An alternative approach for the numerical solution of seemingly unrelated regression equations models

E.J. Kontoghiorghes
Centre for Insurance and Investment, City University Business School, London, U.K.

M.R.B. Clarke
Department of Computer Science, Queen Mary and Westfield College, London, U.K.

Received May 1993
Revised November 1993

Abstract: An alternative approach to compute the coefficients of a Seemingly Unrelated Regression Equations (SURE) model is proposed. Orthogonal transformations are employed to avoid the difficulties in directly computing the inverse of the variance–covariance matrix (or its estimate) which often lead to unnecessary loss of accuracy. The solution of the special SURE model where the problem is constrained so that the regressors in each equation contain the regressors in previous equations as a proper subset, is considered in detail.

Keywords: SURE model; Variance–covariance matrix; Orthogonal transformations.

1. Introduction

We are concerned with the problem of estimating a system of regression equations where the random disturbances are correlated with each other. That is, the equations are linked statistically, even though not structurally, through the non-diagonality of the associated variance–covariance matrix. The expression Seemingly Unrelated Regression Equations (SURE) is used to reflect the fact that the individual equations are in fact related to one another, even though...
superficially they may not seem to be. The precise mathematical details of the SURE model are given next.

The SURE model that we are concerned with comprises $M$ regression equations

$$y_i = X_i \beta_i + u_i \quad (i = 1, \ldots, M),$$

(1)

where $y_i \in \mathbb{R}^T$ is the dependent variable, $X_i \in \mathbb{R}^{T \times k_i}$ is the observation matrix, $\beta_i \in \mathbb{R}^{k_i}$ is the vector of regression coefficients and $u_i \in \mathbb{R}^T$ is a vector of random disturbances. The basic assumptions underlying the SURE model (1) are $\mathbb{E}(u_i) = 0$, $\mathbb{E}(u_i u_j^T) = \sigma_{ij} I_T$ and $\lim_{T \to \infty} (X_i^T X_j / T)$ exists.

A more compact form of the SURE model is

$$\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_M
\end{bmatrix} =
\begin{bmatrix}
  X_1 \\
  X_2 \\
  \vdots \\
  X_M
\end{bmatrix}
\begin{bmatrix}
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_M
\end{bmatrix} +
\begin{bmatrix}
  u_1 \\
  u_2 \\
  \vdots \\
  u_M
\end{bmatrix},

(2)

where $y \in \mathbb{R}^{MT}$, $X \in \mathbb{R}^{MT \times K}$, $\beta \in \mathbb{R}^K$, $u \in \mathbb{R}^{MT}$ and $K = \sum_{i=1}^M k_i$. Also the assumptions concerning disturbances can be written as $\mathbb{E}(u) = 0$ and $\mathbb{E}(uu^T) = \Sigma \otimes I_T$, where $\otimes$ denotes the usual Kronecker product and $\Sigma = [\sigma_{ij}]$ is an $M \times M$ matrix.

Various least squares estimators have been proposed for estimating the parameters of the SURE model [13]. The most popular are the Ordinary Least Squares (OLS) and General Least Squares (GLS) approach. The OLS approach implicitly assumes that the regression equations in (1) are independent of each other and estimates the parameters of the model equation by equation. The GLS approach utilizes the additional prior information about correlation between the disturbances that is fundamental to the SURE specification and estimates the parameters of all the equations jointly. In [13] it was shown that the GLS estimator

$$b_G = (X^T (\Sigma^{-1} \otimes I_T) X)^{-1} X^T (\Sigma^{-1} \otimes I_T) y$$

(3)

is the best linear unbiased estimator (blue) of $\beta$. In some cases the OLS and GLS estimators are identical, e.g. $\forall i, j \quad X_i = X_j$ and $\forall i \neq j \quad \sigma_{ij} = 0$. More details can be found in [12,13].

In general $\Sigma$ is unknown and thus the estimator of $\beta$ will be unobservable. Zellner in [15] proposed an estimator of $\beta$ based on $b_G$ where $\Sigma$ is replaced by an estimator matrix $S \in \mathbb{R}^{M \times M}$. Thus $b_F = (X^T (S^{-1} \otimes I_T) X)^{-1} X^T (S^{-1} \otimes I_T) y$ is a feasible GLS (FGLS) estimator of $\beta$. Methods for constructing $S$ based on residuals obtained by the application of OLS are shown in [15,16].
2. An alternative procedure for estimating the SURE model

A number of methods have been proposed to compute $b_G$ in (3) or its feasible equivalent $b_F$ [14,15]. These methods require the expensive computation of the inverse of $\Sigma$ which often leads to unnecessary loss of accuracy when $\Sigma$ is ill-conditioned. Here an alternative approach for estimating the coefficient parameters of the SURE model is proposed. With this approach the coefficients in all equations of the SURE model are estimated simultaneously, by applying Paige’s generalized least squares method to the whole system of equations (2) [6–8]. This approach avoids the difficulty in directly computing the inverse of $\Sigma$ (or its estimate $S$) and it can also be used to give the blue of $\beta$ in (2) when $\Sigma$ is singular [5].

