ISSN 1745-8587

Birkbeck UNIVERSITY OF LONDON

Department of Economics, Mathematics and Statistics

Birkbeck Working Papers in Economics & Finance

BWPEF 1606

EM Estimation of Dynamic Panel Data Models with Heteroskedastic Random Coefficients

Andrea Nocera
Birkbeck, University of London

November 2016

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

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

Andrea Nocera
Birkbeck, University of London*

1st November 2016

Abstract

In this paper, we show how to combine the EM algorithm with the Restricted Maximum Likelihood (REML) method to estimate iteratively both the average effects and the unit-specific coefficients as well as the variance components in a wide class of dynamic heterogeneous panel data models. The estimation procedure can also be adapted to allow for cross-section dependence. Compared to existing methods, our approach allows for heteroskedastic random coefficients, and leads to an unbiased estimation of the variance components of the model without running into the problem of non-positive definite covariance matrices typically encountered in random coefficients models. 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.

JEL Classification: C13, C33, C63, F34, G15, H63.
Keywords: EM algorithm, restricted maximum likelihood, correlated random coefficient models, dynamic heterogeneous panels, debt intolerance, sovereign credit spreads.

*I am very grateful to Ron Smith and Zacharias Psaradakis for their careful reading, comments and advice. I would also like to thank Ivan Petrella and participants at the "New Trends and Developments in Econometrics" 2016 Conference organized by the Banco de Portugal, at the IAAE 2016 Annual Conference, and at Bristol Econometric Study Group 2016. I am also thankful to seminar participants at the Deutsche Bundesbank, and at Birkbeck in 2015 for their comments on a preliminary version of this paper. All remaining errors are mine.
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. Swamy focuses on estimating the average effects 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 unit or to describe the past behavior of a particular individual.[1] 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 as well as the variance components. The proposed estimation procedure is quite general, as we consider a broad framework which incorporates various panel data models as special case, and can accommodate recent developments in the dynamic heterogeneous panels literature, such as the CS-ARDL model developed by Chudik and Pesaran (2015). We also review the existing sampling and Bayesian methods commonly used to estimate heterogeneous dynamic panels, to highlight similarities and differences with the EM-REML approach.

Both the EM and the REML are commonly used tools to estimate linear mixed models but have been neglected by the literature on panel data with random coefficients.[2] We illustrate the merits of the EM-REML approach in estimating a general class of dynamic heterogeneous panels. The EM algorithm has also 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 effects population ("alphas") while the other coefficients of the model ("betas") 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

[1] Joint estimation of the individual parameters and their mean has been proposed by Lee and Griffiths (1979). Joint estimation in a Bayesian setting has been suggested by Lindley and Smith (1972), and has been further studied by Smith (1973), Maddala et al. (1997) and Hsiao et al. (1999). A good survey of the literature is provided by Hsiao and Pesaran (2004) and in Smith and Fuertes (2016).

[2] For discussions on EM and REML estimation of linear mixed models, see Harivlle (1977), Searle and Quaas (1978), Laird and Ware (1982), Pawitan (2001), and McLachlan and Krishnan (2008), among others.

an expression for the likelihood of the model accordingly. More importantly, our goal is to illustrate the advantages of the EM-REML approach in estimating heterogeneous panel data models, compared to the existing methods.

First, estimating heterogeneous panels by EM-REML yields an unbiased estimation of the variance components. This is important as the unbiased 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 a term to obtain a non-negative definite matrix. Although not unbiased, this alternative estimator is consistent when T tends to infinity. Lee and Griffiths (1979) derive 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. However, I demonstrate that their estimate of the coefficients residuals' variance-covariance matrix does not satisfy the law of total variance. 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 later on). 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. As a result, our approach should be preferred when T is relatively small. Another interesting feature of the EM (compared to the papers mentioned in this paragraph) 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.

Many economic applications involve behavioural 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. However, following Hsiao et al. (1999), we consider the first p observations of the dependent variable as fixed. We then derive an expression for the joint likelihood of the data and the random coefficients. Under such assumptions the lagged dependent variables are no longer endogenous, and the EM-REML yields unbiased estimators of the coefficients. The choice of using the first p observations as presample is also motivated by the need of directly comparing our approach with existing methods which have been developed only for the static case (e.g. Lee and Griffiths (1979) and Swamy (1970)).[3] At the same time, as discussed in Anderson and Hsiao (1981, 1982) regarding the first p observations as fixed can be a strong assumption for finite T.[4] Nevertheless, as it will be shown in the Monte Carlo

[3]The case where the initial p observations are treated as random is being investigated in a separate paper.

[4]Whenever treating the initial observations as fixed might be questionable, we resort to unbiasedness

analysis, the proposed method has good properties when estimating dynamic panels even when the sample size is relatively small.[5] Compared to Swamy and the Mean Group estimators, the EM-REML method leads to remarkable reduction of the bias of the estimates of the coefficients of the model and their variances. In view of the above reasons, the EM-REML approach should be regarded as a valid alternative to Bayesian estimation (e.g. Maddala et al. (1997) and Hsiao et al. (1999)) 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 (especially when the random coefficients have heteroskedastic variances). At the same time, a drawback of the Bayesian approach is that, when sample sizes are small (relative to the number of parameters being estimated), the prior choice will have a heavy weight on the posterior, which will consequently be far from being data dominated (Kass and Wasserman, 1996). The second merit of our proposed method is to overcome this problem while performing well even in relatively small samples.

Third, 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).[6] Ignoring heteroskedasticity when it is present still yields 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 errors or only the unit time-varying error components are assumed heteroskedastic can be seen as special case. Heteroskedasticity may occur in the Swamy's random coefficient model because of omitted information in the random coefficients equation and in many economic applications in which it may be more realistic to model the variance of the random coefficients as varying across units, conditional on some explanatory variables.[7]

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).

[5]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.

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

[7]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. While there is a large literature on the empirical determinants of sovereign yield spreads there is no work, to the best of our knowledge, which tries to quantify the sensitivity of financial markets during episode of debt growth.[8] Our analysis helps explain why some middle-income countries are considered to be riskier and unable to tolerate their debt, despite the fact that their debt-to-GDP ratios are considerably lower than those of several high-income countries. It provides a bridge between the strand of the literature on the problem of "debt intolerance", pioneered by Reinhart et al. (2003), and the aforementioned works on the determinants of sovereign risk spreads. First, we show that financial markets reactions to an increase in government debt are highly heterogeneous. We then model such reactions as function of macroeconomic 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.

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 shows how to perform the two steps of the EM algorithm, called the E-step and the M-step. We compare the EM-REML approach with alternative methods in Section 5. Results from Monte Carlo experiments are shown in Section 6. In Section 7, 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:[9]

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

[8] The effects of macroeconomic fundamentals on sovereign credit spreads are examined in Akitoby and Stratmann (2008), Bellas et al. (2010), Edwards (1984), Eichengreen and Mody (2000) and Hilscher and Nosbusch (2010), among others.

[9] 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. For the same reason, we assume that $T = T_i$, for all $i = 1, .., N$ although the results are also valid for an unbalanced panel.

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 aK-dimensional vector of unknown slope coefficients . 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 $s = 1,..,p$. Using the first p observations as presample, equation (1) can be rewritten as

$$y_i = Z_i \psi_i + \varepsilon_i, \tag{2}$$

where

$$\underset{((T-p)\times 1)}{y_i} = \begin{bmatrix} y_{ip+1} & \cdots & y_{iT} \end{bmatrix}', \qquad \underset{(K^* \times 1)}{\psi_i} = \begin{bmatrix} \psi'_{i0} & \cdots & \psi'_{ip} \end{bmatrix}',$$

$$\underset{((T-p)\times K^*)}{Z_i} = \begin{bmatrix} Z'_{ip+1} & \cdots & Z'_{iT} \end{bmatrix}', \qquad \underset{(1 \times K^*)}{Z_{it}} = \begin{bmatrix} \tilde{x}_{it} & z_{it-1} & \cdots & z_{it-p} \end{bmatrix},$$

with $K^* = (K+1)(p+1)$. Following Hsiao et al. (1993), in order to provide a more general framework which incorporates various panel data models as special case, we partition Z_i and ψ_i as

$$Z_i = \begin{bmatrix} \bar{Z}_i & \underline{Z}_i \end{bmatrix}, \qquad \psi_i = \begin{bmatrix} \psi_{1i} \\ \psi_{2i} \end{bmatrix},$$

where $\bar{Z}_i$ is $(T-p) \times k_1^*$ and $\underline{Z}_i$ is $(T-p) \times k_2^*$, with $K^* = k_1^* + k_2^*$. The coefficients ψ_{1i}'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:

$$\psi_{1i} = \Gamma_1 f_{1i} + \gamma_i, \tag{3}$$

where γ_i is a $k_1^* \times 1$ vector of random coefficients errors, Γ_1 and is a $(k_1^* \times l_1)$ matrix of unknown fixed parameters, f_{1i} is a $l_1 \times 1$ vector of observed explanatory variables that do not vary over time (for instance, Smith and Fuertes (2016) suggest using the group means of the x_{it}'s). The first element of f_{1i} is one to allow for an intercept. The coefficients of $\underline{Z}_i$ are non-stochastic and subject to

