MEMORANDUM

No 20/99

Estimating Regression Systems from unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure

By
Erik Biørn

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/sosoek/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 19	By Knut Røed and Steinar Strøm: Progresssive Taxes and the Labour Market – Is the Trade-Off between Equality and Efficiency Inevitable? 39 p.
No 18	By Erik Hernæs: Fewer in Number but Harder to Employ: Incidence and Duration of Unemployment in an Economic Upswing. 28 p.
No 17	By Erik Hernæs, Marte Sollie and Steinar Strøm: Early retirement and Economic Incentives. 27 p.
No 16	By Oddbjørn Raaum and Tom Erik Aabø: The effect of schooling on earnings: The role of family background studied by a large sample of Norwegian twins. 31 p.
No 15	By T. Hægeland, T.J. Klette and K.G. Salvanes: Declining returns to education in Norway? Comparing estimates across chorts, sectors and overtime. 21 p.
No 14	By Knut Røed and Tao Zhang:Unemployment Duration in a Non-Stationary Macroeconomic Environment. 32 p.
No 13	By Tor Jakob Klette, Jarle Møen and Zvi Griliches: Do subsidies to commercial R&D reducemarket failures? Microeconometric evaluation studies. 36 p.
No 12	By Steinar Holden: Wage setting under different Monetary regimes. 40 p.
No 11	By Finn R. Førsund and Kjell Ove Kalhagen: Efficiency and Productivity of Norwegian Colleges. 38 p.
No 10	By Asbjørn Rødseth and Ragnar Nymoen: Nordic wage formation and unemployment seven years later. 52 p.

A complete list of this memo-series is available in a PDF® format at:
http://www..sv.uio.no/sosoek/memo/

ESTIMATING REGRESSION SYSTEMS FROM UNBALANCED PANEL DATA: A STEPWISE MAXIMUM LIKELIHOOD PROCEDURE *)

by

ERIK BIØRN

ABSTRACT

In this paper, we consider the formulation and estimation of systems of regression equations with random individual effects in the intercept terms from unbalanced panel data, *i.e.*, panel data where the individual time series have unequal length. Generalized Least Squares (GLS) estimation and Maximum Likelihood (ML) estimation are discussed. A stepwise algorithm for solving the ML problem is developed.

Keywords: Panel Data. Unbalanced panels. Regression equation systems. Maximum Likelihood. Heterogeneity. Covariance estimation

JEL classification: C13, C23, C33

---

* I thank Terje Skjerpen and Knut Wangen for valuable comments.

1 Introduction

Systems of regression equations and methods for estimating such systems have a rather long history in econometrics. Notable examples, with cross equational coefficient constraints, are linear systems of factor demand equations derived from producers' cost minimizing or profit maximizing behaviour and systems of demand functions for consumption commodities derived from consumers' cost minimizing or utility maximizing behaviour. The reduced form of a linear structural model is also, formally, a system of regression equations, since the left hand variable is the only endogenous variable in each equation.

In this paper, we consider the specification and estimation of regression systems with random individual effects from unbalanced panel data, *i.e.*, panel data where the individual time series have unequal length. This approach is more general than the standard approach to the estimation of regression equations from panel data. A large number of papers and textbook chapters discuss single equation models with balanced panel data and random effects [see, *e.g.*, Greene (1997, chapter 14)]. Single equation models with unbalanced panel data and random individual effects are discussed in Biørn (1981) and Baltagi (1985). Systems of regression equations for balanced panel data with random individual and period specific effects are discussed in Avery (1977) and Baltagi (1980).[1] The model and methods to be considered in the present paper is a generalization of the models and methods in the papers above, except that time specific random effects are ignored.[2]

The model and notation is presented in Section 2, and we take a first look at the estimation of the covariance matrices of the error terms. Two versions of (feasible) Generalized Least Squares (GLS) estimation – considered as a preliminary to Maximum Likelihood (ML) estimation – are discussed in Section 3. In Section 4, we first describe the ML problem, in two variants, and next derive a stepwise switching algorithm for implementing their solution.[3] The relationship to the GLS problem is discussed.

[1] Models with randomly varying coefficients in addition to randomly varying intercepts will not be considered in this paper. Biørn (1999) elaborates Maximum Likelihood estimation for such, more general models for unbalanced panel data.

[2] The joint occurrence of unbalanced panel data and random two-way effects raises special problems and will not be considered here. For most practical problems, random individual heterogeneity is far more important than random time specific heterogeneity – at least for genuine micro data for individuals, households, or firms. Quadratic unbiased and Maximum Likelihood estimation of a single equation combining unbalanced panel data and random two-way effects is considered in Wansbeek and Kapteyn (1989). Fixed period specific effects can be included without notable problems.

[3] This algorithm may be programmed in matrix oriented software codes, *e.g.*, Gauss.

2 Model and notation

Consider a system of G regression equations, indexed by $g = 1, \ldots, G$, with observations from an unbalanced panel with N individuals, indexed by $i = 1, \ldots, N$. The individuals are observed in at least one and at most P periods. Let N_p denote the number of individuals observed in p periods (not necessary the same and not necessarily consecutive), $p = 1, \ldots, P$, and let n be the total number of observations, *i.e.*, $N = \sum_{p=1}^{P} N_p$ and $n = \sum_{p=1}^{P} N_p p$.

Assume that the individuals are ordered in P groups such that the N_1 individuals observed once come first, the N_2 individuals observed twice come second, ..., the N_P individuals observed P times come last. Let M_p be the cumulated number of individuals observed up to p times, *i.e.*,

$$M_p = N_1 + N_2 + \cdots + N_p, \quad p = 1, \ldots, P.$$

In particular, $M_1 = N_1$ and $M_P = N$. Let I_p denote the index set of the individuals observed p times, *i.e.*,

$$\text{(1)} \qquad \begin{cases} I_1 = [1, \ldots, M_1], \\ I_2 = [M_1+1, \ldots, M_2], \\ \vdots \\ I_P = [M_{P-1}+1, \ldots, M_P]. \end{cases}$$

We may formally consider I_1 as a cross section and $I_2, I_3, \ldots, I_P$ as balanced subpanels with $2, 3, \ldots, P$ observations of each individual, respectively.

The g'th equation for individual i, observation t – specifying H_g regressors (including a one to represent the intercept term) and unobserved, additive, random individual heterogeneity – is

$$\text{(2)} \qquad y_{git} = \boldsymbol{x}_{git}\boldsymbol{\beta}_g + \alpha_{gi} + u_{git}, \quad g = 1, \ldots, G;\ i \in I_p;\ t = 1, \ldots, p;\ p = 1, \ldots, P,$$

where y_{git} is the left hand side variable, $\boldsymbol{x}_{git}$, of dimension $(1 \times H_g)$, is the regressor vector, and u_{git} is the genuine disturbance in the g'th equation specific to individual i, observation t. Finally, $\boldsymbol{\beta}_g$ is the $(H_g \times 1)$ coefficient vector in the g'th equation (including the intercept term), common to all individuals, and α_{gi} is a latent effect specific to individual i in the g'th equation.

Stacking the G equations for each observation (i, t), we have

$$\text{(3)} \qquad \begin{bmatrix} y_{1it} \\ \vdots \\ y_{Git} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_{1it} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{x}_{Git} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_G \end{bmatrix} + \begin{bmatrix} \alpha_{1i} \\ \vdots \\ \alpha_{Gi} \end{bmatrix} + \begin{bmatrix} u_{1it} \\ \vdots \\ u_{Git} \end{bmatrix},$$

or in compact notation,[4]

$$\boldsymbol{y}_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{\alpha}_i + \boldsymbol{u}_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{it}, \tag{4}$$

where

$$\boldsymbol{\epsilon}_{it} = \begin{bmatrix} \epsilon_{1it} \\ \vdots \\ \epsilon_{Git} \end{bmatrix} = \boldsymbol{\alpha}_i + \boldsymbol{u}_{it}. \tag{5}$$

The dimension of the stacked (block-diagonal) regressor matrix $\boldsymbol{X}_{it}$ and the coefficient vector $\boldsymbol{\beta}$ are $(G \times H)$ and $(H \times 1)$, respectively, where $H = \sum_{g=1}^{G} H_g$. We assume[5]

$$\mathsf{E}(\boldsymbol{\alpha}_i) = \mathbf{0}_{G,1}, \qquad \mathsf{E}(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j') = \delta_{ij}\boldsymbol{\Sigma}_\alpha, \tag{6}$$

$$\mathsf{E}(\boldsymbol{u}_{it}) = \mathbf{0}_{G,1}, \qquad \mathsf{E}(\boldsymbol{u}_{it}\boldsymbol{u}_{js}') = \delta_{ij}\delta_{ts}\boldsymbol{\Sigma}_u, \tag{7}$$

$$\boldsymbol{X}_{it}, \boldsymbol{\alpha}_i, \boldsymbol{u}_{it} \quad \text{are uncorrelated.} \tag{8}$$

where the δ's are Kronecker deltas and

$$\boldsymbol{\Sigma}_\alpha = \begin{bmatrix} \sigma_{11}^\alpha & \cdots & \sigma_{1G}^\alpha \\ \vdots & & \vdots \\ \sigma_{G1}^\alpha & \cdots & \sigma_{GG}^\alpha \end{bmatrix}, \qquad \boldsymbol{\Sigma}_u = \begin{bmatrix} \sigma_{11}^u & \cdots & \sigma_{1G}^u \\ \vdots & & \vdots \\ \sigma_{G1}^u & \cdots & \sigma_{GG}^u \end{bmatrix}.$$

