EM Estimation of Dynamic Panel Data Models with Heteroskedastic Random Coefficients.

Andrea Nocera
Birkbeck, University of London*

11th July 2016

Abstract

In this paper, we show how to combine the EM algorithm with the Restricted Maximum Likelihood (REML) method to estimate dynamic heterogeneous panel data models. The EM-REML approach allows us to estimate iteratively both the average effects and the individual coefficients. It yields tractable closed form solutions. Compared to Swamy's random coefficients model, our method allows the random coefficients to have heteroskedastic variances and leads to an unbiased estimation of the variance components of the model. The estimation procedure can also be adapted to allow for cross-section dependence. Monte Carlo simulations reveal that the proposed estimator has good properties even in small samples. A novel approach to investigate heterogeneity of the sensitivity of sovereign spreads to government debt is presented.

*I am very grateful to Ron Smith and Z. Psaradakis for their comments and advice. I would like to thank seminar participants at the IAAE 2016 Annual Conference and at the "New Trends and Developments in Econometrics" 2016 Conference organized by the Banco de Portugal.
E-mail address: a.nocera@mail.bbk.ac.uk

1 Introduction

Nowadays panels in which both N (the number of units) and T (the number of time periods) are large, are quite common. As shown by Pesaran and Smith (1995), when regression coefficients differ across units, pooling and aggregating in a dynamic model give inconsistent and misleading estimates of the coefficients. As a solution, they propose estimating N time series separately. The expected value of the unit-specific coefficients can be estimated by averaging the OLS estimates for each unit. This procedure is called Mean Group estimation. Alternatively, if one sees the coefficients as randomly drawn from a common distribution, one can apply Swamy (1970) GLS estimation, that yields a weighted average of the individual OLS estimates. A good survey of the literature is provided by Hsiao and Pesaran (2004) and in Smith and Fuertes (2016). Swamy (1970) focuses on estimating the average effect while the random coefficients' residuals are treated as nuisance effects and conditioned out of the problem. However, the estimation of the random components of the model becomes crucial if the researcher wishes to predict future values of the dependent variable for a given individual or to describe the past behavior of a particular individual. Joint estimation of the average effect and individual parameters has been proposed by Lee and Griffiths (1979) and by Lindley and Smith (1972) in a Bayesian setting and has been further studied by Hsiao et al. (1998) and Maddala et al. (1997).

In this paper, following the seminal papers of Dempster et al. (1977) and Patterson and Thompson (1971), we propose to estimate dynamic heterogeneous panels by combining the EM algorithm with the Restricted Maximum Likelihood estimation, to obtain tractable closed form solutions of both fixed and random coefficients and the variance components. While Swamy (1970) GLS estimator can be obtained by maximizing the marginal likelihood function, Lee and Griffiths (1979) show that the random and fixed coefficients can be estimated by minimizing the sum of the weighted sum of squared residuals (SSR) of the regression model and the weighted SSR of the random coefficients' equation. In our framework, we derive an expression for the joint likelihood of the observed data and the random coefficients which will be used to infer not only on the average effects and unit-specific coefficients but also on their variances. Another interesting feature of the EM is that it allows us to make inference on the random coefficients population. Indeed, in general, it gives a probability distribution over the missing data.

The EM algorithm has recently gained attention in the finance literature. Harvey and Liu (2016) suggest a similar approach to ours to evaluate investment fund managers. The authors focus on estimating the fund-specific random intercepts population while the slope coefficients of the model are assumed to be fixed. Instead, we provide a more general framework where both the intercept and slope parameters are a function of a set of explanatory variables and are randomly drawn from a certain distribution. We derive an expression for the likelihood of the model accordingly. More importantly, our goals is to illustrate the advantages of the EM-REML approach in estimating heterogeneous panel data models, compared to the existing methods.

First, in the static case, estimating heterogeneous panels by EM-REML yields an unbiased estimation of the variance components. Therefore, this approach has an advantage when T is relatively small. In particular, the estimator of the variance-covariance matrix of the random coefficients proposed by Swamy (1970) is often negative definite. In such cases, the author suggests eliminating the second term of the right-hand side to obtain a non-negative definite matrix. Although not unbiased, this alternative estimator is consistent when T tends to infinity. Lee and Griffiths (1979) also derives a recursive system of equations as a solution to the maximization of the likelihood function of the data which incorporates the prior likelihood of the random coefficients. Differently from the latter, we consider the joint likelihood of the observed data and the random coefficients as an incomplete data problem (in a sense which will be more clear hereafter). We show that maximizing the expected value of the joint likelihood function with respect to the conditional distribution of the random coefficients residuals given the observed data is necessary to obtain an unbiased estimator of the random coefficients covariance matrix.

Many economic applications involve behavioral relationships which are dynamic in nature. Therefore, we define the data generating process as an ARDL panel model since one of the advantages of panel data is that they shed light on the dynamics of adjustment. However, including lagged dependent variables among the regressors raises a problem of endogeneity since they are a function of the individual effects. Consequently, the estimates of the coefficients will be biased and inconsistent even for large N and even if the error terms are not serially correlated. Hence, we resort to unbiasedness properties in the static case, while relying on the consistency properties (which depends upon T being large) when lagged values of the dependent variable are included among the regressors. This approach is in line with Maddala et al. (1997). Nevertheless, as it will be shown in the Monte Carlo analysis, the proposed method has good properties when estimating dynamic panels even when the sample size is relatively small. Compared to Swamy and the Mean Group estimators, the EM-REML method leads to a remarkable reduction of the bias of the estimates of the coefficients of the model and their variances. As will be clear later on, we need that $T > p + rank(W)$, where W is the matrix of explanatory variables including lagged values of the dependent variable and p is the number of lags included in the model.

In view of the above reasons, the EM-REML approach should be regarded as a valid alternative to bayesian estimation (as described in Maddala et al. (1997) and Hsiao et al. (1998)) in those cases in which the researcher wishes to make inference on the coefficients distribution while having little knowledge on what a sensible prior might be.

Second, our approach allows the conditional variances of the random coefficients residuals to have heteroskedasticity of unknown functional form and thus can be seen as a generalization of the one-way error component model where both the random effects and the regression disturbances are heteroskedastic, as described in Baltagi (2005).[1] Ignoring heteroskedasticity when it is present will still result in consistent estimates of the regression coefficients. Nevertheless, these estimates will not be efficient and their standard errors will be biased. The specifications where either only the random coefficients residuals or only the unit time-varying error components are assumed heteroskedastic can be seen as special case. Heteroskedasticity may occur in many economic applications in which it may be more realistic to model the variance of the random coefficients as varying across units.

For example, as shown in Mian and Sufi (2014), households with less wealth and higher debt are characterized by higher marginal propensity to consume (MPC). Similarly, the variance of the reaction of consumption to a shock in income may differ across individuals. For example, one could expect that the variation of unexplained MPC increases with debt and decreases with wealth, just as the MPC increases with debt and decreases with accumulated wealth. Households who own assets and who do not face any borrowing constraint can easily smooth their consumption. Furthermore, some of the determinants of MPC for wealthy households may have no explanatory power for MPC of «poor» households and/or viceversa. In such cases, the estimated variances of the unobserved idiosyncratic components of the random coefficients may vary largely across units.

In this paper, the proposed econometric methodology is used to study the determinants of the sensitivity of sovereign spreads with respect to government debt. First, we show that financial markets reactions to a shock in government spending are highly heterogenous. We then model such reactions as function of macroeconomics fundamentals and a set of explanatory variables which reflect the history of government debt and economic crises of various forms. We find that while country-specific macroeconomic indicators are significant determinants of sovereign credit risk, they do not have any significant impact on the sensitivity of spreads to debt. On the other hand, history of repayment plays an important role. A 1% increase in the percentage of year in default or restructuring domestic debt is associated with a 0.52% increase in the additional risk premium in response to an increase in debt.

Finally, the proposed estimation procedure is quite general and can accomodates recent developments in the dynamic heterogeneous panels literature, such as the CS-ARDL model developed by Chudik and Pesaran (2015).

The paper is organized as follows. Section 2 describes the regression model and its main assumptions. In Section 3 an expression for the likelihood of the complete data, which includes both the observed and the missing data, is derived. The Restricted Likelihood is also derived. Section 4 illustrates the use of EM algorithm and show how to perform the two steps of the EM algorithm, called the E-step and the M-step. Results from Monte Carlo

[1] This literature assumes T is small and N is large.

experiments are shown in Section 5. In Section 6, an application of the econometric model is reported. Finally, we conclude.

2 The Dynamic Heterogeneous Panel Model

We assume that the dependent variables, y_{it} 's, are generated by an ARDL(p,p) Panel Model[2]

$$y_{it} = c_i + \sum_{s=1}^{p} \phi_{is} y_{it-s} + \sum_{s=0}^{p} x'_{it-s} \beta_{is} + \varepsilon_{it} \tag{1}$$

for $i = 1, .., N$ and $t = 1, .., T$. x_{it} is a $K \times 1$ vector of exogenous regressors, $\phi_{is} \in \mathbb{R}$ and β_{is} is $K \times 1$. Let $\tilde{x}_{it} = (1,\ x'_{it})$ and $z_{it-s} = (y_{it-s},\ x'_{it-s})$ be two row vectors of explanatory variables both of dimension $1 \times (K+1)$, while $\psi_{i0} = (c_i,\ \beta'_{i0})'$ and $\psi_{is} = (\phi_{is},\ \beta'_{is})'$ are both $(K+1) \times 1$ vectors of coefficients, for $s = 1, .., p$. Let

$$\begin{aligned} \underset{(K^* \times 1)}{\psi_i} &= \begin{bmatrix} \psi'_{i0} & \cdots & \psi'_{ip} \end{bmatrix}' \\ \underset{(1 \times K^*)}{Z_{it}} &= \begin{bmatrix} \tilde{x}_{it} & z_{it-1} & \cdots & z_{it-p} \end{bmatrix} \end{aligned} \tag{2}$$

with $K^* = (K+1)(p+1)$. Using the first p observations as presample, equation (1) can be rewritten as

$$y_i = Z_i \psi_i + \varepsilon_i \tag{3}$$

where

$$\begin{aligned} \underset{((T-p) \times 1)}{y_i} &= \begin{bmatrix} y_{ip+1} & \cdots & y_{iT} \end{bmatrix}' \\ \underset{((T-p) \times K^*)}{Z_i} &= \begin{bmatrix} Z'_{ip+1} & \cdots & Z'_{iT} \end{bmatrix}' \end{aligned}$$

The coefficients ψ_i's are assumed to be constant over time but differ randomly across units. Individual-specific characteristics are the main source of heterogeneity in the parameters

$$\psi_i = \Gamma f_i + \gamma_i \tag{4}$$

[2]The analysis also holds for ARDL(p_i,q_i) in general. To make notation easier we set $p = p_i = q_j$ for all i and j.

where $\gamma_i = (\gamma'_{i0}, .., \gamma'_{ip})'$ is a $(K^* \times 1)$ vector of random coefficients residuals, Γ is a $K^* \times l$ matrix of unknown fixed parameters and f_i is a $l \times 1$ vector of observed explanatory variables that do not vary over time (for instance, Smith (2016) suggests using the group means of the x_{it}'s). The first element of f_i is one to allow for an intercept.

Equation (4) can be rewritten as

$$\psi_i = (f'_i \otimes I_{K^*})\,\bar{\Gamma} + \gamma_i \tag{5}$$

where $\bar{\Gamma} = vec\,(\Gamma)$, which is a K^*l-dimensional vector and $F_i = (f'_i \otimes I_{K^*})$ is a $K^* \times K^*l$ matrix.

Swamy (1970) random coefficients equation is a special case of (5), with $f_i = 1$ for all i and $\bar{\Gamma} = \psi$ is a $K^* \times 1$ vector of coefficients.

Substituting (5) into (3) yields

$$y_i = W_i\bar{\Gamma} + Z_i\gamma_i + \varepsilon_i \tag{6}$$

for $i = 1, .., N$, where $W_i = Z_iF_i$.

We assume that:

(i) The regression disturbances are independently distributed with zero means and variances that are constant over time but differ across units:

$$\varepsilon_{it} \sim i.i.d.N(0, \sigma^2_{\varepsilon_i}) \tag{7}$$

(ii) ψ_{is} and ε_{jt} are independent $\forall t, s$ and $\forall i, j$.

(iii) The regressors x_{it} and f_i are independent of the ε_{it} and γ_i.

(iv) The vector containing the random coefficients' residuals, $\gamma = (\gamma'_1, .., \gamma'_N)'$, is normally distributed as

$$\gamma \sim N(0, \Theta_\gamma) \tag{8}$$

where

$$\underset{(NK^* \times NK^*)}{\Theta_\gamma} = \begin{bmatrix} \triangle_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \triangle_N \end{bmatrix} \tag{9}$$

We allow $var\,(\gamma_i \mid f_i)$ to be different from $var\,(\gamma_j \mid f_j)$. In other words $\triangle_i \neq \triangle_j$, for $i \neq j$. Indeed, under assumption (4), it is likely that the variance of the random coefficients residuals is systematically larger for some units than for others depending on the values of the f_i's. For this reason, we allow for heteroskedasticity of unknown form. This phenomenon might be often observed in practice.

For example, when explaining the determinants of sovereign credit risks, the variance of the reaction of spreads to an increase in the debt-to-GDP ratio may be much higher for those countries with a weak repayment history in financial markets. Given higher uncertainty, financial markets are quite sensitive to even small shocks, making their decisions more volatile. Heteroskedasticity may also arise from the simple fact that the explanatory power of f_i in (4) varies largely across countries. Reputation, institutional features and other country-specific fundamentals may be important explanatory factors for some but not for all the countries under study. If the underlying factors which explain the sensitivity of spreads differ across units, treating the unobserved idiosyncratic components of the random coefficients as if they were drawn from an identical distribution can be naive.

Cross-Section Dependence and Estimation of Long-Run Effects. In many economic applications, the assumption of independence (across units) of the error terms may not hold. Such cross-section dependence (CSD) may arise from the fact that the errors are driven by a $r \times 1$ vector of unobserved common factors (ζ_t):

$$\varepsilon_{it} = \tau_i' \zeta_t + \epsilon_{it} \tag{10}$$

where τ_i is a $r \times 1$ vector of factor loadings and ϵ_{it} is an unobserved random error term independently distributed across i and t and which satisfies $E(\epsilon_{it}) = 0$ and $E(\epsilon_{it}^2) = \sigma_{\epsilon_i}^2$.

