Estimating regressions and seemingly unrelated regressions with error component disturbances

Foschi Paolo

University of Bologna

8. February 2005

Estimating Regression and Seemingly Unrelated Regressions with Error Component disturbances

Paolo Foschi *

September 7, 2006

The estimation of regressions models with two-way error component disturbances, is considered for the case where both the random effects are non-spherically distributed. The usual approach that first transforms the effects into uncorrelated ones and then applies within and between transformations, cannot be conveniently applied. Here, it is proposed to revert this scheme by firstly applying the within and between transformations. This results in simple General Linear Model which can be partitioned into three smaller GLMs. Then, by exploiting the structure of the models and using the Generalized QR decomposition as a tool, a computationally efficient and numerically reliable method for estimating the regression parameters is derived. This estimation method is generalized to the case of a system of seemingly unrelated regressions.

1. Introduction

One of the most used model in analysis of panel data is given by the two-way error component regression model [1, 5, 8, 31, 35]. In its basic formulation that model assumes that the time and individual random effects are spherically distributed. In [27] and [29] the authors relaxed that assumption by considering a one-way model with autoregressive (AR) idiosyncratic errors and heteroschedastic individual effects, respectively. Further generalizations followed, specifically, in [2, 10] Moving Average (MA) errors are considered and in [17, 26, 30] an ARMA model for them is assumed. Dynamics for the two-way model has been considered in [12, 13, 19, 28, 33, 34] where autocorrelation is assumed in the time effects and/or the idiosyncratic errors. Here, the two-way error component model with autocorrelations in both the time- and individual-effects is considered. It will be shown that the trasformations used

*paolo.foschi2@unibo.it. Faculty of Economics, University of Bologna, P.le Vittoria 15, I-47100 Forlí, Italy

for the basic two-way model can still be used for this extension keeping the model simple and tractable.

The linear regression model with two-way error component disturbances is given by

$$y_{ti} = \alpha + \sum_{k=1}^{K} x_{tik}\beta_k + u_{ti}, \qquad i = 1,\dots,N,\ t = 1,\dots,T, \tag{1a}$$

with

$$u_{ti} = \lambda_t + \mu_i + \nu_{ti}, \qquad i = 1,\dots,N,\ t = 1,\dots,T, \tag{1b}$$

where μ_i and λ_t denote the unobservable individual and time effect, respectively and ν_{ti} is the idiosyncratic disturbance term. The errors λ_t, μ_i, and ν_{ti} are assumed to have zero mean and to be indipendent each other, even across different observations, that is $\mathrm{E}[\lambda_t\mu_i] = \mathrm{E}[\lambda_t\nu_{sj}] = \mathrm{E}[\mu_i\nu_{sj}] = 0$, for $i, j = 1,\dots,N$ and $s,t = 1,\dots,T$ [8]. Furthremore, the idiosyncratic errors are assumed to be spherically distributed, that is $\mathrm{E}[\nu_{it}^2] = \sigma_\nu^2$ and $\mathrm{E}[\nu_{it}\nu_{js}] = 0$ for $i \neq j$, $s \neq t$, $i, j = 1,\dots,N$, and $s,t = 1,\dots,T$.

The two-way model in (1) can be written in a more compact form as

$$y = \iota_{NT}\alpha + X\beta + u, \tag{2a}$$

with

$$u = (\iota_N \otimes I_T)\lambda + (I_N \otimes \iota_T)\mu + \nu, \tag{2b}$$

where $\iota_n \in \mathbb{R}^n$ denotes a vector with all ones, $\beta \in \mathbb{R}^K$, $\lambda \in \mathbb{R}^T$ and $\mu \in \mathbb{R}^N$ are the vectors with elements β_k, λ_t and μ_i, respectively ($t = 1,2,\dots,T$, $i = 1,2,\dots,N$ and $k = 1,2,\dots,K$). Furthermore, $X = \begin{pmatrix} x_1\ x_2\ \cdots\ x_K \end{pmatrix}$ and $x_k, y, u, v \in \mathbb{R}^{NT}$ are the vectors with elements x_{tik}, y_{ti}, u_{ti} and v_{ti}, respectively, lexicographically sorted for $t = 1,2,\dots,T$ and $i = 1,2,\dots,N$.

The random vectors λ, μ and ν have zero mean and their covariances are given by

$$\mathrm{Cov}\left(\begin{pmatrix}\lambda\\ \mu\\ \nu\end{pmatrix}\right) = \begin{pmatrix}\Psi_\lambda & 0 & 0\\ 0 & \Psi_\mu & 0\\ 0 & 0 & \sigma_\nu^2 I_{NT}\end{pmatrix}, \tag{3}$$

where $\Psi_\lambda \in \mathbb{R}^{T\times T}$ and $\Psi_\mu \in \mathbb{R}^{N\times N}$ are positive semi-definite. It follows that, u has zero mean and covariance matrix given by

$$\Omega \equiv \mathrm{Cov}(u) = J_N \otimes \Psi_\lambda + \Psi_\mu \otimes J_T + \sigma_\nu^2 I_{NT}, \tag{4}$$

where $J_n \equiv \iota_n\iota_n^T$ is a $n \times n$ matrix of all ones. Thus, the two-way random effects regression model (2) can be considered as a General Linear Model (GLM).

The structure of the paper is the following. The next section reviews some results about the basic case of spherically distributed disturbances, that is when $\Psi_\lambda = \sigma_\lambda^2 I_N$ and $\Psi_\mu = \sigma_\mu^2 I_N$.

Then, a compact reformulation of the within and between regressions is introduced and considered. The third section extends this concepts to tackle the estimation in the case of autocorrelated random effects.

Differently to other estimation methods, the approach here proposed can be easily generalized to sets of equations like Seemingly Unrelated Regressions (SUR) or Simultaneous Equations models. The generalization to the SUR case is considered in the forth section. Computationally efficient techniques to estimate the resulting formulations are suggested for each model considered.

Final remarks and directions for future research are reported in the last section.

2. Spherically distributed random effects

In the case of spherically distributed effects the two-way random effects model has already been studied in depth [8]. In that case $\Psi_\lambda = \sigma_\lambda^2 I_T$ and $\Psi_\mu = \sigma_\mu^2 I_N$. Indeed, the variance covariance matrix Ω has only four distinct eigenvalues, $\lambda_1 = \sigma_\nu^2$, $\lambda_2 = T\sigma_\mu^2 + \sigma_\nu^2$, $\lambda_3 = N\sigma_\lambda^2 + \sigma_\nu^2$ and $\lambda_4 = \sigma_\nu^2 + N\sigma_\lambda^2 + T\sigma_\mu^2$, with multiplicity $n_1 = (N-1)(T-1)$, $n_2 = N-1$, $n_3 = (T-1)$ and $n_4 = 1$, respectively. It follows that the eigen-decomposition of Ω is given by

$$\Omega = \lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3 + \lambda_4 Q_4, \tag{5}$$

where $Q_1 = (E_N \otimes E_T)$, $Q_2 = (E_N \otimes \bar{J}_T)$ $Q_3 = (\bar{J}_N \otimes E_T)$ and $Q_4 = (\bar{J}_N \otimes \bar{J}_T)$ are the projects on the four eigenspaces and $\bar{J}_n = J_n/n$ and $E_n = I_n - \bar{J}_n$ are idempotent matrices [37, 36, 31].

