MEMORANDUM

No 6/2001

HOW IS GENERALIZED LEAST SQUARES RELATED TO WITHIN AND BETWEEN ESTIMATORS IN 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.oekonomi.uio.no/ 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 05	By Atle Seierstad: NECESSARY CONDITIONS AND SUFFICIENT CONDITIONS FOR OPTIMAL CONTROL OF PIECEWISE DETERMINISTIC CONTROL SYSTEMS. 39 p.
No 04	By Pål Longva: Out-Migration of Immigrants: Implications for Assimilation Analysis. 48 p.
No 03	By Øystein Kravdal: The Importance of Education for Fertility in Sub-Saharan Africa is Substantially Underestimated When Community Effects are Ignored. 33 p.
No 02	By Geir B. Asheim and Martin L. Weitzman: DOES NNP GROWTH INDICATE WELFARE IMPROVEMENT? 8 p.
No 01	By Tore Schweder and Nils Lid Hjort: Confidence and Likelihood. 35 p.
No 43	By Mads Greaker: Strategic Environmental Policy when the Governments are Threatened by Relocation. 22 p.
No 42	By Øystein Kravdal: The Impact of Individual and Aggregate Unemployment on Fertility in Norway. 36 p.
No 41	By Michael Hoel and Erik Magnus Sæther: Private health care as a supplement to a public health system with waiting time for treatment: 47 p.
No 40	By Geir B. Asheim, Anne Wenche Emblem and Tore Nilssen: Health Insurance: Treatment vs. Compensation. 15 p.
No 39	By Diderik Lund and Tore Nilssen: Cream Skimming, Dregs Skimming, and Pooling: On the Dynamics of Competitive Screening. 21 p.

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

HOW IS GENERALIZED LEAST SQUARES RELATED TO WITHIN AND BETWEEN ESTIMATORS IN UNBALANCED PANEL DATA ? *)

by

ERIK BIØRN

ABSTRACT

For a random effects regression model with unbalanced panel data, we demonstrate that the Generalized Least Squares (GLS) estimator can be expressed as a (matrix) weighted average of estimators which utilize the within individual and the between individual variation in the data set. We thus generalize a relationship familiar for balanced panel data. Specific attention must be given to the intercept of the regression. We also define an estimator containing the GLS, the within individual, and the between individual estimators for balanced and unbalanced data as special cases.

Keywords: Panel Data. Unbalanced panels. Missing observations. Random effects. Generalized Least Squares. Within estimation. Between estimation

JEL classification: C13, C23

---

*) I thank Terje Skjerpen for valuable comments.

1 Introduction

It is well known from textbook expositions of fixed and random effects regression models with balanced panel data that the Ordinary (OLS) and the Generalized Least Squares (GLS) estimators of the coefficient vector can be interpreted as (matrix) weighted averages of the estimators which utilize only the within individual and only the between individual variation in the data set, often denoted as within and between estimators [see Maddala (1977, chapter 14-3) and Hsiao (1986, section 3.3.2)]. Unbalanced situations, however, are more common in practice than balanced ones, in particular when using micro data, due to entry or exit of respondents, non-response, rotation designs, etc. Therefore, the interest of this weighting relationship from a practical point of view is somewhat limited. There exists a growing literature on GLS estimators of random individual effects models in unbalanced situations [see, *e.g.*, Biørn (1981) and Baltagi (1985)]. The question of whether, and possibly how, this estimator can be related to estimators which can be interpreted as within and between estimators has not been addressed in this literature.

Our focus in this note is on the latter question. We demonstrate that a weighting relationship for GLS with unbalanced panel data and random individual effects similar to that in the balanced case exists, provided that the within and the between variation in the data are defined in a suitable way. In deriving this estimator, we show that specific attention must be given to the intercept term of the equation. Finally, we present a general, and easily implementable, estimator which contains the GLS, the OLS, the within individual, and the between individual estimators for balanced and unbalanced situations as special cases.