One way to allow for such common factors and remove the effect of CSD is to add cross-section averages of the dependent and independent variables of the model as shown by Pesaran (2006) in the static case and Chudik and Pesaran (2015) in the dynamic case. The regression model is now given by

$$y_{it} = c_i^* + \sum_{s=1}^{p} \phi_{is} y_{it-s} + \sum_{s=0}^{p} x_{it-s}' \beta_{is} + \sum_{s=0}^{p} \bar{z}_{t-s} \varphi_{is} + \epsilon_{it} \tag{11}$$

where $\bar{z}_{t-s} = (\bar{y}_{t-s},\ \bar{x}_{t-s}')$, $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}$ and $\bar{x}_t = N^{-1} \sum_{i=1}^{N} x_{it}$. Let $\tilde{x}_{it} = (1,\ x_{it}',\ \bar{z}_t)$ and $z_{it-s} = (y_{it-s},\ x_{it-s}',\ \bar{z}_{t-s})$, while $\psi_{i0} = (c_i^*,\ \beta_{i0}',\ \varphi_{i0}')'$ and $\psi_{is} = (\phi_{is},\ \beta_{is}',\ \varphi_{is}')'$. Equation (3) is now replaced by

$$y_i = Z_i \psi_i + \epsilon_i \tag{12}$$

where are Z_i and ψ_i are defined as above.

The vector of long-run effects of a set of regressors on the dependent variables can be estimated as

$$\hat{\theta}_i = \frac{\sum_{s=0}^{p} \hat{\beta}_{is}}{1 - \sum_{s=1}^{p} \hat{\phi}_{is}} \tag{13}$$

where $\hat{\beta}_{is}$ and $\hat{\phi}_{is}$ are the EM-REML estimates obtained as described hereafter.

3 Likelihood of the Complete data

Define the full set of (fixed) parameters to be estimated as

$$\theta = (\bar{\Gamma}', \sigma_{\varepsilon}^2, \omega')' = (\theta_1', \omega')'$$

where $\sigma_{\varepsilon}^2 = (\sigma_{\varepsilon 1}^2, .., \sigma_{\varepsilon N}^2)'$ and ω is a vector of ω_i's which are the vectors containing the non-zero elements of the covariance matrices $\triangle_i$, for $i = 1, .., N$.

We consider the random coefficients residulas, γ_i, as the vector of missing data (for $i = 1, .., N$). To estimate θ we would need to observe the complete data vector $(y', \gamma')'$.

Following the rules of probability, the log-likelihood of the complete data is given by

$$logL(y, \gamma; \theta) = logf(y \mid \gamma; \theta_1) + logf(\gamma; \omega) \tag{14}$$

which is the sum of the conditional log-likelihood of the observed data and the log-likelihood of the missing data.

Using assumptions (8) and (9) the joint log-likelihood of the vector of missing data can be written as[3]

$$logf(\gamma) = \sum_{i=1}^{N} logf(\gamma_i) = c_1 + \frac{1}{2}\sum_{i=1}^{N} log \mid \triangle_i^{-1} \mid - \frac{1}{2}\sum_{i=1}^{N} \gamma_i' \triangle_i^{-1} \gamma_i \tag{15}$$

To derive the likelihood of $y = (y_1', .., y_N')'$ given γ, we regard the value of the first p observations $(y_1, .., y_p)$ as deterministic.[4]

In that case, from (6) we can easily derive the conditional expectation and variance of y_i:

$$\begin{aligned} E(y_i \mid \gamma_i) &= W_i \bar{\Gamma} + Z_i \gamma_i \\ var(y_i \mid \gamma_i) &= var\,(\varepsilon_i) = R_i \end{aligned}$$

Under the assumption that both the regression error terms, ε_i, and the random coefficients residuals, γ_i, are independent and normally distributed, it follows that y_i is normally

[3]To make notation easier, we write $f(\gamma; \omega) = f(\gamma)$ and $f(y \mid \gamma)$ instead of $f(y \mid Z, \gamma; \theta_1)$.

[4]This assumption makes the computation of conditional maximum likelihood estimates much simpler. As noted in Hamilton (1994), as T gets large, the contribution of the first observations to the total likelihood is negligible. He also notes that the exact MLE and conditional MLE have the same large-sample distribution when $\mid \phi \mid < 1$, while only the conditional MLE is consistent when $\mid \phi \mid > 1$. Anderson and Hsiao (1981, 1982) argues against the assumption of fixed initial observations in panel with finite T. However, in line with Hsiao et al (1999), our estimators of the average coefficients have good properties even when T is relatively small as demonstrated by means of Monte Carlo experiments.

distributed and independent of y_j, for $i \neq j$. Therefore, the conditional log-likelihood of the observed data is given by

$$logf(y \mid \gamma) = \sum_{i=1}^{N} logf(y_i \mid \gamma_i) = c_2 - \frac{1}{2}\sum_{i=1}^{N} log \mid R_i \mid -\frac{1}{2}\sum_{i=1}^{N} \varepsilon_i' R_i^{-1} \varepsilon_i \tag{16}$$

where

$$\varepsilon_i = y_i - W_i\bar{\Gamma} - Z_i\gamma_i \tag{17}$$

Having found an explicit formulation for $logf(y \mid \gamma; \theta_1)$ and $logf(\gamma; \omega)$, we can derive an expression for the log-likelihood of the complete data by substituting (16) and (15) into (14).

At this point, we can make two important observations. First, θ_1 and ω are not functionally related, in the sense of Hayashi (2000, Section 7.1). This implies that $logf(\gamma; \omega)$ does not contain any information about θ_1 and similarly $logf(y \mid \gamma; \theta_1)$ does not contain any information about ω.

Second, as stated in Harville (1977), «the ML estimation takes no account of the loss in degrees of freedom that results from estimating the fixed effects» leading to biased estimators. In the next subsection, we eliminate this problem by using the restricted maximum likelihood (REML) approach, developed by Russell and Bradley (1958), Anderson and Bancroft (1952) and Thompson (1962) and described formally by Patterson and Thompson (1971).

3.1 Restricted Likelihood

Following Patterson and Thompson (1971), we can separate $logf(y_i \mid \gamma_i; \theta_1)$ in two parts, L_{1i} and L_{2i} say. By maximizing the former, we can estimate $\sigma^2_{\varepsilon_i}$. An estimate of $\bar{\Gamma}$ is obtained after maximizing L_{2i}.

The two parts can be obtained by defining two matrices S_i and Q_i such that the likelihood of $(y_i \mid \gamma_i)$ (for $i = 1, .., N$) can be decomposed as the product of the likelihoods of $S_i y_i$ and $Q_i y_i$, i.e.

$$logf(y_i \mid \gamma_i; \theta_1) = L_{1i} + L_{2i} \tag{18}$$

Such matrices must satisfy the following properties:

1. The rank of S_i is $T - p - K^*l$ while Q_i is a matrix of rank K^*l .
2. L_{1i} and L_{2i} are statistically independent, i.e.

$$cov\left(S_i y_i, Q_i y_i\right) = 0$$

3. The matrix S_i is chosen so that

$$E\left(S_i y_i\right) = 0, \quad i.e. \quad S_i W_i = 0$$

4. The matrix $Q_i W_i$ must be of rank K^*l.

Finding an expression for L_{1i}. Premutiplying both sides of (6) by S_i, we have

$$E\left(S_i y_i \mid \gamma_i\right) = S_i Z_i \gamma_i, \quad since \quad S_i W_i = 0$$

and

$$var\left(S_i y_i \mid \gamma_i\right) = S_i R_i S_i'$$

Therefore, the conditional log-likelihood of $S_i y_i$ is given by

$$L_{1i} = c_3 - \frac{1}{2} log \mid S_i R_i S_i' \mid - \frac{1}{2}\left(y_i - Z_i \gamma_i\right)' S_i' \left(S_i R_i S_i'\right)^{-1} S_i \left(y_i - Z_i \gamma_i\right) \tag{19}$$

Searle (1978) showed that "it does not matter what matrix S_i of this specification we use; the differentiable part of the log-likelihood is the same for all S_i's". In other words, the log-likelihood L_{1i} can be written without involving S_i. Indeed, as shown in Appendix A.1, equation (19) can be rewritten as

$$L_{1i} = c_3 - \frac{1}{2} log|R_i| - \frac{1}{2} log \mid W_i' R_i^{-1} W_i| - \frac{1}{2} \bar{\varepsilon}_i' R_i^{-1} \bar{\varepsilon}_i \tag{20}$$

where $\bar{\varepsilon}_i = y_i - W_i \hat{\bar{\Gamma}} - Z_i \gamma_i$.

Finding an expression for L_{2i}. Following Patterson and Thompson (1971), we can set $Q_i = W_i' R_i^{-1}$ since it satisfies $cov\left(S_i y_i, Q_i y_i\right) = 0$.

After premutiplying both sides of (6) by Q_i, we have

$$\begin{aligned} E\left(Q_i y_i \mid \gamma_i\right) &= W_i' R_i^{-1}\left(W_i \bar{\Gamma} + Z_i \gamma_i\right) \\ var\left(Q_i y_i \mid \gamma_i\right) &= W_i' R_i^{-1} W_i \end{aligned}$$

The log-likelihood of $Q_i y_i \mid \gamma_i$ is given by

$$L_{2i} = log f(Q_i y_i \mid \gamma_i) = c_4 - \frac{1}{2} log \mid W_i' R_i^{-1} W_i \mid - \frac{1}{2} \varepsilon_i' H_i \varepsilon_i \tag{21}$$

where $H_i = R_i^{-1} W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1}$ and the ε_i's are the regression residuals defined in (17).

4 EM-Algorithm

4.1 Generalities

Using equations (15) and (16), the log-likelihood of the complete data can be rewritten as

$$\begin{aligned} logL(y,\gamma;\theta) &= \textstyle\sum_{i=1}^{N}\{logL(y_i,\gamma_i;\theta)\} \\ &= \textstyle\sum_{i=1}^{N}\{logf(y_i \mid \gamma_i;\theta_1)+logf(\gamma_i;\omega_i)\} \end{aligned}$$

To obtain unbiased estimates of the variance components, following Patterson and Thompson (1971), we consider the complete-data (restricted) log-likelihood:

$$logL(y_i,\gamma_i;\theta) = L_{1i} + L_{2i} + logf(\gamma_i;\omega_i)$$

for $i = 1,..,N$, where $logf(y_i \mid \gamma_i;\theta_1)$ has been decomposed as shown in equation (18). Unfortunately, to obtain closed form solutions of the estimates, we cannot maximize directly the latter. Instead, by using the EM algorithm we are able to compute iteratively maximum likelihood estimates. On each iteration of the algorithm, there are two steps. The first step, called E-step, consists in finding the conditional expected value of the complete-data log-likelihood. Let $\theta^{(0)}$ be some initial value for θ. On the first iteration, this step requires computing

$$\begin{aligned} Q = Q(\theta;\theta^{(0)}) &= E_{\theta^{(0)}}\{logL(y,\gamma;\theta) \mid y\} \\ &= \textstyle\sum_{i=1}^{N} E_{\theta^{(0)}}\{logL(y_i,\gamma_i;\theta) \mid y_i\} = \sum_{i=1}^{N} Q_i \end{aligned} \tag{22}$$

where

$$Q_i = Q_i(\theta;\theta^{(0)}) \equiv E_{\theta^{(0)}}\{logL(y_i,\gamma_i;\theta) \mid y_i\} = Q_{1i} + Q_{2i} + Q_{3i}$$

and

$$\begin{aligned} Q_{1i} &= E_{\theta^{(0)}}\{L_{1i} \mid y_i\} \\ Q_{2i} &= E_{\theta^{(0)}}\{L_{2i} \mid y_i\} \\ Q_{3i} &= E_{\theta^{(0)}}\{logf(\gamma_i;\omega_i) \mid y_i\} \end{aligned} \tag{23}$$

In practice, we replace the missing variables, i.e. the random coefficients residuals (γ_i), by their conditional expectation given the observed data y_i and the current fit for θ.

The second step (M-Step) consists of maximizing $Q(\theta;\theta^{(0)})$ with respect to the parameters of interest, θ. That is, we choose $\theta^{(1)}$ such that

$$Q(\theta^{(1)};\theta^{(0)}) \geq Q(\theta;\theta^{(0)})$$

Once obtained the updated vector, we go to the E-step and iterate until convergence. In particular, on the bth iteration, the E-step requires the calculation of

$$Q(\theta;\theta^{(b-1)}) = E_{\theta^{(b-1)}}\left\{logL(y,\gamma;\theta) \mid y\right\}$$

while the M-step chooses $\theta^{(b)}$ as

$$\theta^{(b)} = \underset{\theta}{arg\ max}\ Q(\theta;\theta^{(b-1)})$$

More precisely, the estimates can be obtained as follows:

$$\begin{aligned}
\sigma^{2(b)}_{\varepsilon_i} &= \underset{\sigma^2_{\varepsilon_i}}{arg\ max}\ Q_{1i} \\
\bar{\Gamma}^{(b)} &= \underset{\bar{\Gamma}}{arg\ max}\ \textstyle\sum_{i=1}^{N} Q_{2i} \\
\omega^{(b)}_i &= \underset{\omega_i}{arg\ max}\ Q_{3i}
\end{aligned}$$

As noted by Patterson and Thompson (1971) and Harville (1977), no information is lost by basing inferences for $\sigma^2_{\varepsilon_i}$ only on Q_{1i} instead of $Q_{1i} + Q_{2i}$. Starting from suitable initial parameter values, the E- and M-steps are repeated until convergence, i.e. until the difference

$$L(y;\theta^{(b)}) - L(y;\theta^{(b-1)})$$

changes by an arbitrarily small amount, where $L(y;\theta)$ denotes the likelihood of the observed data.

4.2 Best Linear Unbiased Prediction

Within the EM algorithm, the random coefficients residuals, γ_i, are estimated by Best Linear Unbiased Prediction. Indeed, the E-step substitutes the γ_i's by their conditional expectation given the observed data y_i and the current fit for θ.

As shown in Appendix A.2, the conditional expectation of γ_i given the data is

$$\begin{aligned}
\hat{\gamma}_i = E\left(\gamma_i \mid y_i\right) &= \triangle_i Z_i' \left(Z_i \triangle_i Z_i' + R_i\right)^{-1} \left(y_i - W_i\bar{\Gamma}\right) \\
&= \left(Z_i' R_i^{-1} Z_i + \triangle_i^{-1}\right)^{-1} Z_i' R_i^{-1} \left(y_i - W_i\bar{\Gamma}\right)
\end{aligned} \tag{24}$$

which is the argument that maximizes the complete data likelihood, as defined in (14), with respect to γ_i. The conditional variance of γ_i is given by

$$V_{\gamma_i} = var\left(\gamma_i \mid y_i\right) = \left(Z_i' R_i^{-1} Z_i + \triangle_i^{-1}\right)^{-1} \tag{25}$$

which is equivalent to the inverse of $I\left(\gamma_i\right) = Z_i' R_i^{-1} Z_i + \triangle_i^{-1}$, the observed Fisher information matrix obtained by taking the second derivative of the log-likelihood of the complete data with respect to γ_i .

