MEMORANDUM

No 28/2001

The Efficiency of Panel Data Estimators: GLS Versus Estimators Which Do Not Depend on Variance Components

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 latest 10 Memoranda:

No 27	Qaisar Farooq Akram and Ragnar Nymoen: Employment behaviour in slack and tight labour markets. 34 pp.
No 26	Snorre Kverndokk, Knut Einar Rosendahl and Thomas F. Rutherford: Climate policies and induced technological change: Which to choose the carrot or the stick? 25 pp.
No 25	Olav Bjerkholt: Tracing Haavelmo’s steps from Confluence Analysis to the Probability Approach. 33 pp.
No 24	Sverre A.C. Kittelsen, Guri Galtung Kjæserud and Odd Jarle Kvamme: Errors in Survey Based Quality Evaluation Variables in Efficiency Models of Primary Care Physicians. 23 pp.
No 23	Ove Wolfgang: Eco-Correlation in Acidification Scenarios. 52 pp.
No 22	Øystein Kravdal: The High Fertility of College EducatWomen in Norway: An Artefact of the ’Piecemeal Approach’. 14 pp.
No 21	Yngve Willassen: On the Economics of the Optimal Fallow-Cultivation Cycle. 19 pp.
No 20	Michael Hoel: Allocating greenhouse gas emissions among countries with mobile populations. 14 pp.
No 19	Tore Nilssen and Lars Sørgard: Who Are the Advertisers? 17 pp.
No 18	Tore Nilssen and Lars Sørgard: The TV Industry: Advertising and Programming. 35 pp.

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

THE EFFICIENCY OF PANEL DATA ESTIMATORS: GLS VERSUS ESTIMATORS WHICH DO NOT DEPEND ON VARIANCE COMPONENTS

by

ERIK BIØRN

ABSTRACT

For a balanced panel data regression model with random effects, we discuss the efficiency of the GLS estimator relative to the OLS, the between, and the within estimators. Focus is on how the efficiency responds to changes in (a) the relative variance components and (b) the composition of the regressor covariance matrix which into between and within variation. Both one-way and two-way models are considered. For the one-way, one regressor model, we show that (i) OLS has maximal inefficiency relative to GLS when the within and between individual variation in the regressor account for the same part of the total variation, (ii) the between estimator is always less efficient than the OLS estimator. For the two-way, one regressor model, the between individual (between period) estimator is more efficient than OLS if the between period (between individual) share of the total variation in the regressor and/or the time specific (individual specific) disturbance variance component are sufficiently large. Illustrations relating to marginal budget shares in household consumption are given.

Keywords: Panel Data. Variance Components. Efficiency. Generalized Least Squares. Within estimation. Between estimation

JEL classification: C13, C23

1 Introduction

According to a familiar textbook result in econometrics, the Generalized Least Squares (GLS) estimator of a regression coefficient vector in the case with a non-scalar disturbance covariance matrix is Minimum Variance Linear Unbiased [Greene (2000, section 11.3)]. When applied to a panel data model with random effects, this extension of Gauss-Markov's theorem implies that the GLS is more efficient than the Within (W), the Between (B), and the Ordinary Least Squares (OLS) estimators. Strict GLS, however, is an 'impracticable' method, as it presumes knowledge of the (relative) disturbance variances, which is rarely available, and the efficiency of *Feasible* GLS relative to W, B, and OLS when the variance components are estimated from residuals from a finite (and often small) number of units and/or periods, is not in general known. Some results are available, though [Swamy and Arora (1972), Maddala and Mount (1973), Taylor (1980), and Baltagi (1981)], *inter alia*, based on Monte Carlo simulations when considering alternative ways of estimating the variance components, but they cannot be easily summarized.

In this paper, we investigate, for a balanced panel data model, the most important determinants of the efficiency of strict GLS relative to estimators which do not presume the disturbance variances to be known, or estimated. To the author's knowledge, this issue has not been discussed in the literature, apart from certain special cases. Focus is on how the efficiency of GLS and other panel data estimators responds to changes in (a) the relative variances of the disturbance components, (b) the number of individuals and periods, and (c) the composition of the regressor covariance matrix in between and within variation. Both one-way and two-way models of the latent heterogeneity are considered.

We examine on the one hand the inefficiency of W, B, and OLS relative to GLS, on the other hand the relative efficiency of W, B, and OLS. Since strict GLS may be unobtainable, choosing between W, B, and OLS is of considerable practical interest, not least from the point of view of robustness. First, the consistency of W, unlike the consistency of OLS, is robust to correlation between the latent random effects and the covariate vector [see Hsiao (1986, section 3.4)]. Second, even if OLS often outperforms B in terms of finite sample efficiency, the consistency of the between individual (between period) estimator, unlike the OLS, is robust to errors of measurement in the regressors when the number of periods (individuals) goes to infinity [see Biørn (1996, section 10.2.3)].

The paper is organized as follows. In Section 2, we define four estimators for the one-way random effects model and derive general efficiency results for a regression model of arbitrary dimension. More detailed results are given for the one regressor (or orthogonal regressors) case. We find that OLS is maximally inefficient relative to GLS when half of the variation in the regressor is within individual and the other half is between individual variation. For this model, we show that the between individual estimator is less efficient than OLS, since the latter exploits both between and within variation in the data.

In Section 3 we extend the analysis to the two-way model and present five estimators. We show, for the one regressor case, that OLS is most inefficient relative to GLS (i) in the presence of between period variation in the regressor when less than half of the variation in the regressor is between individual variation, and (ii) in the presence of between individual variation in the regressor when less than half of the variation in the regressor is between period variation. While for the one-way model the between estimator is uniformly less efficient than OLS, this is not generally true for the two-way model. In the one regressor case, we find for the latter that the *between individual (between period)* estimator is more efficient than OLS if the *between period (between individual)* share of the total variation in the regressor and/or the *time specific (individual specific)* part of the disturbance variance exceeds a certain level. The ranking of the strict OLS and the two-way fixed effects OLS (double within) estimator is also discussed. Empirical illustrations relating to marginal budget shares in household consumption are given. Concluding remarks follow in Section 4.

2 Results for the one-way model

2.1 Model and estimators

Assume that we have a panel data set in which N ($\geq$ 2) individuals are observed in T ($\geq$ 2) periods, and consider the model

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

where i indicates individual, t indicates period, $\boldsymbol{x}_{it}$ is a (row) vector of regressors, $\boldsymbol{\beta}$ its (column) vector of coefficients, α_i is an individual specific random effect, and u_{it} is a disturbance. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_N')'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}', \ldots, \boldsymbol{x}_{iT}')'$,$\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_N')'$, etc, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$, and $\boldsymbol{e}_m$ be the $(m \times 1)$ vector of ones. Compactly, the model can then be written

$$(2)\qquad \begin{array}{l} \boldsymbol{y} = \boldsymbol{e}_{NT}k + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} = \boldsymbol{e}_T \otimes \boldsymbol{\alpha} + \boldsymbol{u}, \\ \mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}, \quad \mathsf{V}(\boldsymbol{\epsilon}) = \mathsf{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Omega}, \end{array}$$

where

$$(3)\qquad \boldsymbol{\Omega} = \boldsymbol{I}_N \otimes (\sigma_\alpha^2 \boldsymbol{e}_T\boldsymbol{e}_T' + \sigma^2\boldsymbol{I}_T) = \boldsymbol{I}_N \otimes [\sigma^2\boldsymbol{K}_T + (\sigma^2 + T\sigma_\alpha^2)\boldsymbol{J}_T],$$

$\otimes$ is the Kronecker product operator, $\boldsymbol{J}_m = (\boldsymbol{e}_m \boldsymbol{e}_m')/m$ and $\boldsymbol{K}_m = \boldsymbol{I}_m - \boldsymbol{J}_m$, $m = 1, 2, \ldots$. The latter matrices are idempotent and have orthogonal columns.

