

Estimating large-scale general linear and seemingly unrelated regressions models after deleting observations

Stella Hadjiantoni¹ and Erricos John Kontoghiorghes^{1,2}

¹School of Economics and Finance, Queen Mary, University of London

²Faculty of Management and Economics, Cyprus University of Technology

Abstract

A new numerical method to solve the downdating problem (and variants thereof), namely removing the effect of some observations from the generalized least squares (GLS) estimator of the general linear model (GLM) after it has been estimated, is extensively investigated. It is verified that the solution of the downdated least squares problem can be obtained from the estimation of an equivalent GLM, where the original model is updated with the imaginary deleted observations. This updated GLM has a non positive definite dispersion matrix which comprises complex covariance values and it is proved herein to yield the same normal equations as the downdated model. Additionally, the problem of deleting observations from the seemingly unrelated regressions model is addressed, demonstrating the direct applicability of this method to other multivariate linear models. The algorithms which implement the novel downdating method utilize efficiently the previous computations from the estimation of the original model. As a result, the computational cost is significantly reduced. This shows the great usability potential of the downdating method in computationally intensive problems. The downdating algorithms have been applied to real and synthetic data to illustrate their efficiency.

Keywords: downdating, generalized least squares, singular dispersion matrix, seemingly unrelated regressions, updating

1 Introduction

The re-estimation of a model is often required in several applications when changes are incorporated into the data (Chambers, 1971; Chavas, 1982; Rader and Steinhardt, 1986; Björck, 1996; Belsley et al., 2004; Yanev and Kontoghiorghes, 2009). Observations often have to be deleted from a model in order to remove the effect of this data from the least squares solution. In many time series problems, for example, as data arrive sequentially, old data are deleted and new data are added creating a moving (rolling) window of the data (Clark and McCracken, 2009; Pesaran and Pick, 2011; Rossi and Inoue, 2012; Pesaran et al., 2013). Similarly in statistics, a series of least squares problems is often needs to be solved when observations in a data set are identified as outliers, faulty or

Corresponding author: S. Hadjiantoni, School of Economics and Finance, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. Email address: s.hadjiantoni@qmul.ac.uk

can be influential in the estimation of the model and regression diagnostics are performed (Cook, 1977; Christensen et al., 1992; Preisser and Qaqish, 1996; Belsley et al., 2004; Haslett and Dillane, 2004; Preisser and Perin, 2007; Zhu et al., 2012). However, in the case of large-scale linear models which are intractable to implement, it is not advantageous to perform all computations afresh especially when the procedure of deleting and/or adding observations is executed sequentially. Instead, computationally attractive and numerically accurate sequential methods are preferable.

The problem of removing the effect of some observations from the least squares solution, herein referred to as the downdating problem, has been previously considered for the ordinary linear model (Golub, 1969; Chambers, 1971; Gragg et al., 1979; Rader and Steinhardt, 1986; Elden and Park, 1994; Björck, 1996; Golub and Van Loan, 1996; Yanev and Kontoghiorghes, 2008). Formulae which provide the change in the generalized least squares (GLS) estimator when observations are removed for deletion diagnostics were proposed (Martin, 1992; Bhimasankaram et al., 1995; Haslett, 1999; Jammalamadaka and Sengupta, 1999, 2007). These techniques are theoretically sound. However, they rely heavily on matrix inverses which are numerically unstable and expensive to compute. Hence, these methods cannot be employed to solve computationally demanding estimation problems. Additionally, when the dispersion matrix is singular, generalized inverses impose restrictions in deriving the best linear unbiased estimator (BLUE) of the model (Zyskind, 1967; Zyskind and Martin, 1969; Kourouklis and Paige, 1981).

Herein, the downdating of the general linear model (GLM) is thoroughly investigated. A novel numerical method to estimate the GLM after deleting observations is introduced. The proposed method updates the original model with the imaginary deleted observations, resulting in an updated GLM with a non positive definite dispersion matrix. It has been proved theoretically that the proposed model yields the GLS estimator that would be derived if the downdated model -the model with the remaining observations after the deletion- is estimated afresh. In contrast with existing methods, the new downdating method is based on strategies which are numerically and computationally efficient and they are shown to yield the BLUE of the GLM. That is, the proposed model is processed within the context of generalized linear least squares where the inversion of a matrix is postponed until the last step of the procedure. Additionally, the computational cost is reduced by utilizing efficiently previous computations from the estimation of the original model and by employing hyperbolic and unitary transformations. This method solves the downdating problem of the GLM even when the dispersion matrix is singular.

Moreover, within the framework of the sliding window method in time series analysis and the k-fold cross validation procedures, a method for the simultaneous addition and deletion of data has also been considered (Pollock, 2003; Yanev and Kontoghiorghes, 2009). In addition, the problem of deleting observations from the seemingly unrelated regressions (SUR) model has also been explored and various cases have been examined. The SUR model is a system of regression equations which takes into account contemporaneous correlation (Zellner, 1962). It is a special case of the GLM where the data have Kronecker structure. The SUR model has many applications in time series analysis problems (Kmenta and Gilbert, 1970; Judge et al., 1985; Chib and Greenberg, 1995; Smith and Kohn, 2000; Griffiths and Valenzuela, 2006) and is also applied in studies for the analysis of longitudinal data (Verbyla and Venables, 1988; Rochon, 1996; Shieh, 2000). The computational efficiency of the proposed downdating algorithms has been investigated using real and synthetic data.

Consider the general linear model (GLM)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma^2\boldsymbol{\Omega}), \quad (1.1)$$

where $\mathbf{y} \in \Re^M$ is the dependent vector, $\mathbf{X} \in \Re^{M \times K}$, $M > K$ is the exogenous data matrix with full column rank, $\boldsymbol{\beta} \in \Re^K$ is the vector of parameters to be estimated and $\boldsymbol{\epsilon} \in \Re^M$ is the error term with zero mean and positive definite variance covariance matrix $\sigma^2\boldsymbol{\Omega}$ (Rao, 2002). As shown in Aitken

(1934), the BLUE, say $\hat{\boldsymbol{\beta}}$, for $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}. \quad (1.2)$$

This is the generalized least squares (GLS) solution of $\operatorname{argmin}_{\boldsymbol{\beta}} \|\mathbf{C}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|^2$, where $\boldsymbol{\Omega} = \mathbf{C}\mathbf{C}^T$ is the Cholesky decomposition and $\|\cdot\|$ denotes the Euclidean norm. However, the solution of the latter becomes computationally expensive and can be unstable when $\boldsymbol{\Omega}$ is ill-conditioned (Paige, 1978, 1979; Kourouklis and Paige, 1981).

The downdating of the GLM can be defined as estimating the GLM after a number of observations is deleted. In particular, let \mathbf{y} , \mathbf{X} and $\boldsymbol{\epsilon}$ in (1.1) be partitioned as follows

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}, \quad (1.3)$$

where $\mathbf{X}_1 \in \mathfrak{R}^{M_1 \times K}$, $\mathbf{X}_2 \in \mathfrak{R}^{M_2 \times K}$, $M = M_1 + M_2$ and \mathbf{y} , $\boldsymbol{\epsilon}$ are partitioned accordingly. Then, the variance covariance matrix $\boldsymbol{\Omega}$ in (1.1) and its upper triangular Cholesky factor \mathbf{C} , i.e. $\boldsymbol{\Omega} = \mathbf{C}\mathbf{C}^T$, (Kontoghiorghes, 2000) are partitioned respectively as

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{0} & \mathbf{C}_{22} \end{pmatrix}. \quad (1.4)$$

Here $\boldsymbol{\Omega}_{11}$ and $\boldsymbol{\Omega}_{22}$ are the dispersion matrices of the first M_1 and the remaining M_2 observations, so that \mathbf{C}_{22} is the upper triangular Cholesky factor of $\boldsymbol{\Omega}_{22}$, i.e. $\boldsymbol{\Omega}_{22} = \mathbf{C}_{22}\mathbf{C}_{22}^T$.

Next, consider estimating the downdated GLM (DGLM)

$$\mathbf{y}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2, \quad \boldsymbol{\epsilon}_2 \sim (\mathbf{0}, \sigma^2 \boldsymbol{\Omega}_{22}), \quad (1.5)$$

after (1.1) has been estimated. Without loss of generality it is assumed that the first M_1 observations are deleted and also that $M_2 > K$, with \mathbf{X}_2 being full column rank. Thus the downdated GLS estimator is given by $\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2^T \boldsymbol{\Omega}_{22}^{-1} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \boldsymbol{\Omega}_{22}^{-1} \mathbf{y}_2$ which is the solution of the GLS problem $\operatorname{argmin}_{\boldsymbol{\beta}_2} \|\mathbf{C}_{22}^{-1}(\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)\|^2$.

In the next section a new method that solves the downdating problem of the GLM is presented. Within this framework, the downdating of the GLM when the dispersion matrix $\boldsymbol{\Omega}$ is singular and also the simultaneous addition and deletion of data are discussed. The problem of deleting observations from the SUR model is investigated in section 3. Section 4 discusses the computational results. Finally, conclusions and future work are presented in section 5.

2 Downdating the General Linear Model

An alternative approach for the estimation of the downdated observations GLM (1.5) is to estimate an updated GLM where the observations added offset those that should be deleted. The main difficulty when deleting observations from a GLM is that a block of columns and a block of rows must be deleted from the dispersion matrix, so that the variance of the deleted observations and the covariances related to the deleted observations no longer affect the solution of the least squares problem. The aim here is to derive an equivalent model to the DGLM in (1.5) that will yield the same estimator but will utilize previous computations that have been obtained when estimating the original model (1.1).

For an ordinary linear model (OLM) the downdated least squares estimator is obtained if the model is updated with the imaginary deleted observations (Golub and Van Loan, 1996; Yanev and Kontoghiorghes, 2008). That is, consider the models (1.1) and (1.5) with $\boldsymbol{\Omega} = \mathbf{I}_M$, $\boldsymbol{\Omega}_{22} = \mathbf{I}_{M_2}$ where their ordinary least squares solution is given by $\hat{\boldsymbol{\beta}}_{OLS} = \mathbf{R}^{-1} \tilde{\mathbf{y}}$ and $\hat{\boldsymbol{\beta}}_{2OLS} = \mathbf{R}_2^{-1} \tilde{\mathbf{y}}_2$. Here \mathbf{R} and \mathbf{R}_2 emanate from the QR decompositions (QRD)

$$\mathbf{Q}^T(\mathbf{X} \ \mathbf{y}) = \begin{pmatrix} \mathbf{R} & \tilde{\mathbf{y}} \\ \mathbf{0} & \hat{\mathbf{y}} \end{pmatrix} \quad \text{and} \quad \mathbf{Q}_2^T(\mathbf{X}_2 \ \mathbf{y}_2) = \begin{pmatrix} \mathbf{R}_2 & \tilde{\mathbf{y}}_2 \\ \mathbf{0} & \hat{\mathbf{y}}_2 \end{pmatrix}.$$

Then, \mathbf{R}_2 can also be obtained by forming the hyperbolic QR decomposition (HQRD)

$$\tilde{\mathbf{Q}}^H \begin{pmatrix} \imath \mathbf{X}_1 & \imath \mathbf{y}_1 \\ \mathbf{R} & \tilde{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \imath \tilde{\mathbf{y}}_2 \\ \mathbf{R}_2 & \tilde{\mathbf{y}}_2 \end{pmatrix}, \quad (2.1)$$

where $\imath^2 = -1$ and $(\cdot)^H$ denotes the conjugate transpose of a matrix.