These two formulae have an empirical Bayesian interpretation. Given that γ is random, the likelihood $f(\gamma)$ can be thought as the "prior" density of γ. The posterior distribution of the latter is Normal with mean and variance given by (24) and (25) respectively.

Moreover, it can be noted from the first equality of (24) that this expression is equivalent to the predictor of the random coefficients residuals derived in Lee and Griffiths (1979) and Lindley and Smith (1972). Two differences emerge. The first concerns the way $\bar{\Gamma}$ and the other parameters are estimated. The second is that here we allow $\triangle_i \neq \triangle_j$, for $i \neq j$.

Given the current fit for θ at iteration b, we get

$$\hat{\gamma}_i^{(b)} = E_{\theta^{(b-1)}}\left(\gamma_i \mid y_i\right) = V_{\gamma_i}^{(b)} Z_i' R_{i(b-1)}^{-1}\left(y_i - W_i \bar{\Gamma}^{(b-1)}\right) \tag{26}$$

where

$$V_{\gamma_i}^{(b)} = var\left(\gamma_i \mid y_i, \theta^{(b-1)}\right) = \left(Z_i' R_{i(b-1)}^{-1} Z_i + \triangle_{i(b-1)}^{-1}\right)^{-1} \tag{27}$$

It is worth noting at this point some differences between the EM algorithm and Bayesian estimation. The EM gives a probability distribution over the unobserved data (i.e. the random coefficients residuals, γ) together with a point estimate for θ, the vector of average coefficients and variance components of the model. The latter is treated as being random in a fully Bayesian version.

The advantage compared to a Bayesian approach would be that there is no need to specify prior means and variances, the choice of which may not be always obvious and can have a large effect on the results when the sample size is small. Instead, within the EM algorithm, we can start with any initial value. The choice of the latter, differently from the prior choice in a Bayesian framework, does not affect the final result. As shown in Dempster, Laird, and Rubin (1977), the incomplete-data likelihood function $L(y;\theta)$ does not decrease after an EM iteration, that is $L(y;\theta^{(b)}) \geq L(y;\theta^{(b-1)})$ for $b = 1, 2, ...$ Nevertheless, this property does not guarantee convergence of the EM algorithm since it can get trapped in a local maximum. In complex case, Pawitan (2001) suggests to try several starting values or to start with a sensible estimate.

4.3 E-step

We can now derive the final expression for $Q_i(\theta)$. Detailed computations are shown in Appendix A.3.

E-step for L_{1i}. To obtain Q_{1i}, we take conditional expectation of both sides of (20). Substituting

$$E_{\theta^{(b-1)}}\left(\bar{\varepsilon}_i' R_i^{-1}\bar{\varepsilon}_i \mid y_i\right) = Tr\left(Z_i' R_i^{-1} Z_i V_{\gamma_i}^{(b)}\right) + \hat{\bar{\varepsilon}}_i' R_i^{-1}\hat{\bar{\varepsilon}}_i$$

where $\hat{\bar{\varepsilon}}_i = y_i - W_i\bar{\Gamma}^{(b)} - Z_i\hat{\gamma}_i^{(b)}$, yields

$$\begin{aligned} Q_{1i} = E_{\theta^{(b-1)}}\left(L_{1i} \mid y_i\right) \quad = \quad & c_3 - \tfrac{1}{2}log|R_i| - \tfrac{1}{2}log \mid W_i' R_i^{-1} W_i| \\ & -\tfrac{1}{2}Tr\left(Z_i' R_i^{-1} Z_i V_{\gamma_i}^{(b)}\right) - \tfrac{1}{2}\hat{\bar{\varepsilon}}_i' R_i^{-1}\hat{\bar{\varepsilon}}_i \end{aligned} \tag{28}$$

E-step for L_{2i}. To obtain Q_{2i}, we take the conditional expectation of (21). Substituting

$$E_{\theta^{(b-1)}}\left(\varepsilon_i' H_i \varepsilon_i \mid y_i\right) = Tr\left(Z_i' H_i Z_i V_{\gamma_i}^{(b)}\right) + \hat{\varepsilon}_i' H_i \hat{\varepsilon}_i$$

where $\hat{\varepsilon}_i = y_i - W_i\bar{\Gamma} - Z_i\hat{\gamma}_i^{(b)}$, yields

$$\begin{aligned} Q_{2i} = E_{\theta^{(b-1)}}\left(L_{2i} \mid y_i\right) \quad = \quad & c_4 - \tfrac{1}{2}log \mid W_i' R_i^{-1} W_i \mid \\ & -\tfrac{1}{2}Tr\left(Z_i' H_i Z_i V_{\gamma_i}^{(b)}\right) - \tfrac{1}{2}\hat{\varepsilon}_i' H_i \hat{\varepsilon}_i \end{aligned} \tag{29}$$

E-step for $logf(\gamma_i)$. Substituting

$$E_{\theta^{(b-1)}}\left(\gamma_i' \triangle_i^{-1}\gamma_i \mid y_i\right) = Tr\left(\triangle_i^{-1} V_{\gamma_i}^{(b)}\right) + \hat{\gamma}_i^{(b)'}\triangle_i^{-1}\hat{\gamma}_i^{(b)}$$

into $logf(\gamma_i)$, as defined in (15), yields

$$\begin{aligned} Q_{3i} = E_{\theta^{(b-1)}}\left(logf(\gamma_i) \mid y_i\right) \quad = \quad & -\tfrac{K^*}{2}log 2\pi + \tfrac{1}{2}log \mid \triangle_i^{-1} \mid \\ & -\tfrac{1}{2}Tr\left(\triangle_i^{-1} V_{\gamma_i}^{(b)}\right) - \tfrac{1}{2}\hat{\gamma}_i^{(b)'}\triangle_i^{-1}\hat{\gamma}_i^{(b)} \end{aligned} \tag{30}$$

4.4 M-step

The M-Step consists in maximizing (22) with respect to the parameters of interest, contained in θ.

Estimation of the Average Effect. An estimate of $\bar{\Gamma}$ can be obtained by maximizing $Q(\theta;\theta^{(b-1)})$ with respect to $\bar{\Gamma}$. This reduces to solving

$$\frac{\partial Q(\theta;\theta^{(b-1)})}{\partial\bar{\Gamma}} = \frac{\partial}{\partial\bar{\Gamma}}\left(-\frac{1}{2}\sum_{i=1}^{N}\hat{\varepsilon}_i' H_i \hat{\varepsilon}_i\right) = 0$$

The solution is

$$\bar{\Gamma}^{(b)} = \left(\sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1} W_i\right)^{-1} \sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1}\left(y_i - Z_i\hat{\gamma}_i^{(b)}\right) \tag{31}$$

This is equivalent to the GLS estimation of $\bar{\Gamma}$ when the model is given by

$$y_i^* = W_i\bar{\Gamma} + \varepsilon_i$$

where $y_i^* = y_i - Z_i\gamma_i$, as if the γ_i's where known.

Unlike the Newton-Raphson and related methods, the EM algorithm does not automatically provide an estimate of the covariance matrix of the maximum likelihood estimate. However, in our random coefficient model, the Fisher information matrix $I\left(\bar{\Gamma}^{(B)}\right)$ can be easily derived by evaluating analytically the second-order derivatives of the incomplete-data log-likelihood (e.g. $logf(y;\theta)$) since computations are not tedious. Therefore, the standard errors of $\bar{\Gamma}^{(B)}$ can be computed as the square root of the diagonal elements of

$$I\left(\bar{\Gamma}^{(B)}\right)^{-1} = \left(\sum_{i=1}^{N} W_i' V_{i(B)}^{-1} W_i\right)^{-1} \tag{32}$$

where $V_i = var(y_i) = Z_i\triangle_i Z_i' + R_i$ while B denotes the last iteration of the algorithm.

Estimation of the Variances of the Residual Terms. An estimate of $\sigma^2_{\varepsilon_i}$ can be derived by maximizing (22). Because Q_{3i} is not a function of $\sigma^2_{\varepsilon_i}$ and given that no information is lost by neglecting Q_{2i} (as shown by Patterson and Thompson, 1971), we focus only on Q_{1i}, as defined in (28). Substituting $R_i = var\left(\varepsilon_i\right) = \sigma^2_{\varepsilon_i} I_{T-p}$ into (28) and equating the first derivative of the latter with respect to $\sigma^2_{\varepsilon_i}$ to zero, yields

$$\sigma^{2(b)}_{\varepsilon_i} = \frac{\hat{\hat{\varepsilon}}_i'\hat{\hat{\varepsilon}}_i + Tr\left(Z_i' Z_i V_{\gamma_i}^{(b)}\right)}{T - p - r(W_i)} \tag{33}$$

where $\hat{\hat{\varepsilon}}_i = y_i - W_i\bar{\Gamma}^{(b)} - Z_i\hat{\gamma}_i^{(b)}$. A necessary condition to be satisfied is $T > p + rank(W_i)$.

Estimation of the Variance of the Random Coefficients. Under the Law of Total Variance, the unconditional variance of γ_i can be written as

$$\begin{aligned} \triangle_i = var\left(\gamma_i\right) &= var\left[E\left(\gamma_i \mid y_i\right)\right] + E\left[var\left(\gamma_i \mid y_i\right)\right] \\ &= var\left(\hat{\gamma}_i\right) + E\left(V_{\gamma_i}\right) \end{aligned} \tag{34}$$

Therefore, it can be shown that

$$\hat{\triangle}_i = \hat{\gamma}_i \hat{\gamma}_i' + V_{\gamma_i} \tag{35}$$

is an unbiased estimator of $\triangle_i$.[5] Indeed, taking expectation of both sides of (35) and using (34)

$$E\left(\hat{\triangle}_i\right) = E\left(\hat{\gamma}_i\hat{\gamma}_i'\right) + E\left(V_{\gamma_i}\right) = var\left(\hat{\gamma}_i\right) + E\left(V_{\gamma_i}\right) = \triangle_i$$

It turns out that the EM-REML estimator of the variance-covariance matrix of the random coefficients residuals (which is the argument which maximizes (30) with respect to $\triangle_i$ for $i = 1, ., N$) is equal to[6]

$$\triangle_i^{(b)} = \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} + V_{\gamma_i}^{(b)} \tag{36}$$

which can be obtained by substituting the unknown parameters in (35) with their current fit in the EM algorithm.

We can now compare (36) with the Swamy (1970) and Lee and Griffiths (1979) estimators of the random coefficients residuals' variance-covariance matrix .

Under the assumption that $\triangle_i = \triangle_j = \triangle$, $\forall i.j$, Swamy (1970) suggested estimating $\triangle = var\left(\gamma_i\right)$ as

$$\begin{aligned} \hat{\triangle} = & \tfrac{1}{N-1}\sum_{i=1}^N \left(\hat{\psi}_i - N^{-1}\sum_{i=1}^N \hat{\psi}_i\right)\left(\hat{\psi}_i - N^{-1}\sum_{i=1}^N \hat{\psi}_i\right)' \\ & -\tfrac{1}{N}\sum_{i=1}^N \hat{\sigma}_i^2 (Z_i'Z_i)^{-1} \end{aligned} \tag{37}$$

where $\hat{\psi}_i$ are obtained by estimating N time series separately by OLS and

$$\hat{\sigma}_i^2 = \frac{\left(y_i - Z_i\hat{\psi}_i\right)'\left(y_i - Z_i\hat{\psi}_i\right)}{T - K^*} \tag{38}$$

are the OLS estimated variances of the residual terms. However, this estimator is not necessarily nonnegative definite. Therefore, if that is the case the author suggests considering only

$$\hat{\triangle} = \frac{1}{N-1}\sum_{i=1}^N \left(\hat{\psi}_i - N^{-1}\sum_{i=1}^N \hat{\psi}_i\right)\left(\hat{\psi}_i - N^{-1}\sum_{i=1}^N \hat{\psi}_i\right)' \tag{39}$$

[5] Note that $var\left(\hat{\gamma}_i\right) = E\left(\hat{\gamma}_i\hat{\gamma}_i'\right) + E\left(\hat{\gamma}_i\right)E\left(\hat{\gamma}_i'\right) = E\left(\hat{\gamma}_i\hat{\gamma}_i'\right)$ since $E\left(\hat{\gamma}_i\right) = E_{y_i}\left(E\left(\gamma_i \mid y_i\right)\right) = E\left(\gamma_i\right) = 0$.

[6] See Appendix A.4, for step by step computations.

Although not unbiased, the latter estimator is nonnegative definite and consistent when T tends to infinity.

When the variances are unknown, Lee and Griffiths (1979) suggest maximizing the joint likelihood of the random coefficients and the observed data given in (14) with respect to the unknown parameters of the model, to get the following iterative solutions of the variance components:

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{\left(\hat{\hat{\varepsilon}}'_i \hat{\hat{\varepsilon}}_i\right)}{T}$$

and

$$\hat{\triangle} = \frac{1}{N} \sum_{i=1}^{N} \hat{\gamma}_i \hat{\gamma}'_i \tag{40}$$

We have seen that within the EM algorithm the random coefficients residuals, γ_i, are considered as missing data and replaced by their conditional expectation given the data, which yields the BLUP of γ_i. By doing so, the incomplete data problem becomes a complete data one. At the same time, it has been shown that maximizing the joint likelihood of the observed data and random coefficients residuals, as given in (14), with respect to γ_i yields an estimator $\hat{\gamma}_i$ which is equivalent to the BLUP of γ_i. This is the approach used by Lee and Griffiths. Nevertheless, ignoring that the joint likelihood, $f(y_i, \gamma_i)$, is an incomplete data problem and considering the random coefficients residuals as parameters to be estimated comes at a price. Indeed, in the latter case, the expected value of the estimated random coefficients residuals variance-covariance matrix, (40), is only equal to the variance of the conditional expectation of γ_i, i.e. $var\left[E\left(\gamma_i \mid y_i\right)\right]$, and therefore is a biased estimator of $\triangle$, the unconditional variance of γ_i. Instead, by maximizing the condtional expectation of the complete-data (restricted) likelihood the EM-REML algorithm yields an unbiased estimator of $\triangle$. Consequently, our approach has an advantage over both Swamy (1970) and Lee and Griffiths (1979) when T is not too large.