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

$$\boldsymbol{W}_{ZQ} = \boldsymbol{Z}'(\boldsymbol{I}_N \otimes \boldsymbol{K}_T)\boldsymbol{Q} = \sum_{i=1}^N \sum_{t=1}^T (\boldsymbol{z}_{it} - \bar{\boldsymbol{z}}_{i\cdot})'(\boldsymbol{q}_{it} - \bar{\boldsymbol{q}}_{i\cdot}),$$
$$\boldsymbol{B}_{ZQ} = \boldsymbol{Z}'(\boldsymbol{K}_N \otimes \boldsymbol{J}_T)\boldsymbol{Q} = T \sum_{i=1}^N (\bar{\boldsymbol{z}}_{i\cdot} - \bar{\boldsymbol{z}})'(\bar{\boldsymbol{q}}_{i\cdot} - \bar{\boldsymbol{q}}),$$
$$\boldsymbol{T}_{ZQ} = \boldsymbol{Z}'(\boldsymbol{I}_{NT} - \boldsymbol{J}_{NT})\boldsymbol{Q} = \sum_{i=1}^N \sum_{t=1}^T (\boldsymbol{z}_{it} - \bar{\boldsymbol{z}})'(\boldsymbol{q}_{it} - \bar{\boldsymbol{q}}) = \boldsymbol{W}_{ZQ} + \boldsymbol{B}_{ZQ},$$

where $\bar{\boldsymbol{z}}_{i\cdot} = T^{-1}\sum_{t=1}^T \boldsymbol{z}_{it}$, $\bar{\boldsymbol{z}} = (NT)^{-1}\sum_{i=1}^N \sum_{t=1}^T \boldsymbol{z}_{it}$, etc. The columns of $\boldsymbol{W}_{ZQ}$ and $\boldsymbol{B}_{ZQ}$ are orthogonal since $\boldsymbol{K}_T$ and $\boldsymbol{J}_T$ have this property. Four estimators of $\boldsymbol{\beta}$, familiar from the panel data literature [see, *e.g.*, Hsiao (1986, chapter 3)], are considered:

$$\text{(4)} \qquad \widehat{\boldsymbol{\beta}}_W = \boldsymbol{W}_{XX}^{-1}\boldsymbol{W}_{XY},$$
$$\text{(5)} \qquad \widehat{\boldsymbol{\beta}}_B = \boldsymbol{B}_{XX}^{-1}\boldsymbol{B}_{XY},$$
$$\text{(6)} \qquad \widehat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{T}_{XX}^{-1}\boldsymbol{T}_{XY} = (\boldsymbol{W}_{XX} + \boldsymbol{B}_{XX})^{-1}(\boldsymbol{W}_{XY} + \boldsymbol{B}_{XY}),$$
$$\text{(7)} \qquad \widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{W}_{XX} + \theta_B \boldsymbol{B}_{XX})^{-1}(\boldsymbol{W}_{XY} + \theta_B \boldsymbol{B}_{XY}),$$

where

$$\text{(8)} \qquad \theta_B = \frac{\sigma^2}{\sigma^2 + T\sigma_\alpha^2}.$$

Here $\widehat{\boldsymbol{\beta}}_W$, the within individual estimator, is the Gauss-Markov estimator if the individual specific effects α_i are treated as fixed and unknown, $\widehat{\boldsymbol{\beta}}_{GLS}$ is the Gauss-Markov estimator if they are treated as random, $\widehat{\boldsymbol{\beta}}_{OLS}$ is the Gauss-Markov estimator when no heterogeneity occurs, and $\widehat{\boldsymbol{\beta}}_B$, the between individual estimator, is the OLS estimator constructed from individual specific means of the observations [see Hsiao (1986, section 3.3.2)]. The full within estimator will only exist when no regressor is time invariant, since otherwise $\boldsymbol{W}_{XX}$ has zero rows and columns.

Let λ_W and λ_B be non-negative scalar constants and define the more general estimator

$$\text{(9)} \qquad \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\lambda_W, \lambda_B) = [\lambda_W \boldsymbol{W}_{XX} + \lambda_B \boldsymbol{B}_{XX}]^{-1}[\lambda_W \boldsymbol{W}_{XY} + \lambda_B \boldsymbol{B}_{XY}].$$

Obviously,

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_W &= \widehat{\boldsymbol{\beta}}(1,0), \\ \widehat{\boldsymbol{\beta}}_B &= \widehat{\boldsymbol{\beta}}(0,1), \\ \widehat{\boldsymbol{\beta}}_{OLS} &= \widehat{\boldsymbol{\beta}}(1,1), \\ \widehat{\boldsymbol{\beta}}_{GLS} &= \widehat{\boldsymbol{\beta}}(1,\theta_B). \end{aligned}$$

Inserting for $\boldsymbol{y}$ from (2) in (9), we obtain

$$\text{(10)} \qquad \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = [\lambda_W \boldsymbol{W}_{XX} + \lambda_B \boldsymbol{B}_{XX}]^{-1}[\lambda_W \boldsymbol{W}_{X\epsilon} + \lambda_B \boldsymbol{B}_{X\epsilon}].$$

2.2 General efficiency results

From (2) and (3) we obtain

$$\mathsf{E}(\boldsymbol{W}_{X\epsilon}\boldsymbol{W}_{X\epsilon}') = \mathsf{E}([\boldsymbol{X}'(\boldsymbol{I}_N\otimes\boldsymbol{K}_T)\boldsymbol{\epsilon}]\,[\boldsymbol{X}'(\boldsymbol{I}_N\otimes\boldsymbol{K}_T)\boldsymbol{\epsilon}]') = \boldsymbol{X}'(\boldsymbol{I}_N\otimes\boldsymbol{K}_T)\,\boldsymbol{\Omega}\,(\boldsymbol{I}_N\otimes\boldsymbol{K}_T)\boldsymbol{X},$$
$$\mathsf{E}(\boldsymbol{B}_{X\epsilon}\boldsymbol{B}_{X\epsilon}') = \mathsf{E}([\boldsymbol{X}'(\boldsymbol{K}_N\otimes\boldsymbol{J}_T)\boldsymbol{\epsilon}]\,[\boldsymbol{X}'(\boldsymbol{K}_N\otimes\boldsymbol{J}_T)\boldsymbol{\epsilon}]') = \boldsymbol{X}'(\boldsymbol{K}_N\otimes\boldsymbol{J}_T)\,\boldsymbol{\Omega}\,(\boldsymbol{K}_N\otimes\boldsymbol{J}_T)\boldsymbol{X},$$

leading to

$$\begin{aligned}\mathsf{E}(\boldsymbol{W}_{X\epsilon}\boldsymbol{W}_{X\epsilon}') &= \sigma^2\boldsymbol{X}'(\boldsymbol{I}_N\otimes\boldsymbol{K}_T)\boldsymbol{X} &= \sigma^2\boldsymbol{W}_{XX},\\ \mathsf{E}(\boldsymbol{B}_{X\epsilon}\boldsymbol{B}_{X\epsilon}') &= (\sigma^2+T\sigma_\alpha^2)\,\boldsymbol{X}'(\boldsymbol{K}_N\otimes\boldsymbol{J}_T)\boldsymbol{X} &= (\sigma^2+T\sigma_\alpha^2)\,\boldsymbol{B}_{XX}.\end{aligned}$$

