REGRESSION SYSTEMS WITH RANDOM COEFFICIENTS: ESTIMATION PROCEDURES FOR THE UNBALANCED PANEL DATA CASE

ERIK BIØRN
Department of Economics, University of Oslo,
P.O. Box 1095 Blindern, 0317 Oslo, Norway

E-mail: erik.biorn@econ.uio.no

Version of June 18, 2010

Preliminary – please do not quote!

ABSTRACT: A framework for analyzing panel data characterized by a system of regression equations with random heterogeneity in intercepts and coefficients and unbalanced panel data is considered. A Maximum Likelihood (ML) procedure for joint estimation of all parameters is described. Since its implementation in numerical calculations is complicated, simplified procedures are presented. The simplifications in particular concern the estimation of the covariance matrices of the random coefficients. The application and 'anatomy' of the proposed algorithm for modified ML estimation is illustrated by using panel data for output, inputs and costs for 111 manufacturing firms observed up to 22 years.

KEYWORDS: Panel Data. Unbalanced data. Random Coefficients. Heterogeneity. Regression Systems. Iterated Maximum Likelihood

JEL CLASSIFICATION: C33, C51, C63, D24

ACKNOWLEDGEMENTS: I am grateful to Xuehui Han for excellent assistance in the programming of the estimation routines and to Terje Skjerpen for helpful 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 intercepts of the equations ('fixed' or 'random' effects) only. If the heterogeneity has a more complex form, this approach may lead to inefficient (and sometimes 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 constitutes 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 (2008) for 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 intercept heterogeneity for balanced panels. Biørn (1981), Baltagi (1985), and Wansbeek and Kapteyn (1989) consider a single regression equation with random intercept heterogeneity for *unbalanced* panels. Systems of regression equations for unbalanced panel data with random *intercept* heterogeneity are considered in Biørn (2004). The model under consideration in the present paper extends all the models mentioned above, except that we will allow for heterogeneity across units only. In general, far less has been done for unbalanced than for 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 curtail an originally unbalanced data set to make it balanced. Our setting is characterized by: (i) a system of linear, static regressions equations, (ii) random unit specific heterogeneity in intercepts and slope coefficients, and (iii) unbalanced panel data. The sample selection rules are assumed to be *ignorable*, *i.e.*, the way the units or groups enter or exit is not related to the model's endogenous variables. See Verbeek and Nijman (1996, section 18.2) for an elaboration of this topic.

The rest of the paper proceeds as follows: The model is presented in Section 2, with specific attention to the treatment of equality constraints on coefficients in different equations. Section 3 describes the main stages in Maximum Likelihood (ML) estimation. A basic difficulty in computer implementation stems from the unbalance of the panel in combination with the rather complex way in which the covariance matrices of the random coefficients enter the likelihood function. In Section 4, we consider a simpler, stepwise procedure for estimation of these covariance matrices. Section 5 presents, as a summing-up, a simplified algorithm for modified ML estimation. An illustration based on cost and input data for Norwegian manufacturing, the firm data having time series length up to 22 years, is presented in Section 6.

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 units are observed in at least 1 and at most P periods. In descriptions of unbalanced panel data sets, the observations from a specific unit i is often indexed as $t = 1, \ldots, T_i$, where T_i is the number of observations from unit i [see, *e.g.*, Baltagi (2008, section 9.3)]. Our notation convention is somewhat different, in that the units are arranged in groups, or blocks, according to the number of times they are observed. Let N_p be the number of units observed in p periods (not necessarily the same and not necessarily contiguous), let (ip) index unit i in block 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, quarter etc.).[1] The total number of units 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 block p $(p = 2, \ldots, P)$ can be considered a balanced panel data set with p observations of each of the N_p units, while the data set in block 1 is a cross-section. Rotating panels are special cases, in which a share of the units included in the panel in one period is replaced by other units in the next period [see, *e.g.*, Biørn (1981)].

Two ways of formulating the model will be described, formulation [A] assuming the G equations to contain disjoint sets of coefficients, and formulation [B] in which some

[1] Subscripts denoting the *calendar* period may be attached. This may be convenient for data documentation and in formulating dynamic models, but will not be necessary for the static model considered here. For example, in a data set with $P = 20$, from the years 1981–2000, some units in the $p = 18$ group may be observed in the years 1981–1998, some in 1982–1999, some in 1981–1990 and 1992–1999, etc.

equations have coefficients in common. We first consider [A], next the modifications needed in [B], and then describe a general formulation which includes both.

Case [A]. When *each equation has a 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 unit (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), and let $\boldsymbol{u}_{g(ip)}$ be the $(p \times 1)$ disturbance vector in Eq. g from unit (ip). We represent heterogeneity, for Eq. g, unit (ip), by the *random coefficient vector* $\boldsymbol{\beta}_{g(ip)}$ (including the intercept) as

$$\text{(2.1)} \qquad \boldsymbol{\beta}_{g(ip)} = \boldsymbol{\beta}_g + \boldsymbol{\delta}_{g(ip)}, \qquad g = 1,\dots,G,\ i = 1,\dots,N_p,\ p = 1,\dots,P,$$

where $\boldsymbol{\beta}_g$ is fixed and $\boldsymbol{\delta}_{g(ip)}$ is a random shift vector. We assume

$$\text{(2.2)} \qquad \begin{array}{ll} \mathsf{E}[\boldsymbol{\delta}_{g(ip)}] = \mathbf{0}_{K_g,1},\ \mathsf{E}[\boldsymbol{\delta}_{g(ip)}\boldsymbol{\delta}'_{h(ip)}] = \boldsymbol{\Sigma}^{\delta}_{gh}, \quad \mathsf{E}[\boldsymbol{u}_{g(ip)}] = \mathbf{0}_{p,1},\ \mathsf{E}[\boldsymbol{u}_{g(ip)}\boldsymbol{u}'_{h(ip)}] = \sigma^{u}_{gh}\boldsymbol{I}_p, & \\ \boldsymbol{X}_{g(ip)} \perp \boldsymbol{u}_{g(ip)} \perp \boldsymbol{\delta}_{g(ip)}, & g, h = 1,\dots,G, \end{array}$$

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

$$\text{(2.3)} \qquad \begin{aligned} \boldsymbol{y}_{g(ip)} &= \boldsymbol{X}_{g(ip)}\boldsymbol{\beta}_{g(ip)} + \boldsymbol{u}_{g(ip)} = \boldsymbol{X}_{g(ip)}\boldsymbol{\beta}_g + \boldsymbol{\eta}_{g(ip)}, \\ \boldsymbol{\eta}_{g(ip)} &= \boldsymbol{X}_{g(ip)}\boldsymbol{\delta}_{g(ip)} + \boldsymbol{u}_{g(ip)}, \end{aligned}$$

where we interpret $\boldsymbol{\eta}_{g(ip)}$ as a *gross disturbance vector*, representing both the genuine disturbances and the random coefficient variation. It follows from (2.2) that these gross disturbance vectors are heteroskedastic and independent across units, with[2]

$$\text{(2.4)} \qquad \mathsf{E}[\boldsymbol{\eta}_{g(ip)}] = \mathbf{0}_{p,1}, \qquad \mathsf{E}[\boldsymbol{\eta}_{g(ip)}\boldsymbol{\eta}'_{h(ip)}] = \boldsymbol{X}_{g(ip)}\boldsymbol{\Sigma}^{\delta}_{gh}\boldsymbol{X}'_{h(ip)} + \sigma^{u}_{gh}\boldsymbol{I}_p.$$