It follows from (6) – (8) that the composite disturbance vectors (5) satisfy

$$\mathsf{E}(\boldsymbol{\epsilon}_{it}) = \mathbf{0}_{G,1}, \qquad \mathsf{E}(\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{js}) = \delta_{ij}(\boldsymbol{\Sigma}_\alpha + \delta_{ts}\boldsymbol{\Sigma}_u). \tag{9}$$

The individual specific mean of the $\boldsymbol{\epsilon}$'s for individual i is given by

$$\bar{\boldsymbol{\epsilon}}_{i\cdot} = \begin{cases} \boldsymbol{\epsilon}_{i1} & \text{for } i \in I_1, \\ (1/2)\sum_{t=1}^{2} \boldsymbol{\epsilon}_{it} & \text{for } i \in I_2, \\ \vdots & \\ (1/P)\sum_{t=1}^{P} \boldsymbol{\epsilon}_{it} & \text{for } i \in I_P, \end{cases} \tag{10}$$

and its global mean is

$$\bar{\boldsymbol{\epsilon}} = \frac{1}{n}\sum_{p=1}^{P}\sum_{i\in I_p}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it} = \frac{1}{n}\sum_{p=1}^{P}\sum_{i\in I_p} p\,\bar{\boldsymbol{\epsilon}}_{i\cdot}\ . \tag{11}$$

---

[4]This way of writing the model can be given an interpretation which is more general than (3), since some kinds of cross-equational constraints may be included. If the coefficient vectors are not disjoint across equations because at least one coefficient occurs in at least two equations, we can (i) redefine $\boldsymbol{\beta}$ as the complete coefficient vector (without duplication) and (ii) redefine the regressor matrix as $\boldsymbol{X}_{it} = [\boldsymbol{x}_{1it}', \ldots, \boldsymbol{x}_{Git}']'$, where the k'th element of x_{git} is redefined to contain the observations on the variable in the g'th equation which corresponds to the k'th coefficient in $\boldsymbol{\beta}$. If the latter coefficient does not occur in the g'th equation, the k'th element of $\boldsymbol{x}_{git}$ is set to zero.

[5]Without loss of generality, we can set $\mathsf{E}(\boldsymbol{\alpha}_i)$ to zero, since a non-zero value can be included in the intercept term.

The $(G \times G)$ matrices of total, within individual, and between individual (co)variation in the $\boldsymbol{\epsilon}$'s of the different equations can be expressed as

$$(12) \qquad \boldsymbol{T}_{\epsilon\epsilon} = \sum_{p=1}^{P} \sum_{i \in I_p} \sum_{t=1}^{p} (\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}})',$$

$$(13) \qquad \boldsymbol{W}_{\epsilon\epsilon} = \sum_{p=1}^{P} \sum_{i \in I_p} \sum_{t=1}^{p} (\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{i\cdot})(\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{i\cdot})',$$

$$(14) \qquad \boldsymbol{B}_{\epsilon\epsilon} = \sum_{p=1}^{P} \sum_{i \in I_p} p\,(\bar{\boldsymbol{\epsilon}}_{i\cdot} - \bar{\boldsymbol{\epsilon}})(\bar{\boldsymbol{\epsilon}}_{i\cdot} - \bar{\boldsymbol{\epsilon}})',$$

respectively.[6] These definitions imply that the total variation can be decomposed into variation within and between individuals as

$$(15) \qquad \boldsymbol{T}_{\epsilon\epsilon} = \boldsymbol{W}_{\epsilon\epsilon} + \boldsymbol{B}_{\epsilon\epsilon},$$

in the same way as for a balanced data set. The total, within, and between matrices of the $\boldsymbol{y}$'s and $\boldsymbol{x}$'s are defined and decomposed similarly.

We show in Appendix A that

$$(16) \qquad \mathsf{E}(\boldsymbol{W}_{\epsilon\epsilon}) = (n - N)\boldsymbol{\Sigma}_u,$$

$$(17) \qquad \mathsf{E}(\boldsymbol{B}_{\epsilon\epsilon}) = (N - 1)\boldsymbol{\Sigma}_u + \left(n - \frac{\sum_{p=1}^{P} N_p p^2}{n}\right) \boldsymbol{\Sigma}_\alpha.$$

Hence,

$$(18) \qquad \widehat{\boldsymbol{\Sigma}}_u = \frac{\boldsymbol{W}_{\epsilon\epsilon}}{n - N},$$

$$(19) \qquad \widehat{\boldsymbol{\Sigma}}_\alpha = \frac{\boldsymbol{B}_{\epsilon\epsilon} - \frac{N-1}{n-N}\boldsymbol{W}_{\epsilon\epsilon}}{n - \frac{\sum_{p=1}^{P} N_p p^2}{n}}$$

would be unbiased estimators of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ if the disturbances $\boldsymbol{\epsilon}_{it}$ were known.[7]

The total, within individual, and between individual (co)variation in the $\boldsymbol{\epsilon}$'s of the individuals observed p times are expressed as[8]

$$(20) \qquad \boldsymbol{T}_{\epsilon\epsilon(p)} = \sum_{i \in I_p} \sum_{t=1}^{p} (\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{(p)})(\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{(p)})',$$

[6] The double sum $\sum_{p=1}^{P} \sum_{i \in I_p}$ in (11) – (14) corresponds to the summation across individuals (*e.g.*, $\sum_{i=1}^{N}$) for a standard balanced design, in which only one I_p is non-empty.

[7] Consistent residuals can replace the ϵ's in (18) and (19) to obtain consistent estimates of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ in practice. See Sections 3 and 4.

[8] Note, by comparing (12) – (14) with (20) – (22), that $\boldsymbol{W}_{\epsilon\epsilon} = \sum_{p=1}^{P} \boldsymbol{W}_{\epsilon\epsilon(p)}$, but similar relationships do not exist between $\boldsymbol{B}_{\epsilon\epsilon}$ and the $\boldsymbol{B}_{\epsilon\epsilon(p)}$'s and between $\boldsymbol{T}_{\epsilon\epsilon}$ and the $\boldsymbol{T}_{\epsilon\epsilon(p)}$'s, because the unweighted group means occur in the latter and the global means (weighted group means) occur in the former.

(21) $$\boldsymbol{W}_{\epsilon\epsilon(p)} = \sum_{i\in I_p}\sum_{t=1}^{p}(\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{i\cdot})(\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{i\cdot})',$$

(22) $$\boldsymbol{B}_{\epsilon\epsilon(p)} = p\sum_{i\in I_p}(\bar{\boldsymbol{\epsilon}}_{i\cdot} - \bar{\boldsymbol{\epsilon}}_{(p)})(\bar{\boldsymbol{\epsilon}}_{i\cdot} - \bar{\boldsymbol{\epsilon}}_{(p)})',$$

respectively, where

(23) $$\bar{\boldsymbol{\epsilon}}_{(p)} = \frac{1}{N_p p}\sum_{i\in I_p}\sum_{t=1}^{p}\boldsymbol{\epsilon}_{it} = \frac{1}{N_p}\sum_{i\in I_p} p\,\bar{\boldsymbol{\epsilon}}_{i\cdot},$$

and we find that the total variation of the individuals observed p times can be decomposed into within and between individual variation as

(24) $$\boldsymbol{T}_{\epsilon\epsilon(p)} = \boldsymbol{W}_{\epsilon\epsilon(p)} + \boldsymbol{B}_{\epsilon\epsilon(p)}.$$

The total, within, and between matrices of the $\boldsymbol{y}$'s and $\boldsymbol{x}$'s are defined similarly.

