UNIVERSITÀ CATTOLICA del Sacro Cuore

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

Heteroskedastic Stratified Two-way Error Component Models of Single Equations and Seemingly Unrelated Regressions Systems

Silvia Platoni
Daniele Moro
Paolo Sckokai
Laura Barbieri

Quaderno n. 117/settembre 2016

VP VITA E PENSIERO

Università Cattolica del Sacro Cuore

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

Heteroskedastic Stratified Two-way Error Component Models of Single Equations and Seemingly Unrelated Regressions Systems

Silvia Platoni
Daniele Moro
Paolo Sckokai
Laura Barbieri

Quaderno n. 117/settembre 2016

VP VITA E PENSIERO

Silvia Platoni, Dipartimento di Scienze Economiche e Sociali, Università Cattolica del Sacro Cuore, Piacenza.

Daniele Moro, Dipartimento di Economia Agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza.

Paolo Sckokai, Dipartimento di Economia Agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza.

Laura Barbieri, Dipartimento di Scienze Economiche e Sociali, Università Cattolica del Sacro Cuore, Piacenza.

✉ silvia.platoni@unicatt.it

✉ daniele.moro@unicatt.it

✉ paolo.sckokai@unicatt.it

✉ laura.barbieri@unicatt.it

I quaderni possono essere richiesti a:
Dipartimento di Scienze Economiche e Sociali,
Università Cattolica del Sacro Cuore
Via Emilia Parmense 84 - 29122 Piacenza - Tel. 0523 599.342
http://dipartimenti.unicatt.it/dises

✉ dises-pc@unicatt.it

www.vitaepensiero.it

ISBN 978-88-343-3273-3

Abstract

A relevant issue in panel data estimation is heteroskedasticity, which often occurs when the sample size is large and individual units are of varying size. Furthermore, many of the available panel data sets are unbalanced in nature, because of attrition or accretion, and micro-econometric models applied to panel data are frequently multi-equation models. This paper considers the general least squares estimation of heteroskedastic stratified two-way error component model of both single equations and seemingly unrelated regressions (SUR) systems (with cross-equations restrictions) on unbalanced panel data. The derived heteroskedastic estimators improve the estimation efficiency, with the SUR procedures performing better than the single-equation procedures.

Keywords: Unbalanced panels; ECM; SUR; Heteroskedasticity.
JEL classification: C13; C23; C33.

1. INTRODUCTION

In applied econometrics, there is an increasing use of panel data, that Baltagi (2013, page 1) defines as "the pooling of observations on a cross-section of households, countries, firms, etc. over several time periods". The reason for this increasing use is that panel data sets are more informative, since they often provide richer and more disaggregated information. Furthermore, they allow to model individual heterogeneity and to address aggregation issues. Finally, since they span over several time periods, they also allow to describe the dynamics of the phenomena under study.

The error component model (ECM) is the standard approach to the estimation of individual and time effects in econometric single-equation models based on panel data (see Baltagi, 2013, for a review of the methods). Many of the available data sets are unbalanced in nature, that is, not all the individuals are observed over the whole time period. Several and different reasons, such as attrition or accretion, may produce an incomplete panel data set. Therefore, standard single-equation ECMs have been extended to the econometric treatment of unbalanced panel data: Biørn (1981) and Baltagi (1985) discussed the single-equation one-way ECM, Wansbeek and Kapteyn (1989) and Davis (2002) extended such estimation method to the two and multi-way cases.

Although often discarded in empirical applications, a relevant issue in panel data estimation is heteroskedasticity, which often occurs when the sample size is large and observations differ in "size characteristic" (i.e., the level of the variables). Under this perspective, heteroskedasticity arises from the fact that the degree to which a relationship may explain actual observations is likely to depend on individual specific characteristics. On the other hand, the error variance may also systematically vary across observations of similar size and, in practice, the two different sources of heteroskedasticity may be simultaneously present (see Lejeune, 1996, 2004). This means that heteroskedasticity is the rule rather than the exception when dealing with individual data concerning households or firms. Assuming homoskedastic disturbances when heteroskedasticity is present will still result in consistent estimates of the regression coefficients, but these estimates will not be efficient. Also, the standard errors of the fixed-effect (FE) estimates will be biased and robust standard errors should be computed in order to correct for the possible presence of heteroskedasticity.

Several authors have analyzed the problem of heteroskedasticity in balanced panel data, usually considering a single-equation regression model with one-way disturbances $\varepsilon_{it} = \mu_i + u_{it}$[1]. Baltagi and Griffin (1988) are concerned with the estimation of a random-effect (RE) model allowing for heteroskedasticity on the individual-specific error term $\text{Var}(\mu_i) = \varphi_i$. In contrast, Rao et al. (1981), Magnus (1982), Baltagi (1988), and Wansbeek (1989) adopt a symmetrically opposite specification allowing for heteroskedasticity on the remainder error term $\text{Var}(u_{it}) = \psi_i$.

As Mazodier and Trognon (1978) pointed out, if the φ_i's are unknown, then there is no hope to estimate them from the data: even if the μ_i's were observed, it would be impossible to estimate their variances from only one observation on each individual disturbance. Therefore, the model proposed by Baltagi and Griffin (1988) suffers from the incidental parameters problem[2]

[1] While all these papers assume constant slope coefficients, Bresson et al. (2006, 2011) allow variations in parameters across cross-sectional units in order to take into account the between individual heterogeneity. Hence, these authors derive a hierarchical Bayesian panel data estimator for a random coefficient model (RCM), where heteroskedasticity is modeled following both the RCMs on panel data proposed by Hsiao and Pesaran (2004) and Chib (2008) and the general heteroskedastic one-way ECM proposed by Randolph (1988), who assumes that both the individual-specific term μ_i and the remainder error term u_{it} are heteroskedastic.

[2] Neyman and Scott (1948) study maximum likelihood (ML) estimation of

(see Phillips, 2003; Baltagi, 2013). Furthermore, also the models allowing for heteroskedasticity on the remainder error term u_{it} suffer from the incidental parameters problem when the time dimension of the panel is short.

There are two possible solutions to avoid the incidental parameters problem (see Baltagi, 2013): either to allow the variances to change across strata (i.e., stratified *ECM*s) or, if the variables that determine heteroskedasticity are known, to specify parametric variance functions (i.e., adaptive estimation of heteroskedasticity of unknown form).

Mazodier and Trognon (1978) proposed a stratified two-way *ECM*, i.e., $\varepsilon_{it} = \mu_i + \nu_t + u_{it}$, on balanced panels in which both the individual-specific effect μ_i and time-specific effect ν_t variances are constant within subsets of observations (or strata), but are allowed to change across strata. More recently, Phillips (2003) considers a stratified one-way *ECM*, again on balanced panels, where the variances of the individual-specific effect μ_i are allowed to change not across individuals but across strata, and provides an expectation-maximization (*EM*) algorithm to estimate the model's parameters.

Li and Stengos (1994) derive an adaptive estimator for the heteroskedastic one-way *ECM* using balanced panel data where heteroskedasticity is placed on the remainder error term, and hence, $\text{Var}\,(u_{it}|\mathbf{x}_{it}) = \psi\,(\mathbf{x}_{it}) \equiv \psi_{it}$[3]. More recently, Roy (2002) derives a similar adaptive estimator where heteroskedasticity is placed on the individual-specific term rather than the remainder disturbance, and hence, $\text{Var}\,(\mu_i|\bar{\mathbf{x}}_{i\cdot}) = \varphi\,(\bar{\mathbf{x}}_{i\cdot}) \equiv \varphi_i$. Baltagi et al.

models having both structural and incidental parameters: while the structural parameters can be consistently estimated, the incidental parameters cannot be consistently estimated. These authors show that the estimation of the *ML* model is inconsistent (or partially inconsistent) if the model contains nuisance or incidental parameters which increase in number with the sample size.

[3]Throughout the paper, all vectors and matrices are in boldface.

(2005) check the sensitivity of these two adaptive heteroskedastic estimators to misspecification of the form of heteroskedasticity, showing that misleading inference may occur when heteroskedasticity is present in both components. Therefore, accounting for both sources of heteroskedasticity seems to be very important in empirical work.

Indeed, if heteroskedasticity is due to differences in size characteristic across statistical units (i.e., individuals, households, firms or countries), then both error components are expected to be heteroskedastic, and it may be difficult to argue that only one component of the error term is heteroskedastic but not the other (see Bresson et al., 2006, 2011). To this end, Randolph (1988), working on unbalanced panel data, allows for a more general heteroskedastic single-equation one-way *ECM*, assuming that both the individual-specific and remainder error terms are heteroskedastic, i.e., $\mathrm{Var}(\mu_i) = \varphi_i$ and $\mathrm{E}(\mathbf{uu}') = \mathrm{diag}[\psi_{it}]$. Lejeune (1996, 2004) is concerned with the estimation and specification testing of a full heteroskedastic one-way *ECM*, in the spirit of Randolph (1988) and Baltagi et al. (2005), and specifies parametrically the variance functions. Baltagi et al. (2006), in the spirit of Randolph (1988) and Lejeune (1996, 2004), derive a joint Lagrange multiplier (*LM*) test for homoskedasticity against the alternative of heteroskedasticity both in the individual-specific term μ_i and in the remainder error term u_{it}.

Micro-econometric models applied to panel data are often multi-equation models. Primal and dual production models are a common case, when systems of input demands and/or output supply equations have to be estimated; the same is true for systems of demand equations in consumer analysis. Baltagi (1980) and Magnus (1982) extended the estimation procedure of the single-equation model to the case of seemingly unrelated regressions (*SUR*s) for balanced panels; Biørn (2004) proposed a parsimonious technique to estimate one-way *SUR* systems on unbalanced panel data; Platoni et al. (2012) extended the procedure suggested by Biørn (2004) to the two-way case. Although heteroskedasticity is a frequent and relevant issue also in the multi-equation models applied to (unbalanced) panel data, to our knowledge very few papers concerning heteroskedastic *SUR* systems have been published. A relevant exception is Verbon (1980), who derived a *LM* test for heteroskedasticity in a model of *SUR* equations for balanced panels.

In order to fill this gap in the literature, this paper extends previous results to the estimation of heteroskedastic stratified two-way *ECM*, i.e., $\varepsilon_{it} = \mu_i + \nu_t + u_{it}$, on unbalanced panel data[4] in the case of *SUR* systems (with cross-equations restrictions). The individual-specific effect μ_i and remainder error term u_{it} variances and covariances are constant within strata, but are allowed to change across strata. Indeed, the variance and covariance estimations in two-way *SUR* systems are implemented, starting from the straightforward extension of the heteroskedastic single-equation stratified *ECM* from the one-way to the two-way case. Moreover, the estimations are implemented by two methods: the quadratic unbiased estimation (*QUE*) procedure suggested by Wansbeek and Kapteyn (1989) and the within-between (*WB*) procedure proposed by Biørn (2004).

The remainder of the paper proceeds as follows. While Section 2 describes the heteroskedastic two-way estimation for single equations, Section 3 extends the analysis to the corresponding estimation for *SUR* systems. Finally, Section 4 provides simulation results, Section 5 discusses a possible extension to heteroskedasticity on the time-specific error term, and Section 6 draws some conclusions.

[4] The estimation procedures proposed here can definitely be applied also to balanced panel data.

2. HETEROSKEDASTIC STRATIFIED SINGLE-EQUATION TWO-WAY ECM

We start by considering an unbalanced panel characterized by a total of n observations, with N individuals (indexed $i = 1, \ldots, N$) observed over T periods (indexed $t = 1, \ldots, T$). Let T_i denote the number of times the individual i is observed and N_t the number of individuals observed in period t. Hence, $\sum_i T_i = \sum_t N_t = n$.

In the following we consider the regression model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mu_i + \nu_t + u_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}, \tag{1}$$

where $\mathbf{x}_{it}$ is a $1 \times k$ vector of explanatory variables and $\boldsymbol{\beta}$ a $k \times 1$ vector of parameters, μ_i is the individual-specific effect, ν_t the time-specific effect, and u_{it} the remainder error term; in the *RE* model ε_{it} is the composite error term.

Using the $n \times N$ matrix $\boldsymbol{\Delta}_\mu$ and the $n \times T$ matrix $\boldsymbol{\Delta}_\nu$, that are matrices of indicator variables denoting observations on individuals and time periods respectively, we can define the $N \times N$ diagonal matrix $\boldsymbol{\Delta}_N \equiv \boldsymbol{\Delta}'_\mu \boldsymbol{\Delta}_\mu$ (diagonal elements correspond to the T_i's) and the $T \times T$ diagonal matrix $\boldsymbol{\Delta}_T \equiv \boldsymbol{\Delta}'_\nu \boldsymbol{\Delta}_\nu$ (diagonal elements correspond to the N_t's), as well as the $T \times N$ matrix of zeros and ones $\boldsymbol{\Delta}_{TN} \equiv \boldsymbol{\Delta}'_\nu \boldsymbol{\Delta}_\mu$, indicating the absence or presence of an individual in a certain time period. Hence, using matrix notation, we can write:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}_\mu \boldsymbol{\mu} + \boldsymbol{\Delta}_\nu \boldsymbol{\nu} + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{X}$ is a $n \times k$ matrix of explanatory variables.

Let us assume there exists a meaningful stratification of observations[5]. Hence, the unbalanced panel can also be characterized by A strata (indexed $a = 1, \ldots, A$), with $\hat{N}_a$ the number of

[5] In empirical work the number of strata is unidentified. Therefore, it is necessary to use a selection procedure, such as the Akaike (1974) information criterion, to determine the number of strata.

individuals belonging to stratum a (indexed $\hat{\imath}_a = \hat{1}_a, \ldots, \hat{N}_a$) and $\hat{I}_a$ the set of individuals belonging to stratum a. Therefore, the number of observations related to stratum a is $\hat{n}_a = \sum_{i \in \hat{I}_a} T_i = \sum_{\hat{\imath}_a = \hat{1}_a}^{\hat{N}_a} T_{\hat{\imath}_a}$. Hence, $\sum_{a=1}^{A} \hat{N}_a = N$ and $\sum_{a=1}^{A} \hat{n}_a = n$.

Using the $n \times A$ matrix $\mathbf{\Delta}_\alpha$ of indicator variables denoting observations on strata, we can define the $A \times A$ diagonal matrix $\mathbf{\Delta}_A \equiv \mathbf{\Delta}'_\alpha \mathbf{\Delta}_\alpha$ (diagonal elements correspond to the $\hat{n}_a$'s) and the $A \times N$ matrix of zeros and ones $\mathbf{\Delta}_{AN} \equiv \mathbf{\Delta}'_\alpha \mathbf{\Delta}_\mu \mathbf{\Delta}_N^{-1}$, indicating the absence or presence of an individual in a certain stratum (notice that $\mathbf{\Delta}'_\alpha \mathbf{\Delta}_\mu$ is a matrix of zeros and T_i's for $i \in \hat{I}_a$ or $T_{\hat{\imath}_a}$'s).

As Mazodier and Trognon (1978) and Phillips (2003), we assume the individual-specific error and remainder error variances are constant within stratum but change across strata. Hence, heteroskedasticity on the individual-specific disturbance implies $\mu_i \sim (0, \varphi_a)$, while heteroskedasticity on the remainder error term implies $u_{it} \sim (0, \psi_a)$.