Case [B]. When *some coefficients occur in at least two equations* – reflecting for instance cross-equational (symmetry) constraints resulting from micro units' optimizing behaviour – the total number of free coefficients is less than $\sum_{g=1}^{G} K_g$. Such (deterministic) coefficient restrictions are assumed to affect both components of (2.1). We stack, for unit (ip), the $\boldsymbol{y}$s, the $\boldsymbol{u}$s, and the $\boldsymbol{\eta}$s by equations and define

$$\boldsymbol{y}_{(ip)} = [\boldsymbol{y}'_{1(ip)},\dots,\boldsymbol{y}'_{G(ip)}]',\ \boldsymbol{u}_{(ip)} = [\boldsymbol{u}'_{1(ip)},\dots,\boldsymbol{u}'_{G(ip)}]',\ \boldsymbol{\eta}_{(ip)} = [\boldsymbol{\eta}'_{1(ip)},\dots,\boldsymbol{\eta}'_{G(ip)}]'$$

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

Then we can in general rewrite (2.1) as

$$\text{(2.5)} \qquad \boldsymbol{\beta}_{(ip)} = \boldsymbol{\beta} + \boldsymbol{\delta}_{(ip)}, \qquad p = 1,\dots,P,$$

where $\boldsymbol{\beta}_{(ip)}$ is the random $(K \times 1)$ vector containing *all* the coefficients, $\boldsymbol{\beta}$ is a fixed vector and $\boldsymbol{\delta}_{(ip)}$ is its random shift vector. Accordingly, we extend the definition of

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

$\boldsymbol{X}_{g(ip)}$ as the $(p \times K)$ matrix of regressors in the gth equation whose kth column contains the observations on the variable which corresponds to the kth coefficient in $\boldsymbol{\beta}_{(ip)}$ $(k = 1, \dots, K)$. If the latter coefficient does not occur in the gth equation, the kth column of $\boldsymbol{X}_{g(ip)}$ is set to zero. We further replace (2.2) and (2.3) by

$$\text{(2.6)} \qquad \mathsf{E}[\boldsymbol{\delta}_{(ip)}] = \mathbf{0}_{K,1}, \ \mathsf{E}[\boldsymbol{\delta}_{(ip)}\boldsymbol{\delta}'_{(ip)}] = \boldsymbol{\Sigma}^{\delta}, \quad \mathsf{E}[\boldsymbol{u}_{(ip)}] = \mathbf{0}_{Gp,1}, \ \mathsf{E}[\boldsymbol{u}_{(ip)}\boldsymbol{u}'_{(ip)}] = \boldsymbol{I}_p \otimes \boldsymbol{\Sigma}^{u}, \\ \boldsymbol{X}_{(ip)} \perp \boldsymbol{u}_{(ip)} \perp \boldsymbol{\delta}_{(ip)},$$

$$\text{(2.7)} \qquad \boldsymbol{y}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\beta} + \boldsymbol{u}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\beta} + \boldsymbol{\eta}_{(ip)}, \\ \boldsymbol{\eta}_{(ip)} = \boldsymbol{X}_{(ip)}\boldsymbol{\delta}_{(ip)} + \boldsymbol{u}_{(ip)},$$

where

$$\boldsymbol{X}_{(ip)} = [\boldsymbol{X}'_{1(ip)}, \dots, \boldsymbol{X}'_{G(ip)}]',$$

so that (2.4) is generalized to

$$\text{(2.8)} \qquad \mathsf{E}[\boldsymbol{\eta}_{(ip)}] = \mathbf{0}_{Gp,1}, \quad \mathsf{E}[\boldsymbol{\eta}_{(ip)}\boldsymbol{\eta}'_{(ip)}] = \boldsymbol{X}_{(ip)}\boldsymbol{\Sigma}^{\delta}\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \boldsymbol{\Sigma}^{u} = \boldsymbol{\Omega}_{(ip)},$$

where $\otimes$ is the Kronecker product operator and $\boldsymbol{\Omega}_{(ip)}$ is defined by the last equality. This is the general setup, valid for both Case [A] and Case [B].

In case [A], where $K = \sum_{g=1}^{G} K_g$, we specifically have

$$\boldsymbol{X}_{(ip)} = \begin{bmatrix} \boldsymbol{X}_{1(ip)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{X}_{G(ip)} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_G \end{bmatrix}, \ \boldsymbol{\eta}_{(ip)} = \begin{bmatrix} \boldsymbol{\eta}_{1(ip)} \\ \vdots \\ \boldsymbol{\eta}_{G(ip)} \end{bmatrix}, \ \boldsymbol{\Sigma}^{\delta} = \begin{bmatrix} \boldsymbol{\Sigma}^{\delta}_{11} & \cdots & \boldsymbol{\Sigma}^{\delta}_{1G} \\ \vdots & & \vdots \\ \boldsymbol{\Sigma}^{\delta}_{G1} & \cdots & \boldsymbol{\Sigma}^{\delta}_{GG} \end{bmatrix}.$$

In case [B], where $K < \sum_{g=1}^{G} K_g$, we, in contrast, include zeros when defining the submatrices $\boldsymbol{X}_{g(ip)}$ and let $\boldsymbol{X}_{(ip)} = [\boldsymbol{X}'_{1(ip)}, \cdots, \boldsymbol{X}'_{G(ip)}]'$, which is no longer block-diagonal.

3 The Maximum Likelihood problem

We now describe the Maximum Likelihood problem for joint estimation of the coefficients and the disturbance covariance matrices, specified as (2.7)–(2.8), and consider the main stages of its solution. We make the additional assumption that the random components of the coefficients and the disturbances are *normally* distributed:

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

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 (2.8). The log-density function of $(\boldsymbol{y}_{(ip)}|\boldsymbol{X}_{(ip)})$ is

$$\mathcal{L}_{(ip)} = -\tfrac{Gp}{2}\ln(2\pi) - \tfrac{1}{2}\ln|\boldsymbol{\Omega}_{(ip)}| - \tfrac{1}{2}Q_{(ip)},$$

where

$$\text{(3.1)} \qquad Q_{(ip)} = [\boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\boldsymbol{\beta}]'\boldsymbol{\Omega}^{-1}_{(ip)}[\boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\boldsymbol{\beta}] = \boldsymbol{\eta}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{\eta}_{(ip)}.$$

The log-likelihood function of all $\boldsymbol{y}$s conditional on all $\boldsymbol{X}$s for block p and the overall log-likelihood function can be written as, respectively,

$$\text{(3.2)} \quad \mathcal{L}_{(p)} = \textstyle\sum_{i=1}^{N_p} \mathcal{L}_{(ip)} = -\frac{GN_p p}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N_p}\ln|\boldsymbol{\Omega}_{(ip)}| - \frac{1}{2}\sum_{i=1}^{N_p} Q_{(ip)},$$

$$\text{(3.3)} \quad \mathcal{L} = \textstyle\sum_{p=1}^{P} \mathcal{L}_{(p)} = -\frac{Gn}{2}\ln(2\pi) - \frac{1}{2}\sum_{p=1}^{P}\sum_{i=1}^{N_p}\ln|\boldsymbol{\Omega}_{(ip)}| - \frac{1}{2}\sum_{p=1}^{P}\sum_{i=1}^{N_p} Q_{(ip)}.$$

Two Maximum Likelihood (ML) problems are of interest: block-specific estimation and full estimation using observations from all blocks. The ML estimators of $(\boldsymbol{\beta}, \boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta)$ *for block* p are the values that maximize $\mathcal{L}_{(p)}$, the ML estimators based on the complete data set are the values that maximize $\mathcal{L} = \sum_{p=1}^{P} \mathcal{L}_{(p)}$.