$$\psi_{2i} = \Gamma_2 f_{2i}, \tag{4}$$

where Γ_2 is a $(k_2^* \times l_2)$ matrix of unknown fixed parameters, and f_{2i} is a $l_2 \times 1$ vectors of observed unit-specific characteristics. Equations (3) and (4) can be rewritten as

$$\begin{aligned} \psi_{1i} &= \left(f'_{1i} \otimes I_{k_1^*}\right) \bar{\Gamma}_1 + \gamma_i, \\ \psi_{2i} &= \left(f'_{2i} \otimes I_{k_2^*}\right) \bar{\Gamma}_2, \end{aligned} \tag{5}$$

where $\bar{\Gamma}_j = vec(\Gamma_j)$, which is a $k_j^* l_j$-dimensional vector and $F_{ji} = \left(f'_{ji} \otimes I_{k_j^*}\right)$ is a $k_j^* \times k_j^* l_j$ matrix, for $j = 1, 2$. Substituting (5) into (2) yields

$$y_i = W_i \bar{\Gamma} + \bar{Z}_i \gamma_i + \varepsilon_i, \tag{6}$$

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

$$\underset{(T-p)\times \bar{K}}{W_i} = \begin{bmatrix} \bar{Z}_i F_{1i} & \underline{Z}_i F_{2i} \end{bmatrix}, \qquad \underset{\bar{K}\times 1}{\bar{\Gamma}} = \begin{bmatrix} \bar{\Gamma}_1 \\ \bar{\Gamma}_2 \end{bmatrix},$$

with $\bar{K} = (k_1^* l_1 + k_2^* l_2)$. We assume that:

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

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

(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

$$\underset{(Nk_1^* \times Nk_1^*)}{\Theta_\gamma} = \begin{bmatrix} \triangle_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \triangle_N \end{bmatrix}. \tag{9}$$

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 (3), 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 (3) 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. Moreover, 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 (2) is now replaced by

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

where Z_i and ψ_i are defined as above.

Estimation of Long-Run Effects. 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.

Special Cases. Many panel data models can be derived as special cases of the model described above. Among others:

1. Models in which all the coefficients are stochastic and depend on individual-specific characteristics can be obtained from (6) by setting $\underline{Z}_i = 0$.

2. Swamy (1970) random coefficients model requires $\underline{Z}_i = 0$, and $f_{1i} = 1$ for all $i = 1, .., N$, while $\bar{\Gamma} = \psi$ is a $K^* \times 1$ vector of coefficients. Finally, $\triangle_i = \triangle$, for all i.

3. The correlated random effects (CRE) model proposed by Mundlak (1978) and Chamberlain (1982) can be obtained by setting $\bar{Z}_i = \iota$ (where ι is a vector of ones), f_{1i} contains $\bar{x}_i$, the average over time of the x_{it}'s; $f_{2i} = 1$ for all i, which implies that $\psi_{2i} = \psi_2$ is a vector of common coefficients..

4. Error-components models (as described in Baltagi (2005) and in Hsiao (2003)) which are a special case of the CRE model with $f_{1i} = 1$ for all i and $\Gamma_1 \equiv c \in \mathbb{R}$. Typically, it is assumed that both the random effects and the error terms are homoskedastic. In the static case, Mazodier and Trognon (1978) assume the c_i are heteroskedastic, i.e. $c_i \sim (0, \sigma^2_{c_i})$ while $\varepsilon_{it} \sim IID(0, \sigma^2_\varepsilon)$. Roy (2002) considers the case where $E\left(c_i \mid \bar{x}_i\right) = 0$ and $var(c_i \mid \bar{x}_i) = g(\bar{x}_i) = \omega_i$, which means that the conditional variance of the unit-specific error term has heteroskedasticity of unknown functional. Randolph (1988) considers the case where both c_i and ε_{it} are heteroskedastic.

5. Model with interaction terms (e.g. Friedrich (1982)): $\bar{Z}_i = 0$ and for instance $f_{2i} = (1, d_i)'$ where d_i is a dichotomous independent variable.

6. Common Model for all cross-sectional units: $\bar{Z}_i = 0$ and $f_{2i} = 1$ for all i.[10]

3 Likelihood of the Complete Data

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

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

where $\sigma^2_\varepsilon = (\sigma^2_{\varepsilon 1}, .., \sigma^2_{\varepsilon N})'$ 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 residuals, γ, as the vector of missing data and $(y', \gamma')'$ as the complete data vector. 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[11]

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

[10] Models 5 and 6 do not involve any random coefficients and do not require the use of the EM algorithm.
[11] To make notation easier, we write $f(\gamma; \omega) = f(\gamma)$ and $f(y \mid \gamma)$ instead of $f(y \mid Z, \gamma; \theta_1)$.

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.[12] In that case, from (6) we can easily derive the conditional expectation and variance of y_i which are given by $E(y_i \mid \gamma_i) = W_i\bar{\Gamma} + \bar{Z}_i\gamma_i$ and $var(y_i \mid \gamma_i) = var(\varepsilon_i) = R_i$, respectively. 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 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

$$\varepsilon_i = y_i - W_i\bar{\Gamma} - \bar{Z}_i\gamma_i. \tag{17}$$

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, 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}. 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

[12] 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 the absolute value of the autoregressive coefficient of a Gaussian AR(1) process is less than one, $\mid \phi \mid < 1$, 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.

likelihoods of $S_i y_i$ and $Q_i y_i$, i.e.

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

Such matrices must satisfy the following properties: (i) the rank of S_i is $T - p - \bar{K}$ while Q_i is a matrix of rank $\bar{K}$, (ii) L_{1i} and L_{2i} are statistically independent, i.e. $cov\,(S_i y_i, Q_i y_i) = 0$, (iii) the matrix S_i is chosen so that $E\,(S_i y_i) = 0$, i.e. $S_i W_i = 0$, and (iv) the matrix $Q_i W_i$ must be of rank $\bar{K}$.[13]

Finding an expression for L_{1i}. Premutiplying both sides of (6) by S_i, we have $E\,(S_i y_i \mid \gamma_i) = S_i \bar{Z}_i \gamma_i$, since $S_i W_i = 0$ and $var\,(S_i y_i \mid \gamma_i) = S_i R_i S_i'$. Therefore, the conditional log-likelihood of $S_i y_i$ is given by

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

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, 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 $\bar{\varepsilon}_i = y_i - W_i \hat{\bar{\Gamma}} - \bar{Z}_i \gamma_i$.

Finding an expression for L_{2i}. Following Patterson and Thompson (1971), we can set $Q_i = W_i' R_i^{-1}$ since it satisfies $cov\,(S_i y_i, Q_i y_i) = 0$. After premutiplying both sides of (6) by Q_i, we have $E\,(Q_i y_i \mid \gamma_i) = W_i' R_i^{-1}\left(W_i \bar{\Gamma} + Z_i \gamma_i\right)$ and $var\,(Q_i y_i \mid \gamma_i) = W_i' R_i^{-1} W_i$. The log-likelihood of $Q_i y_i \mid \gamma_i$ is given by

$$L_{2i} = c_4 - \frac{1}{2} log \mid W_i' R_i^{-1} W_i \mid - \frac{1}{2} \varepsilon_i' H_i \varepsilon_i, \tag{21}$$

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).

[13] See Appendix 9.1 for detailed computations.

4 EM-Algorithm

4.1 Generalities

Using equations (14), (15) and (16), the log-likelihood of the complete data can be rewritten as

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

Lee and Griffiths (1979) obtain iterative estimates of θ and γ by maximizing directly the latter. Instead, we argue in favour of using the EM algorithm to compute maximum likelihood estimates as this method has some added advantages. First, as established in Dempster et al. (1977), the EM algorithm assures that each iteration increases the likelihood. Second, as it will be shown in the next sections, contrary to Lee and Griffiths approach which delivers $var\left\{E\left(\gamma_i \mid y_i\right)\right\}$ as an estimator of $var(\gamma_i)$, the unconditional variance of the γ_i, the EM algorithm yields an unbiased estimator of the latter. Finally, the EM allows us to make inference on the random coefficients population.

Moreover, to obtain unbiased estimates of the variances of the time-varying disturbances, following Patterson and Thompson (1971), we consider the complete-data (restricted) log-likelihood:

$$logL(y_i,\gamma_i;\theta)=L_{1i}+L_{2i}+logf(\gamma_i;\omega_i), \tag{22}$$

for $i=1,..,N$, where $logf(y_i \mid \gamma_i;\theta_1)$ has been decomposed as shown in equation (18).

