Heteroscedastic stratified two-way EC models of single equations and SUR systems

SILVIA PLATONI¹, LAURA BARBIERI¹, DANIELE MORO², AND PAOLO SCKOKAI²

¹Dipartimento di Scienze economiche e sociali and
²Dipartimento di Economia agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza, Italy

A relevant issue in panel data estimation is heteroscedasticity, which often occurs when the sample 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 the heteroscedastic stratified two-way error component (EC) models of both single equations and seemingly unrelated regressions (SUR) systems (with cross-equations restrictions) on unbalanced panel data. The derived heteroscedastic estimators of both single equations and SUR systems improve the estimation efficiency.

KEYWORDS. Unbalanced panel, EC model, SUR, heteroscedasticity.

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 (EC) model 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 EC models have been extended to the econometric treatment of unbalanced panel data: Bistri (1981) and Baltagi (1985)
discussed the single-equation one-way EC model, 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 heteroscedasticity, which often occurs when the sample is large and observations differ in “size characteristic” (i.e., the level of the variables). Under this perspective, heteroscedasticity 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 heteroscedasticity may be simultaneously present (see Lejeune, 1996, 2004). This means that heteroscedasticity is the rule rather than the exception when dealing with individual data concerning households or firms. Assuming homoscedastic disturbances when heteroscedasticity 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 heteroscedasticity.

Several authors have analyzed the problem of heteroscedasticity in balanced panel data, usually considering a single-equation regression model with one-way disturbances $e_{it} = \mu_i + u_{it}$.

Baltagi and Griffin (1988) are concerned with the estimation of a random-effect (RE) model allowing for heteroscedasticity on the individual-specific error term $\text{var}(\mu_i) = \phi^2_i$. In contrast, Rao et al. (1981), Magnus (1982), Baltagi (1988), and Wansbeek (1989) adopt a symmetrically opposite specification allowing for heteroscedasticity on the remainder error term $\text{var}(u_{it}) = \psi^2_i$.

As Mazodier and Trognon (1978) pointed out, if the $\phi^2_i$’s are unknown, then there is no hope to estimate them from the data: even if the $\mu_i$’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 (see Phillips, 2003; Baltagi, 2013). Furthermore, also the models allowing for heteroscedasticity 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 EC models) or, if the variables that determine heteroscedasticity are known, to specify parametric variance functions (i.e., adaptive estimation of heteroscedasticity of unknown form).

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 heteroscedasticity is modeled following both the RCMs on panel data proposed by Hsiao and Pesaran (2004) and Chib (2008) and the general heteroscedastic one-way EC model proposed by Randolph (1988), who assumes that both the individual-specific term $\mu_i$ and the remainder error term $u_{it}$ are heteroscedastic.

2 Neyman and Scott (1948) study maximum likelihood (ML) estimation of 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.
Mazodier and Trognon (1978) proposed a stratified two-way EC model, i.e., \( \varepsilon_{it} = \mu_i + \nu_t + u_{it} \), on balanced panels in which both the individual-specific effect \( \mu_i \) and time-specific effect \( \nu_t \) 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 EC model, again on balanced panels, where the variances of the individual-specific effect \( \mu_i \) 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 heteroscedastic one-way EC model using balanced panel data where heteroscedasticity is placed on the remainder error term, and hence, \( \text{var}(u_{it}|x_{it}) = \psi^2_{it} \). Later, Roy (2002) derives a similar adaptive estimator where heteroscedasticity is placed on the individual-specific term rather than the remainder disturbance, and hence, \( \text{var}(\mu_i|x_{it}) = \phi^2_i \).

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

Indeed, if heteroscedasticity 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 heteroscedastic, and it may be difficult to argue that only one component of the error term is heteroscedastic 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 heteroscedastic single-equation one-way EC model, assuming that both the individual-specific and remainder error terms are heteroscedastic, i.e., \( \text{var}(\mu_i) = \phi^2_i \) and \( E(uu^T) = \text{diag}(\psi^2_{it}) \). Lejeune (1996, 2004) is concerned with the estimation and specification testing of a full heteroscedastic one-way EC model, 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 homoscedasticity against the alternative of heteroscedasticity both in the individual-specific term \( \mu_i \) 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 (SURs) for balanced panels; Bistron (2004) proposed a parsimonious technique to estimate one-way SUR systems on unbalanced panel data; Platoni et al. (2012) extended the procedure suggested by Bistron (2004) to the two-way case. Although heteroscedasticity is a frequent and relevant issue also in the multi-equation models applied to (unbalanced) panel data, to our knowledge very few papers concerning heteroscedastic SUR systems have been published. A relevant exception is Verbon (1980), who derived a LM test for heteroscedasticity in a model of SUR equations for

---

3 Throughout the paper, all vectors and matrices are in non-italics.
balanced panels.

In order to fill this gap in the literature, this paper extends previous results to the estimation of the heteroscedastic stratified two-way EC model, i.e., \( \varepsilon_{it} = \mu_i + \nu_t + u_{it} \), on unbalanced panel data\(^4\) in case of both single equations and SUR systems (with cross-equations restrictions). The individual-specific effect \( \mu_i \) and remainder error term \( u_{it} \) variances and covariances are constant within strata, but they are allowed to change across strata. Indeed, the variance and covariance estimations in two-way SUR systems are implemented, starting from the extension of the two-way single-equation EC model from the homoscedastic to the heteroscedastic stratified case. Moreover, the estimation is 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 heteroscedastic two-way estimation for single equations, Section 3 extends the analysis to the corresponding estimation for SUR systems. Finally, Section 4 draws some conclusions.

2. Heteroscedastic Single-Equation Two-Way EC Model

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:

\[
y_{it} = x_{it}^T \beta + \mu_i + \nu_t + u_{it} = x_{it}^T \beta + \varepsilon_{it}, \tag{1}
\]

where \( x_{it} \) is a \( k \times 1 \) vector of explanatory variables and \( \beta \) a \( k \times 1 \) vector of parameters, \( \mu_i \) the individual-specific effect, \( \nu_t \) the time-specific effect, and \( u_{it} \) the remainder error term; in the RE model \( \varepsilon_{it} \) is the composite error term.

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

\[
y = X\beta + \Delta \mu \mu + \Delta \nu \nu + u = X\beta + \varepsilon, \tag{2}
\]

where \( X \) 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 \( N_a \) the number of individuals belonging to stratum \( a \). Moreover, the number of observations related to stratum \( a \) is \( n_a = \sum_{i \in I_a} T_i \), with \( I_a \) the set of individuals belonging to

\(^4\) The estimation procedures proposed here can definitely be applied also to balanced panel data.

\(^5\) 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.
stratum $a$.\(^6\)

Using the $n \times A$ matrix $\Delta_a$ of indicator variables denoting observations on strata, we can define the $A \times A$ diagonal matrix $\Delta \equiv \Delta^T \Delta_a$ (diagonal elements correspond to the $n_a$’s) and the $A \times N$ matrix of zeros and ones $\Delta_{AN} \equiv \Delta^T \Delta_a \Delta^{-1}$, indicating the absence or presence of an individual in a certain stratum (notice that $\Delta^T \Delta$ is a matrix of zeros and $T_i$’s for $i \in I_a$).

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, heteroscedasticity on the individual-specific disturbance implies $\mu_i \sim (0, \varphi^2_a)$, while heteroscedasticity on the remainder error term implies $u_{it} \sim (0, \psi^2_{it})$.

2.1 Robust two-way FE

In the FE model the individual-specific term $\mu_i$ and the time-specific term $v_t$ are parameters to be estimated. Therefore, heteroscedasticity is placed only on the remainder error $u_{it}$ by assuming $u_{it} \sim (0, \psi^2_{it})$. The Within ($W$) estimator\(^7\) is:

$$\hat{\beta}^W = (X^T Q_{\Delta} X)^{-1} X^T Q_{\Delta} Y,$$

(3)

where the $n \times n$ matrix $Q_{\Delta}$ on which the two-way $EC$ model transformation is based is:

$$Q_{\Delta} = Q_A - P_B = Q_A - Q_A \Delta_{\nu} Q^{-}\Delta^T \Delta A,$$

(4)

with $Q_A = I_n - P_A$, $P_A = \Delta_{\mu} \Delta_{\nu}^{-1} \Delta^T \mu$, $Q = \Delta^T \Delta \Delta_{\nu}$, and $Q^{-}$ the generalized inverse (see Wansbeek and Kapteyn, 1989; Davis, 2002).\(^8\)

Under the assumptions of strict exogeneity, consistency, homoscedasticity and no serial correlation (see assumptions FE.1-FE.3 in Appendix A of Platoni et al., 2012), the $W$ estimator is consistent and asymptotically normal (see Wooldridge, 2010) with

$$\text{var} (\hat{\beta}^W) = \hat{\sigma}_{\mu}^2 (X^T Q_{\Delta} X)^{-1},$$

(5)

where $\hat{\sigma}_{\mu}^2$ is the estimator of $\sigma_{\mu}^2$. However, relaxing the homoscedasticity assumption (see assumption FE.3 in Appendix A), the expression (5) 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 $EC$ model, and proposed also by Baltagi (2013). If we stack

---

\(^6\) Note that $\sum_{a=1}^{A} N_a = N$ and $\sum_{a=1}^{A} n_a = n$.

\(^7\) The number of explanatory variables, obviously without the intercept, is $k - 1$.

\(^8\) For a $FE$ model the number of fixed-effect parameters $\mu_1, \ldots, \mu_{N}$ and $v_1, \ldots, v_T$ 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 $v_1, \ldots, v_T$ 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 $\mu_i$ 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 in linear models, is to transform the model by the $W$ transformation (i.e., $y_{it}$ and the $(k - 1) \times 1$ vector $x_{it}$ are demeaned), as we have done in deriving our estimation.

5
the observations for each individual $i$, we can write:

$$
\begin{align*}
\tilde{y}_i & = (E_{it} - E_{it}D_iQ^{-1}D_i^TE_{it}) y_i, \\
X_i & = (E_{it} - E_{it}D_iQ^{-1}D_i^TE_{it}) X_i,
\end{align*}
$$

where $E_{it} = I_{it} - \bar{I}_t$, with $I_{it}$ an identity matrix of dimension $T_i$, $\bar{I}_t = \frac{I_T}{T}$, and $J_{it}$ a matrix of ones of dimension $T_i$, and $D_i$ is the $T_i \times T$ matrix obtained from the $T \times T$ identity matrix $I_T$ 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 $\tilde{e}_i = \tilde{y}_i - \bar{X}_i \hat{\beta}^W$ and the robust asymptotic variance-covariance matrix of $\hat{\beta}^W$ is:

$$
\text{var} (\hat{\beta}^W) = (X^TQ\Delta X)^{-1}\sum_{i=1}^{N} X_i^T \tilde{e}_i \tilde{e}_i^T (X^TQ\Delta X)^{-1}.
$$

However, since $u_{it} \sim (0, \psi_i^2)$, it is possible to obtain robust standard errors also by stacking the observations for each stratum $a$, as described later in Appendix C.

### 2.2 GLS estimation

In the $RE$ model, not only the remainder error $u_{it}$, but also the individual-specific error $\mu_i$ and the time-specific error $v_t$ are random variables.

If we assume that the variances of $\mu_i$, $v_t$, and $u_{it}$ are known, then the general least squares (GLS) estimator\(^9\) for $\beta$, obtained by minimizing $\epsilon_{it}^T \Omega^{-1} \epsilon_{it}$ where $\Omega$ is the $n \times n$ variance-covariance matrix, is given by:

$$
\hat{\beta}^\text{GLS} = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}y.
$$

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

$$
\Omega = \sigma_u^2 I_n + \sigma_\mu^2 \Delta_\mu \Delta_\mu^T + \sigma_\nu^2 \Delta_\nu \Delta_\nu^T,
$$

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

With general heteroscedasticity (see assumption RE.3 in Appendix B), that is $\mu_i \sim (0, \varphi_i^2)$ and $u_{it} \sim (0, \psi_i^2)$, the matrix $\Omega$ in (9) is modified to:

$$
\Omega = \Psi + \Delta_\mu \Phi \Delta_\mu^T + \sigma_\nu^2 \Delta_\nu \Delta_\nu^T,
$$

with the $n \times n$ matrix\(^10\) $\Psi = \text{diag} \left( \Delta_{\mu_{AA}} \Delta_{\mu_{AB}} \psi \right)$, and the $A \times 1$ vector $\psi = (\psi_1^2, \psi_2^2, \ldots, \psi_A^2)^T$, the $N \times N$ matrix $\Phi = \text{diag} \left( \Delta_{\mu_{AB}} \varphi \right)$, and the $A \times 1$ vector $\varphi = (\varphi_1^2, \varphi_2^2, \ldots, \varphi_A^2)^T$.

