G-1)kM\big(M+2(G+3/2)(M-k+1)/3\big) \\ &\simeq G^2kM\big(M+2G(M-k+1)/3\big). \end{aligned}$$

5.B.4 Algorithm 3

The complexity of the interleaving Algorithm 8 is determined by that of steps 6, 7 and 9. Step 6 applies Q_i^T from the left of an $M \times ks$ matrix and Q_i from the right of an $M(G-s-1) \times M$ matrix. The complexity of this step is thus

$$T_{QR}^2(M,k)(ks + M(G-s-1)) = 2k(2M-k+1)(M(G-1) - s(M-k)).$$

For all the iterations $s = 0, \ldots, G-1$ the complexity is evaluated to

$$\begin{aligned}\sum_{s=0}^{G-1} 2k(2M-k+1)(M(G-1) - s(M-k)) &= G(G-1)k(M+k)(2M-k+1)\\ &\simeq G^2k(M+k)(2M-k+1).\end{aligned}$$

Now, the complexity of step 7 is given by that of the URQDs (5.28) and of (5.25b), i.e.,

$$\begin{aligned}&\sum_{i=1}^{s} \big(T_{UQR}^1(M,k) + (k(s-i+1) + M(G-s+1))T_{UQR}^2(M,k)\big)\\ &\qquad = 2k(M-k+1)\big(ks(s+4) + 2Ms(G-s+1)\big).\end{aligned}$$

Therefore, for all the iterations $s = 0, \ldots, G-1$ the complexity becomes

$$\begin{aligned}&\sum_{s=0}^{G-1} 2k(M-k+1)\big(ks(s+4) + 2Ms(G-s+1)\big)\\ &\qquad = G(G+1)k(M-k+1)\big(G(k+M) + 13k + 4M\big)/3\\ &\qquad \simeq G^3k(M+k)(M-k+1)/3.\end{aligned}$$

Step 9 consists of multiplying an $M(G-s-1) \times sk$ matrix with a vector using $2M(G-s-1)sk$ flops. For all the iterations $s = 0, \ldots, G-1$ the number of flops required is

$$\sum_{s=0}^{G-1} 2Mks(G-s-1) = G(G-1)(G-2)kM/3 \simeq G^3kM/3.$$

Thus, the total complexity of steps 6, 7 and 9, and thus, of Algorithm 8, is given by

$$T_3(G,M,k) \simeq G^2kM\big(M + G(M-k+2)/3\big) + G^2k^2\big(M + G(M-k+1)/3\big).$$

5.B.5 Algorithm 4

For the complexity analysis of Algorithm 9 it is assumed that $M_1 = k$ and that $M_s = \alpha = (M-k)/(p-1)$, $s = 2, \ldots, p$. Under these assumptions the complexity of step 2 is approximately $T_2(G,k,k) = G^2k^3 + 2/3G^3k^2$. Now, the complexities of computing the UQRD (5.36a), the tranformation

$$Q_{(s)}^T \begin{pmatrix} W_{AA}^{(s-1)} & 0 \\ 0 & C \otimes I_{M_s} \end{pmatrix}$$

and the RQD (5.36b) are, respectively,

$$\begin{aligned} G\, T_{UQR}^1(k+\alpha,k) &= 2Gk^2(\alpha+1), \\ G(G-1)(k+\alpha)T_{UQR}^2(k+\alpha,k)/2 &= 2G(G-1)k(\alpha+1)(k+\alpha) \\ &\simeq 2G^2k(\alpha+1)(k+\alpha) \end{aligned}$$

and

$$\begin{aligned} T_{diag}(G,k,\alpha,\alpha) &= G(G+1)k(\alpha+1)(2(k+\alpha)(G+1)/3-k+2) \\ &\simeq 2G^3k(\alpha+1)(k+\alpha)/3. \end{aligned}$$

It follows that the complexity of step 6 is dominated by that of computing the RQD (5.36b), i.e. $T_{diag}(G,k,\alpha,\alpha)$.

Finally the complexities of steps 7 and 8 are, respectively, $\alpha^3/3$ and $2G^2k\alpha$, which are marginal respect to that of step 6. Thus, the complexity of Algorithm 9 is given by

$$T_4(G,M,k,\alpha) \simeq G^2k^3 + 2G^3k^2/3 + 2G^3k(\alpha+1)(k+\alpha)(M-k)/(3\alpha).$$

For $\alpha = 1$, this reduces to

$$T_4(G,M,k,1) \simeq G^2k^3 + 2G^3k^2\big(2M-2k+1\big)/3.$$

Notice that, if $M \gg k$, then $T_4(G,M,k) = 4G^3k^2(k+1)M/3$, which is a linear function of the sample size.

5.C A computationally efficient method for solving SUR models with orthogonal regressors

---

Abstract:

A computationally efficient method to estimate seemingly unrelated regression equations models with orthogonal regressors is presented. The method considers the estimation problem as a generalized linear least squares problem (GLLSP). The basic tool for solving the GLLSP is the generalized QR decomposition of the block-diagonal exogenous matrix and Cholesky factor $C \otimes I_T$ of the covariance matrix of the disturbances. Exploiting the orthogonality property of the regressors the estimation problem is reduced into smaller and independent GLLSPs. The solution of each of the smaller GLLSPs is obtained by a single-column modification of C. This reduces significantly the computational burden of the standard estimation procedure, especially when the iterative feasible estimator of the model is needed. The covariance matrix of the estimators is also derived.

