MEMORANDUM

No 27/99

Random Coefficients in Regression Equation Systems: The Case with unbalanced Panel Data

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 26	By Gunnar Bårdsen, Eilev S. Jansen and Ragnar Nymoen: Econometric inflation targeting. 38 p.
No 25	By Cheetal K Chand and Karl O. Moene: Rent Grabbing and Russia's Economic Collapse. 35 p.
No 24	By Rolf Golombek and Espen R. Moen: Do Voluntary Agreements Lead to Cost Efficiency? 45 p.
No 23	By Atle Seierstad: Necessary conditions involving generalized directional derivatives for optimal control of nonsmooth retarded Volterra integral equations in Banach space. 68 p.
No 22	By Karl Ove Moene and Michael Wallerstein: Inequality, Social Insurance and Redistribution. 40 p.
No 21	By Kai Leitemo: Inflation Targeting Strategies in small open Economies. 44 p.
No 20	By Erik Biørn: Estimating Regression Systems from unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure. 18 p.
No 19	By Knut Røed and Steinar Strøm: Progressive 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.

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

RANDOM COEFFICIENTS IN REGRESSION EQUATION SYSTEMS: THE CASE WITH UNBALANCED PANEL DATA*)

by

ERIK BIØRN

ABSTRACT

We consider a framework for analyzing panel data characterized by: (i) a system of regressions equations, (ii) random individual heterogeneity in both intercepts and slope coefficients, and (iii) unbalanced panel data, *i.e.*, panel data where the individual time series have unequal length. A Maximum Likelihood (ML) procedure for joint estimation of all parameters is described. Since it is complicated to implement in numerical calculations, we consider simplified procedures, in particular for estimating the covariance matrices of the random coefficients. An algorithm for modified ML estimation of all parameters is presented.

Keywords: Panel Data. Unbalanced Panels. Random Coefficients. Heterogeneity. Regression Equation Systems. Maximum Likelihood
JEL classification: C13, C23, C33

---

* I thank Jørgen Aasness and Terje Skjerpen for valuable comments.

1 Introduction

A challenge in the analysis of economic relationships by means of micro data in general and panel data in particular is how to treat heterogeneity regarding the form of the relationships across the units or groups in the data set. Many researchers assume a common coefficient structure, possibly allowing for unit specific (or time specific) differences in the intercept terms of the equations ('fixed' or 'random' effects) only. If the heterogeneity has a more complex form, this approach may lead to inefficient (and even inconsistent) estimation of the slope coefficients, and invalid inference.

A more appealing modelling approach is to allow for heterogeneity not only in the intercepts, but also in the slope coefficients. We may, for instance, want to investigate heterogeneity in returns to scale coefficients and elasticities of substitution across firms in factor demand, Engel and Cournot elasticities across households in commodity demand, or accelerator coefficients in investment equations. The challenge then becomes how to construct a model which is sufficiently flexible without being overparametrized. The *fixed coefficients* approach, in which each unit has its distinct coefficient vector, with no assumptions made about its variation between units, is very flexible, but may easily suffer from this overparametrization problem; the number of degrees of freedom may be too low to permit reliable inference. The *random coefficients* approach, in which specific assumptions are made about the distribution from which the unit specific coefficients are 'drawn', is far more parsimonious in general. The common expectation vector of these coefficients represents, in a precise way, the coefficients of an average unit, *e.g.*, the average scale elasticity or the average Engel elasticity, while its covariance matrix gives readily interpretable measures of the degree of heterogeneity. Moreover, the random coefficients approach may be considered a parsimonious way of representing certain kinds of disturbance heteroskedasticity in panel data analysis.

There is a growing number of methodological papers in the econometric literature dealing with this random coefficient problem for balanced panel data situations [see Longford (1995) and Hsiao (1996) for recent surveys]. Early contributions to the econometric literature on random coefficients for linear, static single regression equations with *balanced* panel data are Swamy (1970, 1971, 1974), Hsiao (1975), and Swamy and Mehta (1977). Estimation problems for the covariance matrices of such models are discussed in Wansbeek and Kapteyn (1982). Avery (1977) and Baltagi (1980) consider systems of regression equations with random effects in the intercept term for balanced panels. Biørn (1981), Baltagi (1985), and Wansbeek and Kapteyn (1989) consider a single regression equation with random effects in the intercept term for *unbalanced* panels. The model under consideration in the present paper can be considered a generalization of all the models in the papers mentioned above, except that we will allow for heterogeneity along the individual dimension only. Sometimes a two-way decomposition with symmetric treatment of individual and time effects is allowed for. An empirical application of some of the procedures described in this paper is presented in Biørn, Lindquist, and Skjerpen (1998). In general, far less has been done for unbalanced than for the formally simpler balanced cases. This is surprising, since in practice, the latter is the exception rather than the rule. We may waste a lot of observations if we force ourselves to use a balanced subpanel constructed from an originally unbalanced data set.

In this paper, we consider a framework for analyzing panel data with the following characteristics: (i) a system of linear, static regressions equations, (ii) random individual heterogeneity in both intercepts and slope coefficients, structured via their first and second order moments, and (iii) unbalanced panel data.[1] The model is presented in Section 2. We show a way to treat constraints on coefficients in different equations. Section 3 describes the main stages in Maximum Likelihood (ML) estimation and its relationship to Generalized Least Squares (GLS). A basic difficulty with applying ML in the present context stems from the unbalancedness of the panel in combination with the rather complex way in which the covariance matrices of the random slack variables enter the likelihood function. In Section 4, we consider a simpler, stepwise procedure for estimation of these covariance matrices. Finally, Section 5 summarizes the preceeding sections and presents a simplified algorithm for modified ML estimation.