We find[9]

(25) $$\mathsf{E}(\boldsymbol{W}_{\epsilon\epsilon(p)}) = N_p(p-1)\boldsymbol{\Sigma}_u,$$

(26) $$\mathsf{E}(\boldsymbol{B}_{\epsilon\epsilon(p)}) = (N_p - 1)\boldsymbol{\Sigma}_u + p(N_p - 1)\boldsymbol{\Sigma}_\alpha,$$

and hence

(27) $$\widehat{\boldsymbol{\Sigma}}_{u(p)} = \frac{\boldsymbol{W}_{\epsilon\epsilon(p)}}{N_p(p-1)},$$

(28) $$\widehat{\boldsymbol{\Sigma}}_{\alpha(p)} = \frac{1}{p}\left[\frac{\boldsymbol{B}_{\epsilon\epsilon(p)}}{N_p - 1} - \frac{\boldsymbol{W}_{\epsilon\epsilon(p)}}{N_p(p-1)}\right],$$

(29) $$\widehat{\boldsymbol{\Sigma}}_{(p)} = \widehat{\boldsymbol{\Sigma}}_{u(p)} + p\widehat{\boldsymbol{\Sigma}}_{\alpha(p)} = \frac{\boldsymbol{B}_{\epsilon\epsilon(p)}}{N_p - 1}$$

would be unbiased estimators of $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_\alpha$, and $\boldsymbol{\Sigma}_{(p)}$ based on the disturbances $\boldsymbol{\epsilon}_{it}$ from the individuals observed p times if these disturbances were known.

Let now $\boldsymbol{y}_{i(p)}$, $\boldsymbol{X}_{i(p)}$, and $\boldsymbol{\epsilon}_{i(p)}$ be the stacked $(Gp \times 1)$ vector, $(Gp \times H)$ matrix, and $(Gp \times 1)$ vector of $\boldsymbol{y}$'s, $\boldsymbol{X}$'s, and $\boldsymbol{\epsilon}$'s, respectively, corresponding to the p observations of individual $i \in I_p$, *i.e.*,

(30) $$\boldsymbol{y}_{i(p)} = \begin{bmatrix} \boldsymbol{y}_{i1} \\ \vdots \\ \boldsymbol{y}_{ip} \end{bmatrix},\quad \boldsymbol{X}_{i(p)} = \begin{bmatrix} \boldsymbol{X}_{i1} \\ \vdots \\ \boldsymbol{X}_{ip} \end{bmatrix},\quad \boldsymbol{\epsilon}_{i(p)} = \begin{bmatrix} \boldsymbol{\epsilon}_{i1} \\ \vdots \\ \boldsymbol{\epsilon}_{ip} \end{bmatrix}\quad \text{for } i \in I_p,\quad p = 1, \ldots, P.$$

Then (4) can be rewritten as

(31) $$\boldsymbol{y}_{i(p)} = \boldsymbol{X}_{i(p)}\boldsymbol{\beta} + (\boldsymbol{e}_p \otimes \boldsymbol{\alpha}_i) + \boldsymbol{u}_{i(p)} = \boldsymbol{X}_{i(p)}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{i(p)},$$

[9] These expressions can be obtained by, *e.g.*, replacing N, n, and $\sum_{p=1}^{P} N_p p^2$ in (16) and (17) by respectively N_p, $N_p p$, and $N_p p^2$.

where $\boldsymbol{e}_p$ is the $(p \times 1)$ vector of ones. From (9) it follows that

$$\mathsf{E}(\boldsymbol{\epsilon}_{i(p)}) = \mathbf{0}_{Gp,1}, \qquad \mathsf{E}(\boldsymbol{\epsilon}_{i(p)}\boldsymbol{\epsilon}'_{i(p)}) = \boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_u + \boldsymbol{E}_p \otimes \boldsymbol{\Sigma}_\alpha = \boldsymbol{\Omega}_{\epsilon(p)}, \tag{32}$$

where $\boldsymbol{I}_p$ is the p dimensional identity matrix, $\boldsymbol{E}_p = \boldsymbol{e}_p\boldsymbol{e}'_p$ is the $(p \times p)$ matrix with all elements equal to one, and $\boldsymbol{\Omega}_{\epsilon(p)}$ is the $(Gp \times Gp)$ matrix defined by the last equality. The expression for this disturbance covariance matrix can be rewritten as

$$\boldsymbol{\Omega}_{\epsilon(p)} = \boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u + \boldsymbol{A}_p \otimes (\boldsymbol{\Sigma}_u + p\,\boldsymbol{\Sigma}_\alpha), \tag{33}$$

where $\boldsymbol{A}_p = (1/p)\boldsymbol{E}_p$ and $\boldsymbol{B}_p = \boldsymbol{I}_p - (1/p)\boldsymbol{E}_p$. The latter two matrices are symmetric and idempotent and have orthogonal columns.

In Sections 3 and 4, we discuss stepwise procedures for Generalized Least Squares and Maximum Likelihood estimation of the coefficient vector $\boldsymbol{\beta}$.

3 Generalized Least Squares estimation

The Generalized Least Squares (GLS) problem for estimating the joint coefficient vector $\boldsymbol{\beta}$ when the covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ are known is to minimize the quadratic form

$$Q = \sum_{p=1}^{P}\sum_{i\in I_p} \boldsymbol{\epsilon}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\boldsymbol{\epsilon}_{i(p)} = \sum_{p=1}^{P}\sum_{i\in I_p} (\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}_{\epsilon(p)}(\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}) \tag{34}$$

with respect to $\boldsymbol{\beta}$ for given $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$. Since it follows from the properties of $\boldsymbol{A}_p$ and $\boldsymbol{B}_p$ that

$$\boldsymbol{\Omega}^{-1}_{\epsilon(p)} = \boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1} + \boldsymbol{A}_p \otimes (\boldsymbol{\Sigma}_u + p\,\boldsymbol{\Sigma}_\alpha)^{-1}, \tag{35}$$

we can rewrite Q as

$$Q = \sum_{p=1}^{P}\sum_{i\in I_p} \boldsymbol{\epsilon}'_{i(p)}[\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}]\boldsymbol{\epsilon}_{i(p)} + \sum_{p=1}^{P}\sum_{i\in I_p} \boldsymbol{\epsilon}'_{i(p)}[\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}]\boldsymbol{\epsilon}_{i(p)}, \tag{36}$$

where

$$\boldsymbol{\Sigma}_{(p)} = \boldsymbol{\Sigma}_u + p\,\boldsymbol{\Sigma}_\alpha, \qquad p = 1,\ldots,P. \tag{37}$$

Let us consider the GLS problem (i) when we utilize the full data set, and (ii) when we utilize the data from the individuals observed p times only.

GLS estimation for all observations

The solution to the problem of minimizing Q obtained from $\partial Q/\partial\boldsymbol{\beta} = \mathbf{0}$, *i.e.*, the GLS estimator of $\boldsymbol{\beta}$ for known $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$, is

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left[\sum_{p=1}^{P}\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\boldsymbol{X}_{i(p)}\right]^{-1} \left[\sum_{p=1}^{P}\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\boldsymbol{y}_{i(p)}\right] \tag{38}$$

$$= \left[\sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} [\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}] \boldsymbol{X}_{i(p)} + \sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} [\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}] \boldsymbol{X}_{i(p)} \right]^{-1}$$
$$\times \left[\sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} [\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}] \boldsymbol{y}_{i(p)} + \sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} [\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}] \boldsymbol{y}_{i(p)} \right].$$

Its covariance matrix, which can be obtained from the standard GLS formula, is

$$\text{(39)} \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) = \left[\sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} \boldsymbol{\Omega}_{\epsilon(p)}^{-1} \boldsymbol{X}_{i(p)} \right]^{-1}$$
$$= \left[\sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} [\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}] \boldsymbol{X}_{i(p)} + \sum_{p=1}^{P} \sum_{i \in I_p} \boldsymbol{X}'_{i(p)} [\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}] \boldsymbol{X}_{i(p)} \right]^{-1}.$$

Usually, $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ are unknown. Then the following *stepwise GLS procedure* can be used for *estimating* $\boldsymbol{\beta}$*,* $\boldsymbol{\Sigma}_u$*, and* $\boldsymbol{\Sigma}_\alpha$ *jointly*:

Step 1: Run OLS separately on all G equations in (2), using all observations on $\boldsymbol{y}_{it}$ and $\boldsymbol{X}_{it}$. Stacking the estimators, we get the joint estimator vector $\widehat{\boldsymbol{\beta}}_{OLS}$. Form the corresponding vectors of residuals $\widehat{\boldsymbol{\epsilon}}_{it} = \boldsymbol{y}_{it} - \boldsymbol{X}_{it}\widehat{\boldsymbol{\beta}}_{OLS}$ for all i and t. These residuals are consistent.

Step 2: Compute within and between matrices of residuals by inserting $\boldsymbol{\epsilon}_{it} = \widehat{\boldsymbol{\epsilon}}_{it}$ in (13) and (14). Let the estimators obtained be denoted as $\boldsymbol{W}_{\hat{\epsilon}\hat{\epsilon}}$ and $\boldsymbol{B}_{\hat{\epsilon}\hat{\epsilon}}$.