Combining these expressions with (10) it follows that the covariance matrix of $\widehat{\boldsymbol{\beta}}$ (conditional on $\boldsymbol{X}$), is

$$\begin{aligned}(11)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}) &= \mathsf{E}[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})']\\ &= [\lambda_W\boldsymbol{W}_{XX}+\lambda_B\boldsymbol{B}_{XX}]^{-1}\,[\lambda_W^2\sigma^2\boldsymbol{W}_{XX}+\lambda_B^2(\sigma^2+T\sigma_\alpha^2)\boldsymbol{B}_{XX}]\\ &\qquad\times[\lambda_W\boldsymbol{W}_{XX}+\lambda_B\boldsymbol{B}_{XX}]^{-1}.\end{aligned}$$

This expression can be used to rank unbiased estimators with different (λ_W,λ_B) constellations. Sometimes, one estimator, $\widehat{\boldsymbol{\beta}}_1$, is uniformly superior to another, $\widehat{\boldsymbol{\beta}}_2$, if $\mathsf{V}(\widehat{\boldsymbol{\beta}}_2)-\mathsf{V}(\widehat{\boldsymbol{\beta}}_1)$ is positive definite for any $\boldsymbol{X}$. In particular, we have

$$(12)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_W) = \left[\frac{\boldsymbol{W}_{XX}}{\sigma^2}\right]^{-1},$$

$$(13)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) = \left[\frac{\boldsymbol{B}_{XX}}{\sigma^2+T\sigma_\alpha^2}\right]^{-1},$$

$$\begin{aligned}(14)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) &= \boldsymbol{T}_{XX}^{-1}[\sigma^2\boldsymbol{W}_{XX}+(\sigma^2+T\sigma_\alpha^2)\boldsymbol{B}_{XX}]\boldsymbol{T}_{XX}^{-1}\\ &= \left[\frac{\boldsymbol{T}_{XX}}{\sigma^2}\right]^{-1}+T\sigma_\alpha^2\boldsymbol{T}_{XX}^{-1}\boldsymbol{B}_{XX}\boldsymbol{T}_{XX}^{-1},\end{aligned}$$

$$(15)\qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) = \left[\frac{\boldsymbol{W}_{XX}}{\sigma^2}+\frac{\boldsymbol{B}_{XX}}{\sigma^2+T\sigma_\alpha^2}\right]^{-1}.$$

Using Magnus and Neudecker (1988, chapter 1, Theorem 24), it follows that

$$\begin{aligned}\mathsf{V}(\widehat{\boldsymbol{\beta}}_W)-\mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \text{ is pos. def.} &\iff \frac{\boldsymbol{B}_{XX}}{\sigma^2+T\sigma_\alpha^2} \text{ is pos. def.},\\ \mathsf{V}(\widehat{\boldsymbol{\beta}}_B)-\mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \text{ is pos. def.} &\iff \frac{\boldsymbol{W}_{XX}}{\sigma^2} \text{ is pos. def.}\end{aligned}$$

This implies that $\widehat{\boldsymbol{\beta}}_{GLS}$ is strictly more efficient than both $\widehat{\boldsymbol{\beta}}_W$ and $\widehat{\boldsymbol{\beta}}_B$ when both $\boldsymbol{B}_{XX}$ and $\boldsymbol{W}_{XX}$ are positive definite and σ^2 and σ_α^2 are positive and finite and T is finite. Furthermore, we can write $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})-\mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS})$ as a matrix product which can be shown [for instance by using results in Horn and Johnson (1985, section 7.6)] to be positive definite, in agreement with Gauss-Markov's theorem.

2.3 GLS versus OLS: The one regressor case

Consider the relative efficiency of OLS and GLS in the one regressor case, $K = 1$. The results we derive below are not strictly confined to this case, however; they are also valid, for each regression coefficient, if all regressors are orthogonal so that $\boldsymbol{T}_{XX}$, $\boldsymbol{B}_{XX}$, and $\boldsymbol{W}_{XX}$ are diagonal matrices. Let, for $K = 1$, $b = \boldsymbol{B}_{XX}/\boldsymbol{T}_{XX}$, *i.e.*, the share of the total variation in the regressor (or a typical regressor, under orthogonality) which is between individual variation. It then follows from (8), (14), and (15) that

$$\begin{aligned} e_{OLS} = \frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS})} &= [\sigma^2(1-b) + (\sigma^2 + T\sigma_\alpha^2)b]\left[\frac{1-b}{\sigma^2} + \frac{b}{\sigma^2 + T\sigma_\alpha^2}\right] \\ &= (1-b)^2 + b(1-b)\left(\theta_B + \frac{1}{\theta_B}\right) + b^2, \end{aligned}$$

i.e.,

$$e_{OLS} = \left[1 + b\left(\theta_B + \frac{1}{\theta_B} - 1\right)\right](1-b) + b^2. \tag{16}$$

Since

$$\begin{aligned} \frac{\partial e_{OLS}}{\partial \theta_B} &= b(1-b)\left(1 - \frac{1}{\theta_B^2}\right), \\ \frac{\partial e_{OLS}}{\partial b} &= (1-\theta_B)\left(\frac{1}{\theta_B} - 1\right)(1-2b), \end{aligned}$$

the efficiency function (16) has the following properties: (i) It is convex in θ_B when $0 < b < 1$ and attains its minimum, one, for $\theta_B = 1$, *i.e.*, $\sigma_\alpha^2 = 0$ (which is obvious since OLS and GLS coincide in this case). (ii) It is concave in b when $\theta_B < 1$ and attains its maximum, $(1/4)(1+\theta_B)(1/\theta_B + 1)$, for $b = \frac{1}{2}$. This means that $\widehat{\boldsymbol{\beta}}_{OLS}$, for given T and σ_α^2 (> 0) is maximally inefficient relative to $\widehat{\boldsymbol{\beta}}_{GLS}$ when half of the variation in the regressor is within individual and the other half is between individual variation. When $b = 0$ or $b = 1$, $e_{OLS} = 1$ for any θ_B.

2.4 The ranking of the within, between, and OLS estimators

We see from (12) – (14) that the ranking of $\widehat{\boldsymbol{\beta}}_B$, $\widehat{\boldsymbol{\beta}}_W$, and $\widehat{\boldsymbol{\beta}}_{OLS}$ depends in the general case on σ^2, $T\sigma_\alpha^2$, $\boldsymbol{B}_{XX}$, and $\boldsymbol{W}_{XX}$. The one regressor (or orthogonal regressors) case is most transparent, and we consider this case specifically. From (8) and (12) – (14) we obtain the following variance ratios when $K=1$:

$$\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_W)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)} = \frac{b\theta_B}{1-b} > (<)\ 1 \iff 1-b < (>) \frac{\theta_B}{1+\theta_B},$$

$$\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_W)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} = \frac{1}{(1-b)\left[1+b\left(\frac{1}{\theta_B}-1\right)\right]} > (<)\ 1 \iff 1-b < (>) \frac{\theta_B}{1-\theta_B},$$

$$\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} = \frac{1}{b\theta_B\left[1+b\left(\frac{1}{\theta_B}-1\right)\right]} > 1 \iff 0<b<1.$$

If $0<\theta_B<1$ and $0<b<1$, then $\widehat{\boldsymbol{\beta}}_B$ is always less efficient than $\widehat{\boldsymbol{\beta}}_{OLS}$, $\widehat{\boldsymbol{\beta}}_{OLS}$ is less efficient than $\widehat{\boldsymbol{\beta}}_{GLS}$, and $\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)/\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})$ is monotonically declining in b. The ranking of $\widehat{\boldsymbol{\beta}}_W$ depends on b.