Now consider premultiplying the GLM (1.1) with \mathbf{C}^{-1} in order to transform it to an OLM. From the partitioning of \mathbf{C} in (1.4), it follows (see Björck (1996)) that

$$\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C}_{11}^{-1} & -\mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \\ \mathbf{0} & \mathbf{C}_{22}^{-1} \end{pmatrix}.$$

Thus, the transformed GLM (TGLM) is given by

$$\begin{pmatrix} \mathbf{C}_{11}^{-1} \mathbf{y}_1 - \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{y}_2 \\ \mathbf{C}_{22}^{-1} \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11}^{-1} \mathbf{X}_1 - \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{X}_2 \\ \mathbf{C}_{22}^{-1} \mathbf{X}_2 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{C}_{11}^{-1} \boldsymbol{\epsilon}_1 - \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \boldsymbol{\epsilon}_2 \\ \mathbf{C}_{22}^{-1} \boldsymbol{\epsilon}_2 \end{pmatrix} \quad (2.2)$$

or conformably written as

$$\begin{pmatrix} \check{\mathbf{y}}_1 \\ \check{\mathbf{y}}_2 \end{pmatrix} = \begin{pmatrix} \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_2 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \check{\boldsymbol{\epsilon}}_1 \\ \check{\boldsymbol{\epsilon}}_2 \end{pmatrix}, \quad \begin{pmatrix} \check{\boldsymbol{\epsilon}}_1 \\ \check{\boldsymbol{\epsilon}}_2 \end{pmatrix} \sim \left(\mathbf{0}, \sigma^2 \begin{pmatrix} \mathbf{I}_{M_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_2} \end{pmatrix} \right). \quad (2.3)$$

Observe from (2.2) that on deleting the first M_1 observations from the GLM (1.1), the resulting model is equivalent to that obtained by deleting the first M_1 observations from the TGLM (2.3).

Next, consider the model

$$\begin{pmatrix} \imath \check{\mathbf{y}}_1 \\ \check{\mathbf{y}}_1 \\ \check{\mathbf{y}}_2 \end{pmatrix} = \begin{pmatrix} \imath \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_2 \end{pmatrix} \boldsymbol{\beta}_2 + \begin{pmatrix} \imath \check{\boldsymbol{\epsilon}}_1 \\ \check{\boldsymbol{\epsilon}}_1 \\ \check{\boldsymbol{\epsilon}}_2 \end{pmatrix}, \quad \begin{pmatrix} \imath \check{\boldsymbol{\epsilon}}_1 \\ \check{\boldsymbol{\epsilon}}_1 \\ \check{\boldsymbol{\epsilon}}_2 \end{pmatrix} \sim \left(\mathbf{0}, \sigma^2 \begin{pmatrix} -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{M_2} \end{pmatrix} \right) \quad (2.4)$$

which has the GLS solution

$$\begin{aligned} \hat{\boldsymbol{\beta}}_2 &= \left(\begin{pmatrix} \imath \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_2 \end{pmatrix}^H \begin{pmatrix} -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{M_2} \end{pmatrix} \begin{pmatrix} \imath \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} \imath \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_1 \\ \check{\mathbf{X}}_2 \end{pmatrix}^H \begin{pmatrix} -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{M_2} \end{pmatrix} \begin{pmatrix} \imath \check{\mathbf{y}}_1 \\ \check{\mathbf{y}}_1 \\ \check{\mathbf{y}}_2 \end{pmatrix} \\ &= \left(\check{\mathbf{X}}_2^T \check{\mathbf{X}}_2 \right)^{-1} \check{\mathbf{X}}_2^T \check{\mathbf{y}}_2 \\ &= \left(\mathbf{X}_2^T \mathbf{C}_{22}^{-T} \mathbf{C}_{22}^{-1} \mathbf{X}_2 \right)^{-1} \mathbf{X}_2^T \mathbf{C}_{22}^{-T} \mathbf{C}_{22}^{-1} \mathbf{y}_2 \\ &= \left(\mathbf{X}_2^T \boldsymbol{\Omega}_{22}^{-1} \mathbf{X}_2 \right)^{-1} \mathbf{X}_2^T \boldsymbol{\Omega}_{22}^{-1} \mathbf{y}_2 \end{aligned}$$

on using $\check{\mathbf{y}}_2 = \mathbf{C}_{22}^{-1} \mathbf{y}_2$ and $\check{\mathbf{X}}_2 = \mathbf{C}_{22}^{-1} \mathbf{X}_2$ from (2.2). The latter is the GLS solution of the downdated GLM (1.5).

The OLM model in (2.4) is written as

$$\mathbf{C}_d^{-1} \begin{pmatrix} \imath \mathbf{y}_1 \\ \mathbf{y} \end{pmatrix} = \mathbf{C}_d^{-1} \begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X} \end{pmatrix} \boldsymbol{\beta}_2 + \mathbf{C}_d^{-1} \begin{pmatrix} \imath \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon} \end{pmatrix},$$

where

$$\mathbf{C}_d^{-1} = \begin{pmatrix} \mathbf{C}_{11}^{-1} & \mathbf{0} & -\imath \mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \\ \mathbf{0} & \mathbf{C}_{11}^{-1} & -\mathbf{C}_{11}^{-1} \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{22}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{0} & \imath \mathbf{C}_{12} \\ \mathbf{0} & \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{22} \end{pmatrix}^{-1}.$$

Thus (2.4) is equivalent to the GLM, updated with the imaginary deleted observations, given by

$$\begin{pmatrix} \imath \mathbf{y}_1 \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X} \end{pmatrix} \beta_2 + \begin{pmatrix} \imath \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon} \end{pmatrix}, \quad \begin{pmatrix} \imath \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon} \end{pmatrix} \sim (\mathbf{0}, \sigma^2 \boldsymbol{\Omega}_d). \quad (2.5)$$

Here the variance covariance matrix $\boldsymbol{\Omega}_d$ is given by

$$\boldsymbol{\Omega}_d = \begin{pmatrix} \boldsymbol{\Omega}_{33} & \imath \mathbf{V}_d \\ -\imath \mathbf{V}_d^T & \boldsymbol{\Omega} \end{pmatrix} = \mathbf{C}_d \boldsymbol{\Psi} \mathbf{C}_d^H, \quad \mathbf{C}_d = \begin{pmatrix} \mathbf{C}_{11} & \imath \tilde{\mathbf{C}}_{12} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}, \quad (2.6)$$

where $\tilde{\mathbf{C}}_{12} = (\mathbf{0} \ \mathbf{C}_{12})$ and $\boldsymbol{\Psi}$ is the signature matrix

$$\boldsymbol{\Psi} = \text{diag}(-\mathbf{I}_{M_1}, \mathbf{I}_M) = \begin{pmatrix} -\mathbf{I}_{M_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M \end{pmatrix}. \quad (2.7)$$

Thus $\boldsymbol{\Omega}_{33} = \mathbf{C}_{12} \mathbf{C}_{12}^T - \mathbf{C}_{11} \mathbf{C}_{11}^T$ and $\mathbf{V}_d = (\mathbf{C}_{12} \mathbf{C}_{12}^T \ \mathbf{C}_{12} \mathbf{C}_{22}^T)$. Observe that the normal equations of (2.5) give the GLS estimator of the DGLM (1.5), that is,

$$\begin{aligned} \hat{\beta}_2 &= \left(\begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X} \end{pmatrix}^H \begin{pmatrix} \boldsymbol{\Omega}_{33} & \imath \mathbf{V}_d \\ -\imath \mathbf{V}_d^T & \boldsymbol{\Omega} \end{pmatrix}^{-1} \begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X} \end{pmatrix} \right)^{-1} \begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X} \end{pmatrix}^H \begin{pmatrix} \boldsymbol{\Omega}_{33} & \imath \mathbf{V}_d \\ -\imath \mathbf{V}_d^T & \boldsymbol{\Omega} \end{pmatrix}^{-1} \begin{pmatrix} \imath \mathbf{y}_1 \\ \mathbf{y} \end{pmatrix} \\ &= (\mathbf{X}_2^T \boldsymbol{\Omega}_{22}^{-1} \mathbf{X}_2)^{-1} \mathbf{X}_2^T \boldsymbol{\Omega}_{22}^{-1} \mathbf{y}_2. \end{aligned}$$

This implies that the updated GLM (2.5) is equivalent to the downdated GLM (1.5).

2.1 The Generalized Linear Least Squares Approach

The GLS estimator of the GLM (1.1) is the solution of the generalized linear least squares problem (GLLSP)

$$\underset{v, \beta}{\text{argmin}} \ \|v\|^2 \quad \text{subject to} \quad \mathbf{y} = \mathbf{X}\beta + \mathbf{C}v, \quad (2.8)$$

where $\boldsymbol{\Omega} = \mathbf{C}\mathbf{C}^T$, \mathbf{C} is the upper triangular Cholesky factor and $v \sim (\mathbf{0}, \sigma^2 \mathbf{I}_M)$ is defined to give $\boldsymbol{\epsilon} = \mathbf{C}v$ (Paige, 1978, 1979; Kourouklis and Paige, 1981). In order to solve the GLLSP (2.8), let the computation of the generalized QR decomposition (GQRD) of \mathbf{X} and \mathbf{C} be given by

$$\mathbf{Q}^T (\mathbf{X} \ \mathbf{y}) = \begin{pmatrix} \mathbf{R} & \tilde{\mathbf{y}} \\ \mathbf{0} & \hat{\mathbf{y}} \end{pmatrix} \begin{matrix} K \\ M - K \end{matrix}, \quad \text{with} \quad \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix} \begin{matrix} K \\ M - K \end{matrix}, \quad (2.9a)$$

$$(\mathbf{Q}^T \mathbf{C}) \mathbf{P} = \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{pmatrix} \begin{matrix} K \\ M - K \end{matrix}, \quad (2.9b)$$

where \mathbf{R} and \mathbf{U} are upper triangular, non singular matrices and $\mathbf{Q}, \mathbf{P} \in \Re^{M \times M}$ are orthogonal matrices. The GLLSP (2.8) is then equivalent to

$$\underset{\tilde{v}, \hat{v}, \beta}{\text{argmin}} \ \left\| \begin{pmatrix} \tilde{v} \\ \hat{v} \end{pmatrix} \right\|^2 \quad \text{subject to} \quad \begin{pmatrix} \tilde{\mathbf{y}} \\ \hat{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix} \beta + \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{pmatrix} \begin{pmatrix} \tilde{v} \\ \hat{v} \end{pmatrix}, \quad (2.10)$$

where $\mathbf{P}^T v = (\tilde{v}^T \ \hat{v}^T)$. The second part of the partitioned constraint in (2.10) gives $\hat{v} = \mathbf{U}_{22}^{-1} \hat{\mathbf{y}}$. Setting $\tilde{v} = \mathbf{0}$, the estimator is obtained by solving the triangular system $\hat{\beta} = \mathbf{R}^{-1} (\tilde{\mathbf{y}} - \mathbf{U}_{12} \hat{v})$.