Step 3: Estimate $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_\alpha$, and $\boldsymbol{\Sigma}_{(p)}$ by inserting $\boldsymbol{W}_{\epsilon\epsilon} = \boldsymbol{W}_{\hat{\epsilon}\hat{\epsilon}}$ and $\boldsymbol{B}_{\epsilon\epsilon} = \boldsymbol{B}_{\hat{\epsilon}\hat{\epsilon}}$ in (18), (19), and (37) for $p = 1, \ldots, P$. Let the estimators obtained be denoted as $\widehat{\boldsymbol{\Sigma}}_u$, $\widehat{\boldsymbol{\Sigma}}_\alpha$, and $\widehat{\boldsymbol{\Sigma}}_{(p)}$.

Step 4: Compute the (feasible) GLS estimator of $\boldsymbol{\beta}$ by inserting $\boldsymbol{\Sigma}_u = \widehat{\boldsymbol{\Sigma}}_u$ and $\boldsymbol{\Sigma}_{(p)} = \widehat{\boldsymbol{\Sigma}}_{(p)}$ for $p = 1, \ldots, P$ in (38).

This algorithm can be iterated by recomputing the residuals from the GLS estimators and repeating steps 2 – 4 until convergence, according to some criterion.

GLS estimation for the individuals observed p times

We may alternatively – or as a preliminary to full GLS estimation – apply GLS on the observations for the individuals observed p times, in the following denoted as *group* p, and do this *separately for* $p = 1, \ldots, P$. We then minimize the part of the quadratic form Q which relates to the individuals observed p times, *i.e.*,

$$\text{(40)} \qquad Q_{(p)} = \sum_{i \in I_p} \boldsymbol{\epsilon}'_{i(p)} [\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}] \boldsymbol{\epsilon}_{i(p)} + \sum_{i \in I_p} \boldsymbol{\epsilon}'_{i(p)} [\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}] \boldsymbol{\epsilon}_{i(p)}.$$

The solution to the (conditional) problem of minimizing $Q_{(p)}$ obtained from $\partial Q_{(p)}/\partial\boldsymbol{\beta} = \mathbf{0}$, *i.e.*, the GLS estimator of $\boldsymbol{\beta}$ for group p for known $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$, is

$$
\begin{aligned}
(41) \qquad \widehat{\boldsymbol{\beta}}_{GLS(p)} &= \left[\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\boldsymbol{X}_{i(p)}\right]^{-1}\left[\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\boldsymbol{y}_{i(p)}\right] \\
&= \left[\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}[\boldsymbol{B}_p\otimes\boldsymbol{\Sigma}_u^{-1}]\boldsymbol{X}_{i(p)} + \sum_{i\in I_p} \boldsymbol{X}'_{i(p)}[\boldsymbol{A}_p\otimes\boldsymbol{\Sigma}_{(p)}^{-1}]\boldsymbol{X}_{i(p)}\right]^{-1} \\
&\quad\times\left[\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}[\boldsymbol{B}_p\otimes\boldsymbol{\Sigma}_u^{-1}]\boldsymbol{y}_{i(p)} + \sum_{i\in I_p} \boldsymbol{X}'_{i(p)}[\boldsymbol{A}_p\otimes\boldsymbol{\Sigma}_{(p)}^{-1}]\boldsymbol{y}_{i(p)}\right].
\end{aligned}
$$

Its covariance matrix is

$$
\begin{aligned}
(42) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS(p)}) &= \left[\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\boldsymbol{X}_{i(p)}\right]^{-1} \\
&= \left[\sum_{i\in I_p} \boldsymbol{X}'_{i(p)}[\boldsymbol{B}_p\otimes\boldsymbol{\Sigma}_u^{-1}]\boldsymbol{X}_{i(p)} + \sum_{i\in I_p} \boldsymbol{X}'_{i(p)}[\boldsymbol{A}_p\otimes\boldsymbol{\Sigma}_{(p)}^{-1}]\boldsymbol{X}_{i(p)}\right]^{-1}.
\end{aligned}
$$

Note that since $A_1 = 1$ and $B_1 = 0$, we have $\boldsymbol{\Omega}^{-1}_{\epsilon(1)} = \boldsymbol{\Sigma}^{-1}_{(1)} = (\boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_\alpha)^{-1}$ and $Q_{(1)} = \sum_{i\in I_1}\boldsymbol{\epsilon}'_{i(1)}(\boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_\alpha)^{-1}\boldsymbol{\epsilon}_{i(1)}$. The expressions for $\widehat{\boldsymbol{\beta}}_{GLS(1)}$ and $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS(1)})$ are simplified accordingly.

If $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ are unknown, we can proceed as follows:

Step 1(p): Run OLS separately on all G equations in (2) using the observations for $i \in I_p$ and all t. Stacking the estimators, we get the joint estimator vector $\widehat{\boldsymbol{\beta}}_{OLS(p)}$. Form the corresponding vectors of residuals $\widehat{\boldsymbol{\epsilon}}_{it} = \boldsymbol{y}_{it} - \boldsymbol{X}_{it}\widehat{\boldsymbol{\beta}}_{OLS(p)}$ for $i \in I_p$ and all t. These residuals are consistent.

Step 2(p): Compute within and between matrices of residuals by for group p by inserting $\boldsymbol{\epsilon}_{it} = \widehat{\boldsymbol{\epsilon}}_{it}$ in (21) and (22) for $i \in I_p$. Let the values be denoted as $\boldsymbol{W}_{\hat{\epsilon}\hat{\epsilon}(p)}$ and $\boldsymbol{B}_{\hat{\epsilon}\hat{\epsilon}(p)}$.

Step 3(p): Estimate $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_\alpha$, and $\boldsymbol{\Sigma}_{(p)}$, by inserting $\boldsymbol{W}_{\epsilon\epsilon(p)} = \boldsymbol{W}_{\hat{\epsilon}\hat{\epsilon}(p)}$ and $\boldsymbol{B}_{\epsilon\epsilon(p)} = \boldsymbol{B}_{\hat{\epsilon}\hat{\epsilon}(p)}$ in (27), (28), and (29) if $p = 2,\ldots,P$. Let the estimators obtained be denoted as $\widehat{\boldsymbol{\Sigma}}_{u(p)}$, $\widehat{\boldsymbol{\Sigma}}_{\alpha(p)}$, and $\widehat{\boldsymbol{\Sigma}}_{(p)}$. If $p = 1$, $\boldsymbol{\Sigma}_{(1)} = \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_\alpha$ is estimable, but not $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$.

Step 4(p): Compute the (feasible) GLS estimator of $\boldsymbol{\beta}$ for group p by inserting $\boldsymbol{\Sigma}_u = \widehat{\boldsymbol{\Sigma}}_{u(p)}$ and $\boldsymbol{\Sigma}_{(p)} = \widehat{\boldsymbol{\Sigma}}_{(p)}$ for $p = 2,\ldots,P$ and $\widehat{\boldsymbol{\Sigma}}_{(1)}$ for $p = 1$ in (41).

This algorithm can be iterated by recomputing the residuals from the GLS estimators and repeating steps 2(p) – 4(p) until convergence, according to some criterion.

It follows from (38), (39), (41), and (42) that the overall estimator $\widehat{\boldsymbol{\beta}}_{GLS}$ can be interpreted as a compromise, a matrix weighted average of the group specific estimators $\widehat{\boldsymbol{\beta}}_{GLS(1)}, \dots, \widehat{\boldsymbol{\beta}}_{GLS(P)}$.[10] Their weights are the inverse of their respective covariance matrices:

$$
\widehat{\boldsymbol{\beta}}_{GLS} = \left[\sum_{p=1}^{P} \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS(p)})^{-1}\right]^{-1} \left[\sum_{p=1}^{P} \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS(p)})^{-1}\widehat{\boldsymbol{\beta}}_{GLS(p)}\right]. \tag{43}
$$

4 Maximum Likelihood estimation

In this section, we consider the Maximum Likelihood method for joint estimation of the coefficient vectors and the disturbance covariance matrices. We make the assumption of *normality* of the individual effects and the disturbances and replace (6) and (7) by

$$
\boldsymbol{\alpha}_i \sim \mathsf{IIN}\,(\mathbf{0}_{G,1}, \boldsymbol{\Sigma}_\alpha), \qquad \boldsymbol{u}_{it} \sim \mathsf{IIN}\,(\mathbf{0}_{G,1}, \boldsymbol{\Sigma}_u). \tag{44}
$$