We find, by inserting from (8), that the relative efficiency of the four estimators depends on the share of the variation in the regressor which is within individual variation, $1-b$, as follows [these inequalities generalize results in Malinvaud (1978, chapter 8.4(ii))]:

$$(17)\qquad \begin{array}{ll} \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) \geq \mathsf{V}(\widehat{\boldsymbol{\beta}}_W) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) & \text{if}\ \ \frac{\sigma^2}{T\sigma_\alpha^2} \leq 1-b < 1, \\ \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) \geq \mathsf{V}(\widehat{\boldsymbol{\beta}}_W) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) & \text{if}\ \ \frac{\sigma^2}{2\sigma^2+T\sigma_\alpha^2} \leq 1-b < \frac{\sigma^2}{T\sigma_\alpha^2}, \\ \mathsf{V}(\widehat{\boldsymbol{\beta}}_W) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) & \text{if}\ \ 0 < 1-b < \frac{\sigma^2}{2\sigma^2+T\sigma_\alpha^2}. \end{array}$$

The larger T or σ_α^2/σ^2 for given b is, the more likely is the first inequality to be satisfied (if, at the limit, $T\to\infty$, then $\theta_B\to\infty$, so that $\widehat{\boldsymbol{\beta}}_W$ and $\widehat{\boldsymbol{\beta}}_{GLS}$ coincide). The larger b, for given T and σ_α^2/σ^2, is, the more likely is the last inequality to be satisfied (if, at the limit, $b\to 1$, then $\widehat{\boldsymbol{\beta}}_B$, $\widehat{\boldsymbol{\beta}}_{OLS}$ and $\widehat{\boldsymbol{\beta}}_{GLS}$ coincide and $\widehat{\boldsymbol{\beta}}_W$ is undefined). From the point of view of robustness, the first inequality in (17) is particularly interesting: If $1-b > \sigma^2/(T\sigma_\alpha^2)$, then $\widehat{\boldsymbol{\beta}}_W$ is not only more efficient than $\widehat{\boldsymbol{\beta}}_{OLS}$, it is also robust to violation of the assumption that α_i and $\boldsymbol{x}_{it}$ are independent and neither depends on variance components. If $1-b<\sigma^2/(T\sigma_\alpha^2)$, there will be a trade-off between efficiency and robustness for these two estimators.

3 Results for the two-way model

We now extend the analysis to the two-way model.

3.1 Model and estimators

Consider the model

$$(18)\qquad \begin{aligned} & y_{it} = k + \boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i + \gamma_t + u_{it}, \\ & \alpha_i \sim \mathsf{IID}(0,\sigma_\alpha^2), \quad \gamma_t \sim \mathsf{IID}(0,\sigma_\gamma^2), \quad u_{it} \sim \mathsf{IID}(0,\sigma^2), \\ & \alpha_i, \gamma_t, u_{it}, \boldsymbol{x}_{it} \ \text{ are independent for all } i,t, \end{aligned} \qquad \begin{aligned} & i = 1,\dots,N, \\ & t = 1,\dots,T, \end{aligned}$$

where γ_t is a period specific random effect and the other symbols have the same interpretation as in Section 2. Compactly, the model can be written as

$$(19)\qquad \begin{aligned} & \boldsymbol{y} = \boldsymbol{e}_{NT}k + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} = (\boldsymbol{\alpha}\otimes\boldsymbol{e}_T) + (\boldsymbol{e}_N\otimes\boldsymbol{\gamma}) + \boldsymbol{u}, \\ & \mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}, \qquad \mathsf{V}(\boldsymbol{\epsilon}) = \boldsymbol{\Omega}, \end{aligned}$$

where $\boldsymbol{\gamma} = (\gamma_1,\dots,\gamma_T)'$ and

$$(20)\quad \begin{aligned} \boldsymbol{\Omega} &= \sigma_\alpha^2 \boldsymbol{I}_N \otimes (\boldsymbol{e}_T\boldsymbol{e}_T') + \sigma_\gamma^2(\boldsymbol{e}_N\boldsymbol{e}_N')\otimes\boldsymbol{I}_T + \sigma^2(\boldsymbol{I}_N\otimes\boldsymbol{I}_T) \\ &= \sigma^2(\boldsymbol{K}_N\otimes\boldsymbol{K}_T) + (\sigma^2+T\sigma_\alpha^2)(\boldsymbol{K}_N\otimes\boldsymbol{J}_T) + (\sigma^2+N\sigma_\gamma^2)(\boldsymbol{J}_N\otimes\boldsymbol{K}_T) \\ &\qquad + (\sigma^2+T\sigma_\alpha^2+N\sigma_\gamma^2)(\boldsymbol{J}_N\otimes\boldsymbol{J}_T), \end{aligned}$$

see Fuller and Battese (1974, section 2).

We define $\boldsymbol{B}_{ZQ}$ and $\boldsymbol{T}_{ZQ}$ as in Section 2 and use the following notation for the residual (*i.e.*, double within) and the between period covariation for arbitrary matrices of panel data, $\boldsymbol{Z}$ and $\boldsymbol{Q}$:

$$\begin{aligned} \boldsymbol{R}_{ZQ} &= \boldsymbol{Z}'(\boldsymbol{K}_N\otimes\boldsymbol{K}_T)\boldsymbol{Q} = \textstyle\sum_{i=1}^N\sum_{t=1}^T(\boldsymbol{z}_{it}-\bar{\boldsymbol{z}}_{i\cdot}-\bar{\boldsymbol{z}}_{\cdot t}+\bar{\boldsymbol{z}})'(\boldsymbol{q}_{it}-\bar{\boldsymbol{q}}_{i\cdot}-\bar{\boldsymbol{q}}_{\cdot t}+\bar{\boldsymbol{q}}), \\ \boldsymbol{C}_{ZQ} &= \boldsymbol{Z}'(\boldsymbol{J}_N\otimes\boldsymbol{K}_T)\boldsymbol{Q} = N\textstyle\sum_{t=1}^T(\bar{\boldsymbol{z}}_{\cdot t}-\bar{\boldsymbol{z}})'(\bar{\boldsymbol{q}}_{\cdot t}-\bar{\boldsymbol{q}}), \end{aligned}$$

where $\bar{\boldsymbol{z}}_{\cdot t} = N^{-1}\sum_{i=1}^N \boldsymbol{z}_{it}$, $\boldsymbol{T}_{ZQ} = \boldsymbol{R}_{ZQ} + \boldsymbol{B}_{ZQ} + \boldsymbol{C}_{ZQ}$, and the columns of $\boldsymbol{R}_{ZQ}$, $\boldsymbol{B}_{ZQ}$, and $\boldsymbol{C}_{ZQ}$ are orthogonal. Now, five estimators of $\boldsymbol{\beta}$ are considered:

$$\begin{aligned} &(21)\qquad \widehat{\boldsymbol{\beta}}_R = \boldsymbol{R}_{XX}^{-1}\boldsymbol{R}_{XY}, \\ &(22)\qquad \widehat{\boldsymbol{\beta}}_B = \boldsymbol{B}_{XX}^{-1}\boldsymbol{B}_{XY}, \\ &(23)\qquad \widehat{\boldsymbol{\beta}}_C = \boldsymbol{C}_{XX}^{-1}\boldsymbol{C}_{XY}, \\ &(24)\qquad \widehat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{T}_{XX}^{-1}\boldsymbol{T}_{XY} = (\boldsymbol{R}_{XX}+\boldsymbol{B}_{XX}+\boldsymbol{C}_{XX})^{-1}(\boldsymbol{R}_{XY}+\boldsymbol{B}_{XY}+\boldsymbol{C}_{XY}), \\ &(25)\qquad \widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{R}_{XX}+\theta_B\boldsymbol{B}_{XX}+\theta_C\boldsymbol{C}_{XX})^{-1}(\boldsymbol{R}_{XY}+\theta_B\boldsymbol{B}_{XY}+\theta_C\boldsymbol{C}_{XY}), \end{aligned}$$

where

$$(26)\qquad \theta_B = \frac{\sigma^2}{\sigma^2+T\sigma_\alpha^2}, \qquad \theta_C = \frac{\sigma^2}{\sigma^2+N\sigma_\gamma^2}.$$