4.5 EM-REML Algorithm - Complete Iterations

The EM algorithm steps can be summarised as follows. We start with some initial guess: $\psi^{(0)}$, $\triangle_{i(0)}$ and $R_{i(0)} = \sigma^{2(0)}_{\varepsilon_i} I_{T-p}$. In the simplest case the researcher may wish to set $\psi^{(0)}$ as the pooled OLS while $\triangle_{i(0)}$ and $R_{i(0)}$ can be set equal to the identity matrix for all i. Alternatively, initial estimates which should guarantee faster convergence are the Swamy GLS estimator for $\psi^{(0)}$, (39) for $\triangle_{i(0)}$ and (38) for the $\sigma^{2(0)}_{\varepsilon_i}$'s. Then, for $b = 1, 2, ..$

1. First, we compute $E_{\theta^{(b-1)}}\left(\gamma_i \mid y_i\right)$ and $var\left(\gamma_i \mid y_i, \theta^{(b-1)}\right)$:

$$V^{(b)}_{\gamma_i} = \left(Z'_i R^{-1}_{i(b-1)} Z_i + \triangle^{-1}_{i(b-1)}\right)^{-1} \tag{41}$$

$$\hat{\gamma}_i^{(b)} = V_{\gamma_i}^{(b)} Z_i' R_{i(b-1)}^{-1} \left(y_i - W_i \bar{\Gamma}^{(b-1)} \right) \tag{42}$$

2. The average coefficients are given by

$$\bar{\Gamma}^{(b)} = \left(\sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1} W_i \right)^{-1} \sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1} \left(y_i - Z_i \hat{\gamma}_i^{(b)} \right) \tag{43}$$

3. Finally, we can compute, the variance components:

$$\sigma_{\varepsilon_i}^{2(b)} = \frac{\hat{\hat{\varepsilon}}_i' \hat{\hat{\varepsilon}}_i + Tr \left(Z_i' Z_i V_{\gamma_i}^{(b)} \right)}{T - p - r(W_i)} \tag{44}$$

where $\hat{\hat{\varepsilon}}_i = y_i - W_i \bar{\Gamma}^{(b)} - Z_i \hat{\gamma}_i^{(b)}$ and

$$\triangle_i^{(b)} = V_{\gamma_i}^{(b)} + \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} \tag{45}$$

The iterations continue until the difference $L(y;\theta^{(b)}) - L(y;\theta^{(b-1)})$ changes only by an arbitrary small amount, where $L(y;\theta)$ is the likelihood of the observed data. After convergence, the variance-covariance matrix of $\bar{\Gamma}^{(B)}$ (where B denotes the last iteration) can be computed as

$$v\hat{a}r \left(\bar{\Gamma}^{(B)} \right) = \left(\sum_{i=1}^{N} W_i' V_{i(B)}^{-1} W_i \right)^{-1} \tag{46}$$

where $V_i = var(y_i) = Z_i \triangle_i Z_i' + R_i$.

5 Monte-Carlo Results

In this section, we employ Monte-Carlo experiments to examine and compare the small sample properties of the proposed EM-REML method versus the most commonly used techniques in panel time series analysis, such as Swamy's random coefficient model and the Mean Group estimation proposed by Pesaran and Smith (1995) with a particular focus on the bias of the average effects and of the variance components of the models.

The data generating process used in the Monte Carlo analysis is given by

$$\begin{aligned} y_{it} &= c_i + \beta_i x_{it} + \phi_i y_{it-1} + \varepsilon_{it} \\ x_{it} &= c_{x,i}(1-\rho) + \rho x_{it-1} + u_{it} \end{aligned} \tag{47}$$

where

$$\begin{array}{rcl} u_{it} & \sim & N(0,1) \\ \varepsilon_{it} & \sim & i.i.d.N(0,\sigma^2_{\varepsilon_i}) \\ c_{x,i} & \sim & N(1,1) \end{array} \tag{48}$$

We set $\rho = 0.6$. Once generated, the x_{it} are taken as fixed across different replications. The regression residuals' standard deviation (σ_{ε_i}) are assumed to be uniformly distributed in the interval [0.5, 1.5]. The coefficients differ randomly across units according to

$$\begin{array}{rcl} c_i & = & \mu + \gamma_{1i} \\ \beta_i & = & \beta + \gamma_{2i} \\ \phi_i & = & \phi + \gamma_{3i} \end{array} \tag{49}$$

where $\psi = (\mu, \beta, \phi) = (0.2, 0.1, 0.5)$. Moreover, we assume that $\gamma_{ji} \sim IN(0, \sigma^2_{\gamma_j,i})$, for $j = 1, 2, 3$, where

$$\sigma^2_{\gamma_j,i} = var\left(\gamma_{ji}\right) = \left(\theta_j \bar{x}_i\right)^2$$

with $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$ and $\theta_j \sim \chi_{(1)}$. We set θ_3 to be the smallest in order to avoid explosive behaviours. For each $i = 1, .., N$ we eliminate the first 200 observations generated in the experiments to minimise the effect of initial observations.

The results shown in Appendix A.5 are based on 1000 replications. Tables 1 to 3 report the bias of each coefficient, the standard errors of such biases and an overall measure of the bias which is chosen to be the norm of the bias of ψ. The root mean square errors (RMSE) are also given. Regarding the variance components $\sigma^2_{\gamma_j,i}$, instead of providing the bias of each estimator[7] for $i = 1, .., N$, we consider the the norm of the bias of the $N \times 1$ vector $\hat{\sigma}^2_{\gamma_j}$, whose ith element is $\hat{\sigma}^2_{\gamma_j,i}$, for $j = 1, 2, 3$. Similarly, we report the average across units of the RMSE of the estimators of the variance components $\hat{\sigma}^2_{\gamma_j,i}$.

The EM-REML does quite well even in small samples. It outperforms both Swamy's and the Mean Group estimator in term of bias of both the average effects and the variance components.

As shown in Table 4, when $T = 10$, the bias of autoregressive coefficients estimated by EM-REML vary between -0.086 and 0.026 as N goes from 10 to 80. On the contrary, when estimating the model using Swamy GLS method, the bias takes values between -0.249 and -0.196. The bias is even larger when considering the Mean Group estimator, between -0.330 and -0.290. The advantages persist when T increases. Results when $T = 80$ are reported in Table 6. Focusing again on the autoregressive coefficient, the bias ranges between 0.014 and 0.001 for the EM-REML case. The bias varies from -0.021 to -0.005 when using

[7] In the columns «N-Time Series - MG» of Table 1 to 3, the estimated variances are obtained estimating N time series separately.

Swamy's estimator, and from -0.043 and -0.037 when using the MG estimator. The RMSE associated to ϕ_i is much smaller when estimating the model by EM-REML when $T = 10$. This advantage reduces when $T = 30$. Instead, when $T = 80$ and N is equal to 10 or 30, the RMSE is relatively smaller when using Swamy's estimator. The advantages remain when comparing the EM-REML approach with the MG estimation. The bias of the variance of random coefficients residuals is smaller across different size of T and N when our proposed approached is used. A smaller bias is sometimes associated to larger RMSE.

In general, these gains can be explained by two factors. First, in most of the experiments the Swamy's covariance matrix estimator given in (37) is negative definite. Using (39) instead of the latter results in a biased estimator. Only when both T and N are large the probability of (37) being negative definite are small. Although in large samples the differences with the EM-REML reduces, the latter continues to have an advantage. In fact, the EM-REML variances' estimator is the most efficient among the competitors since it accounts for heteroskedasticity of the variance of the random coefficients. Ignoring such heteroskedasticity yields biased estimators of the variance components.

6 Application

6.1 Introduction

In this section, we demonstrate an application to motivate the use of EM-REML estimation of dynamic heterogeneous panels.

Reinhart et al. (2003), studying sovereigns' credit histories since the early nineteenth century, argue that an important subgroup of middle-income countries has been "systematically" afflicted by what they call "debt intolerance". Even though their debt-to-GDP ratios are considerably lower than those of several high-income countries, these economies are considered to be riskier and unable to tolerate as much debt.

We corroborate this argument by first showing that the response of sovereign spreads to changes in government debt (which we also refer to as the "sensitivity" of financial markets during episodes of debt growth) is highly heterogeneous. It is only statistically significant for a small sub-group of countries. We ask why this is so by modelling the sensitivity of spreads as function of macroeconomics fundamentals and a set of explanatory variables which reflect the history of government debt and economic crises of various forms. The more pervasive the phenomenon of serial default is (i.e. the weaker the reputation), the stronger the reaction of financial markets when debt increases. We quantify such reactions.

We depart from the literature on the determinants of sovereign spreads in several ways.[8]

[8]See for instance, Akitoby and Stratmann (2008), Bellas et al. (2010), Edwards (1984), Eichengreen and

First, instead of considering only one group of countries (e.g. emerging markets), we collect quarterly data for a panel of 17 emerging market economies and 21 developed countries over 22 years (1994Q1-2015Q4).[9] Second, given that we are comparing countries with very different characteristics, even within group, we allow for heterogeneity rather than pooling.

Finally, the focus of this paper is on understanding which factors determine the additional risk premium to charge during episodes of debt growth. Assume that sovereign spreads are a function of debt-to-GDP ratio, a proxy for history of default and other macroeconomic fundamentals. Rather than looking at how spreads change with respect to one variable while debt-to-GDP and the remaining covariates are held constant (i.e. partial effect), we investigate which country characteristics significantly affect the magnitude of sovereign spreads' reaction to changes in debt. Studying the sensitivity of financial markets during episode of debt growth is crucial to understand how emerging markets can borrow at level comparable to more developed economies without having to pay unsustainable interest rates and therefore it is important to shed some light on the debt intolerance problem.

6.2 The Empirical Model

Following Edwards (1984), we assume that the spread over U.S. (or German) Treasuries can be explained by a set of macroeconomic indicators. We focus on real GDP growth, the growth rate of CPI and the general gross government debt as a percentage of GDP. J.P. Morgan's Emerging Markets Bond Index Global (EMBI Global) is our measure of government bond yields for emerging markets.

Because linear interdependencies may exist among these time series, we can assume they follow a VAR(p) process. Given that the spreads are observed at a daily frequency, it is reasonable to think that they react near-instantaneously to shocks and news. Therefore, considering the variables under study, we can assume that the economy possesses a recursive structure where spreads are ordered last. The last equation of the recursive system can be written as

$$y_{it} = \phi_i y_{it-1} + x'_{it}\beta_i + \mu_i + \varepsilon_{it} \tag{50}$$

for $i = 1, .., N$ and $t = 1, .., T$. We study both the case where government spreads (y_{it}) and debt-to-GDP are in first difference and the case where they are not differenced. Given that they lead to similar conclusions, we only report results from the first case.

The number of lags has been selected using AIC and BIC criteria, which give very similar results.

Mody (2000) and Hilscher and Nosbusch (2010), among others.

[9] The panel is slightly unbalanced. The individual time observations vary between $60 \leq T_i \leq 87$.

6.3 Parameter Equality Tests

Before estimating the model, we employ some homogeneity tests to show that both the slope and the intercept parameters are heterogenous across countries. Accounting for such heterogeneity is very important. Indeed, as shown in Pesaran and Smith (1995), if the DGP includes lagged values of the dependent variables among the explanatory variables, as it is in our case, then pooling give inconsistent and potentially highly misleading estimates of the coefficients when the coefficients differ across units. This problem does not arise in the static case, where pooling estimation give unbiased estimates of coefficient means when the coefficients differ randomly. We then show that also the random coefficients variances differ across units. Accounting for such heteroskedastictiy is important when testing hypotheses. In fact, although consistent, the estimates of the regression coefficients which ignore heteroskedasticity will not be efficient and their standard errors will be biased.

6.3.1 Test for Heterogeneous Coefficients

To test the null hypothesis $H_0: \ \psi_1 = .. = \psi_N = \psi$ (i.e. to test whether the coefficient vectors ψ_i are constant across units), we can use the following test proposed in Swamy (1970)[10]:

$$F = \frac{1}{(N-1)} \sum_{i=1}^{N} F_i \sim F\left(K^*(N-1), (\sum_{i=1}^{N} T_i - NK^*) \right) \tag{51}$$

where

$$F_i = \frac{\left(\hat{\psi}_i - \hat{\psi}\right)' Z_i' Z_i \left(\hat{\psi}_i - \hat{\psi}\right)}{K^* \hat{\sigma}^2_{\varepsilon_i}}$$

and

$$\hat{\psi} = \left(\sum_{i=1}^{N} \frac{Z_i' Z_i}{\hat{\sigma}_i^2} \right)^{-1} \left(\sum_{i=1}^{N} \frac{Z_i' Z_i}{\hat{\sigma}_i^2} \hat{\psi}_i \right) = \left(\sum_{i=1}^{N} \frac{Z_i' Z_i}{\hat{\sigma}_i^2} \right)^{-1} \left(\sum_{i=1}^{N} \frac{Z_i' y_i}{\hat{\sigma}_i^2} \right)$$

K^{-*} is the dimension of ψ. The $\hat{\psi}_i$'s can be obtained by estimating N time series separately by OLS.

This test is appropriate in our case, since it should be used when T is large relative to N.

For 185 and 2822 degree of freedoms, the F-value that leaves exactly 0.01 of the area under the F curve in the right tail of the distribution is approximately 1.32.[11] Because our test has a value of 5.1852, we are able to reject the null of homogenous slope and intercept parameters.

---

[10] Swamy propsed $F = \frac{1}{(N-1)} \sum_{i=1}^{N} F_i \sim F\left(K^*(N-1), N(T-K^*)\right)$.

[11] The 1% significance level has been arbitraly chosen.

6.3.2 Test for Heteroskedastic Variances

Once rejected the hypothesis of homogeneity of the coefficients across countries, we can test whether they have heteroskedastic variances or constant variances across units. One way to proceed is to use the Likelihood Ratio Test defined as

$$LR = 2\left[logL(\hat{\theta}) - logL\left(\hat{\theta}_r\right)\right] \tag{52}$$

where $\hat{\theta}$ is the unrestricted MLE, obtained estimating the model by EM-REML algorithm under the assumption of heteroskedastic variances, i.e. $\gamma_{is} \sim IN(0, \triangle_{is})$. On the other hand, $\hat{\theta}_r$ is the restricted MLE obtained from the EM-REML estimation under the assumption of homoskedastic variances, i.e. $\gamma_{is} \sim IN(0, \triangle_s)$, $\forall i$. When $\triangle_i = \triangle$ for all i, the iterations illustrated in Section 4.5 still hold. Equation (45) has to be replaced by

$$\triangle^{(b)} = \frac{1}{N}\sum_{i=1}^{N} V_{\gamma_i}^{(b)} + \frac{1}{N}\sum_{i=1}^{N} \hat{\gamma}_i^{(b)}\hat{\gamma}_i^{(b)\prime} \tag{53}$$

For a 0.01 level test and with $(N-1)\cdot(p+1)$ restrictions (in our case $N = 38$ and the number of lags is $p = 1$), the critical value for a Chi-squared distribution is less than 112.33. Given that our LR test has a value of 151.21, we reject the null of homoskedasticity at the 1% level. at the 1% level.