This implies that the GLS solution of the updated GLM (2.5) can be obtained from the GLLSP

$$\underset{v_1, v, \beta_2}{\text{argmin}} \ \left\| \begin{pmatrix} \imath v_1 \\ v \end{pmatrix} \right\|_h \quad \text{subject to} \quad \begin{pmatrix} \imath \mathbf{y}_1 \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X} \end{pmatrix} \beta_2 + \begin{pmatrix} \mathbf{C}_{11} & \imath \tilde{\mathbf{C}}_{12} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \imath v_1 \\ v \end{pmatrix}, \quad (2.11)$$

where $\|\mathbf{x}\|_h = \mathbf{x}^H \boldsymbol{\Phi} \mathbf{x}$ is the hyperbolic norm of a complex column vector \mathbf{x} (Rader and Steinhardt, 1986). It is assumed that the solution of the GLLSP (2.8) as well as the matrices \mathbf{X} and \mathbf{R} in (2.9a) are available. Utilizing the GQRD (2.9), yields

$$\operatorname{argmin}_{v_1, \hat{v}, \hat{v}, \beta_2} \left\| \begin{pmatrix} {}_i v_1 \\ \hat{v} \\ \hat{v} \end{pmatrix} \right\|_h \text{ subject to } \begin{pmatrix} {}_i y_1 \\ \hat{y} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} {}_i X_1 \\ R \\ 0 \end{pmatrix} \beta_2 + \begin{pmatrix} C_{11} & {}_i C_A & {}_i C_B \\ 0 & U_{11} & U_{12} \\ 0 & 0 & U_{22} \end{pmatrix} \begin{pmatrix} {}_i v_1 \\ \hat{v} \\ \hat{v} \end{pmatrix}, \quad (2.12)$$

where $(C_A \ C_B) = \tilde{C}_{12} P$. In (2.10), $\hat{v} = U_{22}^{-1} \hat{y}$ and thus (2.12) is reduced to

$$\operatorname{argmin}_{v_1, \hat{v}, \beta_2} \left\| \begin{pmatrix} {}_i v_1 \\ \hat{v} \end{pmatrix} \right\|_h \text{ subject to } \begin{pmatrix} {}_i \tilde{y}_1 \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} {}_i X_1 \\ R \end{pmatrix} \beta_2 + \begin{pmatrix} C_{11} & {}_i C_A \\ 0 & U_{11} \end{pmatrix} \begin{pmatrix} {}_i v_1 \\ \hat{v} \end{pmatrix}, \quad (2.13)$$

where $\tilde{y} = \hat{y} - U_{12} \hat{v}$ and $\tilde{y}_1 = y_1 - C_B \hat{v}$. For the solution of (2.13) consider the HQRD

$$\tilde{Q}^H \begin{pmatrix} {}_i X_1 & {}_i \tilde{y}_1 \\ R & \tilde{y} \end{pmatrix} = \begin{pmatrix} 0 & {}_i \hat{z} \\ R_2 & \tilde{z} \end{pmatrix} \begin{matrix} M_1 \\ K \end{matrix} \quad (2.14)$$

and the product

$$\tilde{Q}^H \begin{pmatrix} C_{11} & {}_i C_A \\ 0 & U_{11} \end{pmatrix} = \begin{pmatrix} D_{11} & {}_i D_{12} \\ {}_i D_{21} & D_{22} \end{pmatrix}. \quad (2.15)$$

Here \tilde{Q} is a $(M_1 + K) \times (M_1 + K)$ hypernormal matrix with respect to the signature matrix $\operatorname{diag}(-I_{M_1}, I_K)$ and is defined as the product of K hyperbolic Householder transformations (Rader and Steinhardt, 1986). The efficient computation of the HQRD has been studied and a block strategy which is based on the compact form of the hyperbolic matrices has been proposed (Yanev and Kontoghiorghes, 2008).

Now, premultiplying the constraints of (2.13) by \tilde{Q}^H yields

$$\operatorname{argmin}_{v_1, \hat{v}, \beta_2} \left\| \begin{pmatrix} {}_i v_1 \\ \hat{v} \end{pmatrix} \right\|_h \text{ subject to } \begin{pmatrix} {}_i \hat{z} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} 0 \\ R_2 \end{pmatrix} \beta_2 + \begin{pmatrix} D_{11} & {}_i D_{12} \\ {}_i D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} {}_i v_1 \\ \hat{v} \end{pmatrix}. \quad (2.16)$$

For the solution of (2.16), let the RQ decomposition (RQD) of $({}_i D_{11} \ D_{12})$ be given by

$$\begin{pmatrix} D_{11} & {}_i D_{12} \end{pmatrix} \tilde{P} = (\tilde{U}_{11} \ 0),$$

$$({}_i D_{21} \ D_{22}) \tilde{P} = ({}_i \tilde{U}_{21} \ \tilde{U}_{22})$$

and

$$\tilde{P}^H \begin{pmatrix} {}_i v_1 \\ \hat{v} \end{pmatrix} = \begin{pmatrix} {}_i \hat{v} \\ \tilde{v} \end{pmatrix},$$

where $\tilde{U}_{11} \in \mathfrak{R}^{M_1 \times M_1}$ is a non singular, upper triangular matrix and $\tilde{P} \in \mathfrak{S}^{(K+M_1) \times (K+M_1)}$ is a unitary matrix and is defined as the product of M_1 Householder transformations. Thus the GLLSP (2.16) is written as

$$\operatorname{argmin}_{\hat{v}, \tilde{v}, \beta_2} \left\| \begin{pmatrix} {}_i \hat{v} \\ \tilde{v} \end{pmatrix} \right\|_h \text{ subject to } \begin{pmatrix} {}_i \hat{z} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} 0 \\ R_2 \end{pmatrix} \beta_2 + \begin{pmatrix} \tilde{U}_{11} & 0 \\ {}_i \tilde{U}_{21} & \tilde{U}_{22} \end{pmatrix} \begin{pmatrix} {}_i \hat{v} \\ \tilde{v} \end{pmatrix}. \quad (2.17)$$

From the constraints in (2.17) it follows that $\hat{v} = \tilde{U}_{11}^{-1} \hat{z}$. Setting $\tilde{v} = 0$, the estimator is obtained by the solution of the triangular system $R_2 \hat{\beta}_2 = \tilde{z} + \tilde{U}_{21} \hat{v}$. The downdating strategy is summarized in Algorithm 2.1.

In many cases, it is possible that the variance covariance matrix is singular (Zyskind and Martin, 1969; Judge et al., 1985; Deaton, 1986; Bai and Shi, 2011; Holly et al., 2011). Therefore, consider downdating the GLM (1.1) by deleting the first M_1 observations when Ω is singular, in which case the GLS estimator in (1.2) does not exist, and let $\operatorname{rank}(\Omega) = m < M$. The GLLSP (2.8) gives the BLUE of the GLM where now $C \in \mathfrak{R}^{M \times m}$ and $v \sim (0, \sigma^2 I_m)$ (Kourouklis and Paige, 1981). In this case for the solution of the GLLSP (2.8), the QRD (2.9a) and the complete QRD of $Q_2^T C$ are required, that is,

$$G^T (Q_2^T C) \tilde{P} = \begin{pmatrix} 0 & S_{22} \\ 0 & 0 \end{pmatrix}, \quad (2.18)$$

Algorithm 2.1: Estimating the DGLM (2.5) after solving (2.8) using the GQRD (2.9)

1. Let $\tilde{\mathbf{y}}, \mathbf{R}, \mathbf{U}_{11}, \mathbf{P}$ and $\hat{\mathbf{v}}$ from the solution of the GLLSP (2.8).
2. Compute $(\mathbf{C}_A \ \mathbf{C}_B) = \tilde{\mathbf{C}}_{12}\mathbf{P}$.
3. Compute $\tilde{\mathbf{y}}_1 = \mathbf{y}_1 - \mathbf{C}_B\hat{\mathbf{v}}$ where $\hat{\mathbf{v}}$ is given by $\hat{\mathbf{v}} = \mathbf{U}_{22}^{-1}\tilde{\mathbf{y}}$.
4. Compute the HQRD (2.14) and the product in (2.15).
5. Compute the RQD $(\mathbf{D}_{11} \ \imath\mathbf{D}_{12})\tilde{\mathbf{P}} = (\tilde{\mathbf{U}}_{11} \ \mathbf{0})$.
6. Compute the product $(\imath\mathbf{D}_{21} \ \mathbf{D}_{22})\tilde{\mathbf{P}} = (\imath\tilde{\mathbf{U}}_{21} \ \tilde{\mathbf{U}}_{22})$.
7. Solve the triangular system $\mathbf{R}_2\hat{\boldsymbol{\beta}}_2 = \tilde{\mathbf{z}} + \tilde{\mathbf{U}}_{21}\hat{\mathbf{v}}$ for $\hat{\boldsymbol{\beta}}_2$.

where $\mathbf{S}_{22} \in \Re^{q \times q}$ is upper triangular and non singular, $q = \text{rank}(\mathbf{Q}_2^T\mathbf{C})$, $\mathbf{G} \in \Re^{(M-K) \times (M-K)}$ and $\bar{\mathbf{P}} \in \Re^{m \times m}$ are orthogonal. Now if $\mathbf{G}^T\hat{\mathbf{y}} = (\hat{\mathbf{y}}_1^T \ \hat{\mathbf{y}}_2^T)^T$ and $\mathbf{Q}_1^T\mathbf{C}\bar{\mathbf{P}} = (\mathbf{S}_{11} \ \mathbf{S}_{12})$, then the GLLSP is equivalent to

$$\underset{\tilde{\mathbf{u}}, \hat{\mathbf{u}}, \boldsymbol{\beta}}{\text{argmin}} \left\| \begin{pmatrix} \tilde{\mathbf{u}} \\ \hat{\mathbf{u}} \end{pmatrix} \right\|^2 \text{ subject to } \begin{pmatrix} \tilde{\mathbf{y}} \\ \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{0} & \mathbf{S}_{22} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{u}} \\ \hat{\mathbf{u}} \end{pmatrix}, \quad (2.19)$$

where $\bar{\mathbf{P}}^T\mathbf{v} = (\tilde{\mathbf{u}}^T \ \hat{\mathbf{u}}^T)^T$. It follows that $\hat{\mathbf{y}}_2 = \mathbf{0}$, otherwise the model is inconsistent, $\hat{\mathbf{u}} = \mathbf{S}_{22}^{-1}\hat{\mathbf{y}}_1$, and the BLUE is $\hat{\boldsymbol{\beta}} = \mathbf{R}^{-1}(\tilde{\mathbf{y}} - \mathbf{S}_{12}\hat{\mathbf{u}})$.