2.1. Robust Two-way FE

In the *FE* model the individual-specific term μ_i and the time-specific term ν_t are parameters to be estimated. Therefore, heteroskedasticity is placed only on the remainder error u_{it} by assuming $u_{it} \sim (0, \psi_a)$. The Within (W) transformation[6] of the

[6]For a *FE* model the number of fixed-effect parameters $\mu_1, \ldots, \mu_N$ and $\nu_1, \ldots, \nu_T$ increases with the number of individuals N and periods T, respectively. Hence, the conventional asymptotic result cannot be applied: if $N \to \infty$, then estimates of the parameters $\mu_1, \ldots, \mu_N$ are necessarily inconsistent for a fixed T (see Wang and Ho, 2010), and if $T \to \infty$, then estimates of the parameters $\nu_1, \ldots, \nu_T$ are necessarily inconsistent for a fixed N. Therefore, when the time dimension of the panel is short, the noise in the estimation of the incidental parameters μ_i contaminates the *ML* estimates of the structural parameters (see Bester and Hansen, 2016). The literature proposes some solutions to the incidental parameters problem for some of the models, usually relying on removing the incidental parameters before estimations (see Wang and Ho, 2010). One popular approach, widely used

two-way *ECM* is based on the $n \times n$ matrix:

$$\mathbf{Q}_{[\Delta]} = \mathbf{Q}_A - \mathbf{P}_B = \mathbf{Q}_A - \mathbf{Q}_A \boldsymbol{\Delta}_\nu \mathbf{Q}^- \boldsymbol{\Delta}'_\nu \mathbf{Q}_A, \tag{3}$$

where $\mathbf{Q}_A = \mathbf{I}_n - \mathbf{P}_A$, $\mathbf{P}_A = \boldsymbol{\Delta}_\mu \boldsymbol{\Delta}_N^{-1} \boldsymbol{\Delta}'_\mu$, and $\mathbf{Q} = \boldsymbol{\Delta}'_\nu \mathbf{Q}_A \boldsymbol{\Delta}_\nu$, with $\mathbf{Q}^-$ the generalized inverse (see Wansbeek and Kapteyn, 1989; Davis, 2002). Therefore, the *W* estimator is:

$$\hat{\boldsymbol{\beta}}^W = \left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{y}, \tag{4}$$

where the number of explanatory variables, obviously without the intercept, is $k-1$.

Moreover, the following assumptions are made[7].

FE.1 Strict exogeneity The set of $(k-1) \cdot T_i$ explanatory variables for each individual $\mathbf{x}_{i(N)} \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT_i})$ is uncorrelated with the idiosyncratic error u_{it} and the set of $(k-1) \cdot N_t$ explanatory variables in each time period $\mathbf{x}_{t(T)} \equiv (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \ldots, \mathbf{x}_{N_t t})$ is also uncorrelated with the same idiosyncratic error u_{it}:

$$\begin{aligned} \mathrm{E}\left(u_{it} \left| \mathbf{x} \right., \mu_i, \nu_t\right) &= \mathrm{E}\left(u_{it} \left| \mathbf{x}_{i(N)} \right., \mu_i, \nu_t\right) \\ &= \mathrm{E}\left(u_{it} \left| \mathbf{x}_{t(T)} \right., \mu_i, \nu_t\right) = 0, \end{aligned}$$

with $\mathbf{x} \equiv (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1T_1}, \mathbf{x}_{21}, \ldots, \mathbf{x}_{2T_2}, \ldots, \mathbf{x}_{N1}, \ldots, \mathbf{x}_{NT_N})$.

FE.2 Consistency The *W* estimator in (4) is asymptotically well behaved, in the sense that the "adjusted" $(k-1) \times (k-1)$ outer product matrix $\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}$ has the appropriate rank:

$$\operatorname{rank}\left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right) = k-1.$$

in linear models, is to transform the model by the *W* transformation (i.e., y_{it} and the $1 \times (k-1)$ vector $\mathbf{x}_{it}$ are demeaned), as we have done in deriving our estimation.

[7]Details on the assumptions FE.1 and FE.2 can be found in Appendix A of Platoni et al. (2012).

Under assumptions FE.1 and FE.2 the W estimator is consistent and asymptotically normal (see Wooldridge, 2010). But without assuming homoskedasticity and no serial correlation (i.e., the assumption FE.3 in Appendix A of Platoni et al., 2012), the expression

$$\operatorname{Var}\left(\hat{\boldsymbol{\beta}}^{W}\right) = \hat{\sigma}_u^2 \cdot \left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right)^{-1}, \tag{5}$$

with $\hat{\sigma}_u^2$ the estimator of σ_u^2, gives an improper variance-covariance matrix estimator (see Wooldridge, 2010).

To obtain robust standard errors we follow the simple method suggested by Arellano (1987) for the one-way *ECM*, and proposed also by Baltagi (2013). If we stack the observations for each individual i, we can write:

$$\begin{aligned} \widetilde{\mathbf{y}}_i &= \left(\mathbf{E}_{T_i} - \mathbf{E}_{T_i}\mathbf{D}_i\mathbf{Q}^{-}\mathbf{D}_i'\mathbf{E}_{T_i}\right)\mathbf{y}_i, \\ \widetilde{\mathbf{X}}_i &= \left(\mathbf{E}_{T_i} - \mathbf{E}_{T_i}\mathbf{D}_i\mathbf{Q}^{-}\mathbf{D}_i'\mathbf{E}_{T_i}\right)\mathbf{X}_i, \end{aligned} \tag{6}$$

where $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$, with $\mathbf{I}_{T_i}$ an identity matrix of dimension T_i, $\bar{\mathbf{J}}_{T_i} = \frac{\mathbf{J}_{T_i}}{T_i}$, and $\mathbf{J}_{T_i}$ a matrix of ones of dimension T_i, and $\mathbf{D}_i$ is the $T_i \times T$ matrix obtained from the $T \times T$ identity matrix $\mathbf{I}_T$ by omitting the rows corresponding to periods in which the individual i is not observed. Therefore, we can compute the $T_i \times 1$ vector $\widetilde{\mathbf{e}}_i = \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{X}}_i\hat{\boldsymbol{\beta}}^{W}$ and the robust asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}^{W}$ is estimated by:

$$\operatorname{Var}\left(\hat{\boldsymbol{\beta}}^{W}\right) = \left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right)^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{e}}_i\widetilde{\mathbf{e}}_i'\widetilde{\mathbf{X}}_i \left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right)^{-1}. \tag{7}$$

However, since $u_{it} \sim (0, \psi_a)$, it is possible to obtain robust standard errors also by stacking the observations for each stratum a, as described later in Appendix A.1.

2.2. GLS Estimation

In the *RE* model, not only the remainder error u_{it}, but also the individual-specific error μ_i and the time-specific error ν_t are random variables.

If we assume that the variances of μ_i, ν_t, and u_{it} are known, then we can write the general least squares (GLS) estimator for $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}^{GLS} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}, \tag{8}$$

where the number of explanatory variables is k, as the problem of minimizing $\boldsymbol{\varepsilon}'_{it}\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_{it}$, where $\boldsymbol{\Omega}$ is the $n \times n$ variance-covariance matrix.

The following assumptions are made[8].

RE.1.a Strict exogeneity Same definition as assumption FE.1.

RE.1.b and RE.1.c Orthogonality conditions Both μ_i and ν_t are orthogonal to the corresponding sets of explanatory variables, that is the $k \cdot T_i$ explanatory variables for each individual $\mathbf{x}_{i(N)}$ and the $k \cdot N_t$ explanatory variables in each time period $\mathbf{x}_{t(T)}$:

$$\mathrm{E}\left(\mu_i \left|\mathbf{x}_{i(N)}\right.\right) = \mathrm{E}\left(\mu_i\right) = 0 \text{ and } \mathrm{E}\left(\nu_t \left|\mathbf{x}_{t(T)}\right.\right) = \mathrm{E}\left(\nu_t\right) = 0.$$

RE.2 Rank condition The $k \times k$ weighted outer product matrix $\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}$ has the appropriate rank, ensuring the GLS estimator in (8) is consistent:

$$\operatorname{rank}\left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right) = k.$$

Assuming homoskedasticity and no serial correlation (i.e., the assumption RE.3 in Appendix B of Platoni et al., 2012), the variance-covariance matrix $\boldsymbol{\Omega}$ has the following form:

$$\boldsymbol{\Omega} = \sigma_u^2 \cdot \mathbf{I}_n + \sigma_\mu^2 \cdot \boldsymbol{\Delta}_\mu\boldsymbol{\Delta}'_\mu + \sigma_\nu^2 \cdot \boldsymbol{\Delta}_\nu\boldsymbol{\Delta}'_\nu, \tag{9}$$

[8]Details on the assumptions RE.1 and RE.2 can be found in Appendix B of Platoni et al. (2012).

and the GLS estimator in (8) is efficient. However, assuming homoskedastic μ_i and/or u_{it} when heteroskedasticity is present will still result in consistent estimates of the regression coefficients, but these estimates will not be efficient.

With general heteroskedasticity, that is $\mu_i \sim (0, \varphi_a)$ and $u_{it} \sim (0, \psi_a)$, the matrix $\mathbf{\Omega}$ in (9) is modified to:

$$\mathbf{\Omega} = \mathbf{\Psi} + \mathbf{\Delta}_\mu \mathbf{\Phi} \mathbf{\Delta}'_\mu + \sigma^2_\nu \cdot \mathbf{\Delta}_\nu \mathbf{\Delta}'_\nu, \tag{10}$$

with the $N \times N$ matrix $\mathbf{\Phi} = \text{diag}\left[\mathbf{\Delta}'_{AN} \boldsymbol{\varphi}\right]$, the $A \times 1$ vector $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_A)'$, the $n \times n$ matrix $\mathbf{\Psi} = \text{diag}\left[\mathbf{\Delta}_\mu \mathbf{\Delta}'_{AN} \boldsymbol{\psi}\right]$, and the $A \times 1$ vector $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_A)'$.

The ANOVA-type quadratic unbiased estimator of the variance components based on the W residuals in the homoskedastic case (9) is determined in Wansbeek and Kapteyn (1989) and Davis (2002). The estimation of the components of the variance-covariance matrix $\mathbf{\Omega}$ in the heteroskedastic case (10) can be obtained modifying the QUE procedure suggested by Wansbeek and Kapteyn (1989).

The QUE procedure considers the $n \times 1$ residuals $\mathbf{e} \equiv \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^W$ from the W estimator in (4), where $\mathbf{X}$ is a matrix of dimension $n \times (k-1)$, since it does not include the intercept. If we assume that the $n \times k$ matrix $\mathbf{X}$ in (8) contains a vector of ones, we have to define the $n \times 1$ consistent centered residuals $\mathbf{f} \equiv \mathbf{E}_n \mathbf{e} = \mathbf{e} - \bar{e}$, where $\mathbf{E}_n = \mathbf{I}_n - \bar{\mathbf{J}}_n$, with $\mathbf{I}_n$ being an identity matrix of dimension n, $\bar{\mathbf{J}}_n = \frac{\mathbf{J}_n}{n}$, and $\mathbf{J}_n$ a matrix of ones of dimension n (see Wansbeek and Kapteyn, 1989). Moreover, we have to define also the $\hat{n}_a \times 1$ consistent centered residuals $\mathbf{f}_a = \mathbf{H}_a \mathbf{f}$, with $\mathbf{H}_a$ the $\hat{n}_a \times n$ matrix obtained from the identity matrix $\mathbf{I}_n$ by omitting the rows referring to observations not related to stratum a, and the matrix $\bar{\mathbf{J}}_{\hat{n}_a} = \frac{\mathbf{J}_{\hat{n}_a}}{\hat{n}_a}$ with $\mathbf{J}_{\hat{n}_a}$ a matrix of ones of dimension $\hat{n}_a$.

The adapted QUEs for $\mathbf{\Psi}$, $\mathbf{\Phi}$, and σ^2_ν is obtained by equating:

$$
\begin{aligned}
q_{a(n)} &\equiv \mathbf{f}'\mathbf{Q}_{[\Delta]}\mathbf{H}_a'\mathbf{H}_a\mathbf{Q}_{[\Delta]}\mathbf{f} \rightarrow \sum_{a=1}^{A} q_{a(n)} = q_n \equiv \mathbf{f}'\mathbf{Q}_{[\Delta]}\mathbf{f}, \\
q_{a(N)} &\equiv \mathbf{f}_a'\bar{\mathbf{J}}_{\hat{n}_a}\mathbf{f}_a \rightarrow \sum_{a=1}^{A} q_{a(N)} = q_N \equiv \mathbf{f}'\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}_N^{-1}\boldsymbol{\Delta}_\mu'\mathbf{f}, \\
q_T &\equiv \mathbf{f}'\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}_T^{-1}\boldsymbol{\Delta}_\nu'\mathbf{f},
\end{aligned} \tag{11}
$$

to their expected values (see Wansbeek and Kapteyn, 1989; Davis, 2002). For more details on the expressions in (11), see Appendix A.2.

With the $n \times n$ matrix $\mathbf{M} \equiv \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_{[\Delta]}$ (and then by definition $\mathbf{e} = \mathbf{My} = \mathbf{M}\boldsymbol{\varepsilon}$ and $\mathbf{ff}' = \mathbf{E}_n\mathbf{ee}'\mathbf{E}_n = \mathbf{E}_n\mathbf{M}\boldsymbol{\Omega}\mathbf{M}'\mathbf{E}_n$), the expected value of $q_{a(n)}$ is:

$$
\begin{aligned}
\mathrm{E}\left(q_{a(n)}\right) &= \mathrm{tr}\left(\mathbf{H}_a\mathbf{Q}_{[\Delta]}\mathbf{E}_n\mathbf{M}\boldsymbol{\Omega}\mathbf{M}'\mathbf{E}_n\mathbf{Q}_{[\Delta]}\mathbf{H}_a'\right) \\
&= \left(\hat{n}_a - \hat{N}_a - \tau_a\right) \cdot \psi_a - k_a \cdot \bar{\psi},
\end{aligned} \tag{12}
$$

where $\tau_a \equiv \hat{n}_a - \hat{N}_a - \mathrm{tr}(\mathbf{H}_a\mathbf{Q}_{[\Delta]}\mathbf{H}_a')$, with $\sum_{a=1}^{A} \tau_a = T - 1$, $k_a \equiv \mathrm{tr}[(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{H}_a'\mathbf{H}_a\mathbf{Q}_{[\Delta]}\mathbf{X}]$, with $\sum_{a=1}^{A} k_a = k-1$, and $\bar{\psi} \approx \sigma_u^2$ is obtained by equating q_n to its expected value (see Wansbeek and Kapteyn, 1989; Davis, 2002):

$$
\mathrm{E}\left(q_n\right) = [n - N - (T-1) - (k-1)] \cdot \sigma_u^2. \tag{13}
$$

Hence, the estimator of ψ_a is:

$$
\hat{\psi}_a = \frac{q_{a(n)} + k_a \cdot \hat{\sigma}_u^2}{\hat{n}_a - \hat{N}_a - \tau_a}. \tag{14}
$$