On each iteration of the EM 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 bth iteration, for $b=1,2,..$, the E-step requires computing the conditional expectation of the $logL(y,\gamma;\theta)$ given y, using $\theta^{(b-1)}$ for θ, which is given by

$$\begin{aligned} Q=Q(\theta;\theta^{(b-1)}) &= E_{\theta^{(b-1)}}\left\{logL(y,\gamma;\theta)\mid y\right\} \\ &= \sum_{i=1}^{N}E_{\theta^{(b-1)}}\left\{logL(y_i,\gamma_i;\theta)\mid y_i\right\} \;=\; \sum_{i=1}^{N}Q_i, \end{aligned} \tag{23}$$

where

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

and

$$\begin{aligned} Q_{1i} &= E_{\theta^{(b-1)}}\left\{L_{1i}\mid y_i\right\}, \\ Q_{2i} &= E_{\theta^{(b-1)}}\left\{L_{2i}\mid y_i\right\}, \\ Q_{3i} &= E_{\theta^{(b-1)}}\left\{logf(\gamma_i;\omega_i)\mid y_i\right\}. \end{aligned} \tag{24}$$

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 $Q(\theta;\theta^{(b-1)})$ with respect to the parameters of interest, θ. That is, we choose $\theta^{(b)}$ such that $Q(\theta^{(b)};\theta^{(b-1)}) \geq Q(\theta;\theta^{(b-1)})$. In other words, the M-step chooses $\theta^{(b)}$ as

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

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.[14] The conditional expectation of γ_i given the data is

$$\begin{aligned} \hat{\gamma}_i = E\left(\gamma_i \mid y_i\right) &= \triangle_i \bar{Z}_i' \left(\bar{Z}_i \triangle_i \bar{Z}_i' + R_i\right)^{-1} (y_i - W_i\bar{\Gamma}) \\ &= \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle_i^{-1}\right)^{-1} \bar{Z}_i' R_i^{-1} \left(y_i - W_i\bar{\Gamma}\right), \end{aligned} \tag{25}$$

which is also the argument that maximizes the complete data likelihood, as defined in (14), with respect to γ_i. It can be noted from the first equality of (25) that this expression is equivalent to the predictor of the random coefficients residuals derived in Lee and Griffiths (1979), Lindley and Smith (1972) and Smith (1973). Two differences emerge. The first concerns the way the coefficients and the variances components are estimated. The second is that here we allow $\triangle_i \neq \triangle_j$, for $i \neq j$.

The conditional variance of γ_i is given by

$$V_{\gamma_i} = var\left(\gamma_i \mid y_i\right) = \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle_i^{-1}\right)^{-1}, \tag{26}$$

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

[14] For details, see Appendix 9.2.

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 (25) and (26), respectively.

4.3 E-step

At each iteration, the E-step requires the calculation of the conditional expectation of (22) given the observed data and the current fit for the parameters, to obtain an expression for $Q_i(\theta)$, for $i = 1, .., N$.[15]

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

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

where $\hat{\bar{\varepsilon}}_i = y_i - W_i \bar{\Gamma}^{(b)} - \bar{Z}_i \hat{\gamma}_i^{(b)}$, into $E_{\theta^{(b-1)}} \{L_{1i} \mid y_i\}$, yields

$$Q_{1i} = E_{\theta^{(b-1)}}\left(L_{1i} \mid y_i\right) = \begin{array}{l} c_3 - \frac{1}{2} log|R_i| - \frac{1}{2} log \mid W_i' R_i^{-1} W_i| \\ -\frac{1}{2} Tr\left(\bar{Z}_i' R_i^{-1} \bar{Z}_i V_{\gamma_i}^{(b)}\right) - \frac{1}{2} \hat{\bar{\varepsilon}}_i' R_i^{-1} \hat{\bar{\varepsilon}}_i. \end{array} \tag{27}$$

where $\hat{\gamma}_i^{(b)}$ and $V_{\gamma_i}^{(b)}$ are given by (25) and (26) respectively, after substituting the current fit for θ at iteration each iteration $b = 1, 2, ...$

To obtain Q_{2i}, we take the conditional expectation of (21). Substituting

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

where $\hat{\varepsilon}_i = y_i - W_i \bar{\Gamma} - \bar{Z}_i \hat{\gamma}_i^{(b)}$, into $E_{\theta^{(b-1)}} \{L_{2i} \mid y_i\}$, yields

$$Q_{2i} = E_{\theta^{(b-1)}}\left(L_{2i} \mid y_i\right) = \begin{array}{c} c_4 - \frac{1}{2} log \mid W_i' R_i^{-1} W_i \mid \\ -\frac{1}{2} Tr\left(\bar{Z}_i' H_i \bar{Z}_i V_{\gamma_i}^{(b)}\right) - \frac{1}{2} \hat{\varepsilon}_i' H_i \hat{\varepsilon}_i. \end{array} \tag{28}$$

Finally, substituting

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

into $E_{\theta^{(b-1)}} \{log f(\gamma_i) \mid y_i\}$, yields

$$Q_{3i} = E_{\theta^{(b-1)}}\left(log f(\gamma_i) \mid y\right) = \begin{array}{c} -\frac{K^*}{2} log 2\pi + \frac{1}{2} log \mid \triangle_i^{-1} \mid \\ -\frac{1}{2} Tr\left(\triangle_i^{-1} V_{\gamma_i}^{(b)}\right) - \frac{1}{2} \hat{\gamma}_i^{(b)'} \triangle_i^{-1} \hat{\gamma}_i^{(b)} \end{array} \tag{29}$$

[15] Detailed computations are shown in Appendix 9.3.

4.4 M-step

The M-Step consists in maximizing (23) 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

$$\bar{\Gamma}^{(b)} = \left(\sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1} W_i\right)^{-1} \sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1}\left(y_i - \bar{Z}_i\hat{\gamma}_i^{(b)}\right). \tag{30}$$

which is equivalent to the GLS estimation of $\bar{\Gamma}$ when the model is given by $y_i^* = W_i\bar{\Gamma} + \varepsilon_i$, where $y_i^* = y_i - \bar{Z}_i\gamma_i$, 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 marginal log-likelihood of the observed data (e.g. $logf(y;\theta)$) since computations are not complicated. Therefore, after convergence, the standard errors of $\bar{\Gamma}^{(B)}$ can be computed as the square root of the diagonal elements of

$$I\left(\bar{\Gamma}^{(B)}\right)^{-1} = \left(\sum_{i=1}^{N} W_i' V_{i(B)}^{-1} W_i\right)^{-1} \tag{31}$$

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 (23). Because Q_{3i} is not a function of $\sigma^2_{\varepsilon_i}$ and given that no information is lost by neglecting Q_{2i} (as noted by Patterson and Thompson (1971) and Harville (1977)), we base inferences for $\sigma^2_{\varepsilon_i}$ only on Q_{1i}, which is defined in (27).

Substituting $R_i = var\left(\varepsilon_i\right) = \sigma^2_{\varepsilon_i} I_{T-p}$ into (27) and equating the first derivative of the latter with respect to $\sigma^2_{\varepsilon_i}$ to zero, yields

$$\sigma^{2(b)}_{\varepsilon_i} = \frac{\hat{\hat{\varepsilon}}_i'\hat{\hat{\varepsilon}}_i + Tr\left(\bar{Z}_i'\bar{Z}_i V_{\gamma_i}^{(b)}\right)}{T-p-r(W_i)} \tag{32}$$

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

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

$$\begin{aligned} \triangle_i = var(\gamma_i) &= var[E(\gamma_i \mid y_i)] + E[var(\gamma_i \mid y_i)] \\ &= var(\hat{\gamma}_i) + E(V_{\gamma_i}). \end{aligned} \tag{33}$$

Therefore, it can be shown that

$$\hat{\triangle}_i = \hat{\gamma}_i \hat{\gamma}_i' + V_{\gamma_i} \tag{34}$$

is an unbiased estimator of $\triangle_i$. Indeed, taking expectation of both sides of (34) and using (33), we get

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

Notably, the EM estimator of the variance-covariance matrix of the random coefficients residuals (which is the argument which maximizes (29) with respect to $\triangle_i$, for $i = 1, .., N$) is equal to

$$\triangle_i^{(b)} = \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} + V_{\gamma_i}^{(b)}, \tag{35}$$

which is equivalent to (34) after substituting the unknown parameters with their current fit in the EM algorithm.[16]

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 $R_{i(0)} = \sigma_{\varepsilon_i}^{2(0)} I_{T-p}$. One can use Swamy (1970) estimators, which are reported in the next Section. Then, for $b = 1, 2, ..$

1. Given the current fit for θ at iteration b, we compute $var\left(\gamma_i \mid y_i, \theta^{(b-1)}\right)$ and $E_{\theta^{(b-1)}}(\gamma_i \mid y_i)$, which are given by

$$\begin{aligned} V_{\gamma_i}^{(b)} &= \left(\bar{Z}_i' R_{i(b-1)}^{-1} \bar{Z}_i + \triangle_{i(b-1)}^{-1}\right)^{-1}, \\ \hat{\gamma}_i^{(b)} &= V_{\gamma_i}^{(b)} \bar{Z}_i' R_{i(b-1)}^{-1} \left(y_i - W_i \bar{\Gamma}^{(b-1)}\right), \end{aligned}$$

respectively.

[16] See Appendix 9.4, for more details.

2. The average coefficients are given by

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

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

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

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

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

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.