5.C.1 Introduction

Consider the seemingly unrelated regressions (SUR) model in the compact form

$$\operatorname{Vec}(Y) = \left(\oplus_{i=1}^{G} X_i\right) \operatorname{Vec}\left(\{\beta_i\}_G\right) + \operatorname{Vec}(U), \tag{5.58}$$

where $Y = \left(y_1 \dots y_G\right)$ and $U = \left(u_1 \dots u_G\right)$ are the $T \times G$ matrices of the endogenous and disturbance vectors, respectively, $\oplus_{i=1}^{G} X_i$ is equivalent to the exogenous block diagonal matrix $\operatorname{diag}(X_1, X_2, \dots, X_G)$, $X_i \in \mathbb{R}^{T \times k_i}$ has full column rank, $\{\beta_i\}_G$ denotes the set of coefficient vectors $\beta_1, \dots, \beta_G$ and $\operatorname{Vec}(\cdot)$ is the vector operator that stacks the columns of a matrix or set of vectors [3, 46, 69]. The disturbance vector $\operatorname{Vec}(U)$ has zero mean and dispersion matrix $\Sigma \otimes I_T$, where $\Sigma = [\sigma_{ij}]$ is positive definite and $\otimes$ denotes the Kronecker product operator. For notational convenience the subscript G in the set operator $\{\cdot\}$ is dropped and $\oplus_{i=1}^{G}$ will be abbreviated to $\oplus_i$.

The best linear unbiased estimator (BLUE) of the SUR model derives from the solution of the Generalized linear least squares problem (GLLSP)

$$\underset{V,\{\beta_i\}}{\operatorname{argmin}} \left\|V\right\|_F \text{ subject to } \operatorname{Vec}(Y) = (\oplus_i X_i) \operatorname{Vec}\left(\{\beta_i\}\right) + (C \otimes I_T) \operatorname{Vec}(V), \tag{5.59}$$

where $\Sigma = CC^T$, $U = VC^T$ and $\|\cdot\|_F$ demotes the Frobenius norm [46, 61, 63]. Although this formulation allows for singular Σ, without loss of generality, it will be assumed that Σ is non-singular and C is its upper triangular Cholesky factor. The solution of (5.59) can be obtained using the generalized QR decomposition (GQRD):

$$Q^T(\oplus_i X_i) = \begin{array}{cl} \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} & \begin{array}{l} K \\ GT-K \end{array} \end{array} \tag{5.60a}$$

and

$$Q^T(C \otimes I_T)P = \begin{array}{cc} \begin{array}{cc} K & GT-K \end{array} & \\ \begin{pmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{pmatrix} & \begin{array}{l} K \\ GT-K \end{array} \end{array}, \tag{5.60b}$$

where R_i, W_{11} and W_{22} are upper triangular, $Q \in \Re^{GT \times GT}$ and $P \in \Re^{GT \times GT}$ are orthogonal and $K = \sum_{i=1}^{G} k_i$ [63]. The orthogonal matrix Q is defined as

$$Q = \begin{pmatrix} \oplus_i Q_i & \oplus_i \widehat{Q}_i \end{pmatrix}, \tag{5.61}$$

where

$$X_i = \begin{pmatrix} Q_i & \widehat{Q}_i \end{pmatrix} \begin{array}{cl} \begin{pmatrix} R_i \\ 0 \end{pmatrix} & \begin{array}{l} k_i \\ T-k_i \end{array} \end{array} \tag{5.62}$$

is the QR decomposition (QRD) of X_i [46, pp. 117-123].

Using the GQRD (5.60) the GLLSP (5.59) can be written as

$$\underset{\{\tilde{v}_i\},\{\hat{v}_i\},\{\beta_i\}}{\text{argmin}} \sum_{i=1}^{G} \left(\|\tilde{v}_i\| + \|\hat{v}_i\|\right) \text{ subject to}$$

$$\begin{pmatrix} \text{Vec}(\{\tilde{y}_i\}) \\ \text{Vec}(\{\hat{y}_i\}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i \\ 0 \end{pmatrix} \text{Vec}(\{\beta_i\}) + \begin{pmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{pmatrix} \begin{pmatrix} \text{Vec}(\{\tilde{v}_i\}) \\ \text{Vec}(\{\hat{v}_i\}) \end{pmatrix},$$

where

$$P^T \text{Vec}(V) = \begin{pmatrix} \text{Vec}(\{\tilde{v}_i\}) \\ \text{Vec}(\{\hat{v}_i\}) \end{pmatrix},$$

$Q_i^T y = \tilde{y}_i$, $\widehat{Q}_i^T y = \hat{y}_i$, $\tilde{y}_i, \tilde{v}_i \in \Re^{k_i}$, $\hat{y}_i, \hat{v}_i \in \Re^{T-k_i}$ $(i = 1, \ldots, G)$ and $\| \cdot \|$ is the Euclidean norm. Thus, $\mathrm{Vec}\,(\{\tilde{v}_i\}) = 0$, and the solution of (5.59) comes from solving the triangular system

$$\begin{pmatrix} \mathrm{Vec}\,(\{\tilde{y}_i\}) \\ \mathrm{Vec}\,(\{\hat{y}_i\}) \end{pmatrix} = \begin{pmatrix} \oplus_i R_i & W_{12} \\ 0 & W_{22} \end{pmatrix} \begin{pmatrix} \mathrm{Vec}\,(\{\beta_i\}) \\ \mathrm{Vec}\,(\{\hat{v}_i\}) \end{pmatrix}.$$

In practice, the covariance matrix of the SUR model is typically unknown. Here, an iterative procedure is used to derive a feasible estimator of the coefficients. Given an initial consistent estimator of Σ, an estimator of $\mathrm{Vec}\,(\{\beta_i\})$ is computed, and from the residual of the estimated coefficients, another estimator of Σ is derived. This procedure is repeated until convergence [86]. Thus, (5.60a) is computed only once, and the computational cost of the iterative estimation procedure is dominated by (5.60b), which needs to be computed at each iteration for different estimators of C.

It has previously been shown how to reduce the computational burden of the iterative estimation procedures when the SUR model has common or proper subset regressors [26, 46, 48]. Often inferences and comparisons about the SUR estimators are made under the assumption of orthogonal regressors (hereafter abbreviated to SUR-OR) [70, 79, 87]. The SUR-OR has the condition that the regressors in any two equations obey

$$X_i^T X_j = 0, \qquad i, j = 1, \ldots, G \text{ and } i \neq j. \tag{5.63}$$

Giles and Srivastava have investigated the efficiency of the BLUE of the SUR model in various special cases (e.g. variance inequalities and positivity of correlations constraints, restrictions on the parameters, and heteroscedastic disturbances) by studying the corresponding BLUE in the SUR-OR model [79, pp. 22, 87-104, 137-139, 245-251].