Notice that, by the eigen-decomposition (5), the powers of Ω are given by

$$\Omega^p = \sum_{i=1}^{4} \lambda_i^p Q_i, \tag{6}$$

for $p \in \mathbb{R}$. Thus, simple explicit expressions for inverse and the square root of Ω exist and can be used for the estimation of the regression parameters. For example, the GLM (2a) is equivalent to the OLM

$$\sigma_\nu \Omega^{-\frac{1}{2}} y = \sigma_\nu \Omega^{-\frac{1}{2}} \iota_{NT} \alpha + \sigma_\nu \Omega^{-\frac{1}{2}} X\beta + v, \qquad v \sim (0, \sigma_\nu^2 I_{NT}), \tag{7}$$

and thus the Best Linear Unbiased Estimator (BLUE) for the parameters α and β is computed by Ordinary Least Squares (OLS) [15, 16].

Alternatively, premultiplying (2a) by $\begin{pmatrix} Q_1 & Q_2 & Q_3 & Q_4 \end{pmatrix}^T$ gives the GLM

$$\begin{pmatrix} Q_1 y \\ Q_2 y \\ Q_3 y \\ Q_4 y \end{pmatrix} = \begin{pmatrix} 0 & Q_1 X \\ 0 & Q_2 X \\ 0 & Q_3 X \\ i_{NT} & Q_4 X \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \bar{u}_3 \\ \bar{u}_4 \end{pmatrix}, \tag{8}$$

where $\bar{u}_i \sim (0, \lambda_i Q_i)$ and $\mathrm{E}[\bar{u}_i \bar{u}_j^T] = 0$ for $i \neq j$ and $i, j = 1, 2, 3, 4$. The first three blocks correspond, respectively, to the Within, Between-individuals and Between-time periods regressions.

It is easy to show that the last block of observations in (8) is un-influential for the GLS and OLS estimators of β and thus it can be dropped when estimating β. Then, given an estimator $\hat{\beta}$ for β, α is estimated by as $\hat{\alpha} = \iota_{NT}^T(y - X\hat{\beta})$ and the residual vector corresponding to $\bar{u}_4$ is null.

Notice that, the ith block of (8) contains NT observations, while its covariance matrix has rank n_i ($i = 1,2,3,4$). Thus, this approach is not optimal and the computational complexity and memory requirements grow by a factor of four. A more parsimonious approach consists on projecting the observations on the $\mathbb{R}^{n_i}$ ($i = 1,2,3,4$) eigen-spaces and reformulate the GLM (2a) on these spaces, rather than in the original $\mathbb{R}^{NT}$ space. This can be done by considering the orthonormal matrices $P_i \in \mathbb{R}^{NT\times n_i}$, $i = 1,2,3,4$ defined by

$$P_1 = W_N \otimes W_T, \qquad P_2 = W_N \otimes w_T, \qquad P_3 = w_N \otimes W_T \qquad \text{and} \qquad P_4 = w_N \otimes w_T$$

where $\begin{pmatrix} w_n & W_n \end{pmatrix} \in \mathbb{R}^{n\times n}$ is the orthogonal matrix such that $\bar{J}_n = w_n w_n^T$ and $E_n = W_n W_n^T$. It follows that

$$Q_i = P_i P_i^T, \qquad P_i^T P_i = I \qquad \text{and} \qquad P_i^T P_j = 0, \tag{9}$$

Notice that, $w_n = \iota_n/\sqrt{n}$ is uniqely defined while W_n can be choosen with some freedom. Convenient choices for W_n are discussed in a more general setting in appendix A.

Now, by premultiplying the GLM (2a) by the orthogonal matrix $P^T = (P_1\ P_2\ P_3\ P_4)^T$ gives the equivalent GLM

$$\begin{pmatrix} P_1^T y \\ P_2^T y \\ P_3^T y \\ P_4^T y \end{pmatrix} = \begin{pmatrix} 0 & P_1^T X \\ 0 & P_2^T X \\ 0 & P_3^T X \\ \sqrt{NT} & P_4^T X \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} P_1^T u \\ P_2^T u \\ P_3^T u \\ P_4^T u \end{pmatrix},$$

or, with the appropriate substitutions,

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 0 & X_1 \\ 0 & X_2 \\ 0 & X_3 \\ \sqrt{NT} & X_4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}, \tag{10}$$

where $u_i \sim (0, \lambda_i I_{n_i})$ are uncorrelated. Here, the covariance matrix of the disturbances of (10) is diagonal and, thus, the BLUE for the parameters can be computed by a Weighted Least Squares (WLS) procedure. Again, the last observation can be dropped when computing the estimator for β.

3. Autocorrelated random effects

Let now consider the general case of autocorrelated random effects. Since, the individual- and time-effects, λ and μ, are not assumed to be spherically distributed, the variance-covariance matrix of the disturbances Ω is given by (4) and, in general, the eigendecomposition (5) does not hold anymore. However, it is still convenient to consider how the structure of the GLM (10) becomes.

Observing that $W_n^T J_n = 0$, $w_n^T J_n w_n = 1$ and recalling that P_i, $i = 1,2,3,4$ are mutually orthogonal, it can be verified that the variance-covariance matrix of $\begin{pmatrix} u_1^T & u_2^T & u_3^T & u_4 \end{pmatrix}$ is given by

$$\bar{\Omega} = \begin{pmatrix} \sigma_\nu^2 I_{n_1} & 0 & 0 & 0 \\ 0 & \Omega_2 & 0 & \omega_{42} \\ 0 & 0 & \Omega_3 & \omega_{43} \\ 0 & \omega_{42}^T & \omega_{43}^T & \omega_4 \end{pmatrix}, \tag{11}$$

where $\Omega_i = P_i^T \Omega P_i$ and $\omega_{4i} = P_i^T \Omega P_4$, for $i = 1,2,3,4$. More specifically,

$$\Omega_1 = \sigma_\nu^2 I_{n_1}, \tag{12a}$$

$$\Omega_2 = T W_N^T \Psi_\mu W_N + \sigma_\nu^2 I_{n_2}, \qquad \omega_{42} = T W_N^T \Psi_\mu w_N, \tag{12b}$$

$$\Omega_3 = N W_T^T \Psi_\lambda W_T + \sigma_\nu^2 I_{n_3}, \qquad \omega_{43} = N W_T^T \Psi_\lambda w_T \tag{12c}$$

and

$$\omega_4 = N w_T^T \Psi_\lambda w_T + T w_N^T \Psi_\mu w_N + \sigma_\nu^2. \tag{12d}$$