6.4 Comparison

We now compare the results obtained estimating (50) by EM-REML versus Swamy (1970) and the Mean Group method. In particular, the average effects (and their T-test between parentheses) are shown in the following table:[12]

The second column of the table reports the results assuming that the random coefficients have homoskedastic variances (even though the null hypothesis of homoskedasticity has been rejected).[13]

As expected economic growth suggests that a country can "easily" services its existing debt burden over time and therefore has a negative and significant impact on spreads at the 1% confidence level. The impact is larger when using the Mean Group estimation and the Swamy estimator, although in the latter case statistical significance only holds at the 10% level.

[12] It is known that the main drawback of the EM algorithm is its slow rate of convergence. However, in this particular application the rate of convergence is pretty fast, about 120 seconds.

[13] After expressing the coefficients as function of the explanatory variables, the LR test has a value of 107.1300 . We reject the null of homoskedasticity at the 2% level.

Table 1: Determinants of sovereign risk: EM, Swamy and Mean Group Estimates.

	EM-REML	**EM-REML**^{2}	**Swamy**	**MG**
Constant	0.002	0.006	0.066	0.089
	(0.253)	(0.805)	(0.944)	(1.342)
RGDP growth	**-0.016*****	**-0.016*****	**-0.044***	**-0.064*****
	(-2.750)	(-2.957)	(-1.747)	(-2.795)
Inflation	**0.019****	0.010	-0.010	-0.012
	(2.013)	(1.098)	(-0.276)	(-0.346)
Debt/GDP	-0.002	-0.003	0.016	0.025
	(-0.587)	(-1.020)	(0.727)	(1.223)
Lag Dep V.	**0.068*****	**0.090****	0.055	0.040
	(2.950)	(2.520)	(1.462)	(1.208)

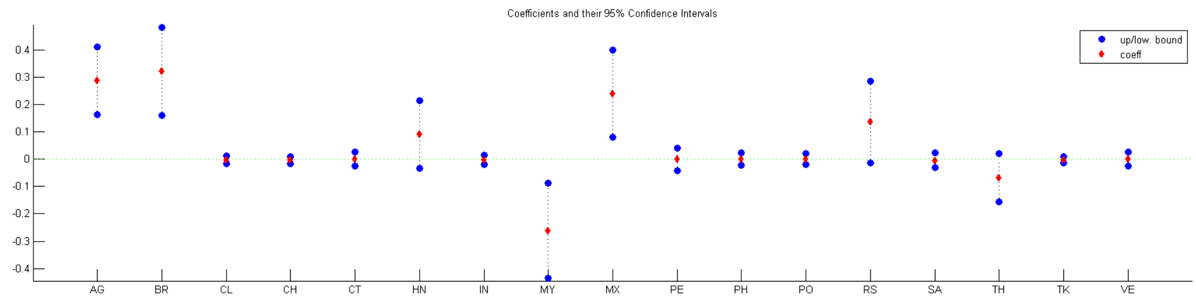
T-test between parentheses. The second column reports results when ignoring heteroskedasticity. Simbols ***, **, and * denotes significance at 1%, 5% and 10% respectively.

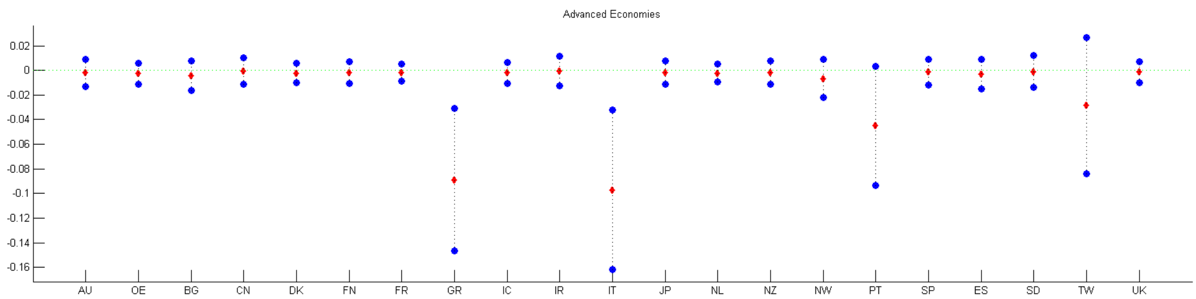
Only when accounting for heteroskedasticity in the random coefficients residuals using the EM-REML approach, spreads are found to be positively correlated with inflation rate. Indeed, high growth rates of inflation may reflect the inability of a government to finance its current budgetary expenses through taxes or further debt issuance. Moreover, the EM-REML estimation gives more predictive power to the autoregressive components compared to the other models. The coefficients on debt-to-GDP are not significant in all the four cases. This is in contrast with the literature on the determinants of sovereign spreads which find a significant positive correlation between spreads and debt. This difference can be explained by the fact that (i) we consider quarterly data rather than annual, (ii) we study both developed and emerging economies rather than just the latter, (iii) our model includes lagged values of the dependent variable and finally (iv) our estimation accounts for heterogeneity rather than pooling. The implications of neglected heterogeneity and dynamics are studied in Haque et al. (2000). Focusing on cross-country savings regressions, the authors find that ignoring differences across countries can lead to overestimating the influence of certain factors on the private savings rate. At the same time, one can obtain highly significant, but spurious, nonlinear effects for some of the potential determinants.

6.5 EM-REML Estimation and Shrinkage.

As shown in Section 4.2, the unobserved idiosyncratic components of the random coefficients, γ_i, are estimated by Best Linear Unbiased Prediction. This choice arises naturally in the EM algorithm and has the advantage over estimating N time series separately because BLUPs are shrinkage estimators. Indeed, they tend to be closer to zero than the estimated effects would be if they were computed by treating a random coefficient as if it were fixed. For instance, Maddala et al. (1997), estimating short-run and long-run elasticities of residential demand for electricity and natural gas, find that individual heterogeneous state estimates are difficult to interpret and have the wrong signs. Therefore, they suggest shrinkage estimators (instead of heterogeneous or homogeneous parameter estimates) if one is interested in obtaining elasticity estimates for each state since these give more reliable results and are superior for prediction purposes.

Focusing on the relationship between debt and spreads, the individual coefficients $\hat{\psi}_{ik} = \hat{\psi}_k + \hat{\gamma}_{ik}$ and their 95% confidence bands are shown below.





The sensitivity of the spread with respect to debt-to-GDP ratio is statistically significant only for a handful of countries, among which Argentina, Brazil and Mexico. The coefficients for Hungary and Russia are also positive but not significant. Surprisingly, Malaysia, Greece and Italy show a negative and significant correlation between the first-difference of spread and debt.[14] One could argue, that the latter two countries have benefited from joining the

[14]This is not the case when spreads and debt are not in first-difference. The estimated coefficients get close to zero and not statistically significant

eurozone. By doing so, their government and public sector agencies were allowed to increase the external obligations at rates which were lower than those they would have paid as single unit.[15]

It is interesting to compare the results described above with those obtained following Lee and Griffiths (1979), while using Swamy (1970) estimator of the random coefficients covariance. In that case the BLU predictions are given by

$$\tilde{\psi}_i = \hat{\psi}_{GLS} + \hat{\triangle} Z_i'(Z_i \hat{\triangle} Z_i' + \hat{\sigma}_i^2 I_T)^{-1}(y_i - Z_i \hat{\psi}_{GLS}) \tag{54}$$

where $\hat{\triangle}$ has been defined in (39), $\hat{\psi}_i$ are the individual OLS estimates and $\hat{\psi}_{GLS}$ is the Swamy's GLS estimator. The variance-covariance matrix of $\tilde{\psi}_i$ is given by

$$var\left(\tilde{\psi}_i\right) = var\left(\hat{\psi}_{GLS}\right) + (I - A_i)\left[\hat{\sigma}_{ii}\left(Z_i'Z_i\right)^{-1} - var\left(\hat{\psi}_{GLS}\right)\right](I - A_i)' \tag{55}$$

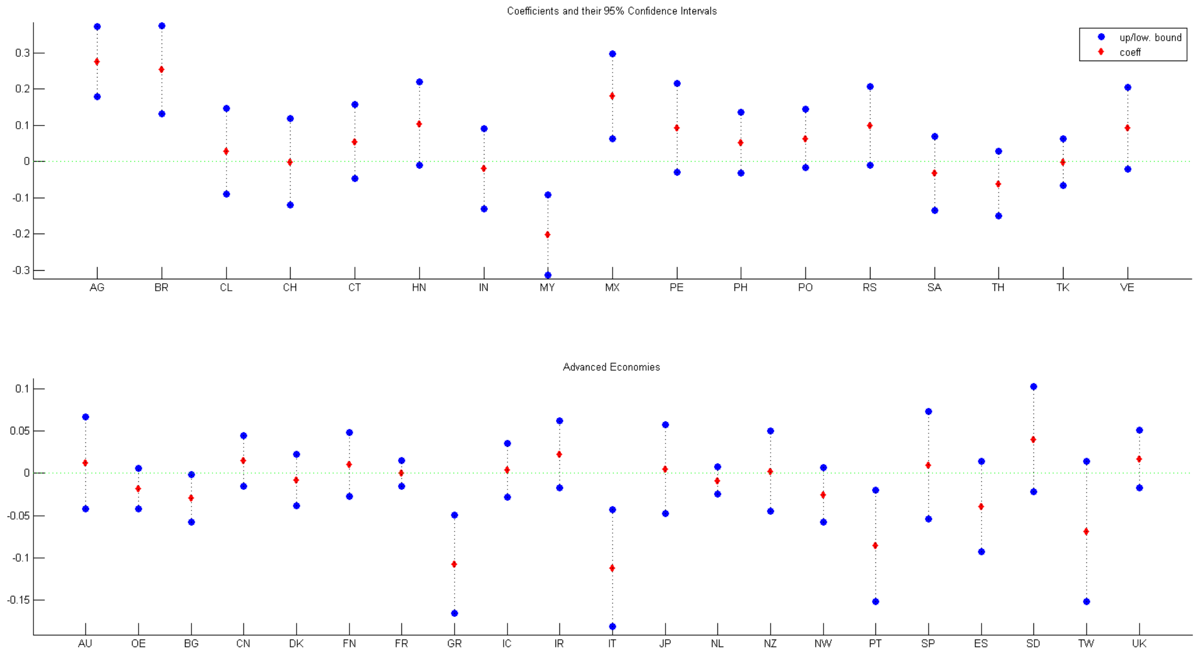
where

$$A_i = \left(\hat{\triangle}^{-1} + \hat{\sigma}_i^{-2} Z_i' Z_i\right)^{-1} \hat{\triangle}^{-1}$$

In this particular application $\hat{\triangle}$ is not positive-definite, therefore we use the asymptotic estimator given by (39).

As before, we focus on the relationship between spreads and debt:

[15] One could test this hypothesis by allowing for time-varying coefficients. We leave open the question for future research.



In this application, individual estimates obtained using Lee and Griffiths' approach are close to the OLS estimates of the individual coefficients. This can be disadvantageous in those cases where as argued in Maddala et al. (1997) estimating N heterogeneous time series yield inaccurate estimates and even wrong signs for the coefficients.

Comparing the two figures, it emerges that estimating the parameters of the model by EM-REML results into random coefficients residuals estimates which are shrunk to zero in a more effective way. The advantage of this approach is that it exploits the joint likelihood of the observed and unobserved data and allows for heteroskedasticity of the random coefficients residuals.

6.6 The Sensitivity of Spreads to Debt

We now explore why the sensitivity of spreads to debt differs significantly across countries by modelling the latter as a function of selected explanatory variables. We ask which factors influence financial markets decision when evaluating the credit worthiness of the borrower and setting interest rate during episodes of government debt growth.

First, using Reinhart and Rogoff (2011) historical time series on country's creditworthiness and financial turmoil, we model the random coefficients as function of a common constant and the percentage of years (between 1980 and 2010) in default or restructuring domestic and external debt. Results are shown below.

Table 2: Determinants of sensitivity of spreads: EM-REML Estimates.

	const	**% y Dom Def**	**% y Ext Def**
c_i	0.000	0.151	-0.007
	(0.023)	(0.228)	(-0.043)
$\beta_i^{(gdp)}$	**-0.014****	-0.550	-0.088
	(-2.375)	(-1.431)	(-0.888)
$\beta_i^{(cpi)}$	**0.021****	-0.383	0.067
	(2.039)	(-1.016)	(0.549)
$\beta_i^{(debt)}$	-0.003	**0.520*****	**0.158****
	(-0.981)	(2.986)	(2.217)
ϕ_i	**0.091*****	-0.574	0.025
	(3.403)	(-1.568)	(0.152)

T-test between parentheses. Simbols ***, **, and * denotes significance at 1%, 5% and 10% respectively.

A 1% increase in the percentage of year in default or restructuring domestic debt is associated with a 0.52% increase in the sensitivity of spreads to debt. History of repayment plays an important role. "Bad" reputation leads to high sensitivity of spreads to debt. As a consequence, relatively small increase in debt-to-GDP may lead to unsustainable interest rates which cannot be tolerated.

The above analysis is robust when augmenting the regression with additional explanatory variables. In particular, we consider the percentage of years (from 1980 to 2010) where a country faces an annual inflation rate of 20 percent or higher and the percentage of years (1980-2010) in which an annual depreciation versus the US dollar (or another relevant anchor currency) of 15 percent or more occurs.[16] We also includes measures of macroeconomic fundamentals such as the average (and standard deviation of) real GDP growth, rate of currency depreciation, inflation rate and Current Account to GDP growth. The average (first difference of) general gross government debt to GDP ratio and its standard deviation are used as a measure of sudden increases in debt's level. In Table **??**, we focus on the coefficients equation corresponding to the sensitivity of spreads to debt and report results from using different specifications.[17] Standard deviations over the sample period under considerations

[16]See Reinhart and Rogoff (2009) for more details.

[17]Other factors such as political instability and the composition of debt are currently being tested.

Table 3: Determinants of sensitivity of spreads to government debt: EM-REML Estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	-0.003	-0.004	0.008	0.003	0.001	-0.005	-0.019
	(-0.981)	(-0.795)	(0.505)	(0.225)	(0.038)	(-0.263)	(-1.029)
% y Curr Crisis		0.011					
		(0.128)					
% y Infl Crisis		-0.02					
		(-0.235)					
% y Dom Def	**0.52**	**0.646**	**0.653**	**0.478**	**0.577**	**0.541**	
	(2.986)	(3.151)	(3.167)	(2.452)	(2.841)	(2.615)	
% y Ext Def	**0.158**	**0.168**	**0.161**	**0.182**	**0.19**	**0.196**	
	(2.217)	(2.135)	(2.189)	(2.63)	(2.743)	(2.828)	
Volatility FX			-0.002	-0.004	-0.004	-0.002	0.002
			(-0.736)	(-1.052)	(-0.828)	(-0.464)	(0.301)
Volatility Debt/GDP				0.007	0.006	0.005	0.004
				(1.085)	(0.86)	(0.708)	(0.605)
Volatility Infl					-0.009	-0.013	-0.005
					(-0.901)	(-1.245)	(-0.438)
Volatility RGDP					0.007	0.018	0.014
					(0.738)	(1.296)	(0.998)
Volatility CA/GDP						-0.005	-0.006
						(-0.982)	(-1.096)

T-test between parentheses.

are used as measure of volatility. Including averages rather than volatility leads to very similar conclusions. Therefore, we do not report them.