The expected value of $q_{a(N)}$ is:

$$
\begin{aligned}
&\mathrm{E}\left(q_{a(N)}\right) = \mathrm{tr}\left(\bar{\mathbf{J}}_{\hat{n}_a}\mathbf{H}_a\mathbf{E}_n\mathbf{M}\boldsymbol{\Omega}\mathbf{M}'\mathbf{E}_n\mathbf{H}_a'\right) \\
&= \left(\hat{N}_a - 2 \cdot \tfrac{\hat{n}_a}{n}\right) \cdot \psi_a + \left(k_{N_a} - k_{0_a} + \tfrac{\hat{n}_a}{n} \cdot k_0 + \tfrac{\hat{n}_a}{n}\right) \cdot \bar{\psi} \\
&+ \left(\hat{n}_a - 2 \cdot \lambda_{\mu_a}\right) \cdot \varphi_a + \tfrac{\hat{n}_a}{n} \cdot \lambda_\mu \cdot \bar{\varphi} + \left(\hat{N}_a - 2 \cdot \lambda_{\nu_a} + \tfrac{\hat{n}_a}{n} \cdot \lambda_\nu\right) \cdot \sigma_\nu^2,
\end{aligned} \tag{15}
$$

where $k_{N_a} \equiv \mathrm{tr}[(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}_a'\bar{\mathbf{J}}_{\hat{n}_a}\mathbf{X}_a]$, $k_0 \equiv \frac{\boldsymbol{\iota}_n'\mathbf{X}(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\iota}_n}{n}$, $k_{0_a} \equiv 2 \cdot \frac{\boldsymbol{\iota}_n'\mathbf{X}(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}_a'\boldsymbol{\iota}_{\hat{n}_a}}{n} = 2 \cdot \frac{\boldsymbol{\iota}_{\hat{n}_a}'\mathbf{X}_a(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\iota}_n}{n}$, with $\boldsymbol{\iota}_n$ and $\boldsymbol{\iota}_{\hat{n}_a}$ vectors of ones of dimension n and $\hat{n}_a$ respectively, $\lambda_\mu \equiv \frac{\boldsymbol{\iota}_n'\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}_\mu'\boldsymbol{\iota}_n}{n} = \frac{\sum_{i=1}^N T_i^2}{n}$, $\lambda_{\mu_a} \cdot \varphi_a \equiv \frac{\boldsymbol{\iota}_n'\boldsymbol{\Delta}_\mu\boldsymbol{\Phi}\boldsymbol{\Delta}_\mu'\mathbf{H}_a'\boldsymbol{\iota}_{\hat{n}_a}}{n} = \frac{\sum_{i\in\hat{I}_a} T_i^2}{n} \cdot \varphi_a = \frac{\sum_{\hat{\imath}_a=\hat{1}_a}^{\hat{N}_a} T_{\hat{\imath}_a}^2}{n} \cdot \varphi_a$, $\lambda_\nu \equiv \frac{\boldsymbol{\iota}_n'\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}_\nu'\boldsymbol{\iota}_n}{n} = \frac{\sum_{t=1}^T N_t^2}{n}$, $\lambda_{\nu_a} \equiv \frac{\boldsymbol{\iota}_n'\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}_\nu'\mathbf{H}_a'\boldsymbol{\iota}_{\hat{n}_a}}{n} = \frac{\sum_{t\in\hat{J}_a} N_t}{n}$, with $\hat{J}_a$ the set of periods individuals belonging to stratum a are observed, and $\bar{\varphi} \approx \sigma_\mu^2$ is obtained jointly with σ_ν^2 by equating q_N and q_T to their expected values (see Wansbeek and Kapteyn, 1989; Davis, 2002):

$$\begin{aligned} \mathrm{E}\left(q_N\right) = \quad & (N + k_N - k_0 - 1) \cdot \sigma_u^2 + (n - \lambda_\mu) \cdot \sigma_\mu^2 + \\ & (N - \lambda_\nu) \cdot \sigma_\nu^2, \\ \mathrm{E}\left(q_T\right) = \quad & (T + k_T - k_0 - 1) \cdot \sigma_u^2 + (T - \lambda_\mu) \cdot \sigma_\mu^2 + \\ & (n - \lambda_\nu) \cdot \sigma_\nu^2, \end{aligned} \tag{16}$$

with $k_N \equiv \mathrm{tr}[(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta}_\mu\boldsymbol{\Delta}_N\boldsymbol{\Delta}_\mu'\mathbf{X}]$ and $k_T \equiv \mathrm{tr}[(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}_T\boldsymbol{\Delta}_\nu'\mathbf{X}]$. Hence, the estimator of φ_a is:

$$\begin{aligned} \hat{\varphi}_a = & \frac{q_{a(N)} - \frac{\hat{n}_a}{n} \cdot \lambda_\mu \cdot \hat{\sigma}_\mu^2 - \left(\hat{N}_a - 2 \cdot \lambda_{\nu_a} + \frac{\hat{n}_a}{n} \cdot \lambda_\nu\right) \cdot \hat{\sigma}_\nu^2}{\hat{n}_a - 2 \cdot \lambda_{\mu_a}} \\ & + \frac{-\left(\hat{N}_a - 2 \cdot \frac{\hat{n}_a}{n}\right) \cdot \hat{\psi}_a - \left(k_{N_a} - k_{0_a} + \frac{\hat{n}_a}{n} \cdot k_0 + \frac{\hat{n}_a}{n}\right) \cdot \hat{\sigma}_u^2}{\hat{n}_a - 2 \cdot \lambda_{\mu_a}}. \end{aligned} \tag{17}$$

Simpler heteroskedastic schemes (i.e., heteroskedasticity only on the individual-specific disturbance and on the remainder error) can be obtained combining results for the general scheme with those for the homoskedastic case, although when we consider the case of heteroskedasticity only on the individual-specific disturbance, the expected value of $q_{a(N)}$ is obtained differently as:

$$\begin{aligned} \mathrm{E}\left(q_{a(N)}\right) = & \left(\hat{N}_a + k_{N_a} - k_{0_a} + \frac{\hat{n}_a}{n} \cdot k_0 - \frac{\hat{n}_a}{n}\right) \cdot \sigma_u^2 \\ & + (\hat{n}_a - 2 \cdot \lambda_{\mu_a}) \cdot \varphi_a + \frac{\hat{n}_a}{n} \cdot \lambda_\mu \cdot \bar{\varphi} + \left(\hat{N}_a - 2 \cdot \lambda_{\nu_a} + \frac{\hat{n}_a}{n} \cdot \lambda_\nu\right) \cdot \sigma_\nu^2, \end{aligned} \tag{18}$$

and, therefore,

$$\hat{\varphi}_a = \frac{q_{a(N)} - \frac{\hat{n}_a}{n} \cdot \lambda_\mu \cdot \hat{\sigma}^2_\mu - \left(\hat{N}_a - 2 \cdot \lambda_{\nu_a} + \frac{\hat{n}_a}{n} \cdot \lambda_\nu\right) \cdot \hat{\sigma}^2_\nu}{\hat{n}_a - 2 \cdot \lambda_{\mu_a}} + \frac{-\left(\hat{N}_a + k_{N_a} - k_{0_a} + \frac{\hat{n}_a}{n} \cdot k_0 - \frac{\hat{n}_a}{n}\right) \cdot \hat{\sigma}^2_u}{\hat{n}_a - 2 \cdot \lambda_{\mu_a}}. \tag{19}$$

3. HETEROSKEDASTIC STRATIFIED TWO-WAY SUR SYSTEMS

When systems of equations have to be estimated, as it is the case of *SUR* systems, single-equation estimation techniques are not appropriate. In order to estimate heteroskedastic two-way *SUR* systems we extend the procedure in Biørn (2004), with individuals grouped according to the number of times they are observed.

3.1. Model and Notation

Let $\widetilde{N}_p$ denote the number of individuals observed exactly in p periods, with $p = 1, \ldots, T$. Hence $\sum_p \widetilde{N}_p = N$ and $\sum_p \left(\widetilde{N}_p \cdot p\right) = n$. Moreover, let $N_{a,p}$ denote the number of individuals belonging to stratum a and observed in p periods; therefore, $\sum_a N_{a,p} = \widetilde{N}_p$ and $\sum_p \sum_a N_{a,p} = N$.

We assume that the T groups of individuals are ordered such that the $\widetilde{N}_1$ individuals observed once come first, the $\widetilde{N}_2$ individuals observed twice come second, etc. Hence, with $C_p = \sum_{h=1}^{p} \widetilde{N}_h$ being the cumulated number of individuals observed at most p times, the index sets of the individuals observed exactly p times can be written as $I_p = \{C_{p-1} + 1, \ldots, C_p\}$. Note that I_1 may be considered as a pure cross section and I_p, with $p \geq 2$, as a pseudo-balanced panel with p observations for each individual. This structure allows us to use a number of results derived for the two-way *SUR* in the balanced case.

If k_m is the number of regressors for equation m, the total number of regressors for the system is $K = \sum_{m=1}^{M} k_m$. Stacking the M equations, indexed by $m = 1, \ldots, M$, for the observation (i, t) we have:

$$\mathbf{y}_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \boldsymbol{\mu}_i + \boldsymbol{\nu}_t + \mathbf{u}_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{20}$$

where the $M \times K$ matrix of explanatory variables is $\mathbf{X}_{it} = \mathrm{diag}[\mathbf{x}_{1it}, \ldots, \mathbf{x}_{Mit}]$ and the $K \times 1$ vector of parameters is $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_M)'$ and where $\boldsymbol{\mu}_i \equiv (\mu_{1i}, \ldots, \mu_{Mi})'$, $\boldsymbol{\nu}_t \equiv (\nu_{1t}, \ldots, \nu_{Mt})'$, and $\mathbf{u}_{it} \equiv (u_{1it}, \ldots, u_{Mit})'$. If we do not have cross-equation restrictions, we can assume $\mathrm{E}(u_{mit}|\mathbf{x}_{1it}, \mathbf{x}_{2it}, \ldots, \mathbf{x}_{Mit}) = 0$, and then $\mathrm{E}(y_{mit}|\mathbf{x}_{1it}, \mathbf{x}_{2it}, \ldots, \mathbf{x}_{Mit}) = \mathrm{E}(y_{mit}|\mathbf{x}_{mit}) = \mathbf{x}_{mit}\boldsymbol{\beta}_m$. On the contrary, if we have cross-equation restrictions[9], we can only assume $\mathrm{E}(\mathbf{u}_{it}|\mathbf{x}_{it}) = 0$, where $\mathbf{x}_{it} \equiv (\mathbf{x}_{1it}, \mathbf{x}_{2it}, \ldots, \mathbf{x}_{Mit})$.

With heteroskedasticity on both the individual-specific disturbance and the remainder error, we assume for $i \in \hat{I}_a$:

$$\begin{aligned}
\mathrm{E}\left(\mu_{mi}, \mu_{ji'}\right) &= \begin{cases} \varphi_{a_{mj}} & i = i' \\ 0 & i \neq i', \end{cases} \\
\mathrm{E}\left(\nu_{mt}, \nu_{jt'}\right) &= \begin{cases} \sigma^2_{\nu_{mj}} & t = t' \\ 0 & t \neq t', \end{cases} \\
\mathrm{E}\left(u_{mit}, u_{ji't'}\right) &= \begin{cases} \psi_{a_{mj}} & i = i' \text{ and } t = t' \\ 0 & i \neq i' \text{ and/or } t \neq t'. \end{cases}
\end{aligned} \tag{21}$$

Let us consider the $NM \times 1$ vector $\boldsymbol{\mu} \equiv (\boldsymbol{\mu}'_1, \ldots, \boldsymbol{\mu}'_N)'$, the $TM \times 1$ vector $\boldsymbol{\nu} \equiv (\boldsymbol{\nu}'_1, \ldots, \boldsymbol{\nu}'_T)'$, and the $nM \times 1$ vector $\mathbf{u} \equiv (\mathbf{u}'_{11}, \mathbf{u}'_{12}, \ldots, \mathbf{u}'_{1T_1}, \mathbf{u}'_{21}, \ldots, \mathbf{u}'_{NT_N})'$. Since the $M \times 1$ vectors $\mathbf{u}_{it} \sim$

[9] As Biørn (2004) suggests, with cross-equations restrictions we can redefine $\boldsymbol{\beta}$ as the complete $K \times 1$ coefficient vector (without duplication) and the $M \times K$ regression matrix as $\mathbf{X}_{it} = (\mathbf{x}'_{1it}, \mathbf{x}'_{2it}, \ldots, \mathbf{x}'_{Mit})'$, where the k^{th} element of the $1 \times k_m$ vector $\mathbf{x}_{mit}$ either contains the observation on the variable in the m^{th} equation which corresponds to the k^{th} coefficient in $\boldsymbol{\beta}$ or is zero if the k^{th} coefficient does not occur in the m^{th} equation.

$(0, \boldsymbol{\Sigma}_{\psi_a})$, the $M \times 1$ vectors $\boldsymbol{\mu}_i \sim (0, \boldsymbol{\Sigma}_{\varphi_a})$, and the $TM \times 1$ vector $\boldsymbol{\nu} \sim (0, \boldsymbol{\Sigma}_\nu)$, with the $M \times M$ matrices $\boldsymbol{\Sigma}_{\psi_a} = \left[\psi_{a_{mj}}\right]$, $\boldsymbol{\Sigma}_{\varphi_a} = \left[\varphi_{a_{mj}}\right]$, and $\boldsymbol{\Sigma}_\nu = [\sigma^2_{\nu_{mj}}]$, we can assume that the expected values of the vectors $\mathbf{u}_{it}$, $\boldsymbol{\mu}_i$, and $\boldsymbol{\nu}_t$ are zero and their covariance matrices are equal to $\boldsymbol{\Sigma}_{\psi_a}$, $\boldsymbol{\Sigma}_{\varphi_a}$, and $\boldsymbol{\Sigma}_\nu$. It follows that $\mathrm{E}[\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}'_{i't'}] = \delta_{ii'}\boldsymbol{\Sigma}_{\varphi_a} + \delta_{tt'}\boldsymbol{\Sigma}_\nu + \delta_{ii'}\delta_{tt'}\boldsymbol{\Sigma}_{\psi_a}$, with $\delta_{ii'} = 1$ for $i = i'$ and $\delta_{ii'} = 0$ for $i \neq i'$, $\delta_{tt'} = 1$ for $t = t'$ and $\delta_{tt'} = 0$ for $t \neq t'$.

As in Biørn (2004), let us consider the $pM \times 1$ vector of independent variables $\mathbf{y}_{i(p)} \equiv (\mathbf{y}'_{i1}, \ldots, \mathbf{y}'_{ip})'$, the $pM \times K$ matrix of explanatory variables $\mathbf{X}_{i(p)} \equiv (\mathbf{X}'_{i1}, \ldots, \mathbf{X}'_{ip})'$, and the $pM \times 1$ vector of composite error terms $\boldsymbol{\varepsilon}_{i(p)} \equiv (\boldsymbol{\varepsilon}'_{i1}, \ldots, \boldsymbol{\varepsilon}'_{ip})'$ for $i \in I_p$. If we define the $pM \times TM$ matrix $\boldsymbol{\Delta}_{i(p)}$, indicating in which period t the individual i of the group p is observed, and if we consider the $TM \times 1$ vector $\boldsymbol{\nu}$, for the individual $i \in I_p$ we can define the $pM \times 1$ vector $\boldsymbol{\nu}_{i(p)} \equiv \boldsymbol{\Delta}_{i(p)}\boldsymbol{\nu}$ and write the model:

$$\mathbf{y}_{i(p)} = \mathbf{X}_{i(p)}\boldsymbol{\beta} + (\boldsymbol{\iota}_p \otimes \boldsymbol{\mu}_i) + \boldsymbol{\nu}_{i(p)} + \mathbf{u}_{i(p)} = \mathbf{X}_{i(p)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{i(p)}, \quad (22)$$

where $\boldsymbol{\iota}_p$ is a $p \times 1$ vector of ones (see Platoni et al., 2012).

The $pM \times pM$ heteroskedastic variance-covariance matrix of the $pM \times 1$ composite error terms $\boldsymbol{\varepsilon}_{i(a,p)}$ for the individual $i \in I_{a,p}$, with $I_{a,p} = \hat{I}_a \cap I_p$ the set of individuals belonging to stratum a and observed in p periods, is given by:

$$\boldsymbol{\Omega}_{i(a,p)} = \mathbf{E}_p \otimes (\boldsymbol{\Sigma}_{\psi_a} + \boldsymbol{\Sigma}_\nu) + \bar{\mathbf{J}}_p \otimes (\boldsymbol{\Sigma}_{\psi_a} + \boldsymbol{\Sigma}_\nu + p \cdot \boldsymbol{\Sigma}_{\varphi_a}), \quad (23)$$

where $\mathbf{I}_p$ is an identity matrix of dimension p, $\mathbf{J}_p$ a matrix of ones of dimension p, $\mathbf{E}_p = \mathbf{I}_p - \bar{\mathbf{J}}_p$, and $\bar{\mathbf{J}}_p = \frac{\mathbf{J}_p}{p}$. Since $\mathbf{E}_p$ and $\bar{\mathbf{J}}_p$ are symmetric, idempotent, and have orthogonal columns, the inverse of the variance-covariance matrix of the individual i belonging to stratum a and to group p is:

$$\boldsymbol{\Omega}^{-1}_{i(a,p)} = \mathbf{E}_p \otimes (\boldsymbol{\Sigma}_{\psi_a} + \boldsymbol{\Sigma}_\nu)^{-1} + \bar{\mathbf{J}}_p \otimes (\boldsymbol{\Sigma}_{\psi_a} + \boldsymbol{\Sigma}_\nu + p \cdot \boldsymbol{\Sigma}_{\varphi_a})^{-1}. \quad (24)$$

This specification nests simpler heteroskedastic schemes as well as the homoskedastic case by replacing $\mathbf{\Sigma}_{\varphi_a}$ with $\mathbf{\Sigma}_{\mu}$ and/or $\mathbf{\Sigma}_{\psi_a}$ with $\mathbf{\Sigma}_u$.