Notice that, by setting $H_\mu \equiv \bar{W}_N^T \Psi_\mu \bar{W}_N$, $H_\lambda \equiv \bar{W}_T^T \Psi_\lambda \bar{W}_T$ with $\bar{W}_n = \begin{pmatrix} w_n & W_n \end{pmatrix}^T$, and partitioning

$$H_\mu = \begin{pmatrix} h_1^\mu & (h_{12}^\mu)^T \\ h_{12}^\mu & H_2^\mu \end{pmatrix}, \qquad \text{and} \qquad H_\lambda = \begin{pmatrix} h_1^\lambda & (h_{12}^\lambda)^T \\ h_{12}^\lambda & H_2^\lambda \end{pmatrix},$$

from (11) and (12) it follows that

$$\bar{\Omega} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & T H_2^\mu & 0 & T h_{12}^\mu \\ 0 & 0 & N H_2^\lambda & N h_{12}^\lambda \\ 0 & T (h_{12}^\mu)^T & N (h_{12}^\lambda)^T & T h_1^\mu + N h_1^\lambda \end{pmatrix} + \sigma_\nu^2 I_{NT}. \tag{13}$$

Now, let consider the Cholesky factorizations

$$T H_\mu + \sigma_\nu^2 I_{n_1+1} = C_\mu C_\mu^T \qquad \text{and} \qquad N H_\lambda + \sigma_\nu^2 I_{n_2+1} = C_\lambda C_\lambda^T, \tag{14}$$

where C_μ and C_λ are upper triangular[1].Then, partitioning the Cholesky factors as

$$C_\mu = \begin{pmatrix} c_\mu & c_{24}^T \\ 0 & C_2 \end{pmatrix} \qquad \text{and} \qquad C_\lambda = \begin{pmatrix} c_\lambda & c_{34}^T \\ 0 & C_3 \end{pmatrix},$$

[1] From the positive semi-definiteness of Ψ_μ and Ψ_λ, it follows that $T H_\mu + \sigma_\nu^2 I_{n_1+1}$ and $N H_\lambda + \sigma_\nu^2 I_{n_2+1}$ are positive definite and, thus, these Cholesky factorizations always exists.

allows to derive the Cholesky factor of $\bar{\Omega} = \bar{C}\bar{C}^T$ as

$$\bar{C} = \begin{pmatrix} \sigma_\nu I_{n_1} & 0 & 0 & 0 \\ 0 & C_2 & 0 & 0 \\ 0 & 0 & C_3 & 0 \\ 0 & c_{24}^T & c_{34}^T & c_4 \end{pmatrix}, \tag{15}$$

where $c_4^2 = c_\mu^2 + c_\lambda^2 - \sigma_\nu^2$.

Notice that when μ is spherically distributed, that is when $\Psi_\mu = \sigma_\mu^2 I_N$, the Cholesky factor C_2 becomes a diagonal matrix and c_{24} vanish. In fact, in that case $H_\mu = \sigma_\mu^2 I_N$ and, by (14), $C_\mu = \sqrt{T\sigma_\mu^2 + \sigma_\nu^2}\, I_N$. Analogously, when λ is spherically distributed, $C_3 = \sqrt{T\sigma_\lambda^2 + \sigma_\nu^2}\, I_{n_2}$ and c_{34} vanishes. The zero elements that arises under these cases should be taken into account for a computationally efficient implementation of the estimation algorithms.

The GLS estimator for the GLM (10) with $\bar{\Omega}$, the disturbances covariance matrix, given by (11) derives from the solution of the Generalized Least Squares problem (GLLSP)

$$\operatorname*{argmin}_{\alpha,\beta} \; \|v_1\|^2 + \|v_2\|^2 + \|v_3\|^2 + \|v_4\|^2, \quad \text{subject to}$$
$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 0 & X_1 \\ 0 & X_2 \\ 0 & X_3 \\ \sqrt{NT} & X_4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \sigma_\nu I_{n_1} & 0 & 0 & 0 \\ 0 & C_2 & 0 & 0 \\ 0 & 0 & C_3 & 0 \\ 0 & c_{42}^T & c_{43}^T & c_4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}, \tag{16}$$

where $v_i \sim (0, I_{n_i})$ are uncorrelated [21, 24, 32]. It follows that $\hat{\beta}$, the GLS estimator for β, comes from the solution the GLLSP

$$\operatorname*{argmin}_{\beta} \; \|v_1\|^2 + \|v_2\|^2 + \|v_3\|^2, \quad \text{subject to}$$
$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \beta + \begin{pmatrix} \sigma_\nu I_{n_1} & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & C_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \tag{17}$$

and that the GLS estimator for α is given by

$$\hat{\alpha} = (y_4 - X_4\hat{\beta} - c_{42}^T\hat{v}_2 - c_{43}^T\hat{v}_3)/\sqrt{NT}$$

where $\hat{v}_2$, $\hat{v}_3$ and the GLS estimator $\hat{\beta}$ come from the solution of the GLLSP (17). Notice that, only Cholesky and orthogonal factorizations, and no matrix inversions, have been used to formulate the GLLSP (17). Thus, this approach results numerically stable even in the case of nearly singular covariance matrices.

The GLLSP (17) is naturally solved by using the Generalized QR decomposition (GQRD) of the matrices

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} C_1 & 0 & 0 \\ 0 & C_2 & 0 \\ 0 & 0 & C_3 \end{pmatrix},$$

where $C_1 = \sigma_\nu I_{n_1}$. Alternatively, given the block structure of the Cholesky factor, a convenient strategy consists on computing the GQRDs of X_i and C_i indipendetly, for $i = 1,2,3$ and then use an updating GQRD techniques to retrieve the whole GQRDs. Specifically, let consider the GQRDs

$$\begin{pmatrix} \hat{Q}_i^T \\ \bar{Q}_i^T \end{pmatrix} X_i = \begin{pmatrix} R_i \\ 0 \end{pmatrix}, \qquad \text{and} \qquad \begin{pmatrix} \hat{Q}_i^T \\ \bar{Q}_i^T \end{pmatrix} C_i \begin{pmatrix} \hat{P}_i & \bar{P}_i \end{pmatrix} = \begin{pmatrix} \hat{C}_i & \tilde{C}_i \\ 0 & \bar{C}_i \end{pmatrix}, \tag{18}$$

for $i = 1,2,3$ and let

$$\begin{pmatrix} \hat{y}_i \\ \bar{y}_i \end{pmatrix} = \begin{pmatrix} \hat{Q}_i^T \\ \bar{Q}_i^T \end{pmatrix} y_i.$$

Notice that, $C_1 = \sigma_\nu I_{n_1}$ and thus the first GQRD is actually a simple QR decomposition, that is $\hat{P}_1 = \hat{Q}_1$, $\bar{P}_1 = \bar{Q}_1$, $\hat{C}_1 = \sigma_\nu I_K$, $\bar{C}_1 = \sigma_\nu I_{n_1-K}$ and $\tilde{C}_1 = 0$.