2 Model and notation

We consider a linear, static regression model with G equations, indexed by $g = 1, \ldots, G$, equation g having K_g regressors. The data are from an unbalanced panel, in which the individuals are observed in at least 1 and at most P periods. In descriptions of unbalanced panel data sets, the observations of a specific individual i is often indexed as $t = 1, \ldots, T_i$, where T_i is the number of observations of individual i [see, *e.g.*, Baltagi (1995, section 9.3)]. Our notation convention is this paper is somewhat different. The individuals are assumed to be arranged in groups according to the number of times they are observed. Let N_p be the number of individuals observed in p periods (not necessarily the same and not necessarily consecutive), let (ip) index individual i in group p $(i = 1, \ldots, N_p;\ p = 1, \ldots, P)$, and let t index the number of the running observation $(t = 1, \ldots, p)$. In unbalanced panels, t differs from the calendar period (year, quar-

[1] We assume that the selection rules for the unbalanced panels are *ignorable*, *i.e.*, the way in which the units or groups enter or exit is not related to the endogenous variables in the model. See Verbeek and Nijman (1996, section 18.2) for an elaboration of this topic.

ter etc.).[2] The total number of individuals and the total number of observations are then $N = \sum_{p=1}^{P} N_p$ and $n = \sum_{p=1}^{P} N_p p$, respectively. Formally, the data set in group p ($p = 2, \dots, P$) can be considered a balanced panel data set with p observations of each of the N_p units, while the data set in group 1 is a cross section. Rotating panels are special cases of this kind of data, in which a share of the individuals which are included in the panel in one period is replaced by a fresh sample drawn in the next period [see, *e.g.*, Biørn (1981)].

Two ways of formulating the model are convenient, depending on whether (a) the G equations contain disjoint sets of coefficients or (b) some, or all, of the equations have some coefficients in common. We first consider case (a), next describe the modifications needed in case (b), and then describe a general formulation which includes both (a) and (b).

When *each equation has its distinct coefficient vector*, the total number of coefficients is $K = \sum_{g=1}^{G} K_g$. Let the $(p \times 1)$ vector of observations of the regressand in eq. g from individual (ip) be $\boldsymbol{y}_{g(ip)}$, let its $(p \times K_g)$ regressor matrix be $\boldsymbol{X}_{g(ip)}$ (including a vector of ones associated with the intercept term), and let $\boldsymbol{u}_{g(ip)}$ be the $(p \times 1)$ vector of disturbances in eq. g from individual (ip). We allow for individual heterogeneity and represent it, for equation g and individual (ip), by the *random individual coefficient vector* $\boldsymbol{\beta}_{g(ip)}$ (including the intercept) as

$$\boldsymbol{\beta}_{g(ip)} = \boldsymbol{\beta}_g + \boldsymbol{\delta}_{g(ip)}, \tag{1}$$

where $\boldsymbol{\beta}_g$ is a fixed constant vector and $\boldsymbol{\delta}_{g(ip)}$ its random shift variable. We assume that $\boldsymbol{X}_{g(ip)}$, $\boldsymbol{u}_{g(ip)}$, and $\boldsymbol{\delta}_{g(ip)}$ are mutually independent, that $\boldsymbol{u}_{g(ip)}$ and $\boldsymbol{\delta}_{g(ip)}$ are independent and homoskedastic across (ip), but correlated across g, and that

$$\mathsf{E}[\boldsymbol{\delta}_{g(ip)}] = \mathbf{0}_{K_g,1}, \quad \mathsf{E}[\boldsymbol{\delta}_{g(ip)}\boldsymbol{\delta}'_{h(ip)}] = \boldsymbol{\Sigma}^{\delta}_{gh}, \tag{2}$$

$$\mathsf{E}[\boldsymbol{u}_{g(ip)}] = \mathbf{0}_{p,1}, \quad \mathsf{E}[\boldsymbol{u}_{g(ip)}\boldsymbol{u}'_{h(ip)}] = \sigma^{u}_{gh}\boldsymbol{I}_p, \qquad g,h = 1,\dots,G, \tag{3}$$

where $\mathbf{0}_{m,n}$ is the $(m \times n)$ zero matrix and $\boldsymbol{I}_p$ is the $(p \times p)$ identity matrix. Eq. g for individual (ip) is

$$\boldsymbol{y}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\beta}_{g(ip)} + \boldsymbol{u}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\beta}_g + \boldsymbol{\eta}_{g(ip)}, \qquad g = 1,\dots,G,\ i = 1,\dots,N_p,\ p = 1,\dots,P, \tag{4}$$

[2] Subscripts symbolizing the *calendar* period may be attached. This may be convenient for data documentation purposes and in formulating dynamic models, but will not be necessary for the static model we consider in this paper. For example, in a data set with $P = 20$, from the years 1971 – 1990, some individuals in the $p = 18$ group may be observed in the years 1971 – 1988, some in 1972 – 1989, etc. Others in the $p = 18$ group may be observed with gaps in the series, *e.g.*, in 1971 – 1980 and 1982 – 1989, etc. This may be indicated by separate subscripts.

where we interpret

$$\boldsymbol{\eta}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\delta}_{g(ip)} + \boldsymbol{u}_{g(ip)} \tag{5}$$

as a *gross disturbance vector*, representing both the genuine disturbances and the random variation in the coefficients. It follows from (2) and (3) that these gross disturbance vectors are independent across individuals and heteroskedastic, with[3]

$$\mathsf{E}[\boldsymbol{\eta}_{g(ip)}] = \boldsymbol{0}_{p,1}, \quad \mathsf{E}[\boldsymbol{\eta}_{g(ip)}\boldsymbol{\eta}'_{h(ip)}] = \boldsymbol{X}_{g(ip)}\boldsymbol{\Sigma}^{\delta}_{gh}\boldsymbol{X}'_{h(ip)} + \sigma^{u}_{gh}\boldsymbol{I}_p. \tag{6}$$

The heteroskedasticity of $\boldsymbol{\eta}_{g(ip)}$ is due to the random components of the slope coefficients.

When *some coefficients occur in at least two equations* – which may reflect, for instance, cross-equational (*e.g.*, symmetry) constraints resulting from micro units' optimizing behaviour[4] – the total number of free coefficients, K, is less than $\sum_{g=1}^{G} K_g$. We then replace (1) by

$$\boldsymbol{\beta}_{(ip)} = \boldsymbol{\beta} + \boldsymbol{\delta}_{(ip)}, \tag{7}$$

where $\boldsymbol{\beta}_{(ip)}$ is the random $(K \times 1)$ vector containing *all* the coefficients in the model, $\boldsymbol{\beta}$ is a fixed vector and $\boldsymbol{\delta}_{(ip)}$ is its random shift variable. We reinterpret $\boldsymbol{X}_{g(ip)}$ as the $(p \times K)$ matrix (including a vector of ones associated with the intercept term) of regressors in the g'th equation whose k'th column contains the observations on the variable corresponding to the k'th coefficient in $\boldsymbol{\beta}_{(ip)}$ $(k = 1, \ldots, K)$. If the latter coefficient does not occur in the g'th equation, the k'th column of $\boldsymbol{X}_{g(ip)}$ is set to zero.[5] We retain (3) and replace (2), (4), and (5) by