Then, the $\boldsymbol{\epsilon}_{i(p)}|\boldsymbol{X}_{i(p)}$'s are stochastically independent across $i(p)$ and distributed as $\mathsf{N}(\mathbf{0}_{Gp,1}, \boldsymbol{\Omega}_{\epsilon(p)})$, where $\boldsymbol{\Omega}_{\epsilon(p)}$ is defined in (33). This implies that the log-density function of $\boldsymbol{y}_{i(p)}|\boldsymbol{X}_{i(p)}$, *i.e.*, for individual i, $i \in I_p$, is

$$
L_{i(p)} = -\frac{Gp}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Omega}_{\epsilon(p)}| - \frac{1}{2}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]'\boldsymbol{\Omega}_{\epsilon(p)}^{-1}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}], \tag{45}
$$

where $|\boldsymbol{\Omega}_{\epsilon(p)}| = |\boldsymbol{\Sigma}_{(p)}||\boldsymbol{\Sigma}_u|^{p-1}$ [cf. (B.2)] and $\boldsymbol{\Omega}_{\epsilon(p)}^{-1}$ is given by (35). The log-likelihood function of all $\boldsymbol{y}$'s conditional on all $\boldsymbol{X}$'s for the individuals which are observed p times then becomes

$$
\begin{aligned}
L_{(p)} = \sum_{i=1}^{N_p} L_{i(p)} &= -\frac{GN_p p}{2}\ln(2\pi) - \frac{N_p}{2}\ln|\boldsymbol{\Omega}_{\epsilon(p)}| \\
&\quad -\frac{1}{2}\sum_{i\in I_p}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]'\boldsymbol{\Omega}_{\epsilon(p)}^{-1}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}],
\end{aligned} \tag{46}
$$

and the log-likelihood function of all $\boldsymbol{y}$'s conditional on all $\boldsymbol{X}$'s in the full data set is

$$
\begin{aligned}
L = \sum_{p=1}^{P} L_{(p)} &= -\frac{Gn}{2}\ln(2\pi) - \frac{1}{2}\sum_{p=1}^{P} N_p \ln|\boldsymbol{\Omega}_{\epsilon(p)}| \\
&\quad -\frac{1}{2}\sum_{p=1}^{P}\sum_{i\in I_p}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]'\boldsymbol{\Omega}_{\epsilon(p)}^{-1}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}].
\end{aligned} \tag{47}
$$

Two related Maximum Likelihood (ML) estimation problems will be considered. The first is the ML problem based on the balanced subsample of the individuals *in group*

[10] These P group specific estimators are independent.

p *only.* The ML estimators of $(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_u,)$ for these individuals are the values that maximize $L_{(p)}$. The second is the ML problem based on the *complete data set.* The ML estimators of $(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_u)$ in the latter case are obtained by maximizing L, and, like the corresponding GLS estimators, they may be interpreted as *compromise values* of the (conditional) group specific estimators for $p = 1, 2, \ldots, P$. The structure of the first problem (when $p > 1$) is similar to the ML problem for a system of regression equations for balanced panel data since group p formally represents a balanced data set. The structure of the second, full ML, problem is more complicated since the individuals included are observed a varying number of times, and hence different 'gross' disturbance covariance matrices, $\boldsymbol{\Omega}_{\epsilon(p)}$, all of which are functions of the same 'basic' matrices, $\boldsymbol{\Sigma}_\alpha$ and $\boldsymbol{\Sigma}_u$, are involved.

We now elaborate the solution to both problems.

The ML problem for group p

We write the log-density function $L_{(p)}$ as

$$L_{(p)} = -\frac{GN_p p}{2}\ln(2\pi) - \frac{N_p}{2}\ln(|\boldsymbol{\Omega}_{\epsilon(p)}|) - \frac{1}{2}Q_{(p)}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_\alpha), \tag{48}$$

where

$$Q_{(p)} = Q_{(p)}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_\alpha) = \sum_{i \in I_p} [\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]' \boldsymbol{\Omega}_{\epsilon(p)}^{-1} [\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]. \tag{49}$$

We split the problem of maximizing L_p into *two conditional subproblems*:

Subproblem A: Maximization of $L_{(p)}$ *with respect to* $\boldsymbol{\beta}$ *for given* $\boldsymbol{\Sigma}_u$ *and* $\boldsymbol{\Sigma}_\alpha$.
Subproblem B: Maximization of $L_{(p)}$ *with respect to* $\boldsymbol{\Sigma}_u$ *and* $\boldsymbol{\Sigma}_\alpha$ *for given* $\boldsymbol{\beta}$.

The joint solution to A and B defines the solution to the problem of maximizing L_p.

For solving subproblem B, it is convenient to arrange the disturbances from individual i, $i \in I_p$, in all G equations and p observations in the $(G \times p)$ matrix

$$\widetilde{\boldsymbol{E}}_{i(p)} = [\boldsymbol{\epsilon}_{i1}, \ldots, \boldsymbol{\epsilon}_{ip}] = \begin{bmatrix} \epsilon_{1i1} & \cdots & \epsilon_{1ip} \\ \vdots & & \vdots \\ \epsilon_{Gi1} & \cdots & \epsilon_{Gip} \end{bmatrix}, \quad i \in I_p, \tag{50}$$

so that the $(Gp \times 1)$-vector $\boldsymbol{\epsilon}_{i(p)}$, $i \in I_p$, defined in (30), can be written as

$$\boldsymbol{\epsilon}_{i(p)} = \text{vec}(\widetilde{\boldsymbol{E}}_{i(p)}), \qquad i \in I_p, \tag{51}$$

where 'vec' is the vectorization operator. Then we have

$$\widetilde{\boldsymbol{E}}_{i(p)}\widetilde{\boldsymbol{E}}'_{i(p)} = \begin{bmatrix} \sum_t \epsilon_{1it}^2 & \cdots & \sum_t \epsilon_{1it}\epsilon_{Git} \\ \vdots & & \vdots \\ \sum_t \epsilon_{Git}\epsilon_{1it} & \cdots & \sum_t \epsilon_{Git}^2 \end{bmatrix},$$

$$\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{B}_p\widetilde{\boldsymbol{E}}'_{i(p)} = \begin{bmatrix} \sum_t(\epsilon_{1it}-\bar{\epsilon}_{1i\cdot})^2 & \cdots & \sum_t(\epsilon_{1it}-\bar{\epsilon}_{1i\cdot})(\epsilon_{Git}-\bar{\epsilon}_{Gi\cdot}) \\ \vdots & & \vdots \\ \sum_t(\epsilon_{Git}-\bar{\epsilon}_{Gi\cdot})(\epsilon_{1it}-\bar{\epsilon}_{1i\cdot}) & \cdots & \sum_t(\epsilon_{Git}-\bar{\epsilon}_{Gi\cdot})^2 \end{bmatrix},$$

$$\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{A}_p\widetilde{\boldsymbol{E}}'_{i(p)} = \begin{bmatrix} p\bar{\epsilon}_{1i\cdot}^2 & \cdots & p\bar{\epsilon}_{1i\cdot}\bar{\epsilon}_{Gi\cdot} \\ \vdots & & \vdots \\ p\bar{\epsilon}_{Gi\cdot}\bar{\epsilon}_{1i\cdot} & \cdots & p\bar{\epsilon}_{Gi\cdot}^2 \end{bmatrix},$$

$$i \in I_p,$$

where $\bar{\epsilon}_{gi\cdot} = (1/p)\sum_{t=1}^{p}\epsilon_{git}$, $i \in I_p$. Adding each of these three expressions across i, $i \in I_p$, we obtain, respectively,

$$\widetilde{\boldsymbol{T}}_{\epsilon\epsilon(p)} = \sum_{i\in I_p}\widetilde{\boldsymbol{E}}_{i(p)}\widetilde{\boldsymbol{E}}'_{i(p)}, \tag{52}$$

$$\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)} = \sum_{i\in I_p}\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{B}_p\widetilde{\boldsymbol{E}}'_{i(p)}, \tag{53}$$

$$\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)} = \sum_{i\in I_p}\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{A}_p\widetilde{\boldsymbol{E}}'_{i(p)}. \tag{54}$$

We see that $\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)} + \widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)} = \widetilde{\boldsymbol{T}}_{\epsilon\epsilon(p)}$ since $\boldsymbol{A}_p$ and $\boldsymbol{B}_p$ add to the identity matrix $\boldsymbol{I}_p$, and that $\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)} = \boldsymbol{W}_{\epsilon\epsilon(p)}$, where $\boldsymbol{W}_{\epsilon\epsilon(p)}$ is the within individual matrix defined in (21). We also see that $\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}$ and $\widetilde{\boldsymbol{T}}_{\epsilon\epsilon(p)}$ differ from the between individual and total matrices $\boldsymbol{B}_{\epsilon\epsilon(p)}$ and $\boldsymbol{T}_{\epsilon\epsilon(p)}$ as defined in (22) and (20), but coincide with them when the group mean vector $\bar{\boldsymbol{\epsilon}}_{(p)}$ (which is zero asymptotically) is omitted from both expressions.

Subproblem A is identical with the GLS problem for group p, since maximization of $L_{(p)}$ with respect to $\boldsymbol{\beta}$ for given $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ is equivalent to minimization of $Q_{(p)}$, considered in section 3. This gives (41) as the expression for the ML estimator of $\boldsymbol{\beta}$ conditionally on $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ for group p.