Next, premultiplying the ith block of the constraints (17) by $\begin{pmatrix} \hat{Q}_i & \bar{Q}_i \end{pmatrix}^T$ and rearranging it gives the equivalent GLLSP

$$\operatorname*{argmin}_{\beta} \sum_{i=1,2,3} \|\hat{v}_i\|^2 + \|\bar{v}_i\|^2, \quad \text{subject to}$$

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ 0 \\ 0 \\ 0 \end{pmatrix} \beta + \begin{pmatrix} \sigma_\nu I_K & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{C}_2 & 0 & 0 & \tilde{C}_2 & 0 \\ 0 & 0 & \hat{C}_3 & 0 & 0 & \tilde{C}_3 \\ 0 & 0 & 0 & \sigma_\nu I_{n_1-K} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{C}_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \bar{C}_3 \end{pmatrix} \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \hat{v}_3 \\ \bar{v}_1 \\ \bar{v}_2 \\ \bar{v}_3 \end{pmatrix}, \tag{19}$$

where $\hat{v}_i = \hat{P}_i^T v_i$ and $\bar{v}_i = \bar{P}_i^T v_i$. It follows that, when $\hat{C}_2$ and $\hat{C}_3$ are non-singular, $\bar{v}_i = \bar{C}_i^{-1} \bar{y}_i$, $i = 1,2,3$ and thus the GLLSP (19) is equivalent to the smaller in size $3K \times 3K$ GLLSP

$$\operatorname*{argmin}_{\beta} \sum_{i=1,2,3} \|\hat{v}_i\|^2, \quad \text{subject to}$$

$$\begin{pmatrix} \check{y}_1 \\ \check{y}_2 \\ \check{y}_3 \end{pmatrix} = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} \beta + \begin{pmatrix} \sigma_\nu I_K & 0 & 0 \\ 0 & \hat{C}_2 & 0 \\ 0 & 0 & \hat{C}_3 \end{pmatrix} \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \hat{v}_3 \end{pmatrix}, \tag{20}$$

where $\check{y}_i = \hat{y}_i - \tilde{C}_i \bar{C}_i^{-1} \bar{y}_i$, for $i = 1,2,3$. Finally this GLLSP can be solved by means of the GQRD of the regressor and Cholesky factor matrices. Algorithm 1 resumes the steps needed for the estimation of the parameters.

Algorithm 1 Estimation of the two-way non-spherically distributed error component model.

1: Compute $X_i = P_i^T X$ and $y_i = P_i^T y$.
2: Compute the Cholesky factorizations in (14).
3: Compute the GQRDs (18) and compute $\check{y}_i = \hat{y}_i - \hat{\bar{C}}_i \bar{C}_i^{-1} \bar{y}_i$
4: Obtain the estimator $\hat{\beta}$ by solving (20) with a GQRD approach

In order to derive the computational complexity of this approach, let firstly recall that the cost of computing the QRD of an $M \times K$ matrix is $O(K^2 M)$ flops, that of computing the GQRD of two matrices of dimension $M \times K$ and $M \times M$ is $O(M^3)$ flops, while the computation of the Choleky factor of an $M \times M$ matrix require $O(M^3)$ flops [18]. Let consider Now, the most expensive steps. The computation of the Cholesky factor C_2 and C_3, in step 2, needs $O(n_2^3 + n_3^3)$ flops, step 3 require $O(K^2 n_1 + n_2^3 + n_3^3)$ flops, (since one of the GQRDs is a simple QRD). Finally, the last step need $O(K^3)$ flops for the computation of the corresponding GQRD. Thus, the overall cost is given by

$$O(K^2 n_1 + n_2^3 + n_3^3) = O(NTK^2 + K^3 + N^3 + T^3) \tag{21}$$

flops, which is remarkably smaller than the cost, $O(N^3 T^3)$ flops, required for computing the GQRD corresponding to the original model (2). A more computationally efficient algorithm can be designed by using updating GQRD techniques to exploit the upper-triangular structure of the blocks of the matrices in the GLLSP (20) [38].

Notice that, when the parameters are reestimated for different covariance parameters, the QRDs in the GQRDs of step 3 are already available. Most notably, only the second and third RQDs in (18) need to be computed and thus, the cost of re-estimate the parameters reduces to $O(K^3 + N^3 + T^3)$.

4. SUR Model with two-way error component disturbances

Let generalize the linear regression model (2) to the set of Seemingly Unrelated Regressions (SUR) with Error Component disturbances (SUR-EC)

$$y_j = \iota_{NT} \alpha_j + X_j \beta_j + u_j, \qquad j = 1, \dots, G \tag{22a}$$

with

$$u_j = (\iota_N \otimes I_T)\lambda_j + (I_N \otimes \iota_T)\mu_j + \nu_j, \qquad j = 1, \dots, G \tag{22b}$$

where $y_j, u, \nu_j \in \mathbb{R}^{NT}$, $X_j \in \mathbb{R}^{NT \times K_j}$, $\alpha_j \in \mathbb{R}$, $\beta_j \in \mathbb{R}^{K_j}$, $\lambda_j \in \mathbb{R}^T$ and $\mu_j \in \mathbb{R}^N$. Furthermore, the random effects λ_j, μ_j and ν_j have zero mean and covariances given by

$$\mathrm{Cov}\left(\begin{pmatrix} \lambda_i \\ \mu_i \\ \nu_i \end{pmatrix}, \begin{pmatrix} \lambda_j \\ \mu_j \\ \nu_j \end{pmatrix} \right) = \begin{pmatrix} \Psi_{ij}^{\lambda} & 0 & 0 \\ 0 & \Psi_{ij}^{\mu} & 0 \\ 0 & 0 & \sigma_{ij}^{\nu} I_{NT} \end{pmatrix} \tag{23}$$

for $i, j = 1, \ldots, G$.

The estimation of the SUR model (22) is approached by following the method proposed in the previous section for the single equation case. That is, each equation in (22a) is premultiplyied by P^T and that results in the set of regressions

$$\begin{pmatrix} y_{1,i} \\ y_{2,i} \\ y_{3,i} \\ y_{4,i} \end{pmatrix} = \begin{pmatrix} 0 & X_{1,i} \\ 0 & X_{2,i} \\ 0 & X_{3,i} \\ \sqrt{NT} & X_{4,i} \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + \begin{pmatrix} u_{1,i} \\ u_{2,i} \\ u_{3,i} \\ u_{4,i} \end{pmatrix}, \qquad i = 1, \ldots, G \tag{24}$$

where $y_{l,i} = P_l^T y_i$, $X_{l,i} = P_l^T X_i$ and $u_{l,i} = P_l^T u_i$ for $l = 1, \ldots, 4$ and $i = 1, \ldots, G$.