For the downdating problem, assume that the estimator for $\boldsymbol{\beta}$ has already been obtained from the solution of the GLLSP as it is formulated in (2.19). The partitioning in (1.3) and (1.4) for $\mathbf{y}, \mathbf{X}, \boldsymbol{\epsilon}$ and $\boldsymbol{\Omega}$ will be used, but now \mathbf{C} is partitioned as

$$\mathbf{C} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{B}_{22} \end{pmatrix},$$

so that $\mathbf{B}_{11} \in \Re^{M_1 \times m_1}$, $\mathbf{B}_{22} = \mathbf{C}_{22} \in \Re^{M_2 \times m_2}$ and $m_1 + m_2 = m$. The downdating of the GLM (1.1) with singular $\boldsymbol{\Omega}$, after previous computation have been efficiently utilized, is obtained from the solution of the GLLSP

$$\underset{\mathbf{u}_1, \tilde{\mathbf{u}}, \boldsymbol{\beta}_2}{\text{argmin}} \left\| \begin{pmatrix} \imath\mathbf{u}_1 \\ \tilde{\mathbf{u}} \end{pmatrix} \right\|_h \text{ subject to } \begin{pmatrix} \imath\tilde{\mathbf{y}}_1 \\ \tilde{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \imath\mathbf{X}_1 \\ \mathbf{R} \end{pmatrix} \boldsymbol{\beta}_2 + \begin{pmatrix} \mathbf{B}_{11} & \imath\mathbf{B}_1 \\ \mathbf{0} & \mathbf{S}_{11} \end{pmatrix} \begin{pmatrix} \imath\mathbf{u}_1 \\ \tilde{\mathbf{u}} \end{pmatrix},$$

where $(\mathbf{B}_1 \ \mathbf{B}_2) = (\mathbf{0} \ \mathbf{B}_{12})\bar{\mathbf{P}}$, $\bar{\mathbf{P}}$ is the orthogonal matrix from the complete QRD (2.18), $\tilde{\mathbf{y}} = \tilde{\mathbf{y}} - \mathbf{S}_{12}\hat{\mathbf{u}}$ and $\tilde{\mathbf{y}}_1 = \mathbf{y}_1 - \mathbf{B}_2\hat{\mathbf{u}}$. The solution of the latter GLLSP is analogous to that of (2.13) (Hadjiantoni, 2015).

Often, the deletion of observations (from a model) is accompanied by the addition of new observations. Consider the GLM (1.1) and assume that the first M_1 observations will be deleted while M_u new observations become available. Let the new observations, added to the GLM (1.1), be denoted by

$$\mathbf{y}_u = \mathbf{X}_u\boldsymbol{\beta}_u + \boldsymbol{\epsilon}_u, \quad \boldsymbol{\epsilon}_u \sim (\mathbf{0}, \sigma^2\boldsymbol{\Omega}_{uu})$$

where $\mathbf{y}_u, \boldsymbol{\epsilon}_u \in \Re^{M_u}$ and $\mathbf{X}_u \in \Re^{M_u \times K}$. It is assumed that $\boldsymbol{\epsilon}_u$ and $\boldsymbol{\epsilon}$ are correlated. The GLM to be estimated is given by

$$\begin{pmatrix} \mathbf{y}_u \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_u \\ \mathbf{X}_2 \end{pmatrix} \tilde{\boldsymbol{\beta}} + \begin{pmatrix} \boldsymbol{\epsilon}_u \\ \boldsymbol{\epsilon}_2 \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{\epsilon}_u \\ \boldsymbol{\epsilon}_2 \end{pmatrix} \sim \left(\mathbf{0}, \sigma^2 \begin{pmatrix} \boldsymbol{\Omega}_{uu} & \boldsymbol{\Omega}_{u2} \\ \boldsymbol{\Omega}_{2u} & \boldsymbol{\Omega}_{22} \end{pmatrix} \right). \quad (2.20)$$

Suppose that the GLM (1.1) has been estimated. The problem of estimating (2.20) is equivalent to estimate an updated GLM where the new observations are added and those to be excluded from the model are removed. Consider the up-downdated GLM (UDGLM) where the first M_1 observations of the GLM (1.1) are deleted and the M_u new observations are added, namely

$$\begin{pmatrix} \imath \mathbf{y}_1 \\ \mathbf{y}_u \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \imath \mathbf{X}_1 \\ \mathbf{X}_u \\ \mathbf{X} \end{pmatrix} \tilde{\boldsymbol{\beta}} + \begin{pmatrix} \imath \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_u \\ \boldsymbol{\epsilon} \end{pmatrix}, \quad \begin{pmatrix} \imath \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_u \\ \boldsymbol{\epsilon} \end{pmatrix} \sim (\mathbf{0}, \sigma^2 \tilde{\boldsymbol{\Omega}}). \quad (2.21)$$

Here $\tilde{\boldsymbol{\Omega}}$ is given by

$$\tilde{\boldsymbol{\Omega}} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{0} & \imath \tilde{\mathbf{C}}_{12} \\ \mathbf{0} & \mathbf{C}_{uu} & \tilde{\mathbf{C}}_{u2} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix} \hat{\boldsymbol{\Phi}} \begin{pmatrix} \mathbf{C}_{11} & \mathbf{0} & \imath \tilde{\mathbf{C}}_{12} \\ \mathbf{0} & \mathbf{C}_{uu} & \tilde{\mathbf{C}}_{u2} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{pmatrix}^H,$$

where $\tilde{\mathbf{C}}_{12} = (\mathbf{0} \ \mathbf{C}_{12})$, $\tilde{\mathbf{C}}_{u2} = (\mathbf{0} \ \mathbf{C}_{u2})$ and $\hat{\boldsymbol{\Phi}} = \text{diag}(-\mathbf{I}_{M_1}, \mathbf{I}_{M_u+M})$. The normal equations of the UDGLM (2.21) confirm that the GLS solution is the same as the one obtained from the GLM (2.20). The UDGLM (2.21) is then formulated to the equivalent GLLSP which is solved in a way similar to (2.11) (Hadjiantoni, 2015).

3 Seemingly Unrelated Regressions Model

The seemingly unrelated regressions (SUR) model is a special case of the GLM (1.1) and is defined as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \quad i = 1, \dots, G,$$

or in compact form as

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_G \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & & & \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ & & & \mathbf{X}_G \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_G \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_G \end{pmatrix}$$

or

$$\text{vec}(\mathbf{Y}) = (\oplus_{i=1}^G \mathbf{X}_i) \boldsymbol{\beta} + \text{vec}(\mathbf{E}), \quad (3.1)$$

where $\mathbf{y}_i \in \mathfrak{R}^M$ are the response vectors, $\mathbf{X}_i \in \mathfrak{R}^{M \times k_i}$ are the exogenous matrices with full column rank, $\boldsymbol{\beta}_i \in \mathfrak{R}^{k_i}$ are the vectors of parameters to be estimated and $\mathbf{e}_i \in \mathfrak{R}^M$ are the disturbance vectors (Zellner, 1962; Telser, 1964; Srivastava and Dwivedi, 1979; Srivastava and Giles, 1987). For the error terms we have that $\mathbb{E}(\mathbf{e}_i) = \mathbf{0}$, $\mathbb{E}(\mathbf{e}_i \mathbf{e}_i^T) = \sigma_{ii} \mathbf{I}_M$, $\mathbb{E}(\mathbf{e}_i \mathbf{e}_j^T) = \sigma_{ij} \mathbf{I}_M$, that is, the disturbances are contemporaneously correlated across regressions. Therefore, $\mathbf{Y} = (\mathbf{y}_1 \ \dots \ \mathbf{y}_G)$, $\oplus_{i=1}^G \mathbf{X}_i = \text{diag}(\mathbf{X}_1 \ \dots \ \mathbf{X}_G)$ where \oplus stands for direct sum, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T \ \dots \ \boldsymbol{\beta}_G^T)^T$, $\mathbf{E} = (\mathbf{e}_1 \ \dots \ \mathbf{e}_G)$ and vec is the column stack operator. The disturbance term $\text{vec}(\mathbf{E}) \sim (\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_M)$ where $\boldsymbol{\Sigma} = [\sigma_{ij}] \in \mathfrak{R}^{G \times G}$ is a symmetric positive definite matrix and \otimes denotes the Kronecker product. Note that $\oplus_{i=1}^G$ will be abbreviated by \oplus_i for notational convenience and $\{\cdot\}$ will be used to denote a set of vectors. The numerical and computational strategies for modifying the SUR model after adding observations have been previously investigated (Kontoghiorghes and Clarke, 1995; Foschi and Kontoghiorghes, 2002; Foschi et al., 2003). That is, the GLLSP (2.8) now corresponds to

$$\underset{\mathbf{V}, \boldsymbol{\beta}}{\text{argmin}} \|\mathbf{V}\|_F^2 \quad \text{subject to } \text{vec}(\mathbf{Y}) = (\oplus_i \mathbf{X}_i) \boldsymbol{\beta} + (\mathbf{C} \otimes \mathbf{I}_M) \text{vec}(\mathbf{V}), \quad (3.2)$$

where $\boldsymbol{\Sigma} = \mathbf{C} \mathbf{C}^T$, \mathbf{C} is the upper triangular Cholesky factor, $\mathbf{V} \sim (\mathbf{0}, \mathbf{I}_{GM})$ is defined to give $\mathbf{E} = \mathbf{V} \mathbf{C}^T$ and $\|\cdot\|_F$ denotes the Frobenius norm (Kontoghiorghes, 2000). For the solution of the GLLSP (3.2), the GQRD of $\oplus_i \mathbf{X}_i$ and $\mathbf{C} \otimes \mathbf{I}_M$, used in the solution of the latter, is given by

$$\mathbf{Q}^T (\oplus_i \mathbf{X}_i) = \begin{pmatrix} \oplus_i \mathbf{R}_i \\ \mathbf{0} \end{pmatrix} \begin{matrix} K \\ GM - K \end{matrix}, \quad (3.3a)$$

$$\mathbf{Q}^T (\mathbf{C} \otimes \mathbf{I}_M) \mathbf{P} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{pmatrix} \begin{matrix} K \\ GM - K, \end{matrix} \quad (3.3b)$$

where $\mathbf{R}_i \in \mathfrak{R}^{k_i \times k_i}$ is the upper triangular factor of the QRD of \mathbf{X}_i , namely

$$\mathbf{Q}_i^T (\mathbf{X}_i \ \mathbf{y}_i) = \begin{pmatrix} \mathbf{R}_i & \tilde{\mathbf{y}}_i \\ \mathbf{0} & \hat{\mathbf{y}}_i \end{pmatrix} \begin{matrix} k_i \\ M - k_i, \end{matrix}$$

$\mathbf{Q} \in \mathfrak{R}^{GM \times GM}$, $\mathbf{Q}_i \in \mathfrak{R}^{M \times M}$ are orthogonal and $K = \sum_{i=1}^G k_i$. The GLLSP (3.2) is then equivalent to

$$\underset{\tilde{\mathbf{v}}_i, \hat{\mathbf{v}}_i, \boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^G (\|\tilde{\mathbf{v}}_i\|^2 + \|\hat{\mathbf{v}}_i\|^2) \quad \text{subject to} \quad (3.4)$$

$$\begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{y}}_i\}) \\ \operatorname{vec}(\{\hat{\mathbf{y}}_i\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_i \\ \mathbf{0} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_i\}) \\ \operatorname{vec}(\{\hat{\mathbf{v}}_i\}) \end{pmatrix},$$

where $\mathbf{P}^T \operatorname{vec}(\mathbf{V}) = (\operatorname{vec}(\{\tilde{\mathbf{v}}_i\})^T \ \operatorname{vec}(\{\hat{\mathbf{v}}_i\})^T)^T$. The second constraint in (3.4) gives $\operatorname{vec}(\{\hat{\mathbf{v}}_i\}) = \mathbf{U}_{22}^{-1} \operatorname{vec}(\{\hat{\mathbf{y}}_i\})$; $\operatorname{vec}(\{\tilde{\mathbf{v}}_i\})$ is set to zero in order to minimize the argument in (3.4), and the estimator is obtained from the solution of the triangular system $(\oplus_i \mathbf{R}_i) \hat{\boldsymbol{\beta}} = \operatorname{vec}(\{\tilde{\mathbf{y}}_i\}) - \mathbf{U}_{12} \operatorname{vec}(\{\hat{\mathbf{v}}_i\})$.