5 Comparison between EM-REML Estimation and Alternative Methods.

5.1 Average Effect

Representing (30) and (25) as a system of two equations, Searle (1978, eq. 3.17) demonstrated that these two formulae can be rewritten as

$$\hat{\hat{\Gamma}} = \left(\sum_{i=1}^{N} W_i' V_i^{-1} W_i\right)^{-1} \sum_{i=1}^{N} W_i' V_i^{-1} y_i, \tag{36}$$

$$\hat{\gamma}_i = \triangle_i \bar{Z}_i' V_i^{-1} \left(y_i - W_i \hat{\hat{\Gamma}}\right), \tag{37}$$

respectively. Note that $\hat{\hat{\Gamma}}$ is the estimator which maximizes the log-likelihood function constructed by referring to the marginal distribution of the dependent variable. When $f_i = 1$ for all i, and $W_i = \bar{Z}_i$, equation (36) is equivalent to the Swamy (1970) GLS estimator.[17] The latter can be rewritten as a weighted average of the least squares estimates of individual units:

$$\hat{\hat{\Gamma}} = \sum_{i=1}^{N} \Psi_i \hat{\psi}_i, \tag{38}$$

[17]To obtain exact equivalence, one should also restrict $\triangle_i = \triangle$, for all $i = 1, .., N$.

where

$$\begin{array}{rcl}\Psi_i & = & \left\{\sum_{i=1}^{N}[\triangle_i+\sigma^2_{\varepsilon_i}(\bar{Z}_i'\bar{Z}_i)^{-1}]^{-1}\right\}^{-1}[\triangle_i+\sigma^2_{\varepsilon_i}(\bar{Z}_i'\bar{Z}_i)^{-1}]^{-1}, \\ \hat{\psi}_i & = & (\bar{Z}_i'\bar{Z}_i)^{-1}\bar{Z}_i'y_i. \end{array} \quad (39)$$

Swamy's estimator is a two-step procedure, which requires first to estimate N time series separately as if the individual coefficients were fixed and all different in each cross-section. Instead, the EM-REML is an iterative method which shrinks the unit-specific parameters toward a common mean. Moreover, Swamy's estimator is the minimizer of the weighted sum of squared errors (or the maximizer of the marginal likelihood of y) and being developed for static models, ignore that the lagged dependent variables are correlated with the random coefficients' errors. Instead, the EM-REML estimator maximizes (iteratively) the joint likelihood of the data and the random coefficients' errors. Maddala et al. (1997) argue in favour of iterative procedures when the model includes lagged dependent variables since, as indicated in Amemiya and Fuller (1967), Maddala (1971) and Pagan (1986), when estimating dynamic models, the two-step estimators based on any consistent estimators of $\sigma^2_{\varepsilon_i}$ and $\triangle$ are consistent but not efficient.

Hsiao et al. (1999) show that $\hat{\bar{\Gamma}}$ is the posterior mean of $\bar{\Gamma}$ in a Bayesian approach which assumes the prior distribution of $\bar{\Gamma}$ is normal with mean μ and variance Ψ, with $\Psi^{-1}=0$. Another important contribution of the aforementioned paper is to establish that the Bayes estimator $\hat{\bar{\Gamma}}$ is asymptotic equivalent to the mean group estimator proposed by Pesaran and Smith (1995), as $T\to\infty$, $N\to\infty$ and $\sqrt{N}/T^{\frac{3}{2}}\to 0$.

5.2 Unit-Specific Parameters

Without loss of generality, for comparison purposes, let us focus on the case where $f_{1i}=1$, $\forall i$ and $\underline{Z}_i=0$. Substituting (36) and (25) into (5) yields the best linear unbiased predictor of ψ_i, which Lee and Griffiths (1979) have shown to be equal to[18]

$$\begin{array}{rcl}\tilde{\psi}_i & = & \hat{\bar{\Gamma}}+\hat{\gamma}_i \\ & = & \left(\bar{Z}_i'R_i^{-1}\bar{Z}_i+\triangle_i^{-1}\right)^{-1}\left(\left(\bar{Z}_i'R_i^{-1}\bar{Z}_i\right)\hat{\psi}_i+\triangle_i^{-1}\hat{\bar{\Gamma}}\right). \end{array} \quad (40)$$

The latter expression is also equivalent to the empirical Bayes estimator of ψ_i, described in Maddala et al. (1997) although the latter is a two-step estimator and does not involve any iteration. Interestingly, as shown in Smith (1973), the Swamy GLS estimator, given in (36), can be rewritten as a simple average of the $\tilde{\psi}_i$:

$$\hat{\bar{\Gamma}}=\frac{1}{N}\sum_{i=1}^{N}\tilde{\psi}_i. \quad (41)$$

[18]Lee and Griffiths (1979) assume $\triangle_i=\triangle$, $\forall i$.

Assessing the Errors of Estimation. In the general case, the standard errors of the predictor of ψ_{1i}, defined in (5), can be computed as the square root of the diagonal elements of[19]

$$var\left(\tilde{\psi}_{1i} - \psi_{1i}\right) = F_{1i}\Xi F_{1i}' + var\left(\hat{\gamma}_i - \gamma_i\right) - F_{1i}\Lambda - \Lambda' F_{1i}', \tag{42}$$

where

$$\Lambda = cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \gamma_i\right) = \Xi W_i' V_i^{-1} \bar{Z}_i \triangle_i,$$

$$var\left(\hat{\gamma}_i - \gamma_i\right) = \triangle_i \left[I - \bar{Z}_i' V_i^{-1}\left(I + W_i \Xi W_i' V_i^{-1}\right)\bar{Z}_i \triangle_i\right],$$

and $\Xi = var\left(\hat{\bar{\Gamma}}\right)$ as defined in (31). Equation (40) is just a special case of (5), and its standard errors can also be obtained from (42) after substituting $F_{1i} = I$, for all i.

At the same time, one can exploit the fact that the EM algorithm provides a distribution over the random coefficients residuals. For instance, we suggest drawing S samples from

$$\gamma_i^{(s)} \mid y_i \sim N\left(\hat{\gamma}_i, V_{\gamma_i}\right) \tag{43}$$

where $\hat{\gamma}_i$ and V_{γ_i} are given by (25) and (26) respectively, to then report histograms for each unit for comparison and diagnostic purposes. Moreover, if we go to extremes, assuming prior ignorance on $\bar{\Gamma}$, as in Smith (1973) and Maddala et al. (1997), $\bar{\Gamma}$ can be drawn from its posterior distribution given by

$$\bar{\Gamma}^{(s)} \mid y \sim N\left(\hat{\bar{\Gamma}}, \Xi\right) \tag{44}$$

where $\hat{\bar{\Gamma}}$ and Ξ are given by (36) and (31) respectively. It follows that the individual coefficients, as defined in (5) can be drawn from the following Gaussian distribution:

$$\psi_{1i}^{(s)} \mid y_i \sim N\left(F_{1i}\hat{\bar{\Gamma}} + \hat{\gamma}_i, F_{1i}\Xi F_{1i}' + V_{\gamma_i}\right) \tag{45}$$

for $s = 1, .., S$.

Comparison between EM and a Full Bayesian Implementation. It is worth noting at this point some differences between the EM algorithm and Bayesian estimation. The EM gives a probability distribution over 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 full Bayesian version. However, this may come at a

[19] This expression is equivalent to the one proposed by Lee and Griffiths (1979). See Appendix 9.5 for computations.

cost in applications where the time dimension is small. As discussed in Kass and Wasserman (1996), when sample sizes are small (relative to the number of parameters being estimated) the prior choice will have a heavy weight on the posterior, which will consequently be far from being data dominated. At the same time, using a purely ‘’noninformative” prior (in the sense of Koop (2003)) may have the undesirable property that this prior ”density” does not integrate to one, which in turn may raise many of the problems discussed in the Bayesian literature.

The advantage of the EM compared to the iterative Bayesian approach developed by Lindley and Smith (1992)[20] and the Gibbs sampling-based approach suggested in Hsiao et al. (1999), 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. In fact, while the Bayesian point estimates incorporate prior information, the EM-REML estimates do not involve the starting values (chosen to initiate the algorithm). One can start with any initial value. As shown in Dempster et al. (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 cases, Pawitan (2001) suggests to try several starting values or to start with a sensible estimate. However, the EM and the full Bayesian method should be seen as complementary. Depending on the particular application, the researcher may decide which of the two approaches is more suitable.

5.3 Variance Components

We now compare (35) with the Swamy (1970) and Lee and Griffiths (1979) estimators of the random coefficients residuals' variance-covariance matrix.