To solve *subproblem B*, we need expressions for the derivatives of $L_{(p)}$ with respect to $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$. In Appendix B, we show that

$$\begin{aligned} \frac{\partial L_{(p)}}{\partial \boldsymbol{\Sigma}_u} &= -\frac{1}{2}[N_p\boldsymbol{\Sigma}_{(p)}^{-1} + N_p(p-1)\boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1} - \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}], \\ \frac{\partial L_{(p)}}{\partial \boldsymbol{\Sigma}_\alpha} &= -\frac{1}{2}[N_p p\,\boldsymbol{\Sigma}_{(p)}^{-1} - p\,\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1}]. \end{aligned} \tag{55}$$

From the first order conditions which solve subproblem B, *i.e.*, $\partial L_{(p)}/\partial \boldsymbol{\Sigma}_u = \partial L_{(p)}/\partial \boldsymbol{\Sigma}_\alpha = \mathbf{0}_{GG}$, we get

$$
\begin{aligned}
&N_p \boldsymbol{\Sigma}_{(p)}^{-1} + N_p(p-1)\boldsymbol{\Sigma}_u^{-1} = \boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1} + \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}, \\
&N_p p\, \boldsymbol{\Sigma}_{(p)}^{-1} = p\, \boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1},
\end{aligned}
\tag{56}
$$

which give the ML estimators of the covariance matrices conditionally on $\boldsymbol{\beta}$ for group p,

$$
\widehat{\boldsymbol{\Sigma}}_{(p)} = \widehat{\boldsymbol{\Sigma}}_{(p)}(\boldsymbol{\beta}) = \frac{\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{N_p}, \tag{57}
$$

$$
\widehat{\boldsymbol{\Sigma}}_{u(p)} = \widehat{\boldsymbol{\Sigma}}_{u(p)}(\boldsymbol{\beta}) = \frac{\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}}{N_p(p-1)}. \tag{58}
$$

From this it follows, using (37), that

$$
\widehat{\boldsymbol{\Sigma}}_{\alpha(p)} = \widehat{\boldsymbol{\Sigma}}_{\alpha(p)}(\boldsymbol{\beta}) = \frac{1}{N_p p}\left[\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)} - \frac{\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}}{p-1}\right], \tag{59}
$$

$$
\widehat{\boldsymbol{\Sigma}}_{u(p)} + \widehat{\boldsymbol{\Sigma}}_{\alpha(p)} = \frac{\widetilde{\boldsymbol{T}}_{\epsilon\epsilon(p)}}{N_p p}. \tag{60}
$$

The above estimators of $\boldsymbol{\Sigma}_u$, $\boldsymbol{\Sigma}_\alpha$, and $\boldsymbol{\Sigma}_{(p)}$ are approximately equal to the "estimators" given in (27), (28), and (29) when N_p is not too small.

The complete stepwise, switching algorithm for solving *jointly* subproblems A and B of the ML problem for group p then consists of switching between the GLS estimator (41) and the estimators of the covariance matrices (58) and (59). This algorithm is iterated until convergence. Under certain conditions, this kind of zig-zag procedure will converge to the ML estimator; see Oberhofer and Kmenta (1974).

The ML problem for the complete data set

We write the complete log-density function L as

$$
L = -\frac{Gn}{2}\ln(2\pi) - \frac{1}{2}\sum_{p=1} N_p \ln(|\boldsymbol{\Omega}_{\epsilon(p)}|) - \frac{1}{2}Q(\boldsymbol{\beta}, \boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_\alpha), \tag{61}
$$

where

$$
Q = Q(\boldsymbol{\beta}, \boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_\alpha) = \sum_{p=1}^{P}\sum_{i\in I_p}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]'\boldsymbol{\Omega}_{\epsilon(p)}[\boldsymbol{y}_{i(p)} - \boldsymbol{X}_{i(p)}\boldsymbol{\beta}]. \tag{62}
$$

We split the problem of maximizing L into *two conditional subproblems*:

Subproblem A: Maximization of L with respect to $\boldsymbol{\beta}$ for given $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$.
Subproblem B: Maximization of L with respect to $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ for given $\boldsymbol{\beta}$.

The joint solution to A and B defines the solution to the problem of maximizing L.

Subproblem A is identical with the GLS problem for all individuals, since maximization of L with respect to $\boldsymbol{\beta}$ for given $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ is equivalent to minimization of Q, considered in section 3. This gives (38) as the expression for the conditional ML estimator of $\boldsymbol{\beta}$.

To solve *subproblem B*, we need expressions for the derivatives of L with respect to $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$. From (47) and (55) we find

$$\begin{aligned}\frac{\partial L}{\partial \boldsymbol{\Sigma}_u} &= -\frac{1}{2}\sum_{p=1}^{P}[N_p\boldsymbol{\Sigma}_{(p)}^{-1} + N_p(p-1)\boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1} - \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}],\\ \frac{\partial L}{\partial \boldsymbol{\Sigma}_\alpha} &= -\frac{1}{2}\sum_{p=1}^{P}[N_p p\,\boldsymbol{\Sigma}_{(p)}^{-1} - p\,\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1}].\end{aligned} \tag{63}$$

The first order conditions which solve subproblem B, *i.e.*, $\partial L/\partial\boldsymbol{\Sigma}_u = \partial L/\partial\boldsymbol{\Sigma}_\alpha = \mathbf{0}_{GG}$, then reduce to

$$\begin{aligned}&\sum_{p=1}^{P}[N_p\boldsymbol{\Sigma}_{(p)}^{-1} + N_p(p-1)\boldsymbol{\Sigma}_u^{-1}] = \sum_{p=1}^{P}[\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1} + \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}],\\ &\sum_{p=1}^{P} N_p p\,\boldsymbol{\Sigma}_{(p)}^{-1} = \sum_{p=1}^{P} p\,\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1}.\end{aligned} \tag{64}$$

To obtain the ML estimators, these two sets of non-linear equations, with $\boldsymbol{\Sigma}_{(p)} = \boldsymbol{\Sigma}_u + p\,\boldsymbol{\Sigma}_\alpha$ inserted [cf. (37)], have to be solved jointly for $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$. Unlike the situation with group specific estimation [cf. (57) – (60)], no closed form solution to subproblem B exists.

Inserting for $\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}$ and $\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}$ from (57) and (58) we can express the first order conditions (64) in terms of the group specific estimators of $\boldsymbol{\Sigma}_{(p)}$ and $\boldsymbol{\Sigma}_u$, denoted as $\widehat{\boldsymbol{\Sigma}}_{(p)}$ and $\widehat{\boldsymbol{\Sigma}}_{u(p)}$. We then get

$$\begin{aligned}&\sum_{p=1}^{P} N_p\boldsymbol{\Sigma}_{(p)}^{-1}[\boldsymbol{I}_G - \widehat{\boldsymbol{\Sigma}}_{(p)}\boldsymbol{\Sigma}_{(p)}^{-1}] = -\sum_{p=1}^{P} N_p(p-1)\boldsymbol{\Sigma}_u^{-1}[\boldsymbol{I}_G - \widehat{\boldsymbol{\Sigma}}_{u(p)}\boldsymbol{\Sigma}_u^{-1}],\\ &\sum_{p=1}^{P} N_p p\boldsymbol{\Sigma}_{(p)}^{-1}[\boldsymbol{I}_G - \widehat{\boldsymbol{\Sigma}}_{(p)}\boldsymbol{\Sigma}_{(p)}^{-1}] = \mathbf{0}_{GG},\end{aligned} \tag{65}$$

where $\boldsymbol{\Sigma}_{(p)} = \boldsymbol{\Sigma}_u + p\boldsymbol{\Sigma}_\alpha$. This way of writing the first order conditions shows more explicitly than (64) the compromise nature of the overall estimators of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$.

The complete stepwise, switching algorithm for solving *jointly* subproblems A and B of the ML problem for the complete unbalanced data set consists of switching between GLS estimator (38) and the solution to (64). This algorithm is iterated until convergence. The solution to the non-linear subproblem (64) may require separate iteration loops.