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

This latter procedure considers the $n \times 1$ residuals $e \equiv y - X\hat{\beta}^W$ from the $W$ estimator in (3), where $X$ is a matrix of dimension $n \times (k-1)$, since it does not include the intercept. Given that the $n \times k$ matrix $X$ in (8) contains a vector of ones, we have

---

\(^9\) Note that the number of explanatory variables, obviously including the intercept, is $k$.

\(^10\) This matrix and the related vector $\psi$ have been already defined in Appendix A.
to define the $n \times 1$ consistent centered residuals $f \equiv E_a e = e - \bar{e}$, where $E_a = I_a - J_n$, with $I_a$ being an identity matrix of dimension $n$, $J_n = \frac{1}{n}$, and $J_n$ a matrix of ones of dimension $n$. Moreover, we have to define also the $n_a \times 1$ consistent centered residuals $f_a = H_a f$, with $H_a$ the $n_a \times n$ matrix obtained from the identity matrix $I_n$ by omitting the rows referring to observations not related to stratum $a$, and the matrix $J_{n_a} = \frac{J_n}{n_a}$, with $J_{n_a}$ a matrix of ones of dimension $n_a$.

The adapted $QUEs$ for $\Psi, \Phi$, and $\sigma^2$ is obtained by equating:

$$
q_{n_a} \equiv f^T Q_\Delta H^T_a H_a Q_\Delta f \quad \rightarrow \quad \sum_{a=1}^A q_{n_a} = q_n \equiv f^T Q_\Delta f,
$$

$$
q_{N_a} \equiv f^T J_{n_a} f_a \quad \rightarrow \quad \sum_{a=1}^A q_{N_a} = q_N \equiv f^T \Delta_\mu \Delta_N^{-1} \Delta_\mu^T f,
$$

$$
q_r = f^T \Delta_\nu \Delta_r^{-1} \Delta_r^T f,
$$

to their expected values. For more details on the identities in (11), see the formula (37) in Appendix D.

Hence, the estimator of $\psi^2$ is:

$$\hat{\psi}^2_a = \frac{q_{n_a} + k_a \hat{\sigma}^2_a}{n_a - N_a - \tau_a},$$

where $k_a \equiv \text{tr}[X^T Q_\Delta X]^{-1} X^T Q_\Delta H^T_a H_a Q_\Delta X]$, with $\sum_{a=1}^A k_a = k - 1$, $\tau_a \equiv n_a - N_a - \text{tr}(H_a Q_\Delta H^T_a)$, with $\sum_{a=1}^A \tau_a = T - 1$. The estimated variance $\hat{\sigma}^2_a$ is obtained by equating $q_{n_a}$ to its expected value (see Wansbeek and Kapteyn, 1989). Furthermore, the estimator of $\phi^2_a$ is:

$$\hat{\phi}^2_a = \frac{q_{N_a} - (N_a - 2\nu_a) \hat{\psi}^2_a}{n_a - 2\lambda_{\mu a} - \nu_a \lambda_{\nu a} \hat{\sigma}^2_a} \left( N_a - 2\lambda_{\nu a} - \frac{\nu_a}{n_a} \lambda_{\nu a} \hat{\sigma}^2_a \right),$$

where $k_{N_a} \equiv \text{tr}[X^T Q_\Delta X]^{-1} X^T J_{n_a} X_n]$, $k_0 \equiv \frac{\text{tr}[X^T Q_\Delta X]^{-1} X^T}{n_a}$, $k_{0a} \equiv \frac{2 \text{tr}[X^T Q_\Delta X]^{-1} X^T L_a}{n_a}$, $k_\nu \equiv \frac{\text{tr}[X^T Q_\Delta X]^{-1} X^T}{n_a}$, $k_{\nu a} \equiv \frac{2 \text{tr}[X^T Q_\Delta X]^{-1} X^T L_a}{n_a}$, $\lambda_{\mu a} \equiv \frac{\text{tr}[X^T Q_\Delta X]^{-1} X^T L_a}{n_a}$, $\lambda_{\nu a} \equiv \frac{\text{tr}[X^T Q_\Delta X]^{-1} X^T L_a}{n_a}$, $\mu_{\nu a} \equiv \frac{\text{tr}[X^T Q_\Delta X]^{-1} X^T L_a}{n_a}$, with $J_a$ the set of periods in which individuals belonging to stratum $a$ are observed. The estimated variances $\hat{\sigma}^2_a$ and $\hat{\phi}^2_a$ are obtained jointly by equating $q_{N_a}$ and $q_r$ to their expected values (see Wansbeek and Kapteyn, 1989).

Simpler heteroscedastic schemes (i.e., heteroscedasticity only on the individual-specific disturbance or on the remainder error) can be obtained combining results for the general scheme with those for the homoscedastic case, although when we consider the case of heteroscedasticity only on the individual-specific disturbance the expected value of $q_{N_a}$ and the estimated variance $\hat{\phi}^2_a$ are obtained differently as detailed in equations (42)-(43) in Appendix D.
2.3 Monte Carlo experiment – single-equation case

In order to analyze the performances of the proposed techniques, we develop a simple simulation\(^\text{11}\) on

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon, \]

where \( \beta_0 = 10, \beta_1 = -3, \beta_2 = 8, \text{ and } \beta_3 = -2. \)

We assume unbalanced panels with a large number of individuals \((N = 250 \text{ and } N = 500)\) extended over a rather long time period \((T = 12)\). This should mimic a real world situation of a large unbalanced panel for which the two-way \( \text{EC} \) model is the appropriate one.

Moreover, the experiment is implemented by considering as strata the deciles of the independent variable \( x_2 \). The homoscedastic time variance is \( \sigma^2 = 6.271 \), while the heteroscedastic variances have been generated with \( \phi_a^2 = \sigma^2 (1 + \lambda x_2)^2 \), where \( \sigma^2 = 6.488 \), and \( \psi^2 = \sigma^2 (1 + \lambda x_2)^2 \), where \( \sigma^2 = 6.039; \lambda \) is assigned values 0, 1, and 2, where \( \lambda = 0 \) denotes the homoscedastic case and the degree of heteroscedasticity increases as the value of \( \lambda \) becomes larger.\(^\text{12}\)

Finally, the independent variables’ values \( x_{k\ell} \) \((k = 1, 2, 3)\) 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. \((2012)\):

\[ x_{k\ell} = 0.1r + 0.5x_{k\ell-1} + \omega_{k\ell}, \quad k = 1, 2, 3 \]

with \( \omega_{k\ell} \) following the uniform distribution \([-\frac{1}{2}, \frac{1}{2}]\) and \( x_{k0} = 5 + 10\omega_{k0} \).

In order to construct the unbalanced panels, we adopt the procedure currently used for rotating panels, in which we have approximately the same number of individuals every time period: a fixed percentage of individuals \((20\% \text{ in our case}^{\text{13}})\) is replaced each time period, but they can re-enter the sample in later periods. Thus, if the number of individuals is \( N = 250 \) then the number of observations is \( n = 1031 \), if the number of individuals is \( N = 500 \) then he number of observations is \( n = 2062 \).

The results of a 2000-run simulation\(^\text{14}\) are shown in Table 1 and Table 2\(^\text{15}\).

Table 1 reports the estimated variances \( \hat{\psi}_a^2 \) and \( \hat{\phi}_a^2 \), being the latter computed on the basis of a remainder error either homoscedastic \( (\phi_a^2) \) or heteroscedastic \( (\psi_a^2) \). As one can notice right away, if \( \lambda \) is equal to 1 or 2 \((\text{i.e. in the heteroscedastic cases})\) the

\(^{11}\) The simulations have been implemented with the econometric software \( \text{TSP} \) version 5.1.

\(^{12}\) Whereas data have been generated by specifying the same parametric variance functions as in Li and Stengos \((1994)\) and Roy \((2002)\), the proposed estimation method proves to be effective also in the case of heteroscedasticity of unknown form, if the strata are identified by using a proper selection procedure, such as the Akaike \((1974)\) information criterion.

\(^{13}\) Also in Wansbeek and Kapteyn \((1989)\) each period 20\% of the households in the panel is removed randomly.

\(^{14}\) With \( N = 250 \) the average numbers of observations for each stratum \( a \) are \( \bar{n}_{a1} = 78, \bar{n}_{a2} = 113, \bar{n}_{a3} = 134, \bar{n}_{a4} = 144, \bar{n}_{a5} = 145, \bar{n}_{a6} = 141, \bar{n}_{a7} = 116, \bar{n}_{a8} = 77, \bar{n}_{a9} = 51, \bar{n}_{a10} = 32 \); and with \( N = 500 \) they are \( \bar{n}_{a1} = 155, \bar{n}_{a2} = 226, \bar{n}_{a3} = 269, \bar{n}_{a4} = 287, \bar{n}_{a5} = 290, \bar{n}_{a6} = 282, \bar{n}_{a7} = 232, \bar{n}_{a8} = 153, \bar{n}_{a9} = 104, \text{ and } \bar{n}_{a10} = 64. \)

\(^{15}\) As in Baltagi and Griffin \((1988)\) and Phillips \((2003)\), negative variance estimates are replaced by zero.

\(^{16}\) Whereas data have been generated such that the individual-specific error \( u_i \) and the time-specific error \( \nu_t \) are random variables. Table 2 displays also the results of the two-way \( \text{FE} \) and robust two-way \( \text{FE} \) estimations to check the method suggested in subsection 1. Moreover, note that the two-way \( \text{FE} \) residuals are used in the \( \text{QUE} \) procedure of the \( \text{GLS} \) estimation \((\text{and both in the } \text{QUE} \text{ and } \text{WB} \text{ procedures of the } \text{SUR} \text{ systems estimation in the following section 3).} \)
estimated variance $\hat{\sigma}_{\phi}^2 (\psi^2_a)$ is closer than the estimated variance $\hat{\sigma}^2 (\hat{\sigma}^2_a)$ to the true value $\varphi^2_a$. Moreover, when $\lambda$ is equal to 0 (homoscedastic case), the heteroscedastic procedures allow to obtain estimated variances $\psi^2_a$ and $\hat{\phi}^2_a$ that do not substantially vary among strata, and that are very close to the estimated values $\hat{\sigma}^2_a$ and $\hat{\sigma}^2_a$ obtained through the homoscedastic procedure (also reported in Table 1).

Table 1. Simulation results on single-equation two-way EC model: estimated variances $\psi^2_a$ and $\hat{\phi}^2_a$

| $\lambda$ = 0 | $\lambda$ = 1 | $\lambda$ = 2 |
|----------------|----------------|----------------|
| $N = 250$, $T = 12$, and $n = 1031$ | $N = 500$, $T = 12$, and $n = 2062$ | $N = 1000$, $T = 12$, and $n = 5000$ |
| $\psi^2_a$ | $\psi^2_a$ | $\psi^2_a$ |
| $\hat{\phi}^2_a$ | $\hat{\phi}^2_a$ | $\hat{\phi}^2_a$ |
| $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ |
| $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ |
| $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ |
| $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ | $\hat{\sigma}^2_a$ |

Note: $\psi^2_a$ and $\hat{\phi}^2_a$ are the true values of the variances, $\hat{\sigma}^2_a$ are the estimated variances of the remainder error $\hat{\mu}_a$, $\phi^2_a$ are the estimated variances of the individual-specific error $\hat{\mu}_i$ computed on the basis of a remainder error either homoscedastic ($\hat{\sigma}^2_a$) or heteroscedastic ($\hat{\sigma}^2_a$).

Table 2 shows that the heteroscedastic procedures allow to obtain standard errors
lower than those obtained through the homoscedastic procedure if $\lambda = 1, 2$, but higher standard errors if $\lambda = 0$. However, in the latter case (i.e., the homoscedastic case) if the number of individuals (and thus the number of observations) increases, then the standard errors computed with the heteroscedastic procedures become closer to the standard errors computed with the homoscedastic procedure.