Assuming that $\triangle_i = \triangle$, $\forall i$, Swamy (1970) suggested estimating $var\left(\gamma_i\right)$ as

$$\begin{aligned} \hat{\triangle} \;\; = \;\; & \tfrac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\psi}_i - N^{-1}\sum_{i=1}^{N}\hat{\psi}_i\right)\left(\hat{\psi}_i - N^{-1}\sum_{i=1}^{N}\hat{\psi}_i\right)' \\ & -\tfrac{1}{N}\sum_{i=1}^{N}\hat{\sigma}^2_{\varepsilon_i}(\bar{Z}_i'\bar{Z}_i)^{-1} \end{aligned} \tag{46}$$

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

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{T-K^*}\left(y_i - \bar{Z}_i\hat{\psi}_i\right)'\left(y_i - \bar{Z}_i\hat{\psi}_i\right) \tag{47}$$

are the OLS estimated variances of the residual terms. However, (46) is not necessarily

[20] See Maddala et al. (1997) for a further discussion on the iterative Bayesian approach.

nonnegative definite. Therefore, if that is the case the author suggests considering only

$$\hat{\triangle}_S = \frac{1}{N-1} \sum_{i=1}^{N} \left(\hat{\psi}_i - N^{-1} \sum_{i=1}^{N} \hat{\psi}_i \right) \left(\hat{\psi}_i - N^{-1} \sum_{i=1}^{N} \hat{\psi}_i \right)' \tag{48}$$

Although not unbiased, the latter estimator is nonnegative definite and consistent when T tends to infinity. This estimator is also used in the empirical Bayesian approach and in Lee and Griffiths' «modified mixed estimation» procedure. 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:[21]

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{T} \left(y_i - \bar{Z}_i \tilde{\psi}_i \right)' \left(y_i - \bar{Z}_i \tilde{\psi}_i \right) \tag{49}$$

where $\tilde{\psi}_i$ is given by (40), and

$$\hat{\triangle}_{LG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\gamma}_i \hat{\gamma}_i' \tag{50}$$

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. At the same time, we have seen that the latter is equivalent to the argument which maximizes the joint likelihood of the observed data and random coefficients residuals, given in (14). This is the approach followed by Lee and Griffiths (1979). We argue in favor of treating the joint likelihood as an incomplete data problem to then applying the EM algorithm to obtain maximum likelihood estimates because, among the other reasons highlighted in Section 4, the expected value of (50) does not satisfy the law of total variance while the EM 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. In fact, assuming homoskedasticity (for comparison purposes), we can establish that

$$E\left(\hat{\triangle}_{LG}\right) \leq E\left(\hat{\triangle}_{EM}\right) \equiv \triangle \leq E\left(\hat{\triangle}_S\right) \tag{51}$$

where

$$\hat{\triangle}_{EM} = \frac{1}{N} \sum_{i=1}^{N} \{ V_{\gamma_i} + \hat{\gamma}_i \hat{\gamma}_i' \} \tag{52}$$

is the EM estimate when $\triangle_i = \triangle$, for $i = 1, .., N$. Result (51) is of relevance when T is small, because $\triangle$ appears not only in both the formula for the average effect and the predicted

[21] In this Section, we omit the superscript $b = 1, 2, ...$ in $\tilde{\psi}_i^{(b)}$ and $\hat{\gamma}_i^{(b)}$, for ease of exposition even though the solutions are iterative.

random coefficients residuals but also in their standard errors. Testing hypothesis crucially depends on correctly estimating the random coefficients residuals variances.

Finally, we report the Bayes mode of the posterior distribution of $\triangle$ and $\sigma^2_{\varepsilon_i}$ suggested by Lindley and Smith (1972) and Smith (1973), which are equal to

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{T + \upsilon_i + 2} \left\{ \upsilon_i \lambda_i + \left(y_i - \bar{Z}_i \tilde{\psi}_i \right)' \left(y_i - \bar{Z}_i \tilde{\psi}_i \right) \right\} \tag{53}$$

$$\bar{\triangle} = \frac{1}{N + \rho - K^* - 2} \left\{ \Upsilon + \sum_{i=1}^{N} \hat{\gamma}_i \hat{\gamma}_i' \right\} \tag{54}$$

respectively, under the assumption that $\triangle^{-1}$ has a Wishart distribution, with ρ degrees of freedom and matrix Υ and $\sigma^2_{\varepsilon_i}$ follows a $\chi^{(2)}$ with prior parameters υ_i and λ_i, and is independent of $\triangle$. Note from (40) that $\hat{\gamma}_i = \tilde{\psi}_i - \hat{\bar{\Gamma}}$, when assuming a noninformative prior for $\bar{\Gamma}$. Smith (1973) suggests vague priors by setting $\rho = 1$ and Υ to be a diagonal matrix with small positive entries (such as .001). Note that, by setting $\rho = K^* + 2$, $\upsilon_i = -p - r\left(W_i\right) - 2$ and $\upsilon_i \lambda_i = Tr\left(\bar{Z}_i' \bar{Z}_i \Upsilon \right)$, we can consider (53) and (54) as a close approximation to the EM-REML estimates given by (32) and (52) respectively.

6 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 some 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

$$\begin{aligned} y_{it} &= c_i + \beta_i x_{it} + \phi_i y_{it-1} + \varepsilon_{it} \\ x_{it} &= c_{x,i}(1-\rho) + \rho x_{it-1} + u_{it} \end{aligned} \tag{55}$$

where

$$\begin{aligned} u_{it} &\sim N(0,1) \\ \varepsilon_{it} &\sim i.i.d.N(0, \sigma^2_{\varepsilon_i}) \\ c_{x,i} &\sim N(1,1) \end{aligned} \tag{56}$$

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

$$\begin{aligned} c_i &= \mu + \gamma_{1i} \\ \beta_i &= \beta + \gamma_{2i} \\ \phi_i &= \phi + \gamma_{3i} \end{aligned} \tag{57}$$

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 behaviour. For each $i = 1, .., N$ we eliminate the first 200 observations generated in the experiments to minimize the effect of initial observations.

The results are based on 1000 replications. Tables 4, 5, and 6 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) for T equal to 10, 30, and 80 respectively. 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 for $i = 1, .., N$, we consider the Euclidean 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 $\hat{\sigma}^2_{\gamma_j,i}$.[22]

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 1, 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 3. 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 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

[22] In the columns "N-Time Series - MG" of Tables 4, 5, and 6, the estimated variances are obtained estimating N time series separately.

Swamy's covariance matrix estimator given in (46) is negative definite. Using (48) instead of the latter results in a biased estimator. Only when both T and N are large the probability of (46) being negative definite are small. Although in large samples the differences with the EM-REML reduces, the latter continues to have an advantage. Infact, 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. Monte Carlo experiments corroborate Maddala et al. (1997) argument in favor of iterative procedures to two-step estimators and confirm Hsiao et al. (1999) finding that the MG estimator is unlikely to be an appropriate estimator when either T or N is small.

7 Application

Reinhart et al. (2003), studying sovereigns' credit histories since the early nineteenth century, argue that an important portion 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 subgroup 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.[23] 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).[24] 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

[23]See for instance, Akitoby and Stratmann (2008), Bellas et al. (2010), Edwards (1984), Eichengreen and Mody (2000) and Hilscher and Nosbusch (2010), among others.

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

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.

7.1 The Empirical Model

Following Edwards (1984), we assume that the spread over U.S. (or Germany) 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

$$y_{it} = \phi_i y_{it-1} + x_{it}' \beta_i + \mu_i + \varepsilon_{it} \tag{58}$$

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. The panel data model in matrix notation can be written as in (2) where all the coefficients are random and follow (3). When doing parameter equality tests and comparing the EM-REML to alternative methods, we set $f_{1i} = 1$ for all $i = 1, .., N$, to then extend the analysis to the case where f_{1i} is a $l \times 1$ vector of unit-specific explanatory variables.

7.2 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 latter differ across units. This problem does not arise in the static case, where pooling estimation give unbiased estimates of coefficient means when they differ randomly. We then show that the random coefficients variances also 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.

7.2.1 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):

$$F = \frac{1}{(N-1)} \sum_{i=1}^{N} F_i \sim F\left(K^*(N-1), (\sum_{i=1}^{N} T_i - NK^*) \right) \tag{59}$$

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.[25] Because our test has a value of 5.1852, we are able to reject the null of homogenous slope and intercept parameters.

7.2.2 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

$$LR = 2 \left[logL(\hat{\theta}) - logL\left(\hat{\theta}_r\right) \right] \tag{60}$$

where $\hat{\theta}$ is the unrestricted MLE, obtained estimating the model by EM-REML algorithm under the assumption of heteroskedastic variances, i.e. $\gamma_i \sim IN(0, \triangle_i)$. On the other hand,

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

$\hat{\theta}_r$ is the restricted MLE obtained from the EM-REML estimation under the assumption of homoskedastic variances, i.e. $\gamma_i \sim IN(0, \triangle)$, $\forall i$. When $\triangle_i = \triangle$ for all i, the iterations illustrated in Section 4.5 still hold. Equation (35) has to be replaced by (52).

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.[26]

7.3 Comparison

We now compare the results obtained estimating (58) 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 Table 1.[27]

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.

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).

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

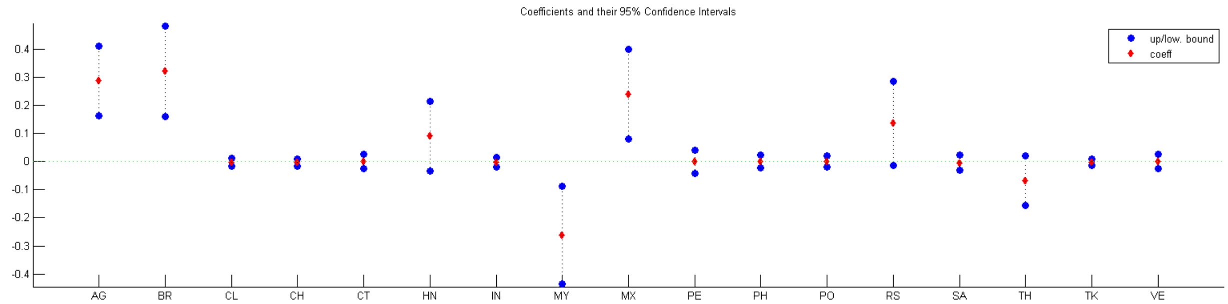
[26]When expressing the coefficients as function of the explanatory variables, the LR test has a value of 107.1300 . We reject the null of homoskedasticity at the 2% level.