Let $WW^T$ be the Choleski decomposition of $\Sigma$ and $C = [c_{ij}] \in \mathbb{R}^{M \times m}$ be obtained by removing the zero columns of the lower triangular $W \in \mathbb{R}^{M \times M}$, where $\text{rank}(\Sigma) = m \leq M$. As shown in [5,7] the blue of $\beta$, say $\tilde{\beta}$, is the solution of

$$\arg\min_{\beta} \beta^T V \beta \quad \text{subject to} \quad y = X\beta + (C \otimes I_T)V,$$

where the random $mT$ element vector $V$ satisfies $(C \otimes I_T)V = u$, $\mathcal{G}(V) = 0$ and $\mathcal{G}(VV^T) = I_{mT}$. For

$$Q^T[y, X_i] = \begin{bmatrix} \hat{y}_i & 0 \\ \tilde{y}_i & L_i \end{bmatrix}^T k_i,$$

where $Q_i \in \mathbb{R}^{T \times T}$ is orthogonal and $L_i$ is lower triangular non-singular, that is $X_i$ is of full rank ($i = 1, \ldots, M$), let

$$Q^T = \begin{bmatrix} Q_{1}^T \\ \vdots \\ Q_{M}^T \end{bmatrix},$$

and $\Pi \in \mathbb{R}^{MT \times MT}$ be a permutation matrix such that

$$\Pi^TQ^T[y, X] = \Pi^T \begin{bmatrix} \hat{y}_1 & 0 \\ \tilde{y}_1 & L_1 \\ \hat{y}_2 & 0 \\ \tilde{y}_2 & L_2 \\ \vdots & \vdots \\ \hat{y}_M & 0 \\ \tilde{y}_M & L_M \end{bmatrix},$$

$$= \begin{bmatrix} \hat{y} \\ \tilde{y} \\ L \\ 1 \end{bmatrix}^T K,$$
Also let

\[
\Pi^T Q^T = \begin{bmatrix} Q_A^T \\ Q_B^T \end{bmatrix} \begin{bmatrix} MT - K \\ K \end{bmatrix},
\]

\[
G^T \begin{bmatrix} \hat{\gamma} \\ Q_A^T(C \otimes I_T) \end{bmatrix} = \begin{bmatrix} 1 \\ \hat{\gamma}_1^+ \\ L_1 \\ L_{11} \end{bmatrix}, \tag{8}
\]

\[
Q_B^T(C \otimes I_T)P = \begin{bmatrix} L_{21} \\ L_{22} \end{bmatrix}, \tag{9}
\]

and

\[
P^T V = \begin{bmatrix} \hat{\nu} \\ \hat{V} \end{bmatrix}, \tag{10}
\]

where \( G \) and \( P \) are orthogonal matrices with dimensions \((MT - K) \times (MT - K)\) and \( mT \times mT \) respectively, \( \text{rank}(Q_A^T(C \otimes I_T)) = t \) and \( L_{11} \) is lower triangular non-singular. Efficient algorithms to compute (8) when \( \Sigma \) is non-singular can be found in [1–4, 6, 8].

Using (7) to (10), (4) can equivalently be written as

\[
\begin{aligned}
\arg\min_{\beta} \left( \hat{\gamma}^T \hat{\gamma} + \hat{\nu}^T \hat{V} \right) \quad \text{subject to} \quad & \begin{cases} 
\hat{\gamma}_1^+ = 0, \\
\hat{\gamma}_2^+ = L_{11} \hat{\nu}, \\
\hat{y} = L\beta + L_{21} \hat{\nu} + L_{22} \hat{V}.
\end{cases}
\end{aligned} \tag{11}
\]

Clearly if \( \hat{\gamma}_1^+ \) is not null the constraints in (11) are inconsistent and the SURE model has no solution. Given that \( \hat{\gamma}_1^+ = 0, \) \( \hat{\nu} \) can be derived from solving the triangular system in the second constraint. In the third constraint for minimizing \( \nu^T V \) the arbitrary \( \hat{\nu} \) is set to zero and thus, \( \hat{\beta} \) is computed by solving the lower triangular system

\[
L_1 \hat{\beta} = \hat{y} - L_{21} \hat{\nu}. \tag{12}
\]