$$\mathsf{E}[\boldsymbol{\delta}_{(ip)}] = \boldsymbol{0}_{K,1}, \quad \mathsf{E}[\boldsymbol{\delta}_{(ip)}\boldsymbol{\delta}'_{(ip)}] = \boldsymbol{\Sigma}^{\delta}, \tag{8}$$

$$\boldsymbol{y}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\beta}_{(ip)} + \boldsymbol{u}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\beta} + \boldsymbol{\eta}_{g(ip)}, \tag{9}$$

$$\boldsymbol{\eta}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\delta}_{(ip)} + \boldsymbol{u}_{g(ip)}, \tag{10}$$

$$g = 1, \ldots, G,\ i = 1, \ldots, N_p,\ p = 1, \ldots, P,$$

It follows from (3), (8), and (10) that with this modification of the model, (6) is replaced by

$$\mathsf{E}[\boldsymbol{\eta}_{g(ip)}] = \boldsymbol{0}_{p,1}, \quad \mathsf{E}[\boldsymbol{\eta}_{g(ip)}\boldsymbol{\eta}'_{h(ip)}] = \boldsymbol{X}_{g(ip)}\boldsymbol{\Sigma}^{\delta}\boldsymbol{X}'_{h(ip)} + \sigma^{u}_{gh}\boldsymbol{I}_p. \tag{11}$$

We stack, for each individual, the $\boldsymbol{y}$'s, $\boldsymbol{\delta}$'s, $\boldsymbol{u}$'s, and $\boldsymbol{\eta}$'s by equations, and define, for individual (ip),

$$\begin{aligned} \boldsymbol{y}_{(ip)} &= (\boldsymbol{y}'_{1(ip)}, \ldots, \boldsymbol{y}'_{G(ip)})', \quad \boldsymbol{\delta}_{(ip)} = (\boldsymbol{\delta}'_{1(ip)}, \ldots, \boldsymbol{\delta}'_{G(ip)})', \\ \boldsymbol{u}_{(ip)} &= (\boldsymbol{u}'_{1(ip)}, \ldots, \boldsymbol{u}'_{G(ip)})', \quad \boldsymbol{\eta}_{(ip)} = (\boldsymbol{\eta}'_{1(ip)}, \ldots, \boldsymbol{\eta}'_{G(ip)})' \end{aligned}$$

[3]Strictly, these properties hold conditionally on $(\boldsymbol{X}_{g(ip)}, \boldsymbol{X}_{h(ip)})$.

[4]These (deterministic) coefficient restrictions relate to both the expectation and the random part of the coefficients.

[5]This interpretation is, of course, also valid when different equations have different regressors.

and let

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

In the case where each equation has its distinct coefficient vector, we define

$$\boldsymbol{X}_{(ip)} = \begin{bmatrix} \boldsymbol{X}_{1(ip)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{X}_{G(ip)} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{1} \\ \vdots \\ \boldsymbol{\beta}_{G} \end{bmatrix}, \quad \boldsymbol{\Sigma}^{\delta} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{\delta} & \cdots & \boldsymbol{\Sigma}_{1G}^{\delta} \\ \vdots & & \vdots \\ \boldsymbol{\Sigma}_{G1}^{\delta} & \cdots & \boldsymbol{\Sigma}_{GG}^{\delta} \end{bmatrix}.$$

In the case where some coefficients occur in at least two equations, we define $\boldsymbol{X}_{(ip)}$ as

$$\boldsymbol{X}_{(ip)} = \begin{bmatrix} \boldsymbol{X}_{1(ip)} \\ \vdots \\ \boldsymbol{X}_{G(ip)} \end{bmatrix}$$

and reinterpret $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{\delta}$ as explained above. We can then write the *model for both cases* as

$$\boldsymbol{y}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\beta} + \boldsymbol{\eta}_{(ip)}, \qquad \boldsymbol{\eta}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\delta}_{(ip)} + \boldsymbol{u}_{(ip)}, \tag{12}$$

$$\boldsymbol{\delta}_{(ip)} \sim \mathsf{IID}(\mathbf{0}_{K,1}, \boldsymbol{\Sigma}^{\delta}), \qquad \boldsymbol{u}_{(ip)} \sim \mathsf{IID}(\mathbf{0}_{Gp,1}, \boldsymbol{I}_{p} \otimes \boldsymbol{\Sigma}^{u}), \tag{13}$$

$$\mathsf{E}[\boldsymbol{\eta}_{(ip)}] = \mathbf{0}_{Gp}, \quad \mathsf{E}[\boldsymbol{\eta}_{(ip)}\boldsymbol{\eta}_{(ip)}'] = \boldsymbol{\Omega}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\Sigma}^{\delta}\boldsymbol{X}_{(ip)}' + \boldsymbol{I}_{p} \otimes \boldsymbol{\Sigma}^{u}, \tag{14}$$

where $\otimes$ is the Kronecker product operator and $\boldsymbol{\Omega}_{(ip)}$ is the gross disturbance covariance matrix of individual (ip) defined by the last equality in (14).

3 The Maximum Likelihood problem

We now describe the Maximum Likelihood problem for joint estimation of the coefficients and the disturbance covariance matrices in the model (12) – (14), and consider the main stages in its solution. We make the additional assumption that the random components of the coefficients and the disturbances are *normally* distributed and replace (13) by

$$\boldsymbol{\delta}_{(ip)} \sim \mathsf{IIN}\,(\mathbf{0}_{K,1}, \boldsymbol{\Sigma}^{\delta}), \qquad \boldsymbol{u}_{(ip)} \sim \mathsf{IIN}\,(\mathbf{0}_{Gp,1}, \boldsymbol{I}_{p} \otimes \boldsymbol{\Sigma}^{u}).$$

Then the $\boldsymbol{\eta}_{(ip)}|\boldsymbol{X}_{(ip)}$'s are independent across (ip) and distributed as $\mathsf{N}(\mathbf{0}_{Gp,1}, \boldsymbol{\Omega}_{(ip)})$, with $\boldsymbol{\Omega}_{(ip)}$ defined as in (14). The log-density function of $\boldsymbol{y}_{(ip)}|\boldsymbol{X}_{(ip)}$ is

$$L_{(ip)} = -\,\frac{Gp}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Omega}_{(ip)}| - \frac{1}{2}[\boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\boldsymbol{\beta}]'\boldsymbol{\Omega}_{(ip)}^{-1}[\boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\boldsymbol{\beta}].$$