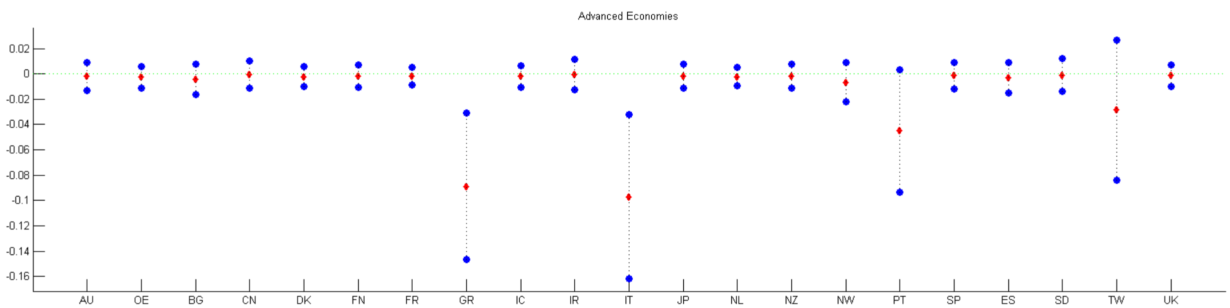
[27]It is known that the main drawback of the EM algorithm is its slow rate of convergence. However, in this particular application the rate of convergence is pretty fast, less than 15 seconds.

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. 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.

7.4 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.[28] One could argue, that the latter two countries have benefited from joining the 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.[29]

7.5 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

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

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

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 in Table 2.

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. "% y Dom Def" ("% y Ext Def") denotes the percentage of year in default or restructuring domestic (external) debt; ϕ_i is the autoregressive coefficient; $\beta_i^{(k)}$ is the sensitivity of spread to the kth variable.

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 spread. 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.[30] 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

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

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 Inflation					-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. "% y Curr Crisis" ("% y Infl Crisis") denote the percentage of years with annual inflation of 20% or higher and with an annual depreciation vs US dollar of 15% or more, respectively. "% y Dom Def" ("% y Ext Def") denotes the percentage of year in default or restructuring domestic (external) debt.

used as a measure of sudden increases in debt's level. In Table 3, we focus on the coefficients equation corresponding to the sensitivity of spreads to debt and report results from using different specifications.[31] Standard deviations over the sample period under considerations 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. Interestingly, domestic defaults have a larger economic impact than external ones. Our finding that domestic defaults play a significant role in explaining changes in the sensitivity of

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

spreads corroborates Reinhart and Rogoff (2010) argument: ”when ignored domestic debt obligations are taken into account, fiscal duress at the time of default is often revealed to be quite severe”. Second, country-specific macroeconomic indicators do not play any significant role in explaining the reactions of financial markets 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[32], we find no evidence that such variables affect 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. On the contrary, reputation in financial markets is crucial.

8 Conclusion

We show how to combine the EM algorithm with the restricted maximum likelihood (REML) approach to estimate a general class of dynamic heterogeneous panels. We also review the existing sampling and Bayesian methods commonly used to estimate random coefficients panel data models, to highlight similarities and differences with the EM-REML approach. Our method allows us to estimate iteratively both fixed and random coefficients, as well as the variance components. Among its interesting features, it belongs to the class of shrinkage estimators and it gives a probability distribution over the random coefficients residuals. Compared to existing methods, the EM-REML algorithm yields an unbiased estimator of the variance components. This is important, especially when T is small, given that both the estimator of the average effect and the predictor of the random coefficients residuals are a function of the variance components. Similarly, over- or under-estimating the latter, affects the estimated standard errors and may undermine any hypothesis testing. Monte Carlo experiments confirm that our approach has good properties 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. Second, the proposed method allows for heteroskedastic random coefficients and thus offers 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

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

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. Our findings indicate that countries who have defaulted in the past may find it difficult to finance government 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. As a consequence, their debt-to-GDP ratios remain considerably lower than those of several high-income countries. The unanswered question is how to escape such a "trap".

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
$\parallel Bias\,(\psi) \parallel$	**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
$\parallel Bias\,(var\,(\gamma_1)) \parallel$	0.247	0.234	0.303	0.098	0.476	0.633	3.967	9.375	14.976	5.716	16.276	19.694
$\parallel Bias\,(var\,(\gamma_2)) \parallel$	0.252	0.196	0.274	0.056	0.273	0.367	0.998	1.619	2.514	1.222	1.754	3.166
$\parallel Bias\,(var\,(\gamma_3)) \parallel$	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
% Negative Definite							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
$\parallel Bias\,(\psi) \parallel$	**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
$\parallel Bias\,(var\,(\gamma_1)) \parallel$	0.102	0.140	0.137	0.251	0.395	0.438	0.720	1.139	0.974	0.564	1.356	1.299
$\parallel Bias\,(var\,(\gamma_2)) \parallel$	0.082	0.184	0.244	0.137	0.224	0.243	0.171	0.300	0.309	0.133	0.262	0.320
$\parallel Bias\,(var\,(\gamma_3)) \parallel$	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
% Negative Definite							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
$\parallel Bias\,(\psi) \parallel$	**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
$\parallel Bias\,(var\,(\gamma_1)) \parallel$	0.083	0.039	0.235	0.433	0.152	0.704	0.521	0.190	0.677	0.207	0.101	0.413
$\parallel Bias\,(var\,(\gamma_2)) \parallel$	0.178	0.074	0.167	0.237	0.082	0.396	0.243	0.089	0.396	0.281	0.095	0.443
$\parallel Bias\,(var\,(\gamma_3)) \parallel$	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
% Negative Definite							0.740	0.320	0.000			

9 Appendix

9.1 Matrix Computations for REML

9.1.1 A Choice for S_i

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

$$M_i = I - W_i \left(W_i' W_i\right)^{-1} W_i'. \tag{61}$$

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

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

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

where U_i is an orthogonal matrix. Searle and Quaas (1978) defines A_i to be the first $T-p-\bar{K}$ columns of U_i'. It follows that $M_i = A_i A_i'$ and $A_i' A_i = I$. Premultiplying M_i by A, we get

$$M_i A_i = A_i, \qquad A_i' M_i = A_i' \ . \tag{62}$$

Since U' is orthogonal and non-singular, A_i' has full rank and $A_i' W_i = 0$. As stated in Searle and Quaas (1978), using (62), it can be shown that $A_i \left(A_i' R_i A_i\right)^{-1} A_i'$ is the Moore-Penrose inverse of $M_i R_i M_i$:

$$\left(M_i R_i M_i\right)^{+} = A_i \left(A_i' R_i A_i\right)^{-1} A_i'. \tag{63}$$

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

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

$$P_i = R_i^{-1} - R_i^{-1} W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1}, \tag{64}$$

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

$$\begin{aligned} R_i P_i &= I - W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i' R_i^{-1}, \\ P_i R_i &= I - R_i^{-1} W_i \left(W_i' R_i^{-1} W_i\right)^{-1} W_i'. \end{aligned} \tag{65}$$

Therefore,

$$P_i R_i P_i = P_i, \tag{66}$$

and also $(P_i R_i)^2 = P_i R_i$. It follows that $tr(P_i R_i) = r(P_i R_i) = r(P_i) = T - p - \bar{K}$.

Relationship between M_i and P_i. Using (61) and the fact that $P_iW_i = 0$, it can be seen that

$$P_iM_i = P_i = M_iP_i. \tag{67}$$

Furthermore, post-multiplying (65) by M_i and using $M_iW_i = 0$ and $W_i'M_i = 0$, we get $P_iR_iM_i = M_i$. Post-multiplying (67) by R_iM_i

$$P_iM_iR_iM_i = P_iR_iM_i = M_iP_iR_iM_i = M_i^2 = M_i. \tag{68}$$

From (67) and (68), Searle (1978) establishes P_i as the Moore-Penrose inverse of $M_iR_iM_i$:

$$P_i = (M_iR_iM_i)^+. \tag{69}$$

Since $(M_iR_iM_i)^+$ is unique, (63) and (69) imply that

$$P_i = (M_iR_iM_i)^+ = A_i\left(A_i'R_iA_i\right)^{-1}A_i'. \tag{70}$$

9.1.2 Some Lemmas from Searle and Quaas (1978)

Lemma 1. Searle and Quaas (1978) shows that $S_i = F_i'A_i'$ for some non-singular F_i'. It follows that

$$\begin{aligned} S_i'\left(S_iR_iS_i'\right)^{-1}S_i &= A_iF_i\left(F_i'A_i'R_iA_iF_i\right)^{-1}F_i'A_i' \\ &= A_i\left(A_i'R_iA_i\right)^{-1}A_i \quad = \quad P_i. \end{aligned} \tag{71}$$

where the last equality follows from (70).