Here $\widehat{\boldsymbol{\beta}}_R$, the residual (double within) estimator, is the Gauss-Markov estimator if the individual specific and period specific effects α_i and γ_t are all treated as fixed, $\widehat{\boldsymbol{\beta}}_{GLS}$ is the

Gauss-Markov estimator if they are all treated as random [see Fuller and Battese (1974, section 3) and Mátyás (1996, section 4.2.2)], $\widehat{\boldsymbol{\beta}}_{OLS}$ is the Gauss-Markov estimator when no heterogeneity occurs, $\widehat{\boldsymbol{\beta}}_B$, the between individual estimator is, as in Section 2, the OLS estimator constructed from individual specific means, and $\widehat{\boldsymbol{\beta}}_C$, the between period estimator, is the symmetric estimator constructed from period specific means. The full residual estimator will only exist when no regressor is individual or time invariant, since otherwise $\boldsymbol{R}_{XX}$ has zero rows and columns.

Let λ_R, λ_B, and λ_C be non-negative scalar constants and consider the more general estimator

$$(27)\qquad \begin{aligned}\widehat{\boldsymbol{\beta}} &= \widehat{\boldsymbol{\beta}}(\lambda_R, \lambda_B, \lambda_C)\\ &= [\lambda_R \boldsymbol{R}_{XX} + \lambda_B \boldsymbol{B}_{XX} + \lambda_C \boldsymbol{C}_{XX}]^{-1}[\lambda_R \boldsymbol{R}_{XY} + \lambda_B \boldsymbol{B}_{XY} + \lambda_C \boldsymbol{C}_{XY}],\end{aligned}$$

which also generalizes (9) since latter corresponds to $\lambda_R = \lambda_C = \lambda_W$. Obviously,

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_R &= \widehat{\boldsymbol{\beta}}(1,0,0),\\ \widehat{\boldsymbol{\beta}}_B &= \widehat{\boldsymbol{\beta}}(0,1,0),\\ \widehat{\boldsymbol{\beta}}_C &= \widehat{\boldsymbol{\beta}}(0,0,1),\\ \widehat{\boldsymbol{\beta}}_{OLS} &= \widehat{\boldsymbol{\beta}}(1,1,1),\\ \widehat{\boldsymbol{\beta}}_{GLS} &= \widehat{\boldsymbol{\beta}}(1,\theta_B,\theta_C).\end{aligned}$$

Inserting for $\boldsymbol{y}$ from (19) in (27), we obtain

$$(28)\qquad \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = [\lambda_R \boldsymbol{R}_{XX} + \lambda_B \boldsymbol{B}_{XX} + \lambda_C \boldsymbol{C}_{XX}]^{-1}[\lambda_R \boldsymbol{R}_{X\epsilon} + \lambda_B \boldsymbol{B}_{X\epsilon} + \lambda_C \boldsymbol{C}_{X\epsilon}].$$

3.2 General efficiency results

From (19) and (20) we obtain

$$\begin{aligned}\mathsf{E}(\boldsymbol{R}_{X\epsilon}\boldsymbol{R}'_{X\epsilon}) &= \mathsf{E}([\boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{K}_T)\boldsymbol{\epsilon}][\boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{K}_T)\boldsymbol{\epsilon}]') = \boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{K}_T)\boldsymbol{\Omega}(\boldsymbol{K}_N \otimes \boldsymbol{K}_T)\boldsymbol{X},\\ \mathsf{E}(\boldsymbol{B}_{X\epsilon}\boldsymbol{B}'_{X\epsilon}) &= \mathsf{E}([\boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{J}_T)\boldsymbol{\epsilon}][\boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{J}_T)\boldsymbol{\epsilon}]') = \boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{J}_T)\boldsymbol{\Omega}(\boldsymbol{K}_N \otimes \boldsymbol{J}_T)\boldsymbol{X},\\ \mathsf{E}(\boldsymbol{C}_{X\epsilon}\boldsymbol{C}'_{X\epsilon}) &= \mathsf{E}([\boldsymbol{X}'(\boldsymbol{J}_N \otimes \boldsymbol{K}_T)\boldsymbol{\epsilon}][\boldsymbol{X}'(\boldsymbol{J}_N \otimes \boldsymbol{K}_T)\boldsymbol{\epsilon}]') = \boldsymbol{X}'(\boldsymbol{J}_N \otimes \boldsymbol{K}_T)\boldsymbol{\Omega}(\boldsymbol{J}_N \otimes \boldsymbol{K}_T)\boldsymbol{X},\end{aligned}$$

leading to

$$\begin{aligned}\mathsf{E}(\boldsymbol{R}_{X\epsilon}\boldsymbol{R}'_{X\epsilon}) &= \sigma^2 \boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{K}_T)\boldsymbol{X} &&= \sigma^2 \boldsymbol{R}_{XX},\\ \mathsf{E}(\boldsymbol{B}_{X\epsilon}\boldsymbol{B}'_{X\epsilon}) &= (\sigma^2 + T\sigma_\alpha^2)\boldsymbol{X}'(\boldsymbol{K}_N \otimes \boldsymbol{J}_T)\boldsymbol{X} &&= (\sigma^2 + T\sigma_\alpha^2)\boldsymbol{B}_{XX},\\ \mathsf{E}(\boldsymbol{C}_{X\epsilon}\boldsymbol{C}'_{X\epsilon}) &= (\sigma^2 + N\sigma_\gamma^2)\boldsymbol{X}'(\boldsymbol{J}_N \otimes \boldsymbol{K}_T)\boldsymbol{X} &&= (\sigma^2 + N\sigma_\gamma^2)\boldsymbol{C}_{XX}.\end{aligned}$$

Combining these expressions with (28) it follows that

$$
\begin{aligned}
(29) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}) &= \mathsf{E}[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'] \\
&= [\lambda_R \boldsymbol{R}_{XX} + \lambda_B \boldsymbol{B}_{XX} + \lambda_C \boldsymbol{C}_{XX}]^{-1} \\
&\qquad \times [\lambda_R^2 \sigma^2 \boldsymbol{R}_{XX} + \lambda_B^2(\sigma^2 + T\sigma_\alpha^2)\boldsymbol{B}_{XX} + \lambda_C^2(\sigma^2 + N\sigma_\gamma^2)\boldsymbol{C}_{XX}] \\
&\qquad\qquad \times [\lambda_R \boldsymbol{R}_{XX} + \lambda_B \boldsymbol{B}_{XX} + \lambda_C \boldsymbol{C}_{XX}]^{-1}.
\end{aligned}
$$