2 Model and estimators

Consider a one-way error components regression model for unbalanced panel data in which individual i ($i = 1, \ldots, N$) is observed in T_i periods (not all equal), and let t denote the observation number (which differs from the calendar period if the starting period of the individuals differ or if gaps occur in the time series of some of them):

$$\begin{aligned} & y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + k + \epsilon_{it}, \quad \epsilon_{it} = \alpha_i + u_{it}, \\ & \alpha_i \sim \mathsf{IID}(0, \sigma_\alpha^2), \quad u_{it} \sim \mathsf{IID}(0, \sigma^2), \qquad i = 1, \ldots, N;\ t = 1, \ldots, T_i, \\ & \alpha_i, u_{it}, \boldsymbol{x}_{it} \ \text{ are independent for all } i, t, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_{it}$ is a (row) vector of regressors, $\boldsymbol{\beta}$ its (column) vector of coefficients, α_i an individual specific random effect, and u_{it} a disturbance. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT_i})'$, $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_N')'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT_i}')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_N')'$, etc., and let $\boldsymbol{I}_m$ be the m

dimensional identity matrix and $\boldsymbol{e}_m$ the $(m \times 1)$ vector of ones. The number of observations, *i.e.*, the number of rows in $\boldsymbol{y}$ and $\boldsymbol{X}$, is $n = \sum_{i=1}^{N} T_i$. Compactly, the model can then be written

$$\begin{aligned} &\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}_n k + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} = [\boldsymbol{e}_{T_1}'\alpha_1, \ldots, \boldsymbol{e}_{T_N}'\alpha_N]' + \boldsymbol{u}, \\ &\mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}_{n,1}, \quad \mathsf{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Omega}, \end{aligned} \tag{2}$$

where

$$\boldsymbol{\Omega} = \mathrm{diag}(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_N), \tag{3}$$

$$\boldsymbol{\Omega}_i = \mathsf{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i') = \sigma_\alpha^2 \boldsymbol{e}_{T_i}\boldsymbol{e}_{T_i}' + \sigma^2 \boldsymbol{I}_{T_i} = \sigma^2 \boldsymbol{K}_{T_i} + (\sigma^2 + T_i\sigma_\alpha^2)\boldsymbol{J}_{T_i}, \qquad i = 1, \ldots, N, \tag{4}$$

'diag' denoting a block-diagonal matrix, $\boldsymbol{J}_m = (\boldsymbol{e}_m\boldsymbol{e}_m')/m$, and $\boldsymbol{K}_m = \boldsymbol{I}_m - \boldsymbol{J}_m$, $m = 1, 2, \ldots$. Since the latter two matrices are idempotent and have orthogonal columns, we simply have

$$\boldsymbol{\Omega}_i^{-1} = \frac{1}{\sigma^2}(\boldsymbol{K}_{T_i} + \theta_i \boldsymbol{J}_{T_i}), \tag{5}$$

where

$$\theta_i = \frac{\sigma^2}{\sigma^2 + T_i\sigma_\alpha^2}, \qquad i = 1, \ldots, N. \tag{6}$$

We use the following notation for the within individual, the between individual, and the total covariation in arbitrary matrices, $\boldsymbol{Z}$ and $\boldsymbol{Q}$, constructed in the same way as $\boldsymbol{X}$ above:

$$\begin{aligned} \boldsymbol{W}_{ZQ} &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(\boldsymbol{z}_{it} - \bar{\boldsymbol{z}}_{i\cdot})'(\boldsymbol{q}_{it} - \bar{\boldsymbol{q}}_{i\cdot}) = \boldsymbol{Z}'\mathrm{diag}(\boldsymbol{K}_{T_1}, \ldots, \boldsymbol{K}_{T_N})\boldsymbol{Q}, \\ \boldsymbol{B}_{ZQ} &= \sum_{i=1}^{N} T_i(\bar{\boldsymbol{z}}_{i\cdot} - \bar{\boldsymbol{z}})'(\bar{\boldsymbol{q}}_{i\cdot} - \bar{\boldsymbol{q}}) = \boldsymbol{Z}'\mathrm{diag}(\boldsymbol{J}_{T_1}, \ldots, \boldsymbol{J}_{T_N})\boldsymbol{Q} - \boldsymbol{Z}'\boldsymbol{J}_n\boldsymbol{Q}, \\ \boldsymbol{T}_{ZQ} &= \sum_{i=1}^{N}\sum_{t=1}^{T_i}(\boldsymbol{z}_{it} - \bar{\boldsymbol{z}})'(\boldsymbol{q}_{it} - \bar{\boldsymbol{q}}) = \boldsymbol{W}_{ZQ} + \boldsymbol{B}_{ZQ} = \boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{J}_n)\boldsymbol{Q}, \end{aligned}$$

where $\bar{\boldsymbol{z}}_{i\cdot} = T_i^{-1}\sum_{t=1}^{T_i}\boldsymbol{z}_{it}$ and $\bar{\boldsymbol{z}} = n^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\boldsymbol{z}_{it} = n^{-1}\sum_{i=1}^{N} T_i\bar{\boldsymbol{z}}_{i\cdot}$.