The purpose of this work is to investigate numerically stable and computationally efficient methods for solving SUR-OR. The proposed estimation procedure exploits this orthogonality property and computes the BLUE by solving G independent GLLSPs. In solving the smaller GLLSPs, a single-column modification of the triangular $C \in \Re^{G \times G}$ is performed. Thus, the costly computation of the RQ decomposition (RQD) of the $GT \times GT$ matrix $Q^T(C \otimes I_T)$ in (5.60b) is avoided.

In the next Subsection the solution of the SUR-OR model using the GLLSP approach is considered, and it is shown how the problem can be reduced to G independent smaller GLLSPs. In Section 5.C.2 an efficient method to compute the GQRDs by exploiting the structure of the GLLSPs is presented. The advantages of this method for computing the iterative feasible estimator are discussed. The computation of the covariance matrix of the estimator is derived in Section 5.C.4. Finally, conclusions and future work are presented in Section 5.C.5.

5.C.2 The SUR model with orthogonal regressors

Consider now the solution of the GLLSP (5.59) for deriving the BLUE of the SUR-OR model, that is, the SUR model (5.58) with orthogonal regressors (5.63). From the QRD (5.62) and property (5.63) it follows that

$$Q_j^T Q_i = 0, \qquad i \neq j \tag{5.64}$$

and

$$Q_j^T X_i = 0, \qquad i \neq j. \tag{5.65}$$

Notice that the submatrix Q_i of the orthogonal matrix Q in the QRD (5.62) that corresponds to the range space of X_i, is unique, apart for the sign of its elements, while $\widehat{Q}_i$ – the submatrix of the orthogonal matrix that corresponds to the null space of X_i – is not. Considering this and using (5.65) the submatrix $\widehat{Q}_i$ can be defined as

$$\widehat{Q}_i = \begin{pmatrix} Q_0 & Q_1 & \dots & Q_{i-1} & Q_{i+1} & \dots & Q_G \end{pmatrix}, \tag{5.66}$$

where $Q_0 \in \Re^{k_0}$ is the orthogonal complement of the matrix $\begin{pmatrix} Q_1 & Q_2 & \dots & Q_G \end{pmatrix}$ and $k_0 = T - K$. This implies that (5.64) and (5.65) hold for Q_0 ($j = 0$).

Multiplying the constraints of the GLLSP (5.59) from the right by the transposed of the orthogonal matrix

$$Q = \begin{pmatrix} I_G \otimes Q_0 & I_G \otimes Q_1 & \dots & I_G \otimes Q_G \end{pmatrix}$$

gives

$$\operatorname*{argmin}_{\widetilde{V},\{\beta_i\}} \|\widetilde{V}\|_F \quad \text{subject to}$$

$$\operatorname{Vec}(\widetilde{Y}) = Q^T(\oplus_i X_i)\operatorname{Vec}(\{\beta_i\}) + Q^T(C \otimes I_T)Q\operatorname{Vec}(\widetilde{V}), \tag{5.67}$$

where $\widetilde{Y} = \begin{pmatrix} \widetilde{Y}_0 & \widetilde{Y}_1 & \cdots & \widetilde{Y}_G \end{pmatrix}$, $\widetilde{V} = \begin{pmatrix} \widetilde{V}_0 & \widetilde{V}_1 & \cdots & \widetilde{V}_G \end{pmatrix}$, $\widetilde{Y}_i = Q_i^T Y$ and $\widetilde{V}_i = Q_i^T V$ for $i = 0, 1, \dots, G$. The constraints in (5.67) can be written as the set of $G + 1$ constraints

$$\begin{aligned}\operatorname{Vec}(\widetilde{Y}_i) =& (I_G \otimes Q_i^T)(\oplus_j X_j)\operatorname{Vec}(\{\beta_j\}) \\ &+ \sum_{j=1}^{G}(I_G \otimes Q_i^T)(C \otimes I_T)(I_G \otimes Q_j)\operatorname{Vec}(\widetilde{V}_j), \quad i = 0, 1, \dots, G.\end{aligned} \tag{5.68}$$

However, since for $i = 1, \ldots, G$

$$(I_G \otimes Q_i^T)(\oplus_j X_j)\,\mathrm{Vec}(\{\beta_j\}) = \mathrm{Vec}(\{Q_i^T X_j \beta_j\}) = \begin{pmatrix} 0 \\ R_i \\ 0 \end{pmatrix} \beta_i,$$

$$(I_G \otimes Q_0^T)(\oplus_j X_j)\,\mathrm{Vec}(\{\beta_j\}) = \mathrm{Vec}(\{Q_0^T X_j \beta_j\}) = 0$$

and

$$(I_G \otimes Q_i^T)(C \otimes I_T)(I_G \otimes Q_j) = (C \otimes Q_i^T Q_j) = \begin{cases} 0 & \text{if } i \neq j, \\ C \otimes I_{k_i} & \text{if } i = j, \end{cases}$$

the constraints (5.68) can be written as

$$\mathrm{Vec}(\widetilde{Y}_i) = \begin{pmatrix} 0 \\ R_i \\ 0 \end{pmatrix} \beta_i + (C \otimes I_{k_i})\,\mathrm{Vec}(\widetilde{V}_i), \quad \text{for } i = 1, \ldots, G, \tag{5.69a}$$

and

$$\mathrm{Vec}(\widetilde{Y}_0) = (C \otimes I_{T-K})\,\mathrm{Vec}(\widetilde{V}_0). \tag{5.69b}$$