The covariance matrix of $\begin{pmatrix} u_{1,i}^T & u_{2,i}^T & u_{3,i}^T & u_{4,i}^T \end{pmatrix}$ and $\begin{pmatrix} u_{1,j}^T & u_{2,j}^T & u_{3,j}^T & u_{4,j}^T \end{pmatrix}$ has the same structure of $\bar{\Omega}$ in (11) and is given by

$$\bar{\Omega}_{ij} = \begin{pmatrix} \sigma_{ij}^{\nu} I_{n_1} & 0 & 0 & 0 \\ 0 & \Omega_{2,ij} & 0 & \omega_{42,ij} \\ 0 & 0 & \Omega_{3,ij} & \omega_{43,ij} \\ 0 & \omega_{42,ij}^T & \omega_{43,ij}^T & \omega_{4,ij} \end{pmatrix} \tag{25}$$

where

$$\Omega_{2,ij} = TW_N^T \Psi_{ij}^{\mu} W_N + \sigma_{ij}^{\nu} I_{n_2}, \qquad \omega_{42,ij} = TW_N^T \Psi_{ij}^{\mu} w_N$$
$$\Omega_{3,ij} = NW_T^T \Psi_{ij}^{\lambda} W_T + \sigma_{ij}^{\nu} I_{n_3}, \qquad \omega_{43,ij} = NW_T^T \Psi_{ij}^{\lambda} w_T$$

and

$$w_{4,ij} = Tw_N^T \Psi_{ij}^{\mu} w_N + Nw_T^T \Psi_{ij}^{\lambda} w_T + \sigma_{ij}^{\nu},$$

for $i, j = 1, \ldots, G$.

Now, the system of regressions can be reassembled as the equivalent GLM

$$\begin{pmatrix} \breve{y}_1 \\ \breve{y}_2 \\ \breve{y}_3 \\ \breve{y}_4 \end{pmatrix} = \begin{pmatrix} 0 & \oplus_i X_{1,i} \\ 0 & \oplus_i X_{1,i} \\ 0 & \oplus_i X_{1,i} \\ \sqrt{NT}\, I_G & \oplus_i X_{1,i} \end{pmatrix} \begin{pmatrix} \mathrm{Vec}\{\alpha_i\} \\ \mathrm{Vec}\{\beta_i\} \end{pmatrix} + \begin{pmatrix} \breve{u}_1 \\ \breve{u}_2 \\ \breve{u}_3 \\ \breve{u}_4 \end{pmatrix}$$

where $\breve{y}_l = \mathrm{Vec}\{y_{l,i}\}$, $\breve{u}_l = \mathrm{Vec}\{u_{l,i}\}$, and the disturbances have dispersion matrix given by

$$\mathrm{Cov}(\begin{pmatrix} \breve{u}_1 \\ \breve{u}_2 \\ \breve{u}_3 \\ \breve{u}_4 \end{pmatrix}) = \breve{\Omega} = \begin{pmatrix} \Sigma_{\nu} \otimes I_{n_1} & 0 & 0 & 0 \\ 0 & \bar{\Omega}_2 & 0 & \bar{\Omega}_{42} \\ 0 & 0 & \bar{\Omega}_3 & \bar{\Omega}_{43} \\ 0 & \bar{\Omega}_{42}^T & \bar{\Omega}_{43}^T & \bar{\Omega}_4 \end{pmatrix}, \tag{26}$$

where

$$\bar{\Omega}_x = \begin{pmatrix} \Omega_{x,11} & \cdots & \Omega_{x,1G} \\ \vdots & & \vdots \\ \Omega_{x,G1} & \cdots & \Omega_{x,GG} \end{pmatrix}, \qquad \bar{\Omega}_{4x} = \begin{pmatrix} \omega_{4x,11} & \cdots & \omega_{4x,G1} \\ \vdots & & \vdots \\ \omega_{4x,1G} & \cdots & \omega_{4x,GG} \end{pmatrix}$$

for $x = 2, 3$ and

$$\bar{\Omega}_4 = \begin{pmatrix} \omega_{4,11} & \cdots & \omega_{4,1G} \\ \vdots & & \vdots \\ \omega_{4,G1} & \cdots & \omega_{4,GG} \end{pmatrix}.$$

Now, in parallel with (13) for the univariate case, $\check{\Omega}$ can be written as

$$\check{\Omega} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & TH_2^{\mu} & 0 & TH_{12}^{\mu} \\ 0 & 0 & NH_2^{\lambda} & NH_{12}^{\lambda} \\ 0 & T(H_{12}^{\mu})^T & N(H_{12}^{\lambda})^T & TH_1^{\mu} + NH_1^{\lambda} \end{pmatrix} + \begin{pmatrix} \Sigma_\nu \otimes I_{n_1} & 0 & 0 & 0 \\ 0 & \Sigma_\nu \otimes I_{n_2} & 0 & 0 \\ 0 & 0 & \Sigma_\nu \otimes I_{n_3} & 0 \\ 0 & 0 & 0 & \Sigma_\nu \end{pmatrix} \tag{27}$$

where

$$\begin{aligned} H_1^{\mu} &= (I_G \otimes w_N)^T \bar{\Psi}_{\mu} (I_G \otimes w_N), & H_1^{\lambda} &= (I_G \otimes w_N)^T \bar{\Psi}_{\lambda} (I_G \otimes w_N), \\ H_{12}^{\mu} &= (I_G \otimes w_N)^T \bar{\Psi}_{\mu} (I_G \otimes W_N), & H_{12}^{\lambda} &= (I_G \otimes w_T)^T \bar{\Psi}_{\lambda} (I_G \otimes W_T), \\ H_2^{\mu} &= (I_G \otimes W_T)^T \bar{\Psi}_{\mu} (I_G \otimes W_T), & H_2^{\lambda} &= (I_G \otimes W_T)^T \bar{\Psi}_{\lambda} (I_G \otimes W_T). \end{aligned}$$

Let

$$TH_{\mu} + \begin{pmatrix} \Sigma_\nu & 0 \\ 0 & \Sigma_\nu \otimes I_{n_2} \end{pmatrix} = C_{\mu} C_{\mu}^T \quad \text{and} \quad NH_{\lambda} + \begin{pmatrix} \Sigma_\nu & 0 \\ 0 & \Sigma_\nu \otimes I_{n_3} \end{pmatrix} = C_{\lambda} C_{\lambda}^T \tag{28}$$

and let partition

$$C_{\mu} = \begin{pmatrix} \check{C}_{\mu} & \check{C}_{42} \\ 0 & \check{C}_2 \end{pmatrix} \begin{matrix} G \\ Gn_2 \end{matrix} \quad \text{and} \quad C_{\lambda} = \begin{pmatrix} \check{C}_{\lambda} & \check{C}_{43} \\ 0 & \check{C}_3 \end{pmatrix} \begin{matrix} G \\ Gn_2 \end{matrix}.$$