### Table 2. Simulation results on single-equation two-way EC model: standard errors of the estimated parameters and (average) estimated variances of the error components

| $\lambda$ | $N = 250, T = 12, n = 1031$ | $N = 500, T = 12, n = 2062$ |
|-----------|----------------------------|----------------------------|
| FE       | true value | RE QUE heteroscedasticity on $\mu$ | (a) | (b) | (c) | (d) | (e) | (f) | RE QUE heteroscedasticity on $\mu$ | (a) | (b) | (c) | (d) | (e) | (f) |
| $\beta_0$ | 0.756 0.248 0.248 0.247 | 0.731 0.176 0.176 0.176 |
| $\beta_1$ | 0.132 0.129 0.108 0.122 0.122 0.121 | 0.093 0.092 0.076 0.086 0.086 0.086 |
| $\beta_2$ | 0.132 0.129 0.108 0.121 0.121 0.120 | 0.092 0.092 0.076 0.086 0.086 0.086 |
| $\phi^2$ | 6.488 6.516 6.516 6.519 6.526 6.488 | 6.521 6.521 6.521 6.521 6.524 |
| $\sigma_i^2$ | 6.271 6.044 6.044 6.044 6.044 6.043 | 6.039 6.043 6.043 6.043 6.043 6.041 |
| $\phi^3$ | 46.040 49.792 49.792 49.748 46.339 46.071 | 49.784 49.784 49.755 46.263 |
| $\sigma_i^3$ | 6.271 6.265 6.271 6.271 6.234 6.239 |

Note: Parameters estimation based on (a-b) the estimated homoscedastic variance $\hat{\sigma}_i^2$; (c) the estimated homoscedastic variances $\hat{\sigma}_i^2$; (d) the estimated homoscedastic variances $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ and heteroscedastic variances $\hat{\psi}_u^2$, whose the average value is $\hat{\psi}_u^2$; (e) the estimated homoscedastic variances $\hat{\psi}_u^2$ and $\hat{\psi}_v^2$ and heteroscedastic variances $\hat{\phi}_u^2(\hat{\phi}_v^2)$, whose the average value is $\hat{\phi}_u^2$; (f) the estimated homoscedastic variances $\hat{\sigma}_u^2$ and heteroscedastic variances $\hat{\psi}_u^2$ and $\hat{\phi}_u^2(\hat{\phi}_v^2)$.

Focusing on the heteroscedastic cases, considering heteroscedasticity only on the remainder error (columns (d)) allows to obtain standard errors that are lower than the standard errors obtained considering heteroscedasticity only on the individual-specific effect (columns (e)). In other words, misspecifying the form of heteroscedasticity can be costly when heteroscedasticity is assumed only on the individual-specific effect. These findings confirm the conclusions in Baltagi et al. (2005). Obviously, the smallest standard errors are obtained implementing the estimation procedure which considers both heteroscedasticity types (columns (f)).

As in Li and Stengos (1994), Roy (2002), and Baltagi et al. (2005), we consider the relative efficiency of the different estimators, computed as the ratio of the mean square
error (MSE) of the estimator under consideration to the MSE of the true GLS estimator. Results are reported in Table 3.

| \( \lambda = 0 \) | homoscedasticity | heteroscedasticity on \( \epsilon_{it} = \mu_i + \nu_t + u_{it} \) | \( \lambda = 1 \) | homoscedasticity | heteroscedasticity on \( \epsilon_{it} = \mu_i + \nu_t + u_{it} \) | \( \lambda = 2 \) | homoscedasticity | heteroscedasticity on \( \epsilon_{it} = \mu_i + \nu_t + u_{it} \) |
|----------------|-----------------|-------------------|----------------|-----------------|-------------------|----------------|----------------|-------------------|
| 1.0025 | 1.0001 | 1.0000 | 1.0001 | 1.0013 | 1.0000 | 1.0000 | 1.0000 |
| 0.9997 | 0.9989 | 1.0000 | 1.0006 | 0.9998 | 0.9994 | 1.0000 |
| 1.0002 | 0.9997 | 0.9982 | 1.0002 | 1.0001 | 0.9998 | 0.9991 | 1.0000 |

Note: Relative efficiency is defined as the ratio of the MSE of the estimator under consideration to the MSE of the true GLS estimator (computed considering the true variances \( \sigma_1^2 \), \( \sigma_2^2 \), and \( \sigma_3^2 \)). Note that values of the ratio both larger and smaller than 1 indicate a loss in efficiency: if the ratio is larger than 1, then the absolute value of the composite error term \( \epsilon_{it} = \mu_i + \nu_t + u_{it} \) is larger than the true value; and if the ratio is smaller than 1, then the absolute value of the composite error term \( \epsilon_{it} \) is smaller than the true value.

We see that there are improvements in relative MSE numbers as the sample size increases, especially when we refer to the homoscedastic estimator. Furthermore, confirming our previous remarks, misspecifying the form of heteroscedasticity may be costly when only the individual-specific effect is considered heteroscedastic, especially if the sample size is small. Besides, as already observed in the comments to Table 2, the most efficient estimator is the one that considers both the remainder error and the individual-specific effect heteroscedastic.

### 3. Heteroscedastic 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 heteroscedastic two-way SUR systems we extend the procedure in Bistri (2004), with individuals grouped according to the number of times they are observed.

#### 3.1 Model and notation

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

We assume that the \( T \) groups of individuals are ordered such that the \( N_{p=1} \) individuals observed once come first, the \( N_{p=2} \) individuals observed twice come second, etc. Hence, with \( C_p = \sum_{h=1}^p N_h \) 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 \( I_{p=1} \) 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...
systems 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 \( m = 1, \ldots, M \), for the observation \((i,t)\) we have:

\[
y_{it} = X_{it} \beta + \mu_i + \nu_t + u_{it} = X_{it} \beta + \epsilon,
\]

where the \( M \times K \) matrix of explanatory variables is \( X_{it} = \text{diag}(x_{it1}^T, \ldots, x_{itm}^T) \) and the \( K \times 1 \) vector of parameters is \( \beta = (\beta_1, \ldots, \beta_M)^T \) and where \( \mu_i \equiv (\mu_{i1}, \ldots, \mu_{im})^T \), \( \nu_t \equiv (\nu_{1t}, \ldots, \nu_{Mt})^T \), and \( u_{it} \equiv (u_{1it}, \ldots, u_{Mit})^T \). If we do not have cross-equation restrictions, we can assume \( E(u_{mit} | X_{it1}, X_{it2}, \ldots, X_{itM}) = 0 \), and then \( E(y_{mit} | X_{it1}, X_{it2}, \ldots, X_{itM}) = E(y_{mti} | x_{mit}) = x_{mit} \beta_m \). On the contrary, if we have cross-equation restrictions\(^{17}\), we can only assume \( E(u_{it} | x_{it}^T) = 0 \), where \( x_{it} \equiv (x_{it1}^T, x_{it2}^T, \ldots, x_{itM}^T)^T \).

With heteroscedasticity on both the individual-specific disturbance and the remainder error, for \( i \in I_a \) across the regression equations \( m \) and \( j \), we assume that:

\[
E\left( \mu_{mit}, \mu_{jt'} \right) = \begin{cases} \Phi_{a,mj} & i = i' \text{ and } t = t' \\ 0 & i \neq i' \text{ or } t \neq t' \end{cases},
\]

\[
E\left( \nu_{mti}, \nu_{jt'} \right) = \begin{cases} \sigma_{\nu,mj} & t = t' \\ 0 & t \neq t' \end{cases},
\]

\[
E\left( u_{mit}, u_{jt'} \right) = \begin{cases} \Psi_{a,mj} & i = i' \text{ and } t = t' \\ 0 & i \neq i' \text{ or } t \neq t' \end{cases}.
\]

Let us consider the \( NM \times 1 \) vector \( \mu \equiv (\mu_1^T, \ldots, \mu_N^T)^T \), the \( TM \times 1 \) vector \( \nu \equiv (\nu_1^T, \ldots, \nu_T^T)^T \), and the \( nM \times 1 \) vector \( u \equiv (u_{11}^T, \ldots, u_{1N}^T, u_{21}^T, \ldots, u_{2N}^T)^T \). Since the \( M \times 1 \) vectors \( u_{it} \sim (0, \Psi_{u}) \), the \( M \times 1 \) vectors \( \mu_{it} \sim (0, \Phi_{a}) \), and the \( TM \times 1 \) vector \( \nu \sim (0, \Sigma_{\nu}) \), with the \( M \times M \) matrices \( \Psi_{u} = [\Psi_{a,mj}], \Phi_{a} = [\Phi_{a,mj}], \) and \( \Sigma_{\nu} = [\sigma_{\nu,mj}] \), we can assume that the expected values of the vectors \( u_{it}, \mu_{it}, \) and \( \nu_t \) are zero and their covariance matrices are equal to \( \Psi_{u}, \Phi_{a}, \) and \( \Sigma_{\nu} \). It follows that \( E(\xi_{it} \xi_{jt'}) = \delta_{i't'} \Phi_{a} + \delta_{i't'} \Sigma_{\nu} + \delta_{i't'} \delta_{j't'} \Psi_{u} \), with \( \delta_{i't'} = 1 \) for \( i = i' \) and \( t = t' \), \( \delta_{i't'} = 0 \) for \( i \neq i' \) or \( t \neq t' \), \( \delta_{i't'} = 0 \) for \( t = t' \) and \( \delta_{i't'} = 0 \) or \( t \neq t' \).

As in Biørn (2004), let us consider the \( pM \times 1 \) vector of independent variables \( y_{i(p)} \equiv (y_{i1}^T, \ldots, y_{ip}^T)^T \), the \( pM \times K \) matrix of explanatory variables \( X_{i(p)} \equiv (X_{i1}^T, \ldots, X_{ip}^T)^T \), and the \( pM \times 1 \) vector of composite error terms \( \epsilon_{i(p)} \equiv (\epsilon_{i1}, \ldots, \epsilon_{ip})^T \) for \( i \in I_p \). If we define the \( pM \times TM \) matrix \( \Delta_{i(p)} \), indicating in which period \( t \) the individual \( i \) of the group \( p \) is observed, and if we consider the \( TM \times 1 \) vector \( \nu \), for the individual \( i \in I_p \), we can define the \( pM \times 1 \) vector \( \nu_{i(p)} = \Delta_{i(p)} \nu \) and write the model:

\[
y_{i(p)} = X_{i(p)} \beta + (\iota_p \otimes \mu_i + \nu_{i(p)} + u_{i(p)} = X_{i(p)} \beta + \epsilon_{i(p)},
\]

where \( \iota_p \) is a \( p \times 1 \) vector of ones (see Platoni et al., 2012).

The \( pM \times pM \) heteroscedastic variance-covariance matrix of the \( pM \times 1 \) composite error terms \( \epsilon_{i(a,p)} \) for the individual \( i \in I_{a,p} \), with \( I_{a,p} = I_a \cap I_p \) the set of individuals

\(^{17}\) As Biørn (2004) suggests, with cross-equations restrictions we can redefine \( \beta \) as the complete \( K \times 1 \) coefficient vector (without duplication) and the \( M \times K \) regression matrix as \( X_{p} = (x_{11t}, \ldots, x_{km}^T) \), where the \( kth \) element of the \( k_m \times 1 \) vector \( x_{ma} \) either contains the observation on the variable in the \( mth \) equation which corresponds to the \( kth \) coefficient in \( \beta \) or is zero if the \( kth \) coefficient does not occur in the \( mth \) equation.
belonging to stratum $a$ and observed in $p$ periods, is given by:

$$\Omega_{a,p} = E_p \otimes (\Psi_a + \Sigma_v) + \bar{J}_p \otimes (\Psi_a + \Sigma_v + p\Phi_a), \quad (17)$$

where $E_p = I_p - \bar{J}_p$ (with $I_p$ identity matrix of dimension $p$) and $\bar{J}_p = \frac{J_p}{p}$ (with $J_p$ matrix of ones of dimension $p$). Since $E_p$ and $\bar{J}_p$ are symmetric, idempotent, and have orthogonal columns, the inverse of the variance-covariance matrix of the individuals belonging to stratum $a$ and group $p$ is:

$$\Omega_{a,p}^{-1} = E_p \otimes (\Psi_a + \Sigma_v)^{-1} + \bar{J}_p \otimes (\Psi_a + \Sigma_v + p\Phi_a)^{-1}. \quad (18)$$

This specification nests simpler heteroscedastic schemes as well as the homoscedastic case by replacing $\Phi_a$ with $\Sigma_\mu$ and/or $\Psi_a$ with $\Sigma_u$.

If we assume that $\Psi_a$, $\Phi_a$, and $\Sigma_v$ are known, then in the heteroscedastic case we can write the GLS estimator for the $K \times 1$ vector of parameters $\beta$ as the problem of minimizing:

$$\sum_{p=1}^{T} \sum_{a=1}^{A} \sum_{i \in I_{a,p}} \epsilon_i^T \Omega_{a,p}^{-1} \epsilon_i. \quad (19)$$

where, for sake of simplicity and since there is no risk of ambiguity, $\epsilon_i$ is used instead of $\epsilon_{i(a,p)}$.