Furthermore, the constraints (5.69) are equivalent to

$$\widetilde{Y}_0 = \widetilde{V}_0 C^T, \tag{5.70a}$$

and

$$\widetilde{Y}_i = \begin{pmatrix} 0 & R_i\beta_i & 0 \end{pmatrix} + \widetilde{V}_i C^T, \quad \text{for } i = 1, \ldots G. \tag{5.70b}$$

The solution of the triangular system (5.70a) is not required as it does not provide any useful information. Thus Q_0 in (5.66) need not be determined. The constraints (5.70b) are structurally and statistically unrelated. Therefore, from $\|\widetilde{V}\|_F^2 = \sum_{i=0}^{G} \|\widetilde{V}_i\|_F^2$, it follows that the GLLSP (5.67) can be reduced into the G smaller and independent GLLSPs

$$\operatorname*{argmin}_{\widetilde{V}_i, \beta_i} \left\|\widetilde{V}_i\right\|_F \quad \text{subject to} \quad \widetilde{Y}_i = \begin{pmatrix} 0 & R_i\beta_i & 0 \end{pmatrix} + \widetilde{V}_i C^T, \quad \text{for } i = 1, \ldots G.$$

5.C.3 The solution of the ith GLLSP

Consider the solution of the ith ($i = 1, \ldots, G$) GLLSP

$$\min_{\widetilde{V}_i, \beta_i} \|\widetilde{V}_i\|_F \quad \text{subject to} \quad \widetilde{Y}_i = \begin{pmatrix} 0 & R_i\beta_i & 0 \end{pmatrix} + \widetilde{V}_i C^T. \tag{5.71}$$

Let $\widetilde{Y}_i$, $\widetilde{V}_i$ and C be partitioned, respectively, as

$$\widetilde{Y}_i = \begin{array}{c} \begin{array}{ccc} \scriptstyle i-1 & \scriptstyle 1 & \scriptstyle G-i \end{array} \\ \begin{pmatrix} \widetilde{Y}_A^{(i)} & \widetilde{y}_i^{(i)} & \widetilde{Y}_B^{(i)} \end{pmatrix} \end{array}, \qquad \widetilde{V}_i = \begin{array}{c} \begin{array}{ccc} \scriptstyle i-1 & \scriptstyle 1 & \scriptstyle G-i \end{array} \\ \begin{pmatrix} \widetilde{V}_A^{(i)} & \widetilde{v}_i^{(i)} & \widetilde{V}_B^{(i)} \end{pmatrix} \end{array},$$

and

$$C = \begin{array}{c} \begin{array}{ccc} \scriptstyle i-1 & \scriptstyle 1 & \scriptstyle G-i \end{array} \\ \begin{pmatrix} C_{AA} & \xi & C_{AB} \\ 0^T & c_{ii} & \eta^T \\ 0 & 0 & C_{BB} \end{pmatrix} \end{array} \begin{array}{l} \scriptstyle i-1 \\ \scriptstyle 1 \\ \scriptstyle G-i \end{array} .$$

Furthermore, let $W^{(i)} = G_{i-1} \ldots G_1$, where G_j ($j = 1, \ldots, i-1$) denotes a Givens rotation that annihilates c_{ji}, i.e., ξ_j, when applied from the right of C. The rotation G_j affects only the ith and jth columns of C. That is,

$$\widehat{C}^{(i)} = CW^{(i)} = \begin{pmatrix} \widehat{C}_{AA}^{(i)} & 0 & C_{AB} \\ \lambda^{(i)T} & \hat{c}_{ii} & \eta^T \\ 0 & 0 & C_{BB} \end{pmatrix}, \tag{5.72}$$

where $\widehat{C}_{AA}^{(i)}$ and C_{BB} are upper triangular, and $\lambda^{(i)}$ is the fill-in. This implies that the GLLSP (5.71) can be equivalently written as

$$\operatorname*{argmin}_{\widehat{V}_A^{(i)}, \hat{v}_i^{(i)}, \widetilde{V}_B^{(i)}, \beta_i} \left(\|\widehat{V}_A^{(i)}\|_F + \|\hat{v}_i^{(i)}\|_2 + \|\widetilde{V}_B^{(i)}\|_F\right) \quad \text{subject to}$$

$$\begin{pmatrix} \widetilde{Y}_A^{(i)} & \widetilde{y}_i^{(i)} & \widetilde{Y}_B^{(i)} \end{pmatrix} = \begin{pmatrix} 0 & R_i\beta_i & 0 \end{pmatrix} + \begin{pmatrix} \widehat{V}_A^{(i)} & \hat{v}_i^{(i)} & \widetilde{V}_B^{(i)} \end{pmatrix} \begin{pmatrix} \widehat{C}_{AA}^{(i)T} & \lambda^{(i)} & 0 \\ 0^T & \hat{c}_{ii} & 0^T \\ C_{AB}^T & \eta & C_{BB}^T \end{pmatrix}, \tag{5.73}$$

where

$$\widetilde{V}_i W^{(i)} = \begin{pmatrix} \widehat{V}_A^{(i)} & \hat{v}_i^{(i)} & \widetilde{V}_B^{(i)} \end{pmatrix}.$$

From (5.73) it follows that $\hat{V}_A^{(i)}$ and $\widetilde{V}_B^{(i)}$ can be computed by solving the triangular system

$$\begin{pmatrix} \widehat{V}_A^{(i)} & \widetilde{V}_B^{(i)} \end{pmatrix} \begin{pmatrix} \widehat{C}_{AA}^{(i)T} & 0 \\ C_{AB}^T & C_{BB}^T \end{pmatrix} = \begin{pmatrix} \widetilde{Y}_A^{(i)} & \widetilde{Y}_B^{(i)} \end{pmatrix},$$

while the random vector $\hat{v}_i^{(i)} = 0$ is set to zero in order to minimize the objective function. Therefore, the estimator for β_i derives from the solution of the triangular system

$$R_i \hat{\beta}_i = \tilde{y}_i^{(i)} - \widehat{V}_A^{(i)} \lambda^{(i)} - \widetilde{V}_B^{(i)} \eta. \tag{5.74}$$