Let $\widetilde{\boldsymbol{X}}_i = (\boldsymbol{X}_i \vdots \boldsymbol{e}_{T_i})$ and $\widetilde{\boldsymbol{X}} = (\widetilde{\boldsymbol{X}}_1', \ldots, \widetilde{\boldsymbol{X}}_N')'$. In the following we do not, however, include the intercept term and the ones attached to it in the coefficient vectors and regressor matrices, as in, *e.g.*, Baltagi (1995, section 9.2), but specify them explicitly in the formulae. This is essential in defining between estimators and decomposing the GLS estimator into within and between estimators for the unbalanced case.

The OLS and GLS estimators of $(\boldsymbol{\beta}'\,k)'$ are

$$\begin{aligned} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{OLS} \\ \widehat{k}_{OLS} \end{bmatrix} &= (\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}(\widetilde{\boldsymbol{X}}'\boldsymbol{y}) = \begin{bmatrix} \sum \boldsymbol{X}_i'\boldsymbol{X}_i & \sum \boldsymbol{X}_i'\boldsymbol{e}_{T_i} \\ \sum \boldsymbol{e}_{T_i}'\boldsymbol{X}_i & \sum \boldsymbol{e}_{T_i}'\boldsymbol{e}_{T_i} \end{bmatrix}^{-1} \begin{bmatrix} \sum \boldsymbol{X}_i'\boldsymbol{y}_i \\ \sum \boldsymbol{e}_{T_i}'\boldsymbol{y}_i \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{W}_{XX} + \sum T_i\bar{\boldsymbol{x}}_{i\cdot}'\bar{\boldsymbol{x}}_{i\cdot} & \sum T_i\bar{\boldsymbol{x}}_{i\cdot}' \\ \sum T_i\bar{\boldsymbol{x}}_{i\cdot} & n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{W}_{XY} + \sum T_i\bar{\boldsymbol{x}}_{i\cdot}'\bar{y}_{i\cdot} \\ \sum T_i\bar{y}_{i\cdot} \end{bmatrix}. \end{aligned} \tag{7}$$

and

$$(8)\qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{GLS} \\ \widehat{k}_{GLS} \end{bmatrix} = (\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}(\widetilde{\boldsymbol{X}}'\boldsymbol{y})$$

$$= \begin{bmatrix} \sum \boldsymbol{X}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{X}_i & \sum \boldsymbol{X}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{e}_{T_i} \\ \sum \boldsymbol{e}_{T_i}'\boldsymbol{\Omega}_i^{-1}\boldsymbol{X}_i & \sum \boldsymbol{e}_{T_i}'\boldsymbol{\Omega}_i^{-1}\boldsymbol{e}_{T_i} \end{bmatrix}^{-1} \begin{bmatrix} \sum \boldsymbol{X}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{y}_i \\ \sum \boldsymbol{e}_{T_i}'\boldsymbol{\Omega}_i^{-1}\boldsymbol{y}_i \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{W}_{XX} + \sum \theta_i T_i \bar{\boldsymbol{x}}_{i\cdot}'\bar{\boldsymbol{x}}_{i\cdot} & \sum \theta_i T_i \bar{\boldsymbol{x}}_{i\cdot}' \\ \sum \theta_i T_i \bar{\boldsymbol{x}}_{i\cdot} & \sum \theta_i T_i \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{W}_{XY} + \sum \theta_i T_i \bar{\boldsymbol{x}}_{i\cdot}'\bar{y}_{i\cdot} \\ \sum \theta_i T_i \bar{y}_{i\cdot} \end{bmatrix},$$

respectively, the last equality following from (5). Since the formula for the partitioned inverse [see, *e.g.*, Lütkepohl (1996, section 3.5.3)] implies

$$(9)\qquad \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b}' \\ \boldsymbol{b} & c \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{Q} & -\boldsymbol{Q}\frac{\boldsymbol{b}'}{c} \\ -\frac{\boldsymbol{b}}{c}\boldsymbol{Q} & \frac{\boldsymbol{b}}{c}\boldsymbol{Q}\frac{\boldsymbol{b}'}{c} + \frac{1}{c} \end{bmatrix}, \qquad \boldsymbol{Q} = \left(\boldsymbol{A} - \frac{\boldsymbol{b}'\boldsymbol{b}}{c}\right)^{-1},$$

where $\boldsymbol{A}$ is a symmetric matrix, $\boldsymbol{b}$ a row vector, and c a scalar, (7) and (8) can be written as