3.1 DOWNDATING THE SUR MODEL

The problem of deleting observations from the SUR model is now considered. Assume that an equal number of observations, say the last d observations, are removed from every regression and let \mathbf{y}_i , \mathbf{X}_i and \mathbf{e}_i in (3.1) be partitioned as

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_i^* \\ \mathbf{y}_i^{(d)} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} \mathbf{X}_i^* \\ \mathbf{X}_i^{(d)} \end{pmatrix}, \quad \mathbf{e}_i = \begin{pmatrix} \mathbf{e}_i^* \\ \mathbf{e}_i^{(d)} \end{pmatrix},$$

where $\mathbf{X}_i^* \in \mathfrak{R}^{M^* \times k_i}$, $\mathbf{X}_i^{(d)} \in \mathfrak{R}^{d \times k_i}$ for $i = 1, \dots, G$ and $M = M^* + d$. Note that the observations in the SUR model can be reordered and thus (3.1) is expressed in the equivalent form:

$$\begin{pmatrix} \operatorname{vec}(\mathbf{Y}^*) \\ \operatorname{vec}(\mathbf{Y}^{(d)}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_i^* \\ \oplus_i \mathbf{X}_i^{(d)} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \operatorname{vec}(\mathbf{E}^*) \\ \operatorname{vec}(\mathbf{E}^{(d)}) \end{pmatrix}, \quad \begin{pmatrix} \operatorname{vec}(\mathbf{E}^*) \\ \operatorname{vec}(\mathbf{E}^{(d)}) \end{pmatrix} \sim \left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Sigma} \otimes \mathbf{I}_{M^*} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \otimes \mathbf{I}_d \end{pmatrix} \right). \quad (3.5)$$

The downdated SUR (DSUR) model to be estimated is then given by

$$\mathbf{y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta}_i^* + \mathbf{e}_i^*, \quad \mathbf{e}_i^* \sim (\mathbf{0}, \sigma_{ii} \mathbf{I}_{M^*}), \quad i = 1, \dots, G$$

or by the equivalent compact form

$$\operatorname{vec}(\mathbf{Y}^*) = (\oplus_i \mathbf{X}_i^*) \boldsymbol{\beta}^* + \operatorname{vec}(\mathbf{E}^*), \quad \operatorname{vec}(\mathbf{E}^*) \sim (\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_{M^*}). \quad (3.6)$$

Analogous to the method for the downdating of the GLM, the DSUR model (3.6) is equivalent to

$$\begin{pmatrix} \operatorname{vec}(\mathbf{Y}) \\ \operatorname{vec}(\imath \mathbf{Y}^{(d)}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_i \\ \oplus_i \imath \mathbf{X}_i^{(d)} \end{pmatrix} \boldsymbol{\beta}^* + \begin{pmatrix} \operatorname{vec}(\mathbf{E}) \\ \operatorname{vec}(\imath \mathbf{E}^{(d)}) \end{pmatrix}, \quad \begin{pmatrix} \operatorname{vec}(\mathbf{E}) \\ \operatorname{vec}(\imath \mathbf{E}^{(d)}) \end{pmatrix} \sim \left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Sigma} \otimes \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \otimes (-\mathbf{I}_d) \end{pmatrix} \right).$$

Now using the partitioning in (3.5) for (3.1) and following the same procedure as for the downdating of the GLM (see section 2), it can be shown that the latter gives the same GLS solution as the DSUR model (3.6). Furthermore its GLS solution can be obtained from the solution of the corresponding GLLSP, namely

$$\underset{\mathbf{v}, \mathbf{V}^{(d)}, \boldsymbol{\beta}^*}{\operatorname{argmin}} \|\mathbf{V}\|_F^2 - \|\mathbf{V}^{(d)}\|_F^2 \quad \text{subject to} \quad (3.7)$$

$$\begin{pmatrix} \operatorname{vec}(\mathbf{Y}) \\ \operatorname{vec}(\imath \mathbf{Y}^{(d)}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{X}_i \\ \oplus_i \imath \mathbf{X}_i^{(d)} \end{pmatrix} \boldsymbol{\beta}^* + \begin{pmatrix} \mathbf{C} \otimes \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \otimes \mathbf{I}_d \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\mathbf{V}) \\ \operatorname{vec}(\imath \mathbf{V}^{(d)}) \end{pmatrix},$$

where \mathbf{C} is the upper triangular Cholesky factor such that $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^T$ as in (3.2). Using the GQRD (3.3) the latter GLLSP becomes

$$\operatorname{argmin}_{\tilde{\mathbf{v}}_i, \hat{\mathbf{v}}_i, \mathbf{V}^{(d)}, \beta^*} \sum_{i=1}^G (\|\tilde{\mathbf{v}}_i\|^2 + \|\hat{\mathbf{v}}_i\|^2) - \|\mathbf{V}^{(d)}\|_F^2 \text{ subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{y}}_i\}) \\ \operatorname{vec}(\{\hat{\mathbf{y}}_i\}) \\ \operatorname{vec}({}_i\mathbf{Y}^{(d)}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_i \\ \mathbf{0} \\ \oplus_i {}_i\mathbf{X}_i^{(d)} \end{pmatrix} \beta^* + \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \otimes \mathbf{I}_d \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_i\}) \\ \operatorname{vec}(\{\hat{\mathbf{v}}_i\}) \\ \operatorname{vec}({}_i\mathbf{V}^{(d)}) \end{pmatrix}$$

which reduces to the equivalent GLLSP

$$\operatorname{argmin}_{\tilde{\mathbf{v}}_i, \mathbf{V}^{(d)}, \beta^*} \sum_{i=1}^G \|\tilde{\mathbf{v}}_i\|^2 - \|\mathbf{V}^{(d)}\|_F^2 \text{ subject to}$$

$$\begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{y}}_i\}) \\ \operatorname{vec}({}_i\mathbf{Y}^{(d)}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_i \\ \oplus_i {}_i\mathbf{X}_i^{(d)} \end{pmatrix} \beta^* + \begin{pmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \otimes \mathbf{I}_d \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_i\}) \\ \operatorname{vec}({}_i\mathbf{V}^{(d)}) \end{pmatrix}. \quad (3.8)$$

The solution of (3.8) is obtained by the HQRD

$$\tilde{\mathbf{Q}}^H \begin{pmatrix} \oplus_i \mathbf{R}_i \\ \oplus_i {}_i\mathbf{X}_i^{(d)} \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_i^* \\ \mathbf{0} \end{pmatrix} \begin{matrix} K \\ Gd \end{matrix} \quad (3.9)$$

and the RQD

$$\left(\tilde{\mathbf{Q}}^H \begin{pmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \otimes \mathbf{I}_d \end{pmatrix} \right) \tilde{\mathbf{P}} = \begin{pmatrix} \tilde{\mathbf{U}}_{11} & {}_i\tilde{\mathbf{U}}_{12} \\ \mathbf{0} & \tilde{\mathbf{U}}_{22} \end{pmatrix}, \quad (3.10)$$

where $\tilde{\mathbf{Q}}$ is a hyperbolic matrix with respect to the signature matrix $\operatorname{diag}(\mathbf{I}_K, -\mathbf{I}_{Gd})$ and $\tilde{\mathbf{P}}$ is a unitary matrix of order $\mathbf{I}_{Gd} + \mathbf{I}_K$, \mathbf{R}_i^* , $i = 1, \dots, G$, are the upper triangular and non singular factors obtained when the QRD of $\oplus_i \mathbf{X}_i^*$ is computed, and $\tilde{\mathbf{U}}_{22}$ is an upper triangular and non singular matrix. Then, the GLLSP (3.8) is equivalent to

$$\operatorname{argmin}_{\tilde{\mathbf{u}}_i, \hat{\mathbf{u}}_i, \beta^*} \sum_{i=1}^G \left\| \begin{pmatrix} \tilde{\mathbf{v}}_i \\ {}_i\hat{\mathbf{v}}_i \end{pmatrix} \right\|_h \text{ subject to } \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{z}}_i\}) \\ \operatorname{vec}(\{{}_i\hat{\mathbf{z}}_i\}) \end{pmatrix} = \begin{pmatrix} \oplus_i \mathbf{R}_i^* \\ \mathbf{0} \end{pmatrix} \beta^* + \begin{pmatrix} \tilde{\mathbf{U}}_{11} & {}_i\tilde{\mathbf{U}}_{12} \\ \mathbf{0} & \tilde{\mathbf{U}}_{22} \end{pmatrix} \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_i\}) \\ \operatorname{vec}(\{{}_i\hat{\mathbf{v}}_i\}) \end{pmatrix},$$

where

$$\tilde{\mathbf{Q}}^H \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{y}}_i\}) \\ \operatorname{vec}({}_i\mathbf{Y}^{(d)}) \end{pmatrix} = \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{z}}_i\}) \\ \operatorname{vec}(\{{}_i\hat{\mathbf{z}}_i\}) \end{pmatrix} \text{ and } \tilde{\mathbf{P}}^H \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_i\}) \\ {}_i\operatorname{vec}(\mathbf{V}^{(d)}) \end{pmatrix} = \begin{pmatrix} \operatorname{vec}(\{\tilde{\mathbf{v}}_i\}) \\ \operatorname{vec}(\{\hat{\mathbf{v}}_i\}) \end{pmatrix}.$$

The solution of the latter GLLSP is similar to that of the GLLSP (3.4). The same formulation holds in (3.1) for non singular and singular dispersion matrix $\mathbf{\Sigma}$. Here, only the estimation of the SUR model and the DSUR when the dispersion matrix is non singular is presented. The method can be extended for singular $\mathbf{\Sigma}$ as in the case of the GLM in section 2.1. The steps for the downdating method of the SUR model are summarized in Algorithm 3.1.

Algorithm 3.1: Estimating the DSUR model (3.6) after solving (3.2) using the GQRD (3.3)

1. Let $\tilde{\mathbf{y}}_i$, \mathbf{R}_i , $i = 1, \dots, G$ and \mathbf{U}_{11} from the solution of the GLLSP (3.2).
2. Compute the HQRD (3.9).
3. Compute the RQD (3.10).
4. Solve the triangular system $\operatorname{vec}(\{\hat{\mathbf{z}}_i\}) = \tilde{\mathbf{U}}_{22} \operatorname{vec}(\{\hat{\mathbf{v}}_i\})$ for $\operatorname{vec}(\{\hat{\mathbf{v}}_i\})$.
5. Solve the triangular system $(\oplus_i \mathbf{R}_i^*) \hat{\beta}^* = \operatorname{vec}(\{\hat{\mathbf{z}}_i\}) + \tilde{\mathbf{U}}_{12} \operatorname{vec}(\{\hat{\mathbf{v}}_i\})$ for $\hat{\beta}^*$.