In the *single equation case*, $G = 1$, in which $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$ are scalars, denoted as σ_u^2 and σ_α^2, respectively, things become a little more transparent. Then (64) can be simplified to

$$
\begin{aligned}
&\sum_{p=1}^{P}\left[\frac{N_p}{\sigma_u^2 + p\sigma_\alpha^2} + \frac{N_p(p-1)}{\sigma_u^2}\right] = \sum_{p=1}^{P}\left[\frac{\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(\sigma_u^2 + p\sigma_\alpha^2)^2} + \frac{\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}}{(\sigma_u^2)^2}\right], \\
&\sum_{p=1}^{P}\left[\frac{N_p p}{\sigma_u^2 + p\sigma_\alpha^2}\right] = \sum_{p=1}^{P}\left[\frac{p\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(\sigma_u^2 + p\sigma_\alpha^2)^2}\right],
\end{aligned} \tag{66}
$$

where now $\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}$ and $\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}$ are scalars. Let $\sigma^2 = \sigma_u^2 + \sigma_\alpha^2$ be the variance of the total disturbance and $\rho = \sigma_\alpha^2/\sigma^2$ be the share of this variance which represents individual heterogeneity. Multiplying through (66) by σ^2, we can rewrite the first order conditions as two non-linear equations in σ^2 and ρ:

$$
\begin{aligned}
&\sigma^2 \sum_{p=1}^{P}\left[\frac{N_p}{1-\rho+p\rho} + \frac{N_p(p-1)}{1-\rho}\right] = \sum_{p=1}^{P}\left[\frac{\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(1-\rho+p\rho)^2} + \frac{\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}}{(1-\rho)^2}\right], \\
&\sigma^2 \sum_{p=1}^{P}\left[\frac{N_p p}{1-\rho+p\rho}\right] = \sum_{p=1}^{P}\left[\frac{p\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(1-\rho+p\rho)^2}\right].
\end{aligned} \tag{67}
$$

Eliminating σ^2, we get the following non-linear equation in the share variable ρ:

$$
\begin{aligned}
&\sum_{p=1}^{P}\left[\frac{N_p}{1-\rho+p\rho} + \frac{N_p(p-1)}{1-\rho}\right]\sum_{p=1}^{P}\left[\frac{p\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(1-\rho+p\rho)^2}\right] \\
&\qquad = \sum_{p=1}^{P}\left[\frac{N_p p}{1-\rho+p\rho}\right]\sum_{p=1}^{P}\left[\frac{\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(1-\rho+p\rho)^2} + \frac{\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}}{(1-\rho)^2}\right].
\end{aligned} \tag{68}
$$

In practical applications, this equation can, for instance, be solved, in a first stage, by a grid-search procedure across $\rho = (0,1)$. In the second stage, we can insert the value obtained from the first stage into (67) and solve for σ^2. This gives

$$
\sigma^2 = \frac{\sum_{p=1}^{P}\left[\frac{p\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(1-\rho+p\rho)^2}\right]}{\sum_{p=1}^{P}\left[\frac{N_p p}{1-\rho+p\rho}\right]} = \frac{\sum_{p=1}^{P}\left[\frac{\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}}{(1-\rho+p\rho)^2} + \frac{\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}}{(1-\rho)^2}\right]}{\sum_{p=1}^{P}\left[\frac{N_p}{1-\rho+p\rho} + \frac{N_p(p-1)}{1-\rho}\right]}. \tag{69}
$$

Appendix A: Proof of (16) and (17)

In this appendix, we demonstrate (16) and (17), which are used in deriving unbiased estimators of the covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$. We consider the $\boldsymbol{\epsilon}_{it}$'s as observable. Let T_i denote the number of observations of individual i, *i.e.*, $T_i = p$ for $i \in I_p$. The within individual and the between individual variation, defined in (13) and (14), can then be written as

$$\text{(A.1)} \qquad \boldsymbol{W}_{\epsilon\epsilon} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{i\cdot})(\boldsymbol{\epsilon}_{it} - \bar{\boldsymbol{\epsilon}}_{i\cdot})', \qquad \boldsymbol{B}_{\epsilon\epsilon} = \sum_{i=1}^{N} T_i(\bar{\boldsymbol{\epsilon}}_{i\cdot} - \bar{\boldsymbol{\epsilon}})(\bar{\boldsymbol{\epsilon}}_{i\cdot} - \bar{\boldsymbol{\epsilon}})',$$

where

$$\text{(A.2)} \qquad \bar{\boldsymbol{\epsilon}}_{i\cdot} = \frac{1}{T_i} \sum_{t=1}^{T_i} \boldsymbol{\epsilon}_{it}, \qquad \bar{\boldsymbol{\epsilon}} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \boldsymbol{\epsilon}_{it}}{\sum_{i=1}^{N} T_i} = \frac{\sum_{i=1}^{N} T_i \bar{\boldsymbol{\epsilon}}_{i\cdot}}{\sum_{i=1}^{N} T_i}.$$

We will first show that

$$\text{(A.3)} \qquad \mathsf{E}(\boldsymbol{W}_{\epsilon\epsilon}) = (\textstyle\sum T_i - N)\boldsymbol{\Sigma}_u = (n - N)\boldsymbol{\Sigma}_u,$$

$$\text{(A.4)} \qquad \mathsf{E}(\boldsymbol{B}_{\epsilon\epsilon}) = (N-1)\boldsymbol{\Sigma}_u + \left(\sum T_i - \frac{\sum T_i^2}{\sum T_i}\right)\boldsymbol{\Sigma}_\alpha = (N-1)\boldsymbol{\Sigma}_u + \left(n - \frac{\sum T_i^2}{n}\right)\boldsymbol{\Sigma}_\alpha.$$

The proof of (A.3) – (A.4) is as follows: Since the α_i's and the u_{it}'s are independent,

$$\text{(A.5)} \qquad \mathsf{E}(\boldsymbol{W}_{\epsilon\epsilon}) = \mathsf{E}(\boldsymbol{W}_{uu}),$$

$$\text{(A.6)} \qquad \mathsf{E}(\boldsymbol{B}_{\epsilon\epsilon}) = \mathsf{E}(\boldsymbol{B}_{\alpha\alpha}) + \mathsf{E}(\boldsymbol{B}_{uu}),$$

where

$$\boldsymbol{W}_{uu} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (\boldsymbol{u}_{it} - \bar{\boldsymbol{u}}_{i\cdot})(\boldsymbol{u}_{it} - \bar{\boldsymbol{u}}_{i\cdot})' = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \boldsymbol{u}_{it}\boldsymbol{u}_{it}' - \sum_{i=1}^{N} T_i \bar{\boldsymbol{u}}_{i\cdot}\bar{\boldsymbol{u}}_{i\cdot}',$$

$$\boldsymbol{B}_{uu} = \sum_{i=1}^{N} T_i(\bar{\boldsymbol{u}}_{i\cdot} - \bar{\boldsymbol{u}})(\bar{\boldsymbol{u}}_{i\cdot} - \bar{\boldsymbol{u}})' = \sum_{i=1}^{N} T_i \bar{\boldsymbol{u}}_{i\cdot}\bar{\boldsymbol{u}}_{i\cdot}' - \left(\sum_{i=1}^{N} T_i\right)\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}',$$

$$\boldsymbol{B}_{\alpha\alpha} = \sum_{i=1}^{N} T_i(\boldsymbol{\alpha}_i - \bar{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_i - \bar{\boldsymbol{\alpha}})' = \sum_{i=1}^{N} T_i \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i' - \left(\sum_{i=1}^{N} T_i\right)\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}',$$

$$\bar{\boldsymbol{u}}_{i\cdot} = \frac{1}{T_i} \sum_{t=1}^{T_i} \boldsymbol{u}_{it}, \qquad \bar{\boldsymbol{u}} = \frac{\sum_{i=1}^{N} T_i \bar{\boldsymbol{u}}_{i\cdot}}{\sum_{i=1}^{N} T_i}, \qquad \bar{\boldsymbol{\alpha}} = \frac{\sum_{i=1}^{N} T_i \boldsymbol{\alpha}_i}{\sum_{i=1}^{N} T_i}.$$

From (6) – (8) it follows that

$$\mathsf{E}(\bar{\boldsymbol{u}}_{i\cdot}\bar{\boldsymbol{u}}_{i\cdot}') = \frac{\boldsymbol{\Sigma}_u}{T_i}, \qquad \mathsf{E}(\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}') = \frac{\boldsymbol{\Sigma}_u}{\sum_i T_i}, \qquad \mathsf{E}(\bar{\boldsymbol{\alpha}}\bar{\boldsymbol{\alpha}}') = \frac{(\sum T_i^2)\boldsymbol{\Sigma}_\alpha}{(\sum T_i)^2}.$$

This implies

$$\begin{aligned}
\mathsf{E}(\boldsymbol{W}_{uu}) &= (\textstyle\sum T_i - N)\boldsymbol{\Sigma}_u = (n - N)\boldsymbol{\Sigma}_u, \\
\mathsf{E}(\boldsymbol{B}_{uu}) &= (N-1)\boldsymbol{\Sigma}_u, \\
\mathsf{E}(\boldsymbol{B}_{\alpha\alpha}) &= \left(\sum T_i - \frac{\sum T_i^2}{\sum T_i}\right)\boldsymbol{\Sigma}_\alpha = \left(n - \frac{\sum T_i^2}{n}\right)\boldsymbol{\Sigma}_\alpha.
\end{aligned}$$

Combining these expressions with (A.5) and (A.6) completes the proof of (A.3) and (A.4).