Even the block specific problem is more complicated than the ML problem for systems of regression equations with balanced panel data, deterministic coefficients and random intercepts, confer Avery (1977) and Baltagi (1980). The reason is that different units have different $\boldsymbol{\Omega}_{(ip)}$ matrices, depending on $\boldsymbol{X}_{(ip)}$. The estimation problem for the complete unbalanced panel is still more complicated since the $\boldsymbol{y}$, $\boldsymbol{X}$, and $\boldsymbol{\Omega}$ matrices occurring have different number of rows, reflecting the different number of observations of the units. Although the dimensions of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ are the same for all units, the dimensions of $\boldsymbol{X}_{(ip)}$ and $\boldsymbol{I}_p$, and hence of $\boldsymbol{\Omega}_{(ip)}$, differ. We briefly discuss these two problems.

ML ESTIMATION FOR BLOCK p.

We set the derivatives of $\mathcal{L}_{(p)}$ with respect to $\boldsymbol{\beta}, \boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta$ equal to zero and obtain

$$\sum_{i=1}^{N_p} \left(\frac{\partial Q_{(ip)}}{\partial \boldsymbol{\beta}} \right) = \mathbf{0}_{K,1}, \tag{3.4}$$

$$\begin{aligned} \sum_{i=1}^{N_p} \left[\frac{\partial \ln |\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^u} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^u} \right] &= \mathbf{0}_{G,G}, \\ \sum_{i=1}^{N_p} \left[\frac{\partial \ln |\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^\delta} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^\delta} \right] &= \mathbf{0}_{K,K}. \end{aligned} \tag{3.5}$$

These first-order conditions define the solution to the ML problem for block p, each p giving a distinct estimator. Conditions (3.4) coincide with those that solve the GLS problem for $\boldsymbol{\beta}$ for block p, conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, and we find the solution

$$\widehat{\boldsymbol{\beta}}_{(p)}^{GLS} = [\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{X}_{(ip)}]^{-1} [\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{y}_{(ip)}]. \tag{3.6}$$

By inserting $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_{(p)}^{GLS}$ we obtain from (3.1) and (3.2) the concentrated log-likelihood functions for block p. Each function can be maximized with respect to $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ to give the block specific estimators of these covariance matrices.

ML ESTIMATION FOR ALL BLOCKS JOINTLY.

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

$$\sum_{p=1}^{P} \sum_{i=1}^{N_p} \left[\frac{\partial Q_{(ip)}}{\partial \boldsymbol{\beta}} \right] = \mathbf{0}_{K,1}, \tag{3.7}$$

$$\begin{aligned} \sum_{p=1}^{P} \sum_{i=1}^{N_p} \left[\frac{\partial \ln |\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^u} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^u} \right] &= \mathbf{0}_{G,G}, \\ \sum_{p=1}^{P} \sum_{i=1}^{N_p} \left[\frac{\partial \ln |\boldsymbol{\Omega}_{(ip)}|}{\partial \boldsymbol{\Sigma}^\delta} + \frac{\partial Q_{(ip)}}{\partial \boldsymbol{\Sigma}^\delta} \right] &= \mathbf{0}_{K,K}. \end{aligned} \tag{3.8}$$

They define the solution to the overall ML problem. Conditions (3.7) coincide with those that solve the full GLS problem for $\boldsymbol{\beta}$, conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, and we find

$$\widehat{\boldsymbol{\beta}}^{GLS} = [\sum_{p=1}^{P} \sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{X}_{(ip)}]^{-1} [\sum_{p=1}^{P} \sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)} \boldsymbol{\Omega}_{(ip)}^{-1} \boldsymbol{y}_{(ip)}]. \tag{3.9}$$

By inserting $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{GLS}$, we obtain from (3.1) and (3.3) the concentrated log-likelihood function, which can be maximized with respect to $\boldsymbol{\Sigma}^u$, and $\boldsymbol{\Sigma}^\delta$ to give the estimators of these covariance matrices.

4 Simplified estimation procedures

To implement ML as outlined above in computations based on analytical matrix derivatives is difficult. Below we present simplified, stepwise procedures. The procedures for estimating $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, in particular, are simpler than those following by differentiating the concentrated log-likelihood functions corresponding to (3.5) and (3.8). We describe the full procedure as a four-step algorithm, suitable for computer programming, with the following elements:

A. Initial estimation of $\boldsymbol{\beta}_{(ip)}$ and $\boldsymbol{\beta}$.
B. Initial estimation of $\boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta$ from disturbances and coefficient-slacks.
C. Revised estimation of $\boldsymbol{\beta}_{(ip)}$ and $\boldsymbol{\beta}$.
D. Revised estimation of $\boldsymbol{\Sigma}^u, \boldsymbol{\Sigma}^\delta$ from revised disturbances and coefficient-slacks.

A. First step OLS estimation of $\boldsymbol{\beta}$

Consider first the estimation of the expected coefficient vector $\boldsymbol{\beta}$. We start by computing *unit specific* OLS estimators separately for all units for which a sufficient number of observations to permit such estimation exist. This means that in each equation, the number of observations p must exceed the number of coefficients, including the intercept[3] Let q denote the lowest value of p permitting OLS estimation of all G equations. The estimator of the coefficient vector for unit (ip) – when we formally *condition inference on* $\boldsymbol{\beta}_{(ip)}$ – is

$$(4.1) \quad \widehat{\boldsymbol{\beta}}_{(ip)} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{1(ip)} \\ \vdots \\ \widehat{\boldsymbol{\beta}}_{G(ip)} \end{bmatrix} = [\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{y}_{(ip)}] = \begin{bmatrix} [\boldsymbol{X}'_{1(ip)}\boldsymbol{X}_{1(ip)}]^{-1}[\boldsymbol{X}'_{1(ip)}\boldsymbol{y}_{1(ip)}] \\ \vdots \\ [\boldsymbol{X}'_{G(ip)}\boldsymbol{X}_{G(ip)}]^{-1}[\boldsymbol{X}'_{G(ip)}\boldsymbol{y}_{G(ip)}] \end{bmatrix}.$$

Inserting for $\boldsymbol{y}_{(ip)}$ from (2.7) we find that $\widehat{\boldsymbol{\beta}}_{(ip)}$ is unbiased *for the expected coefficient vector* $\boldsymbol{\beta}$ with covariance matrix

$$(4.2) \quad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)}) = [\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}_{(ip)}\boldsymbol{X}_{(ip)}][\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}.$$

Conditional on $\boldsymbol{\beta}_{(ip)}$ it is unbiased with covariance matrix

$$\mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)}|\boldsymbol{\beta}_{(ip)}) = [\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}(\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}^u)\boldsymbol{X}_{(ip)}][\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1},$$

which if $\boldsymbol{\Sigma}^u$ is diagonal can be simplified to

$$(4.3) \quad \mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)}|\boldsymbol{\beta}_{(ip)}) = \begin{bmatrix} \sigma^u_{11}(\boldsymbol{X}'_{1(ip)}\boldsymbol{X}_{1(ip)})^{-1} \\ \vdots \\ \sigma^u_{GG}(\boldsymbol{X}'_{G(ip)}\boldsymbol{X}_{G(ip)})^{-1} \end{bmatrix}.$$

[3]We here neglect possible cross-equational coefficient restrictions.