This expression can be used to rank unbiased estimators with different $(\lambda_R, \lambda_B, \lambda_C)$ constellations. Sometimes, one estimator, $\widehat{\boldsymbol{\beta}}_1$, is uniformly superior to another, $\widehat{\boldsymbol{\beta}}_2$, if $\mathsf{V}(\widehat{\boldsymbol{\beta}}_2) - \mathsf{V}(\widehat{\boldsymbol{\beta}}_1)$ is positive definite for any $\boldsymbol{X}$. In particular, we have

$$
\begin{aligned}
(30) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_R) &= \left[\frac{\boldsymbol{R}_{XX}}{\sigma^2}\right]^{-1}, \\
(31) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) &= \left[\frac{\boldsymbol{B}_{XX}}{\sigma^2 + T\sigma_\alpha^2}\right]^{-1}, \\
(32) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_C) &= \left[\frac{\boldsymbol{C}_{XX}}{\sigma^2 + N\sigma_\gamma^2}\right]^{-1}, \\
(33) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) &= \boldsymbol{T}_{XX}^{-1}[\sigma^2 \boldsymbol{R}_{XX} + (\sigma^2 + T\sigma_\alpha^2)\boldsymbol{B}_{XX} + (\sigma^2 + N\sigma_\gamma^2)\boldsymbol{C}_{XX}]\boldsymbol{T}_{XX}^{-1} \\
&= \left[\frac{\boldsymbol{T}_{XX}}{\sigma^2}\right]^{-1} + \boldsymbol{T}_{XX}^{-1}[T\sigma_\alpha^2 \boldsymbol{B}_{XX} + N\sigma_\gamma^2 \boldsymbol{C}_{XX}]\boldsymbol{T}_{XX}^{-1}, \\
(34) \qquad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) &= \left[\frac{\boldsymbol{R}_{XX}}{\sigma^2} + \frac{\boldsymbol{B}_{XX}}{\sigma^2 + T\sigma_\alpha^2} + \frac{\boldsymbol{C}_{XX}}{\sigma^2 + N\sigma_\gamma^2}\right]^{-1}.
\end{aligned}
$$

Using Magnus and Neudecker (1988, chapter 1, Theorem 24), it follows that

$$
\begin{aligned}
\mathsf{V}(\widehat{\boldsymbol{\beta}}_R) - \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \text{ is pos. def.} &\iff \frac{\boldsymbol{B}_{XX}}{\sigma^2 + T\sigma_\alpha^2} + \frac{\boldsymbol{C}_{XX}}{\sigma^2 + N\sigma_\gamma^2} \text{ is pos. def.}, \\
\mathsf{V}(\widehat{\boldsymbol{\beta}}_B) - \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \text{ is pos. def.} &\iff \frac{\boldsymbol{R}_{XX}}{\sigma^2} + \frac{\boldsymbol{C}_{XX}}{\sigma^2 + N\sigma_\gamma^2} \text{ is pos. def.}, \\
\mathsf{V}(\widehat{\boldsymbol{\beta}}_C) - \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \text{ is pos. def.} &\iff \frac{\boldsymbol{R}_{XX}}{\sigma^2} + \frac{\boldsymbol{B}_{XX}}{\sigma^2 + T\sigma_\alpha^2}, \text{ is pos. def.}
\end{aligned}
$$

This implies that $\widehat{\boldsymbol{\beta}}_{GLS}$ is strictly more efficient than both $\widehat{\boldsymbol{\beta}}_R$, $\widehat{\boldsymbol{\beta}}_B$, and $\widehat{\boldsymbol{\beta}}_C$ when both $\boldsymbol{B}_{XX}$, $\boldsymbol{C}_{XX}$, and $\boldsymbol{R}_{XX}$ are positive definite and σ^2, σ_α^2, and σ_γ^2 are positive and finite and T and N are finite. Furthermore, we can write $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) - \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS})$ as a matrix product which can be shown [for instance by using results in Horn and Johnson (1985, section 7.6)] to be positive definite, in agreement with Gauss-Markov's theorem.

3.3 GLS versus OLS: The one regressor case

Consider the relative efficiency of OLS and GLS in the one regressor case, $K = 1$. The results we derive below are not strictly confined to this case, however; they are also valid, for each regression coefficient, if all regressors are orthogonal so that $\boldsymbol{T}_{XX}$, $\boldsymbol{B}_{XX}$, $\boldsymbol{C}_{XX}$, and $\boldsymbol{R}_{XX}$ are diagonal matrices. Let, for $K = 1$, $b = \boldsymbol{B}_{XX}/\boldsymbol{T}_{XX}$ and $c = \boldsymbol{C}_{XX}/\boldsymbol{T}_{XX}$, *i.e.*, the shares of the total variation in the regressor (or a typical regressor, under orthogonality) which are between individual and between period variation, respectively. It then follows from (26), (33), and (34) that

$$\begin{aligned} e_{OLS} &= \frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS})} = [\sigma^2(1-b-c) + (\sigma^2 + T\sigma_\alpha^2)b + (\sigma^2 + N\sigma_\gamma^2)c] \\ &\quad \times \left[\frac{1-b-c}{\sigma^2} + \frac{b}{\sigma^2 + T\sigma_\alpha^2} + \frac{c}{\sigma^2 + N\sigma_\gamma^2}\right] \\ &= (1-b-c)^2 + b^2 + c^2 + bc\left(\frac{\theta_B}{\theta_C} + \frac{\theta_C}{\theta_B}\right) \\ &\quad + \left[b\left(\theta_B + \frac{1}{\theta_B}\right) + c\left(\theta_C + \frac{1}{\theta_C}\right)\right](1-b-c), \end{aligned}$$

i.e.,

$$\begin{aligned} e_{OLS} &= \left[1 + b\left(\theta_B + \frac{1}{\theta_B} - 1\right) + c\left(\theta_C + \frac{1}{\theta_C} - 1\right)\right](1-b-c) \\ &\quad + b^2 + bc\left(\frac{\theta_B}{\theta_C} + \frac{\theta_C}{\theta_B}\right) + c^2. \end{aligned} \tag{35}$$

Not unexpectedly, this efficiency function is more complicated than (16) for the one-way model. In the particular case where $\theta_B = \theta_C = \theta$, *i.e.*, $T\sigma_\alpha^2 = N\sigma_\gamma^2$, we have, however

$$e_{OLS} = \left[1 + (b+c)\left(\theta + \frac{1}{\theta} - 1\right)\right](1-b-c) + (b+c)^2,$$

which depends on θ and $b + c$ only, and resembles (16).

Since, in general,

$$\begin{aligned} \frac{\partial e_{OLS}}{\partial \theta_B} &= b(1-b-c)\left(1 - \frac{1}{\theta_B^2}\right) + \frac{bc}{\theta_C}\left(1 - \frac{\theta_C^2}{\theta_B^2}\right), \\ \frac{\partial e_{OLS}}{\partial \theta_C} &= c(1-b-c)\left(1 - \frac{1}{\theta_C^2}\right) + \frac{bc}{\theta_B}\left(1 - \frac{\theta_B^2}{\theta_C^2}\right), \\ \frac{\partial e_{OLS}}{\partial b} &= (1-\theta_B)\left[(1-2b)\left(\frac{1}{\theta_B} - 1\right) - c(1-\theta_C)\left(\frac{1}{\theta_B} + \frac{1}{\theta_C}\right)\right], \\ \frac{\partial e_{OLS}}{\partial c} &= (1-\theta_C)\left[(1-2c)\left(\frac{1}{\theta_C} - 1\right) - b(1-\theta_B)\left(\frac{1}{\theta_B} + \frac{1}{\theta_C}\right)\right], \end{aligned}$$