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

$$\hat{\beta}_{GLS}^{p} = \left( \sum_{a=1}^{A} \sum_{i \in I_{a,p}} X_i^T \Omega_{a,p}^{-1} X_i \right)^{-1} \sum_{a=1}^{A} \sum_{i \in I_{a,p}} X_i^T \Omega_{a,p}^{-1} Y_i, \quad (20)$$

while the full GLS estimator is:

$$\hat{\beta}_{GLS} = \left( \sum_{p=1}^{T} \sum_{a=1}^{A} \sum_{i \in I_{a,p}} X_i^T \Omega_{a,p}^{-1} X_i \right)^{-1} \sum_{p=1}^{T} \sum_{a=1}^{A} \sum_{i \in I_{a,p}} X_i^T \Omega_{a,p}^{-1} Y_i, \quad (21)$$

where $X_i$ 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 $\Psi_a$, $\Phi_a$, and $\Sigma_v$. This can be achieved adopting either the QUE procedure suggested by Wansbeek and Kapteyn (1989) for the homoscedastic single-equation case or the within-between (WB) procedure suggested by Biørn (2004) for the homoscedastic one-way SUR system. In the following sub-sections we modify both procedures making them suitable for the heteroscedastic two-way SUR system.
The QUE procedure

The QUE procedure considers the \( n \times 1 \) residuals \( e_m = y_m - X_m \hat{\beta}_m^W \) from the \( W \) estimator in (3) for the equation \( m = 1, \ldots, M \), where \( X_m \) is a matrix of dimension \( n \times (k_m - 1) \). If we assume that the \( n \times k_m \) matrix \( X_m \) contains a vector of ones, then we have to define the \( n \times 1 \) consistent centered residuals \( f_m = E_y e_m = e_m - \bar{e}_m \).

With heteroscedasticity, we can obtain the adapted QUEs for \( \Psi_{m,j} \), \( \Phi_{m,j} \), and \( \sigma_{\nu,mj} \) by equating:

\[
q_{n,mj} = f_j^T \Delta_0 H_0^T H_0 \Delta f_m \rightarrow \sum_{a=1}^{\Lambda} q_{n,mj} = q_{n,mj} = f_j^T \Delta_0 f_m,
\]

\[
q_{N,mj} = f_j^T \bar{J}_N f_{nu} \rightarrow \sum_{a=1}^{\Lambda} q_{N,mj} = q_{N,mj} = f_j^T \Delta_\Lambda \Delta^{-1}_\Lambda f_m, \tag{22}
\]

\[
q_{T,mj} = f_j^T \Delta_\nu \Delta^{-1}_\nu f_m,
\]

to their expected values. The identities in (22) can be further detailed as already done in formula (37), Appendix D, for the identities in (11).

Hence, the estimator of \( \psi_{u,mj} \) is:

\[
\hat{\psi}_{u,mj} = \frac{q_{n,mj} + (k_{u,m} + k_{u,j} - k_{u,mj}) \bar{\sigma}_{u,mj}}{n_a - N_a - \tau_a} \tag{23}
\]

where \( k_{u,m} = \text{tr}[(X_m^T \Delta_0 X_m)^{-1} X_m^T \Delta_0 X_j (X_j^T \Delta_0 X_j)^{-1} X_j^T \Delta_0 H_0^T H_0 \Delta X_m] \), \( \sum_{a=1}^{\Lambda} k_{a,mj} = k_{mj} \) and \( k_{mj} = \text{tr}[(X_m^T \Delta_0 X_m)^{-1} X_m^T \Delta_0 X_j (X_j^T \Delta_0 X_j)^{-1} X_j^T \Delta_0 X_m] \). The estimated variance-covariance \( \hat{\sigma}_{u,mj} \) is obtained by equating \( q_{n,mj} \) to its expected value (see Platoni et al., 2012). Furthermore, the estimator of \( \phi_{u,mj} \) is:

\[
\hat{\phi}_{u,mj} = \frac{q_{n,mj} - (N_a - 2 \hat{\lambda}_u) \hat{\psi}_{u,mj} - (k_{N,mj} - k_{0,m} + \hat{\mu} \hat{\sigma}_{u,mj} + \frac{\hat{\mu}_{v}}{\hat{\nu}}) \hat{\sigma}_{u,mj}}{n_a - 2 \hat{\lambda}_u \hat{\sigma}_{u,mj}} + \frac{-\frac{\hat{\mu}_V}{\hat{\nu}} \lambda_{\mu} \hat{\sigma}_{u,mj} - (N_a - 2 \lambda_{\nu} + \frac{\hat{\mu}_{\nu}}{\hat{\nu}} \lambda_{\nu}) \hat{\sigma}_{v,mj}}{n_a - 2 \lambda_{\nu} \hat{\sigma}_{v,mj}}, \tag{24}
\]

where \( k_{N,mj} = \text{tr}[(X_m^T \Delta_0 X_m)^{-1} X_m^T \Delta_0 X_j (X_j^T \Delta_0 X_j)^{-1} X_j^T \bar{J}_N f_{nu}] \), \( k_{0,mj} = \frac{1}{n} \sum_{a=1}^{\Lambda} X_m (X_m^T \Delta_0 X_m)^{-1} X_m^T \Delta_0 X_j (X_j^T \Delta_0 X_j)^{-1} X_j^T \bar{J}_N f_{nu} \), and \( k_{0,mj} = \frac{1}{n} \sum_{a=1}^{\Lambda} X_m (X_m^T \Delta_0 X_m)^{-1} X_m^T \Delta_0 X_j (X_j^T \Delta_0 X_j)^{-1} X_j^T \bar{J}_N f_{nu} \). The estimated variance-covariance \( \hat{\sigma}_{u,mj} \) is obtained jointly with \( \hat{\sigma}_{v,mj} \) by equating \( q_{N,mj} \) and \( q_{T,mj} \) to their expected values (see Platoni et al., 2012).

As in the single-equation case, simpler heteroscedastic scheme (i.e., heteroscedasticity only on the individual-specific disturbance or on the remainder error) can be obtained combining results for the general scheme with those for the homoscedastic case, although when we consider the case of heteroscedasticity only on the individual-specific disturbance the expected value of \( q_{N,mj} \) and the estimated variance-covariance \( \hat{\phi}_{u,mj} \) are obtained differently (see equations (48)-(49) in Appendix D).
The WB procedure

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

\[
W_e = \sum_{a=1}^{A} W_{e_a} = \sum_{a=1}^{A} \sum_{i \in I_a} \sum_{t=1}^{T_i} (\varepsilon_{at} - \bar{\varepsilon}_i - \bar{\varepsilon}_d)(\varepsilon_{at} - \bar{\varepsilon}_i - \bar{\varepsilon}_d)^T,
\]

\[
B_{C}^{e} = \sum_{a=1}^{A} B_{C_{e_a}} = \sum_{a=1}^{A} \sum_{i \in I_a} T_i (\bar{\varepsilon}_i - \bar{\varepsilon})(\bar{\varepsilon}_i - \bar{\varepsilon})^T,
\]

\[
B_{F}^{e} = \sum_{t=1}^{T} N_t (\bar{\varepsilon}_d - \bar{\varepsilon})(\bar{\varepsilon}_d - \bar{\varepsilon})^T,
\]

where for each equation $m$ we have $\bar{\varepsilon}_{mi} = \sum_{i=1}^{N_m} \varepsilon_{mi} / N_m$, $\bar{\varepsilon}_{md} = \sum_{m=1}^{M} \varepsilon_{md} / N_m$, and $\bar{\varepsilon}_m = \sum_{m=1}^{M} \varepsilon_{mi} / n$.

Because the $u_{it}$'s, the $\mu$'s, and the $\nu$'s are independent, from the equations in (25) we can write:

\[
E(W_{e_a}) = E(W_{u_{a}}),
\]

\[
E(B_{C_{e_a}}^{e}) = E(B_{C_{u_{a}}}) + E(B_{C_{\nu_{a}}}),
\]

\[
E(B_{F_{e_a}}^{e}) = E(B_{F_{u_{a}}}) + E(B_{F_{\nu}}),
\]

where the within individuals (co)variation is:

\[
W_{u_{a}} = \sum_{i \in I_a} T_i (u_{it} - \bar{u}_i - \bar{u}_d)(u_{it} - \bar{u}_i - \bar{u}_d)^T
\]

\[
= \sum_{i \in I_a} \sum_{t=1}^{T_i} u_{it} u_{it}^T - \sum_{i \in I_a} T_i \bar{u}_i \bar{u}_i^T - \sum_{i \in I_a} \sum_{t=1}^{T_i} \bar{u}_d \bar{u}_d^T,
\]

the between individuals (co)variations are:

\[
B_{C_{\mu_{a}}}^{e} = \sum_{i \in I_a} T_i (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T = \sum_{i \in I_a} T_i \bar{\mu}_i \bar{\mu}_i^T - \sum_{i \in I_a} T_i \bar{\mu} \bar{\mu}^T,
\]

\[
B_{C_{\nu_{a}}}^{e} = \sum_{i \in I_a} T_i (\nu_i - \bar{\nu})(\nu_i - \bar{\nu})^T = \sum_{i \in I_a} T_i \bar{\nu}_i \bar{\nu}_i^T - \sum_{i \in I_a} T_i \bar{\mu} \bar{\mu}^T,
\]

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

\[
B_{F_{\nu}}^{e} = \sum_{t=1}^{T} N_t (\nu_t - \bar{\nu})(\nu_t - \bar{\nu})^T = \sum_{t=1}^{T} N_t \bar{\nu}_t \bar{\nu}_t^T - n \bar{\nu} \bar{\nu}^T,
\]

\[
B_{F_{\mu}}^{e} = \sum_{t=1}^{T} N_t (\mu_t - \bar{\mu})(\mu_t - \bar{\mu})^T = \sum_{t=1}^{T} N_t \bar{\mu}_t \bar{\mu}_t^T - n \bar{\mu} \bar{\mu}^T,
\]

where $\bar{\mu}_{mis} = \sum_{i \in I_a} \bar{\mu}_{it} / T_i$, $\bar{\mu}_{md} = \sum_{m=1}^{M} \bar{\mu}_{mi} / n$, $\bar{\mu}_m = \sum_{m=1}^{M} \bar{\mu}_{mi} / n = \sum_{m=1}^{M} (T_i \bar{\mu}_{mi}) / n$ or $\bar{\mu}_m = \sum_{m=1}^{M} \bar{\mu}_{mi} / n$ (see Bjorn, 2004; Platoni et al., 2012).

Since for $i \in I_a$ we have $E(\varepsilon_{at} \varepsilon_{at}^T) = \delta_{at} \Phi_a + \delta_{at} \Sigma_{\nu} + \delta_{at} \delta_{at} \Psi_{at}$, where $E(u_{it} u_{it}^T) = \delta_{it} \Phi_i$, $E(u_{it} u_{it}^T) = \delta_{it} \Sigma_{\nu}$, and $E(u_{it} u_{it}^T) = \delta_{at} \Psi_{at}$.
\( \delta_\alpha \delta_\nu \Psi_\alpha, \ E(\mu_1 \mu_1^T) = \delta_\nu \Phi_\alpha, \) and \( E(\gamma_i \gamma_i') = \delta_\nu \Sigma_\nu, \) it follows that \( E(\bar{u}_i \bar{u}_i^T) = \frac{\Psi}{T}, \)

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} = \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} \approx \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} = \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} \approx \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} = \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} \approx \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} = \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]

\[
E(\bar{u}_i \bar{u}_i^T) = \frac{\Sigma_i}{n_i} \approx \frac{\Sigma_i}{N_i}, \quad \text{with } I_1 \text{ the set of individuals observed in period } t,
\]
the different $M$ equations are the following:

$$W_f = \sum_{a=1}^{A} W_{f_a} = \sum_{a=1}^{A} \sum_{i=1}^{T_i} (f_{a1} - \bar{f}_{i1}) (f_{a2} - \bar{f}_{i2})^T,$$

$$B_{f}^C = \sum_{a=1}^{A} B_{f_a}^C = \sum_{a=1}^{A} \sum_{i=1}^{T_i} T_i \bar{f}_{i} (\bar{f}_{i} - \bar{f})^T,$$

$$B_{T}^F = \sum_{i=1}^{T} N_i (\bar{f}_{i} - \bar{f}) (\bar{f}_{i} - \bar{f})^T,$$

where for each equation $m$ we have $f_{ma} = \sum_{i=1}^{T_i} f_{mi}$, $f_{md} = \sum_{i=1}^{T_i} f_{mi}$, and $f_m = \sum_{i=1}^{T_i} f_{mi}$. Given that:

$$E(W_{f_a}) = (n_a - N_a) \Psi_a - \sum_{i=1}^{T_i} \frac{1}{N_a} \bar{f}_{i},$$

$$E(B_{f_a}^C) = \sum_{i=1}^{T} N_i \bar{f}_{i} - \sum_{i=1}^{T} \frac{1}{N_a} \bar{f}_{i},$$

$$E(B_{T}^F) = (n - \sum_{i=1}^{T} \frac{N_i}{N}) \Sigma_v + (T - 1) \Psi,$$

where $J_i$ is the set of periods in which individual $i$ is observed and with $\Psi \approx \Sigma_a$ and $\bar{f} \approx \Sigma_a$, we can conclude that the estimators in (30) and (31), with $W_{f_a}$ instead of $W_{e_a}$ and $B_{f_a}^C$ instead of $B_{e_a}^C$ respectively, are consistent estimators of $\Psi_a$ and $\Phi_a$. As mentioned above, both the consistent estimators of $\Sigma_a$ and $\Sigma_{\mu}$ and the consistent estimator of $\Sigma_v$ are derived as in the homoscedastic case (see Bistron, 2004; Platoni et al., 2012). Finally, with heteroscedasticity only on the individual-specific disturbance, the expected value $E(B_{f_a}^C)$ is given by the equation (51) in Appendix E.