3.2 DOWNDATING THE UNBALANCED SUR MODEL

Consider the case where an arbitrary number of observations has been deleted from an arbitrary number of regressions in the SUR model (3.1) which has resulted in the unbalanced SUR (USUR) model

$$\check{\mathbf{y}}_j = \check{\mathbf{X}}_j \check{\boldsymbol{\beta}}_j + \check{\mathbf{e}}_j, \quad j = 1, \dots, G, \quad (3.11)$$

where $\check{\mathbf{X}}_j \in \mathfrak{R}^{t_j \times k_j}$, $\check{\mathbf{y}}_j, \check{\mathbf{e}}_j \in \mathfrak{R}^{t_j}$, $\check{\mathbf{e}}_j \sim (\mathbf{0}, \sigma_{jj}^2 \mathbf{I}_{t_j})$ and $\mathbb{E}(\check{\mathbf{e}}_i \check{\mathbf{e}}_j^T) = \sigma_{ij} (\mathbf{I}_{t_i} \mathbf{0}_{t_i \times (t_j - t_i)})$ for $t_i < t_j$. Assume that $t_i < t_j$ for $i < j$, $i, j = 1, \dots, G$ and let $\tau = \sum_{j=1}^G t_j$. The estimation of the USUR model using the method of GLLSP has been previously considered in Foschi and Kontoghiorghe (2002). Specifically, in order to apply the GLLSP approach for the USUR model in (3.11) the observations of each regression are partitioned and reordered according to

$$\check{\mathbf{y}}_i = \begin{pmatrix} \check{\mathbf{y}}_{1i} \\ \check{\mathbf{y}}_{2i} \\ \vdots \\ \check{\mathbf{y}}_{ii} \end{pmatrix} \begin{matrix} m_1 \\ m_2 \\ \vdots \\ m_i \end{matrix}, \quad \check{\mathbf{X}}_i = \begin{pmatrix} \check{\mathbf{X}}_{1i} \\ \check{\mathbf{X}}_{2i} \\ \vdots \\ \check{\mathbf{X}}_{ii} \end{pmatrix} \begin{matrix} m_1 \\ m_2 \\ \vdots \\ m_i \end{matrix}, \quad \check{\mathbf{e}}_i = \begin{pmatrix} \check{\mathbf{e}}_{1i} \\ \check{\mathbf{e}}_{2i} \\ \vdots \\ \check{\mathbf{e}}_{ii} \end{pmatrix} \begin{matrix} m_1 \\ m_2 \\ \vdots \\ m_i \end{matrix},$$

where $m_1 = t_1$ and $m_i = t_i - t_{i-1}$, $i = 2, \dots, G$. Now the USUR model (3.11) is reformulated as

$$\begin{pmatrix} \text{vec}(\{\check{\mathbf{y}}_{1j}\}) \\ \vdots \\ \text{vec}(\{\check{\mathbf{y}}_{ij}\}) \\ \vdots \\ \text{vec}(\{\check{\mathbf{y}}_{Gj}\}) \end{pmatrix} = \begin{pmatrix} (0_{\kappa_1} \oplus_j \check{\mathbf{X}}_{1j}) \\ \vdots \\ (0_{\kappa_i} \oplus_j \check{\mathbf{X}}_{ij}) \\ \vdots \\ (0_{\kappa_G} \oplus_j \check{\mathbf{X}}_{Gj}) \end{pmatrix} \check{\boldsymbol{\beta}} + \begin{pmatrix} \text{vec}(\{\check{\mathbf{e}}_{1j}\}) \\ \vdots \\ \text{vec}(\{\check{\mathbf{e}}_{ij}\}) \\ \vdots \\ \text{vec}(\{\check{\mathbf{e}}_{Gj}\}) \end{pmatrix}, \quad j = i, \dots, G, \quad (3.12)$$

or in compact form as

$$\check{\mathbf{y}} = \check{\mathbf{X}} \check{\boldsymbol{\beta}} + \check{\mathbf{e}},$$

where $\kappa_1 = 0$ and $\kappa_j = \sum_{l=1}^{j-1} k_l$, $j = 2, \dots, G$. The disturbance term in (3.12) has zero mean and dispersion matrix $\check{\boldsymbol{\Omega}}$ which is block diagonal with entries $\boldsymbol{\Sigma}_{i:G, i:G} \otimes \mathbf{I}_{m_i}$, where $\boldsymbol{\Sigma}_{i:G, i:G}$ is the $(G - i + 1) \times (G - i + 1)$ submatrix of $\boldsymbol{\Sigma}$ starting at position (i, i) , $i = 1, \dots, G$. The corresponding GLLSP is given by

$$\underset{\check{\mathbf{v}}, \check{\boldsymbol{\beta}}}{\text{argmin}} \|\check{\mathbf{v}}\|^2 \quad \text{subject to } \check{\mathbf{y}} = \check{\mathbf{X}} \check{\boldsymbol{\beta}} + \check{\mathbf{C}} \check{\mathbf{v}}, \quad (3.13)$$

where $\check{\mathbf{e}} = \check{\mathbf{C}} \check{\mathbf{v}}$ and $\check{\mathbf{v}} \sim (\mathbf{0}, \mathbf{I}_\tau)$. Note that $\check{\mathbf{C}}$ is block diagonal with entry to the i th block, the upper triangular matrix $\mathbf{C}_{i:G, i:G}$ such that $\check{\boldsymbol{\Omega}} = \check{\mathbf{C}} \check{\mathbf{C}}^T$ (Foschi and Kontoghiorghe, 2002). The solution of the GLLSP (3.13) is obtained from the GQRD of $\check{\mathbf{X}}$ and $\check{\mathbf{C}}$.

Assume now that further observations are to be deleted from the USUR model (3.11). Let the $\check{\mathbf{y}}_i$, $\check{\mathbf{X}}_i$ and $\check{\mathbf{e}}_i$ in (3.11) be partitioned as

$$\check{\mathbf{y}}_i = \begin{pmatrix} \check{\mathbf{y}}_i^* \\ \check{\mathbf{y}}_i^{(d)} \end{pmatrix}, \quad \check{\mathbf{X}}_i = \begin{pmatrix} \check{\mathbf{X}}_i^* \\ \check{\mathbf{X}}_i^{(d)} \end{pmatrix}, \quad \check{\mathbf{e}}_i = \begin{pmatrix} \check{\mathbf{e}}_i^* \\ \check{\mathbf{e}}_i^{(d)} \end{pmatrix}, \quad (3.14)$$

where $\check{\mathbf{X}}_i^* \in \mathfrak{R}^{(t_i - d_i) \times k_i}$ are the remaining observations in the i th regression and $\check{\mathbf{X}}_i^{(d)} \in \mathfrak{R}^{d_i \times k_i}$ are the observations to be deleted from the i th regression. Note that $d_i \geq d_j \geq 0$ for $i < j$, $i, j = 1, \dots, G$ and let $\delta = \sum_{i=1}^G d_i$. Suppose that the system of the deleted observations, as defined by (3.14), has been reformulated as in (3.12); it is given by

$$\check{\mathbf{y}}^{(d)} = \check{\mathbf{X}}^{(d)} \check{\boldsymbol{\beta}}^{(d)} + \check{\mathbf{e}}^{(d)},$$

where $\check{\mathbf{e}}^{(d)}$ has zero mean and dispersion matrix $\check{\boldsymbol{\Omega}}^{(d)}$. Then, the downdating USUR model is equivalent to the updated USUR model

$$\begin{pmatrix} \check{\mathbf{y}} \\ \imath \check{\mathbf{y}}^{(d)} \end{pmatrix} = \begin{pmatrix} \check{\mathbf{X}} \\ \imath \check{\mathbf{X}}^{(d)} \end{pmatrix} \check{\boldsymbol{\beta}}^* + \begin{pmatrix} \check{\mathbf{e}} \\ \imath \check{\mathbf{e}}^{(d)} \end{pmatrix}, \quad \begin{pmatrix} \check{\mathbf{e}} \\ \imath \check{\mathbf{e}}^{(d)} \end{pmatrix} \sim \left(\mathbf{0}, \begin{pmatrix} \check{\boldsymbol{\Omega}} & \mathbf{0} \\ \mathbf{0} & -\check{\boldsymbol{\Omega}}^{(d)} \end{pmatrix} \right). \quad (3.15)$$

The SUR model (3.15) is thus equivalent to the GLLSP

$$\underset{\check{\mathbf{v}}, \check{\mathbf{v}}^{(d)}, \check{\boldsymbol{\beta}}^*}{\operatorname{argmin}} \left\| \begin{pmatrix} \check{\mathbf{v}} \\ \imath \check{\mathbf{v}}^{(d)} \end{pmatrix} \right\|_h \quad \text{subject to} \quad \begin{pmatrix} \check{\mathbf{y}} \\ \imath \check{\mathbf{y}}^{(d)} \end{pmatrix} = \begin{pmatrix} \check{\mathbf{X}} \\ \imath \check{\mathbf{X}}^{(d)} \end{pmatrix} \check{\boldsymbol{\beta}}^* + \begin{pmatrix} \check{\mathbf{C}} & \mathbf{0} \\ \mathbf{0} & \check{\mathbf{C}}^{(d)} \end{pmatrix} \begin{pmatrix} \check{\mathbf{v}} \\ \imath \check{\mathbf{v}}^{(d)} \end{pmatrix}, \quad (3.16)$$

where $\check{\mathbf{v}}^{(d)} \sim (\mathbf{0}, \mathbf{I}_\delta)$ is defined such that $\check{\mathbf{e}}^{(d)} = \check{\mathbf{C}}^{(d)} \check{\mathbf{v}}^{(d)}$, $\check{\boldsymbol{\Omega}}^{(d)} = \check{\mathbf{C}}^{(d)} (\check{\mathbf{C}}^{(d)})^T$. Previous computations from the solution of (3.13) are utilized. The solution of (3.16) follows analogously to that of (3.7) and the algorithm for the estimation of (3.15) is similar to Algorithm 3.1.

4 Computational Results

When modifying a model by either deleting and/or adding observations or regressors the main aim is to utilize previous computations in order to minimize the number of arithmetic operations and storage locations (Björck, 1996). Therefore, in order to examine the computational efficiency of Algorithm 2.1 compared to other algorithms that solve the same problem afresh, its computational complexity has been studied. Moreover, the development of efficient algorithms requires that the special structure of the matrices be exploited (Yanev and Kontoghiorghes, 2008). Within this context, a strategy for the efficient execution of Algorithm 2.1 has been developed.

For the analysis and comparison of the computational results, four algorithms, herein referred to as Alg. 1, Alg. 1b, Alg. 2 and Alg. 2b, have been considered. They are:

Alg. 1: Solves afresh the problem by computing the GQRD (2.9) and solving the triangular system $\hat{\boldsymbol{\beta}} = \mathbf{R}^{-1} (\tilde{\mathbf{y}} - \mathbf{U}_{12} \hat{\mathbf{v}})$.

Alg. 1b: This is Alg. 1 but using a block recursive strategy.

Alg. 2: This is Algorithm 2.1, which is given in section 2.1.

Alg. 2b: This is Algorithm 2.1 but with the special structure of the matrices in steps 2-5 being exploited.

The theoretical complexities in terms of flops of the four algorithms have been computed in accordance with Golub and Van Loan (1996). Their order of complexity is given approximately by $4/3g^3K^3$, $2gM_1K^2$, $4n^2K^3$ and $nK^3(5n + 2g)$, respectively, where $g = M/K$ and $n = M_1/K$ when M_1 is large. In particular, the case where the number of the deleted observations M_1 is smaller than the regressors K and also the instance where M_1 is a multiple of K have been examined in detail. Specifically when $M_1 < K$, Alg. 1 estimates the model by solving the equivalent GLLSP and is compared with Alg. 2 which employs directly the downdating strategy. When $M_1 = nK$, Alg. 1b and Alg. 2b can be employed. Hence in this case, Alg. 1b is used and its results are compared with Alg. 2b which applies the same strategy but is based on the proposed downdating method, i.e. Algorithm 2.1. In the case where $M_1 < K$, the complexity analysis indicates that Alg. 2 is more efficient when $M_1 < 2/3g^2K$. In addition, when $M_1 = nK$, Alg. 2b, performs better when $h > n(1 + \sqrt{21})/4$, where $h = M_2/K$.