If we assume that $\mathbf{\Sigma}_{\psi_a}$, $\mathbf{\Sigma}_{\varphi_a}$, and $\mathbf{\Sigma}_{\nu}$ are known, then in the heteroskedastic case we can write the *GLS* estimator for the $K \times 1$ vector of parameters $\boldsymbol{\beta}$ as the problem of minimizing:

$$\sum_{p=1}^{T}\sum_{a=1}^{A}\sum_{i\in I_{a,p}} \boldsymbol{\varepsilon}'_{i(a,p)}\mathbf{\Omega}^{-1}_{i(a,p)}\boldsymbol{\varepsilon}_{i(a,p)}. \tag{25}$$

If we apply *GLS* on the observations for the individuals observed p times we obtain:

$$\hat{\boldsymbol{\beta}}_p^{GLS} = \left(\sum_{a=1}^{A}\sum_{i\in I_{a,p}} \mathbf{X}'_{i(a,p)}\mathbf{\Omega}^{-1}_{i(a,p)}\mathbf{X}_{i(a,p)}\right)^{-1} \sum_{a=1}^{A}\sum_{i\in I_{a,p}} \mathbf{X}'_{i(a,p)}\mathbf{\Omega}^{-1}_{i(a,p)}\mathbf{y}_{i(a,p)}, \tag{26}$$

while the full *GLS* estimator is:

$$\hat{\boldsymbol{\beta}}^{GLS} = \left(\sum_{p=1}^{T}\sum_{a=1}^{A}\sum_{i\in I_{a,p}} \mathbf{X}'_{i(a,p)}\mathbf{\Omega}^{-1}_{i(a,p)}\mathbf{X}_{i(a,p)}\right)^{-1} \sum_{p=1}^{T}\sum_{a=1}^{A}\sum_{i\in I_{a,p}} \mathbf{X}'_{i(a,p)}\mathbf{\Omega}^{-1}_{i(a,p)}\mathbf{y}_{i(a,p)}, \tag{27}$$

where $\mathbf{X}_{i(a,p)}$ is the $pM \times K$ matrix of explanatory variables related to individual $i \in I_{a,p}$.

3.2. Estimation of the Covariance Matrices

The next step is to find an appropriate technique to estimate the components of the variance-covariance matrices of the two-way *SUR* system $\mathbf{\Sigma}_{\psi_a}$, $\mathbf{\Sigma}_{\varphi_a}$, and $\mathbf{\Sigma}_{\nu}$. This can be achieved adopting either the *QUE* procedure suggested by Wansbeek and Kapteyn

(1989) for the homoskedastic single-equation case or the within-between (WB) procedure suggested by Biørn (2004) for the homoskedastic one-way SUR system. In the following sub-sections we modify both procedures making them suitable for the heteroskedastic two-way SUR system.

The QUE Procedure

The QUE procedure considers the $n \times 1$ residuals $\mathbf{e}_m \equiv \mathbf{y}_m - \mathbf{X}_m \hat{\boldsymbol{\beta}}_m^W$ from the W estimator in (4) for the equation $m = 1, \ldots, M$, where $\mathbf{X}_m$ is a matrix of dimension $n \times (k_m - 1)$. If we assume that the $n \times k_m$ matrix $\mathbf{X}_m$ contains a vector of ones, then we have to define the $n \times 1$ consistent centered residuals $\mathbf{f}_m \equiv \mathbf{E}_n \cdot \mathbf{e}_m = \mathbf{e}_m - \bar{e}_m$ (see Wansbeek and Kapteyn, 1989).

With heteroskedasticity, we can obtain the adapted QUEs for $\boldsymbol{\Psi}_{mj}$, $\boldsymbol{\Phi}_{mj}$, and $\sigma^2_{\nu_{mj}}$ by equating:

$$\begin{aligned}
q_{a(n)_{mj}} &\equiv \mathbf{f}_j' \mathbf{Q}_{[\Delta]} \mathbf{H}_a' \mathbf{H}_a \mathbf{Q}_{[\Delta]} \mathbf{f}_m \\
&\rightarrow \sum_{a=1}^{A} q_{a(n)_{mj}} = q_{n_{mj}} \equiv \mathbf{f}_j' \mathbf{Q}_{[\Delta]} \mathbf{f}_m, \\
q_{a(N)_{mj}} &\equiv \mathbf{f}_{a_j}' \bar{\mathbf{J}}_{\hat{n}_a} \mathbf{f}_{a_m} \rightarrow \sum_{a=1}^{A} q_{a(N)_{mj}} = q_{N_{mj}} \equiv \mathbf{f}_j' \boldsymbol{\Delta}_\mu \boldsymbol{\Delta}_N^{-1} \boldsymbol{\Delta}_\mu' \mathbf{f}_m, \\
q_{T_{mj}} &\equiv \mathbf{f}_j' \boldsymbol{\Delta}_\nu \boldsymbol{\Delta}_T^{-1} \boldsymbol{\Delta}_\nu' \mathbf{f}_m,
\end{aligned} \tag{28}$$

to their expected values (see Wansbeek and Kapteyn, 1989; Davis, 2002). The expressions in (28) can be further detailed as already done in (A.3) for the expressions in (11).

With the $n \times n$ matrix $\mathbf{M}_m \equiv \mathbf{I}_n - \mathbf{X}_m (\mathbf{X}_m' \mathbf{Q}_{[\Delta]} \mathbf{X}_m)^{-1} \mathbf{X}_m' \mathbf{Q}_{[\Delta]}$ (and then by definition $\mathbf{e}_m = \mathbf{M}_m \mathbf{y}_m = \mathbf{M}_m \boldsymbol{\varepsilon}_m$ and $\mathbf{f}_m \mathbf{f}_j' = \mathbf{E}_n \mathbf{e}_m \mathbf{e}_j' \mathbf{E}_n = \mathbf{E}_n \mathbf{M}_m \boldsymbol{\Omega}_{mj} \mathbf{M}_j' \mathbf{E}_n$), the expected value of $q_{a(n)_{mj}}$ is:

$$\begin{aligned}
\mathrm{E}\left(q_{a(n)_{mj}}\right) &= \mathrm{tr}\left(\mathbf{H}_a \mathbf{Q}_{[\Delta]} \mathbf{E}_n \mathbf{M}_m \boldsymbol{\Omega}_{mj} \mathbf{M}_j' \mathbf{E}_n \mathbf{Q}_{[\Delta]} \mathbf{H}_a'\right) \\
&= \left(\hat{n}_a - \hat{N}_a - \tau_a\right) \cdot \psi_{a_{mj}} - \left(k_{a_m} + k_{a_j} - k_{a_{mj}}\right) \cdot \bar{\psi}_{mj},
\end{aligned} \tag{29}$$

where $k_{a_{mj}} \equiv \mathrm{tr}[(\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_j(\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_j)^{-1}\mathbf{X}'_j \mathbf{Q}_{[\Delta]}\mathbf{H}'_a\mathbf{H}_a\mathbf{Q}_{[\Delta]}\mathbf{X}_m]$, with $\sum_{a=1}^{A} k_{a_{mj}} = k_{mj}$ and $k_{mj} \equiv \mathrm{tr}[(\mathbf{X}'_m \mathbf{Q}_{[\Delta]}\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_j(\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_j)^{-1}\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_m]$, and $\bar{\psi}_{mj} \approx \sigma^2_{u_{mj}}$ is obtained by equating $q_{n_{mj}}$ to its expected value (see Platoni et al., 2012):

$$\mathrm{E}\left(q_{n_{mj}}\right) = [n - N - (T-1) - (k_m - 1) - (k_j - 1) + k_{mj}] \cdot \sigma^2_{u_{mj}}. \tag{30}$$

Hence, the estimator of $\psi_{a_{mj}}$ is:

$$\hat{\psi}_{a_{mj}} = \frac{q_{a(n)_{mj}} + \left(k_{a_m} + k_{a_j} - k_{a_{mj}}\right) \cdot \hat{\sigma}^2_{u_{mj}}}{\hat{n}_a - \hat{N}_a - \tau_a}. \tag{31}$$

The expected value of $q_{a(N)_{mj}}$ is:

$$\begin{aligned}\mathrm{E}\big(q_{a(N)_{mj}}\big) &= \mathrm{tr}\left(\bar{\mathbf{J}}_{\hat{n}_a}\mathbf{H}_a\mathbf{E}_n\mathbf{M}_m\mathbf{\Omega}_{mj}\mathbf{M}'_j\mathbf{E}_n\mathbf{H}'_a\right) \\ &= \left(k_{N_{a_{mj}}} - k_{0_{a_{mj}}} + \tfrac{\hat{n}_a}{n} \cdot k_{0_{mj}} + \tfrac{\hat{n}_a}{n}\right) \cdot \bar{\psi}_{mj} \\ &+ \left(\hat{N}_a - 2 \cdot \tfrac{\hat{n}_a}{n}\right) \cdot \psi_{a_{mj}} + \tfrac{\hat{n}_a}{n} \cdot \lambda_\mu \cdot \bar{\varphi}_{mj} \\ &+ (\hat{n}_a - 2 \cdot \lambda_{\mu_a}) \cdot \varphi_{a_{mj}} + \left(\hat{N}_a - 2 \cdot \lambda_{\nu_a} + \tfrac{\hat{n}_a}{n} \cdot \lambda_\nu\right) \cdot \sigma^2_{\nu_{mj}},\end{aligned} \tag{32}$$

where $k_{a_{mj}} \equiv \mathrm{tr}[(\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_j(\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_j)^{-1}\mathbf{X}'_j \mathbf{Q}_{[\Delta]}\mathbf{H}'_a\mathbf{H}_a\mathbf{Q}_{[\Delta]}\mathbf{X}_m]$, with $\sum_{a=1}^{A} k_{a_{mj}} = k_{mj}$ and $k_{mj} \equiv \mathrm{tr}[(\mathbf{X}'_m \mathbf{Q}_{[\Delta]}\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_j(\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_j)^{-1}\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_m]$, and $\bar{\psi}_{mj} \approx \sigma^2_{u_{mj}}$ is obtained by equating $q_{n_{mj}}$ to its expected value (see Platoni et al., 2012):

$$\begin{aligned}\mathrm{E}\left(q_{N_{mj}}\right) = &\left(N + k_{N_{mj}} - k_{0_{mj}} - 1\right) \cdot \sigma^2_u \\ &+ (n - \lambda_\mu) \cdot \sigma^2_{\mu_{mj}} + (N - \lambda_\nu) \cdot \sigma^2_{\nu_{mj}}, \\ \mathrm{E}\left(q_{T_{mj}}\right) = &\left(T + k_{T_{mj}} - k_{0_{mj}} - 1\right) \cdot \sigma^2_u \\ &+ (T - \lambda_\mu) \cdot \sigma^2_{\mu_{mj}} + (n - \lambda_\nu) \cdot \sigma^2_{\nu_{mj}},\end{aligned} \tag{33}$$

with $k_{N_{mj}} \equiv \mathrm{tr}[(\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_j)^{-1}\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_m(\mathbf{X}'_m\mathbf{Q}_{[\Delta]}\mathbf{X}_m)^{-1}\mathbf{X}'_m\mathbf{\Delta}_\mu \mathbf{\Delta}_N\mathbf{\Delta}'_\mu\mathbf{X}_j]$ and $k_{T_{mj}} \equiv \mathrm{tr}[(\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_j)^{-1}\mathbf{X}'_j\mathbf{Q}_{[\Delta]}\mathbf{X}_m(\mathbf{X}'_m\mathbf{Q}_{[\Delta]}$

$\mathbf{X}_m)^{-1}\mathbf{X}'_m\boldsymbol{\Delta}_\nu\boldsymbol{\Delta}_T\boldsymbol{\Delta}'_\nu\mathbf{X}_j]$. Hence, the estimator of $\varphi_{a_{mj}}$ is:

$$\begin{aligned}\hat{\varphi}_{a_{mj}} =& \frac{q_{a(N)_{mj}} - \frac{\hat{n}_a}{n}\cdot\lambda_\mu\cdot\hat{\sigma}^2_{\mu_{mj}} - \left(\hat{N}_a - 2\cdot\lambda_{\nu_a} + \frac{\hat{n}_a}{n}\cdot\lambda_\nu\right)\cdot\hat{\sigma}^2_{\nu_{mj}}}{\hat{n}_a - 2\cdot\lambda_{\mu_a}} \\ &+ \frac{-\left(\hat{N}_a - 2\cdot\frac{\hat{n}_a}{n}\right)\cdot\hat{\psi}_{a_{mj}}}{\hat{n}_a - 2\cdot\lambda_{\mu_a}} \\ &+ \frac{-\left(k_{N_{a_{mj}}} - k_{0_{a_{mj}}} + \frac{\hat{n}_a}{n}\cdot k_{0_{mj}} + \frac{\hat{n}_a}{n}\right)\cdot\hat{\sigma}^2_{u_{mj}}}{\hat{n}_a - 2\cdot\lambda_{\mu_a}}. \end{aligned} \tag{34}$$

As in the single-equation case, simpler heteroskedastic scheme (i.e., heteroskedasticity only on the individual-specific disturbance and on the remainder error) can be obtained combining results for the general scheme with those for the homoskedastic case, although when we consider the case of heteroskedasticity only on the individual-specific disturbance the expected value of $q_{a(N)_{mj}}$ is obtained differently as:

$$\begin{aligned}\mathrm{E}\left(q_{a(N)_{mj}}\right) =& (\hat{n}_a - 2\cdot\lambda_{\mu_a})\cdot\varphi_{a_{mj}} + \frac{\hat{n}_a}{n}\cdot\lambda_\mu\cdot\bar{\varphi}_{mj} \\ &+ \left(\hat{N}_a - 2\cdot\lambda_{\nu_a} + \frac{\hat{n}_a}{n}\cdot\lambda_\nu\right)\cdot\sigma^2_{\nu_{mj}} \\ &+ \left(\hat{N}_a + k_{N_{a_{mj}}} - k_{0_{a_{mj}}} + \frac{\hat{n}_a}{n}\cdot k_{0_{mj}} - \frac{\hat{n}_a}{n}\right)\cdot\sigma^2_{u_{mj}} \end{aligned} \tag{35}$$

and, therefore,

$$\begin{aligned}\hat{\varphi}_{a_{mj}} =& \frac{q_{a(N)_{mj}} - \frac{\hat{n}_a}{n}\cdot\lambda_\mu\cdot\hat{\sigma}^2_{\mu_{mj}} - \left(\hat{N}_a - 2\cdot\lambda_{\nu_a} + \frac{\hat{n}_a}{n}\cdot\lambda_\nu\right)\cdot\hat{\sigma}^2_{\nu_{mj}}}{\hat{n}_a - 2\cdot\lambda_{\mu_a}} \\ &+ \frac{-\left(\hat{N}_a + k_{N_{a_{mj}}} - k_{0_{a_{mj}}} + \frac{\hat{n}_a}{n}\cdot k_{0_{mj}} - \frac{\hat{n}_a}{n}\right)\cdot\hat{\sigma}^2_{u_{mj}}}{\hat{n}_a - 2\cdot\lambda_{\mu_a}}. \end{aligned} \tag{36}$$

The WB Procedure

With heteroskedastic two-way systems of equations, the $M \times M$ matrices of within individuals, between individuals, and between times (co)variations in the $\boldsymbol{\varepsilon}$'s of the M equations are the following:

$$
\begin{aligned}
\mathbf{W}_{\varepsilon} &= \sum_{a=1}^{A} \mathbf{W}_{\varepsilon_a} = \sum_{a=1}^{A} \sum_{i\in \hat{I}_a} \sum_{t=1}^{T_i} \left(\boldsymbol{\varepsilon}_{it} - \bar{\boldsymbol{\varepsilon}}_{i\cdot} - \bar{\boldsymbol{\varepsilon}}_{\cdot t}\right)\left(\boldsymbol{\varepsilon}_{it} - \bar{\boldsymbol{\varepsilon}}_{i\cdot} - \bar{\boldsymbol{\varepsilon}}_{\cdot t}\right)', \\
\mathbf{B}_{\varepsilon}^{C} &= \sum_{a=1}^{A} \mathbf{B}_{\varepsilon_a}^{C} = \sum_{a=1}^{A} \sum_{i\in \hat{I}_a} T_i \left(\bar{\boldsymbol{\varepsilon}}_{i\cdot} - \bar{\boldsymbol{\varepsilon}}\right)\left(\bar{\boldsymbol{\varepsilon}}_{i\cdot} - \bar{\boldsymbol{\varepsilon}}\right)', \\
\mathbf{B}_{\varepsilon}^{T} &= \sum_{t=1}^{T} N_t \left(\bar{\boldsymbol{\varepsilon}}_{\cdot t} - \bar{\boldsymbol{\varepsilon}}\right)\left(\bar{\boldsymbol{\varepsilon}}_{\cdot t} - \bar{\boldsymbol{\varepsilon}}\right)',
\end{aligned}
\tag{37}
$$

where for each equation m we have $\bar{\varepsilon}_{mi\cdot} = \frac{\sum_{t=1}^{T_i} \varepsilon_{mit}}{T_i}$, $\bar{\varepsilon}_{m\cdot t} = \frac{\sum_{i=1}^{N_t} \varepsilon_{mit}}{N_t}$, and $\bar{\varepsilon}_m = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} \varepsilon_{mit}}{n} = \frac{\sum_{i=1}^{N}(T_i \cdot \bar{\varepsilon}_{mi\cdot})}{n}$ or $\bar{\varepsilon}_m = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N_t} \varepsilon_{mit}}{n} = \frac{\sum_{t=1}^{T}(N_t \cdot \bar{\varepsilon}_{m\cdot t})}{n}$.