To take advantage of the block diagonal form of \( L \) let \( \hat{\nu}^T L_{21} = [g_1^T \cdots g_M^T] \) where \( \hat{g}_i \in \mathbb{R}^k \). The solution of (12) is then equivalent to solving

\[
L_i \hat{\beta}_i = \hat{y}_i \quad \text{for} \quad i = 1, \ldots, M, \tag{13}
\]

where \( \hat{\beta}^T = [\hat{\beta}_1^T \cdots \hat{\beta}_M^T] \). Note that in the extreme case of \( t = 0 \) which implies \( \hat{\gamma} = Q_A^T(C \otimes I_T) = 0 \), the arbitrary \( \nu \) is set to zero in order to minimize \( \nu^T V \). Thus, \( \hat{\beta} = L^{-1} \hat{y} \) can be derived using the OLS approach.
A reliable representation of $\mathcal{V}(\hat{\beta})$, the variance–covariance matrix of $\hat{\beta}$, is given by

$$L\mathcal{V}(\hat{\beta})L^T = L_{22}L_{22}^T.$$  

(14)

The derivation of the latter is shown in [5,7] where it is also stated that the computation of $L^{-1}L_{22}$ could lead to loss of accuracy if $L$ is ill-conditioned. Let the matrices $\mathcal{V}(\hat{\beta})$ and $L_{22}$ be partitioned as

$$\mathcal{V}(\hat{\beta}) = \begin{bmatrix}
U_{11} & U_{12} & \ldots & U_{1M} \\
U_{21} & U_{22} & \ldots & U_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
U_{M1} & U_{M2} & \ldots & U_{MM}
\end{bmatrix} \quad \text{and} \quad L_{22} = \begin{bmatrix}
\Lambda_1 \\
\Lambda_2 \\
\vdots \\
\Lambda_M
\end{bmatrix},$$  

(15)

respectively, where $U_{ij} \in \mathbb{R}^{k_i \times k_j}$ and $\Lambda_i \in \mathbb{R}^{k_i \times (mT-k_i)}$ (observe that $U_{ii} = \mathcal{V}(\hat{\beta}_i)$ and $U_{ij} - \mathcal{E}(\hat{\beta}_i - \hat{\beta}_j)\mathcal{E}(\hat{\beta}_j - \hat{\beta}_j)^T = U_{ii}^T$). From (14) and (15) it follows that

$$L_{1i}U_{ij}L_{j}^T = \Lambda_i \Lambda_j^T.$$  

(16)

The following iterative algorithm summarizes the steps required to compute $\hat{\beta}$ and $\mathcal{V}(\hat{\beta})$, as $\Sigma$ has to be estimated.

1. Compute $Q_i, \hat{y}_i, \tilde{y}_i$ and $L_i$ as in (5).
2. Compute $\Pi$ as in (7).
3. Until convergence do
   - Compute $\Sigma$ by $S$.
   - Compute the Choleski decomposition $S = WW^T$.
   - Eliminate the $(M - m)$ zero columns of $W$ to obtain $C$ where $\text{rank}(S) = m$.
   - Compute $P, \hat{y}_1^*, \hat{y}_2^*, L_{11}$ as in (8) and $L_{21}, L_{22}$ as in (9).
   - If $\hat{y}_i^* = 0$ then
     - Compute $V$ from the second constraint in (11).
     - $\forall i$ solve the lower triangular system of equations (13).
     - $\forall i, j$ such that $i \geq j$ compute (16).
   - Else the SURE model is inconsistent and has no solution.
4. End if
5. End until

3. A SURE model with proper subset regressors

In some cases the regressor matrices $X_i$ have specific structures which simplify the computations of the above iterative algorithm. To demonstrate this, consider the special SURE model where $\Sigma$ is non singular and the regressors in each equation contain the regressors in previous equations as a proper subset. That
is, in the SURE model (1) the regression matrix \( X_i \) is defined (or restricted) as

\[
X_i = \begin{bmatrix} \tilde{X}_i & X_{i-1} \end{bmatrix} \quad (i = 2, \cdots, M).
\]

(17)

For \( Q_i = Q_M = \{ i = 1, \ldots, M-1 \} \) the restriction in the regression matrices (17) implies that \( I_i \) in (5) is a submatrix of \( I_M \). If \( Q^T = (I_M \otimes Q_M^T) \) and \( P = (I_M \otimes Q_M) \Pi \), then

\[
\Pi^T(I_M \otimes Q_M^T)(C \otimes I_T)(I_M \otimes Q_M) \Pi = \Pi^T(C \otimes I_T) \Pi
\]

\[
= \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} MT - K \\ MT - K \end{bmatrix}.
\]