In order to validate further the proposed method, a series of experiments based on synthetic data have been carried out. Table 1 shows the execution times, in CPU seconds, of the four algorithms mentioned above for models with the initial number of observations $M = 2500, 5000$, $K = 50, 100$ and for various values of M_1 . The times presented are the average times after solving the same problem 20 times. The execution times show that Algorithm 2.1 for downdating the GLM outperforms existing algorithms which solve the same problem afresh. Furthermore, Figure 1 illustrates the computational

efficiency of the up-downdating strategy, herein referred to as Alg. 3, (see Algorithm 2.2.2 in Hadjiantoni (2015)), when compared with Alg. 1 which solves the same problem afresh, by presenting the ratio of the execution times of the two algorithms. In particular, the execution times of models with initial number of observations $M = 2500, 5000$ and $K = 50, 100, 250, 500$ and different values of M_1 and M_u have been obtained. It is assumed that the same number of observations has been added and deleted simultaneously. The results show that Alg. 3 (up-downdating strategy) outperforms existing algorithms. Observe that the performance of Alg. 3 is significantly higher the smaller is the increment of the rolling window is and also note that the efficiency decreases as the number of parameters in the model increases.

Table 1: Execution times in seconds of the downdating algorithms for the GLM

M_1	K	M_2	Alg. 1			M_1	K	M_2	Alg. 1b		
			afresh	recursive	$\frac{\text{Alg. 1}}{\text{Alg. 2}}$				afresh	recursive	$\frac{\text{Alg. 1b}}{\text{Alg. 2b}}$
1	50	2499	281	7	4071	200	50	2300	10	20	49
25	50	2495	279	8	3508	500	50	2000	7	54	13
50	50	2450	275	9	2993	1000	50	1500	4	141	3
1	100	2499	305	8	4068	200	100	2300	16	37	43
25	100	2495	280	9	2999	500	100	2000	12	108	11
50	100	2450	271	12	2196	1000	100	1500	7	292	2
1	50	4999	2462	29	8548	500	50	4500	55	88	63
25	50	4995	2387	30	7956	1000	50	4000	41	185	22
50	50	4950	2402	32	7505	1500	50	3500	30	331	9
1	100	4999	2489	30	8423	500	100	4500	76	153	50
25	100	4995	2440	32	7565	1000	100	4000	58	365	16
50	100	4950	2398	35	6764	1500	100	3500	43	686	6

Alg. 1 and Alg. 1b solve the problem afresh (Alg. 1b uses a block recursive strategy).

Alg. 2 is Algorithm 2.1 and Alg. 2b is Algorithm 2.1 but the structure of the matrices is exploited.

The times for Alg. 2 and Alg. 2b have been multiplied by 10^2 .

The computational efficiency of the proposed algorithms for the downdating of the SUR and the USUR models has also been examined using macroeconomic data applied to a vector autoregressive (VAR) process. A VAR process $\mathbf{y}_t \in \mathfrak{R}^p$ with q lags is formulated as a SUR model where the dependent variables are the vectors \mathbf{y}_t and the explanatory variables, which are the same in all equations, are given by the matrix $(\mathbf{1} \ \mathbf{y}_{t-1}^T \ \dots \ \mathbf{y}_{t-q}^T)$, and can be efficiently estimated by applying OLS equation by equation (Lütkepohl, 2007). However in the case where zero coefficient constraints are imposed, that is a subset VAR model is to be estimated, the method of GLS needs to be applied (Zellner, 1962; Lütkepohl, 2007).

The data set used here has been originally employed in Lee (1992) and then in Wang (2010). It consists of monthly data of four variables: real stock returns, real interest rates, industrial production growth and inflation rates spanning from July 1926 to December 2000. Herein, the most probable fitted subset VAR model obtained from Wang (2010) in a study for sparse SUR models for the period January 1947 to December 1987 is used. Within this context, leave-one-out experiments when the influence of each observation separately is of concern have been executed and also rolling window estimation for various window sizes has been implemented. For the window estimation of the SUR model, an up-downdating of the model is performed. For the analysis of the computational results in the case of the SUR model, the following four algorithms have been considered:

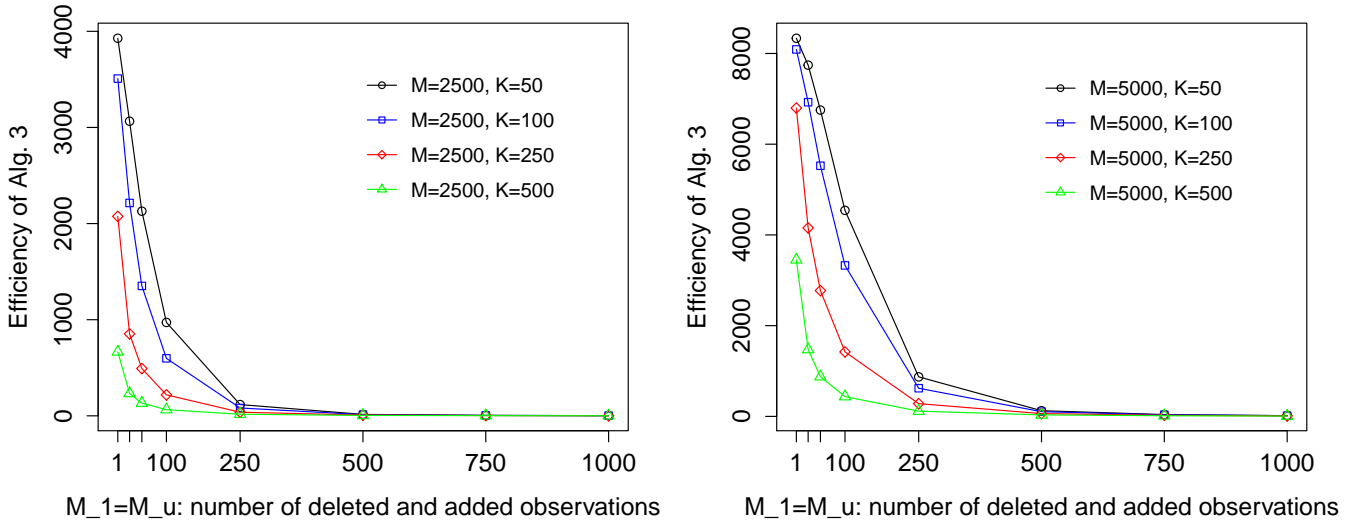


Figure 1: Computational efficiency (execution times ratio) of Alg. 3 (up-dating strategy) when compared with Alg. 1 which solves the same problem afresh. Results for models with $M = 2500; 5000$, $K = 50; 100, 250; 500$ and $M_1 = M_u = 1; 25; 50; 100; 250; 500; 750; 1000$ are presented.

Alg. 4: Solves afresh the problem by computing the GQRD (3.3) and solving the triangular system $(\oplus_i \mathbf{R}_i) \hat{\beta} = \text{vec}(\{\tilde{\mathbf{y}}_i\}) - \mathbf{U}_{12} \text{vec}(\{\hat{\mathbf{v}}_i\})$.

Alg. 4b: Solves afresh the problem by applying the method of Foschi and Kontoghiorghes (2002) for the USUR model after deleting a single observation.

Alg. 5: This is Algorithm 3.1, which is given in section 3.1.

Alg. 5b: This is Algorithm 3.1 which is given in section 3.1 applied for the downdating of the USUR model as presented in section 3.2.

Alg. 6: This is the extended version of Algorithm 3.1 to simultaneously add and delete observations from the model similarly to Alg. 2.2.2 in Hadjiantoni (2015).

Table 2 presents the results of the leave-one-out experiments. The performance of the algorithms is demonstrated using different sample sizes of the data set. Specifically, the case where one observation is deleted from every regression in the model and also the instance where one observation is deleted from one of the regressions have been considered. The first case yields M possibilities equal to the number of observations in each regression of the model whereas the second case gives GM possibilities that correspond to the number of regressions G times the number of observations M in each regression of the model. In the second case an USUR model needs to be estimated. In Table 2 the total times required by Alg. 4 and Alg. 4b and the proposed downdating Alg. 5 and Alg. 5b in order to complete the leave one out experiments when the model consists of $M = 250, 500, 750$ observations are reported. The results show that Alg. 5 and Alg. 5b outperform Alg. 4 and Alg. 4b which solve the same problem afresh. In addition, Figure 2 demonstrates the computational efficiency of Alg. 6 which implements the up-dating method for the window estimation of the SUR model when compared with Alg. 4 which solves the same problem afresh. Particularly, Figure 2 shows the ratio of the execution times

of the two algorithms when the estimation window is rolling over the data set by $h = 1, 3, 6, 12, 24$ (months). Note that herein all the algorithms that have been implemented for the SUR and the USUR model take into account the special sparse structure of the model.

M	Alg. 4	Alg. 5×10^2	$\frac{\text{Alg. 4}}{\text{Alg. 5}}$	M	Alg. 4b	Alg. $5b \times 10^2$	$\frac{\text{Alg. 4b}}{\text{Alg. 5b}}$
	afresh	recursive			afresh	recursive	
250	451	16	2817	250	450	101	446
500	10769	39	27612	500	10584	272	3891
750	59999	60	99998	750	53166	1324	4016

(a) Panel A: SUR model leave-one-out (b) Panel B: USUR model leave-one-out

Table 2: Total execution times in seconds of the algorithms from the leave-one-out experiments. Alg. 4 estimates the model afresh, Alg. 5 and Alg. 5b are the proposed downdating algorithms for the SUR and the USUR models respectively. The times for Alg. 5 and Alg. 5b have been multiplied by 10^2 . The algorithms are implemented on a subset VAR model where the dependent variables are: real stock returns, real interest rates, industrial production growth and inflation rates for the period 01/1947 to 12/1987. Results are presented where the sample size is $M = 250, 500, 750$.

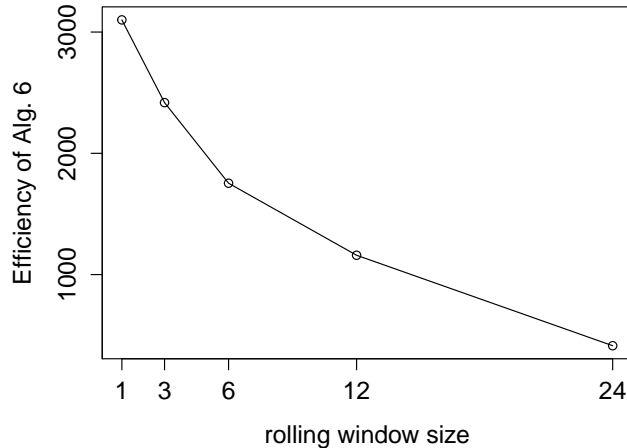


Figure 2: Computational efficiency (execution times ratio) of Alg. 6 when compared with Alg. 4 which solves the same problem afresh. The algorithms are implemented on a subset VAR model where the dependent variables are: real stock returns, real interest rates, industrial production growth and inflation rates for the period 01/1947 to 12/1987. Results when the rolling window size, that is the number of observations added and deleted simultaneously, is $h = 1, 3, 6, 12, 24$ (months) are presented.