Because the $\mathbf{u}_{it}$'s, the $\boldsymbol{\mu}_i$'s, and the $\boldsymbol{\nu}_t$'s are independent, from the equations in (37) we can write:

$$
\begin{aligned}
&\mathrm{E}\left(\mathbf{W}_{\varepsilon_a}\right) = \mathrm{E}\left(\mathbf{W}_{u_a}\right), \\
&\mathrm{E}\left(\mathbf{B}_{\varepsilon_a}^{C}\right) = \mathrm{E}\left(\mathbf{B}_{\mu_a}^{C}\right) + \mathrm{E}\left(\mathbf{B}_{u_a}^{C}\right), \\
&\mathrm{E}\left(\mathbf{B}_{\varepsilon}^{T}\right) = \mathrm{E}\left(\mathbf{B}_{\nu}^{T}\right) + \mathrm{E}\left(\mathbf{B}_{u}^{T}\right),
\end{aligned}
\tag{38}
$$

where the within individuals (co)variation is:

$$
\begin{aligned}
\mathbf{W}_{u_a} &= \sum_{i\in \hat{I}_a} \sum_{t=1}^{T_i} \left(\mathbf{u}_{it} - \bar{\mathbf{u}}_{i\cdot} - \bar{\mathbf{u}}_{\cdot t}\right)\left(\mathbf{u}_{it} - \bar{\mathbf{u}}_{i\cdot} - \bar{\mathbf{u}}_{\cdot t}\right)' \\
&= \sum_{i\in \hat{I}_a} \sum_{t=1}^{T_i} \mathbf{u}_{it}\mathbf{u}_{it}' - \sum_{i\in \hat{I}_a} T_i \bar{\mathbf{u}}_{i\cdot}\bar{\mathbf{u}}_{i\cdot}' - \sum_{i\in \hat{I}_a} \sum_{t=1}^{T_i} \bar{\mathbf{u}}_{\cdot t}\bar{\mathbf{u}}_{\cdot t}',
\end{aligned}
\tag{39}
$$

the between individuals (co)variations are:

$$
\begin{aligned}
\mathbf{B}_{\mu_a}^{C} &= \sum_{i\in \hat{I}_a} T_i \left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}\right)' = \sum_{i\in \hat{I}_a} T_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i' - \sum_{i\in \hat{I}_a} T_i \bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}', \\
\mathbf{B}_{u_a}^{C} &= \sum_{i\in \hat{I}_a} T_i \left(\bar{\mathbf{u}}_{i\cdot} - \bar{\mathbf{u}}\right)\left(\bar{\mathbf{u}}_{i\cdot} - \bar{\mathbf{u}}\right)' = \sum_{i\in \hat{I}_a} T_i \bar{\mathbf{u}}_{i\cdot}\bar{\mathbf{u}}_{i\cdot}' - \sum_{i\in \hat{I}_a} T_i \bar{\mathbf{u}}\bar{\mathbf{u}}',
\end{aligned}
\tag{40}
$$

and the between times (co)variations, as in the homoskedastic case, are:

$$\mathbf{B}_{\nu}^{T}=\sum_{t=1}^{T}N_t\left(\boldsymbol{\nu}_t-\bar{\boldsymbol{\nu}}\right)\left(\boldsymbol{\nu}_t-\bar{\boldsymbol{\nu}}\right)'=\sum_{t=1}^{T}N_t\boldsymbol{\nu}_t\boldsymbol{\nu}_t'-n\bar{\boldsymbol{\nu}}\bar{\boldsymbol{\nu}}', \\ \mathbf{B}_{u}^{T}=\sum_{t=1}^{T}N_t\left(\bar{\mathbf{u}}_{\cdot t}-\bar{\mathbf{u}}\right)\left(\bar{\mathbf{u}}_{\cdot t}-\bar{\mathbf{u}}\right)'=\sum_{t=1}^{T}N_t\bar{\mathbf{u}}_{\cdot t}\bar{\mathbf{u}}_{\cdot t}'-n\bar{\mathbf{u}}\bar{\mathbf{u}}', \tag{41}$$

where $\bar{u}_{mi\cdot}=\frac{\sum_{t=1}^{T_i}u_{mit}}{T_i}$, $\bar{u}_{m\cdot t}=\frac{\sum_{i=1}^{N_t}u_{mit}}{N_t}$, $\bar{u}_m=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}u_{mit}}{n}=\frac{\sum_{i=1}^{N}(T_i\cdot\bar{u}_{mi\cdot})}{n}$ or $\bar{u}_m=\frac{\sum_{t=1}^{T}\sum_{i=1}^{N_t}u_{mit}}{n}=\frac{\sum_{t=1}^{T}(N_t\cdot\bar{u}_{m\cdot t})}{n}$, $\bar{\mu}_m=\frac{\sum_{i=1}^{N}(T_i\cdot\mu_{mi})}{n}$, and $\bar{\nu}_m=\frac{\sum_{t=1}^{T}(N_t\cdot\nu_{mt})}{n}$ (see Biørn, 2004; Platoni et al., 2012).

Since for $i\in\hat{I}_a$ we have $\mathrm{E}(\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}_{i't'}')=\delta_{ii'}\boldsymbol{\Sigma}_{\varphi_a}+\delta_{tt'}\boldsymbol{\Sigma}_{\nu}+\delta_{ii'}\delta_{tt'}\boldsymbol{\Sigma}_{\psi_a}$, where $\mathrm{E}(\mathbf{u}_{it}\mathbf{u}_{i't'}')=\delta_{ii'}\delta_{tt'}\boldsymbol{\Sigma}_{\psi_a}$, $\mathrm{E}(\boldsymbol{\mu}_i\boldsymbol{\mu}_{i'}')=\delta_{ii'}\boldsymbol{\Sigma}_{\varphi_a}$, and $\mathrm{E}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t'}')=\delta_{tt'}\boldsymbol{\Sigma}_{\nu}$, it follows that $\mathrm{E}(\bar{\mathbf{u}}_{i\cdot}\bar{\mathbf{u}}_{i\cdot}')=\frac{\boldsymbol{\Sigma}_{\psi_a}}{T_i}$, $\mathrm{E}(\bar{\mathbf{u}}_{\cdot t}\bar{\mathbf{u}}_{\cdot t}')=\frac{\sum_{i\in I_t}\boldsymbol{\Sigma}_{\psi_a}}{N_t^2}\simeq\frac{\bar{\boldsymbol{\Sigma}}_{\psi}}{N_t}\approx\frac{\boldsymbol{\Sigma}_u}{N_t}$, with I_t the set of individuals observed in period t, $\mathrm{E}(\bar{\mathbf{u}}\bar{\mathbf{u}}')=\frac{\sum_{i=1}^{N}(T_i\cdot\boldsymbol{\Sigma}_{\psi_a})}{n^2}=\frac{\bar{\boldsymbol{\Sigma}}_{\psi}}{n}\approx\frac{\boldsymbol{\Sigma}_u}{n}$, $\mathrm{E}(\bar{\boldsymbol{\mu}}\bar{\boldsymbol{\mu}}')=\frac{\sum_{i=1}^{N}(T_i^2\cdot\boldsymbol{\Sigma}_{\varphi_a})}{(\sum_{i=1}^{N}T_i)^2}=\frac{\sum_{i=1}^{N}T_i^2}{n^2}\cdot\bar{\boldsymbol{\Sigma}}_{\varphi}\approx\frac{\sum_{i=1}^{N}T_i^2}{n^2}\cdot\boldsymbol{\Sigma}_{\mu}$, and $\mathrm{E}(\bar{\boldsymbol{\nu}}\bar{\boldsymbol{\nu}}')=\frac{\sum_{t=1}^{T}N_t^2}{n^2}\cdot\boldsymbol{\Sigma}_{\nu}$. Hence, the $M\times M$ matrices

$$\hat{\boldsymbol{\Sigma}}_{\psi_a}=\frac{\mathbf{W}_{\varepsilon_a}+\sum_{i\in\hat{I}_a}\sum_{t=1}^{T_i}\frac{1}{N_t}\cdot\hat{\boldsymbol{\Sigma}}_u}{\hat{n}_a-\hat{N}_a}, \tag{42}$$

with $\sum_{a=1}^{A}\sum_{i\in\hat{I}_a}\sum_{t=1}^{T_i}\frac{1}{N_t}=T$, and

$$\hat{\boldsymbol{\Sigma}}_{\varphi_a}=\frac{\mathbf{B}_{\varepsilon_a}^{C}+\sum_{i\in\hat{I}_a}\frac{T_i}{n}\cdot\sum_{j=1}^{N}\frac{T_j^2}{n}\cdot\hat{\boldsymbol{\Sigma}}_{\mu}-\hat{N}_a\cdot\hat{\boldsymbol{\Sigma}}_{\psi_a}+\sum_{i\in\hat{I}_a}\frac{T_i}{n}\cdot\hat{\boldsymbol{\Sigma}}_u}{\sum_{i\in\hat{I}_a}T_i} \tag{43}$$

would be unbiased estimators of $\boldsymbol{\Sigma}_{\psi_a}$ and $\boldsymbol{\Sigma}_{\varphi_a}$ if the $\boldsymbol{\varepsilon}$'s were known. Both the estimators of $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_{\mu}$ and the estimator of

$\mathbf{\Sigma}_\nu$ are derived as in the homoskedastic case:

$$\hat{\mathbf{\Sigma}}_u = \frac{\mathbf{W}_\varepsilon}{n - N - T}, \ \hat{\mathbf{\Sigma}}_\mu = \frac{\mathbf{B}_\varepsilon^C - (N-1) \cdot \hat{\mathbf{\Sigma}}_u}{n - \sum_{i=1}^{N} \frac{T_i^2}{n}}, \text{ and}$$
$$\hat{\mathbf{\Sigma}}_\nu = \frac{\mathbf{B}_\varepsilon^T - (T-1) \cdot \hat{\mathbf{\Sigma}}_u}{n - \sum_{t=1}^{T} \frac{N_t^2}{n}}, \tag{44}$$

that would be unbiased estimators of $\mathbf{\Sigma}_u$, $\mathbf{\Sigma}_\mu$, and $\mathbf{\Sigma}_\nu$ if the $\boldsymbol{\varepsilon}$'s were known (see Biørn, 2004; Platoni et al., 2012).

Again, simpler heteroskedastic scheme (i.e., heteroskedasticity only on the individual-specific disturbance and on the remainder error) can be obtained combining results for the general scheme with those for the homoskedastic case, although when we consider the case of heteroskedasticity only on the individual-specific disturbance the estimator is:

$$\hat{\mathbf{\Sigma}}_{\varphi_a} = \frac{\mathbf{B}_{\varepsilon_a}^C + \sum_{i \in \hat{I}_a} \frac{T_i}{n} \cdot \sum_{j=1}^{N} \frac{T_j^2}{n} \cdot \hat{\mathbf{\Sigma}}_\mu - \left(\hat{N}_a - \sum_{i \in \hat{I}_a} \frac{T_i}{n} \right) \cdot \hat{\mathbf{\Sigma}}_u}{\sum_{i \in \hat{I}_a} T_i}, \tag{45}$$

that would be an unbiased estimator of $\mathbf{\Sigma}_{\varphi_a}$ if the $\boldsymbol{\varepsilon}$'s were known[10].

As Biørn (2004) suggested, in empirical applications consistent residuals can replace $\boldsymbol{\varepsilon}$'s in (37) to obtain consistent estimates of $\mathbf{\Sigma}_{\psi_a}$, $\mathbf{\Sigma}_{\varphi_a}$, and $\mathbf{\Sigma}_\nu$. Since the *QUE* procedure is based on the *W* residuals, for coherence also in the *WB* procedure we consider the $M \times 1$ residuals $\mathbf{e}_{it} \equiv \mathbf{y}_{it} - \mathbf{X}_{it}\hat{\boldsymbol{\beta}}^W$ from the *W* estimator in (4) for the individual i in period t, where $\mathbf{X}_{it}$ is

[10] Alternative computations of the estimators (42), (43), and (45) are provided in Appendix A.3.