### 3.3 Monte Carlo experiment – SUR system case

In order to analyze the performances of the proposed techniques, we develop a simple simulation on a three-equation system ($M = 3$). The simulated model is:

$$y_1 = \beta_{10} + \beta_{11} x_1 + \beta_{12} x_2 + \epsilon_1,$$

$$y_2 = \beta_{20} + \beta_{21} x_1 + \beta_{22} x_2 + \beta_{23} x_3 + \epsilon_2,$$

$$y_3 = \beta_{30} + \beta_{32} x_2 + \beta_{33} x_3 + \epsilon_3,$$

where $\beta_1 = (15.6, -3)^T$, $\beta_2 = (10, -3, 8, -2)^T$, and $\beta_3 = (20, -2, 5)^T$.18 Then we also allow the cross equations restrictions $\beta_{12} = \beta_{21}$ and $\beta_{23} = \beta_{32}$.

The independent variables’ values $x_{ik}$ ($k = 1, 2, 3$) have been generated and the unbalanced panel has been constructed according to the same DGP of the single-equation

18 Note that the second equation is the same equation of the single-equation case in subsection 3.
case\textsuperscript{19}. This should mimic a real world situation of a large unbalanced panel for which the two-way SUR system is the appropriate model. Moreover, as in the single-equation case, the experiment is implemented by considering as strata the deciles of the independent variable \( x_2 \). The homoscedastic time variance-covariance matrix is:

\[
\Sigma_v = \begin{bmatrix}
  6.429 & 0.717 & -1.107 \\
  6.271 & 1.235 \\
  9.371 
\end{bmatrix},
\]

while the heteroscedastic variances-covariances \( \varphi_{u,mj} \) and \( \psi_{u,mj} \) have been generated from the matrices:

\[
\Sigma_\mu = \begin{bmatrix}
  9.377 & -1.048 & 1.276 \\
  6.488 & 0.710 \\
  6.207 
\end{bmatrix}
\]

\[
\Sigma_\nu = \begin{bmatrix}
  6.544 & 0.738 & 0.881 \\
  6.039 & -1.232 \\
  9.489 
\end{bmatrix}
\]

with \( \varphi_{u,mj} = \sigma_{\mu,mj}(1 + \lambda \bar{x}_{2j})^2 \) and \( \psi_{u,mj} = \sigma_{u,mj}(1 + \lambda \bar{x}_{2j})^2 \), where \( \sigma_{\mu,mj} \) and \( \sigma_{u,mj} \) are elements of the matrices \( \Sigma_\mu \) and \( \Sigma_\nu \) respectively and \( \bar{x}_{2j} \) is the mean of the independent variable \( x_2 \) over the decile/stratum \( a \).\textsuperscript{20}

The results of a 2000-run simulation are shown in Tables 4 and 5.\textsuperscript{21,22}

Tables 4 and 5 show that, contrary to the single-equation case, the heteroscedastic procedures allow to obtain standard errors lower than those obtained through the homoscedastic procedure in all cases, i.e., not only in the heteroscedastic cases \( \lambda = 1, 2 \), but also in the homoscedastic case \( \lambda = 0 \). However, in the homoscedastic case (i.e., with \( \lambda = 0 \)) the standard errors obtained with the heteroscedastic procedures are very closed to the standard errors computed with the homoscedastic procedure.

Focusing on the heteroscedastic cases (i.e., with \( \lambda = 1, 2 \))

- the smallest standard errors are obtained when the estimation procedure which considers both kinds of heteroscedasticity is implemented;
- though, differently from the single-equation estimation, there is not an evident difference in the loss in efficiency due to the misspecification in the form of heteroscedasticity.

Finally, comparing the standard errors obtained with the QUE procedure (displayed in Table 4) and those obtained with the WB procedure (displayed in Table 5), it is possible to assert that the QUE procedure allows to obtained lower standard errors than those obtained with the WB procedure.

\textsuperscript{19} With \( N = 250 \) the numbers of individuals for each group \( p \) are \( N_{p=1} = 54, N_{p=2} = 43, N_{p=3} = 34, N_{p=4} = 27, N_{p=5} = 22, N_{p=6} = 18, N_{p=7} = 14, N_{p=8} = 11, N_{p=9} = 9, N_{p=10} = 7, N_{p=11} = 6 \), and \( N_{p=12} = 5 \); and with \( N = 500 \) they are \( N_{p=1} = 107, N_{p=2} = 86, N_{p=3} = 69, N_{p=4} = 55, N_{p=5} = 44, N_{p=6} = 35, N_{p=7} = 28, N_{p=8} = 22, N_{p=9} = 18, N_{p=10} = 14, N_{p=11} = 12, \) and \( N_{p=12} = 10 \).

\textsuperscript{20} The correlation among equations verifies the null hypothesis of the Breusch and Pagan (1979) test at \( n = 1,031 \), which is the number of observations when \( N = 250 \).

\textsuperscript{21} As in Baltagi and Griffin (1988) and Phillips (2003), negative variance estimates are replaced by zero.

\textsuperscript{22} Table 7 in Appendix F displays the estimated variances-covariances for the stratum \( a = 5 \).
Table 4. Simulation results on two-way SUR systems - *QUE* procedure: standard errors of the estimated parameters and (average) estimated variances and covariances of the error components

| Parameter | $N = 250$, $T = 12$, and $n = 1031$ | $N = 500$, $T = 12$, and $n = 2062$ |
|-----------|-----------------------------------|-----------------------------------|
| $\beta_{10}$ | 0.333 0.322 0.317 0.316 | 0.235 0.235 0.230 0.229 |
| $\beta_{11}$ | 0.129 0.128 0.127 | 0.092 0.092 0.091 0.091 |
| $\beta_{12}$ | 0.096 0.096 0.094 0.094 | 0.068 0.068 0.067 0.067 |
| $\phi_{1}^2$ | 9.377 9.437 9.437 9.443 | 9.377 9.393 9.393 9.394 |
| $\phi_{12}$ | -1.048 -1.067 -1.067 -1.069 | -1.048 -1.044 -1.044 -1.044 |
| $\phi_{13}$ | 1.276 1.260 1.260 1.258 | 1.276 1.291 1.291 1.294 |
| $\sigma_{11}$ | 6.429 6.292 | 6.429 6.289 |
| $\sigma_{12}$ | 0.717 0.669 | 0.717 0.670 |
| $\sigma_{13}$ | 1.120 1.120 | 1.120 1.120 |
| $\psi_{1}^2$ | 6.544 6.539 6.539 6.539 | 6.544 6.550 6.550 6.550 |
| $\psi_{12}$ | 0.738 0.738 0.740 0.740 | 0.738 0.740 0.739 0.740 |
| $\psi_{13}$ | 0.881 0.880 0.881 0.880 | 0.881 0.885 0.885 0.883 |
| $\beta_{20}$ | 0.316 0.315 0.302 0.301 | 0.224 0.223 0.219 0.218 |
| $\beta_{21}$ | 0.096 0.096 0.094 0.094 | 0.068 0.068 0.067 0.067 |
| $\beta_{22}$ | 0.130 0.129 0.127 0.126 | 0.092 0.091 0.091 0.091 |
| $\phi_{2}^2$ | 6.488 6.516 6.516 6.519 | 6.488 6.521 6.521 6.521 |
| $\phi_{23}$ | 0.710 0.722 0.722 0.720 | 0.710 0.708 0.708 0.709 |
| $\sigma_{21}$ | 6.271 | 6.225 | 6.271 | 6.234 |
| $\sigma_{23}$ | 1.235 | 1.307 | 1.235 | 1.317 |
| $\psi_{2}^2$ | 6.039 6.044 6.044 6.043 | 6.039 6.043 6.043 6.043 |
| $\psi_{23}$ | -1.232 -1.238 -1.238 -1.236 | -1.232 -1.228 -1.228 -1.229 |
| $\beta_{30}$ | 0.340 0.338 0.324 0.323 | 0.241 0.240 0.235 0.235 |
| $\beta_{31}$ | 0.102 0.101 0.099 0.099 | 0.072 0.072 0.071 0.071 |
| $\beta_{32}$ | 0.148 0.147 0.145 0.145 | 0.105 0.105 0.104 0.104 |
| $\phi_{3}^2$ | 6.207 6.189 6.189 6.204 | 6.207 6.230 6.230 6.232 |
| $\phi_{33}$ | 9.371 | 9.475 | 9.371 | 9.452 |
| $\sigma_{31}$ | 9.489 9.484 9.482 9.484 | 9.489 9.484 9.484 9.484 |
| $\psi_{3}^2$ | 9.489 9.484 9.482 9.484 | 9.489 9.484 9.484 9.484 |

$\lambda = 0$

$\lambda = 1$
\[ \bar{\psi}_1 = 42.854, 49.792, 49.792, 49.748, 46.339, 46.071, 49.784, 49.784, 49.755, 46.263 \]
\[ \bar{\psi}_2 = 5.038, 4.371, 4.371, 4.372, 5.022, 5.042, 4.255, 4.255, 4.256, 4.993 \]
\[ \beta_{20} = 0.714, 0.711, 0.581, 0.583, 0.505, 0.505, 0.422, 0.425 \]
\[ \beta_{21} = 0.201, 0.194, 0.197, 0.187, 0.142, 0.138, 0.140, 0.140, 0.134 \]
\[ \beta_{22} = 0.277, 0.285, 0.275, 0.277, 0.195, 0.204, 0.196, 0.196, 0.199 \]
\[ \beta_{23} = 0.217, 0.208, 0.210, 0.197, 0.154, 0.148, 0.151, 0.151, 0.142 \]
\[ \bar{\psi}_1 = 46.040, 49.792, 49.792, 49.748, 46.339, 46.071, 49.784, 49.784, 49.755, 46.263 \]
\[ \bar{\psi}_2 = 5.038, 4.371, 4.371, 4.372, 5.022, 5.042, 4.255, 4.255, 4.256, 4.993 \]
\[ \bar{\psi}_3 = 67.336, 61.614, 67.067, 61.614, 67.067, 67.382, 61.741, 67.287, 61.741, 67.287 \]
\[ \lambda = 2 \]

Note: Parameters estimation based on (a) the estimated homoscedastic vars-Covs \( \hat{\sigma}_{\nu,mj}^2 \), \( \hat{\sigma}_{\mu,mj}^2 \), \( \hat{\sigma}_{u,mj}^2 \); (b) the estimated homoscedastic vars-Covs \( \hat{\sigma}_{\nu,mj}^2 \) and \( \hat{\sigma}_{u,mj}^2 \) and heteroscedastic vars-Covs \( \hat{\psi}_{aj}^2 \), whose the average value is \( \hat{\psi}_{mj} \); (c) the estimated homoscedastic vars-Covs \( \hat{\sigma}_{\nu,mj}^2 \) and \( \hat{\sigma}_{u,mj}^2 \) and heteroscedastic vars-Covs \( \hat{\psi}_{aj}^2 \), whose the average value is \( \hat{\phi}_{mj} \); (d) the estimated homoscedastic vars-Covs \( \hat{\sigma}_{\nu,mj}^2 \) and heteroscedastic vars-Covs \( \hat{\psi}_{aj}^2 \), where \( \hat{\phi}_{mj} \) is the average value of \( \hat{\phi}_{aj} \).