(18)

For \( L_{pq} = \begin{bmatrix} A^{(p,a)}_{11} & A^{(p,a)}_{12} & \ldots & A^{(p,a)}_{1M} \\ A^{(p,a)}_{21} & A^{(p,a)}_{22} & \ldots & A^{(p,a)}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(p,a)}_{M1} & A^{(p,a)}_{M2} & \ldots & A^{(p,a)}_{MM} \end{bmatrix} \quad (p, q = 1, 2),
\]

in the Appendix we proved that

\[
L_{12} = 0, \quad A^{(1,1)}_{ij} = \begin{cases} c_{ii}I_{T-k_i}, & i = j, \\ [c_{ij}I_{T-k_i}, 0], & i > j, \\ 0, & i < j, \end{cases}
\]

(19)

\[
A^{(2,1)}_{ij} = \begin{bmatrix} 0 & c_{ij}I_{k_i-k_j} \\ 0 & 0 \end{bmatrix}, \quad i > j, \quad \text{and} \quad A^{(2,2)}_{ij} = \begin{cases} c_{ii}I_{k_i}, & i = j, \\ 0, & i > j, \\ [c_{ij}I_{k_i}, 0], & i < j, \end{cases}
\]

where the matrices \( A^{(1,1)}_{ij} \), \( A^{(2,1)}_{ij} \) and \( A^{(2,2)}_{ij} \) are of dimensions \((T-k_i) \times (T-k_j)\), \((T-k_i) \times k_j \) and \( k_i \times k_j \) respectively. Note that (19) implies (18) is lower triangular. For \( \hat{V}_T = [\hat{V}_{1T} \cdots \hat{V}_{MT}] \) where \( \hat{V}_i \in \mathbb{R}^{(T-k_i)} \), the solution of the special SURE model is given by

\[
\hat{V}_i = \frac{1}{c_{ii}} \left( \hat{y}_i - \sum_{j=1}^{i-1} c_{ij} \left[ I_{T-k_i} \quad 0 \right] \hat{V}_j \right)
\]

(20)

and

\[
L_i \hat{\beta}_i = \hat{y}_i - \sum_{j=1}^{i-1} c_{ij} \left[ 0 \quad I_{k_i-k_j} \quad 0 \right] \hat{V}_j.
\]

(21)
Within this context, the equivalence of (16) is
\[ L_iU_jL_j^T = \sum_{p=1}^i c_{ij} c_{jp} \begin{bmatrix} 0 & 0 \\ I_{k_p} & 0 \end{bmatrix} \]
where \( i > j \) and a zero dimension denotes a null matrix.

Observe that \( \hat{\beta}_i \) is given by the solution of the OLM \( y_i = X_i\hat{\beta}_i + u_i \) and a linear combination of the residuals \( \hat{V}_j \) \((j = 1, \ldots, i - 1)\). Revankar in [9] considered the above special SURE model when \( M = 2 \), i.e., the SURE model comprised of only two equations. Evaluating the solutions he presented in his paper, (21) is derived. In the case of \( X_g = X_{g+1} = \cdots = X_M \), \( \hat{\beta}_i \) for \( i = 1, \ldots, g - 1 \) can be derived by treating separately the subsystem comprising the first \( g - 1 \) regression equations. This is also true when the restrictions in the first \( g - 1 \) equations are only subsets of \( X_g = \cdots = X_M \) [10,11].

If \( \Sigma \) is singular then some columns of \( C \) will be zero. This implies that (20) will hold if \( c_{ii} \neq 0 \) and for \( c_{ii} = 0 \) the vector \( \hat{V}_j \) is set to zero in order to minimize \( V^TV \). In this case, the special SURE model will be consistent if \( \forall i \) such that \( c_{ii} = 0 \) we have \( \hat{y}_i = \sum_{k=1}^{i-1} c_{ik} [I_{r-k_i} 0] \hat{V}_j \).

4. Comments

Following Paige [6–8], we have proposed an approach based on orthogonal transformations for estimating the coefficient parameters of the SURE model. This approach leads to an algorithm that does not require the expensive computation of matrix inverses which often results to unnecessary loss of accuracy. Furthermore, this algorithm can be used to derive the blue of \( \beta \) in (2) when \( \Sigma \) (or its estimate \( S \)) is singular [5]. The employment of this algorithm to solve the SURE model with regression matrices constrained as in (17), has demonstrated its efficiency. Currently the implementation of the algorithm on a massively parallel system is been considered and the application of the new approach to compute the Three-Stage Least Squares of simultaneous equations is investigated.