5.C.4 The covariance matrix of the estimators

The variance-covariance of the estimators $\hat{\beta}_i$ and $\hat{\beta}_j$ is given by

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \mathrm{E}\big((\hat{\beta}_i - \mathrm{E}(\hat{\beta}_i))(\hat{\beta}_j - \mathrm{E}(\hat{\beta}_j))^T\big).$$

From (5.74) and (5.73) it follows that

$$\hat{\beta}_i - \beta_i = R_i^{-1} \hat{v}_i^{(i)} \hat{c}_{ii} = R_i^{-1} Q_i^T V w_i \hat{c}_{ii} = \hat{c}_{ii} R_i^{-1} \big(w_i^T \otimes Q_i^T\big) \,\mathrm{Vec}(V), \tag{5.75}$$

where w_i is the ith column of $W^{(i)}$ such that $\hat{v}_i^{(i)} = \widetilde{V}^{(i)} w_i = Q_i^T V w_i$.

Since $\mathrm{Vec}(V)$ has zero mean and covariance matrix I_{GT}, it follows from (5.75) that, for $i, j = 1, \ldots, G$,

$$\mathrm{E}(\hat{\beta}_i) = \beta_i$$

and

$$\begin{aligned} \mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) &= \mathrm{E}\big((\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j)^T\big) \\ &= \hat{c}_{ii}\hat{c}_{jj} R_i \big(w_i^T \otimes Q_i^T\big) \,\mathrm{E}\big(\mathrm{Vec}(V)\,\mathrm{Vec}(V)^T\big) \big(w_j \otimes Q_j\big) R_j^T \\ &= \hat{c}_{ii}\hat{c}_{jj} R_i \big(w_i^T w_j \otimes Q_i^T Q_j\big) R_j^T. \end{aligned}$$

Using (5.64), this is reduced to

$$\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \begin{cases} \hat{c}_{ii}^2 (R_i^T R_i)^{-1}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

5.C.5 Conclusion

A computationally efficient method has been proposed to estimate SUR models with orthogonal regressors (SUR-OR model) using the GLLSP approach. Initially this method computes the QRD of the exogenous matrices and then exploits the orthogonality of the regressors to reduce the GLLSP to smaller, independent GLLSPs. To estimate each of the smaller GLLSPs, a sequence of Givens rotations is employed to annihilate the elements in a single column of the triangular matrix C. The computation of the RQD (5.60b) is avoided. This makes the proposed method particularly useful when an iterative estimator of SUR-OR is needed. An expression for the variance-covariance matrix of the estimators shows that there is no correlation among them. This expression is simpler than previously proposed methods [79, pp. 137-139].

The proposed method for computing the BLUE of the SUR-OR model has lower computational complexity than those which compute the BLUE of the SUR model without orthogonality among the regressors. This suggests that the proposed method could be used to derive a biased estimator of the SUR model when the derivation of the BLUE is not computationally feasible. The efficiency of this biased estimator needs to be studied and compared with that of other estimators [20, 22].

The proposed method is intrinsically parallel. The QRDs of the exogenous matrices and the solution of smaller GLLSPs can be obtained simultaneously [46, pp. 29-38]. However, another possibility bears investigation is the design of (parallel) Givens strategies that could operate on the matrix C to produce $\widehat{C}^{(1)}, \ldots, \widehat{C}^{(G)}$ in (5.72) as intermediate results [46, ch. 23]. This kind of parallelism would be efficient for SIMD architectures [47].

Chapter 6

Conclusions

Computationally efficient methods for estimating Seemingly Unrelated Regressions (SUR) models have been proposed and analyzed. The SUR model have been considered in its standard form and with additional assumptions on the regressor matrices and/or the disturbances. Speficially the following models have been investigated:

- the SUR model,
- SUR models deriving from Vector Autoregressive (VAR) processes,
- SUR models having disturbances which are generated by a VAR process,
- SUR models with unequal size observations,
- SUR models with Orthogonal Regressors.

The aforementioned SUR models have been reformulated as Generalized Linear Least Squares Problems (GLLSP). The algorithms which solve the GLLSPs and provide the Best Linear Unbiased Estimators (BLUE) have use the Generalized QR decomposition (GQRD) as the main computational tool [55, 61, 62, 63]. In each model the structure and properties of the matrices involved are exploited. This approach allows for computationally efficient and numerically stable implementations [62, 76]. Furthermore, the proposed algorithms can be extended to detect inconsistencies and handle models having singular covariance matrices [45]. The computational details of the various implementations of each method have been considered.

The VAR model with zero coefficient constraints or Granger–caused variables have been considered as a SUR model with common, or proper subset regressors, respectively [26]. The exogenous matrices of the regression equations comprise columns from a block Toeplitz matrix. Efficient numerial and computational methods that exploit the special structure of the SUR model are proposed. Block versions of the algorithms which are suitable for conventional high performance computers need to be designed [17]. The adaptation of the numerical methods to tackle other models that have similar matrix structures as those proposed here needs to be considered. Currently the Vector Error Correction Model (VECM) and the Johansen procedure for estimating cointegrated systems are investigated. The VECM has a structure similar to that of (2.4), while the Johansen procedure requires the OLS estimation of a linear system having a block Toeplitz structure [36, 37, 58].

The SUR model with vector autoregressive distubances is firstly transformed to a smaller model, then the GQRD is used as the main computational tool to exploit the structure of the matrices and to estimate the reduced-size model [24]. The computational savings are significant for large samples. Computational results confirm the efficiency of the proposed method when compared to solving the general linear model. The estimator of the model is computed iteratively when the covariance matrix of the uncorrelated component or the matrix of AR parameters are unknown and need to be estimated. Some of the computations are unnecessary if only one of these two matrices changes. An iterative estimation procedure that efficiently utilizes computations from previous steps has been developed. The extension of the proposed method when the SUR model has VAR(p) disturbances, that is, when the disturbance matrix U satisfies $U - \sum_{i=1}^{p} Z^i U A_i^T = E$. Furhermore, in the case of autoregressive disturbances the matrix of AR coefficient is diagonal and this allows to simplify the proposed method.