20
| Table 5. Simulation results on two-way SUR systems - WB procedure: standard errors of the estimated parameters and (average) estimated variances and covariances of the error components |
|----------------------------------|-----------------|-----------------|-----------------|
|                                  | $N = 250$, $T = 12$, and $n = 1031$ | $N = 500$, $T = 12$, and $n = 2062$ |
|                                  | true value     | homosc.         | heteroscedasticity on $\nu_{it}$, $\mu_{it}$, $\mu_{it}, u_{it}$ | true value     | homosc.         | heteroscedasticity on $\nu_{it}$, $\mu_{it}$, $\mu_{it}, u_{it}$ |
|                                  | (a)            | (b)             | (c)             | (d)            | (a)            | (b)             | (c)             | (d)            |
| $\lambda = 0$                   |                |                 |                 |                |                |                 |                 |                |
| $\beta_{10}$                    | 0.333          | 0.335           | 0.321           | 0.322          | 0.235          | 0.237           | 0.231           | 0.232          |
| $\beta_{11}$                    | 0.101          | 0.103           | 0.109           | 0.109          | 0.071          | 0.073           | 0.071           | 0.072          |
| $\beta_{12}$                    | 0.137          | 0.141           | 0.136           | 0.139          | 0.096          | 0.100           | 0.097           | 0.098          |
| $\phi_{1}^2$                    | 6.488          | 7.196           | 7.196           | 7.160          | 6.815          | 6.848           | 7.197           | 7.179          |
| $\phi_{12}$                     | 0.710          | 0.858           | 0.858           | 0.726          | 0.710          | 0.846           | 0.846           | 0.727          |
| $\psi_{13}$                     | 6.271          | 6.249           | 6.271           | 6.246          | 1.235          | 1.308           | 1.235           | 1.318          |
| $\lambda = 1$                   |                |                 |                 |                |                |                 |                 |                |
| $\beta_{10}$                    | 0.362          | 0.363           | 0.348           | 0.348          | 0.256          | 0.257           | 0.250           | 0.251          |
| $\beta_{11}$                    | 0.107          | 0.109           | 0.106           | 0.107          | 0.075          | 0.077           | 0.075           | 0.076          |
| $\beta_{12}$                    | 0.156          | 0.158           | 0.155           | 0.155          | 0.110          | 0.112           | 0.110           | 0.111          |
| $\phi_{1}^2$                    | 6.207          | 7.220           | 7.220           | 7.185          | 6.746          | 6.207           | 7.256           | 7.238          |
| $\phi_{12}$                     | 9.371          | 9.488           | 9.371           | 9.459          | 6.429          | 6.797           | 6.429           | 6.527          |
| $\psi_{13}$                     | 46.438         | 44.624          | 49.196          | 44.624         | 49.196         | 46.469          | 44.346          | 48.779          |
| $\psi_{12}$                     | 5.327          | 4.917           | 5.465           | 4.917          | 5.465          | 5.241           | 4.933           | 5.407           |
| $\psi_{13}$                     | 6.252          | 5.579           | 6.031           | 5.579          | 6.031          | 6.256           | 5.546           | 5.923           |

21
\[\lambda = 2\]

| \(\beta_{00}\) | 0.725 | 0.724 | 0.601 | 0.603 | 0.512 | 0.512 | 0.433 | 0.436 |
| \(\beta_{10}\) | 0.205 | 0.200 | 0.202 | 0.193 | 0.144 | 0.142 | 0.143 | 0.138 |
| \(\beta_{20}\) | 0.282 | 0.294 | 0.282 | 0.285 | 0.198 | 0.208 | 0.200 | 0.203 |
| \(\beta_{30}\) | 0.221 | 0.214 | 0.216 | 0.204 | 0.156 | 0.152 | 0.154 | 0.145 |

\[
\Phi_{21} = 46.040 \quad 50.510 \quad 50.510 \quad 50.283 \quad 46.234 \quad 46.071 \quad 50.478 \quad 50.478 \quad 50.359 \quad 46.290
\]

\[
\Phi_{23} = 5.038 \quad 4.487 \quad 4.487 \quad 4.465 \quad 5.000 \quad 5.042 \quad 4.382 \quad 4.382 \quad 4.371 \quad 4.992
\]

\[
\sigma_{\nu, 1}^2 = 6.271 \quad 6.549 \quad 6.271 \quad 6.402
\]

\[
\sigma_{\nu, 2}^2 = 1.235 \quad 1.313 \quad 1.235 \quad 1.314
\]

\[
\psi_{21} = 42.854 \quad 41.156 \quad 45.343 \quad 41.156 \quad 45.343 \quad 42.883 \quad 40.928 \quad 40.928 \quad 40.913 \quad 45.013
\]

\[
\bar{\phi}_{2} = 5.038 \quad 4.487 \quad 4.487 \quad 4.465 \quad 5.000 \quad 5.042 \quad 4.382 \quad 4.382 \quad 4.371 \quad 4.992
\]

\[
\sigma_{\nu, 1}^2 = 9.371 \quad 9.888 \quad 9.371 \quad 9.654
\]

\[
\psi_{21} = 36.736 \quad 64.326 \quad 70.879 \quad 64.326 \quad 70.879 \quad 67.382 \quad 64.079 \quad 70.526 \quad 64.079 \quad 70.526
\]

\[
\bar{\phi}_{2} = 44.046 \quad 50.658 \quad 50.658 \quad 50.567 \quad 44.044 \quad 44.076 \quad 50.902 \quad 50.902 \quad 50.902 \quad 44.044
\]

\[
\sigma_{\nu, 1}^2 = 9.371 \quad 9.888 \quad 9.371 \quad 9.654
\]

\[
\psi_{21} = 36.736 \quad 64.326 \quad 70.879 \quad 64.326 \quad 70.879 \quad 67.382 \quad 64.079 \quad 70.526 \quad 64.079 \quad 70.526
\]

Note: Parameters estimation based on (a) the estimated homoscedastic vars-Covs \(\hat{\sigma}_{\nu, mj}\), \(\hat{\sigma}_{\mu, mj}\), \(\hat{\sigma}_{\nu, mj}\); (b) the estimated homoscedastic vars-Covs \(\hat{\sigma}_{\nu, mj}\) and \(\hat{\sigma}_{\mu, mj}\); (c) the estimated homoscedastic vars-Covs \(\hat{\psi}_{a,mj}\); (d) the estimated homoscedastic vars-Covs \(\hat{\sigma}_{\nu, mj}\) and heteroscedastic vars-Covs \(\hat{\psi}_{a,mj}\), whose the average value is \(\hat{\psi}_{m}\).
As Table 3, Table 6 displays the ratios of the \( MSE \) of the estimators under consideration to the \( MSE \) of the true GLS estimator, i.e. it displays the measures of relative efficiency of the different estimators.

**Table 6. Relative efficiency of two-way SUR systems**

| \( N = 250, T = 12, and n = 1031 \) | \( N = 500, T = 12, and n = 2062 \) |
|--------------------------------------|--------------------------------------|
| homoscedasticity                        | heteroscedasticity on                | homoscedasticity                        | heteroscedasticity on                |
| \( u_\alpha \), \( \mu_\ell \) | \( u_\alpha \), \( \mu_\ell \) | \( u_\alpha \), \( \mu_\ell \) | \( u_\alpha \), \( \mu_\ell \) |
| \( \lambda = 0 \)         | \( y_1 \)     | 1.0001 | 1.0001 | 1.0003 | 1.0011 | 1.0000 | 1.0000 | 1.0001 | 1.0001 |
|                          | \( y_2 \)     | 1.0001 | 1.0001 | 1.0004 | 1.0014 | 1.0000 | 1.0001 | 1.0001 | 1.0001 |
|                          | \( y_3 \)     | 1.0000 | 1.0001 | 1.0003 | 1.0010 | 1.0000 | 1.0001 | 1.0001 | 1.0001 |
| \( \lambda = 1 \)         | \( y_1 \)     | 0.9995 | 0.9996 | 1.0041 | 1.0000 | 0.9997 | 0.9998 | 1.0003 | 1.0000 |
|                          | \( y_2 \)     | 0.9996 | 0.9997 | 1.0070 | 1.0005 | 0.9998 | 0.9998 | 1.0012 | 1.0001 |
|                          | \( y_3 \)     | 0.9993 | 0.9997 | 1.0037 | 1.0002 | 0.9997 | 0.9998 | 1.0008 | 1.0000 |
| \( \lambda = 2 \)         | \( y_1 \)     | 0.9990 | 0.9992 | 1.0022 | 0.9999 | 0.9995 | 0.9996 | 1.0004 | 1.0000 |
|                          | \( y_2 \)     | 0.9992 | 0.9994 | 1.0025 | 1.0001 | 0.9996 | 0.9996 | 1.0010 | 1.0000 |
|                          | \( y_3 \)     | 0.9989 | 0.9994 | 1.0018 | 1.0000 | 0.9995 | 0.9997 | 1.0004 | 1.0000 |

**QUE procedure**

| \( \lambda = 0 \)         | \( y_1 \)     | 1.0000 | 1.0001 | 1.0002 | 1.0002 | 1.0000 | 1.0000 | 1.0001 | 1.0001 |
|                          | \( y_2 \)     | 1.0001 | 1.0001 | 1.0003 | 1.0003 | 1.0000 | 1.0001 | 1.0001 | 1.0001 |
|                          | \( y_3 \)     | 1.0001 | 1.0001 | 1.0003 | 1.0003 | 1.0000 | 1.0001 | 1.0001 | 1.0001 |
| \( \lambda = 1 \)         | \( y_1 \)     | 0.9994 | 0.9995 | 1.0008 | 0.9999 | 0.9997 | 0.9997 | 0.9999 | 1.0000 |
|                          | \( y_2 \)     | 0.9996 | 0.9996 | 1.0016 | 1.0002 | 0.9998 | 0.9998 | 1.0007 | 1.0000 |
|                          | \( y_3 \)     | 0.9993 | 0.9996 | 1.0007 | 1.0001 | 0.9997 | 0.9998 | 1.0004 | 1.0000 |
| \( \lambda = 2 \)         | \( y_1 \)     | 0.9990 | 0.9991 | 1.0012 | 0.9999 | 0.9995 | 0.9995 | 1.0002 | 0.9999 |
|                          | \( y_2 \)     | 0.9992 | 0.9992 | 1.0024 | 1.0007 | 0.9996 | 0.9996 | 1.0009 | 1.0000 |
|                          | \( y_3 \)     | 0.9998 | 0.9993 | 1.0012 | 0.9999 | 0.9995 | 0.9996 | 1.0003 | 0.9999 |

**WB procedure**

| \( \lambda = 0 \)         | \( y_1 \)     | 1.0000 | 1.0001 | 1.0002 | 1.0002 | 1.0000 | 1.0000 | 1.0001 | 1.0001 |
|                          | \( y_2 \)     | 1.0001 | 1.0001 | 1.0003 | 1.0003 | 1.0000 | 1.0001 | 1.0001 | 1.0001 |
|                          | \( y_3 \)     | 1.0001 | 1.0001 | 1.0003 | 1.0003 | 1.0000 | 1.0001 | 1.0001 | 1.0001 |
| \( \lambda = 1 \)         | \( y_1 \)     | 0.9994 | 0.9995 | 1.0008 | 0.9999 | 0.9997 | 0.9997 | 0.9999 | 1.0000 |
|                          | \( y_2 \)     | 0.9996 | 0.9996 | 1.0016 | 1.0002 | 0.9998 | 0.9998 | 1.0007 | 1.0000 |
|                          | \( y_3 \)     | 0.9993 | 0.9996 | 1.0007 | 1.0001 | 0.9997 | 0.9998 | 1.0004 | 1.0000 |
| \( \lambda = 2 \)         | \( y_1 \)     | 0.9990 | 0.9991 | 1.0012 | 0.9999 | 0.9995 | 0.9995 | 1.0002 | 0.9999 |
|                          | \( y_2 \)     | 0.9992 | 0.9992 | 1.0024 | 1.0007 | 0.9996 | 0.9996 | 1.0009 | 1.0000 |
|                          | \( y_3 \)     | 0.9998 | 0.9993 | 1.0012 | 0.9999 | 0.9995 | 0.9996 | 1.0003 | 0.9999 |

*Note: Relative efficiency is defined as the ratio of the MSE of the estimator under consideration to the MSE of the true GLS estimator (computed considering the true vars-Covs \( \psi_{a,nj}, \phi_{a,nj}, \) and \( \sigma_{\nu,mj} \)). Note that values of the ratio both larger and smaller than 1 indicate a loss in efficiency: if the ratio is larger than 1, then the absolute value of the composite error term \( \epsilon = \mu_m + \nu_m + u_{mit} \) is larger than the true value; and if the ratio is smaller than 1, then the absolute value of the composite error term \( \epsilon \) is smaller than the true value.*

This table highlights that, as expected, with \( \lambda = 0 \) the most efficient procedure is the homoscedastic one, whereas with \( \lambda = 1, 2 \) the most efficient procedure is the one that considers both the remainder error and the individual-specific effect heteroscedastic. Note also that if the heteroscedasticity is low (i.e., with \( \lambda = 1 \)) the WB procedure is more efficient than the QUE procedure, whereas if the heteroscedasticity is high (i.e., with \( \lambda = 2 \)) the QUE procedure is more efficient than the WB procedure.
4. 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 EC models on these data, the heteroscedasticity problem may be very common, especially when individuals differ in size.