the efficiency function (35) has the following properties: If $0 < b < 1$, $0 < c < 1$, $0 < b + c < 1$, then e_{OLS} attains its minimum, one, for $\theta_B = \theta_C = 1$, *i.e.*, $\sigma_\alpha^2 = \sigma_\gamma^2 = 0$ (which is obvious since OLS and GLS coincide in this case). If $b = 0, c = 1$ or $b = 1, c = 0$, then $e_{OLS} = 1$ for any θ_B and θ_C. Furthermore,

$$\begin{aligned}
\theta_C = 1, \theta_B < 1, \quad 0 < b < 1 \quad &\Longrightarrow \quad \frac{\partial e_{OLS}}{\partial \theta_B} < 0, \\
\theta_B = 1, \theta_C < 1, \quad 0 < c < 1 \quad &\Longrightarrow \quad \frac{\partial e_{OLS}}{\partial \theta_C} < 0, \\
\theta_B = \theta_C < 1, \quad 0 < b + c < 1 \quad &\Longrightarrow \quad \frac{\partial e_{OLS}}{\partial \theta_B} < 0, \quad \frac{\partial e_{OLS}}{\partial \theta_C} < 0.
\end{aligned}$$

It also follows that

$$\begin{aligned}
\frac{\partial e_{OLS}}{\partial b} > \ (<)\ 0 &\Longleftrightarrow b < \ (>)\ \frac{1}{2}\left[1 - c\frac{1-\theta_C}{1-\theta_B}\left(1 + \frac{\theta_B}{\theta_C}\right)\right] \quad (\theta_B < 1), \\
\frac{\partial e_{OLS}}{\partial c} > \ (<)\ 0 &\Longleftrightarrow c < \ (>)\ \frac{1}{2}\left[1 - b\frac{1-\theta_B}{1-\theta_C}\left(1 + \frac{\theta_C}{\theta_B}\right)\right] \quad (\theta_C < 1), \\
\frac{\partial^2 e_{OLS}}{\partial b^2} &= -\frac{2}{\theta_B}(1-\theta_B)^2, \\
\frac{\partial^2 e_{OLS}}{\partial c^2} &= -\frac{2}{\theta_C}(1-\theta_C)^2, \\
\frac{\partial^2 e_{OLS}}{\partial b \partial c} &= -\frac{\theta_B + \theta_C}{\theta_B \theta_C}(1-\theta_B)(1-\theta_C).
\end{aligned}$$

The efficiency of OLS relative to GLS therefore has the following properties: (i) When $0 < \theta_B < 1$ and $0 < \theta_C < 1$, then e_{OLS} is strictly concave in b and c. (ii) When $c(1-\theta_C) = 0$, then e_{OLS} attains its maximum for $b = \frac{1}{2}$, and when $0 \leq c(1-\theta_C) \leq [\theta_C(1-\theta_B)]/(\theta_B+\theta_C)$, then e_{OLS} attains its maximum for $b \in (0, \frac{1}{2})$. (iii) When $b(1-\theta_B) = 0$, then e_{OLS} attains its maximum for $c = \frac{1}{2}$, and when $0 \leq b(1-\theta_B) \leq [\theta_B(1-\theta_C)]/(\theta_B+\theta_C)$, then e_{OLS} attains its maximum for $c \in (0, \frac{1}{2})$. In the particular case where $\theta_B = \theta_C < 1$, we have

$$\frac{\partial e_{OLS}}{\partial (b+c)} > \ (<)\ 0 \ \Longleftrightarrow \ b + c < \ (>)\ \frac{1}{2},$$

so that OLS is maximally inefficient when the sum of the between individual and the between period variation in the regressor is half of the total variation.

3.4 The ranking of the within, between, and OLS estimators

We see from (30) – (33) that the ranking of $\widehat{\boldsymbol{\beta}}_B$, $\widehat{\boldsymbol{\beta}}_C$, $\widehat{\boldsymbol{\beta}}_R$, and $\widehat{\boldsymbol{\beta}}_{OLS}$ in the general case is determined by σ^2, $T\sigma_\alpha^2$, $N\sigma_\gamma^2$, $\boldsymbol{B}_{XX}$, $\boldsymbol{C}_{XX}$, and $\boldsymbol{R}_{XX}$. Again, the one regressor (or orthogonal regressors) case is most transparent, and we consider this case specifically.

From (26) and (30) – (33) we obtain the following variance ratios when $K = 1$:

$$
\begin{aligned}
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_R)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)} &= \frac{b\theta_B}{1-b-c},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_R)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_C)} &= \frac{c\theta_C}{1-b-c},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_C)} &= \frac{c\theta_C}{b\theta_B},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_R)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} &= \frac{1}{(1-b-c)\left[1+b\left(\frac{1}{\theta_B}-1\right)+c\left(\frac{1}{\theta_C}-1\right)\right]},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} &= \frac{1}{b\theta_B\left[1+b\left(\frac{1}{\theta_B}-1\right)+c\left(\frac{1}{\theta_C}-1\right)\right]},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_C)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} &= \frac{1}{c\theta_C\left[1+b\left(\frac{1}{\theta_B}-1\right)+c\left(\frac{1}{\theta_C}-1\right)\right]}.
\end{aligned}
$$

Rearranging these expressions, defining $\tau_\alpha^2 = \sigma_\alpha^2/\sigma^2$ and $\tau_\gamma^2 = \sigma_\gamma^2/\sigma^2$, and using (26), we find

$$
\begin{aligned}
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_R)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)} > (<)\, 1 &\iff \frac{1-b-c}{1-c} < (>)\, \frac{1}{2+T\tau_\alpha^2},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_R)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_C)} > (<)\, 1 &\iff \frac{1-b-c}{1-b} < (>)\, \frac{1}{2+N\tau_\gamma^2},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_C)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)} > (<)\, 1 &\iff \frac{b}{c} > (<)\, \frac{1+T\tau_\alpha^2}{1+N\tau_\gamma^2},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_R)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} > (<)\, 1 &\iff 1-b-c < (>)\, \frac{1}{1+bT\tau_\alpha^2+cN\tau_\gamma^2},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_B)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} > (<)\, 1 &\iff b < (>)\, \frac{1+T\tau_\alpha^2}{1+bT\tau_\alpha^2+cN\tau_\gamma^2},\\
\frac{\mathsf{V}(\widehat{\boldsymbol{\beta}}_C)}{\mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})} > (<)\, 1 &\iff c < (>)\, \frac{1+N\tau_\gamma^2}{1+bT\tau_\alpha^2+cN\tau_\gamma^2}.
\end{aligned}
$$

From these expressions, we can rank the five estimators by relative efficiency.