a matrix of dimension $M \times (K - M)$. As above, if we assume that the $M \times K$ matrix $\mathbf{X}_{it}$ in (20) always contains M vectors of ones (a vector of ones for each equation m), then we have to define the $M \times 1$ consistent centered residuals $\mathbf{f}_{it} = \mathbf{e}_{it} - \bar{\mathbf{e}}$, where $\bar{e}_m = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} e_{mit}}{n} = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N_t} e_{mit}}{n}$. Therefore, the $M \times M$ matrices of within individuals, between individuals, and between times (co)variations in the $\mathbf{f}$'s of the different M equations are the following:

$$
\begin{aligned}
\mathbf{W}_f &= \sum_{a=1}^{A} \mathbf{W}_{f_a} = \sum_{a=1}^{A} \sum_{i \in \hat{I}_a} \sum_{t=1}^{T_i} \left(\mathbf{f}_{it} - \bar{\mathbf{f}}_{i\cdot} - \bar{\mathbf{f}}_{\cdot t}\right)\left(\mathbf{f}_{it} - \bar{\mathbf{f}}_{i\cdot} - \bar{\mathbf{f}}_{\cdot t}\right)', \\
\mathbf{B}_f^C &= \sum_{a=1}^{A} \mathbf{B}_{f_a}^C = \sum_{a=1}^{A} \sum_{i \in \hat{I}_a} T_i \left(\bar{\mathbf{f}}_{i\cdot} - \bar{\mathbf{f}}\right)\left(\bar{\mathbf{f}}_{i\cdot} - \bar{\mathbf{f}}\right)', \qquad (46) \\
\mathbf{B}_f^T &= \sum_{t=1}^{T} N_t \left(\bar{\mathbf{f}}_{\cdot t} - \bar{\mathbf{f}}\right)\left(\bar{\mathbf{f}}_{\cdot t} - \bar{\mathbf{f}}\right)',
\end{aligned}
$$

where for each equation m we have $\bar{f}_{mi\cdot} = \frac{\sum_{t=1}^{T_i} f_{mit}}{T_i}$, $\bar{f}_{m\cdot t} = \frac{\sum_{i=1}^{N_t} f_{mit}}{N_t}$, and $\bar{f}_m = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} f_{mit}}{n} = \frac{\sum_{i=1}^{N}(T_i \cdot \bar{f}_{mi\cdot})}{n}$ or $\bar{f}_m = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N_t} f_{mit}}{n} = \frac{\sum_{t=1}^{T}(N_t \cdot \bar{f}_{m\cdot t})}{n}$. Given that:

$$
\begin{aligned}
\mathrm{E}\left(\mathbf{W}_{f_a}\right) &= \left(\hat{n}_a - \hat{N}_a\right) \cdot \mathbf{\Sigma}_{\psi_a} - \sum_{i \in \hat{I}_a} \sum_{t \in J_i} \frac{1}{N_t} \cdot \bar{\mathbf{\Sigma}}_{\psi}, \\
\mathrm{E}\left(\mathbf{B}_{f_a}^C\right) &= \sum_{i \in \hat{I}_a} T_i \cdot \mathbf{\Sigma}_{\varphi_a} - \sum_{i \in \hat{I}_a} \frac{T_i}{n} \cdot \sum_{j=1}^{N} \frac{T_j^2}{n} \cdot \bar{\mathbf{\Sigma}}_{\varphi} + \hat{N}_a \cdot \mathbf{\Sigma}_{\psi_a} \\
&\quad - \sum_{i \in \hat{I}_a} \frac{T_i}{n} \cdot \bar{\mathbf{\Sigma}}_{\psi}, \qquad (47) \\
\mathrm{E}\left(\mathbf{B}_f^T\right) &= \left(n - \sum_{t=1}^{T} \frac{N_t^2}{n}\right) \cdot \mathbf{\Sigma}_{\nu} + (T-1) \cdot \bar{\mathbf{\Sigma}}_{\psi},
\end{aligned}
$$

with $\bar{\mathbf{\Sigma}}_{\psi} \approx \mathbf{\Sigma}_u$ and $\bar{\mathbf{\Sigma}}_{\varphi} \approx \mathbf{\Sigma}_{\mu}$, we can conclude that the estimators in (42) and (43), with $\mathbf{W}_{f_a}$ instead of $\mathbf{W}_{\varepsilon_a}$ and $\mathbf{B}_{f_a}^C$ instead of $\mathbf{B}_{\varepsilon_a}^C$ respectively, are consistent estimators of $\mathbf{\Sigma}_{\psi_a}$ and $\mathbf{\Sigma}_{\varphi_a}$. As mentioned above, both the consistent estimators of $\mathbf{\Sigma}_u$

and $\boldsymbol{\Sigma}_\mu$ and the consistent estimator of $\boldsymbol{\Sigma}_\nu$ are derived as in the homoskedastic case (see Biørn, 2004; Platoni et al., 2012). Finally, with heteroskedasticity only on the individual-specific disturbance, we have that:

$$\mathrm{E}\left(\mathbf{B}_{f_a}^C\right) = \begin{array}{l} \sum\limits_{i\in\hat{I}_a} T_i \cdot \boldsymbol{\Sigma}_{\varphi_a} - \sum\limits_{i\in\hat{I}_a} \frac{T_i}{n} \cdot \sum\limits_{j=1}^{N} \frac{T_j^2}{n} \cdot \bar{\boldsymbol{\Sigma}}_\varphi \\ + \left(\hat{N}_a - \sum\limits_{i\in\hat{I}_a} \frac{T_i}{n}\right) \cdot \boldsymbol{\Sigma}_u, \end{array} \tag{48}$$

and therefore the estimator in (45), with $\mathbf{B}_{f_a}^C$ instead of $\mathbf{B}_{\varepsilon_a}^C$, is a consistent estimator of $\boldsymbol{\Sigma}_{\varphi_a}$[11].

4. MONTE CARLO EXPERIMENT

In order to analyze the performances of the proposed techniques, we develop a simple simulation[12] on a three-equation system ($M = 3$). We assume an unbalanced panel with a large number of individuals ($N = 4,000$) extended over a rather long time period ($T = 8$). This should mimic a real world situation of a large unbalanced panel for which the two-way *SUR* system is the appropriate model. The simulated model is:

$$\begin{array}{lllllll} y_1 & = \beta_{10} & +\beta_{11} \cdot x_1 & +\beta_{12} \cdot x_2 & & +\varepsilon_1, \\ y_2 & = \beta_{20} & +\beta_{21} \cdot x_1 & +\beta_{22} \cdot x_2 & +\beta_{23} \cdot x_3 & +\varepsilon_2, \\ y_3 & = \beta_{30} & & +\beta_{32} \cdot x_2 & +\beta_{33} \cdot x_3 & +\varepsilon_3, \end{array}$$

where $\boldsymbol{\beta}_1 = (15, 6, -3)'$, $\boldsymbol{\beta}_2 = (10, -3, 8, -2)'$, and $\boldsymbol{\beta}_3 = (20, -2, 5)'$, implying the cross equations restrictions $\beta_{12} = \beta_{21}$ and $\beta_{23} = \beta_{32}$.

[11] Alternative computations of consistent estimators are provided in Appendix A.3.

[12] The simulation has been implemented with the econometric software *TSP* version 5.1.

Moreover, the experiment is implemented by considering as strata the deciles of the independent variable x_2. The homoskedastic time variance-covariance matrix is:

$$\mathbf{\Sigma}_\nu = \begin{bmatrix} 0.7659 & \text{-}0.0280 & \text{-}0.2242 \\ & 0.7204 & 0.1068 \\ & & 0.7207 \end{bmatrix},$$

while the heteroskedastic variances-covariances $\varphi_{a_{mj}}$ and $\psi_{a_{mj}}$ have been generated from the matrices:

$$\underline{\mathbf{\Sigma}}_\varphi = \begin{bmatrix} 0.8070 & \text{-}0.0056 & 0.2433 \\ & 0.7073 & \text{-}0.0044 \\ & & 0.6928 \end{bmatrix} \text{ and } \underline{\mathbf{\Sigma}}_\psi = \begin{bmatrix} 0.6128 & 0.0571 & 0.1554 \\ & 0.6726 & \text{-}0.0941 \\ & & 0.9218 \end{bmatrix}$$

with $\varphi_{a_{mj}} = \underline{\varphi}_{mj} \cdot \bar{x}^2_{2_a}$ and $\psi_{a_{mj}} = \underline{\psi}_{mj} \cdot \bar{x}^2_{2_a}$, where $\underline{\varphi}_{mj}$ and $\underline{\psi}_{mj}$ are elements of the matrices $\underline{\mathbf{\Sigma}}_\varphi$ and $\underline{\mathbf{\Sigma}}_\psi$ respectively and $\bar{x}_{2_a}$ is the mean of the independent variable x_2 over the decile/stratum a.

Finally, the scalars x_{kit} are generated according to a modified version of the scheme introduced by Nerlove (1971) and used, among others, by Baltagi (1981), Wansbeek and Kapteyn (1989), and Platoni et al. (2012):

$$\begin{aligned} x_{kit} &= 0.1 \cdot t + 0.5 \cdot x_{kit-1} + \omega_{kit}, \quad k = 1, 3 \\ x_{kit} &= 0.1 \cdot t + x_{kit-1} + 0.5 \cdot \omega_{kit}, \quad k = 2 \end{aligned}$$

with ω_{kit} following the uniform distribution $[-\frac{1}{2}, \frac{1}{2}]$ and $x_{ki0} = 5 + 10 \cdot \omega_{ki0}$.

In order to construct the unbalanced panel, we adopt the procedure currently used for rotating panels, in which we have approximately the same number of individuals every year: a fixed percentage of individuals (20% in our case[13]) is replaced each year, but they can re-enter the sample in the following years.

[13] Also in Wansbeek and Kapteyn (1989) each period 20% of the households in the panel is removed randomly.

Thus, for each group p we have the following number of individuals: $\widetilde{N}_1 = 962$, $\widetilde{N}_2 = 769$, $\widetilde{N}_3 = 615$, $\widetilde{N}_4 = 492$, $\widetilde{N}_5 = 394$, $\widetilde{N}_6 = 315$, $\widetilde{N}_7 = 252$, and $\widetilde{N}_8 = 201$ (and then $n = 13,545$).

The results of a 1000-run simulation are shown in Table 1 and Table 2[14].

Table 1 reports average estimates when the homoskedastic specification is employed; the homoskedastic model is estimated under several different specifications (the *FE* one-way and two-way, the *RE* one-way *GLS* and *ML*, the *RE* two-way *GLS* and *ML*, the *SUR* one-way *WB*, and the *SUR* two-way *WB* and *QUE*).

In Table 2, simulation results for the heteroskedastic specifications are reported. We have estimated the six models (the *FE* one-way and two-way robust to heteroskedasticity of unknown form, the *RE* two-way *GLS*, both *WB* and *QUE*, the *SUR* two-way, both *WB* and *QUE*) for three different specifications of the heteroskedastic structure (only on the individual-specific effect μ_i, only on the remainder error u_{it}, and the true specification with both the individual-specific and remainder error terms heteroskedastic). Average estimates for the true parameters are overall similar to those obtained with the homoskedastic specifications, with an average bias within 0.07%. As expected, the *SUR* procedures perform better than the single-equation procedures in all cases, also in those cases with no relevant differences between the *WB* and the *QUE* estimates. Looking at the standard errors of the parameter estimates in the *RE* and *SUR* models, it is evident that accounting for heteroskedasticity will increase efficiency (standard errors in the heteroskedastic models are largely lower than those in the homoskedastic model). Finally, the correct specification (i.e., with both the individual-specific and remainder error terms heteroskedastic) is the most

[14] As in Baltagi and Griffin (1988) and Phillips (2003), negative variance estimates are replaced by zero.

Table 1
Simulation results: means of the estimated parameters and average variances of the error components - homoscedastic case

	True value	FE one-way	FE two-way	RE one-way (GLS)	RE one-way (ML)	RE two-way (GLS)	RE two-way (ML)	SUR one-way WB	SUR two-way WB	SUR two-way QUE
β_{10}	15			14.9991	14.9991	14.9933	14.9934	15.0038	15.0039	15.0039
				(0.1021)	(0.1022)	(0.3130)	(0.2955)	(0.0832)	(0.0840)	(0.0832)
β_{11}	6	5.9988	5.9996	5.9989	5.9989	5.9997	5.9997	5.9990	5.9990	5.9990
		(0.0237)	(0.0228)	(0.0210)	(0.0210)	(0.0204)	(0.0204)	(0.0207)	(0.0209)	(0.0207)
β_{12}	-3	-3.0029	-2.9991	-3.0004	-3.0004	-2.9989	-2.9989	-3.0019	-3.0020	-3.0020
		(0.0476)	(0.0492)	(0.0273)	(0.0273)	(0.0275)	(0.0275)	(0.0174)	(0.0176)	(0.0174)
$\sigma^2_{\mu_{11}}$	11.0511			7.8453	7.8334	7.9299	7.9129	7.9521	8.0087	7.9299
$\sigma^2_{\mu_{12}}$	-0.0767							-0.0690	-0.0695	-0.0674
$\sigma^2_{\mu_{13}}$	3.3318							2.3914	2.3709	2.3932
$\sigma^2_{\nu_{11}}$	0.7659					0.7417	0.6502		0.7440	0.7417
$\sigma^2_{\nu_{12}}$	-0.0280								-0.0276	-0.0276
$\sigma^2_{\nu_{13}}$	-0.2242								-0.2360	-0.2367
$\sigma^2_{u_{11}}$	8.3917	6.7497	6.0254	6.7497	6.7532	6.0254	6.0235	6.7483	6.2082	6.0254
$\sigma^2_{u_{12}}$	0.7819							0.5344	0.5532	0.5601
$\sigma^2_{u_{13}}$	2.1280							1.2914	1.4696	1.5254
β_{20}	10			10.0023	10.0024	9.9851	9.9851	9.9993	9.9991	9.9992
				(0.1020)	(0.1020)	(0.3066)	(0.2896)	(0.1001)	(0.1010)	(0.1001)
β_{21}	-3	-3.0028	-3.0013	-3.0023	-3.0022	-3.0009	-3.0009	-3.0019	-3.0020	-3.0020
		(0.0353)	(0.0338)	(0.0274)	(0.0274)	(0.0265)	(0.0265)	(0.0174)	(0.0176)	(0.0174)
β_{22}	8	7.9897	8.0013	7.9971	7.9971	8.0010	8.0010	7.9974	7.9974	7.9974
		(0.0496)	(0.0516)	(0.0269)	(0.0269)	(0.0271)	(0.0271)	(0.0269)	(0.0271)	(0.0269)
β_{23}	-2	-2.0008	-1.9992	-2.0021	-2.0021	-2.0004	-2.0004	-2.0012	-2.0012	-2.0012
		(0.0353)	(0.0337)	(0.0274)	(0.0274)	(0.0265)	(0.0265)	(0.0177)	(0.0179)	(0.0177)
$\sigma^2_{\mu_{22}}$	9.6858			6.8528	6.8418	6.9327	6.9175	6.9572	7.0091	6.9327
$\sigma^2_{\mu_{23}}$	-0.0603							-0.0515	-0.0456	-0.0550
$\sigma^2_{\nu_{22}}$	0.7204					0.7097	0.6227		0.7117	0.7097
$\sigma^2_{\nu_{23}}$	0.1068								0.1032	0.1032
$\sigma^2_{u_{22}}$	9.2106	7.3063	6.6165	7.3063	7.3090	6.6165	6.6135	7.3040	6.7945	6.6165
$\sigma^2_{u_{23}}$	-1.2886							-0.8257	-0.9030	-0.9276
β_{30}	20			19.9884	19.9883	19.9903	19.9903	19.9952	19.9950	19.9950
				(0.1073)	(0.1074)	(0.3130)	(0.2956)	(0.0836)	(0.0844)	(0.0835)
β_{32}	-2	-1.9972	-1.9991	-1.9989	-1.9989	-1.9993	-1.9993	-2.0012	-2.0012	-2.0012
		(0.0573)	(0.0604)	(0.0287)	(0.0287)	(0.0290)	(0.0290)	(0.0177)	(0.0179)	(0.0177)
β_{33}	5	5.0003	5.0000	5.0011	5.0011	5.0007	5.0007	5.0012	5.0012	5.0012
		(0.0286)	(0.0280)	(0.0243)	(0.0243)	(0.0239)	(0.0239)	(0.0239)	(0.0242)	(0.0239)
$\sigma^2_{\mu_{33}}$	9.4872			6.7199	6.7123	6.8005	6.7920	6.8407	6.8811	6.8005
$\sigma^2_{\nu_{33}}$	0.7207					0.7358	0.6441		0.7378	0.7358
$\sigma^2_{u_{33}}$	12.6231	9.7781	9.0651	9.7781	9.7829	9.0651	9.0622	9.7761	9.2494	9.0651

The numbers in round brackets are the average estimated standard errors. The variances of the estimated parameters are not displayed since they are not significantly different among estimation techniques.

efficient; also note that, at least in our simulation, considering heteroskedasticity only on the individual-specific disturbance does not provide a large gain in efficiency with respect to the homoskedastic specifications.

5. FURTHER EXTENSION: HETEROSKEDASTIC TIME-SPECIFIC EFFECT

Usually panel data are characterized by many individuals and relatively few time periods. Therefore heteroskedasticity on the time-specific error term should not be a frequent issue. However, the procedure proposed in the previous sections can be easily extended also to this case.