$$(10)\qquad \begin{aligned} \widehat{\boldsymbol{\beta}}_{OLS} &= [\boldsymbol{W}_{XX} + \textstyle\sum T_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}})'(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}})]^{-1} \\ &\qquad \times [\boldsymbol{W}_{XY} + \textstyle\sum T_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}})'(\bar{y}_{i\cdot} - \bar{y})], \\ \widehat{k}_{OLS} &= \bar{y} - \bar{\boldsymbol{x}}\widehat{\boldsymbol{\beta}}_{OLS}, \end{aligned}$$

and

$$(11)\qquad \begin{aligned} \widehat{\boldsymbol{\beta}}_{GLS} &= [\boldsymbol{W}_{XX} + \textstyle\sum \theta_i T_i(\bar{\boldsymbol{x}}_{i\cdot} - \tilde{\bar{\boldsymbol{x}}})'(\bar{\boldsymbol{x}}_{i\cdot} - \tilde{\bar{\boldsymbol{x}}})]^{-1} \\ &\qquad \times [\boldsymbol{W}_{XY} + \textstyle\sum \theta_i T_i(\bar{\boldsymbol{x}}_{i\cdot} - \tilde{\bar{\boldsymbol{x}}})'(\bar{y}_{i\cdot} - \tilde{\bar{y}})], \\ \widehat{k}_{GLS} &= \tilde{\bar{y}} - \tilde{\bar{\boldsymbol{x}}}\widehat{\boldsymbol{\beta}}_{GLS}, \end{aligned}$$

where

$$(12)\qquad \bar{\boldsymbol{x}} = \frac{\sum T_i \bar{\boldsymbol{x}}_{i\cdot}}{\sum T_i}, \quad \bar{y} = \frac{\sum T_i \bar{y}_{i\cdot}}{\sum T_i}, \quad \tilde{\bar{\boldsymbol{x}}} = \frac{\sum \theta_i T_i \bar{\boldsymbol{x}}_{i\cdot}}{\sum \theta_i T_i}, \quad \tilde{\bar{y}} = \frac{\sum \theta_i T_i \bar{y}_{i\cdot}}{\sum \theta_i T_i}.$$

Note that the global means occurring in the definitions of the OLS and the GLS estimators differ when T_i depends on i.

The between and within estimators corresponding to $\widehat{\boldsymbol{\beta}}_{OLS}$ and $\widehat{k}_{OLS}$, obtained by running OLS on

$$(13)\qquad \sqrt{T_i}\bar{y}_{i\cdot} = \sqrt{T_i}\bar{\boldsymbol{x}}_{i\cdot}\boldsymbol{\beta} + \sqrt{T_i}k + \sqrt{T_i}\bar{\epsilon}_{i\cdot}, \qquad i = 1, \ldots, N,$$

and on (1) with the $k + \alpha_i$'s considered as N unknown constants, are, respectively,

$$(14)\qquad \begin{aligned} &\widehat{\boldsymbol{\beta}}_B = [\textstyle\sum T_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}})'(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}})]^{-1}[\textstyle\sum T_i(\bar{\boldsymbol{x}}_{i\cdot} - \bar{\boldsymbol{x}})'(\bar{y}_{i\cdot} - \bar{y})] = \boldsymbol{B}_{XX}^{-1}\boldsymbol{B}_{XY}, \\ &\widehat{k}_B = \bar{y} - \bar{\boldsymbol{x}}\widehat{\boldsymbol{\beta}}_B, \end{aligned}$$

and

$$(15)\qquad \widehat{\boldsymbol{\beta}}_W = \boldsymbol{W}_{XX}^{-1}\boldsymbol{W}_{XY}.$$

Note that the disturbances in (13), $\sqrt{T_i}\bar{\epsilon}_{i\cdot}$, are homoskedastic when no individual effects occur ($\sigma_\alpha^2 = 0$), and heteroskedastic otherwise.

3 The relationship between the estimators

We next consider the relationships between the estimators (10), (11), (14), and (15). From (10), (14), and (15) we have

$$\widehat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{W}_{XX} + \boldsymbol{B}_{XX})^{-1}(\boldsymbol{W}_{XX}\widehat{\boldsymbol{\beta}}_W + \boldsymbol{B}_{XX}\widehat{\boldsymbol{\beta}}_B), \tag{16}$$

regardless of whether the panel data set is balanced or unbalanced. In the balanced case, where $T_i = T$ and $\theta_i = \theta = \sigma^2/(\sigma^2 + T\sigma_\alpha^2)$ for all i, it follows from (11), (14) and (15) that

$$\widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{W}_{XX} + \theta\boldsymbol{B}_{XX})^{-1}(\boldsymbol{W}_{XX}\widehat{\boldsymbol{\beta}}_W + \theta\boldsymbol{B}_{XX}\widehat{\boldsymbol{\beta}}_B). \tag{17}$$

We will now derive a relationship for the unbalanced case similar to the latter.