Then the Cholesky factor in $\check{\Omega} = \check{C}\check{C}^T$ is given by

$$\begin{pmatrix} C_\nu \otimes I_{n_1} & 0 & 0 & 0 \\ 0 & \check{C}_2 & 0 & 0 \\ 0 & 0 & \check{C}_3 & 0 \\ 0 & \check{C}_{42} & \check{C}_{43} & \check{C}_4 \end{pmatrix} \begin{matrix} Gn_1 \\ Gn_2 \\ Gn_3 \\ G \end{matrix},$$

where C_ν and C_4 derive, respectively, from the Cholesky decompositions $\Sigma_\nu = C_\nu C_\nu^T$ and

$$\check{C}_\mu \check{C}_\mu^T + \check{C}_\lambda \check{C}_\lambda^T - \Sigma_\nu = \check{C}_4 \check{C}_4^T. \tag{29}$$

In order to compute the GLS estimate of the parameters, let rewrite the estimation problem of SUR model as the GLLSP

$$\underset{\alpha_i,\beta_i,i=1,\ldots,G}{\operatorname{argmin}} \sum_{i=1}^{4} \|\check{\nu}_i\|^2$$

$$\begin{pmatrix} \check{y}_1 \\ \check{y}_2 \\ \check{y}_3 \\ \check{y}_4 \end{pmatrix} = \begin{pmatrix} 0 & \oplus_i X_{1,i} \\ 0 & \oplus_i X_{2,i} \\ 0 & \oplus_i X_{3,i} \\ \sqrt{NT} I_G & \oplus_i X_{4,i} \end{pmatrix} \begin{pmatrix} \mathrm{Vec}\{\alpha_i\} \\ \mathrm{Vec}\{\beta_i\} \end{pmatrix} + \begin{pmatrix} C_\nu \otimes I_{n_1} & 0 & 0 & 0 \\ 0 & \check{C}_2 & 0 & 0 \\ 0 & 0 & \check{C}_3 & 0 \\ 0 & \check{C}_{42} & \check{C}_{43} & \check{C}_4 \end{pmatrix} \begin{pmatrix} \check{\nu}_1 \\ \check{\nu}_2 \\ \check{\nu}_3 \\ \check{\nu}_4 \end{pmatrix}. \tag{30}$$

Like the GLLSP (16), also the GLLSP (30) can be exactly solved in two stages. In the first stage the GLS esimator $\hat{\beta}_i$ are computed as the solution of the GLLSP

$$\underset{\beta_i,i=1,\ldots,G}{\operatorname{argmin}} \sum_{i=1}^{3} \|\check{\nu}_i\|^2 \qquad \text{s.t.}$$

$$\begin{pmatrix} \check{y}_1 \\ \check{y}_2 \\ \check{y}_3 \end{pmatrix} = \begin{pmatrix} \oplus_i X_{1,i} \\ \oplus_i X_{2,i} \\ \oplus_i X_{3,i} \end{pmatrix} \mathrm{Vec}\{\beta_i\} + \begin{pmatrix} C_\nu \otimes I_{n_1} & 0 & 0 \\ 0 & \check{C}_2 & 0 \\ 0 & 0 & \check{C}_3 \end{pmatrix} \begin{pmatrix} \check{\nu}_1 \\ \check{\nu}_2 \\ \check{\nu}_3 \end{pmatrix}. \tag{31}$$

Then, in the second stage the estimator for α are computed as

$$\hat{\alpha} = \frac{1}{\sqrt{NT}} (\check{y}_4 - \mathrm{Vec}\{X_{4,i}\hat{\beta}_i\} - \check{C}_{42}\hat{v}_2 - \check{C}_{43}\hat{v}_3)$$

where $\hat{v}_2, \hat{v}_3$ and the GLS estimator $\hat{\beta}_i$ come from the optimum of the GLLSP (31).

It is clear that the computation of the solution of the GLLSP (31), which as $G(NT-1)$ constraints, represents the most demanding task in the estimation of the SUR-EC model (22) and requires the computation of the GQRD of the matrices

$$\begin{pmatrix} \oplus_i X_{1,i} \\ \oplus_i X_{2,i} \\ \oplus_i X_{3,i} \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} C_\nu \otimes I_{n_1} & 0 & 0 \\ 0 & \check{C}_2 & 0 \\ 0 & 0 & \check{C}_3 \end{pmatrix}. \tag{32}$$

Alternatively, the solution can be derived, following the same approach illustrated in section 3, by using updating GQRDs. An efficient implementation of those factorization should exploit the structure of the matrices involved. Algorithms for computing the GQRD of the first block of the two matrices have already been considered in the context of the estimation of the standard SUR model [14, 20, 21, 23]. Next, when updating this GQRD the algorithm can exploit the upper triangular structure of $\check{C}_2$ and $\check{C}_3$ [22, 38].

Notice that, if $\Psi_{ij}^\mu = \sigma_{ij}^\mu I_N$ or $\Psi_{ij}^\lambda = \sigma_{ij}^\lambda I_T$ for $i, j = 1, \ldots, G$, then $\bar{\Omega}_2 = (T\Sigma_\mu + \Sigma_\nu) \otimes I_{n_2}$ or $\bar{\Omega}_3 = (N\Sigma_\lambda + \Sigma_\nu) \otimes I_{n_3}$ and $\bar{\Omega}_{42} = 0$ or $\bar{\Omega}_{43} = 0$, respectively. The model is simpler also when the effects do not have correlations across equations, for example when $\Psi_{ij}^\mu = 0$ for $i \neq j$, then $\bar{\Omega}_2 = \oplus_i T W_N^T \Psi_{ii}^\mu W_N + \Sigma_\nu \otimes I_{n_1}$ and $\bar{\Omega}_{42} = \oplus_i T W_N^T \Psi_{ii}^\mu w_N$.

4.1. Special Cases

In the following some special cases of the SUR-EC model are considered. In particular various assumptions are imposed on the covariance matrices of individual effects Ψ^{μ}_{ij}, $i, j = 1, \ldots, G$. The resulting simplifications on the matrices involved in the estimation and the design of the procedure is discussed. Similar considerations hold for the time-effects λ_i.

4.1.1. Spherically distributed individual effects: $\Psi^{\mu}_{ij} = \sigma^{\mu}_{ij} I_N$

Let assume that $\Psi^{\mu}_{ij} = \sigma^{\mu}_{ij} I_N$, for $i, j = 1, \ldots, G$ and let denote $\Sigma_\mu \in \mathbb{R}^{G\times G}$ the matrix with elements σ^{μ}_{ij}. Under that assumption $\bar{\Omega}_{42,ij}$ vanishes and the expressions for $\bar{\Omega}_{2,ij}$ simplifies to

$$\bar{\Omega}_{2,ij} = (T\sigma^{\mu}_{ij} + \sigma^{\nu}_{ij}) I_{n_2}.$$

Thus, $\bar{\Omega}_{ij}$ becomes

$$\bar{\Omega}_{ij} = \begin{pmatrix} \sigma^{\nu}_{ij} I_{n_1} & 0 & 0 & 0 \\ 0 & (T\sigma^{\mu}_{ij} + \sigma^{\nu}_{ij}) I_{n_2} & 0 & 0 \\ 0 & 0 & \Omega_{3,ij} & \omega_{43,ij} \\ 0 & 0 & \omega^{T}_{43,ij} & \omega_{4,ij} \end{pmatrix}$$