A first-step estimator of $\boldsymbol{\beta}$ *based on the observations from the units observed p times* is the sample mean of the unit specific OLS estimators as

$$(4.4) \qquad \widehat{\boldsymbol{\beta}}_{(p)} = \tfrac{1}{N_p}\sum_{i=1}^{N_p}\widehat{\boldsymbol{\beta}}_{(ip)} = \tfrac{1}{N_p}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{y}_{(ip)}], \qquad p = q,\dots,P.$$

A corresponding estimator of $\boldsymbol{\beta}$, based on observations from all units observed *at least q times*, is obtained as the unweighted mean of the $\sum_{p=q}^{P} N_p$ unit specific estimators, *i.e.*,

$$(4.5) \qquad \begin{aligned}\widehat{\boldsymbol{\beta}} &= [\textstyle\sum_{p=q}^{P}N_p]^{-1}\sum_{p=q}^{P}\sum_{i=1}^{N_p}\widehat{\boldsymbol{\beta}}_{(ip)}\\ &= [\textstyle\sum_{p=q}^{P}N_p]^{-1}\sum_{p=q}^{P}\sum_{i=1}^{N_p}[\boldsymbol{X}'_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{y}_{(ip)}].\end{aligned}$$

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

Consider next the estimation of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ needed for implementing (4.2). We construct from (4.1) the $(Gp\times 1)$ *OLS residual vector* corresponding to the disturbance vector $\boldsymbol{u}_{(ip)}$ and rearrange it into a $(G\times p)$ matrix $\widehat{\boldsymbol{U}}_{(ip)}$ as follows:

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

Element (g,t) of the matrix $\widehat{\boldsymbol{U}}_{(ip)}$ is residual t of unit (ip) in equation g. From observations on the units observed p times we obtain a *block* p-specific estimate of $\boldsymbol{\Sigma}^u$ by analogous moments in residuals,

$$(4.6) \qquad \widehat{\boldsymbol{\Sigma}}^u_{(p)} = \tfrac{1}{N_p p}\sum_{i=1}^{N_p}\widehat{\boldsymbol{U}}_{(ip)}\widehat{\boldsymbol{U}}'_{(ip)}, \qquad p = q,\dots,P,$$

and, using (4.1) and (4.4), obtain – from the resulting *coefficient-slack* vectors $\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)}$ – a *block* p-specific estimate of the covariance matrix of the coefficient vector, $\boldsymbol{\Sigma}^\delta$, by using its empirical counterpart, *i.e.*,[4]

$$(4.7) \qquad \widehat{\boldsymbol{\Sigma}}^\delta_{(p)} = \tfrac{1}{N_p}\sum_{i=1}^{N_p}(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)})(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}}_{(p)})', \qquad p = q,\dots,P.$$

Inserting $\widehat{\boldsymbol{\Sigma}}^u_{(p)}$ and $\widehat{\boldsymbol{\Sigma}}^\delta_{(p)}$ into the expression for $\boldsymbol{\Omega}_{(ip)}$ given by (2.8), we get the following estimator *based solely on the observations from block p*:

$$(4.8) \qquad \widehat{\boldsymbol{\Omega}}_{(ip)p} = \boldsymbol{X}_{(ip)}\widehat{\boldsymbol{\Sigma}}^\delta_{(p)}\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p\otimes\widehat{\boldsymbol{\Sigma}}^u_{(p)}, \qquad i = 1,\dots,N_p;\ p = q,\dots,P.$$

It can be inserted into (4.2) to give an estimator of $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)})$.

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

$$(4.9) \qquad \widehat{\boldsymbol{\Sigma}}^u = [\textstyle\sum_{p=q}^{P}N_p p]^{-1}\sum_{p=q}^{P}\sum_{i=1}^{N_p}\widehat{\boldsymbol{U}}_{(ip)}\widehat{\boldsymbol{U}}'_{(ip)} = [\sum_{p=q}^{P}N_p p]^{-1}\sum_{p=q}^{P}N_p p\widehat{\boldsymbol{\Sigma}}^u_{(p)},$$

which is a weighted average of the $\widehat{\boldsymbol{\Sigma}}^u_{(p)}$s. The corresponding estimator of $\boldsymbol{\Sigma}^\delta$ is

$$(4.10) \qquad \widehat{\boldsymbol{\Sigma}}^\delta = [\textstyle\sum_{p=q}^{P}N_p]^{-1}\sum_{p=q}^{P}\sum_{i=1}^{N_p}(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(ip)} - \widehat{\boldsymbol{\beta}})',$$

[4]This estimator is positive definite and consistent if both p and N_p go to infinity, although not, however, unbiased in finite samples. Modified estimators for similar balanced situations are considered in Hsiao (2003, pp. 146–147).

which exploits the global variation in the $\widehat{\boldsymbol{\beta}}_{(ip)}$s. Inserting $\widehat{\boldsymbol{\Sigma}}^u$ and $\widehat{\boldsymbol{\Sigma}}^\delta$ in (2.8), we get the following estimator based on all observations:[5]

$$\widehat{\boldsymbol{\Omega}}_{(ip)} = \boldsymbol{X}_{(ip)}\widehat{\boldsymbol{\Sigma}}^{\delta}\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widehat{\boldsymbol{\Sigma}}^{u}, \quad i = 1, \dots, N_p;\ p = 1, \dots, P. \tag{4.11}$$

How is $\widehat{\boldsymbol{\Sigma}}^\delta$ and the $\widehat{\boldsymbol{\Sigma}}^\delta_{(p)}$s related? Since (4.3) and (4.5) imply

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

$\widehat{\boldsymbol{\Sigma}}^\delta$ can be rewritten as

$$\widehat{\boldsymbol{\Sigma}}^{\delta} = [\textstyle\sum_{p=q}^{P} N_p]^{-1} \sum_{p=q}^{P} N_p \widehat{\boldsymbol{\Sigma}}^{\delta}_{(p)} + [\sum_{p=q}^{P} N_p]^{-1} \sum_{p=q}^{P} N_p (\widehat{\boldsymbol{\beta}}_{(p)} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(p)} - \widehat{\boldsymbol{\beta}})', \tag{4.12}$$

showing that it can be interpreted as separated into components representing within- and between-block variation in the $\widehat{\boldsymbol{\beta}}$s. The former is a weighted mean of the block specific estimators, the latter is a positive definite quadratic form.

C. Second step GLS estimation of $\boldsymbol{\beta}$

Once we have estimated the $\boldsymbol{\Omega}_{(ip)}$s from (4.11), (asymptotically) more efficient estimators of the expected coefficient vector $\boldsymbol{\beta}$ can be constructed. In Step A and Step B, the inefficient unit specific *OLS* estimators of the coefficient vector, (4.3), have been our starting point. In this step we turn to the more efficient unit- and block specific GLS estimators. We then replace $\widehat{\boldsymbol{\beta}}_{(ip)}$ by

$$\widetilde{\boldsymbol{\beta}}_{(ip)} = [\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}], \quad i = 1, \dots, N_p;\ p = q, \dots, P. \tag{4.13}$$

We here proceed as if the $\boldsymbol{\Omega}_{(ip)}$s are known. In practice, we may use either the estimators $\widehat{\boldsymbol{\Omega}}_{(ip)}$ or estimate these covariance matrices from recomputed GLS residuals and coefficient-slacks, as will be described below. By using (3.1) it can be shown that this estimator is unbiased, with

$$\mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)}) = [\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}. \tag{4.14}$$

Since it can be shown from (4.1) and (4.14) that $\mathsf{V}(\widehat{\boldsymbol{\beta}}_{(ip)}) - \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})$ is positive definite, $\widetilde{\boldsymbol{\beta}}_{(ip)}$ is more efficient than $\widehat{\boldsymbol{\beta}}_{(ip)}$.