In this paper, we have derived suitable EC model estimators for heteroscedastic two-way single equations 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 heteroscedasticity is not taken into account, especially when both the individual-specific and remainder error components are heteroscedastic.

Appendix A: Fixed effects estimation assumptions

In the FE estimation the following assumptions are made.

FE.1 Strict exogeneity

The set of \((k-1)T_i\) explanatory variables for each individual \(x_{it} \equiv (x_{i1}, x_{i2}, \ldots, x_{iT_i})\) is uncorrelated with the idiosyncratic error \(u_{it}\) and the set of \((k-1)N_t\) explanatory variables in each time period \(x_{ct} \equiv (x_{t1}, x_{t2}, \ldots, x_{NT_t})\) is also uncorrelated with the same idiosyncratic error \(u_{it}\):

\[
E(u_{it} | x, \mu_i, \nu_t) = E(u_{it} | x_{it}, \mu_i, \nu_t) = E(u_{it} | x_{ct}, \mu_i, \nu_t) = 0,
\]

with \(x \equiv (x_{11}, \ldots, x_{1T_1}, x_{21}, \ldots, x_{2T_2}, \ldots, x_{N_1}, \ldots, x_{NT_N})\).

FE.2 Consistency

The \(W\) estimator in (3) is asymptotically well behaved, in the sense that the “adjusted” \((k-1) \times (k-1)\) outer product matrix \(X^T Q_{[\Delta]} X\) has the appropriate rank:

\[
\text{rank}(X^T Q_{[\Delta]} X) = k - 1.
\]

FE.3 No serial correlation

For each stratum \(a\) the conditional variance-covariance matrix of the idiosyncratic error terms \(u_{it}\) coincides with the unconditional one, and it is characterized by constant variances and zero covariances:

\[
E(u_a u_a^T | x_{it}, \mu_i, \nu_t) = \psi_a^2 I_n.
\]

Hence given the \(A \times 1\) vector \(\psi = (\psi_1^2, \psi_2^2, \ldots, \psi_A^2)^T\) we can define the \(n \times n\) matrix \(\Psi = \text{diag}(\Delta_\mu \Delta_\nu^T \psi)\) and the conditional variance-covariance matrix of \(u_a\) is

\[
E(u u^T | x, \mu_i, \nu_t) = \Psi.
\]

Details on the assumptions FE.1 and FE.2 can be found in Appendix A of Platoni et al. (2012).
In the RE estimation the following assumptions are made\textsuperscript{24}.

RE.1 A STRICT EXOGENEITY The set of $kT_i$ explanatory variables for each individual $x_{i0} \equiv (x_{i1}, x_{i2}, \ldots, x_{iT_i})$ is uncorrelated with the idiosyncratic error $u_{it}$ and the set of $kN_t$ explanatory variables in each time period $x_{ot} \equiv (x_{1t}, x_{2t}, \ldots, x_{N_t})$ is also uncorrelated with the same idiosyncratic error $u_{it}$:

$$E(u_{it} | x_{i0}, \mu_i, \nu_t) = E(u_{it} | x_{ot}, \mu_i, \nu_t) = E(u_{it} | x_{i0}) = 0,$$

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

RE.1.B AND RE.1.C ORTHOGONALITY CONDITIONS Both $\mu_i$ and $\nu_t$ are orthogonal to the corresponding sets of explanatory variables, that is the $kT_i$ explanatory variables for each individual $x_{i0}$ and the $kN_t$ explanatory variables in each time period $x_{ot}$:

$$E(\mu_i | x_{i0}) = E(\mu_i) = 0 \text{ and } E(\nu_t | x_{ot}) = E(\nu_t) = 0.$$

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

$$\text{rank} (X^T \Omega^{-1} X) = k.$$

RE.3 NO SERIAL CORRELATION For each stratum $a$ the conditional variance-covariance matrix of the idiosyncratic error terms $u_{it}$ is characterized by constant variances and zero covariances; in addition, whereas the variance of the time-specific effect $\nu_t$ is constant across strata, the variance of the individual-specific effect $\mu_i$ is constant within each stratum $a$:

a. $E(u_{it}^2 | x_{i(a)}, \mu_{i(a)}, \nu_t) = \psi_a^2$,  

b. $E(\mu_{i(a)}^2 | x_{i(a)}) = \phi_a^2$,  

c. $E(\nu_t^2 | x_t) = \sigma_{\nu}^2$.

APPENDIX C: ALTERNATIVE ROBUST STANDARD ERRORS

Let us re-index the individuals belonging to stratum $a$ as $i_a = 1_a, \ldots, N_a$, so that $T_{i_a}$ refers to the number of times the individual $i$ of the stratum $a$ is observed.

Since $u_{it} \sim (0, \psi_a^2)$, it is possible to obtain robust standard errors also by stacking

\textsuperscript{24} Details on the assumptions RE.1 and RE.2 can be found in Appendix B of Platoni et al. (2012).
Therefore, we can compute the $N_a \times 1$ vector $\tilde{e}_a = \bar{y}_a - \bar{X}_a \hat{\beta}_W$ and the robust asymptotic variance-covariance matrix of $\hat{\beta}_W$ is estimated by:

$$\text{var}(\hat{\beta}_W) = (X^T Q_D X)^{-1} \sum_{a=1}^A \left( \bar{X}_a \tilde{e}_a \tilde{e}_a^T \bar{X}_a \right) (X^T Q_D X)^{-1}. \quad (36)$$

**APPENDIX D: TECHNICAL APPENDIX ON QUE PROCEDURES**

D.1 *Adapted QUEs in (11)*

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

$$q_{na} = \left[ \begin{array}{c} f_a - \bar{f}_N \Delta_{\mu_a} - (\bar{f}_N^T \Delta_T - \bar{f}_N^T \Delta_{T_N}) Q^{-1} (\Delta_{\nu_a} - \Delta_{\mu_a} \Delta_{\nu_a}^{-1} \Delta_{T_N}) \end{array} \right]^T,$$

$$q_n = \left[ \begin{array}{c} f_a - \Delta_{\mu_a} \bar{f}_N - (\Delta_{\nu_a} - \Delta_{\mu_a} \Delta_{\nu_a}^{-1} \Delta_{T_N}) Q^{-1} (\bar{f}_N^T \Delta_T - \bar{f}_N^T \Delta_{T_N}) \end{array} \right]^T,$$

$$q_{Na} = \sum_{i \in I_a} T_i \bar{f}_a^2,$$

$$q_N = \left[ \begin{array}{c} \bar{f}_N \Delta_{\nu_a} \bar{f}_N \end{array} \right]^2 = \sum_{i=1}^N \bar{T}_i \bar{f}_a^2 = \sum_{a=1}^A \sum_{i=1}^T \bar{T}_i \bar{f}_a^2,$$

$$q_T = \left( \bar{f}_N \Delta_{\mu_a} \bar{f}_N \right)^2 = \sum_{i=1}^N \bar{T}_i \bar{f}_a^2,$$

where the elements of the $N \times 1$ matrix $\bar{f}_N$, are $\bar{f}_a = \frac{\sum_{i=1}^T \bar{T}_i \bar{f}_a}{\bar{N}_a}$, the elements of the $T \times 1$ matrix $\bar{f}_T$ are $\bar{f}_a = \frac{\sum_{i=1}^N \bar{T}_i \bar{f}_a}{\bar{N}_a}$, $\Delta_{\mu_a} = H_a \Delta_{\mu}$, and $\Delta_{\nu_a} = H_a \Delta_{\nu}$.

D.2 *Expected values in the single-equation case*

Referring to the identities in (11), and considering the $n \times n$ matrix $M = I_n - X(X^T Q_D X)^{-1} X^T Q_D$ (and then by definition $e = M \tilde{e} = M \epsilon$ and $f = X \tilde{e} = X \epsilon$), the expected value of $q_{na}$ is:

$$E(q_{na}) = \sum_{a=1}^A \sum_{i=1}^T \bar{T}_i \bar{f}_a^2 = \left( n_a - N_a - \tau_a \right) \psi_a^2 - k_a \bar{\psi}^2, \quad (38)$$
where \( \tau_a \equiv n_a - N_a - \text{tr}(H_a Q_{\Delta} H_a^T) \), \( k_a \equiv \text{tr}[(X^T Q_{\Delta} X)^{-1}X^T Q_{\Delta} H_a^T H_a Q_{\Delta} X] \), and \( \psi^2 \approx \sigma_a^2 \) is obtained by equating \( q_a \) to its expected value (see Wansbeek and Kapteyn, 1989; Davis, 2002), that is:

\[
E(q_a) = [n - N - (T - 1) - (k - 1)] \sigma_a^2.
\]

Moreover, the expected value of \( q_{N_a} \) is:

\[
E(q_{N_a}) = \text{tr}( \hat{\Omega}_{a} H_a E_a M \Omega M^T E_a H_a^T )
= \left( N_a - 2 \frac{n_a}{n} \right) \psi_a^2 + \left( k_{N_a} - k_0 \right) + n_a \frac{n_0}{n} \psi^2
+ \left( n_a - 2 \lambda_{2 \mu} \right) \phi_a^2 + n_a \lambda_\mu \phi^2,
\]

(40)

where \( \phi^2 \approx \sigma_\mu^2 \) is obtained jointly with \( \sigma_a^2 \) by equating \( q_N \) and \( q_T \) to their expected values, that is:

\[
E(q_N) = (N + k_N - k_0 - 1) \sigma_a^2 + (n - \lambda_\mu) \sigma_\mu^2 + (N - \lambda_\nu) \sigma_\nu^2,
E(q_T) = (T + k_T - k_0 - 1) \sigma_a^2 + (T - \lambda_\mu) \sigma_\mu^2 + (n - \lambda_\nu) \sigma_\nu^2,
\]

(41)

with \( k_N \equiv \text{tr}[(X^T Q_{\Delta} X)^{-1}X^T \Delta_{\mu} H_a \Delta_{\nu} X] \) and \( k_T \equiv \text{tr}[(X^T Q_{\Delta} X)^{-1}X^T \Delta_{\nu} H_a \Delta_{\mu} X] \).