In this section we assume heteroskedasticity is also on the time-specific error term; hence, $\mathrm{Var}(\nu_t) = \vartheta_t$ and $\mathrm{Var}(u_{it}) = \psi_{it}$. If the ϑ_t's are unknown, then there is no hope to estimate them from the data: even if the ν_t's were observed, it would be impossible to estimate their variances from only one observation on each time period disturbance (see Mazodier and Trognon, 1978). Moreover, if the remainder error term u_{it} is heteroskedastic not only by the individual dimension, but also by the time dimension, then it would be impossible to estimate the variances ψ_{it}'s from only one observation on each remainder error.

Let us assume there exist meaningful stratifications of observations both with respect to individuals and with respect to time periods. Hence, the unbalanced panel can be characterized not only by A strata of individuals, but also by B strata of time periods (indexed $b = 1, \ldots B$), with $\check{T}_b$ the number of time periods pertaining to stratum b (indexed $\check{t}_b = \check{1}_b, \ldots, \check{T}_b$) and $\check{J}_b$ the set of time periods $t = 1, \ldots, T$ pertaining to stratum b. Therefore, the number of observations related to stratum b is $\check{n}_b = \sum_{\check{t}_b=\check{1}_b}^{\check{T}_b} N_{\check{t}_b}$. Hence, $\sum_{b=1}^{B} \check{T}_b = T$ and $\sum_{b=1}^{B} \check{n}_b = n$.

Furthermore, it is possible to identify $C = A \times B$ sub-strata

Table 2

Simulation results: means of the estimated parameters and average variances of the error components - heteroskedastic cases

				Heteroskedasticity on μ_i			
	True value	FE one-way robust	FE two-way robust	RE two-way WB	RE two-way QUE	SUR two-way WB	SUR two-way QUE
β_{10}	15			14.9974	14.9975	15.0015	15.0015
				(0.0565)	(0.0557)	(0.0603)	(0.0594)
β_{11}	6	5.9988	5.9996	5.9986	5.9986	5.9989	5.9989
		(0.0238)	(0.0228)	(0.0187)	(0.0185)	(0.0200)	(0.0197)
β_{12}	-3	-3.0029	-2.9991	-2.9994	-2.9994	-3.0019	-3.0019
		(0.0479)	(0.0494)	(0.0147)	(0.0145)	(0.0160)	(0.0158)
$\bar{\varphi}^2_{11}$	11.0511			8.0323	7.9643	8.0323	7.9643
$\bar{\varphi}^2_{12}$	-0.0767					-0.0695	-0.0674
$\bar{\varphi}^2_{13}$	3.3318					2.3702	2.3932
$\sigma^2_{\nu_{11}}$	0.7659			0.7440	0.7417	0.7440	0.7417
$\sigma^2_{\nu_{12}}$	-0.0280					-0.0276	-0.0276
$\sigma^2_{\nu_{13}}$	-0.2242					-0.2360	-0.2367
$\sigma^2_{u_{11}}$	8.3917	6.7497	6.0254	6.2082	6.0254	6.2082	6.0254
$\sigma^2_{u_{12}}$	0.7819					0.5532	0.5601
$\sigma^2_{u_{13}}$	2.1280					1.4696	1.5254
β_{20}	10			9.9930	9.9929	9.9937	9.9936
				(0.0613)	(0.0605)	(0.0671)	(0.0662)
β_{21}	-3	-3.0028	-3.0013	-3.0013	-3.0013	-3.0019	-3.0019
		(0.0354)	(0.0337)	(0.0229)	(0.0227)	(0.0160)	(0.0158)
β_{22}	8	7.9897	8.0013	8.0009	8.0009	7.9981	7.9981
		(0.0502)	(0.0518)	(0.0154)	(0.0152)	(0.0240)	(0.0238)
β_{23}	-2	-2.0008	-1.9992	-2.0034	-2.0034	-2.0008	-2.0008
		(0.0354)	(0.0337)	(0.0229)	(0.0227)	(0.0168)	(0.0166)
$\bar{\varphi}^2_{22}$	9.6858			7.0623	6.9968	7.0623	6.9968
$\bar{\varphi}^2_{23}$	-0.0603					-0.0457	-0.0550
$\sigma^2_{\nu_{22}}$	0.7204			0.7117	0.7097	0.7117	0.7097
$\sigma^2_{\nu_{23}}$	0.1068					0.1032	0.1032
$\sigma^2_{u_{22}}$	9.2106	7.3063	6.6165	6.7945	6.6165	6.7945	6.6165
$\sigma^2_{u_{23}}$	-1.2886					-0.9030	-0.9276
β_{30}	20			19.9894	19.9894	19.9928	19.9929
				(0.0665)	(0.0657)	(0.0644)	(0.0635)
β_{32}	-2	-1.9972	-1.9991	-1.9996	-1.9996	-2.0008	-2.0008
		(0.0578)	(0.0605)	(0.0174)	(0.0172)	(0.0168)	(0.0166)
β_{33}	5	5.0003	5.0000	5.0015	5.0015	5.0012	5.0012
		(0.0288)	(0.0280)	(0.0218)	(0.0216)	(0.0232)	(0.0229)
$\bar{\varphi}^2_{33}$	9.4872			6.9988	6.9312	6.9988	6.9312
$\sigma^2_{\nu_{33}}$	0.7207			0.7378	0.7358	0.7378	0.7358
$\sigma^2_{u_{33}}$	12.6231	9.7781	9.0651	9.2494	9.0651	9.2494	9.0651

The numbers in round brackets are the average estimated standard errors. The variances of the estimated parameters are not displayed since they are not significantly different among estimation techniques.

continued on next page

continued from previous page

Simulation results: means of the estimated parameters and average variances of the error components - heteroskedastic cases

	Heteroskedasticity on u_{it}				Heteroskedasticity on μ_i and u_{it}			
	RE two-way WB	RE two-way QUE	SUR two-way WB	SUR two-way QUE	RE two-way WB	RE two-way QUE	SUR two-way WB	SUR two-way QUE
β_{10}	14.9992	14.9991	15.0031	15.0031	14.9990	14.9990	15.0012	15.0011
	(0.0405)	(0.0392)	(0.0746)	(0.0734)	(0.0397)	(0.0384)	(0.0561)	(0.0545)
β_{11}	5.9989	5.9989	5.9993	5.9993	5.9989	5.9989	5.9992	5.9993
	(0.0162)	(0.0158)	(0.0177)	(0.0172)	(0.0156)	(0.0152)	(0.0169)	(0.0164)
β_{12}	-3.0003	-3.0003	-3.0022	-3.0022	-3.0003	-3.0003	-3.0022	-3.0022
	(0.0132)	(0.0129)	(0.0157)	(0.0154)	(0.0131)	(0.0128)	(0.0144)	(0.0140)
$\sigma^2_{\mu_{11}}$; $\bar{\varphi}^2_{11}$	8.0087	7.9299	8.0087	7.9299	8.0114	7.9352	8.0114	7.9352
$\sigma^2_{\mu_{12}}$; $\bar{\varphi}^2_{12}$			-0.0695	-0.0674			-0.0689	-0.0669
$\sigma^2_{\mu_{13}}$; $\bar{\varphi}^2_{13}$			2.3709	2.3932			2.3716	2.3945
$\sigma^2_{\nu_{11}}$	0.7440	0.7417	0.7440	0.7417	0.7440	0.7417	0.7440	0.7417
$\sigma^2_{\nu_{12}}$			-0.0276	-0.0276			-0.0276	-0.0276
$\sigma^2_{\nu_{13}}$			-0.2360	-0.2367			-0.2360	-0.2367
$\bar{\psi}^2_{11}$	6.2033	6.0202	6.2033	6.0202	6.2033	6.0202	6.2033	6.0202
$\bar{\psi}^2_{12}$			0.5527	0.5596			0.5527	0.5596
$\bar{\psi}^2_{13}$			1.4682	1.5241			1.4682	1.5241
β_{20}	9.9925	9.9924	10.0007	10.0010	9.9924	9.9923	9.9952	9.9952
	(0.0434)	(0.0422)	(0.0881)	(0.0868)	(0.0426)	(0.0414)	(0.0612)	(0.0592)
β_{21}	-3.0019	-3.0019	-3.0022	-3.0022	-3.0018	-3.0018	-3.0022	-3.0022
	(0.0197)	(0.0192)	(0.0157)	(0.0154)	(0.0189)	(0.0184)	(0.0144)	(0.0140)
β_{22}	8.0004	8.0003	7.9965	7.9964	8.0003	8.0002	7.9973	7.9972
	(0.0137)	(0.0134)	(0.0251)	(0.0248)	(0.0136)	(0.0133)	(0.0223)	(0.0219)
β_{23}	-2.0025	-2.0024	-2.0007	-2.0007	-2.0022	-2.0021	-2.0004	-2.0004
	(0.0197)	(0.0192)	(0.0160)	(0.0157)	(0.0189)	(0.0184)	(0.0147)	(0.0143)
$\sigma^2_{\mu_{22}}$; $\bar{\varphi}^2_{22}$	7.0091	6.9327	7.0091	6.9327	7.0127	6.9384	7.0127	6.9384
$\sigma^2_{\mu_{23}}$; $\bar{\varphi}^2_{23}$			-0.0457	-0.0550			-0.0466	-0.0559
$\sigma^2_{\nu_{22}}$	0.7117	0.7097	0.7117	0.7097	0.7117	0.7097	0.7117	0.7097
$\sigma^2_{\nu_{23}}$			0.1032	0.1032			0.1032	0.1032
$\bar{\psi}^2_{22}$	6.7890	6.6107	6.7890	6.6107	6.7890	6.6107	6.7890	6.6107
$\bar{\psi}^2_{23}$			-0.9021	-0.9268			-0.9021	-0.9268
β_{30}	19.9905	19.9905	19.9944	19.9944	19.9906	19.9906	19.9922	19.9921
	(0.0461)	(0.0449)	(0.0742)	(0.0730)	(0.0450)	(0.0438)	(0.0571)	(0.0554)
β_{32}	-1.9996	-1.9996	-2.0007	-2.0007	-1.9996	-1.9996	-2.0004	-2.0004
	(0.0153)	(0.0150)	(0.0160)	(0.0157)	(0.0152)	(0.0149)	(0.0147)	(0.0143)
β_{33}	5.0010	5.0010	5.0006	5.0006	5.0010	5.0010	5.0008	5.0009
	(0.0185)	(0.0181)	(0.0201)	(0.0197)	(0.0178)	(0.0174)	(0.0191)	(0.0187)
$\sigma^2_{\mu_{33}}$; $\bar{\varphi}^2_{33}$	6.8811	6.8005	6.8811	6.8005	6.8868	6.8084	6.8868	6.8084
$\sigma^2_{\nu_{33}}$	0.7378	0.7358	0.7378	0.7358	0.7378	0.7358	0.7378	0.7358
$\bar{\psi}^2_{33}$	9.2418	9.0573	9.2418	9.0573	9.2418	9.0573	9.2418	9.0573

The numbers in round brackets are the average estimated standard errors. The variances of the estimated parameters are not displayed since they are not significantly different among estimation techniques.

of observations; each sub-stratum is characterized by a total of $\breve{n}_c$ observations, with $\breve{N}_c$ individuals (indexed $\breve{i}_c = \breve{1}_c, \ldots, \breve{N}_c$) observed over $\breve{T}_c$ periods (indexed $\breve{t}_c = \breve{1}_c, \ldots, \breve{T}_c$). Hence, the number of observations within the sub-stratum c is $\breve{n}_c = \sum_{\breve{i}_c=\breve{1}_c}^{\breve{N}_c} T_{\breve{i}_c} = \sum_{\breve{t}_c=\breve{1}_c}^{\breve{T}_c} N_{\breve{t}_c}$.

Using the $n \times B$ matrix $\mathbf{\Delta}_\beta$ of indicator variables denoting observations on strata B, we can define the $B \times B$ diagonal matrix $\mathbf{\Delta}_B \equiv \mathbf{\Delta}'_\beta \mathbf{\Delta}_\beta$ (diagonal elements correspond to the $\breve{n}_b$'s) and the $B \times T$ matrix of zeros and ones $\mathbf{\Delta}_{BT} \equiv \mathbf{\Delta}'_\beta \mathbf{\Delta}_\nu \mathbf{\Delta}_T^{-1}$, indicating the absence or presence of a time period in a certain stratum (notice that $\mathbf{\Delta}'_\beta \mathbf{\Delta}_\nu$ is a matrix of zeros and $N_{\breve{t}_b}$'s).

Moreover, using the $n \times C$ matrix $\mathbf{\Delta}_\gamma$ of indicator variables denoting observations on sub-strata C, we can define the $C \times C$ diagonal matrix $\mathbf{\Delta}_C \equiv \mathbf{\Delta}'_\gamma \mathbf{\Delta}_\gamma$ (diagonal elements correspond to the $\breve{n}_c$'s), the $C \times N$ matrix of zeros and ones $\mathbf{\Delta}_{CN} \equiv \mathbf{\Delta}'_\gamma \mathbf{\Delta}_\mu \mathbf{\Delta}_N^{-1}$, indicating the absence or presence of an individual in a certain sub-stratum (notice that $\mathbf{\Delta}'_\gamma \mathbf{\Delta}_\mu$ is a matrix of zeros and $T_{\breve{i}_c}$'s), and the $C \times T$ matrix of zeros and ones $\mathbf{\Delta}_{CT} \equiv \mathbf{\Delta}'_\gamma \mathbf{\Delta}_\nu \mathbf{\Delta}_T^{-1}$, indicating the absence or presence of a time period in a certain sub-stratum (notice that $\mathbf{\Delta}'_\gamma \mathbf{\Delta}_\nu$ is a matrix of zeros and $N_{\breve{t}_c}$'s).

Therefore, variances $\text{Var}(\mu_i)$'s are assumed to be constant within strata A, i.e., $\text{Var}(\mu_i) = \varphi_a$, variances $\text{Var}(\nu_t)$'s are assumed to be constant within strata B, i.e., $\text{Var}(\nu_t) = \vartheta_b$, and variances $\text{Var}(u_{it})$'s are assumed to be constant within sub-strata C, i.e., $\text{Var}(u_{it}) = \psi_c$. Hence, the approach presented in the previous sections can be easily extended to the case in which the time-specific effect is also heteroskedastic.

6. CONCLUSION

The use of panel data is becoming very popular in applied econometrics, since large data sets including many individuals observed for several periods are increasingly accessible and manageable. Most of these data sets are unbalanced panels, since very often not all the individuals are observed over the whole time period. In estimating single-equation or system of equations *ECM*s on these data, the heteroskedasticity problem may be very common, especially when individuals differ in size.

In this paper, we have derived suitable *ECM* estimators for heteroskedastic two-way single-equation and *SUR* systems (with cross-equations restrictions) on unbalanced panel data. Our simulations show that such estimators substantially improve estimation efficiency as compared to the case where heteroskedasticity is not taken into account, especially when both the individual-specific and remainder error components are heteroskedastic. Among the various estimators used in this analysis, the *QUE* procedures and the *WB* procedures perform equally well (both in the single-equation and the *SUR* specifications), resulting in a similar average bias on the estimates of the variance-covariance matrices.