A revised estimator of $\boldsymbol{\beta}$ *based on the observations from the units observed p times*, *i.e.*, for block p, can then be defined as the matrix weighted mean of the unit specific GLS estimators, where the latter are weighted by their respective inverse covariance matrices. Our estimator of $\boldsymbol{\beta}$, revising $\widehat{\boldsymbol{\beta}}_{(p)}$, as given by (4.4), then becomes

$$\begin{aligned}
\boldsymbol{\beta}^*_{(p)} &= [\textstyle\sum_{i=1}^{N_p} \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}]^{-1}[\sum_{i=1}^{N_p} \mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}\widetilde{\boldsymbol{\beta}}_{(ip)}] \\
&= [\textstyle\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{X}_{(ip)}]^{-1}[\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}^{-1}_{(ip)}\boldsymbol{y}_{(ip)}], \quad p = q, \dots, P.
\end{aligned} \tag{4.15}$$

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

This coincides with the strict GLS estimator for block p, $\widehat{\boldsymbol{\beta}}_{(p)}^{GLS}$, given in (3.6) as the solution to the ML problem for block p conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$. From (4.14) it follows, since all $\widetilde{\boldsymbol{\beta}}_{(ip)}$s are uncorrelated, that

$$\mathsf{V}(\boldsymbol{\beta}^*_{(p)}) = [\textstyle\sum_{i=1}^{N_p} \boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}_{(ip)}^{-1}\boldsymbol{X}_{(ip)}]^{-1}. \tag{4.16}$$

Since $\boldsymbol{\beta}^*_{(p)}$ is the strict GLS estimator for block p, we know that $\mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(p)}) - \mathsf{V}(\boldsymbol{\beta}^*_{(p)})$ is positive definite. The corresponding estimator of $\boldsymbol{\beta}$ based on observations from all blocks with $p \geq q$ is

$$\begin{aligned}\boldsymbol{\beta}^* &= [\textstyle\sum_{p=q}^{P}\sum_{i=1}^{N_p}\mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}]^{-1}[\sum_{p=q}^{P}\sum_{i=1}^{N_p}\mathsf{V}(\widetilde{\boldsymbol{\beta}}_{(ip)})^{-1}\widetilde{\boldsymbol{\beta}}_{(ip)}] \\ &= [\textstyle\sum_{p=q}^{P}\sum_{i=1}^{N_p}\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}_{(ip)}^{-1}\boldsymbol{X}_{(ip)}]^{-1}[\sum_{p=q}^{P}\sum_{i=1}^{N_p}\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}_{(ip)}^{-1}\boldsymbol{y}_{(ip)}].\end{aligned} \tag{4.17}$$

It equals $\widehat{\boldsymbol{\beta}}^{GLS}$, given in (3.9), as the solution to the ML problem conditional on $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, except that observations from blocks $1, 2, \dots, q-1$ are omitted. From (4.14) it follows, because all $\widetilde{\boldsymbol{\beta}}_{(ip)}$s are uncorrelated, that

$$\mathsf{V}(\boldsymbol{\beta}^*) = [\textstyle\sum_{p=q}^{P}\sum_{i=1}^{N_p}\boldsymbol{X}'_{(ip)}\boldsymbol{\Omega}_{(ip)}^{-1}\boldsymbol{X}_{(ip)}]^{-1}. \tag{4.18}$$

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

Using the second step estimators of the coefficient vector obtained in Step C, we can revise the estimators of the disturbance covariance matrices and the covariance matrices of the random coefficients obtained in Step B. We construct from the unit-specific GLS estimators $\widetilde{\boldsymbol{\beta}}_{(ip)}$, given by (4.13), the $(Gp \times 1)$ residual vector corresponding to $\boldsymbol{u}_{(ip)}$ and rearrange it into the $(G \times p)$ matrix $\widetilde{\boldsymbol{U}}_{(ip)}$ as follows:

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

The second step estimator of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$ for block p, updating $\widehat{\boldsymbol{\Sigma}}^u_{(p)}$ and $\widehat{\boldsymbol{\Sigma}}^\delta_{(p)}$, given by (4.6) and (4.7) – using now the coefficient-slack vectors $\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*_{(p)}$ – are, respectively

$$\widetilde{\boldsymbol{\Sigma}}^u_{(p)} = \tfrac{1}{N_p p}\textstyle\sum_{i=1}^{N_p}\widetilde{\boldsymbol{U}}_{(ip)}\widetilde{\boldsymbol{U}}'_{(ip)}, \tag{4.19}$$

$$\widetilde{\boldsymbol{\Sigma}}^\delta_{(p)} = \tfrac{1}{N_p}\textstyle\sum_{i=1}^{N_p}(\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*_{(p)})(\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*_{(p)})', \qquad p = q, \dots, P. \tag{4.20}$$

We can then update $\widehat{\boldsymbol{\Omega}}_{(ip)p}$, given by (4.8), by means of

$$\widetilde{\boldsymbol{\Omega}}_{(ip)p} = \boldsymbol{X}_{(ip)}\widetilde{\boldsymbol{\Sigma}}^\delta_{(p)}\boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widetilde{\boldsymbol{\Sigma}}^u_{(p)}, \qquad i = 1, \dots, N_p;\ p = q, \dots, P. \tag{4.21}$$

The second step estimator of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$, updating $\widehat{\boldsymbol{\Sigma}}^u$ and $\widehat{\boldsymbol{\Sigma}}^\delta$, given by (4.9) and (4.10), respectively, are

$$\widetilde{\boldsymbol{\Sigma}}^u = [\textstyle\sum_{p=q}^{P}N_p p]^{-1}\sum_{p=q}^{P}\sum_{i=1}^{N_p}\widetilde{\boldsymbol{U}}_{(ip)}\widetilde{\boldsymbol{U}}'_{(ip)}, \tag{4.22}$$

$$\widetilde{\boldsymbol{\Sigma}}^\delta = [\textstyle\sum_{p=q}^{P}N_p]^{-1}\sum_{p=q}^{P}\sum_{i=1}^{N_p}(\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*)(\widetilde{\boldsymbol{\beta}}_{(ip)} - \boldsymbol{\beta}^*)'. \tag{4.23}$$

We finally update the estimators of $\boldsymbol{\Omega}_{(ip)}$, obtained from (4.11) in Step B, by using

$$\widetilde{\boldsymbol{\Omega}}_{(ip)} = \boldsymbol{X}_{(ip)} \widetilde{\boldsymbol{\Sigma}}^{\delta} \boldsymbol{X}'_{(ip)} + \boldsymbol{I}_p \otimes \widetilde{\boldsymbol{\Sigma}}^{u}, \qquad i = 1, \ldots, N_p;\ p = 1, \ldots, P. \tag{4.24}$$

An extension of this four-step sequence of estimation problems is outlined in Section 5.

5 Stepwise, modified ML algorithms

Modified, stepwise ML estimation algorithms for $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}^u$, and $\boldsymbol{\Sigma}^\delta$ can be constructed from the stepwise procedure described in Section 4. Below, we present such an algorithm. It can be considered a modified ML algorithm, provided it, when iterated according to some criterion, converges towards a unique solution. We give the algorithm for both block specific estimation and estimation from the full data set. Group specific estimates can be obtained for all blocks such that $p \in [q, P]$, where q is the lowest value of p for which OLS estimation is possible (ensures a positive number of degrees of freedom) for all G equations.

Algorithm for one block p:

- **1g:** Estimate by (4.1) for one $p\,(p \geq q)$, $i = 1, \ldots, N_p$, unit-specific estimators of $\boldsymbol{\beta}$. Extract the corresponding OLS residuals.
- **2g:** Compute from (4.4) an estimator of $\boldsymbol{\beta}$ for block p. Compute coefficient slacks.
- **3g:** Compute from (4.6)–(4.7) block specific estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$.
- **4g:** Compute from (4.8) $\widehat{\boldsymbol{\Omega}}_{(ip)p}$ for $i = 1, \ldots, N_p$.
- **5g:** Insert $\boldsymbol{\Omega}_{(ip)} = \widehat{\boldsymbol{\Omega}}_{(ip)p}$ into (4.13) and (4.15) to compute the unit- and block-specific estimators $\widetilde{\boldsymbol{\beta}}_{(ip)}$ and $\boldsymbol{\beta}^*_{(p)}$.
- **6g:** Extract revised residuals and coefficient-slacks and recompute from (4.19)–(4.20) block-specific estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$.
- **7g:** Compute $\widetilde{\boldsymbol{\Omega}}_{(ip)p}$ from (4.21) for $i = 1, \ldots, N_p$.
- **8g:** Insert $\boldsymbol{\Omega}_{(ip)} = \widetilde{\boldsymbol{\Omega}}_{(ip)p}$ into (4.13) and (4.15) and recompute $\widetilde{\boldsymbol{\beta}}_{(ip)}$ and $\boldsymbol{\beta}^*_{(p)}$.