Let $Q_{(ip)}$ be the quadratic form in this expression, *i.e.*,

$$Q_{(ip)} = [\boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\boldsymbol{\beta}]'\boldsymbol{\Omega}_{(ip)}^{-1}[\boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\boldsymbol{\beta}] = \boldsymbol{\eta}'_{(ip)}\boldsymbol{\Omega}_{(ip)}^{-1}\boldsymbol{\eta}_{(ip)}. \tag{15}$$

The log-likelihood function of all $\boldsymbol{y}$'s conditionally on all $\boldsymbol{X}$'s for group p, *i.e.*, the individuals observed p times, can be written as

$$L_{(p)} = \sum_{i=1}^{N_p} L_{(ip)} = -\frac{GN_p p}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N_p}\ln|\boldsymbol{\Omega}_{(ip)}| - \frac{1}{2}\sum_{i=1}^{N_p} Q_{(ip)}, \tag{16}$$

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

$$L = \sum_{p=1}^{P} L_{(p)} = -\frac{Gn}{2}\ln(2\pi) - \frac{1}{2}\sum_{p=1}^{P}\sum_{i=1}^{N_p}\ln|\boldsymbol{\Omega}_{(ip)}| - \frac{1}{2}\sum_{p=1}^{P}\sum_{i=1}^{N_p} Q_{(ip)}. \tag{17}$$

Two Maximum Likelihood (ML) problems are of interest: group specific estimation and joint estimation for all groups. The ML estimators of $(\boldsymbol{\beta}, \boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta)$ *for group* p are the values that maximize $L_{(p)}$. The ML estimators *of* $(\boldsymbol{\beta}, \boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta)$ *based on the complete data set* are the values that maximize L. Both problems are subject to $\boldsymbol{\Omega}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\Sigma}^\delta\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \boldsymbol{\Sigma}^u$ [cf. (14)].

The structure of the group specific problem is more complicated than the ML problem for systems of regression equations in standard situations with balanced panel data and constant coefficients and random intercept terms [cf. Avery (1977) and Baltagi (1980)], since different individuals have different 'gross' disturbance covariance matrices, $\boldsymbol{\Omega}_{(ip)}$, depending on $\boldsymbol{X}_{(ip)}$ when $\boldsymbol{\Sigma}^\delta$ is non-zero. The structure of the joint estimation problem – for the complete unbalanced panel – is still more complicated since the various $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{\Omega}$ matrices have different number of rows, reflecting the different number of observations of the individuals. Although the dimensions of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ are the same for all individuals, the dimensions of $\boldsymbol{X}_{(ip)}$ and $\boldsymbol{I}_p$, and hence of $\boldsymbol{\Omega}_{(ip)}$, differ.

We describe these ML problems in turn.

ML estimation for group p

We set the derivatives of $L_{(p)}$ with respect to $\boldsymbol{\beta}, \boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta$ equal to zero, and obtain the first order conditions

$$\sum_{i=1}^{N_p}\left(\frac{\partial Q_{(ip)}}{\partial \boldsymbol{\beta}}\right) = \mathbf{0}_{K,1}, \tag{18}$$

$$\begin{cases} \displaystyle\sum_{i=1}^{N_p}\left(\frac{\partial \ln|\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^u} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^u}\right) = \mathbf{0}_{G,G}, \\ \displaystyle\sum_{i=1}^{N_p}\left(\frac{\partial \ln|\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^\delta} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^\delta}\right) = \mathbf{0}_{K,K}. \end{cases} \tag{19}$$

These equations define the solution to the ML problem for group p if $p = 2, 3, \ldots, P$, each value of p giving a distinct set of estimators. If $p = 1$, (18) is solvable with respect to $\boldsymbol{\beta}$, conditionally on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, but (19) is unsolvable.

Conditions (18) coincide with the conditions that solve the Generalized Least Squares (GLS) problem for $\boldsymbol{\beta}$ for group p, conditionally on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, and we find that the solution value is

$$(20) \qquad \widehat{\boldsymbol{\beta}}_{(p)}^{GLS} = \left[\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{X}_{(ip)} \right]^{-1} \left[\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{y}_{(ip)} \right].$$

By inserting this value of $\boldsymbol{\beta}$ into the expressions for $Q_{(ip)}$ in (16), we obtain the concentrated log-likelihood function for group p; $p = 2, \ldots, P$. This can be maximized with respect to $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ (subject to the symmetry conditions) to give the group specific estimators of these covariance matrices. We do not elaborate the details of the latter maximization.

ML estimation for all groups jointly

We set the derivatives of L with respect to $\boldsymbol{\beta}, \boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta$ equal to zero and obtain the first order conditions

$$(21) \qquad \sum_{p=1}^{P} \sum_{i=1}^{N_p} \left(\frac{\partial Q_{(ip)}}{\partial \boldsymbol{\beta}} \right) = \mathbf{0}_{K,1},$$

$$(22) \qquad \begin{cases} \displaystyle\sum_{p=1}^{P} \sum_{i=1}^{N_p} \left(\frac{\partial \ln |\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^u} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^u} \right) = \mathbf{0}_{G,G}, \\ \displaystyle\sum_{p=1}^{P} \sum_{i=1}^{N_p} \left(\frac{\partial \ln |\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^\delta} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^\delta} \right) = \mathbf{0}_{K,K}. \end{cases}$$

These equations define the solution to the overall ML problem.

Conditions (21) coincide with the conditions that solve the GLS problem for $\boldsymbol{\beta}$ for the complete data set, conditionally on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, and we find that the solution value is

$$(23) \qquad \widehat{\boldsymbol{\beta}}^{GLS} = \left[\sum_{p=1}^{P} \sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{X}_{(ip)} \right]^{-1} \left[\sum_{p=1}^{P} \sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{y}_{(ip)} \right].$$

By inserting this solution into (17), we obtain the concentrated log-likelihood function, which can be maximized with respect to $\boldsymbol{\beta}, \boldsymbol{\Sigma}^u$, and $\boldsymbol{\Sigma}^\delta$ (subject to the symmetry conditions) to give the estimators of these covariance matrices. We do not elaborate the details of the latter procedure.