The results show that the proposed methods perform significantly better than existing methods that solve the same problem afresh. This computational efficiency is due to the fact that when previous computations are utilized, most operations are performed on matrices of smaller dimensions than those when the computations are executed afresh. The only computationally expensive operation in the proposed Algorithm 2.1 is the product in step 2 with theoretical complexity that depends on the initial number of observations M . However, in the case of the SUR and the USUR models this step is omitted because of the contemporaneous correlation in these models. That is, in the case of the SUR and the USUR model, the computational efficiency of the proposed algorithms is even more impressive and this is because all the operations here depend only on the number of parameters

-which is usually much smaller than the number of observations in the model- and the number of deleted observations from the model.

5 Concluding Remarks and Future Work

The numerical solution of the general linear model (GLM) after a block of data is deleted, known as the downdating problem, has been investigated in depth. The proposed method uses the imaginary unit to remove the effect of the deleted observations from the generalized least squares (GLS) solution. The original GLM is updated with the imaginary deleted observations resulting in a variance covariance matrix which consists of complex covariance values. This updated GLM has been shown to yield the same normal equations as the downdated model to be estimated. In order to obtain this updated model, the GLM is transformed initially to the equivalent ordinary linear model (OLM), which is downdated by using the imaginary unit and by employing a hyperbolic QR decomposition. Then the downdated OLM is again transformed to give the downdated GLM formulation, i.e. (2.5).

The new method computes the downdated GLS estimator by formulating the updated GLM as a generalized linear least squares problem. The main computational tool is the generalized QR decomposition which is realized via hyperbolic Householder and Householder reflections. This new method provides a solution to the downdating problem even when the dispersion matrix is singular. The same strategy has been employed in the solution of the GLM when observations are added and deleted simultaneously. Furthermore, the method has been extended to the solution of the seemingly unrelated regressions (SUR) and the unbalanced SUR (USUR) models when observations are removed.

The algorithms implement computationally efficient strategies that fully exploit the special structure of the matrices. Experimental results have been presented and analyzed. It is shown that this new method (Algorithm 2.1) for downdating the GLM outperforms existing methods which solve the problem from scratch. Especially when the number of the deleted observations is not very big compared to the remaining observations in the downdating model, then Algorithm 2.1 is computationally very efficient. The computational efficiency of the new methods for the downdating of the SUR and the USUR models has been examined using real macroeconomic data by estimating a subset vector autoregressive model. The results show that the proposed downdating algorithms surpass existing ones that estimate the models afresh. This, amplifies further the usefulness of such a method in time series analysis and in econometrics in general. The algorithms have been implemented in R. Therefore their performance can be optimized if they are implemented using high performance numerical libraries such as LAPACK and ScaLAPACK (Anderson et al., 1999; Blackford et al., 1997).

Future work will consider the use of the proposed downdating method for the investigation of influential data in the SUR model when an unequal number of observations is added and deleted from some of the regressions. Additionally, the application of the downdating method, within the context of regression diagnostics and cross-validation problems as well as in time series analysis, where a sequence of least squares problems are solved each time with observations added (new data) and deleted (old data), merits further investigation. Currently the downdating of the SUR model with time varying coefficients and autoregressive disturbances and also the downdating of the simultaneous equations model is under investigation.

Acknowledgments

The authors are grateful to the Associate Editor and the referee for their valuable comments and suggestions. This work is in part supported by the Cyprus University of Technology start up grand.

References

- Aitken, A. C. (1934). On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48.
- Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., and Sorensen, D. (1999). *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition.
- Bai, J. and Shi, S. (2011). Estimating High Dimensional Covariance Matrices and its Applications. *Annals of Economics and Finance*, 12(2):199–215.
- Belsley, D. A., Kuh, A. E., and Welsch, R. E. (2004). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley and Sons, New York.
- Bhimasankaram, P., Sengupta, D., and Ramanathan, S. (1995). Recursive inference in a general linear model. *Sankhya: The Indian Journal of Statistics*, 57(A):227255.
- Björck, Å. (1996). *Numerical methods for least squares problems*. SIAM, Philadelphia.
- Blackford, L. S., Choi, J., Cleary, A., D’Azevedo, E., Demmel, J., Dhillon, I., Dongarra, J., Hammarling, S., Henry, G., Petitet, A., Stanley, K., Walker, D., and Whaley, R. C. (1997). *ScaLAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Chambers, J. M. (1971). Regression updating. *Journal of the American Statistical Association*, 66(336):744–748.
- Chavas, J.-P. (1982). Recursive estimation of simultaneous equation models. *Journal of Econometrics*, 18(2):207–217.
- Chib, S. and Greenberg, E. (1995). Hierarchical analysis of {SUR} models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics*, 68(2):339 – 360.
- Christensen, R., Pearson, L. M., and Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, 34(1):38–45.
- Clark, T. E. and McCracken, M. W. (2009). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, 50(2):363–395.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.
- Deaton, A. (1986). Demand analysis. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 3 of *Handbook of Econometrics*, chapter 30, pages 1767–1839. Elsevier.
- Elden, L. and Park, H. (1994). Block downdating of least squares solutions. *SIAM Journal of Matrix Analysis and Applications*, 15(3):1018–1034.
- Foschi, P., Belsley, D. A., and Kontoghiorghes, E. J. (2003). A comparative study of algorithms for solving seemingly unrelated regressions models. *Computational Statistics & Data Analysis*, 44(1-2):3–35.

- Foschi, P. and Kontoghiorghes, E. J. (2002). Seemingly unrelated regression model with unequal size observations: computational aspects. *Computational Statistics & Data Analysis*, 41(1):211–229.
- Golub, G. H. (1969). Matrix decompositions and statistical calculations. In Milton, R. C. and Neider, J. A., editors, *Statistical Computation*, pages 365–397. Academic Press, New York.
- Golub, G. H. and Van Loan, C. F. (1996). *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, Maryland, 3rd edition.
- Gragg, W. B., Leveque, R. J., and Trangenstein, J. A. (1979). Numerically stable methods for updating regressions. *Journal of the American Statistical Association*, 74(365):161–168.
- Griffiths, W. E. and Valenzuela, M. R. (2006). Gibbs samplers for a set of seemingly unrelated regressions. *Australian & New Zealand Journal of Statistics*, 48(3):335–351.
- Hadjiantoni, S. (2015). *Numerical methods for the recursive estimation of large-scale linear econometric models*. PhD thesis, Queen Mary, University of London.
- Haslett, J. (1999). A simple derivation of deletion diagnostic results for the general linear model with correlated errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):603–609.
- Haslett, J. and Dillane, D. (2004). Application of ‘delete = replace’ to deletion diagnostics for variance component estimation in the linear mixed model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(1):pp. 131–143.
- Holly, S., Pesaran, M. H., and Yamagata, T. (2011). Spatial and temporal diffusion of house prices in the uk. *Journal of Urban Economics*, 69(1):2–23.
- Jammalamadaka, S. and Sengupta, D. (1999). Changes in the general linear model: A unified approach. *Linear Algebra and its Applications*, 289(13):225 – 242.
- Jammalamadaka, S. R. and Sengupta, D. (2007). Inclusion and exclusion of data or parameters in the general linear model. *Statistics & Probability Letters*, 77(12):1235 – 1247.
- Judge, G. G., Griffiths, W. E., Hill, C. H., Lütkepohl, H., and Lee, T.-C. (1985). *The Theory and Practice of Econometrics, 2nd Edition*. Wiley, New York.
- Kmenta, J. and Gilbert, R. F. (1970). Estimation of seemingly unrelated regressions with autoregressive disturbances. *Journal of the American Statistical Association*, 65(329):pp. 186–197.
- Kontoghiorghes, E. J. (2000). *Parallel algorithms for linear models: numerical methods and estimation problems*, volume 15 of *Advances in computational economics*. Kluwer Academic Publishers, Boston.
- Kontoghiorghes, E. J. and Clarke, M. R. B. (1995). An alternative approach for the numerical solution of seemingly unrelated regression equations models. *Computational Statistics & Data Analysis*, 19(4):369–377.
- Kourouklis, S. and Paige, C. C. (1981). A constrained least squares approach to the general Gauss-Markov linear model. *Journal of the American Statistical Association*, 76(375):620–625.
- Lee, B.-S. (1992). Causal relations among stock returns, interest rates, real activity, and inflation. *The Journal of Finance*, 47(4):1591–1603.

- Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer, 1st ed. 2006. corr. 2nd printing edition.
- Martin, R. J. (1992). Leverage, influence and residuals in regression models when observations are correlated. *Communications in Statistics - Theory and Methods*, 21(5):1183–1212.
- Paige, C. C. (1978). Numerically stable computations for general univariate linear models. *Communications in Statistics - Simulation and Computation*, 7(5):437–453.
- Paige, C. C. (1979). Fast numerically stable computations for generalized linear least squares problems. *SIAM Journal on Numerical Analysis*, 16(1):165–171.
- Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, 29(2):307–318.
- Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134 – 152. *Dynamic Econometric Modeling and Forecasting*.
- Pollock, D. S. G. (2003). Recursive estimation in econometrics. *Computational Statistics & Data Analysis*, 44(1-2):37–75.
- Preisser, J. and Perin, J. (2007). Deletion diagnostics for marginal mean and correlation model parameters in estimating equations. *Statistics and Computing*, 17(4):381–393.
- Preisser, J. S. and Qaqish, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika*, 83(3):551–562.
- Rader, C. and Steinhardt, A. (1986). Hyperbolic householder transformations. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(6):1589 – 1602.
- Rao, C. R. (2002). *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc., 2 edition.
- Rochon, J. (1996). Accounting for covariates observed post randomization for discrete and continuous repeated measures data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):pp. 205–219.
- Rossi, B. and Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, 30(3):432–453.
- Shieh, G. (2000). General multivariate linear models for longitudinal studies. *Communications in Statistics - Theory and Methods*, 29(4):735–753.
- Smith, M. and Kohn, R. (2000). Nonparametric seemingly unrelated regression. *Journal of Econometrics*, 98(2):257–281.
- Srivastava, V. and Dwivedi, T. (1979). Estimation of seemingly unrelated regression equations: A brief survey. *Journal of Econometrics*, 10(1):15 – 32.
- Srivastava, V. K. and Giles, D. E. A., editors (1987). *Seemingly unrelated regression equations models: estimation and inference*. Marcel Dekker, Inc., New York, NY, USA.

- Telsler, L. G. (1964). Iterative estimation of a set of linear regression equations. *Journal of the American Statistical Association*, 59(307):pp. 845–862.
- Verbyla, A. P. and Venables, W. N. (1988). An extension of the growth curve model. *Biometrika*, 75(1):pp. 129–138.
- Wang, H. (2010). Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Computational Statistics & Data Analysis*, 54(11):2866 – 2877.
- Yanev, P. I. and Kontoghiorghes, E. J. (2008). Parallel algorithms for downdating the least squares estimator of the regression model. *Parallel Computing*, 34(6-8):451–468.
- Yanev, P. I. and Kontoghiorghes, E. J. (2009). Graph-based strategies for performing the exhaustive and random k-fold cross-validations. *Journal of Computational and Graphical Statistics*, 18(4):894–914.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):pp. 348–368.
- Zhu, H., Ibrahim, J. G., and Cho, H. (2012). Perturbation and scaled cooks distance. *The Annals of Statistics*, 40(2):785–811.
- Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics*, 38(4):pp. 1092–1109.
- Zyskind, G. and Martin, F. B. (1969). On best linear estimation and general gauss-markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM Journal on Applied Mathematics*, 17(6):1190–1202.