Inserting $\sum_{i=1}^N T_i = \sum_{p=1}^P N_p p = n$ and $\sum_{i=1}^N T_i^2 = \sum_{p=1}^P N_p p^2$, we finally get

$$\mathsf{E}(\boldsymbol{W}_{\epsilon\epsilon}) = (n-N)\boldsymbol{\Sigma}_u, \tag{A.7}$$

$$\mathsf{E}(\boldsymbol{B}_{\epsilon\epsilon}) = (N-1)\boldsymbol{\Sigma}_u + \left(n - \frac{\sum_{p=1}^P N_p p^2}{n}\right)\boldsymbol{\Sigma}_\alpha. \tag{A.8}$$

This completes the proof of (16) and (17).

Appendix B: Proof of (55)

In this appendix, we give a proof of the equations for the first derivatives of the log-likelihood function $L_{(p)}$ with respect to the covariance matrices $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\alpha$.

For this purpose, we exploit four matrix results on traces, determinants, and derivatives:

(a) $\mathrm{tr}(\boldsymbol{ABCD}) = \mathrm{tr}(\boldsymbol{CDAB}) = \mathrm{vec}(\boldsymbol{A}')'(\boldsymbol{D}' \otimes \boldsymbol{B})\mathrm{vec}(\boldsymbol{C}) = \mathrm{vec}(\boldsymbol{C}')'(\boldsymbol{B}' \otimes \boldsymbol{D})\mathrm{vec}(\boldsymbol{A})$,

see Lütkepohl (1996, pp. 41 – 42),

$$(b) \qquad |\boldsymbol{A}_p \otimes \boldsymbol{C} + \boldsymbol{B}_p \otimes \boldsymbol{D}| = |\boldsymbol{C}|\,|\boldsymbol{D}|^{p-1},$$

where $\boldsymbol{A}_p = (1/p)\boldsymbol{E}_p$, with rank 1, and $\boldsymbol{B}_p = \boldsymbol{I}_p - (1/p)\boldsymbol{E}_p$, with rank $p-1$, which follows from Magnus (1982, Lemma 2.1),

$$(c) \qquad \frac{\partial \ln |\boldsymbol{A}|}{\partial \boldsymbol{A}} = (\boldsymbol{A}')^{-1},$$

see Magnus and Neudecker (1988, p. 179) or Lütkepohl (1996, p. 182), and

$$(d) \qquad \frac{\partial \mathrm{tr}(\boldsymbol{CB}^{-1})}{\partial \boldsymbol{B}} = -(\boldsymbol{B}^{-1}\boldsymbol{C}\,\boldsymbol{B}^{-1})',$$

see Magnus and Neudecker (1988, p. 178) or Lütkepohl (1996, p. 179).

Let the part of the quadratic form $Q_{(p)}$, defined in (40), which relates to individual i be

$$\begin{aligned}
Q_{i(p)} = \widehat{\boldsymbol{\epsilon}}'_{i(p)}\boldsymbol{\Omega}^{-1}_{\epsilon(p)}\widehat{\boldsymbol{\epsilon}}_{i(p)} &= \widehat{\boldsymbol{\epsilon}}'_{i(p)}[\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}]\widehat{\boldsymbol{\epsilon}}_{i(p)} + \widehat{\boldsymbol{\epsilon}}'_{i(p)}[\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}]\widehat{\boldsymbol{\epsilon}}_{i(p)} \\
&= \mathrm{vec}(\widetilde{\boldsymbol{E}}_{i(p)})'[\boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u^{-1}]\mathrm{vec}(\widetilde{\boldsymbol{E}}_{i(p)}) \\
&\quad + \mathrm{vec}(\widetilde{\boldsymbol{E}}_{i(p)})'[\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)}^{-1}]\mathrm{vec}(\widetilde{\boldsymbol{E}}_{i(p)})),
\end{aligned}$$

when using (51). Using (a), we find by addition

$$\begin{aligned} Q_{(p)} &= \sum_{i\in I_p} \mathrm{tr}(Q_{i(p)}) = \sum_{i\in I_p} \left[\mathrm{tr}[\widetilde{\boldsymbol{E}}'_{i(p)}\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{A}_p] + \mathrm{tr}[\widetilde{\boldsymbol{E}}'_{i(p)}\boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{B}_p]\right] \\ &= \sum_{i\in I_p} \left[\mathrm{tr}[\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{A}_p\widetilde{\boldsymbol{E}}'_{i(p)}\boldsymbol{\Sigma}_{(p)}^{-1}] + \mathrm{tr}[\widetilde{\boldsymbol{E}}_{i(p)}\boldsymbol{B}_p\widetilde{\boldsymbol{E}}'_{i(p)}\boldsymbol{\Sigma}_u^{-1}]\right], \end{aligned}$$

which, when using (53) and (54), can be rewritten as

$$Q_{(p)} = \mathrm{tr}[\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1}] + \mathrm{tr}[\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}]. \tag{B.1}$$

Using (b), we find the following expression for the determinant of $\boldsymbol{\Omega}_{\epsilon(p)}$

$$|\boldsymbol{\Omega}_{\epsilon(p)}| = |\boldsymbol{A}_p \otimes \boldsymbol{\Sigma}_{(p)} + \boldsymbol{B}_p \otimes \boldsymbol{\Sigma}_u| = |\boldsymbol{\Sigma}_{(p)}||\boldsymbol{\Sigma}_u|^{p-1}. \tag{B.2}$$

From (c) and (d) we finally obtain

$$\begin{aligned} \frac{\partial Q_{(p)}}{\partial \boldsymbol{\Sigma}_u} &= -\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1} - \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}, \\ \frac{\partial Q_{(p)}}{\partial \boldsymbol{\Sigma}_\alpha} &= -p\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1}, \end{aligned} \tag{B.3}$$

and

$$\begin{aligned} \frac{\partial \ln|\boldsymbol{\Omega}_{\epsilon(p)}|}{\partial \boldsymbol{\Sigma}_u} &= \frac{\partial \ln|\boldsymbol{\Sigma}_{(p)}|}{\partial \boldsymbol{\Sigma}_u} + (p-1)\frac{\partial \ln|\boldsymbol{\Sigma}_u|}{\partial \boldsymbol{\Sigma}_u} = \boldsymbol{\Sigma}_{(p)}^{-1} + (p-1)\boldsymbol{\Sigma}_u^{-1}, \\ \frac{\partial \ln|\boldsymbol{\Omega}_{\epsilon(p)}|}{\partial \boldsymbol{\Sigma}_\alpha} &= \frac{\partial \ln|\boldsymbol{\Sigma}_{(p)}|}{\partial \boldsymbol{\Sigma}_\alpha} = p\boldsymbol{\Sigma}_{(p)}^{-1}. \end{aligned} \tag{B.4}$$

Collecting these results, we find

$$\begin{aligned} \frac{\partial L_{(p)}}{\partial \boldsymbol{\Sigma}_u} &= -\frac{1}{2}[N_p\boldsymbol{\Sigma}_{(p)}^{-1} + N_p(p-1)\boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1} - \boldsymbol{\Sigma}_u^{-1}\widetilde{\boldsymbol{W}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_u^{-1}], \\ \frac{\partial L_{(p)}}{\partial \boldsymbol{\Sigma}_\alpha} &= -\frac{1}{2}[N_p p\,\boldsymbol{\Sigma}_{(p)}^{-1} - p\,\boldsymbol{\Sigma}_{(p)}^{-1}\widetilde{\boldsymbol{B}}_{\epsilon\epsilon(p)}\boldsymbol{\Sigma}_{(p)}^{-1}]. \end{aligned} \tag{B.5}$$

These are the expressions given in (55).

References

Avery, R.B. (1977): Error Components and Seemingly Unrelated Regressions. *Econometrica*, 45 (1977), 199 – 209.

Baltagi, B.H. (1980): On Seemingly Unrelated Regressions with Error Components. *Econometrica*, 48 (1980), 1547 – 1551.

Baltagi, B.H. (1985): Pooling Cross-Sections with Unequal Time-Series Lengths. *Economics Letters*, 18 (1985), 133 – 136.

Biørn, E. (1981): Estimating Economic Relations from Incomplete Cross-Section/Time-Series Data. *Journal of Econometrics*, 16 (1981), 221 – 236.

Biørn, E. (1999): Random Coefficients in Regression Equation Systems: the Case with Unbalanced Panel Data. Unpublished paper.

Greene, W.H. (1997): *Econometric Analysis,* third edition. London: Prentice Hall, 1997.

Lütkepohl, H. (1996): *Handbook of Matrices.* Chichester: Wiley, 1996.

Magnus, J.R. (1982): Multivariate Error Components Analysis of Linear and Nonlinear Regression Models by Maximum Likelihood. *Journal of Econometrics*, 19 (1982), 239 – 285.

Magnus, J.R. and Neudecker, H. (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: Wiley, 1988.

Oberhofer, W. and Kmenta, J. (1974): A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models. *Econometrica*, 42 (1974), 579 – 590.

Wansbeek, T. and Kapteyn, A. (1989): Estimation of the Error Components Model with Incomplete Panels. *Journal of Econometrics*, 41 (1989), 341 – 361.