The SUR model with unequal size of observations (SUR-USO) has been treated as a GLLSP [23]. The algorithms proposed for its solution use the GQRD by exploiting the block-sparse structure of the matrices. The first algorithm initially computes the QRD of the exogenous matrix by annihilating from bottom to the top blocks of observations which consist of a non-zero block-superdiagonal. The annihilation of the blocks is obtained by orthogonal transformations which do not create any fill-in. These transformations are also applied from the left of the Cholesky factor, which then needs to be retriangularized. The second recursive algorithm interleaves the QRD and RQD of the exogenous and modified Cholesky factors, respectively. This avoids the explicit computation of the RQD and thus, reduces the computational burden of the estimation procedure. For the case of normally distributed disturbances the maximum likelihood estimation has been considered. A closed-form solution of the Cholesky factor of the covariance matrix has been derived by solving the first order Kuhn-Tucker conditions of the likelihood function. This resulted an iterative procedure to estimate the SUR-USO model when the variance–covariance matrix Σ is unknown. Furthermore, this procedure never yields a non-definite estimator for Σ. The proposed method can adapted to estimate unbalanced Panel Data models and extended to tackle Simultaneous Equations and SUR models with missing observations [5, 33, 35].

Four algorithms for solving the SUR model have been compared [20]. The first algorithm estimates an ordinary linear model (OLM) equivalent to the SUR model which have a block triangular regressor matrix. The BLUE of the SUR model results from the least-squares (LS) solution of the OLM. This algorithm produces the LS estimator by exploiting the sparse structure of the matrices and outperforms the corresponding LAPACK subroutine (DGELS), which treats the matrices as full, when the problem is not very small.

The remaining three algorithms compute the BLUE by formulating the SUR model as a GLLSP which is solved by using the GQRD [21, 46, 50]. The first method computes the GQRD of the block diagonal and Kronecker structures of matrix of the exogenous variables and the Cholesky factor of the dispersion matrix, respectively. This method is computationally more efficient than the corresponding LAPACK routine (DGGGLM) that solves the general linear model. The second method solves the GLLSP iteratively. Each iteration solves a smaller sized GLLSP. The main advantage of this method is that it avoids the formation of the computationally expensive retriangularization of the Cholesky factor. Finally, a recursive estimation strategy is proposed. This is found to be the most efficient when the model is not very small. Furthermore, this strategy requires less memory and can thus solve larger problems.

The algorithms are reassessed after an initial orthogonal transformation is made to reduce the SUR model to one of smaller size. This transformation lead to a significant reduction for large-scale models, where the number of observations in each equation is much larger than the total number of regressors. The solution for the SUR, and consequently transformed SUR, model when the regressions have common exogenous factors is currently under investigation. The algorithms for solving the transformed SUR model (5.44) is currently adapted to solve simultaneous equations models [6, 15, 52, 89]. Although the first algorithm is the most computationally efficient, it is numerically less stable than the other three [76]. Thus it may provide a poor solution when the covariance matrix is ill-conditioned and fails if it is singular [45, 55, 61, 62]. In such cases, the GLLSP approach should be used.

The methods presented in this work can be adapted or extended to estimate related econometrics models. These include dynamic simultaneous equations models, structural VAR, panel data models, SUR and simultaneous equations models with one-way and two-way error components disturbances, SUR models with large and sparse regressor matrices, nonlinear SUR models and SUR models with inequality constraints on the parameters. Furthermore, the numerical stability of the algorithms needs further investigation.

Bibliography

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK Users' Guide*. SIAM, Philadelphia, 1992.

[2] E. Anderson, Z. Bai, and J.J. Dongarra. Generalized QR factorization and its applications. *Linear Algebra and its Applications*, 162:243–271, 1992.

[3] H.C. Andrews and J. Kane. Kronecker matrices, computer implementation, and generalized spectra. *Journal of the ACM*, 17(2):260–268, 1970.

[4] B.H. Baltagi. *Econometric Analysis of Panal Data*. John Wiley and Sons, 2nd edition, 2001.

[5] B.H. Baltagi, S.H. Song, and B.C. Jung. Simple lm tests for the unbalanced nested error component regression model. *Econometric Reviews*, 21(2):167–187, 2002.

[6] D.A. Belsley. Paring 3SLS calculations down to manageable proportions. *Computer Science in Economics and Management*, 5:157–169, 1992.

[7] D.A. Belsley, E. Kuh, and R.E. Welsch. *Regression Diagnostics: Identifying Influential Observations and Sources of Collinearity*. John Wiley and Sons, 1980.

[8] Å. Björck. A general updating algorithm for constrained linear least squares problems. *Siam Journal on Scientific and Statistical Computing*, 5(2):394–402, 1984.

[9] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, 1996.

[10] W.M. Bolstad. An estimation of seemingly unrelated regression-model with contemparaneous covariances based on an efficient recursive algorithm. *Communications in statistics-simulation and computation*, 16(3):689–698, 1987.

[11] P. Businger and G.H. Golub. Linear least squares solutions by Householder transformations. *Numerische Mathematik*, 7:269–276, 1965.

[12] J.-P. Chavas. Recursive estimation of simultaneous equation models. *Journal of Econometrics*, 18:207–217, 1982.

[13] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. Discussion. *J. R. Stat. Soc., Ser. B*, 39:1–38, 1977.

[14] W. Dent. Information and computation in simultaneous equations estimation. *Journal of Econometrics*, 4:89–95, 1976.

[15] P.J. Dhrymes. *Topics in Advanced Econometrics*, volume Vol.2: Linear and Nonlinear Simultaneous Equations. Springer–Verlag, New York, 1994.

[16] E. Dinenis and E.J. Kontoghiorghe. Computing solutions of SURE models with linear equality and stochastic constraints. *Journal of Mathematical Modelling and Scientific Computing*, 8:234–239, 1997.