At least three conclusions can be drawn. First, a «good» reputation in financial markets is essential. The percentage of years in defaults or restructuring have a statistically and economically significant effect on the sensitivity of spreads across all the different specifications. Domestic defaults have a larger economic impact than external ones. Second, country-specific macroeconomic indicators do not play any significant role in explaining the reactions of investors to an increase in debt. This suggests that markets decisions during episodes of debt growth may be driven by sentiments (as defined by Eichengreen and Mody, 2000) rather than fundamentals. At the same time, we have seen that this «irrational exuberance» or «excessive» reaction is usually associated with countries with a weak history of repayment. Finally, contrary to the literature which emphasizes the role of volatility of macroeconomic aggregates in explaining sovereign credit risks[18], we find no evidence that such variables affect

[18] See for example, Eaton and Gersovitz (1981), Catao and Kapur (2006) and Hilscher and Nosbuch (2010).

markets when calculating the additional risk premium to charge in response to an increase in debt.

To conclude, while it is common in the literature to find that certain macroeconomic fundamentals are significant predictors of sovereign spreads, we show that they are not significant determinants of the sensitivity of spreads to changes in sovereign debt.

7 Conclusion

We propose estimating dynamic panel model with random coefficients by combining the EM algorithm, popularised by Dempster et al. (1977) with Restricted Maximum Likelihood (REML) approach, developed by Patterson and Thompson (1971). This approach leads to tractable closed form solutions. The unobserved error terms of the random coefficients are estimated by BLUP and their probability distributions are also derived. The main advantage of the EM-REML algorithm is that it yields unbiased estimators of the variance components that are positive definite. As a consequence, as shown in the Monte Carlo analysis, this approach has good properties even in small samples. Second, the proposed method allows for heteroskedastic random coefficients and thus can be seen as a generalization of the one-way error components model where both the variances of the random effects and the regression disturbances have heteroskedasticity of unknown form. Ignoring heteroskedasticity when it is present will still result in consistent estimates of the regression coefficients. Nevertheless, these estimates will not be efficient and their standard errors will be biased therefore affecting the validity of hypothesis testing. An empirical application is also presented. We investigates what causes the sensitivity of spreads to differ significantly across countries by modelling the latter as a function of selected explanatory variables. We ask which factors influence financial markets decision when evaluating the credit worthiness of the borrower and setting the risk premium during episodes of government debt growth. We find that while country-specific macroeconomic indicators do not play any significant role in explaining the sensitivity of spreads to an increase in debt, history of repayment is crucial. "Bad" reputation leads to higher sensitivity of spreads to debt. As a consequence, countries who have defaulted in the past may find it difficult to finance current expenditures by issuing new debt since relatively small increase in debt-to-GDP may lead to a raise in interest rates which may be difficult to tolerate.

A Appendix

A.1 Matrix Computations for REML

A.1.1 A Choice for S_i

The Projection Matrix M_i. One plausible choice for such an S_i, is

$$M_i = I - W_i \left(W_i' W_i\right)^{-1} W_i' \tag{56}$$

Indeed, M_i is of rank $(T-p) - \underline{K}^*$, with $\underline{K}^* \leq K^* l < T - p$, and it satisfies $M_i W_i = 0$. M_i is symmetric and idempotent.

As noted by Searle (1978), its canonical form under orthogonal similarity is given by

$$U M_i U' = \begin{bmatrix} I_{T-p-K^*l} & O \\ O & O \end{bmatrix}$$

where U is an orthogonal matrix. Searle (1978) defines A to be the first $T-p-K^*l$ columns of U'. It follows that $M_i = AA'$ and $A'A = I$. Premultiplying M_i by A, we get $M_i A = AA'A$ and since $A'A = I$, we have

$$M_i A = A, \qquad A' M_i = A' \tag{57}$$

Since U' is orthogonal and non-singular, A' has full rank and $A' W_i = 0$. As stated in Searle (1978), using (57), it can be shown that $A\left(A' R_i A\right)^{-1} A'$ is the Moore-Penrose inverse of $M_i R_i M_i$:

$$\left(M_i R_i M_i\right)^{+} = A\left(A' R_i A\right)^{-1} A' \tag{58}$$

Since A' has full row rank and R_i is positive definite, the inverse of $A' R_i A$ exists.

A generalization of M_i. As shown in Lemma 2.1 of Searle (1978), any linear combination of M_i, $S_i = J M_i$, satisfies $S_i W_i = 0$.

A generalization of M_i is

$$P_i = R_i^{-1} - R_i^{-1} W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1} \tag{59}$$

satisfying $P_i W_i = 0$. From the definition of P_i, it follows that

$$\begin{aligned} R_i P_i &= I - W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1} \\ P_i R_i &= I - R_i^{-1} W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' \end{aligned} \tag{60}$$

Therefore,

$$P_i R_i P_i = P_i \tag{61}$$

and also $(P_i R_i)^2 = P_i R_i$. It follows that $tr(P_i R_i) = r(P_i R_i) = r(P_i) = T - p - K^* l$.

Relationship between M_i and P_i. Using (56) and the fact that $P_i W_i = 0$, it can be seen that

$$P_i M_i = P_i = M_i P_i \tag{62}$$

Furthermore,post-multiplying (60) by M_i and using $M_i W_i = 0$ and $W_i' M_i = 0$, we get $P_i R_i M_i = M_i$. Post-multiplying (62) by $R_i M_i$

$$P_i M_i R_i M_i = P_i R_i M_i = M_i P_i R_i M_i = M_i^2 = M_i \tag{63}$$

From (62) and (63), we can establish P_i as the Moore-Penrose inverse of $M_i R_i M_i$:

$$P = (M_i R_i M_i)^+ \tag{64}$$

Since $(M_i R_i M_i)^+$ is unique, (58) and (64) imply that

$$P_i = (M_i R_i M_i)^+ = A\left(A' R_i A\right)^{-1} A' \tag{65}$$

A.1.2 Some Lemmas from Searle (1978)

Lemma 1. It can be shown that

$$S_i = F' A' \tag{66}$$

for some non-singular F'. Indeed, we have seen that $S_i = J M_i$ where $M_i = AA'$. Therefore, $S_i = JAA'$. Hence, let $F' = JA$, from which it follows that $S_i = F'A'$. Moreover, F' is a square matrix of order $T - p - K^* l$, where $T - p - K^* l = r(S_i) \leq r(F')$. This implies that $r(F') = T - p - K^* l$. Thus, $S_i = F'A'$ is true.

Lemma 2. Using Lemma 1 and (65), it follows that

$$\begin{aligned} S_i' \left(S_i R_i S_i'\right)^{-1} S_i &= AF\left(F' A' R_i A F\right)^{-1} F' A' \\ &= A\left(A' R_i A\right)^{-1} A' = P_i \end{aligned} \tag{67}$$

Lemma 3. Recall that if $A = (m \times m)$, $B = (m \times n)$,$C = (n \times m)$ and D is a $(n \times n)$ matrix, then

$$\begin{aligned} det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \mid D \mid \cdot \mid A - BD^{-1}C \mid \quad if \quad D\, nonsingular \\ &= \mid A \mid \cdot \mid D - CA^{-1}B \mid \quad if \quad A\, nonsingular \end{aligned} \tag{68}$$

Using this property of the determinant, we can show that

$$| AR_iA' |= \frac{| R_i | \cdot | W_i'R_i^{-1}W_i |}{| W_i'W_i |} \tag{69}$$

To prove the latter, let

$$\begin{bmatrix} A' \\ W_i' \end{bmatrix} R_i \begin{bmatrix} A & W_i \end{bmatrix} = \begin{bmatrix} A'R_iA & A'R_iW_i \\ W_i'R_iA & W_i'R_iW_i \end{bmatrix}$$

Taking the determinant of both sides, we get

$$| R_i | \cdot \begin{vmatrix} A'A & A'W_i \\ W_i'A & W_i'W_i \end{vmatrix} =| A'R_iA | \cdot | W_i'R_iW_i - W_i'R_iA \left(A'R_iA\right)^{-1} A'R_iW_i |$$

Using $A'A = I$ and $A'W_i = 0$ and equation (65), we get

$$| R_i || W_i'W_i |=| A'R_iA | \cdot | W_i'R_iW_i - W_i'R_iPR_iW_i |$$

Substituting (60) into the latter equation and then using the property of determinants, $det(AB) = det(A) \cdot det(B)$, yields (69).

Lemma 4. Using the definition $S_i = F'A'$, it can be shown that

$$| S_iR_iS_i' |=| F |^2| A'R_iA | \tag{70}$$

A.1.3 Finding an expression for L_{1i}

Taking the logarithm of both sides of $f(S_iy_i \mid Z_i, \gamma_i)$, yields

$$L_{1i} = c_3 - \frac{1}{2}log \mid S_iR_iS_i' \mid -\frac{1}{2} \left(y_i - Z_i\gamma_i\right)' S_i' \left(S_iR_iS_i'\right)^{-1} S_i \left(y_i - Z_i\gamma_i\right) \tag{71}$$

Using (69) and (70), we have

$$log \mid S_iR_iS_i' |= c + log|R_i| + log \mid W_i'R_i^{-1}W_i| \tag{72}$$

where c includes the terms that do not involve the parameters of interest.

Furthermore, using Lemma 2, we get

$$\begin{aligned} \left(y_i - Z_i\gamma_i\right)' S_i' \left(S_iR_iS_i'\right)^{-1} S_i \left(y_i - Z_i\gamma_i\right) &= \left(y_i - Z_i\gamma_i\right)' P_i \left(y_i - Z_i\gamma_i\right) \\ &= \left(y_i - W_i\hat{\hat{\Gamma}} - Z_i\gamma_i\right)' R_i^{-1} \left(y_i - W_i\hat{\hat{\Gamma}} - Z_i\gamma_i\right) \end{aligned} \tag{73}$$

Substituting (72) and (73) into (71) yields (20).

Proof of Equation (73). Let[19] $\hat{\bar{\Gamma}}$ be the argument that minimizes $\varepsilon_i' R_i^{-1} \varepsilon_i$, where $\varepsilon_i = y_i - W_i \bar{\Gamma} - Z_i \gamma_i$ and $R_i = var(\varepsilon_i)$. The solution to the problem is given by

$$\hat{\bar{\Gamma}} = \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1} (y_i - Z_i \gamma_i)$$

We can now show that

$$\begin{aligned} y_i - W_i \hat{\bar{\Gamma}} - Z_i \gamma_i &= y_i - W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1} (y_i - Z_i \gamma_i) - Z_i \gamma_i \\ &= R_i P_i y_i - R_i P_i Z_i \gamma_i \end{aligned}$$

Therefore, using (61) and after a few computations, we get

$$\begin{aligned} \left(y_i - W_i \hat{\bar{\Gamma}} - Z_i \gamma_i\right)' R_i^{-1} \left(y_i - W_i \hat{\bar{\Gamma}} - Z_i \gamma_i\right) &= (y_i' P_i R_i - \gamma_i' Z_i' P_i R_i) R_i^{-1} (R_i P_i y_i - R_i P_i Z_i \gamma_i) \\ &= y_i' P_i y_i - y_i' P_i Z_i \gamma_i - \gamma_i' Z_i' P_i y_i + \gamma_i' Z_i' P_i Z_i \gamma_i \\ &= (y_i - Z_i \gamma_i)' P_i (y_i - Z_i \gamma_i) \end{aligned}$$

A.1.4 Finding an expression for L_{2i}.

The Choice of Q_i. It can be shown that $Q_i = W_i' R_i^{-1}$ is a plausible choice. We first compute the covariance conditional on γ_i, to then show that the unconditional covariance is equal to zero, i.e. $cov(S_i y_i, Q_i y_i) = 0$.

$$\begin{aligned} cov(S_i y_i, Q_i y_i \mid \gamma_i) &= E(S_i y_i y_i' Q_i' \mid \gamma_i) - E(S_i y_i \mid \gamma_i) E(y_i' Q_i' \mid \gamma_i) \\ &= S_i E(y_i y_i' \mid \gamma_i) Q_i' - (S_i Z_i \gamma_i) \left(\bar{\Gamma}' W_i' + \gamma_i' Z_i'\right) R_i^{-1} W_i \end{aligned} \tag{74}$$

where $E(S_i y_i \mid \gamma_i) = S_i Z_i \gamma_i$ since $S_i W_i = 0$.
Substituting

$$S_i E(y_i y_i' \mid \gamma_i) Q_i' = S_i var(\varepsilon_i) Q_i' = S_i R_i R_i^{-1} W_i = S_i W_i = 0$$

and

$$\begin{aligned} (S_i Z_i \gamma_i) \left(\bar{\Gamma}' W_i' + \gamma_i' Z_i'\right) R_i^{-1} W_i &= S_i Z_i \gamma_i \bar{\Gamma}' W_i' R_i^{-1} W_i \\ &\quad + S_i Z_i \gamma_i \gamma_i' Z_i' R_i^{-1} W_i \end{aligned}$$

into (74), we get

$$\begin{aligned} cov(S_i y_i, Q_i y_i \mid \gamma_i) &= -S_i Z_i \gamma_i \bar{\Gamma}' W_i' R_i^{-1} W_i \\ &\quad - S_i Z_i \gamma_i \gamma_i' Z_i' R_i^{-1} W_i \end{aligned} \tag{75}$$

[19] To make notation easier we focus on $\varepsilon_i' R_i^{-1} \varepsilon_i$ instead of $\sum_{i=1}^{N} \varepsilon_i' R_i^{-1} \varepsilon_i$.