4 Simplified estimation procedures

To implement the full ML procedure outlined above in numerical computations may be complicated. We therefore in this section present simplified, stepwise estimation procedures. The procedures we present for estimating the covariance matrices $\mathbf{\Sigma}^u$ and $\mathbf{\Sigma}^\delta$, in particular, are simpler than solving (19), respectively (22), after having inserted the solution values for $\boldsymbol{\beta}$ (conditionally on $\mathbf{\Sigma}^u$ and $\mathbf{\Sigma}^\delta$) from (20) and (23), respectively. We first describe the estimation of $\boldsymbol{\beta}$, next the estimation of $\mathbf{\Sigma}^u$ and $\mathbf{\Sigma}^\delta$, then the reestimation of $\boldsymbol{\beta}$, and finally the reestimation of $\mathbf{\Sigma}^u$ and $\mathbf{\Sigma}^\delta$.

First step estimation of $\boldsymbol{\beta}$

Consider first the estimation of the expected coefficient vector $\boldsymbol{\beta}$. We start by computing *individual specific* OLS estimators separately for all individuals which have a sufficient number of observations to permit such estimation. This means that in each equation, the number of observations p must exceed the number of coefficients, including the intercept term.[6] Let q denote the lowest value of p for which OLS estimation of all G equations is possible. The estimator of the coefficient vector for individual (ip) (formally treated as fixed) is

$$(24)\quad \widehat{\boldsymbol{\beta}}_{(ip)} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{1(ip)} \\ \vdots \\ \widehat{\boldsymbol{\beta}}_{G(ip)} \end{bmatrix} = [\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}\boldsymbol{X}'_{(ip)}\boldsymbol{y}_{(ip)} = \begin{bmatrix} (\boldsymbol{X}'_{1(ip)}\boldsymbol{X}_{1(ip)})^{-1}\boldsymbol{X}'_{1(ip)}\boldsymbol{y}_{1(ip)} \\ \vdots \\ (\boldsymbol{X}'_{G(ip)}\boldsymbol{X}_{G(ip)})^{-1}\boldsymbol{X}'_{G(ip)}\boldsymbol{y}_{G(ip)} \end{bmatrix}, \qquad i = 1,\dots,N_p;\, p = q,\dots,P.$$

By inserting from (12) and using (14) it follows that the estimator is unbiased, with covariance matrix

$$(25)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)}) = [\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}][\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}.$$

An estimator of the common expectation of the individual coefficient vectors, $\boldsymbol{\beta}$, *based on the observations from the individuals observed p times* is the unweighted sample mean[7] of the individual specific OLS estimators as

$$(26)\qquad \widehat{\boldsymbol{\beta}}_{(p)} = \frac{1}{N_p}\sum_{i=1}^{N_p}\widehat{\boldsymbol{\beta}}_{(ip)} = \frac{1}{N_p}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{y}_{(ip)}], \qquad p = q,\dots,P.$$

From (25) it follows, since all $\widehat{\boldsymbol{\beta}}_{(ip)}$'s are uncorrelated, that

$$(27)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{(p)}) = \frac{1}{N_p^2}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}][\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}.$$

[6] We here neglect possible equality constraints between different equations.

[7] Estimators constructed from weighted means will be considered below.

An estimator of the common expectation of the individual coefficient vectors, $\boldsymbol{\beta}$, *based on observations of all individuals observed at least q times* can be obtained as the unweighted sample mean[8] of all the $N' = \sum_{p=q}^{P} N_p$ individual specific estimators, *i.e.*,

$$(28) \qquad \widehat{\boldsymbol{\beta}} = \frac{1}{N'} \sum_{p=q}^{P} \sum_{i=1}^{N_p} \widehat{\boldsymbol{\beta}}_{(ip)} = \frac{1}{N'} \sum_{p=q}^{P} \sum_{i=1}^{N_p} [\boldsymbol{X}'_{(ip)} \boldsymbol{X}_{(ip)}]^{-1} [\boldsymbol{X}'_{(ip)} \boldsymbol{y}_{(ip)}].$$

From (25) it follows that

$$(29) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}) = \frac{1}{(N')^2} \sum_{p=q}^{P} \sum_{i=1}^{N_p} [\boldsymbol{X}'_{(ip)} \boldsymbol{X}_{(ip)}]^{-1} [\boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}^{-1}_{(ip)} \boldsymbol{X}_{(ip)}] [\boldsymbol{X}'_{(ip)} \boldsymbol{X}_{(ip)}]^{-1}.$$

This is our simplest way of estimating the expected coefficient vector and its covariance matrix conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$.

First step estimation of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$

Consider next the estimation of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$. We construct the $(Gp \times 1)$ *OLS residual vector* corresponding to $\boldsymbol{u}_{(ip)}$,

$$\widehat{\boldsymbol{u}}_{(ip)} = \begin{bmatrix} \widehat{\boldsymbol{u}}_{1(ip)} \\ \vdots \\ \widehat{\boldsymbol{u}}_{G(ip)} \end{bmatrix} = \boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)} \widehat{\boldsymbol{\beta}}_{(ip)},$$

and rearrange it into the $(G \times p)$ matrix

$$\widehat{\boldsymbol{U}}_{(ip)} = \begin{bmatrix} \widehat{\boldsymbol{u}}'_{1(ip)} \\ \vdots \\ \widehat{\boldsymbol{u}}'_{G(ip)} \end{bmatrix}, \qquad \begin{matrix} i = 1, \ldots, N_p, \\ p = q, \ldots, P, \end{matrix}$$

whose element (g, t) is the t'th OLS residual of individual (ip) in the g'th equation. We estimate, *from observations on the individuals observed p times*, the disturbance covariance matrix $\boldsymbol{\Sigma}^u$ by the analogous sample moments in residuals, *i.e.*,

$$(30) \qquad \widehat{\boldsymbol{\Sigma}}^u_{(p)} = \frac{1}{N_p p} \sum_{i=1}^{N_p} \widehat{\boldsymbol{U}}_{(ip)} \widehat{\boldsymbol{U}}'_{(ip)}, \qquad p = q, \ldots, P,$$

and estimate from (24) and (26) the covariance matrices of the random coefficient vector by its empirical counterparts, *i.e.*,[9]

$$(31) \qquad \widehat{\boldsymbol{\Sigma}}^\delta_{(p)} = \frac{1}{N_p} \sum_{i=1}^{N_p} (\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)})(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)})', \qquad p = q, \ldots, P.$$

[8] Alternatively, the $\boldsymbol{\beta}_{(ip)}$'s could have been weighted by p, the number of observations in each group. Other estimators constructed from weighted means will be considered below.

[9] This estimator is consistent if both p and N_p go to infinity and is always positive definite. It is not, however, unbiased in finite samples. Other estimators for formally similar balanced situations exist. See Hsiao (1986, pp. 83 – 84).