[17] J.J. Dongarra, I.S. Duff, D.C. Sorensen, and H.A. van der Vorst. *Numerical Linear Algebra for High-Performance Computers*, volume 7 of *Software, Environments, and Tools*. SIAM, Philadelphia, 1998.

[18] D.W. Fausett and C.T. Fulton. Large least squares problems involving Kronecker products. *SIAM Journal on Matrix Analysis and Applications*, 15:219–227, 1994.

[19] D.W. Fausett, C.T. Fulton, and H. Hashish. Improved parallel QR method for large least squares problems involving Kronecker products. *Journal of Computational and Applied Mathematics*, 78:63–78, 1997.

[20] P. Foschi, D.A. Belsley, and E.J. Kontoghiorghes. A comparative study of algorithms for solving seemingly unrelated regressions models. *Computational Statistic & Data Analysis*, 2003. (In press).

[21] P. Foschi, L. Garin, and E.J. Kontoghiorghes. Numerical and computational methods for solving sur models. In E.J. Kontoghiorghes, B. Rustem, and S.Siokos, editors, *Computational Methods in Decision-Making, Economics and Finance*, Applied Optimization. Kluwer Academic Publishers, 2003.

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

[23] P. Foschi and E.J. Kontoghiorghes. Solution of seemingly unrelated regression models with unequal size of observations. *Computational Statistics & Data Analysis*, 41(1):211–229, 2002.

[24] P. Foschi and E.J. Kontoghiorghes. Estimating seemingly unrelated regression models with vector autoregressive disturbances. *Journal of Economic Dynamics and Control*, 2003. (In press).

[25] P. Foschi and E.J. Kontoghiorghes. Estimating SUR models with orthogonal regressors: computational aspects. *Linear Algebra and its Applications*, 2003. (In press).

[26] P. Foschi and E.J. Kontoghiorghes. Estimation of VAR models: computational aspects. *Computational Economics*, 21(1):3–22, 2003.

[27] K.A. Gallivan, S. Thirumalai, P. Van Dooren, and V. Vermaut. High performance algorithms for toeplitz and block toeplitz matrices. *Linear Algebra and its Applications*, 241–243:343–388, 1996.

[28] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, Maryland, 3ed edition, 1996.

[29] G.H. Golub and J.H. Wilkinson. Note on the iterative refinement of least squares solution. *Numerische Mathematik*, 9:139–148, 1966.

[30] A. Graham. *Kronecker products and matrix calculus: with applications*. Ellis Horwood Series in Mathematics and its Applications. Chichester: Ellis Horwood Limited, Publishers; New York: Halsted Press, a division of John Wiley & Sons., 1986.

[31] M. Gulliksson and P.-Å. Wedin. Modifying the QR decomposition to constrained and weighted linear least squares. *SIAM Journal on Matrix Analysis and Applications*, 13:4:1298–1313, 1992.

[32] James D. Hamilton. *Time series analysis*. Princeton Univesity Press, 1994.

[33] R.R. Hocking and W.B. Smith. Estimation of parameters in the multivariate normal distribution with missing observations. *Journal of the American Statistical Association*, (63):159–173, 1968.

[34] L.S. Jennings. Simultaneous equations estimation (computational aspects). *Journal of Econometrics*, 12:23–39, 1980.

[35] M. Jhun, S.H. Song, and B.C. Jung. Blup in the nested panel regression model with serially correlated errors. *Computational Statistics & Data Analysis*, 2003. (In press).

[36] S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.

[37] S. Johansen. *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press, 1995.

[38] G.G. Judge, W.E. Griffiths, R.C. Hill, H. Lütkepohl, and T.C. Lee. *The Theory and Practice of Econometrics*. Wiley series in Probability and Mathematical Statistics. John Wiley and Sons, second edition, 1985.

[39] T. Kailath, S.-Y. Kung, and M. Morf. Displacement ranks of a matrix. *Bull. Amer. Math. Soc., New Ser.*, 1:769–773, 1979.

[40] T. Kailath and A.H. Sayed. Displacement structure: theory and applications. *SIAM Review*, 37(3):297–386, 1995.

[41] I. Karasalo. A criterion for truncation of the QR decomposition algorithm for the singular linear least squares problem. *BIT*, 14:156–166, 1974.

[42] J. Kmenta and R.F. Gilbert. Small sample properties of alternative estimators of seemingly unrelated regressions. *Journal of the American Statistical Association*, 63:1180–1200, 1968.

[43] J. Kmenta and R.F. Gilbert. Estimation of seemingly unrelated regressions with autoregressive disturbances. *Journal of the American Statistical Association*, 65(329):186–197, 1970.

[44] E.J. Kontoghiorghes. Parallel strategies for computing the orthogonal factorizations used in the estimation of econometric models. *Algorithmica*, 25:58–74, 1999.

[45] E.J. Kontoghiorghes. Inconsistencies and redundancies in SURE models: computational aspects. *Computational Economics*, 16(1+2):63–70, 2000.

[46] E.J. Kontoghiorghes. *Parallel Algorithms for Linear Models: Numerical Methods and Estimation Problems*, volume 15 of *Advances in Computational Economics*. Kluwer Academic Publishers, Boston, MA, 2000.

[47] E.J. Kontoghiorghes. Parallel strategies for rank–k updating of the QR decomposition. *SIAM Journal on Matrix Analysis and Applications*, 22(3):714–725, 2000.

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

[49] E.J. Kontoghiorghes. Computational methods for modifying seemingly unrelated regressions models. *Journal of Computational and Applied Mathematics*, 2003. (forthcoming).

[50] E.J. Kontoghiorghes and M. R. B. Clarke. An alternative approach for the numerical solution of seemingly unrelated regression equations models. *Computational Statistics & Data Analysis*, 19(4):369–377, 1995.

[51] E.J. Kontoghiorghes and E. Dinenis. Solving triangular seemingly unrelated regression equations models on massively parallel systems. In M. Gilli, editor, *Computational Economic Systems: Models, Methods & Econometrics*, volume 5 of *Advances in Computational Economics*, pages 191–201. Kluwer Academic Publishers, 1996.