Similarly, because $\bar{\Psi}_\mu = \Sigma_\mu \otimes I_T$, H^{μ}_{12} will become zero, $H^{\mu}_1 = \Sigma_\mu$ and $H^{\mu}_2 = \Sigma_\mu \otimes I_{n_2}$. Thus, by (28), $\check{C}_\mu$ is the Cholesky factor of $N\Sigma_\mu + \Sigma_\nu$, that is $\check{C}_\mu \hat{C}^{T}_\mu = N\Sigma_\mu + \Sigma_\nu$, and $\check{C}_2 = \check{C}_\mu \otimes I_{n_2}$. It follows that the computation of the estimators for β_i, $i = 1, \ldots, G$, requires now the GQRD of the matrices

$$\begin{pmatrix} \oplus_i X_{1,i} \\ \oplus_i X_{2,i} \\ \oplus_i X_{3,i} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} C_\nu \otimes I_{n_1} & 0 & 0 \\ 0 & \check{C}_\mu \otimes I_{n_2} & 0 \\ 0 & 0 & \check{C}_3 \end{pmatrix},$$

which can be computed as follows. The GQRD of $\oplus_i X_{1,i}$ and $C_\nu \otimes I_{n_1}$ is computed, next the result is updated with the observations in the matrices $\oplus_i X_{2,i}$ and $\check{C}_\mu \otimes I_{n_2}$ and finally the observation in the last blocks of the matrices are added. The first step is identical to the GQRD of a standard SUR model and the second step is the same of that used in the problem of updating a SUR model. Algorithms to tackle these two problems have already been proposed [14, 22].

4.1.2. Individual effects without correlation between equations: $\Psi^{\mu}_{ij} = 0$ for $i \neq j$

In the following it will be assumed that $\Psi^{\mu}_{ij} = 0$ for $i \neq j$ and $i, j = 1, \ldots, G$. Thus, H^{μ}_1, H^{μ}_{12} and H^{μ}_2 in (27) are given by

$$H^{\mu}_1 = \operatorname{diag}(w^{T}_N \Psi^{\mu}_{ii} w_N) \qquad H^{\mu}_{12} = \bigoplus_{i=1}^{G} w^{T}_N \Psi^{\mu}_{ii} W_N \qquad \text{and} \qquad H^{\mu}_1 = \bigoplus_{i=1}^{G} W^{T}_N \Psi^{\mu}_{ii} W_N$$

and to compute C_μ in (28) it is necessary to compute the Cholesky decomposition of the $GN \times GN$ matrix

$$\begin{pmatrix} \Sigma_\nu + TH_1^\mu & T(\oplus_i w_N^T \Psi_{ii}^\mu W_N) \\ T(\oplus_i W_N^T \Psi_{ii}^\mu w_N) & T(\oplus_i W_N^T \Psi_{ii}^\mu W_N) + (\Sigma_\nu \otimes I_{n_2}) \end{pmatrix}$$

which has the sparse structure illustrated in figure 1.

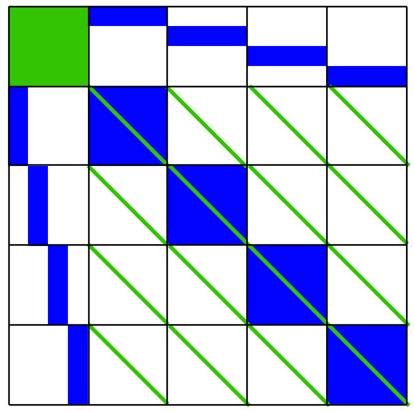

Figure 1: Structure of the matrix in (27), where parts of the matrix which come from elements of Σ_ν and Ψ_{ij}^μ are represented in green and blue, respectively.

5. Conclusions

The estimation of Panel Data models, regressions with two-way error component disurbances, is considered for the case when both the random effects are non-spherically distributed. The usual approach that firstly transforms the effects into uncorrelated errors, for example by applying a Prais-Winsten transformation, and then applies within and between transformations, cannot be conveniently applied when both the effetcs are autocorrelated [8, 9]. The proposed approach reverts this scheme by firstly applying the within and between transformations. The covariance matrix of the resulting General Linear Model (GLM) has a simple structure that allows its partitioning into three smaller GLMs. Furthermore, the within and between transformations considered produce a model which is smaller than those usually derived, allowing for a more computationally efficient estimation. A further reduction in the computations arises when the model is re-estimated for different covariance parameters.

In order to show the advantages of the proposed approach, the same technique is applied to the case of Seemingly Unrelated Regressions with non-spherically distributed error components disturbances [3, 6]. In a similar way the Simultaneous Equation model with error component disturbances can be approached [4, 7, 25, 28].

Future research is needed on the inference side, especially the estimation of the covariance matrices in the present context should be considered. However, it should be noticed that, here the models have been transformed and partitioned into blocks which depend on the single covariance matrices and the residuals can be used to compute, or update, an estimator for them. Another direction of research consists into applying this approach to more specific models of the correlations, like autoregressive or moving average random effects.

Further research is also required for the development of computationally efficient and/or parallel implementation of the estimation algorithms. This is more important in the SUR case where the dimension of the resulting model to be solved can become immediately large as it is given by the product of the number of individuals (N), the number of samples (T) and the number of equations (G).

References

[1] T. Amemiya, *The estimation of the variances in a variance-components model*, International Economic Review **12** (1971), 1–13.

[2] T.W. Anderson and C. Hsiao, *Formulation and estimation of dynamic models using panel data*, Journal of Econometrics **18** (1982), 47–82.

[3] R.B. Avery, *Error components and seemingly unrelated regressions*, Econometrica **45** (1977), 199–209.

[4] P. Balestra and J.Krishnakumar, *Full information estimations of a system of simultaneous equations with error components structure*, Econometric Theory **3** (1987), 223–246.

[5] P. Balestra and M. Nerlove, *Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas*, Econometrica **34** (1966), 585–612.

[6] B.H. Baltagi, *On seemingly unrelated regressions with error components*, Econometrica **48** (1980), 1547–1551.

[7] ______, *Simultaneous equations with error components*, Journal of Econometrics **17** (1981), 189–200.

[8] ______, *Econometric analysis of panal data*, 2nd ed., John Wiley and Sons, 2001.

[9] B.H. Baltagi and Q. Li, *A transformation that will cirumvent the problem of autocorrelation in an error component model*, Journal of Econometrics **48** (1991), 385–393.

[10] ______, *Estimating error component models with general ma(q) disturbances*, Econometric Theory **10** (1994), 396–408.

[11] C.G. Broyden and M.T. Vespucci, *Krylov solvers for linear algebraic systems*, Studies in Computational Mathematics, vol. 11, Elsevier, 2004.

[12] G. Chamberlain, *Multivariate regression models for panel data*, Journal of Econometrics **18** (1982), 5–46.