Inserting $\widehat{\boldsymbol{\Sigma}}^u_{(p)}$ and $\widehat{\boldsymbol{\Sigma}}^\delta_{(p)}$ into (14), we get the following estimator of the covariance matrix $\boldsymbol{\Omega}_{(ip)}$ *based on the observations of the individuals observed p times*:

$$\widehat{\boldsymbol{\Omega}}_{(ip)p} = \boldsymbol{X}_{(ip)}\widehat{\boldsymbol{\Sigma}}^\delta_{(p)}\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widehat{\boldsymbol{\Sigma}}^u_{(p)}, \quad i = 1, \dots, N_p;\ p = q, \dots, P. \tag{32}$$

This estimator can be inserted into (25) and (27) to give estimators of $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)})$ and $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{(p)})$.

An estimator of $\boldsymbol{\Sigma}^u$ *based on observations from all individuals observed at least q times* can be obtained as

$$\widehat{\boldsymbol{\Sigma}}^u = \frac{1}{n'}\sum_{p=q}^{P}\sum_{i=1}^{N_p}\widehat{\boldsymbol{U}}_{(ip)}\widehat{\boldsymbol{U}}'_{(ip)} = \frac{1}{n'}\sum_{p=q}^{P} N_p p \widehat{\boldsymbol{\Sigma}}^u_{(p)}, \tag{33}$$

where $n' = \sum_{p=q}^{P} N_p p$. The corresponding estimator of the covariance matrices of the coefficients is

$$\widehat{\boldsymbol{\Sigma}}^\delta = \frac{1}{N'}\sum_{p=q}^{P}\sum_{i=1}^{N_p}(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}})'. \tag{34}$$

Inserting $\widehat{\boldsymbol{\Sigma}}^u$ and $\widehat{\boldsymbol{\Sigma}}^\delta$ into (14), we get the following estimator of the individual specific gross disturbance matrix $\boldsymbol{\Omega}_{(ip)}$ *based on all observations*:[10]

$$\widehat{\boldsymbol{\Omega}}_{(ip)} = \boldsymbol{X}_{(ip)}\widehat{\boldsymbol{\Sigma}}^\delta\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widehat{\boldsymbol{\Sigma}}^u, \quad i = 1, \dots, N_p;\ p = 1, \dots, P. \tag{35}$$

This estimator can be inserted into (29) to give an estimator of $\mathsf{V}(\widehat{\boldsymbol{\beta}})$.

Since (26) and (28) imply

$$\begin{aligned}\frac{1}{N'}\sum_{p=q}^{P}\sum_{i=1}^{N_p}(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}})' &= \frac{1}{N'}\sum_{p=q}^{P}\sum_{i=1}^{N_p}(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)})(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)})' \\ &\quad + \frac{1}{N'}\sum_{p=q}^{P} N_p(\widehat{\boldsymbol{\beta}}_{(p)} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(p)} - \widehat{\boldsymbol{\beta}})',\end{aligned}$$

we can rewrite $\widehat{\boldsymbol{\Sigma}}^\delta$ as

$$\widehat{\boldsymbol{\Sigma}}^\delta = \frac{1}{N'}\sum_{p=q}^{P} N_p\widehat{\boldsymbol{\Sigma}}^\delta_{(p)} + \frac{1}{N'}\sum_{p=q}^{P} N_p(\widehat{\boldsymbol{\beta}}_{(p)} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(p)} - \widehat{\boldsymbol{\beta}})'. \tag{36}$$

The interpretation of this equation is that it separates $\widehat{\boldsymbol{\Sigma}}^\delta$ into components representing within group variation in the $\boldsymbol{\beta}$'s (first term) and between group variation (second term).

[10] Note that $\widehat{\boldsymbol{\Sigma}}^u$ and $\widehat{\boldsymbol{\Sigma}}^\delta$ are constructed from observations from individuals observed at least q times, whereas $\widehat{\boldsymbol{\Omega}}_{(ip)}$ is constructed for all individuals.

Second step estimation of $\boldsymbol{\beta}$

When the $\boldsymbol{\Omega}_{(ip)}$'s have been estimated from (35), (asymptotically) more efficient estimators of the expected coefficient vector $\boldsymbol{\beta}$ can be constructed. So far, individual specific *OLS* estimators of the coefficient vector have been our point of departure, cf. (24). One modification is to use individual specific GLS estimators instead. Replace $\widehat{\boldsymbol{\beta}}_{(ip)}$ by[11]

$$(37) \quad \widetilde{\boldsymbol{\beta}}_{(ip)} = [\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}], \qquad i = 1, \ldots, N_p;\ p = q, \ldots, P.$$

By inserting from (12) and using (14) it can be shown that this estimator is unbiased and has covariance matrix

$$(38) \qquad \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)}) = [\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}.$$

This estimator vector is more efficient than $\widehat{\boldsymbol{\beta}}_{(ip)}$ since it can be shown from (25) and (38) that $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)}) - \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})$ is positive definit. A revised estimator of $\boldsymbol{\beta}$ *based on the observations from the individuals observed p times* can then be defined as the unweighted mean of the individual specific GLS estimators

$$(39) \quad \widetilde{\boldsymbol{\beta}}_{(p)} = \frac{1}{N_p}\sum_{i=1}^{N_p}\widetilde{\boldsymbol{\beta}}_{(ip)} = \frac{1}{N_p}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}], \qquad p = q, \ldots, P.$$

From (38) it follows, since all $\widetilde{\boldsymbol{\beta}}_{(ip)}$'s are uncorrelated, that

$$(40) \qquad \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(p)}) = \frac{1}{N_p^2}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}.$$

A revised estimator of $\boldsymbol{\beta}$ *based on observations of all individuals observed at least q times* can be obtained as the unweighted mean[12]

$$(41) \quad \widetilde{\boldsymbol{\beta}} = \frac{1}{N'}\sum_{p=q}^{P}\sum_{i=1}^{N_p}\widetilde{\boldsymbol{\beta}}_{(ip)} = \frac{1}{N'}\sum_{p=q}^{P}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}].$$

From (38) it follows, since all $\widetilde{\boldsymbol{\beta}}_{(ip)}$'s are uncorrelated, that

$$(42) \qquad \mathsf{V}(\widetilde{\boldsymbol{\beta}}) = \frac{1}{(N')^2}\sum_{p=q}^{P}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}.$$

Variants of $\widetilde{\boldsymbol{\beta}}_{(p)}$ and $\widetilde{\boldsymbol{\beta}}$ exist. Instead of using the unweighted means of the individual specific GLS estimators, (39) and (41), we may use *matrix weighted means*. Of particular

[11] We here and in the following proceed as if the $\boldsymbol{\Omega}_{(ip)}$'s are known. In practice, we may either use the estimators $\widehat{\boldsymbol{\Omega}}_{(ip)}$ or estimate these covariance matrices from recomputed GLS residuals, as will be described below.