Using the Law of Total Covariance[20], the unconditional covariance can be obtained from

$$\begin{array}{rcc} cov\left(S_i y_i, Q_i y_i\right) & = & E\left[cov\left(S_i y_i, Q_i y_i \mid \gamma_i\right)\right] \\ & & +cov\left(E\left(S_i y_i \mid \gamma_i\right), E\left(Q_i y_i \mid \gamma_i\right)\right) \end{array} \tag{76}$$

Taking expectation of both sides of (75), we get

$$E\left[cov\left(S_i y_i, Q_i y_i \mid \gamma_i\right)\right] = -S_i Z_i \triangle_i Z_i' R_i^{-1} W_i \tag{77}$$

since $\gamma_i \sim N(0, \triangle_i)$. Moreover,

$$\begin{array}{rcc} cov\left(E\left(S_i y_i \mid \gamma_i\right), E\left(Q_i y_i \mid \gamma_i\right)\right) & = & E\left[S_i Z_i \gamma_i \left(W_i' R_i^{-1} W_i \bar{\Gamma} + W_i' R_i^{-1} Z_i \gamma_i\right)'\right] \\ & & -E\left[E\left(S_i y_i\right)\right] E\left[E\left(Q_i y_i\right)'\right] \\ & = & S_i Z_i \triangle_i Z_i' R_i^{-1} W_i \end{array} \tag{78}$$

Therefore, substituting (77) and (78) into (76) we can show that $cov\left(S_i y_i, Q_i y_i\right) = 0$.

A.2 BLUP

Assume that y_i and γ_i are jointly normally distributed.

The conditional expectation of γ_i given the data is

$$\begin{array}{rcc} \hat{\gamma}_i = E\left(\gamma_i \mid y_i\right) & = & E(\gamma_i) + cov(\gamma_i, y_i)\left[var(y_i)\right]^{-1}\left[y_i - E(y_i)\right] \\ & = & c' V_y^{-1}\left(y_i - W_i \bar{\Gamma}\right) \end{array} \tag{79}$$

where by assumption $E(\gamma_i) = 0$.

The conditional variance of γ_i is

$$\begin{array}{rcc} var\left(\gamma_i \mid y_i\right) & = & var(\gamma_i) - cov(\gamma_i, y_i)\left[var(y_i)\right]^{-1} \cdot cov(y_i, \gamma_i) \\ & = & var(\gamma_i) - c' V_y^{-1} c \end{array}$$

Henderson (1984, Chap. 5), showed that

1. the BLP is unbiased:

$$\begin{array}{rcl} E\left(\hat{\gamma}_i\right) & = & E\left[c' V_y^{-1}(y_i - W_i \bar{\Gamma}\right] \\ & = & c' V_y^{-1}\left[E(y_i) - W_i \bar{\Gamma}\right] \quad = \quad 0 = E(\gamma_i) \end{array} \tag{80}$$

since $E(y_i) = W_i \bar{\Gamma}$.

[20] If X, Y and Z are random variables, then $cov\left(X, Y\right) = E\left[cov\left(Y, X \mid Z\right)\right] + cov\left(E\left(X \mid Z\right), E\left(Y \mid Z\right)\right)$

2. The variance of $\hat{\gamma}_i$ is $var\left(\hat{\gamma}_i\right) = var\left[c'V_y^{-1}(y_i - W_i\bar{\Gamma})\right] = c'V_y^{-1}c$.

3. The covariance between $\hat{\gamma}_i$ and γ_i is $c'V_y^{-1}cov\left(\tilde{\gamma}_i, \gamma_i\right) = var\left(\hat{\gamma}_i\right)$, from which it follows that $var\left(\hat{\gamma}_i - \gamma_i\right) = var\left(\hat{\gamma}_i\right) - var\left(\gamma_i\right)$.

4. Finally, the BLP maximizes the correlation between $\hat{\gamma}_i$ and γ_i.

If $E(y_i) = W_i\bar{\Gamma}$ is not known (since $\bar{\Gamma}$ is not known), we need to consider the best linear unbiased prediction (BLUP). However, in the EM-algorithm we substitute $\bar{\Gamma}$ by its guess $\bar{\Gamma}^{(0)}$. As stated in Henderson (1984), if $E(\gamma_i) = 0$, the BLUP is BLP with $\bar{\Gamma}^{(0)}$ substituted for $\bar{\Gamma}$. Furthermore, under normality, BLUP has the same properties as BLP.

A.2.1 Computations

From (6), we know that

$$E(y_i) = W_i\bar{\Gamma} \tag{81}$$

$$V_y = var(y_i) = var\left(W_i\bar{\Gamma} + Z_i\gamma_i + \varepsilon_i\right) = Z_i\triangle_i Z_i' + R_i \tag{82}$$

$$c' = cov(\gamma_i, y_i) = E\left(\gamma_i y_i'\right) = \triangle_i Z_i' \tag{83}$$

If Z_i includes lagged dependent variables among the regressors, then the above results hold only if we assume that the first p observations $(y_1, .., y_p)$ are deterministic and therefore are used as presample.

Substituting (82) and (83) into (79) yields

$$\hat{\gamma}_i = E\left(\gamma_i \mid y_i\right) = \triangle_i Z_i'\left(Z_i\triangle_i Z_i' + R_i\right)^{-1}\left(y_i - W_i\bar{\Gamma}\right)$$

As suggested in Pawitan (2001) and Henderson (1963), using a simple matrix identity we can write

$$\begin{aligned} c'V_y^{-1} = \triangle_i Z_i'\left[Z_i\triangle_i Z_i' + R_i\right]^{-1} &= \left\{\left(Z_i'R_i^{-1}Z_i + \triangle_i^{-1}\right)^{-1}\left(Z_i'R_i^{-1}Z_i + \triangle_i^{-1}\right)\right\}\cdot \\ &\quad \cdot\triangle_i Z_i'\left[Z_i\triangle_i Z_i' + R_i\right]^{-1} \\ &= \left(Z_i'R_i^{-1}Z_i + \triangle_i^{-1}\right)^{-1}\cdot Z_i'R_i^{-1} \end{aligned} \tag{84}$$

Using these results, the conditional expectation of γ_i given the data is

$$E\left(\gamma_i \mid y_i\right) = \left(Z_i'R_i^{-1}Z_i + \triangle_i^{-1}\right)^{-1}\cdot Z_i'R_i^{-1}\left(y_i - W_i\bar{\Gamma}\right) \tag{85}$$

The conditional variance is given by

$$
\begin{aligned}
var\,(\gamma_i \mid y_i) &= var(\gamma_i) - cov(\gamma_i, y_i)\,[var(y_i)]^{-1} \cdot cov(y_i, \gamma_i) \\
&= \triangle_i - \triangle_i Z_i' \left[Z_i \triangle_i Z_i' + R_i\right]^{-1} Z_i \triangle_i
\end{aligned} \tag{86}
$$

Using result (84) the previous formula can be simplified as follows:

$$
\begin{aligned}
var\,(\gamma_i \mid y_i) &= \triangle_i - \triangle_i Z_i' \left[Z_i \triangle_i Z_i' + R_i\right]^{-1} Z_i \triangle_i \\
&= \left(Z_i' R_i^{-1} Z_i + \triangle_i^{-1}\right)^{-1}
\end{aligned}
$$

A.2.2 BLUP as maximizer of Complete Data likelihood

$$
\begin{aligned}
logL(y, \gamma; \theta) &= logf(y \mid \gamma; \theta_1) + logf(\gamma; \omega) \\
&= +\tfrac{1}{2} \textstyle\sum_{i=1}^{N} log \mid \triangle_i^{-1} \mid - \tfrac{1}{2} \sum_{i=1}^{N} \gamma_i' \triangle_i^{-1} \gamma_i \\
&\quad - \tfrac{1}{2} \textstyle\sum_{i=1}^{N} log \mid R_i \mid - \tfrac{1}{2} \sum_{i=1}^{N} \varepsilon_i' R_i^{-1} \varepsilon_i
\end{aligned}
$$

where $\varepsilon_i = y_i - W_i \bar{\Gamma} - Z_i \gamma_i$. Taking the first derivative with respect to γ_i and equating to zero yields

$$
\left(Z_i' R_i^{-1} Z_i + \triangle_i^{-1}\right) \gamma_i = Z_i' R_i^{-1} \left(y_i - W_i \bar{\Gamma}\right)
$$

The observed Fisher information can be obtained by taking the second derivative of the log-likelihood of the complete data with respect to γ_i:

$$
\frac{\partial logL}{\partial \gamma_i \partial \gamma_i'} = -\triangle_i^{-1} - Z_i' R_i^{-1} Z_i
$$

The observed Fisher information is equal to

$$
I\,(\gamma_i) = Z_i' R_i^{-1} Z_i + \triangle_i^{-1}
$$

A.3 E-Step

E-step for L_{2i}. As suggested in Pawitan (2001)

$$
E_{\theta^{(b-1)}}\left(\varepsilon_i' H_i \varepsilon_i \mid y_i\right) = Tr\left[H_i E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right)\right] \tag{87}
$$

To find $E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right)$, recall that for a random variable X, $var(X) = E(XX') - E(X)E(X')$ from which it follows $E(XX') = V + \mu\mu'$. It is clear now that

$$
E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right) = V_{\varepsilon_i} + \hat{\varepsilon}_i \hat{\varepsilon}_i' \tag{88}
$$

where

$$
\begin{aligned}
\hat{\varepsilon}_i = E_{\theta^{(b-1)}}\left(\varepsilon_i \mid y_i\right) &= E_{\theta^{(b-1)}}\left(y_i - W_i \bar{\Gamma} - Z_i \gamma_i \mid y_i\right) \\
&= y_i - W_i \bar{\Gamma} - Z_i \hat{\gamma}_i^{(b)}
\end{aligned}
$$

and

$$
\begin{aligned}
V_{\varepsilon_i} &= var\left(\varepsilon_i \mid y_i;\theta^{(b-1)}\right) &= var\left(y_i - W_i\bar{\Gamma} - Z_i\gamma_i \mid y_i, \theta^{(b-1)}\right) \\
& &= Z_i V_{\gamma_i}^{(b)} Z_i'
\end{aligned} \tag{89}
$$

with $\hat{\gamma}_i^{(b)} = E_{\theta^{(b-1)}}\left(\gamma_i \mid y_i\right)$ and $V_{\gamma_i}^{(b)} = var\left(\gamma_i \mid y_i, \theta^{(b-1)}\right)$.

Substituting (89) into (88) yields

$$
E_{\theta^{(b-1)}}\left(\varepsilon_i\varepsilon_i' \mid y_i\right) = Z_i V_{\gamma_i}^{(b)} Z_i' + \hat{\varepsilon}_i\hat{\varepsilon}_i'
$$

Substituting the latter into (87)

$$
E_{\theta^{(b-1)}}\left(\varepsilon_i' H_i \varepsilon_i \mid y_i\right) = Tr\left(H_i Z_i V_{\gamma_i}^{(b)} Z_i'\right) + Tr\left(H_i \hat{\varepsilon}_i\hat{\varepsilon}_i'\right) \tag{90}
$$

Furthermore, using the properties of trace, we have

$$
E_{\theta^{(b-1)}}\left(\varepsilon_i' H_i \varepsilon_i \mid y_i\right) = Tr\left(Z_i' H_i Z_i V_{\gamma_i}^{(b)}\right) + \hat{\varepsilon}_i' H_i \hat{\varepsilon}_i
$$

We can now write

$$
\begin{aligned}
Q_{2i} = E_{\theta^{(b-1)}}\left(L_{2i} \mid y_i\right) &= c_4 - \tfrac{1}{2} log \mid W_i' R_i^{-1} W_i \mid \\
& -\tfrac{1}{2} Tr\left(Z_i' H_i Z_i V_{\gamma_i}^{(b)}\right) - \tfrac{1}{2}\hat{\varepsilon}_i' H_i \hat{\varepsilon}_i
\end{aligned}
$$

A.4 Estimation of $\triangle_i$

An estimator of $\triangle_i$ can be obtained by maximizing (30) with respect to $\triangle_i$. Before proceeding, we report a few results of matrices differentiation shown in Lutkepohl (1996).

1. X $(m \times m)$ nonsingular, a, b $(m \times 1)$:[21]

$$
\frac{\partial a' X^{-1} b}{\partial X} = -(X^{-1})' a b' (X^{-1})' \tag{91}
$$

2. X $(m \times m)$ nonsingular, A, B $(m \times m)$:[22]

$$
\frac{\partial tr\left(A X^{-1} B\right)}{\partial X} = -\left(X^{-1} B A X^{-1}\right)' \tag{92}
$$

3. X $(m \times m)$, $det(X) > 0$:

$$
\frac{\partial ln \mid X \mid}{\partial X} = (X')^{-1} \tag{93}
$$

[21] Lutkepohl (1996, pag.177 eq. 10).
[22] Lutkepohl (1996, pag.179 eq. 23)

It follows that

$$\begin{array}{ccccccccc} \frac{\partial Q_{3i}}{\partial \triangle_i} & = & -\triangle_i^{-1} & + & \triangle_i^{-1} V_{\gamma_i}^{(b)} \triangle_i^{-1} & + & \triangle_i^{-1} \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} \triangle_i^{-1} & = 0 \\ & & (93) & & (92) & & (91) & \end{array}$$

which implies that

$$\triangle_i^{-1} = \triangle_i^{-1} V_{\gamma_i}^{(b)} \triangle_i^{-1} + \triangle_i^{-1} \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} \triangle_i^{-1}$$

Pre-multiplying and post-multiplying both sides by $\triangle_i$, we get

$$\triangle_i^{(b)} = V_{\gamma_i}^{(b)} + \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} \tag{94}$$

Unbiased Estimator. Using the Law of Total Variance, the unconditional variance of γ_i can be written as

$$\begin{array}{rcl} \triangle_i = var\left(\gamma_i\right) & = & var\left[E\left(\gamma_i \mid y_i\right)\right] + E\left[var\left(\gamma_i \mid y_i\right)\right] \\ & = & var\left(\hat{\gamma}_i\right) + E\left(V_{\gamma_i}\right) \end{array} \tag{95}$$

Therefore, it can be shown that

$$\hat{\triangle}_i = \hat{\gamma}_i \hat{\gamma}_i' + V_{\gamma_i} \tag{96}$$

is an unbiased estimator of $\triangle_i$. Indeed, taking expectation of both sides of (96)[23] and using (95)

$$E\left(\hat{\triangle}_i\right) = E\left(\hat{\gamma}_i \hat{\gamma}_i'\right) + E\left(V_{\gamma_i}\right) = var\left(\hat{\gamma}_i\right) + E\left(V_{\gamma_i}\right) = \triangle_i$$

Another way to prove the latter is shown below:

$$\begin{array}{rcl} E\left(\hat{\triangle}_i\right) & = & E\left(\hat{\gamma}_i \hat{\gamma}_i'\right) + E\left(V_{\gamma_i}\right) \\ & = & E\left\{c' V_{y_i}^{-1} \left(y_i - W_i \bar{\Gamma}\right) \left(y_i - W_i \bar{\Gamma}\right)' V_{y_i}^{-1} c\right\} + \triangle_i - c' V_{y_i}^{-1} c \\ & = & c' V_{y_i}^{-1} c + \triangle_i - c' V_{y_i}^{-1} c = \triangle_i \end{array}$$

A.5 Monte Carlo Analysis

[23]Note that

$$var\left(\tilde{\gamma}_i\right) = E\left(\tilde{\gamma}_i \tilde{\gamma}_i'\right) + E\left(\tilde{\gamma}_i\right) E\left(\tilde{\gamma}_i'\right) = E\left(\tilde{\gamma}_i \tilde{\gamma}_i'\right)$$

since $E\left(\tilde{\gamma}_i\right) = E_{y_i}\left(E\left(\gamma_i \mid y_i\right)\right) = E\left(\gamma_i\right) = 0$.