APPENDIX

A.1. Alternative Robust Standard Errors

Since $u_{it} \sim (0, \psi_a)$, it is possible to obtain robust standard errors also by stacking the observations for each stratum a, and then by writing:

$$\begin{aligned}
\widetilde{\mathbf{y}}_{a(A)} = & \left[\mathrm{diag}[\mathbf{E}_{T_{\hat{\imath}_a}}] - \left(\mathbf{E}_{T_{\hat{1}_a}}\mathbf{D}_{\hat{1}_a}, \ldots, \mathbf{E}_{T_{\hat{N}_a}}\mathbf{D}_{\hat{N}_a}\right)'\right. \\
& \left.\mathbf{Q}^{-}\left(\mathbf{D}'_{\hat{1}_a}\mathbf{E}_{T_{\hat{1}_a}}, \ldots, \mathbf{D}'_{\hat{N}_a}\mathbf{E}_{T_{\hat{N}_a}}\right)\right]\mathbf{y}_{a(A)}, \\
\widetilde{\mathbf{X}}_{a(A)} = & \left[\mathrm{diag}[\mathbf{E}_{T_{\hat{\imath}_a}}] - \left(\mathbf{E}_{T_{\hat{1}_a}}\mathbf{D}_{\hat{1}_a}, \ldots, \mathbf{E}_{T_{\hat{N}_a}}\mathbf{D}_{\hat{N}_a}\right)'\right. \\
& \left.\mathbf{Q}^{-}\left(\mathbf{D}'_{\hat{1}_a}\mathbf{E}_{T_{\hat{1}_a}}, \ldots, \mathbf{D}'_{\hat{N}_a}\mathbf{E}_{T_{\hat{N}_a}}\right)\right]\mathbf{X}_{a(A)}.
\end{aligned} \tag{A.1}$$

Therefore, we can compute the $\hat{n}_a \times 1$ vector $\widetilde{\mathbf{e}}_{a(A)} = \widetilde{\mathbf{y}}_{a(A)} - \widetilde{\mathbf{X}}_{a(A)} \hat{\boldsymbol{\beta}}^W$ and the robust asymptotic variance-covariance matrix of $\hat{\boldsymbol{\beta}}^W$ is estimated by:

$$\begin{aligned}\operatorname{Var}\left(\hat{\boldsymbol{\beta}}^W\right) = & \left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right)^{-1} \sum_{a=1}^{A} \left(\widetilde{\mathbf{X}}'_{a(A)} \widetilde{\mathbf{e}}_{a(A)} \widetilde{\mathbf{e}}'_{a(A)} \widetilde{\mathbf{X}}_{a(A)}\right) \\ & \left(\mathbf{X}'\mathbf{Q}_{[\Delta]}\mathbf{X}\right)^{-1}. \end{aligned} \tag{A.2}$$

A.2. Adapted *QUE*s in (11)

The expressions in (11) can be further detailed as:

$$\begin{aligned}
q_{a(n)} \equiv & \left[\mathbf{f}'_a - \bar{\mathbf{f}}'_{N\cdot}\boldsymbol{\Delta}'_{\mu_a} - \left(\bar{\mathbf{f}}'_{\cdot T}\boldsymbol{\Delta}_T - \bar{\mathbf{f}}'_{N\cdot}\boldsymbol{\Delta}'_{TN}\right)\mathbf{Q}^- \right. \\
& \left.\left(\boldsymbol{\Delta}_{\nu_a} - \boldsymbol{\Delta}_{\mu_a}\boldsymbol{\Delta}_N^{-1}\boldsymbol{\Delta}'_{TN}\right)'\right]\left[\mathbf{f}_a - \boldsymbol{\Delta}_{\mu_a}\bar{\mathbf{f}}_{N\cdot}\right. \\
& \left. - \left(\boldsymbol{\Delta}_{\nu_a} - \boldsymbol{\Delta}_{\mu_a}\boldsymbol{\Delta}_N^{-1}\boldsymbol{\Delta}'_{TN}\right)\mathbf{Q}^- \left(\bar{\mathbf{f}}'_{\cdot T}\boldsymbol{\Delta}_T - \bar{\mathbf{f}}'_{N\cdot}\boldsymbol{\Delta}'_{TN}\right)'\right], \\
q_n \equiv & \underset{1\times n}{\mathbf{f}'}\underset{n\times 1}{\mathbf{f}} - \underset{1\times N}{\bar{\mathbf{f}}'_{N\cdot}}\underset{N\times N}{\boldsymbol{\Delta}_N}\underset{N\times 1}{\bar{\mathbf{f}}_{N\cdot}} - \left(\underset{1\times T}{\bar{\mathbf{f}}'_{\cdot T}}\underset{T\times T}{\boldsymbol{\Delta}_T} - \underset{1\times N}{\bar{\mathbf{f}}'_{N\cdot}}\underset{N\times T}{\boldsymbol{\Delta}'_{TN}}\right)\underset{T\times T}{\mathbf{Q}^-} \\
& \left(\underset{1\times T}{\bar{\mathbf{f}}'_{\cdot T}}\underset{T\times T}{\boldsymbol{\Delta}_T} - \underset{1\times N}{\bar{\mathbf{f}}'_{N\cdot}}\underset{N\times T}{\boldsymbol{\Delta}'_{TN}}\right)', \\
q_{a(N)} \equiv & \sum_{\hat{\imath}_a=\hat{1}_a}^{\hat{N}_a} T_{\hat{\imath}_a}\cdot \bar{f}^2_{\hat{\imath}_a\cdot} = \sum_{i\in \hat{I}_a} T_i \cdot \bar{f}^2_{i\cdot}, \\
q_N \equiv & \underset{1\times N}{\bar{\mathbf{f}}'_{N\cdot}}\underset{N\times N}{\boldsymbol{\Delta}_N}\underset{N\times 1}{\bar{\mathbf{f}}_{N\cdot}} = \sum_{i=1}^{N} T_i \cdot \bar{f}^2_{i\cdot} = \sum_{a=1}^{A}\sum_{\hat{\imath}_a=\hat{1}_a}^{\hat{N}_a} T_{\hat{\imath}_a}\cdot \bar{f}^2_{\hat{\imath}_a\cdot} \\
= & \sum_{a=1}^{A}\sum_{i\in\hat{I}_a} T_i\cdot \bar{f}^2_{i\cdot}, \\
q_T \equiv & \underset{1\times T}{\bar{\mathbf{f}}'_{\cdot T}}\underset{T\times T}{\boldsymbol{\Delta}_T}\underset{T\times 1}{\bar{\mathbf{f}}_{\cdot T}} = \sum_{t=1}^{T} N_t \cdot \bar{f}^2_{t\cdot},
\end{aligned} \tag{A.3}$$

where the elements of the $N \times 1$ matrix $\bar{\mathbf{f}}_{N\cdot}$ are $\bar{f}_{i\cdot} = \frac{\sum_{t=1}^{T_i} f_{it}}{T_i}$, the elements of the $T \times 1$ matrix $\bar{\mathbf{f}}_{\cdot T}$ are $\bar{f}_{\cdot t} = \frac{\sum_{i=1}^{N_t} f_{it}}{N_t}$, $\boldsymbol{\Delta}_{\mu_a} = \mathbf{H}_a\boldsymbol{\Delta}_\mu$, and $\boldsymbol{\Delta}_{\nu_a} = \mathbf{H}_a\boldsymbol{\Delta}_\nu$.

A.3. Alternative (Consistent) Estimators of $\mathbf{\Sigma}_{\psi_a}$ and $\mathbf{\Sigma}_{\varphi_a}$

Estimators of $\mathbf{\Sigma}_{\psi_a}$ in (42) and $\mathbf{\Sigma}_{\varphi_a}$ in (43) and (45) can be calculated by identifying the individuals belonging to stratum a through not the set $\hat{I}_a$ but the index $\hat{\imath}_a = \hat{1}_a, \ldots, \hat{N}_a$:

$$\hat{\mathbf{\Sigma}}_{\psi_a} = \frac{\mathbf{W}_{\varepsilon_a} + \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \sum_{t=1}^{T_{\hat{\imath}_a}} \frac{1}{N_t} \cdot \hat{\mathbf{\Sigma}}_u}{\hat{n}_a - \hat{N}_a}, \tag{A.4}$$

$$\hat{\mathbf{\Sigma}}_{\varphi_a} = \frac{\mathbf{B}^C_{\varepsilon_a} + \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n} \cdot \sum_{i=1}^{N} \frac{T_i^2}{n} \cdot \hat{\mathbf{\Sigma}}_\mu - \hat{N}_a \cdot \hat{\mathbf{\Sigma}}_{\psi_a} + \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n} \cdot \hat{\mathbf{\Sigma}}_u}{\sum_{\hat{\imath}_a=1}^{\hat{N}_a} T_{\hat{\imath}_a}}, \tag{A.5}$$

$$\hat{\mathbf{\Sigma}}_{\varphi_a} = \frac{\mathbf{B}^C_{\varepsilon_a} + \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n} \cdot \sum_{i=1}^{N} \frac{T_i^2}{n} \cdot \hat{\mathbf{\Sigma}}_\mu - \left(\hat{N}_a - \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n}\right) \cdot \hat{\mathbf{\Sigma}}_u}{\sum_{\hat{\imath}_a=1}^{\hat{N}_a} T_{\hat{\imath}_a}}. \tag{A.6}$$

Besides, the expected values of the $M \times M$ matrices of within individuals and between individuals (co)variations in the $\mathbf{f}$'s of the different M equations can be written by identifying the individuals belonging to stratum a through not the set $\hat{I}_a$ as in (47) but the index $\hat{\imath}_a = \hat{1}_a, \ldots, \hat{N}_a$:

$$\begin{aligned} \mathrm{E}(\mathbf{W}_{f_a}) &= \left(\hat{n}_a - \hat{N}_a\right) \cdot \mathbf{\Sigma}_{\psi_a} - \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \sum_{t \in J_{\hat{\imath}_a}} \frac{1}{N_t} \cdot \bar{\mathbf{\Sigma}}_\psi, \\ \mathrm{E}(\mathbf{B}^C_{f_a}) &= \sum_{\hat{\imath}_a=1}^{\hat{N}_a} T_{\hat{\imath}_a} \cdot \mathbf{\Sigma}_{\varphi_a} - \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n} \cdot \sum_{i=1}^{N} \frac{T_i^2}{n} \cdot \bar{\mathbf{\Sigma}}_\varphi \\ &\quad + \hat{N}_a \cdot \mathbf{\Sigma}_{\psi_a} - \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n} \cdot \bar{\mathbf{\Sigma}}_\psi. \end{aligned} \tag{A.7}$$

Therefore, the estimators in (A.4) and (A.5), with $\mathbf{W}_{f_a}$ instead of $\mathbf{W}_{\varepsilon_a}$ and $\mathbf{B}^C_{f_a}$ instead of $\mathbf{B}^C_{\varepsilon_a}$ respectively, are consistent estimators of $\mathbf{\Sigma}_{\psi_a}$ and $\mathbf{\Sigma}_{\varphi_a}$. Finally, with heteroskedasticity only on the individual-specific disturbance, the expected values of the $M \times M$ matrices of between individuals (co)variations in the $\mathbf{f}$'s of the different M equations in (48) becomes:

$$\mathrm{E}\left(\mathbf{B}_{f_a}^{C}\right) = \sum_{\hat{\imath}_a=1}^{\hat{N}_a} T_{\hat{\imath}_a} \cdot \boldsymbol{\Sigma}_{\varphi_a} - \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n} \cdot \sum_{i=1}^{N} \frac{T_i^2}{n} \cdot \bar{\boldsymbol{\Sigma}}_{\varphi} + \left(\hat{N}_a - \sum_{\hat{\imath}_a=1}^{\hat{N}_a} \frac{T_{\hat{\imath}_a}}{n}\right) \cdot \boldsymbol{\Sigma}_u, \qquad \text{(A.8)}$$

and therefore the estimator in (A.6), with $\mathbf{B}_{f_a}^{C}$ instead of $\mathbf{B}_{\varepsilon_a}^{C}$, is a consistent estimator of $\boldsymbol{\Sigma}_{\varphi_a}$.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6):716-723.

Arellano, M. (1987). Computing robust standard errors for within groups estimators. *Oxford Bulletin of Economics and Statistics* 49(4):431-434.

Baltagi, B. H. (1980). On seemingly unrelated regressions with error components. *Econometrica* 48(6):1547-1551.

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(1):21-49.

Baltagi, B. H. (1985). Pooling cross-sections with unequal time series lengths. *Economics Letters* 18(2-3):133-136.

Baltagi, B. H. (1988). An alternative heteroskedastic error components model (problem 88.2.2.). *Econometric Theory* 4(2):349-350.

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*, 5th edition. Chichester, UK: Wiley and Sons.

Baltagi, B. H., Bresson, G., Pirotte, A. (2005). Adaptive estimation of heteroskedastic error component models. *Econometric Reviews* 24(1):39-58.

Baltagi, B. H., Bresson, G., Pirotte, A. (2006). Joint LM test for homoskedasticity in a one-way error component model. *Journal of Econometrics* 134(2):401-417.

Baltagi, B. H., Griffin, J. M. (1988). A generalized error component model with heteroskedastic disturbances. *International Economic Review* 29(4):745-753.

Bester, C. A., Hansen, C. B. (2016). Grouped effects estimators in fixed effects models. *Journal of Econometrics* 190(1):197-208.

Biørn, E. (1981). Estimating economic relations from incomplete cross-section/time-series data. *Journal of Econometrics* 16(2):221-236.

Biørn, E. (2004). Regression systems for unbalanced panel data: A stepwise maximum likelihood procedure. *Journal of Econometrics* 122(2):281-291.

Bresson, G., Hsiao, C., Pirotte, A. (2006). Heteroskedasticity and random coefficient model on panel data. Working Papers ERMES, No. 0601, 51 p.

Bresson, G., Hsiao, C., Pirotte, A. (2011). Assessing the contribution of R&D to total factor productivity: A Bayesian approach to account for heterogeneity and heteroskedasticity. *Advances in Statistical Analysis* 95(4):435-452.

Chib, S., (2008). Panel Data Modeling and Inference: A Bayesian Primer. In Mátyás, L., Sevestre, P., eds. *The Econometrics of Panel Data*, 3rd edition, 479-515, book series *Advanced Studies in Theoretical and Applied Econometrics*, vol. 46. Berlin, GE: Springer-Verlag.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics* 106(1):67-95.

Hsiao, C., Pesaran, M. H. (2004). Random coefficient panel data models. IZA Discussion Paper Series, No. 1236, 39 p.

Lejeune, B. (1996). A full heteroskedastic one-way error components model for incomplete panel: Maximum likelihood estimation and Lagrange multiplier testing. CORE Discussion Paper, Université Catholique de Louvain, No. 1996/006, 28 p.

Lejeune, B. (2004). A full heteroscedastic one-way error components model allowing for unbalanced panel: Pseudo-maximum likelihood estimation and specification testing. CORE Discussion Paper, Université Catholique de Louvain, No. 2004/76, 37 p.

Li, Q., Stengos, T. (1994). Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. *International Economic Review* 35(4):981-1000.

Magnus, J. R. (1982). Multivariate error components analysis of linear and non-linear regression models by maximum likelihood. *Journal of Econometrics* 19(2-3):239-285.

Mazodier, P., Trognon, A. (1978). Heteroskedasticity and stratification in error components models. *Annales de l'INSEE* 30-31:451-482.

Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econometrica* 39(2):359-382.

Neyman, J., Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* 16(1):1-32.

Phillips, R. F. (2003). Estimation of a stratified error-components model. *International Economic Review* 44(2):501-521.

Platoni, S., Sckokai, P., Moro, D. (2012). A note on two-way ECM estimation of SUR systems on unbalanced panel data. *Econometric Reviews* 31(2):119-141.

Randolph, W. C. (1988). A transformation for heteroskedastic error components regression models. *Economics Letters* 27(4):349-354.

Rao, P. S. R. S., Kaplan, J., Cochran, W. C. (1981). Estimators for the one-way random effects model with unequal error variances. *Journal of the American Statistical Association* 76(373):89-97.

Roy, N. (2002). Is adaptive estimation useful for panel models with heteroskedasticity in the individual specific error component? Some Monte Carlo evidence. *Econometric Reviews* 21(2):189-203.

Verbon, H. A. A. (1980). Testing for heteroskedasticity in a model of seemingly unrelated regression equations with variance component. *Economics Letters* 5(2):149-153.

Wang, H.-J., Ho, C.-W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics* 157(2):286-296.

Wansbeek, T. (1989). An alternative heteroskedastic error components model (problem 88.2.2.). *Econometric Theory* 5(2):326.

Wansbeek, T., Kapteyn, A. (1989). Estimation of the error-components model with incomplete panels. *Journal of Econometrics* 41(3):341-361.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section And Panel Data*, 2nd edition. Cambridge, Massachusset, US; London, England, UK: The MIT Press.

Printed by
Gi&Gi srl - Triuggio (MB)
September 2016

9788834332733