Lemma 2. As shown in Lutkepohl (1996, pag. 50 eq. 6), if $A = (m\times m)$, $B = (m\times n)$,$C = (n\times m)$ and D is a $(n\times n)$ matrix, then

$$det\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{aligned} &= \mid D\mid\cdot\mid A - BD^{-1}C\mid \quad if \quad D\,nonsingular \\ &= \mid A\mid\cdot\mid D - CA^{-1}B\mid \quad if \quad A\,nonsingular \end{aligned}. \tag{72}$$

Using this property of the determinant, we can show that

$$\mid A_iR_iA_i'\mid = \frac{\mid R_i\mid\cdot\mid W_i'R_i^{-1}W_i\mid}{\mid W_i'W_i\mid}. \tag{73}$$

To prove the latter, let

$$\begin{bmatrix} A_i' \\ W_i' \end{bmatrix} R_i \begin{bmatrix} A_i & W_i \end{bmatrix} = \begin{bmatrix} A_i'R_iA_i & A_i'R_iW_i \\ W_i'R_iA_i & W_i'R_iW_i \end{bmatrix}.$$

Taking the determinant of both sides, we get

$$|R_i| \cdot \begin{vmatrix} A_i'A_i & A_i'W_i \\ W_i'A_i & W_i'W_i \end{vmatrix} = |A_i'R_iA_i| \cdot |W_i'R_iW_i - W_i'R_iA_i\left(A_i'R_iA_i\right)^{-1}A_i'R_iW_i|\,.$$

Using $A_i'A_i = I$ and $A_i'W_i = 0$ and equation (70), we get

$$|R_i||W_i'W_i| = |A_i'R_iA_i| \cdot |W_i'R_iW_i - W_i'R_iPR_iW_i|\,.$$

Substituting (65) into the latter equation and then using the property of determinants, $det(AB) = det(A) \cdot det(B)$, yields (73).

Lemma 3. Given that $S_i = F_i'A_i$, it can be shown that

$$|S_iR_iS_i'| = |F|^2|A'R_iA|\,. \tag{74}$$

9.1.3 Finding an expression for L_{1i}

Using (73) and (74), we have

$$log\,|S_iR_iS_i'| = c + log|R_i| + log\,|W_i'R_i^{-1}W_i|, \tag{75}$$

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

Furthermore, using (71), we get

$$\begin{aligned}(y_i - Z_i\gamma_i)'S_i'\left(S_iR_iS_i'\right)^{-1}S_i\left(y_i - Z_i\gamma_i\right) &= (y_i - Z_i\gamma_i)'P_i\left(y_i - Z_i\gamma_i\right) \\ &= \left(y_i - W_i\hat{\bar{\Gamma}} - Z_i\gamma_i\right)'R_i^{-1}\left(y_i - W_i\hat{\bar{\Gamma}} - Z_i\gamma_i\right).\end{aligned} \tag{76}$$

Substituting (75) and (76) into (19) yields (20).

Proof of Equation (76). Let $\hat{\bar{\mathbf{\Gamma}}}$ be the argument that minimizes $\varepsilon_i'R_i^{-1}\varepsilon_i$, where $\varepsilon_i = y_i - W_i\bar{\Gamma} - \bar{Z}_i\gamma_i$ and $R_i = var\left(\varepsilon_i\right)$.[33] The solution to the problem is given by

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

It follows that

$$\begin{aligned}y_i - W_i\hat{\bar{\Gamma}} - \bar{Z}_i\gamma_i &= y_i - W_i\left(W_i'R_i^{-1}W_i\right)^{-1}W_i'R_i^{-1}\left(y_i - \bar{Z}_i\gamma_i\right) - \bar{Z}_i\gamma_i \\ &= R_iP_iy_i - R_iP_iZ_i\gamma_i.\end{aligned}$$

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

Therefore, using (66) and after a few computations, we get

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

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

$$\begin{aligned}cov\left(S_i y_i, Q_i y_i \mid \gamma_i\right) &= E\left(S_i y_i y_i' Q_i' \mid \gamma_i\right) - E\left(S_i y_i \mid \gamma_i\right) E\left(y_i' Q_i' \mid \gamma_i\right)\\ &= S_i E\left(y_i y_i' \mid \gamma_i\right) Q_i' - \left(S_i \bar{Z}_i\gamma_i\right)\left(\bar{\Gamma}' W_i' + \gamma_i'\bar{Z}_i'\right) R_i^{-1} W_i,\end{aligned} \tag{77}$$

where $E\left(S_i y_i \mid \gamma_i\right) = S_i\bar{Z}_i\gamma_i$ since $S_i W_i = 0$.
Substituting

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

and

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

into (77), we get

$$\begin{aligned}cov\left(S_i y_i, Q_i y_i \mid \gamma_i\right) &= -S_i\bar{Z}_i\gamma_i\bar{\Gamma}' W_i' R_i^{-1} W_i\\ &\quad - S_i\bar{Z}_i\gamma_i\gamma_i'\bar{Z}_i' R_i^{-1} W_i.\end{aligned} \tag{78}$$

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

$$\begin{aligned}cov\left(S_i y_i, Q_i y_i\right) &= E\left[cov\left(S_i y_i, Q_i y_i \mid \gamma_i\right)\right]\\ &\quad + cov\left(E\left(S_i y_i \mid \gamma_i\right), E\left(Q_i y_i \mid \gamma_i\right)\right).\end{aligned} \tag{79}$$

Taking expectation of both sides of (78), we get

$$E\left[cov\left(S_i y_i, Q_i y_i \mid \gamma_i\right)\right] = -S_i\bar{Z}_i\triangle_i\bar{Z}_i' R_i^{-1} W_i, \tag{80}$$

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

$$\begin{aligned}cov\left(E\left(S_i y_i \mid \gamma_i\right), E\left(Q_i y_i \mid \gamma_i\right)\right) &= E\left[S_i\bar{Z}_i\gamma_i\left(W_i' R_i^{-1} W_i\bar{\Gamma} + W_i' R_i^{-1}\bar{Z}_i\gamma_i\right)'\right]\\ &\quad - E\left[E\left(S_i y_i\right)\right] E\left[E\left(Q_i y_i\right)'\right]\\ &= S_i\bar{Z}_i\triangle_i\bar{Z}_i' R_i^{-1} W_i.\end{aligned} \tag{81}$$

Therefore, substituting (80) and (81) into (79) we can show that $cov\left(S_i y_i, Q_i y_i\right) = 0$.

9.2 BLUP

Conditional Mean and Variance. Under the assumption that y_i and γ_i are jointly normally distributed, the conditional expectation of γ_i given the data is

$$\begin{aligned} \hat{\gamma}_i = E\left(\gamma_i \mid y_i\right) &= E(\gamma_i) + cov(\gamma_i, y_i)\left[var(y_i)\right]^{-1}\left[y_i - E(y_i)\right] \\ &= c' V_i^{-1}\left(y_i - W_i\bar{\Gamma}\right), \end{aligned} \tag{82}$$

where $E(\gamma_i) = 0$, by assumption, $E(y_i) = W_i\bar{\Gamma}$, $V_i = var(y_i) = \bar{Z}_i \triangle_i \bar{Z}_i' + R_i$ and $c' = cov(\gamma_i, y_i) = \triangle_i \bar{Z}_i'$. The conditional variance of γ_i is

$$\begin{aligned} var\left(\gamma_i \mid y_i\right) &= var(\gamma_i) - cov(\gamma_i, y_i)\left[var(y_i)\right]^{-1} \cdot cov(y_i, \gamma_i) \\ &= var(\gamma_i) - c' V_y^{-1} c. \end{aligned} \tag{83}$$

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

$$\begin{aligned} \triangle_i \bar{Z}_i'\left[\bar{Z}_i \triangle_i \bar{Z}_i' + R_i\right]^{-1} &= \left\{\left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle_i^{-1}\right)^{-1}\left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle_i^{-1}\right)\right\} \cdot \\ &\quad \cdot \triangle_i \bar{Z}_i'\left[\bar{Z}_i \triangle_i \bar{Z}_i' + R_i\right]^{-1} \\ &= \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle_i^{-1}\right)^{-1} \cdot \bar{Z}_i' R_i^{-1}. \end{aligned} \tag{84}$$

These results is used in the second equality in equation (25) and substituted into (83) yields (26).

Properties. Henderson (1984, Chap. 5), showed that:
(i) the BLP is unbiased:

$$\begin{aligned} E\left(\hat{\gamma}_i\right) &= E\left[c' V_y^{-1}(y_i - W_i\bar{\Gamma})\right] \\ &= c' V_y^{-1}\left[E(y_i) - W_i\bar{\Gamma}\right] = E(\gamma_i), \end{aligned} \tag{85}$$

since $E(y_i) = W_i\bar{\Gamma}$.
(ii) $var\left(\hat{\gamma}_i\right) = var\left[c' V_i^{-1}(y_i - W_i\bar{\Gamma})\right] = c' V_i^{-1} c$
(iii) $cov\left(\hat{\gamma}_i, \gamma_i\right) = var\left(\hat{\gamma}_i\right)$, from which it follows that $var\left(\hat{\gamma}_i - \gamma_i\right) = var\left(\gamma_i\right) - var\left(\hat{\gamma}_i\right)$.
(iv) the BLUP maximizes the correlation between $\hat{\gamma}_i$ and γ_i.