[12] Alternatively, the $\widetilde{\boldsymbol{\beta}}_{(ip)}$'s could have been weighted by p.

interest is weighting the subestimators by their respective inverse covariance matrices. Our estimator of $\boldsymbol{\beta}$ *based on the observations from the individuals observed p times* then becomes

$$
\begin{aligned}
\boldsymbol{\beta}^*_{(p)} &= \left[\sum_{i=1}^{N_p} \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}\right]^{-1} \left[\sum_{i=1}^{N_p} \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}\widetilde{\boldsymbol{\beta}}_{(ip)}\right] \\
&= \left[\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}\right]^{-1} \left[\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}\right], \quad p = q, \dots, P.
\end{aligned} \tag{43}
$$

This coincides with the strict GLS estimator for group p, $\widehat{\boldsymbol{\beta}}^{GLS}_{(p)}$, given in (20) as the solution to the ML problem for group p conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$. From (38) it follows, since all $\widetilde{\boldsymbol{\beta}}_{(ip)}$'s are uncorrelated, that

$$
\mathsf{V}(\boldsymbol{\beta}^*_{(p)}) = \left[\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}\right]^{-1}. \tag{44}
$$

Since $\boldsymbol{\beta}^*_{(p)}$ is the strict GLS estimator for group p, we know that it is more efficient than $\widetilde{\boldsymbol{\beta}}_{(p)}$ because $\mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(p)}) - \mathsf{V}(\boldsymbol{\beta}^*_{(p)})$ is positive definit. The corresponding estimator of $\boldsymbol{\beta}$ *based on observations of all groups of individuals observed at least q times* is

$$
\begin{aligned}
\boldsymbol{\beta}^* &= \left[\sum_{p=q}^{P}\sum_{i=1}^{N_p} \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}\right]^{-1} \left[\sum_{p=q}^{P}\sum_{i=1}^{N_p} \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}\widetilde{\boldsymbol{\beta}}_{(ip)}\right] \\
&= \left[\sum_{p=q}^{P}\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}\right]^{-1} \left[\sum_{p=q}^{P}\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}\right],
\end{aligned} \tag{45}
$$

which coincides with $\widehat{\boldsymbol{\beta}}^{GLS}$, given in (23) as the solution to the ML problem for all individuals conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$. From (38) it follows, since all $\widetilde{\boldsymbol{\beta}}_{(ip)}$'s are uncorrelated, that

$$
\mathsf{V}(\boldsymbol{\beta}^*) = \left[\sum_{p=q}^{P}\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}\right]^{-1}. \tag{46}
$$

Since $\boldsymbol{\beta}^*$ is the strict GLS estimator, it is more efficient than $\widetilde{\boldsymbol{\beta}}$ because we know that $\mathsf{V}(\widetilde{\boldsymbol{\beta}}) - \mathsf{V}(\boldsymbol{\beta}^*)$ is positive definit.

Second step estimation of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$

The second step estimators of the coefficient vector can be used to revise the estimators of the disturbance covariance matrices and the covariance matrices of the random coefficients obtained in the first step. We construct the $(Gp \times 1)$ residual vector corresponding to $\boldsymbol{u}_{(ip)}$ from

$$\widetilde{\boldsymbol{u}}_{(ip)} = \begin{bmatrix} \widetilde{\boldsymbol{u}}_{1(ip)} \\ \vdots \\ \widetilde{\boldsymbol{u}}_{G(ip)} \end{bmatrix} = \boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\widetilde{\boldsymbol{\beta}}_{(ip)}$$

and rearrange it into the $(G \times p)$ matrix

$$\widetilde{\boldsymbol{U}}_{(ip)} = \begin{bmatrix} \widetilde{\boldsymbol{u}}'_{1(ip)} \\ \vdots \\ \widetilde{\boldsymbol{u}}'_{G(ip)} \end{bmatrix}.$$

The second step estimator of $\boldsymbol{\Sigma}^u$ for group p is

$$\widetilde{\boldsymbol{\Sigma}}^u_{(p)} = \frac{1}{N_p p} \sum_{i=1}^{N_p} \widetilde{\boldsymbol{U}}_{(ip)} \widetilde{\boldsymbol{U}}'_{(ip)}, \qquad p = q, \ldots, P, \tag{47}$$

and the corresponding estimator of $\boldsymbol{\Sigma}^\delta$, when we use $\boldsymbol{\beta}^*_{(p)}$ as the group specific estimator of $\boldsymbol{\beta}$[13] is

$$\widetilde{\boldsymbol{\Sigma}}^\delta_{(p)} = \frac{1}{N_p} \sum_{i=1}^{N_p} (\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*_{(p)})(\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*_{(p)})', \qquad p = q, \ldots, P. \tag{48}$$

Recompute the estimator of $\boldsymbol{\Omega}_{(ip)}$ based on the subpanel for group p by [cf. (32)]

$$\widetilde{\boldsymbol{\Omega}}_{(ip)p} = \boldsymbol{X}_{(ip)} \widetilde{\boldsymbol{\Sigma}}^\delta_{(p)} \boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widetilde{\boldsymbol{\Sigma}}^u_{(p)}, \quad i = 1, \ldots, N_p; \ p = q, \ldots, P. \tag{49}$$

The second step estimator of $\boldsymbol{\Sigma}^u$ based on the complete data set is

$$\widetilde{\boldsymbol{\Sigma}}^u = \frac{1}{n'} \sum_{p=q}^{P} \sum_{i=1}^{N_p} \widetilde{\boldsymbol{U}}_{(ip)} \widetilde{\boldsymbol{U}}'_{(ip)}, \tag{50}$$

and the corresponding estimator of $\boldsymbol{\Sigma}^\delta$, when we use $\boldsymbol{\beta}^*$ as the overall estimator of $\boldsymbol{\beta}$[14] is

$$\widetilde{\boldsymbol{\Sigma}}^\delta = \frac{1}{N'} \sum_{p=q}^{P} \sum_{i=1}^{N_p} (\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*)(\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*)'. \tag{51}$$

Recompute the overall estimator of $\boldsymbol{\Omega}_{(ip)}$ by [cf. (35)]

$$\widetilde{\boldsymbol{\Omega}}_{(ip)} = \boldsymbol{X}_{(ip)} \widetilde{\boldsymbol{\Sigma}}^\delta \boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widetilde{\boldsymbol{\Sigma}}^u, \quad i = 1, \ldots, N_p; \ p = 1, \ldots, P. \tag{52}$$

An iterative application of this four-step sequence of conditional estimation problems will be presented in the final section, as a summing-up of the paper.