Let v_i $(i = 1, \dots, N)$ be an arbitrary weight for individual i and multiply its equation in individual means by the square root of this weight, which generalizes (13) to

$$\sqrt{v_i}\bar{y}_{i\cdot} = \sqrt{v_i}\bar{\boldsymbol{x}}_{i\cdot}\boldsymbol{\beta} + \sqrt{v_i}k + \sqrt{v_i}\bar{\epsilon}_{i\cdot}, \qquad i = 1, \dots, N. \tag{18}$$

Running OLS on this equation, we obtain *generalized* between estimators of $\boldsymbol{\beta}$ and k

$$\begin{bmatrix} \widetilde{\boldsymbol{\beta}}_B(\boldsymbol{v}) \\ \widetilde{k}_B(\boldsymbol{v}) \end{bmatrix} = \begin{bmatrix} \sum v_i \bar{\boldsymbol{x}}_{i\cdot}'\bar{\boldsymbol{x}}_{i\cdot} & \sum v_i \bar{\boldsymbol{x}}_{i\cdot}' \\ \sum v_i \bar{\boldsymbol{x}}_{i\cdot} & \sum v_i \end{bmatrix}^{-1} \begin{bmatrix} \sum v_i \bar{\boldsymbol{x}}_{i\cdot}'\bar{y}_{i\cdot} \\ \sum v_i \bar{y}_{i\cdot} \end{bmatrix}, \tag{19}$$

where $\boldsymbol{v} = (v_1, \dots, v_N)$, which, when we again use (9), leads to

$$\begin{aligned} \widetilde{\boldsymbol{\beta}}_B &= \widetilde{\boldsymbol{\beta}}_B(\boldsymbol{v}) = \left[\sum v_i(\bar{\boldsymbol{x}}_{i\cdot} - \widetilde{\boldsymbol{x}}(\boldsymbol{v}))'(\bar{\boldsymbol{x}}_{i\cdot} - \widetilde{\boldsymbol{x}}(\boldsymbol{v}))\right]^{-1} \left[\sum v_i(\bar{\boldsymbol{x}}_{i\cdot} - \widetilde{\boldsymbol{x}}(\boldsymbol{v}))'(\bar{y}_{i\cdot} - \widetilde{y}(\boldsymbol{v}))\right], \\ \widetilde{k}_B &= \widetilde{k}_B(\boldsymbol{v}) = \widetilde{y}(\boldsymbol{v}) - \widetilde{\boldsymbol{x}}(\boldsymbol{v})\widetilde{\boldsymbol{\beta}}_B(\boldsymbol{v}), \end{aligned} \tag{20}$$

where

$$\widetilde{\boldsymbol{x}}(\boldsymbol{v}) = \frac{\sum v_i \bar{\boldsymbol{x}}_{i\cdot}}{\sum v_i}, \qquad \widetilde{y}(\boldsymbol{v}) = \frac{\sum v_i \bar{y}_{i\cdot}}{\sum v_i}. \tag{21}$$

This brings us to the main result in this note: The estimators $\widehat{\boldsymbol{\beta}}_{OLS}$, $\widehat{\boldsymbol{\beta}}_W$ $\widehat{\boldsymbol{\beta}}_B$, $\widetilde{\boldsymbol{\beta}}_B$, and $\widehat{\boldsymbol{\beta}}_{GLS}$ are all matrix weighted means of $\widehat{\boldsymbol{\beta}}_W$ and $\widetilde{\boldsymbol{\beta}}_B(\boldsymbol{v})$ for a suitable choice of $\boldsymbol{v}$, since they all belong to the class

$$\widehat{\boldsymbol{\beta}}(\lambda_W, \lambda_B, \boldsymbol{v}) = [\lambda_W \boldsymbol{W}_{XX} + \lambda_B \widetilde{\boldsymbol{B}}_{XX}(\boldsymbol{v})]^{-1}[\lambda_W \boldsymbol{W}_{XX}\widehat{\boldsymbol{\beta}}_W + \lambda_B \widetilde{\boldsymbol{B}}_{XX}(\boldsymbol{v})\widetilde{\boldsymbol{\beta}}_B(\boldsymbol{v})], \tag{22}$$

where λ_W and λ_B are arbitrary scalar constants and

$$\widetilde{\boldsymbol{B}}_{XX}(\boldsymbol{v}) = \sum v_i[\bar{\boldsymbol{x}}_{i\cdot} - \widetilde{\boldsymbol{x}}(\boldsymbol{v})]'[\bar{\boldsymbol{x}}_{i\cdot} - \widetilde{\boldsymbol{x}}(\boldsymbol{v})].$$