9.3 E-Step

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

$$E_{\theta^{(b-1)}}\left(\varepsilon_i' H_i \varepsilon_i \mid y_i\right) = Tr\left[H_i E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right)\right]. \tag{86}$$

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

$$E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right) = V_{\varepsilon_i} + \hat{\varepsilon}_i \hat{\varepsilon}_i', \tag{87}$$

where

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

and

$$\begin{aligned} V_{\varepsilon_i} &= var\left(\varepsilon_i \mid y_i; \theta^{(b-1)}\right) &=& var\left(y_i - W_i\bar{\Gamma} - \bar{Z}_i\gamma_i \mid y_i, \theta^{(b-1)}\right) \\ & &=& \bar{Z}_i V_{\gamma_i}^{(b)} \bar{Z}_i', \end{aligned} \tag{88}$$

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 (87) into (86)

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

We can now write

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

Using a similar expedient, we can obtain Q_{1i} and Q_{3i}.

9.4 Estimation of $\triangle_i$

An estimator of $\triangle_i$ can be obtained by maximizing (29) 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 $(m \times 1)$:

$$\frac{\partial a' X^{-1} b}{\partial X} = -(X^{-1})' a b' (X^{-1})'. \tag{89}$$

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

$$\frac{\partial tr\left(A X^{-1} B\right)}{\partial X} = -\left(X^{-1} B A X^{-1}\right)'. \tag{90}$$

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

$$\frac{\partial ln \mid X \mid}{\partial X} = (X')^{-1}. \tag{91}$$

It follows that

$$\frac{\partial Q_{3i}}{\partial \triangle_i} = \underset{(91)}{-\triangle_i^{-1}} + \underset{(90)}{\triangle_i^{-1} V_{\gamma_i}^{(b)} \triangle_i^{-1}} + \underset{(89)}{\triangle_i^{-1} \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} \triangle_i^{-1}} = 0,$$

which implies that

$$\triangle_i^{(b)} = V_{\gamma_i}^{(b)} + \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime}. \tag{92}$$

Unbiased Estimator. It can be shown that

$$\hat{\triangle}_i = \hat{\gamma}_i \hat{\gamma}_i' + V_{\gamma_i}, \tag{93}$$

is an unbiased estimator of $\triangle_i$ since

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

9.5 Standard Errors

The variance-covariance matrix of (40) is given by

$$\begin{aligned} var\left(\tilde{\psi}_{1i} - \psi_{1i}\right) &= F_{1i} var\left(\hat{\bar{\Gamma}}\right) F_{1i}' + var\left(\hat{\gamma}_i - \gamma_i\right) + F_{1i} cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \hat{\gamma}_i - \gamma_i\right) \\ &\quad + \left[F_{1i} cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \hat{\gamma}_i - \gamma_i\right)\right]', \end{aligned} \tag{94}$$

where

$$\begin{aligned} cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \hat{\gamma}_i - \gamma_i\right) &= cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \hat{\gamma}_i\right) - cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \gamma_i\right) \\ &= -\Xi W_i' V_i^{-1} \bar{Z}_i \triangle_i, \end{aligned}$$

since $cov\left(\hat{\bar{\Gamma}}, \hat{\gamma}_i\right) = 0$ and $cov\left(y_i, \hat{\bar{\Gamma}}\right) = var\left(\hat{\bar{\Gamma}}\right) W_i'$.

References

[1] B. Akitoby and T. Stratmann. Fiscal policy and financial markets. *The Economic Journal*, 118(533):1971–1985, 2008.

[2] T. Amemiya and W. A. Fuller. A comparative study of alternative estimators in a distributed lag model. *Econometrica, Journal of the Econometric Society*, pages 509–529, 1967.

[3] T. W. Anderson and C. Hsiao. Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375):598–606, 1981.

[4] T. W. Anderson and C. Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of econometrics*, 18(1):47–82, 1982.

[5] B. Baltagi. *Econometric analysis of panel data*. John Wiley & Sons, 2008.

[6] A. Chudik and M. H. Pesaran. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2):393–420, 2015.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[8] J. Eaton and M. Gersovitz. Debt with potential repudiation: Theoretical and empirical analysis. *The Review of Economic Studies*, 48(2):289–309, 1981.

[9] S. Edwards. Ldc's foreign borrowing and default risk: An empirical investigation, 1983.

[10] B. Eichengreen and A. Mody. What explains changing spreads on emerging-market debt: fundamentals or market sentiment? Technical report, National Bureau of Economic Research, 1998.

[11] R. J. Friedrich. In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, pages 797–833, 1982.

[12] A.-M. Fuertes and R. Smith. *Panel Time Series*. CEMMAP, 2016.

[13] J. D. Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

[14] U. N. Haque, M. H. Pesaran, and S. Sharma. Neglected heterogeneity and dynamics in cross-country savings regressions. 1999.

[15] C. R. Harvey and Y. Liu. Rethinking performance evaluation. Technical report, National Bureau of Economic Research, 2016.

[16] D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338, 1977.

[17] F. Hayashi. Econometrics. 2000. *Princeton University Press. Section*, 1:60–69, 2000.

[18] C. R. Henderson. *Applications of Linear Models in Animal Breeding*, volume 1. 1983.

[19] J. Hilscher and Y. Nosbusch. Determinants of sovereign risk: Macroeconomic fundamentals and the pricing of sovereign debt. *Review of Finance*, page rfq005, 2010.

[20] C. Hsiao. *Analysis of panel data*. Number 54. Cambridge university press, 2014.

[21] C. Hsiao, T. W. Appelbe, and C. R. Dineen. A general framework for panel data models with an application to canadian customer-dialed long distance telephone service. *Journal of Econometrics*, 59(1-2):63–86, 1993.

[22] C. Hsiao and M. H. Pesaran. Random coefficient panel data models. 2004.

[23] C. Hsiao, M. H. Pesaran, A. K. Tahmiscioglu, et al. Bayes estimation of short-run coefficients in dynamic panel data models, 1998.

[24] R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996.

[25] G. Koop, D. J. Poirier, and J. L. Tobias. *Bayesian econometric methods*. Cambridge University Press, 2007.

[26] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[27] L.-F. Lee and W. E. Griffiths. *The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models*. Citeseer, 1979.

[28] D. V. Lindley and A. F. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–41, 1972.

[29] H. Lutkepohl. Handbook of matrices. *Computational Statistics and Data Analysis*, 2(25):243, 1997.

[30] G. Maddala. Generalized least squares with an estimated variance covariance matrix. *Econometrica: Journal of the Econometric Society*, pages 23–33, 1971.

[31] G. S. Maddala, R. P. Trost, H. Li, and F. Joutz. Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics*, 15(1):90–100, 1997.

[32] P. Mazodier and A. Trognon. Heteroscedasticity and stratification in error components models. In *Annales de l'INSEE*, pages 451–482. JSTOR, 1978.

[33] G. McLachlan and T. Krishnan. *The EM algorithm and extensions*, volume 382. John Wiley & Sons, 2007.

[34] A. Mian and A. Sufi. *House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again.* University of Chicago Press, 2015.

[35] Y. Mundlak. On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, pages 69–85, 1978.

[36] A. Pagan. Two stage and related estimators and their applications. *The Review of Economic Studies*, 53(4):517–538, 1986.

[37] H. D. Patterson and R. Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554, 1971.

[38] Y. Pawitan. *In all likelihood: statistical modelling and inference using likelihood.* Oxford University Press, 2001.

[39] H. Pesaran, R. Smith, and K. S. Im. Dynamic linear models for heterogenous panels. In *The econometrics of panel data*, pages 145–195. Springer, 1996.

[40] M. H. Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

[41] M. H. Pesaran and R. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, 68(1):79–113, 1995.

[42] I. Petrova, M. M. G. Papaioannou, and M. D. Bellas. *Determinants of emerging market sovereign bond spreads: fundamentals vs financial stress.* Number 10-281. International Monetary Fund, 2010.

[43] W. C. Randolph. A transformation for heteroscedastic error components regression models. *Economics Letters*, 27(4):349–354, 1988.

[44] C. M. Reinhart. This time is different chartbook: country histories on debt, default, and financial crises. Technical report, National Bureau of Economic Research, 2010.

[45] C. M. Reinhart and K. S. Rogoff. *This time is different: eight centuries of financial folly*. princeton university press, 2009.

[46] C. M. Reinhart and K. S. Rogoff. From financial crash to debt crisis. *The American Economic Review*, 101(5):1676–1706, 2011.

[47] C. M. Reinhart, K. S. Rogoff, and M. A. Savastano. Debt intolerance. Technical report, National Bureau of Economic Research, 2003.

[48] N. Roy. Is adaptive estimation useful for panel models with heteroskedasticity in the individual specific error component? some monte carlo evidence. *Econometric Reviews*, 21(2):189–203, 2002.

[49] S. Searle, R. Quaas, et al. A notebook on variance components: A detailed description of recent methods of estimating variance components, with applications in animal breeding. 1978.

[50] A. F. Smith. A general bayesian linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 67–75, 1973.

[51] P. A. Swamy. Efficient inference in a random coefficient regression model. *Econometrica: Journal of the Econometric Society*, pages 311–323, 1970.

[52] P. A. V. B. Swamy. *Statistical inference in random coefficient regression models*, volume 55. Springer Science & Business Media, 2012.