We do not describe the detailed ranking, as there are a substantial number of possible cases and some parameter constellations are more likely to occur than others. In principle, there is a region in the $(b, c, T\tau_\alpha^2, N\tau_\gamma^2)$ space in which $\widehat{\boldsymbol{\beta}}_R$ is superior to $\widehat{\boldsymbol{\beta}}_B$, $\widehat{\boldsymbol{\beta}}_B$ is superior to $\widehat{\boldsymbol{\beta}}_C$, $\widehat{\boldsymbol{\beta}}_R$ is superior to $\widehat{\boldsymbol{\beta}}_{OLS}$, etc. Genuine panel data from individuals, households, or firms often show substantial individual specific heterogeneity, both in the regressor and in the disturbances, and less pronounced period specific heterogeneity, so that b often by far exceeds c, $1-b-c$ is small (but often larger than c), and σ_α^2 exceeds σ_γ^2. Furthermore,

N is often considerably larger than T. Four realistic cases may then be

$$\mathsf{V}(\widehat{\boldsymbol{\beta}}_C) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_R) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \quad \text{if}$$
$$1 < \frac{1}{1-b-c} < 1 + bT\tau_\alpha^2 + cN\tau_\gamma^2 < \frac{1+T\tau_\alpha^2}{b} < \frac{1+N\tau_\gamma^2}{c},$$

$$\mathsf{V}(\widehat{\boldsymbol{\beta}}_C) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_R) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \quad \text{if}$$
$$1 < 1 + bT\tau_\alpha^2 + cN\tau_\gamma^2 < \frac{1}{1-b-c} < \frac{1+T\tau_\alpha^2}{b} < \frac{1+N\tau_\gamma^2}{c},$$

$$\mathsf{V}(\widehat{\boldsymbol{\beta}}_C) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_R) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \quad \text{if}$$
$$1 < 1 + bT\tau_\alpha^2 + cN\tau_\gamma^2 < \frac{1+T\tau_\alpha^2}{b} < \frac{1}{1-b-c} < \frac{1+N\tau_\gamma^2}{c},$$

$$\mathsf{V}(\widehat{\boldsymbol{\beta}}_R) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_C) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS}) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{GLS}) \quad \text{if}$$
$$1 < 1 + bT\tau_\alpha^2 + cN\tau_\gamma^2 < \frac{1+T\tau_\alpha^2}{b} < \frac{1+N\tau_\gamma^2}{c} < \frac{1}{1-b-c}.$$

Let us consider an *empirical illustration* taken from Biørn (1994, Table 8), in which marginal budget shares for 28 disaggregate consumption commodities (exhausting the complete budget) and their variances are estimated from Norwegian household panel data with $K = 1$, $N = 418$ and $T = 2$. The only regressor variable is total expenditure, which corresponds to $\boldsymbol{x}$. Its shares of the total variation which is between individual, between period, and residual variation in this data set are $b = 80.5\,\%$, $c = 4.2\,\%$, and $1 - b - c = 15.3\,\%$, respectively. (Values of b and c of similar magnitudes are often found for logarithms of outputs and inputs in firm data.) The variance components σ^2, σ_α^2 and σ_γ^2 are commodity specific and are estimated consistently from residuals, as explained in Biørn (1994, p. 142). We find from these estimates of the variance components that the first of the above four sets of inequalities is satisfied for one commodity (tobacco), the second for one commodity (fuel and power), and the third for the remaining 26 commodities (including foods, services, housing, durables etc.). It should come as no surprise that the first inequality is the one to hold for tobacco, since this is a strongly addictive commodity whose value of τ_α^2 is substantially larger than for other commodities. For none of the 28 commodities the fourth set of inequalities is satisfied according to the estimated variance components.

For this two-way, unlike the one-way, model it is possible for the between individual estimator to be more efficient than OLS. This will happen when $cN\tau_\gamma^2$ is so large that $b > (1+T\tau_\alpha^2)/(1+bT\tau_\alpha^2+cN\tau_\gamma^2)$ [if $\tau_\gamma^2 = 0$ or $c = 0$, and $b < 1$, we always have $\mathsf{V}(\widehat{\boldsymbol{\beta}}_B) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})$]. Likewise, it is possible for the between period estimator to be more efficient than OLS. This will happen when $bT\tau_\alpha^2$ is so large that $c > (1+N\tau_\gamma^2)/(1+bT\tau_\alpha^2+cN\tau_\gamma^2)$ [if $\tau_\alpha^2 = 0$ or $b = 0$, and $c < 1$, we always have $\mathsf{V}(\widehat{\boldsymbol{\beta}}_C) > \mathsf{V}(\widehat{\boldsymbol{\beta}}_{OLS})$]. Neither of these situations occur in our marginal budget shares example, however. In fact, the between period estimator has by far the lowest estimated efficiency for all the 28 commodities. The OLS estimator is ranked second for all commodities except one (tobacco), for which it is ranked third and the residual estimator is ranked second.

4 Concluding remarks

Although it is well established that the GLS is the optimal estimator of the coefficient vector in random effects panel data regression models, when the model is correctly specified, we can conclude from the results in this paper that both OLS and various between and within estimators may be of interest for practical purposes. First, in many realistic situations, the estimation efficiency may not be much improved by using GLS instead of one of its competitors. Second, GLS depends on variance components which are rarely available and may be estimable only with substantial margins of errors, so that the Feasible GLS may depart substantially from the strict GLS. Other panel data estimators do not require this kind of information, although it is needed in order to estimate their variances correctly. Third, the consistency of GLS, like OLS, may be vulnerable to model specification errors. For instance, the consistency of within estimators is robust to correlation between the random latent effects and the covariate vector, and the consistency of the between individual (between period) estimator is robust to errors of measurement in the regressors when the number of periods (individuals) goes to infinity.

In the paper, we have, for both the one-way and the two-way random effects models, reconsidered on the one hand the efficiency of the GLS over its competitors, on the other hand the mutual efficiency of the OLS and the within and between estimators. A detailed investigation has been done for the one regressor (or the orthogonal regressors) case, in which the efficiency can be characterized by variance ratios. A further examination of the multiple (non-orthogonal) regressor case is left for future research. Of course, the precise ranking of the estimators is indeterminate unless the disturbance variance components are known. Still, our results may give guidelines about which estimator of the coefficient vector to choose when the relative composition of the (co)variation of the regressors into within and between (co)variation is known and we have estimates of disturbance variance component ratios, for instance obtained from OLS residuals, even if they are inaccurate.

Acknowledgements. The results presented in this note had its origin in lecture notes which I prepared for courses in panel data econometrics at the University of Oslo. I thank Terje Skjerpen for valuable comments.

References

Baltagi, B.H. (1981): Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Error Component Model. *Journal of Econometrics*, 17 (1981), 21 – 49.

Biørn, E. (1994): Moment Estimators and the Estimation of Marginal Budget Shares from Household Panel Data. *Structural Change and Economic Dynamics*, 5 (1994), 133 – 154.

Biørn, E. (1996): Panel Data with Measurement Errors. Chapter 10 in: Mátyás, L., and Sevestre, P. (eds.): *The Econometrics of Panel Data. A Handbook of the Theory with Applications*. Dordrecht: Kluwer, 1996.

Fuller, W.A., and Battese, G.E. (1974): Estimation of Linear Models with Crossed-Error Structure. *Journal of Econometrics*, 2 (1974), 67 – 78.

Greene, W.H. (2000): *Econometric Analysis*, Fourth Edition. London: Prentice Hall, 2000.

Horn, R.A., and Johnson, C.R. (1985): *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

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

Maddala, G.S., and Mount, T.D. (1973): A Comparative Study of Alternative Estimators for Variance Components Models Used in Econometric Applications. *Journal of the American Statistical Association*, 68 (1973), 324 – 328.

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

Malinvaud, E. (1978): *Méthodes Statistiques de l'Économétrie*, 3ème Édition. Paris: Dunod, 1978.

Mátyás, L. (1996): Error Components Models. Chapter 4 in: Mátyás, L., and Sevestre, P. (eds.): *The Econometrics of Panel Data. A Handbook of the Theory with Applications*. Dordrecht: Kluwer, 1996.

Swamy, P.A.V.B., and Arora, S.S. (1972): The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models. *Econometrica*, 40 (1972), 261 – 275.

Taylor, W.E. (1980): Small Sample Considerations in Estimation from Panel Data. *Journal of Econometrics*, 13 (1980), 203 – 223.