[13] ______, *Panel data*, ch. 22, pp. 1247–1318, North-Holland, Amsterdam, 1984.

[14] P. Foschi, D.A. Belsley, and E.J. Kontoghiorghes, *A comparative study of algorithms for solving seemingly unrelated regressions models*, Computational Statistic & Data Analysis **44** (2003), 3–35.

[15] W.A. Fuller and G.E. Battese, *Transformations for estimation of linear models with nested error structure*, Journal of American Statistical Association **68** (1973), 626–632.

[16] ______, *Estimation of linear models with cross-error structure*, Journal of Econometrics **2** (1974), 67–78.

[17] J.W. Galbraith and V. Zinde-Walsh, *Transforming the error-component model for estimation with general arma disturbances*, Journal of Econometrics **66** (1995), 349–355.

[18] G.H. Golub and C.F. Van Loan, *Matrix computations*, 3ed ed., Johns Hopkins University Press, Baltimore, Maryland, 1996.

[19] S. Karlsson and J. Skoglund, *Maximum-likelihood based inference in the two-way random effects model with serially correlated time effects*, Empirical Economics **29** (2004), 79–88.

[20] E.J. Kontoghiorghes, *Inconsistencies and redundancies in SURE models: computational aspects*, Computational Economics **16** (2000), no. 1+2, 63–70.

[21] ______, *Parallel algorithms for linear models: Numerical methods and estimation problems*, Advances in Computational Economics, vol. 15, Kluwer Academic Publishers, Boston, MA, 2000.

[22] ______, *Computational methods for modifying seemingly unrelated regressions models*, Journal of Computational and Applied Mathematics **162** (2004), no. 1, 247–261.

[23] E.J. Kontoghiorghes and M. R. B. Clarke, *An alternative approach for the numerical solution of seemingly unrelated regression equations models*, Computational Statistics & Data Analysis **19** (1995), no. 4, 369–377.

[24] S. Kourouklis and C.C. Paige, *A constrained least squares approach to the general Gauss–Markov linear model*, Journal of the American Statistical Association **76** (1981), no. 375, 620–625.

[25] J. Krishnakumar, *Estimating simultaneous equation models with error components structure*, Springer-Verlag, Berlin, 1988.

[26] L.F. Lee, *Estimation of autocorrelated error components model with panel data, working paper*, Tech. report, Department of Economics, University of Minnesota, 1979.

[27] L.A. Lillard and R.J. Willis, *Dynamic aspects of earning mobility*, Econometrica **47** (1978), 985–1012.

[28] J.R. Magnus and A.D. Wooland, *On the maximum likelihood estimation of multivariate regression models containing serially correlated error components*, Economic Review **29** (1988), 707–725.

[29] P. Mazodier and A. Trognon, *Heteroskedasticity and strification in error component models*, Annales de l'INSEE **30-31** (1978), 451–482.

[30] M.A. McCurdy, *The use of time series processes to model the error structure of earnings in a longitudinal data analysis*, Journal of Econometrics **18** (1982), 83–114.

[31] M. Nerlove, *A note on error components models*, Econometrica **39** (1971), 383–396.

[32] C.C. Paige, *Numerically stable computations for general univariate linear models*, Communications on Statistical and Simulation Computation **7** (1978), no. 5, 437–453.

[33] N.S. Revankar, *Error component models with serial correlated time effects*, Journal of the Indian Statistical Association **17** (1979), 137–160.

[34] J. Skoglund and S. Karlsson, *Specification and estimation of random effects models with correlation of general form*, Tech. Report 433, Stockholm School of Economics, 2001.

[35] T.D. Wallace and A. Hussain, *The use of error components models in combining cross-section and time-series data*, Econometrica **37** (1969), 55–72.

[36] T.J. Wansbeek and A. Kapteyn, *A note on the spectral decomposition and maximum likelihood estimation of anova models with balanced data*, Statistics and Probability **1** (1982), 213–215.

[37] ______, *A simple way to obtain the spectral decomposition of variance components models for balanced data*, Communications in Statistics **A11** (1982), 2105–2112.

[38] P. Yanev and E.J. Kontoghiorghes, *Updating the SUR model*, Tech. report, Univ. of Neuchatel, 2005, to be submitted.

A. Derivation of the eigenvectors

Here, a couple of choices for W_n, and thus P_i, are presented. In the first, the correlations of μ and λ are not taken into account and provide a simple approach for a closed form expression for W_n. Beside its simplicity, its main advantage is the easy updating when new observations or individuals are added. In the second approach, the correlation structure of the random effects are taken into account in order to reduce the non-zero elements of the covariance matrix in (11).

The first choice for W_n is given by

$$W_n = \begin{pmatrix} 1 & 1/2 & 1/3 & \cdots & 1/(n-2) & 1/(n-1) \\ -1 & 1/2 & 1/3 & \cdots & 1/(n-2) & 1/(n-1) \\ 0 & -1 & 1/3 & \cdots & 1/(n-2) & 1/(n-1) \\ 0 & 0 & -1 & \cdots & 1/(n-2) & 1/(n-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1/(n-1) \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix} D_n, \tag{33}$$

where $D_n = \text{diag}(\sqrt{i/(i+1)}, i = 1,2,\ldots,N-1)$. The interpretation is the following. The eigenvector w_n corresponds to the Within transformation and applying w_n^T is equivalent to compute the mean scaled by a factor of $\sqrt{n}$. The matrix W_n corresponds to the Between transformation and applying W_n^T corresponds to compute, for the i-th element of the vector, the deviation from the mean of the previous $i-1$ elements, weighted by $\sqrt{i/(i+1)}$ ($i=2,\ldots,n$). Thus, appplying P_i ($i=1,2,3,4$) consists on either taking the mean or the "deviations from the mean" along time and across individuals. Thus, an advantage of this choice is the easy of updating when new observations are added.

The second choice derives by chosing $\bar{W}_n \equiv \begin{pmatrix} w_n & W_n \end{pmatrix}$ as the orthogonal factor in the QRD of the Krylov matrix $\mathrm{Kr}(w_n, A) \equiv \begin{pmatrix} w_n\, Aw_n & \cdots & A^{n-1}w_n \end{pmatrix}$, where $A = \Psi_\mu, \Psi_\lambda$. Specifically, this can be efficiently computed by using a Lanczos/Arnoldi algorithm [11, 18]. Since A is symmteric, this algorithm allows to compute the orthogonal matrix $\bar{W}_n$ such that $H = \bar{W}_n^T A \bar{W}_n$ is tridiagonal.

Thus, using $A = \Psi_\mu$ and $A = \Psi_\lambda$ for the computation of W_N and W_T, respectively, provides H_μ and H_λ being tridiagonal. Furthermore, their Cholesky factors C_λ and C_μ become upper-bidiagonal, this results in an upper-bidiagonal structure for the Cholesky factor in the GLLSP (16) which can be exploited to derive computationally efficient algorithms for the computation of the GQRDs in (18).