Table 4: EM-REML, Swamy and Mean Group Estimators Properties when $T = 10$

	EM-REML			EM-REML (homos)			Swamy			N-Time Series - MG		
T=10 / N	10	30	80	10	30	80	10	30	80	10	30	80
$Bias\,(c_i)$	**0.033**	**0.004**	**0.008**	**-0.005**	**-0.053**	**-0.068**	**0.104**	**0.119**	**0.119**	**0.140**	**0.174**	**0.191**
$se\,\{Bias\,(c_i)\}$	0.024	0.010	0.006	0.019	0.008	0.005	0.028	0.015	0.011	0.036	0.026	0.016
$Bias\,(\beta_i)$	**0.011**	**0.000**	**0.007**	**0.003**	**-0.020**	**-0.016**	**0.033**	**0.009**	**0.011**	**0.042**	**0.014**	**0.012**
$se\,\{Bias\,(\beta_i)\}$	0.016	0.008	0.004	0.012	0.006	0.003	0.014	0.009	0.005	0.017	0.012	0.007
$Bias\,(\phi_i)$	**-0.086**	**0.009**	**0.026**	**-0.019**	**0.127**	**0.153**	**-0.249**	**-0.218**	**-0.196**	**-0.330**	**-0.312**	**-0.290**
$se\,\{Bias\,(\phi_i)\}$	0.013	0.008	0.004	0.011	0.008	0.004	0.012	0.007	0.005	0.012	0.007	0.004
$\Vert\ Bias\,(\psi)\ \Vert$	**0.093**	**0.009**	**0.028**	**0.020**	**0.139**	**0.168**	**0.272**	**0.248**	**0.230**	**0.361**	**0.357**	**0.348**
$RMSE\,(c_i)$	0.240	0.098	0.064	0.193	0.096	0.083	0.293	0.187	0.158	0.381	0.309	0.249
$RMSE\,(\beta_i)$	0.155	0.081	0.043	0.120	0.066	0.035	0.142	0.087	0.052	0.177	0.119	0.074
$RMSE\,(\phi_i)$	0.154	0.081	0.051	0.109	0.150	0.158	0.276	0.228	0.202	0.351	0.319	0.293
$\Vert\ Bias\,(var\,(\gamma_1))\ \Vert$	0.247	0.234	0.303	0.098	0.476	0.633	3.967	9.375	14.976	5.716	16.276	19.694
$\Vert\ Bias\,(var\,(\gamma_2))\ \Vert$	0.252	0.196	0.274	0.056	0.273	0.367	0.998	1.619	2.514	1.222	1.754	3.166
$\Vert\ Bias\,(var\,(\gamma_3))\ \Vert$	0.096	0.184	0.334	0.013	0.057	0.070	0.378	0.644	1.034	0.491	0.794	1.260
$av\,(RMSE\,\{var\,(\gamma_1)\})$	0.342	0.174	0.169	0.033	0.088	0.071	1.723	1.997	1.825	2.226	3.662	3.070
$av\,(RMSE\,\{var\,(\gamma_2)\})$	0.342	0.173	0.169	0.021	0.051	0.041	0.352	0.316	0.302	0.483	0.396	0.434
$av\,(RMSE\,\{var\,(\gamma_3)\})$	0.125	0.140	0.144	0.013	0.016	0.010	0.134	0.124	0.122	0.189	0.173	0.175
% Neg. Def.							0.970	0.930	0.890			

Table 5: EM-REML, Swamy and Mean Group Estimators Properties when $T = 30$

	EM-REML			EM-REML_homos			Swamy			N_TS-MG		
T=30 / N	10	30	80	10	30	80	10	30	80	10	30	80
$Bias\,(c_i)$	**0.005**	**0.004**	**0.009**	**-0.006**	**-0.007**	**-0.006**	**0.035**	**0.024**	**0.015**	**0.069**	**0.056**	**0.047**
$se\,\{Bias\,(c_i)\}$	0.010	0.006	0.004	0.009	0.005	0.003	0.012	0.006	0.004	0.015	0.009	0.006
$Bias\,(\beta_i)$	**0.023**	**0.006**	**0.001**	**0.013**	**-0.001**	**-0.004**	**0.018**	**0.007**	**0.001**	**0.021**	**0.012**	**0.008**
$se\,\{Bias\,(\beta_i)\}$	0.009	0.005	0.003	0.008	0.005	0.002	0.008	0.005	0.002	0.008	0.005	0.003
$Bias\,(\phi_i)$	**-0.001**	**-0.008**	**-0.004**	**0.012**	**0.008**	**0.013**	**-0.049**	**-0.049**	**-0.030**	**-0.096**	**-0.101**	**-0.101**
$se\,\{Bias\,(\phi_i)\}$	0.007	0.004	0.002	0.006	0.003	0.002	0.006	0.003	0.003	0.006	0.003	0.002
$\Vert\ Bias\,(\psi)\ \Vert$	**0.023**	**0.011**	**0.010**	**0.018**	**0.011**	**0.015**	**0.063**	**0.055**	**0.034**	**0.120**	**0.116**	**0.112**
$RMSE\,(c_i)$	0.103	0.057	0.039	0.090	0.048	0.034	0.123	0.067	0.043	0.163	0.105	0.072
$RMSE\,(\beta_i)$	0.093	0.052	0.030	0.077	0.047	0.024	0.081	0.050	0.025	0.087	0.056	0.029
$RMSE\,(\phi_i)$	0.070	0.041	0.025	0.064	0.035	0.025	0.075	0.060	0.042	0.111	0.106	0.103
$\Vert\ Bias\,(var\,(\gamma_1))\ \Vert$	0.102	0.140	0.137	0.251	0.395	0.438	0.720	1.139	0.974	0.564	1.356	1.299
$\Vert\ Bias\,(var\,(\gamma_2))\ \Vert$	0.082	0.184	0.244	0.137	0.224	0.243	0.171	0.300	0.309	0.133	0.262	0.320
$\Vert\ Bias\,(var\,(\gamma_3))\ \Vert$	0.037	0.070	0.103	0.034	0.055	0.064	0.073	0.128	0.110	0.073	0.142	0.227
$av\,(RMSE\,\{var\,(\gamma_1)\})$	0.158	0.162	0.128	0.083	0.073	0.049	0.298	0.253	0.129	0.205	0.279	0.167
$av\,(RMSE\,\{var\,(\gamma_2)\})$	0.157	0.163	0.130	0.049	0.043	0.028	0.070	0.062	0.039	0.043	0.050	0.038
$av\,(RMSE\,\{var\,(\gamma_3)\})$	0.041	0.048	0.045	0.014	0.012	0.009	0.027	0.027	0.017	0.024	0.027	0.026
% Neg. Def.							0.930	0.740	0.380			

Table 6: EM-REML, Swamy and Mean Group Estimators Properties when $T = 80$

	EM-REML			EM-REML_homos			Swamy			N_TS-MG		
T=80 / N	10	30	80	10	30	80	10	30	80	10	30	80
$Bias\,(c_i)$	**0.000**	**0.005**	**0.003**	**0.002**	**0.001**	**-0.004**	**0.018**	**0.009**	**0.000**	**0.042**	**0.028**	**0.016**
$se\,\{Bias\,(c_i)\}$	0.008	0.005	0.002	0.009	0.004	0.003	0.012	0.005	0.003	0.015	0.005	0.004
$Bias\,(\beta_i)$	**0.000**	**0.002**	**-0.002**	**0.004**	**0.003**	**-0.001**	**0.005**	**0.004**	**-0.001**	**0.005**	**0.007**	**0.002**
$se\,\{Bias\,(\beta_i)\}$	0.008	0.004	0.002	0.009	0.003	0.002	0.009	0.003	0.002	0.009	0.003	0.002
$Bias\,(\phi_i)$	**0.014**	**0.001**	**0.001**	**-0.004**	**0.005**	**0.001**	**-0.021**	**-0.006**	**-0.005**	**-0.043**	**-0.037**	**-0.038**
$se\,\{Bias\,(\phi_i)\}$	0.009	0.003	0.002	0.005	0.002	0.002	0.005	0.002	0.002	0.005	0.002	0.002
$\|\,Bias\,(\psi)\,\|$	**0.014**	**0.005**	**0.003**	**0.006**	**0.006**	**0.004**	**0.028**	**0.012**	**0.005**	**0.060**	**0.047**	**0.041**
$RMSE\,(c_i)$	0.078	0.046	0.025	0.088	0.040	0.027	0.121	0.047	0.029	0.152	0.061	0.041
$RMSE\,(\beta_i)$	0.084	0.038	0.023	0.093	0.035	0.024	0.091	0.035	0.023	0.093	0.035	0.023
$RMSE\,(\phi_i)$	0.088	0.026	0.018	0.053	0.022	0.016	0.055	0.023	0.018	0.064	0.042	0.041
$\|\,Bias\,(var\,(\gamma_1))\,\|$	0.083	0.039	0.235	0.433	0.152	0.704	0.521	0.190	0.677	0.207	0.101	0.413
$\|\,Bias\,(var\,(\gamma_2))\,\|$	0.178	0.074	0.167	0.237	0.082	0.396	0.243	0.089	0.396	0.281	0.095	0.443
$\|\,Bias\,(var\,(\gamma_3))\,\|$	0.041	0.022	0.040	0.059	0.020	0.097	0.059	0.025	0.097	0.063	0.036	0.104
$av\,(RMSE\,\{var\,(\gamma_1)\})$	0.251	0.059	0.120	0.141	0.029	0.079	0.238	0.044	0.081	0.090	0.022	0.053
$av\,(RMSE\,\{var\,(\gamma_2)\})$	0.256	0.060	0.118	0.082	0.016	0.045	0.094	0.019	0.045	0.089	0.017	0.050
$av\,(RMSE\,\{var\,(\gamma_3)\})$	0.047	0.018	0.025	0.021	0.005	0.011	0.022	0.006	0.011	0.020	0.007	0.012
% Neg. Def.							0.740	0.320	0.000			

References

[1] Baltagi, 2005 «Econometric Analysis of Panel Data»

[2] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B

[3] Hsiao, Cheng, 2003. "Analysis of Panel Data". Cambridge University Press

[4] Cheng Hsiao & M. Hashem Pesaran, 2004. "Random Coefficient Panel Data Models," CESifo Working Paper Series 1233, CESifo Group Munich.

[5] Chudik, Alexander & Pesaran, M. Hashem, 2015. "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors," Journal of Econometrics, Elsevier, vol. 188(2), pages 393-420.

[6] Nadeem U. Haque, M. Hashem Pesaran and Sunil Sharma "Neglected Heterogeneity and Dynamics in Cross-Country Savings Regressions" in "Panel Data Econometrics: Future Directions" edited by J. Krishnakumar and E. Ronchetti. Elusiveness Science B.V.

[7] Harville (1977) «Maximum Likelihood Approaches by Variance Component Estimation and to Related Problems». Journal of the American Statistical Association, 72, 320-340.

[8] Hsiao et al. (1995) "Fixed vs Random Effects Specification for Panel Data Analysis" unpublished manuscript

[9] Henderson (1983) "Applications of Linear Models in Animal Breeding"

[10] Hsiao, Pesaran and Tahmiscioglu (1998)."Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models" in Analysis of Panels and Limited Dependent Variables by Hsiao

[11] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu (2002). "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," Journal of Econometrics, 109, 107–150.

[12] Judge et al. 1985 «The Theory and Practice of Econometrics»

[13] Lee, L.F., and W.E. Griffiths (1979). "The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models,"Working Papers in Econometrics and Applied Statistics, No. 1. University of New England

[14] Lindley, D.V., and A.F.M. Smith (1972). “Bayes Estimates for the Linear Model,” and Discussion, Journal of theFormal Statistical Society, Series B, 34, 1–41.

[15] Lutkepohl, 1996. «Handbook of Matrices»

[16] Maddala, G.S., R.P. Trost, H. Li and F. Joutz, 1997, Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators, Journal of Business and Economic Statistics 15, 90–100.

[17] Magnus and Neudecker, 2007. "Matrix Differential Calculus with Applications in Statistics and Econometrics"

[18] G McLachlan and T Krishnan, 2008. "The EM Algorithm and Extensions" 2nd edition

[19] Atif Mian and Amir Sufi, «House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again» (2014), The University of Chicago Press.

[20] Nerlove and Balestra - «The Econometrics of Panel Data»

[21] Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.

[22] Pawitan (2001) - “In All Likelihood: Statistical Modelling and Inference Using Likelihood”.

[23] M. Hashem Pesaran, 2006. "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," Econometrica, Econometric Society, vol. 74(4), pages 967-1012, 07.

[24] MH Pesaran, Y Shin, RP Smith, 1999. «Pooled mean group estimation of dynamic heterogeneous panels». Journal of the American Statistical Association

[25] Pesaran, M. Hashem & Smith, Ron, 1995. "Estimating long-run relationships from dynamic heterogeneous panels," Journal of Econometrics, Elsevier, vol. 68(1), pages 79-113, July.

[26] Hashem Pesaran, Ron Smith and Kyung So Im «Dynamic Linear Models for Heterogenous Panels» in Nerlove and Balestra - «The Econometrics of Panel Data» (Chapt. 8 pag. 145).

[27] Rao, 1973. Linear Statistical Inference and Its Applications, 2nd edition. New York: Wiley.

[28] Robinson G.K. (1991). « That BLUP is a Good Thing: The Estimation of Random Effects». Statistical Science.

[29] Searle S.R. (1978) “A Notebook on Variance Components: A detailed Description of Recent Methods of Estimating Variance Components with Applications in Animal Breeding”.

[30] Smith, A.F.M. (1973). “A General Bayesian Linear Model,” Journal of the Royal Statistical Society, B, 35, 67–75.

[31] Ron P. Smith and Ana-Maria Fuertes (2016), «Panel Time Series», (CEMMAP)

[32] P. A. V. B. Swamy, 1970 «Efficient Inference in a Random Coefficient Regression Model», Econometrica

[33] P. A. V. B. Swamy, 1971, Statistical Inference in Random Coefficient Regression Models, Springer-Verlag, New York.

[34] Stoker, T. M. (1993), "Empirical Approaches to the Problem of Aggregation Over Individuals," Journal of Economic Literature, 31, 1827-1874.

[35] Zellner, A. 1962. «An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias». Journal of the American statistical Association