Steps 6g–8g can be iterated until convergence, according to some criterion.

Algorithm for all blocks combined

- **1:** Estimate by (4.1) unit-specific estimators of $\boldsymbol{\beta}$ for $i = 1, \ldots, N_p;\ p \in [q, P]$, Extract the corresponding OLS residuals.
- **2:** Compute from (4.5) an estimator of $\boldsymbol{\beta}$ from the data for blocks $p \in [q, P]$. Compute coefficient slacks.
- **3:** Compute from (4.9)–(4.10) overall estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$.
- **4:** Compute from (4.11) $\widehat{\boldsymbol{\Omega}}_{(ip)}$ for $i = 1, \ldots, N_p;\ p \in [1, P]$.
- **5:** Insert $\boldsymbol{\Omega}_{(ip)} = \widehat{\boldsymbol{\Omega}}_{(ip)}$ into (4.13) and (4.17) to compute $\widetilde{\boldsymbol{\beta}}_{(ip)}$ and $\boldsymbol{\beta}^*$.
- **6:** Extract revised residuals and coefficient-slacks and recompute from (4.22)–(4.23) overall estimators of $\boldsymbol{\Sigma}^u$ and $\boldsymbol{\Sigma}^\delta$.
- **7:** Compute $\widetilde{\boldsymbol{\Omega}}_{(ip)}$ from (4.24) for $i = 1, \ldots, N_p;\ p \in [1, P]$.
- **8:** Insert $\boldsymbol{\Omega}_{(ip)} = \widetilde{\boldsymbol{\Omega}}_{(ip)}$ into (4.17) and recompute $\boldsymbol{\beta}^*$.

Steps 6–8 can be iterated until convergence, according to some criterion.

If the algorithm 1–8, after iteration of steps 6–8, converges towards a unique solution, it gives our modified ML estimator.

6 An application

Elements in the approach described above are applied to data for the Pulp and Paper industry in Norway for the years 1972–1993. With $T=22$ this is a rather long panel of micro manufacturing data. A substantial part of the data set relate to firms observed in the full 22 years, but sample attrition and accretion as well as gaps in the series have resulted in a data set in which firms responding $p = 1, 2, \ldots, 21$ years are also included, some with low values of N_p. The data set comes from virtually the same source as that used in Biørn, Lindquist, and Skjerpen (2002). For the present illustration not all available observations are exploited, only a selection of about two thirds, $N = 111$ firms, those observed $p = 22, 21, 20, 10, 7, 5$ times – *i.e.*, 6 blocks among the original 22 are included. One reason why we have 'curtailed' the data set in this way is that N_p for several p is quite low, giving potentially 'volatile' estimates in the block-specific regressions. The number of observations from the firms selected is $n=1891$, giving an average of $n/N=17$ observations per firm. The design of this data set is described in Table 1.

Table 1: *Panel design*

p	N_p	$N_p p$
22	61	1342
21	8	168
20	6	120
10	11	110
7	13	91
5	12	60
$\sum$	111	1891

Our illustrative model example has $G=3$ equations, all containing the same $K=2$ regressors. Given the rather small data set, with some units observed only 5 times, it is essential to keep the dimension of the regressor vector small.[6] Within this setting, the model is intended to explain, in a simplistic way, total factor cost per unit of output as well as cost shares for two inputs. The first equation expresses the log of the variable cost per unit of output as functions of the log of output and the log of the material price/labour cost. In this way the potential presence of non-constant returns to scale (non-unitary scale elasticity) can be examined. Equations two and three express, respectively, the cost share of materials and the cost share of labour as functions of the same exogenous variables. Technical change is, for simplicity, disregarded.[7] The specific variable definitions are:

[6] For an elaboration of this issue for a, somewhat related, random coefficient analysis based on a small-sized *balanced* panel data set, see Biørn *et al.* (2010, Section 4.2).

[7] The underlying total cost measure also includes energy cost, which is not modelled in the example. Neither are capital cost and capital input represented in the model. In other words, we have no underlying 'full' factor cost function represented.

Endogenous variables: $y_1, y_2, y_3 =$ *logcx, csm, csl*,
Exogenous variables (in all equations): $x_1, x_2 =$ *logx, logpml*,

where

logcx = log(cost/output),
csm = cost of materials as share of total cost,
csl = cost of labour as share of total cost,
logx = log(output),
logpml = log(material price/labour cost).

Tables 2 through 6 collect results from this application, the *con* columns relating to the intercept of the equations. Supplementary, block-specific results are given in the Appendix.

In Table 2 we first report, as benchmarks, OLS estimates based on all $n = 1891$ observations, along with standard errors *for the u disturbances* ($\widehat{\sigma}_u$) and coefficient standard errors calculated in the 'customary' way neglecting coefficient heterogeneity. In contrast, the empirical means of the n firm-specific slope coefficients estimates – as obtained from (4.1) and (4.5) – and the empirical standard deviations computed from the former are reported in Table 3. As expected, the two sets overall 'means' differ substantially, reflecting, *inter alia*, their different weighting of the firm-specific estimates in the aggregates. These means coincide with respect to sign, but differ substantially in magnitude. The empirical standard deviations in Table 3 signalize considerable slope coefficient heterogeneity.

The block-specific estimates corresponding to the overall estimates in Table 3 are given in the Appendix; see columns 1–4 of the A panels of Table A.1. The block-specific OLS standard errors of regression (SER), computed in the 'customary' way, are given in column 5. In columns 1–4 of the B panels of Table A.1, the block-specific coefficient distributions are described by their estimated skewness and kurtosis. By and large, the kurtosis estimates do not depart substantially from their value under normality, which is 3. The majority of the skewness estimates are in the $(-1, +1)$ range, indicating both left-skewed and right-skewed coefficient distributions (symmetric distributions having zero skewness). Columns 5–6 of the B panels contain the empirical means of the standard error estimates of the firm-specific estimates, while column 7 gives the means of estimated standard errors of the u disturbances, $\widehat{\sigma}_u$, corresponding to the overall estimate in Table 2, column 4. Their orders of magnitude are similar.

The covariance matrix of the 'coefficient-slack' vector $\boldsymbol{\delta}_{(ip)} = \boldsymbol{\beta}_{(ip)} - \boldsymbol{\beta}$, *i.e.*, $\widehat{\boldsymbol{\Sigma}}^{\delta}$, as estimated from (4.1), (4.5) and (4.10), is given in Table 4. The covariance matrix of the 'genuine disturbance vector' $\boldsymbol{u}_{(ip)}$, *i.e.*, $\widehat{\boldsymbol{\Sigma}}^{u}$, as estimated from (4.9), the residual vectors being $\widehat{\boldsymbol{u}}_{(ip)} = \boldsymbol{y}_{(ip)} - \boldsymbol{X}_{(ip)}\widehat{\boldsymbol{\beta}}_{(ip)}$, is given in Table 5.