[13] Alternatively, the less efficient estimator $\widetilde{\boldsymbol{\beta}}_{(p)}$ may be used.

[14] Alternatively, the less efficient estimator $\widetilde{\boldsymbol{\beta}}$ may be used.

5 Summing up: A stepwise, modified ML algorithm

Modified, stepwise ML estimation algorithms for $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}^u$, and $\boldsymbol{\Sigma}^\delta$ can be constructed from the procedures described in Section 4. In this final section, we present one such algorithm. The procedure can be considered a modified ML algorithm, provided it converges towards a unique solution and is iterated until convergence.

We give the algorithm for both group specific estimation and estimation from the complete data set. Each algorithm has eight steps and can be iterated until convergence.

Estimation algorithm for group p

1(p): Estimate $\boldsymbol{\beta}$ from the p observations from individual (ip) for $i = 1, \ldots, N_p$ $(p \geq q)$, by using (24). Compute the corresponding OLS residuals.

2(p): Compute an estimator of $\boldsymbol{\beta}$ for group p from (26).

3(p): Compute group specific estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ from (30) – (31).

4(p): Compute $\widehat{\boldsymbol{\Omega}}_{(ip)p}$ from (32) for $i = 1, \ldots, N_p$ $(p \geq q)$.

5(p): Insert $\boldsymbol{\Omega}_{(ip)} = \widehat{\boldsymbol{\Omega}}_{(ip)p}$ into (37) and (43) to compute the individual specific and group specific estimators $\widetilde{\boldsymbol{\beta}}_{(ip)}$ and $\boldsymbol{\beta}^*_{(p)}$.

6(p): Recompute group specific estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ from (47) – (48).

7(p): Compute $\widetilde{\boldsymbol{\Omega}}_{(ip)p}$ from (49) for $i = 1, \ldots, N_p$ $(p \geq q)$.

8(p): Insert $\boldsymbol{\Omega}_{(ip)} = \widetilde{\boldsymbol{\Omega}}_{(ip)p}$ into (37) and (43) to recompute $\widetilde{\boldsymbol{\beta}}_{(ip)}$ and $\boldsymbol{\beta}^*_{(p)}$.

Steps 6(p) – 8(p) can be repeated until convergence, according to some criterion.

Estimation algorithm for all groups

1: Estimate $\boldsymbol{\beta}$ from the observations from individual (ip) for $i = 1, \ldots, N_p; p = q, \ldots, P$, by using (24). Compute the corresponding OLS residuals.

2: Compute an estimator of $\boldsymbol{\beta}$ from the complete panel data set from (28).

3: Compute overall estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ from (33) – (34).

4: Compute $\widehat{\boldsymbol{\Omega}}_{(ip)}$ from (35) for $i = 1, \ldots, N_p;\ p = 1, \ldots, P$.

5: Insert $\boldsymbol{\Omega}_{(ip)} = \widehat{\boldsymbol{\Omega}}_{(ip)}$ into (45) to compute $\boldsymbol{\beta}^*$.

6: Recompute overall estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ from (50) – (51).

7: Compute $\widetilde{\mathbf{\Omega}}_{(ip)}$ from (52) for $i = 1, \ldots, N_p;\ p = 1, \ldots, P$.

8: Insert $\mathbf{\Omega}_{(ip)} = \widetilde{\mathbf{\Omega}}_{(ip)}$ into (41) to recompute $\boldsymbol{\beta}^*$.

Steps 6 – 8 can be repeated until convergence, according to some criterion. If this algorithm converges towards a unique solution, it leads to our modified ML estimator. It is substantially easier to implement in numerical calculations than the full ML procedure described in Section 3.

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.

Baltagi, B.H. (1995): *Econometric Analysis of Panel Data.* Chichester: Wiley, 1995.

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

Biørn, E., Lindquist, K.G., and Skjerpen, T. (1998): Random Coefficients and Unbalanced Panels: An Application on Data from Norwegian Chemical Plants. Statistics Norway, Discussion Papers No. 235.

Hsiao, C. (1975): Some Estimation Methods for a Random Coefficient Model. *Econometrica*, 43 (1975), 305 – 325.

Hsiao, C. (1986): *Analysis of Panel Data.* Cambridge: Cambridge University Press, 1986.

Hsiao, C. (1996): Random Coefficients Models. Chapter 5 in: Mátyás, L. and Sevestre, P. (eds.): *The Econometrics of Panel Data. Handbook of the Theory with Applications.* Dordrecht: Kluwer.

Longford, N.T. (1995): Random Coefficient Models. Chapter 10 in: Arminger, G., Clogg, C.C., and Sobel, M.E. (eds.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum Press.

Swamy, P.A.V.B. (1970): Efficient Estimation in a Random Coefficient Regression Model. *Econometrica*, 38 (1970), 311 – 323.

Swamy, P.A.V.B. (1971): *Statistical Inference in Random Coefficient Regression Models.* New York: Springer-Verlag, 1971.

Swamy, P.A.V.B. (1974): Linear Models with Random Coefficients. Chapter 5 in: Zarembka, P. (ed.): *Frontiers in Econometrics.* New York: Academic Press.

Swamy, P.A.V.B. and Mehta, J.S. (1977): Estimation of Linear Models with Time and Cross-Sectionally Varying Coefficients. *Journal of the American Statistical Association*, 72 (1977), 890 – 898.

Verbeek, M. and Nijman, T.E. (1996): Incomplete Panels and Selection Bias. Chapter 18 in: Mátyás, L. and Sevestre, P. (eds.): *The Econometrics of Panel Data. A Handbook of the Theory with Applications.* Dordrecht: Kluwer.

Wansbeek, T. and Kapteyn, A. (1982): A Class of Decompositions of the Variance-Covariance Matrix of a Generalized Error Components Model. *Econometrica*, 50 (1982), 713 – 724.

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