[52] E.J. Kontoghiorghes and E. Dinenis. Computing 3SLS solutions of simultaneous equation models with a possible singular variance–covariance matrix. *Computational Economics*, 10:231–250, 1997.

[53] E.J. Kontoghiorghes and E. Dinenis. Towards the parallel implementation of the SURE model estimation algorithm. *Journal of Mathematical Modelling and Scientific Computing*, 8:335–341, 1997.

[54] E.J. Kontoghiorghes and P. Foschi. Algorithms for solving SURE models. Lecture Notes in Computer Science, pages 490–497. Springer Verlag, 2000.

[55] S. Kourouklis and C.C. Paige. A constrained least squares approach to the general Gauss–Markov linear model. *Journal of the American Statistical Association*, 76(375):620–625, 1981.

[56] K. Lahiri and P. Schmidt. On the estimation of triangular structural systems. *Econometrica*, 1978.

[57] C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*. Prentice–Hall Englewood Cliffs, 1974.

[58] H. Lütkepohl. *Introduction to Multiple Time Series Analysis*. Springer–Verlag, second edition, 1993.

[59] H. Lütkepohl. *Handbook of matrices*. Wiley, 1996.

[60] S. Orbe, E. Ferreira, and J. Rodriguez-Poo. An algorithm to estimate time varying parameter sur models under different type of restrictions. *Computational Statistics & Data Analysis*, 2002. (In press).

[61] C.C. Paige. Numerically stable computations for general univariate linear models. *Communications on Statistical and Simulation Computation*, 7(5):437–453, 1978.

[62] C.C. Paige. Computer solution and perturbation analysis of generalized linear least squares problems. *Mathematics of Computation*, 33(145):171–183, 1979.

[63] C.C. Paige. Fast numerically stable computations for generalized linear least squares problems. *SIAM Journal on Numerical Analysis*, 16(1):165–171, 1979.

[64] C.C. Paige. Some aspects of generalized QR factorizations. In M. G. Cox and S. J. Hammarling, editors, *Reliable Numerical Computation*, pages 71–91. Clarendon Press, Oxford, UK, 1990.

[65] R.W. Parks. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association*, 62:500–509, 1967.

[66] D.S.G. Pollock. *The Algebra of Econometrics (Wiley series in Probability and Mathematical Statistics)*. John Wiley and Sons, 1979.

[67] D.S.G. Pollock. *A handbook of time-series analysis, signal processing and dynamics*. Signal processing and its applications. Academic Press, 1999.

[68] C.R. Rao and H. Toutenburg. *Linear Models: Least Squares and Alternatives*. Springer series in Statistics. Springer, 1995.

[69] P.A. Regalia and S.K. Mitra. Kronecker products, unitary matrices and signal processing applications. *SIAM Review*, 31(4):586–613, 1989.

[70] N.S. Revankar. Some finite samples results in the context of two seemingly unrelated regression equations. *Journal of the American Statistical Association*, 69:187–190, 1974.

[71] P. Schmidt. *Econometrics (Statistics: Textbooks and Monographs)*, volume 18. Marcel Dekker, Inc, 1976.

[72] P. Schmidt. Estimation of seemingly unrelated regressions with unequal numbers of observations. *Journal of Econometrics*, (5):365–377, 1977.

[73] P. Schmidt. A note on the estimation of seemingly unrelated regression systems. *Journal of Econometrics*, 7:259–261, 1978.

[74] V.K. Sharma. Estimation of seemingly unrelated regressions with unequal numbers of observation. *Sankhya: The Indian Journal of Statistics*, 55:135–138, 1993.

[75] D M. Smith. *Regression using QR decomposition methods*. PhD thesis, University of Kent, 1991.

[76] I. Söderkvist. On algorithms for generalized least–squares problems with ill–conditioned covariance matrices. *Computational Statistics*, 11(3):303–313, 1996.

[77] J.N. Srivastava and M.K. Zaatar. A monte carlo comparison of four estimators of the dispersion matrix of a bivariate normal population, using incomplete data. *Journal of the American Statistical Association*, 68(341):180–183, 1973.

[78] V.K. Srivastava and T.D. Dwivedi. Estimation of seemingly unrelated regression equations Models: a brief survey. *Journal of Econometrics*, 10:15–32, 1979.

[79] V.K. Srivastava and D.E.A. Giles. *Seemingly Unrelated Regression Equations Models: Estimation and Inference (Statistics: Textbooks and Monographs)*, volume 80. Marcel Dekker, Inc., 1987.

[80] V.K. Srivastava and R. Tiwari. Efficiency of two–stage and three–stage least squares estimators. *Econometrica*, 46(6):1495–1498, 1978.

[81] M. Stewart and P. Van Dooren. Stability issues in the factorization of structured matrices. *SIAM Journal on Matrix Analysis and Applications*, 18(1):104–118, 1997.

[82] A.S. Tanenbaum. *Structured Computer Organization*. Prentice Hall, 4th edition, 1999.

[83] L.G. Telser. Iterative estimation of a set of linear regression equations. *Journal of the American Statistical Association*, 59:845–862, 1964.

[84] H. Theil. *Principles of Econometrics*. John Wiley & Sons, Inc, 1971.

[85] D.A. Turkington. Generalised vec operators and the seemingly unrelated regression equations model with vector correlated disturbances. *Journal of Econometrics*, 99:225–253, 2000.

[86] A. Zellner. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368, 1962.

[87] A. Zellner. Estimators for seemingly unrelated regression equations: some exact finite sample results. *Journal of the American Statistical Association*, 58:977–992, 1963.

[88] A. Zellner. An error–components procedure (ECP) for introducing prior information about covariance matrices and analysis of multivariate regression models. *International Economic Review*, 20(3):679–692, 1979.

[89] A. Zellner and H. Theil. Three–stage least squares: simultaneous estimation of simultaneous equations. *Econometrica*, 30(1):54–78, 1962.