Note that $\widetilde{\boldsymbol{B}}_{XX}(\boldsymbol{v})$ and $\widehat{\boldsymbol{\beta}}(\lambda_W, \lambda_B, \boldsymbol{v})$ are homogeneous in $\boldsymbol{v}$ of degrees one and zero, respectively, while $\widehat{\boldsymbol{\beta}}(\lambda_W, \lambda_B, \boldsymbol{v})$ is homogeneous in (λ_W, λ_B) of degree zero. In particular, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{OLS} &= \widehat{\boldsymbol{\beta}}(1,1,\boldsymbol{T}),\\
\widehat{\boldsymbol{\beta}}_{W} &= \widehat{\boldsymbol{\beta}}(1,0,\boldsymbol{T}) = \widehat{\boldsymbol{\beta}}(1,0,\boldsymbol{\theta T}),\\
\widehat{\boldsymbol{\beta}}_{B} &= \widetilde{\boldsymbol{\beta}}_{B}(\boldsymbol{T}) = \widehat{\boldsymbol{\beta}}(0,1,\boldsymbol{T}),\\
\widetilde{\boldsymbol{\beta}}_{B}(\boldsymbol{\theta T}) &= \widehat{\boldsymbol{\beta}}(0,1,\boldsymbol{\theta T}),\\
\widehat{\boldsymbol{\beta}}_{GLS} &= \widehat{\boldsymbol{\beta}}(1,1,\boldsymbol{\theta T}),
\end{aligned}$$

where $\boldsymbol{T} = (T_1,\ldots,T_N)$ and $\boldsymbol{\theta T} = (\theta_1 T_1,\ldots,\theta_N T_N)$.

In practical applications, the θ_i's have to be estimated, which requires estimation of σ^2 and σ^2_α. This problem, for unbalanced panel data, is discussed in Searle, Casella, and McCulloch (1992, section 3.6) and Biørn (1999, section 3).

4 Conclusion

Our conclusions then are the following:

1. If we define a modified between estimator of $\boldsymbol{\beta}$, (20), by choosing the weight v_i such that the weighted equation in individual means, (18), has disturbances which are homoskedastic with variance σ^2, we obtain the between estimator for the unbalanced panel data set, $\widetilde{\boldsymbol{\beta}}_B(\boldsymbol{\theta T})$. Since $\mathrm{var}(\bar{\epsilon}_{i\cdot}) = \sigma^2_\alpha + \sigma^2/T_i$, this choice inplies $v_i = \theta_i T_i$ $(i = 1,\ldots,N)$.

2. The GLS estimator for the unbalanced case can be interpreted as a matrix weighted mean of $\widehat{\boldsymbol{\beta}}_W$ and $\widetilde{\boldsymbol{\beta}}_B(\boldsymbol{\theta T})$, with weights depending on $\boldsymbol{X}$. Unlike the OLS estimator for the unbalanced case, it cannot, however, in general be interpreted as a matrix weighted mean of $\widehat{\boldsymbol{\beta}}_W$ and $\widehat{\boldsymbol{\beta}}_B$.

3. In the balanced case, in which $\boldsymbol{\theta T} = \theta(T,\ldots,T)$, we have $\widetilde{\boldsymbol{B}}_{XX}(\boldsymbol{\theta T}) = \theta\boldsymbol{B}_{XX}$ and $\widetilde{\boldsymbol{\beta}}_B(\boldsymbol{\theta T}) = \widetilde{\boldsymbol{\beta}}_B(\boldsymbol{T}) = \widehat{\boldsymbol{\beta}}_B$. This gives the familiar decomposition formula (17).

References

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. (1999): Estimating Regression Systems from Unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure. Department of Economics, University of Oslo, Memorandum No. 20/1999.

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

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

Maddala, G.S. (1977): *Econometrics.* Auckland: McGraw-Hill, 1977.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992): *Variance Components.* New York: Wiley, 1992.