Finally, Table 6 gives the overall Feasible Generalized Least Squares (FGLS) estimates, as obtained from (4.17), with their standard errors obtained from the diagonal elements of (4.18). The standard errors exceed those in Table 2 by a large margin. This is as expected, since the former refer to a model which disregards any coefficient heterogeneity, while the latter fully exploit this heterogeneity, 'weighting together' the effects of the firm-to-firm 'coefficient slack' and genuine 'disturbance noise'. We find clear evidence of increasing returns to scale, as *logx* comes out with a significantly negative estimate of the *expected* coefficient in the *logcx* equation, implying an elasticity of cost with respect to output in the (0,1) interval. Increasing the production scale affects the materials cost share negatively and the labour cost share positively, and the effect is significant at 'standard' p-values. Also the logged factor price ratio comes out with significant effects, again in the expected coefficient value sense. Overall, with respect to sign, these results based on FGLS and the random coefficient setup, 'robustify' the results based on simple OLS estimation in Table 2, but the coefficient estimates deviate substantially. Block specific FGLS estimates corresponding to those in Table 6, computed from (4.15) and (4.16), are given in the Appendix, Table A.2. Here most of the estimated expected slope coefficients come out as *insignificant*, except for the $p = 22$ block, for which both the number of firms and the number of observations are by far the largest. One notable exception is that *logx* comes out with an estimate close to zero in the *logcx* equation, suggesting constant returns to scale.

Table 2: *Overall OLS Estimates, for all blocks* ($p = 22, 21, 20, 10, 7, 5$)

$N = \sum N_p = 111$, $n = \sum N_p p = 1891$.

Standard errors, from OLS formula *neglecting coefficient heterogeneity*, in parenthesis

LHS VAR	*con*	*logx*	*logpml*	OLS SER: $\widehat{\sigma}_u$
logcx	0.6609	-0.3008	0.6786	0.2802
	(0.2843)	(0.0094)	(0.0661)	
csm	0.5630	-0.0439	0.0244	0.0340
	(0.0399)	(0.0013)	(0.0093)	
csl	0.5950	0.0283	-0.0400	0.0395
	(0.0455)	(0.0015)	(0.0106)	

Table 3: *Means* ($\widehat{\boldsymbol{\beta}}$) *and Standard deviations of OLS estimates for all firms*

$N = \sum N_p = 111$, $n = \sum N_p p = 1891$.

The means are based on (4.1) and (4.5).

The (across-firm) empirical standard deviations are computed from (4.1).

	$\widehat{\boldsymbol{\beta}}$ = Mean of $\widehat{\boldsymbol{\beta}}_{(ip)}$		Emp.st.dev. of $\widehat{\boldsymbol{\beta}}_{(ip)}$	
LHS VAR	*logx*	*logpml*	*logx*	*logpml*
logcx	-0.2461	0.8074	0.8424	0.9871
csm	-0.0377	0.0823	0.0722	0.1360
csl	0.0369	-0.1138	0.0785	0.1413

Table 4: *Estimate of coefficient covariance matrix,* $\widehat{\boldsymbol{\Sigma}}^{\delta}$

Based on Eq. (4.10)

	LHS VAR: *logcx*			LHS VAR: *csm*			LHS VAR: *csl*		
	con	*logx*	*logpml*	*con*	*logx*	*logpml*	*con*	*logx*	*logpml*
con	82.6957								
logx	-6.7030	0.7096							
logpml	-4.0870	0.0010	0.9744						
con	-2.3112	0.1591	0.1955	0.7651					
logx	0.1653	-0.0144	-0.0075	-0.0429	0.0052				
logpml	0.2056	-0.0082	-0.0316	-0.0813	-0.0004	0.0185			
con	0.2429	0.0143	-0.1378	-0.6685	0.0376	0.0713	0.8655		
logx	0.0079	-0.0018	0.0074	0.0358	-0.0047	0.0009	-0.0512	0.0062	
logpml	-0.1156	0.0056	0.0155	0.0757	0.0005	-0.0174	-0.0840	-0.0008	0.0200

Square root of elements of diag[$\widehat{\boldsymbol{\Sigma}}^{\delta}$]

	RHS VAR:		
LHS VAR:	*con*	*logx*	*logpml*
logcx	9.0937	0.8424	0.9871
csm	0.8747	0.0722	0.1360
csl	0.9303	0.0785	0.1413

Table 5: *Estimated disturbance covariance matrix for full system:* $\widehat{\mathbf{\Sigma}}^u$

Based on Eq. (4.9)

	LHS VAR:		
LHS VAR:	*logcx*	*csm*	*csl*
logcx	0.0785		
csm	-0.0026	0.0012	
csl	0.0008	-0.0011	0.0016

Table 6: *Overall FGLS Coefficient Estimates,* $\boldsymbol{\beta}^*$

Based on Eqs. (4.17)–(4.18). $N = \sum N_p = 111$, $n = \sum N_p p = 1891$.
Standard errors, from estimate of diag$[\mathsf{V}(\boldsymbol{\beta}^*)]^{1/2}$, in parenthesis.

	RHS VAR:		
LHS VAR:	*con*	*logx*	*logpml*
logcx	-1.9173	-0.2158	0.9230
	(0.9393)	(0.0852)	(0.1073)
csm	0.2684	-0.0367	0.0742
	(0.0935)	(0.0076)	(0.0144)
csl	0.8984	0.0327	-0.1112
	(0.1007)	(0.0083)	(0.0152)

References

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

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

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

Baltagi, B.H. (2008): *Econometric Analysis of Panel Data*, fourth edition. Chichester: Wiley.

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

Biørn, E. (2004): Regression Systems for Unbalanced Panel Data: A Stepwise Maximum Likelihood Procedure. *Journal of Econometrics* **122**, 281–291.

Biørn, E., Lindquist, K.-G. and Skjerpen, T. (2002): Heterogeneity in Returns to Scale: A Random Coefficient Analysis with Unbalanced Panel Data. *Journal of Productivity Analysis* **18**, 39–57.

Biørn, E., Hagen, T.P., Iversen, T. and Magnussen, J. (2010): How Different Are Hospitals' Responses to a Financial Reform? The Impact on Efficiency of Activity-Based Financing. *Health Care Management Science* **13**, 1–16.

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

Hsiao, C. (2003): *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Hsiao, C. and Pesaran, M.H. (2008): Random Coefficients Models. Chapter 6 in: Mátyás, L. and Sevestre, P. (eds.): *The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice*. Berlin: Springer.

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**, 311–323.

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

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**, 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**, 713–724.

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

Appendix: Block specific results

Table A.1: *Block-specific OLS results and summary statistics*

$p = 22$ *block*, $N_p = 61, N_p p = 1342$

A. Means and Standard deviations of firm specific coefficient estimates (4.1)

LHS Var.	$\widehat{\beta}_{(p)}$ = Mean of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Emp.st.dev. of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Block OLS SER
logcx	0.0319	1.0723	0.8952	0.8423	1.3806
csm	-0.0447	0.0601	0.0601	0.0915	0.1542
csl	0.0337	-0.1010	0.0659	0.1165	0.1818

B. Across-firm Skewness and Kurtosis of coef. estimates. Means of Std.Err. estimates

LHS Var.	Skewness of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Kurtosis of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_{\widehat{\beta}(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_u$
logcx	0.2072	-0.3719	1.8810	3.0642	0.2995	0.4323	0.3009
csm	0.5425	0.4122	4.1847	4.8145	0.0298	0.0439	0.0312
csl	0.0206	-0.1114	2.9044	3.1033	0.0381	0.0544	0.0379

$p = 21$ *block*, $N_p = 8, N_p p = 168$

A. Means and Standard deviations of firm specific coefficient estimates (4.1)

LHS Var.	$\widehat{\beta}_{(p)}$ = Mean of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Emp.st.dev. of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Block OLS SER
logcx	-0.4422	1.0260	0.5993	0.3999	1.2822
csm	-0.0089	0.0153	0.0635	0.1212	0.2198
csl	-0.0045	-0.0818	0.0734	0.1637	0.2491