Appendix

**Proof by induction of (19).** In the following proof we will use \( \Pi_i \) to denote the matrix \( \Pi \) in (7) when the SURE model comprises \( i \) regression equations and \( C_i \) be an \( i \times i \) lower triangular matrix. For \( M = 2 \) the permutation matrix \( \Pi_2 \) is defined as

\[
\Pi_2^T = \begin{bmatrix}
I_{T-k_1} & 0 & 0 & 0 \\
0 & 0 & I_{T-k_2} & 0 \\
0 & I_{k_1} & 0 & 0 \\
0 & 0 & 0 & I_{k_2}
\end{bmatrix}
\]
such that

\[
\Pi_T^T(C_2 \otimes I_T) \Pi_2 = 
\begin{bmatrix}
    c_{11}I_T-k_1 & 0 & 0 & 0 \\
    c_{21}I_T-k_2 & 0 & c_{22}I_T-k_2 & 0 \\
    0 & 0 & 0 & c_{11}I_k_1 & 0 \\
    0 & c_{21}I_k_2-k_1 & 0 & 0 & 0 \\
    0 & 0 & 0 & c_{21}I_k_1 & c_{22}I_k_2 \\
\end{bmatrix},
\]

which satisfies (19).

**Inductive hypothesis.** For \( M = \lambda(\lambda \geq 2) \) (19) is true, i.e. \( \Pi^T_\lambda(C_\lambda \otimes I_T)\Pi_\lambda \) satisfies (19).

To show that \( \Pi^T_{\lambda+1}(C_{\lambda+1} \otimes I_T)\Pi_{\lambda+1} \) also satisfies (19) let us write \( \Pi^T_{\lambda+1} \) as

\[
\Pi^T_{\lambda+1} = 
\begin{bmatrix}
    I_T-k_1 & 0 & 0 & 0 \\
    0 & 0 & \Pi^T_{\lambda1} \\
    0 & I_k_1 & 0 \\
    0 & 0 & \Pi^T_{\lambda2} \\
\end{bmatrix},
\]

where \([\Pi_{\lambda1}, \Pi_{\lambda2}]\) corresponds to the matrix \( \Pi_\lambda \) of the SURE model comprising equations 2, \ldots, \( \lambda + 1 \). Using the latter we have

\[
\Pi^T_{\lambda+1}(C_{\lambda+1} \otimes I_T)\Pi_{\lambda+1} = 
\begin{bmatrix}
    c_{11}I_T-k_1 & 0 & 0 & 0 \\
    \vdots & 0 & \Pi^T_{\lambda1}(C_\lambda \otimes I_T)\Pi_{\lambda1} & 0 & \Pi^T_{\lambda1}(C_\lambda \otimes I_T)\Pi_{\lambda2} \\
    c_{\lambda+1}I_T-k_\lambda & 0 & c_{11}I_k_1 & 0 \\
    \vdots & 0 & 0 & c_{11}I_k_1 \\
    \vdots & 0 & 0 & c_{11}I_k_1 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    0 & c_{\lambda+1}I_k_{\lambda+1} & 0 & 0 \\
    0 & 0 & c_{\lambda+1}I_k_{\lambda+1} & c_{\lambda+1}I_k_{\lambda+1} \\
\end{bmatrix},
\]

where \( C_\lambda \in \mathbb{R}^{\lambda \times \lambda} \) is the submatrix of \( C_{\lambda+1} \) starting from the element at the position (2, 2). From

\[
\Pi^T_\lambda(C_\lambda \otimes I_T)\Pi_\lambda = 
\begin{bmatrix}
    \Pi^T_{\lambda1}(C_\lambda \otimes I_T)\Pi_{\lambda1} & \Pi^T_{\lambda1}(C_\lambda \otimes I_T)\Pi_{\lambda2} \\
    \Pi^T_{\lambda2}(C_\lambda \otimes I_T)\Pi_{\lambda1} & \Pi^T_{\lambda2}(C_\lambda \otimes I_T)\Pi_{\lambda2} \\
\end{bmatrix}
\]
and using the inductive hypothesis we observe that $H_{k+1}^T (C_{k+1} \otimes I_T) H_{k+1}$ satisfies (19). The latter concludes the proof. $\square$

Acknowledgement

The authors would like to thank the anonymous referees for their helpful and constructive suggestions.

References