In the case heteroscedasticity is only on the individual-specific disturbance the expected value of \( q_{N_a} \) is obtained as follows:

\[
E(q_{N_a}) = \left( N_a + k_{N_a} - k_0 + \frac{n_a}{n} \frac{k_0}{n_0} \right) \sigma_a^2 + \left( n_a - 2 \lambda_{2 \mu} \right) \phi_a^2 + n_a \lambda_\mu \phi^2
+ \left( N_a - 2 \lambda_{2 \nu} + \frac{n_a}{n} \lambda_\nu \right) \sigma_\nu^2,
\]

(42)

and, therefore,

\[
\phi_a^2 = q_{N_a} - \left( N_a + k_{N_a} - k_0 + \frac{n_a}{n} \frac{k_0}{n_0} \right) \sigma_a^2
- \frac{n_a - 2 \lambda_{2 \mu}}{n_a - 2 \lambda_{2 \mu}} \sigma_\mu^2
+ \left( N_a - 2 \lambda_{2 \nu} + \frac{n_a}{n} \lambda_\nu \right) \sigma_\nu^2.
\]

(43)

D.3 Expected values in the SUR systems case

Referring to the identities in (22), and considering the \( n \times n \) matrix \( M_m \equiv I_n - X_m (X_m^T Q_{\Delta} X_m)^{-1}X_m^T Q_{\Delta} \) (and then by definition \( e_m = M_m y_m = M_m \epsilon_m \) and \( f_m e_m^T = E_n e_m e_m^T E_n = E_n M_m Q_{\Delta} M_m^T E_n \)), the expected value of \( q_{n_{a, m_j}} \) is:

\[
E(q_{n_{a, m_j}}) = \text{tr}(H_a Q_{\Delta} E_a M_m M_{m_j}^T E_a Q_{\Delta} H_a^T )
\approx (n_a - N_a - \tau_a) \psi_{a, m_j} - (k_{a, m} + k_{a, j} - k_{a, m_j}) \psi_{m_j},
\]

(44)

where \( k_{a, m} \equiv \text{tr}[(X_m^T Q_{\Delta} X_m)^{-1}X_m^T Q_{\Delta} X_m] \) and \( k_{a, j} \equiv \text{tr}[(X_m^T Q_{\Delta} X_m)^{-1}X_m^T Q_{\Delta} X_m] \), and \( \psi_{m_j} \approx \sigma_{a, m_j} \) is obtained by equat-
ing $q_{n,mj}$ to its expected value (see Platoni et al., 2012):

$$E(q_{n,mj}) = |n - N - (T - 1) - (k_m - 1) - (k_j - 1) + k_{mj}| \sigma_{a,mj}.$$  

(45)

Moreover, the expected value of $q_{N_{a,mj}}$ is:

$$E(q_{N_{a,mj}}) = \text{tr}\left[ (N_a - 2 \frac{n_a}{n}) \psi_{a,mj} + \left( k_{N_{a,mj}} - k_{0_{a,mj}} + \frac{n_a}{n} k_{0_{a,mj}} + \frac{n_a}{n} \right) \psi_{mj} \right] + \left( n_a - 2 \lambda_{\mu} \right) \phi_{a,mj} + \left( n_a - 2 \lambda_{\nu_a} + \frac{n_a}{n} \lambda_{\nu} \right) \sigma_{v,mj},$$  

(46)

where $k_{N_{a,mj}} \equiv \text{tr}[X_m^T Q_m X_m]^{-1} X_m^T Q_m X_j (X_j^T Q_j X_j)^{-1} X_j^T P_j X_{a,mj}]$, $k_{0_{a,mj}} \equiv \sum_{a,mj} (X_m^T Q_m X_m)^{-1} X_m^T Q_m X_j (X_j^T Q_j X_j)^{-1} X_j^T P_j X_{a,mj}$, $k_{0_{a,mj}} \equiv \sum_{a,mj} (X_m^T Q_m X_m)^{-1} X_m^T Q_m X_j (X_j^T Q_j X_j)^{-1} X_j^T P_j X_{a,mj}$, and $\phi_{mj} \approx \sigma_{\mu,mj}$ is obtained jointly with $\sigma_{v,mj}$ by equating $q_{N_{a,mj}}$ and $q_{T,mj}$ to their expected values (see Platoni et al., 2012):

$$E(q_{N_{a,mj}}) = (N + k_{N_{a,mj}} - k_{0_{a,mj}} - 1) \sigma_{a,mj} + (n - \lambda_{\mu}) \sigma_{\mu,mj} + (n - \lambda_{\nu_a}) \sigma_{v,mj},$$  

(47)

$$E(q_{T,mj}) = (T + k_{T,mj} - k_{0_{a,mj}} - 1) \sigma_{a,mj} + (T - \lambda_{\mu}) \sigma_{\mu,mj} + (n - \lambda_{\nu_a}) \sigma_{v,mj},$$  

with $k_{N_{a,mj}} \equiv \text{tr}[X_m^T Q_m X_j]^{-1} X_m^T Q_m X_m (X_m^T Q_m X_m)^{-1} X_m^T \Delta_{\mu} \Delta_{\mu}^T X_j]$ and $k_{T,mj} \equiv \text{tr}[X_j^T Q_j X_j]^{-1} X_j^T Q_j X_m (X_m^T Q_m X_m)^{-1} X_m^T \Delta_{\nu_a} \Delta_{\nu_a}^T X_j].$

In the case heteroscedasticity is only on the individual-specific disturbance, the expected value of $q_{N_{a,mj}}$ is obtained differently as:

$$E(q_{N_{a,mj}}) = \left( N_a + k_{N_{a,mj}} - k_{0_{a,mj}} + \frac{n_a}{n} k_{0_{a,mj}} - \frac{n_a}{n} \right) \sigma_{a,mj}$$  

$$+ \left( n_a - 2 \lambda_{\mu_a} \right) \phi_{a,mj} + \left( n_a - 2 \lambda_{\nu_a} + \frac{n_a}{n} \lambda_{\nu} \right) \sigma_{v,mj},$$  

(48)

and, therefore,

$$\phi_{a,mj} = \frac{q_{N_{a,mj}} - \left( N_a + k_{N_{a,mj}} - k_{0_{a,mj}} + \frac{n_a}{n} k_{0_{a,mj}} - \frac{n_a}{n} \right) \sigma_{a,mj}}{n_a - 2 \lambda_{\mu_a}}$$  

$$+ \frac{- \frac{n_a}{n} \lambda_{\mu} \sigma_{a,mj} - \left( N_a - 2 \lambda_{\nu_a} + \frac{n_a}{n} \lambda_{\nu} \right) \sigma_{v,mj}}{n_a - 2 \lambda_{\mu_a}}.$$  

(49)
APPENDIX E: TECHNICAL APPENDIX ON WB PROCEDURE

In case of heteroscedasticity only on the individual-specific disturbance the estimator is:

\[
\hat{\Phi}_a = B^C_{\varepsilon_a} + \sum_{i \in I_a} \frac{T_i}{n} \sum_{j=1}^{N_a} \frac{T_j^2}{n} \hat{\Sigma}_\mu - \left( N_a - \sum_{i \in I_a} \frac{T_i}{n} \right) \hat{\Sigma}_u \sum_{i \in I_a} T_i, \tag{50}
\]

that would be an unbiased estimator of \(\Phi_a\) if the \(\varepsilon\)'s were known.

Using the centered residuals from the \(W\) estimation, the expected value of the between individuals (co)variations is:

\[
E \left( B^C_{f_a} \right) = \sum_{i \in I_a} T_i \Phi_a - \sum_{i \in I_a} \frac{T_i}{n} \sum_{j=1}^{N_a} \frac{T_j^2}{n} \Phi + \left( N_a - \sum_{i \in I_a} \frac{T_i}{n} \right) \Sigma_u, \tag{51}
\]

and therefore the estimator in (50), with \(B^C_{f_a}\) instead of \(B^C_{\varepsilon_a}\), is a consistent estimator of \(\Phi_a\).

APPENDIX F: ADDITIONAL TABLES

Due to the space limit it would be impossible (and unnecessary) to display 480 variance-covariance matrices as done in Table 1 for the single-equation case. Table 7 displays the estimated variances-covariances for the stratum \(a = 5\).
Simulation results on two-way SUR systems: estimated variances-covariances $\psi_{5,mj}$ and $\hat{\psi}_{5,mj}$

| $\lambda$ | $N = 250$, $T = 12$, and $n = 1031$ | $N = 500$, $T = 12$, and $n = 2062$ |
|---|---|---|
| 0 | 11 | 6.544 9.377 6.547 9.373 6.547 9.372 6.544 9.377 6.563 9.414 9.412 7.338 9.625 9.735 6.544 9.37 7.654 9.414 9.412 7.293 9.708 9.816 |
| 0 | 12 | 0.738 -1.048 0.741 -1.079 -1.079 0.805 -1.031 -1.019 0.73 8 -1.048 0.735 -1.045 -1.044 0.806 -1.003 -0.992 |
| 0 | 13 | 0.881 1.276 0.872 1.212 1.214 0.758 1.141 1.124 0.881 1.276 0.884 1.315 1.315 0.765 1.249 1.230 |
| 0 | 22 | 6.039 6.488 6.038 6.534 6.535 6.802 6.813 6.924 6.039 6.488 6.032 6.525 6.527 6.754 6.829 6.942 |
| 0 | 23 | 11 41.271 59.138 41.325 58.882 59.093 42.697 58.738 59.075 41.265 59.129 41.401 59.116 59.331 42.418 59.208 59.541 |
| 0 | 33 | 9.489 6.207 9.434 6.156 6.166 10.570 6.610 6.783 9.489 6.207 9.434 6.156 6.166 10.570 6.610 6.783 |
| 1 | 11 | 41.271 59.138 41.325 58.882 59.093 42.697 58.738 59.075 41.265 59.129 41.401 59.116 59.331 42.418 59.208 59.541 |
| 1 | 12 | 4.654 -6.609 4.681 -6.784 -6.756 4.713 -6.682 -6.644 4.654 -6.608 4.638 -6.630 -6.597 4.684 -6.561 -6.517 |
| 1 | 13 | 5.556 8.047 5.499 7.606 7.644 5.545 7.480 7.504 5.555 8.046 5.573 8.170 8.198 5.480 8.076 8.085 |
| 1 | 22 | 38.086 40.918 40.964 41.173 39.251 41.011 41.343 38.081 40.912 38.048 40.855 41.077 38.962 41.042 41.382 |
| 1 | 23 | -7.770 4.478 -7.794 4.643 4.599 -7.609 4.663 4.638 -7.769 4.477 -7.655 4.683 -7.567 4.725 4.688 |
| 1 | 33 | 59.844 39.146 59.545 38.501 38.873 61.184 38.723 39.274 59.836 39.140 59.857 39.049 39.377 61.157 39.405 39.910 |
| 2 | 11 | 105.914 151.765 106.110 150.610 151.619 108.699 149.776 150.938 105.884 151.722 106.262 151.181 152.214 107.884 150.922 152.084 |
| 2 | 12 | 11.944 -16.962 12.021 -17.445 -17.319 11.993 -17.245 -17.112 11.941 -16.957 11.903 -17.074 -16.934 11.903 -16.954 -16.804 |
| 2 | 13 | 14.259 20.652 14.120 19.452 19.614 14.219 19.232 19.384 14.255 20.646 14.300 20.841 20.974 14.268 20.697 20.814 |
| 2 | 22 | 97.740 105.007 97.849 104.667 105.633 99.808 104.325 105.437 97.713 104.977 97.656 104.342 105.354 98.982 104.333 105.473 |
| 2 | 23 | -19.940 11.491 -20.012 12.023 11.821 -19.786 11.966 11.779 -19.934 11.488 -19.647 12.075 11.835 -19.464 12.077 11.857 |
| 2 | 33 | 153.578 100.459 152.893 98.094 99.722 155.594 97.926 99.757 153.534 100.431 153.627 99.466 101.008 155.420 99.621 101.356 |

Note: $\psi_{5,mj}$ and $\hat{\psi}_{5,mj}$ are the true values of the var-Covs, $\psi_{5,mj}$ are the estimated vars-Covs of the remainder error $u$, $\hat{\psi}_{5,mj}$ are the estimated var-Covs of the individual-specific error $\mu$, computed on the basis of a remainder error either homoscedastic ($\hat{\sigma}_{u,mj}$) or heteroscedastic ($\psi_{u,mj}$).
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 heteroscedastic error components model (problem 88.2.2.).” *Econometric Theory*, 4 (2), 349-350.

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

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

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

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

Bester, C. A. and C. B. Hansen (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., C. Hsiao, and A. Pirotte (2006), “Heteroskedasticity and random coefficient model on panel data.” Working Papers ERMES, No. 0601, 51 p.

Bresson, G., C. Hsiao, and A. Pirotte (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.

Breusch, T. S. and A. R. Pagan (1979), “A simple test for heteroscedasticity and random coefficient variation.” *Econometrica*, 47 (5), 1287-1294.
Chib, S. (2008), “Panel data modeling and inference: a Bayesian primer.” In The Econometrics of Panel Data (Mátyás, L. and P. Sevestre, eds.), book series Advanced Studies in Theoretical and Applied Econometrics, Vol. 46, Chapter 15, 479-515, 3rd edition. Springer-Verlag, Berlin (GE).

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

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

Lejeune, B. (1996), “A full heteroscedastic 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. and T. Stengos (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. and A. Trognon (1978), “Heteroscedasticity 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. and E. L. Scott (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., P. Sckokai, and D. Moro (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 heteroscedastic error components regression models.” Economics Letters, 27 (4), 349-354.

Rao, P. S. R. S., J. Kaplan, and W. C. Cochran (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 heteroscedasticity in a model of seemingly unrelated regression equations with variance component.” Economics Letters, 5 (2), 149-153.

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

Wansbeek, T. (1989), “An alternative heteroscedastic error components model (problem 88.2.2.).” Econometric Theory, 5 (2), 326.

Wansbeek, T. and A. Kapteyn (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. The MIT Press, Cambridge, Massachusset (US) - London, England (UK).