B. Across-firm Skewness and Kurtosis of coef. estimates. Means of Std.Err. estimates

LHS Var.	Skewness of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Kurtosis of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_{\widehat{\beta}(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_u$
logcx	0.3884	0.4161	1.8163	1.5646	0.3300	0.4736	0.2920
csm	-0.2462	-0.6004	1.6532	3.2693	0.0519	0.0816	0.0471
csl	0.1807	0.7127	1.9881	2.9226	0.0641	0.0964	0.0545

$p = 20$ *block*, $N_p = 6, N_p p = 120$

A. Means and Standards deviation of firm specific coefficient estimates (4.1)

LHS Var.	$\widehat{\beta}_{(p)}$ = Mean of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Emp.st.dev. of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Block OLS SER
logcx	-0.6212	1.0240	0.3854	0.6940	1.3738
csm	-0.0369	-0.0082	0.0403	0.1196	0.1650
csl	0.0466	-0.0372	0.0480	0.1065	0.1936

B. Across-firm Skewness and Kurtosis of coef. estimates. Means of Std.Err. estimates

LHS Var.	Skewness of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Kurtosis of $\widehat{\beta}_{(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_{\widehat{\beta}(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_u$
logcx	1.4469	-0.2405	3.6188	2.0464	0.2352	0.4954	0.3227
csm	0.5037	-0.1589	2.1310	1.8273	0.0272	0.0523	0.0371
csl	0.1325	1.0647	2.9129	2.9540	0.0340	0.0652	0.0442

Table A.1: *Block-specific OLS results and summary statistics (cont.)*

$p = 10$ *block,* $N_p = 11, N_p p = 110$

A. Means and Standard deviations of firm specific coefficient estimates (4.1)

LHS Var.	$\widehat{\boldsymbol{\beta}}_{(p)}$ = Mean of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Emp.st.dev. of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Block OLS SER
logcx	-0.5530	0.5915	0.7356	1.0568	0.5448
csm	-0.0581	0.0873	0.1128	0.1727	0.0992
csl	0.0840	-0.1110	0.1188	0.1527	0.1048

B. Across-firm Skewness and Kurtosis of coef. estimates. Means of Std.Err. estimates

LHS Var.	Skewness of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Kurtosis of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_{\widehat{\boldsymbol{\beta}}(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_u$
logcx	-1.2123	0.0727	4.3715	2.8245	0.3023	0.4993	0.1909
csm	-0.0205	0.2585	2.3166	2.3321	0.0477	0.0848	0.0318
csl	0.1884	-0.4170	2.0474	2.2100	0.0479	0.0889	0.0339

$p = 7$ *block,* $N_p = 13, N_p p = 91$

A. Means and Standard deviations of firm specific coefficient estimates (4.1)

LHS Var.	$\widehat{\boldsymbol{\beta}}_{(p)}$ = Mean of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Emp.st.dev. of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Block OLS SER
logcx	-0.4860	0.5150	0.7932	1.2818	0.4772
csm	-0.0113	0.1620	0.1001	0.1742	0.0707
csl	0.0304	-0.1556	0.1048	0.1757	0.0757

B. Across-firm Skewness and Kurtosis of coef. estimates. Means of Std.Err. estimates

LHS Var.	Skewness of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Kurtosis of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_{\widehat{\boldsymbol{\beta}}(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_u$
logcx	-0.3412	-0.4855	2.8204	2.1651	0.3612	0.7595	0.2092
csm	-0.1380	-0.4044	1.5448	2.5538	0.0458	0.1176	0.0310
csl	0.1039	0.3700	1.5484	3.1027	0.0519	0.1300	0.0346

$p = 5$ *block,* $N_p = 12, N_p p = 60$

A. Means and Standards deviation of firm specific coefficient estimates (4.1)

LHS Var.	$\widehat{\boldsymbol{\beta}}_{(p)}$ = Mean of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Emp.st.dev. of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Block OLS SER
logcx	-0.7996	-0.2786	0.4086	0.9098	0.3544
csm	-0.0320	0.1939	0.0639	0.1728	0.0366
csl	0.0398	-0.1954	0.0669	0.1774	0.0380

B. Across-firm Skewness and Kurtosis of coef. estimates. Means of Std.Err. estimates

LHS Var.	Skewness of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Kurtosis of $\widehat{\boldsymbol{\beta}}_{(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_{\widehat{\boldsymbol{\beta}}(ip)}$ *logx*	*logpml*	Mean of $\widehat{\sigma}_u$
logcx	-0.7242	-0.9904	3.1037	3.0776	0.3047	0.9746	0.2190
csm	-1.6330	0.3938	4.4491	2.3055	0.0342	0.1159	0.0235
csl	1.2148	-0.5350	3.4305	2.7534	0.0344	0.1144	0.0241

Table A.2: *Block-specific FGLS results*

.

Results based on (4.15)–(4.16):
Coefficient Estimates, $\boldsymbol{\beta}^*_{(p)}$.
Standard errors, from estimate of diag$[\mathsf{V}(\boldsymbol{\beta}^*_{(p)})]^{1/2}$, in parenthesis

$p = 22$ *block:* $N_p = 61, N_p p = 1342$

LHS VAR	*con*	*logx*	*logpml*
logcx	-4.8813	-0.0004	1.1254
	(1.2339)	(0.1145)	(0.1366)
csm	0.3620	-0.0407	0.0578
	(0.1213)	(0.0102)	(0.0185)
csl	0.8902	0.0291	-0.0967
	(0.1305)	(0.0112)	(0.0195)

$p = 21$ *block:* $N_p = 8, N_p p = 168$

LHS VAR	*con*	*logx*	*logpml*
logcx	0.0094	-0.4659	1.0674
	(3.5106)	(0.3196)	(0.3845)
csm	0.3254	-0.0101	0.0227
	(0.3498)	(0.0287)	(0.0519)
csl	0.9832	0.0014	-0.0894
	(0.3788)	(0.0316)	(0.0547)

$p = 20$ *block,* $N_p = 6, N_p p = 120$

LHS VAR	*con*	*logx*	*logpml*
logcx	1.0058	-0.5887	1.0756
	(3.8935)	(0.3549)	(0.4376)
csm	0.6654	-0.0345	-0.0074
	(0.3825)	(0.0312)	(0.0592)
csl	0.4978	0.0386	-0.0409
	(0.4113)	(0.0341)	(0.0624)

$p = 10$ *block,* $N_p = 11, N_p p = 110$

LHS VAR	*con*	*logx*	*logpml*
logcx	2.0783	-0.4244	0.6200
	(3.2258)	(0.2822)	(0.3566)
csm	0.4483	-0.0651	0.0887
	(0.3278)	(0.0256)	(0.0474)
csl	0.5040	0.0815	-0.1244
	(0.3552)	(0.0280)	(0.0505)

$p = 7$ *block,* $N_p = 13, N_p p = 91$

LHS VAR	*con*	*logx*	*logpml*
logcx	3.2768	-0.4835	0.2797
	(2.9945)	(0.2551)	(0.3702)
csm	-0.3696	-0.0134	0.1787
	(0.3070)	(0.0231)	(0.0483)
csl	1.3024	0.0228	-0.1917
	(0.3345)	(0.0252)	(0.0520)

$p = 5$ *block,* $N_p = 12, N_p p = 60$

LHS VAR	*con*	*logx*	*logpml*
logcx	5.8443	-0.6110	-0.1336
	(3.0603)	(0.2588)	(0.4164)
csm	-0.3061	-0.0263	0.1787
	(0.3144)	(0.0234)	(0.0536)
csl	1.2911	0.0275	-0.1836
	(0.3415)	(0.0256)	(0.0578)