Panel time series analysis : some theory and applications

Panel Time Series Analysis: Some Theory and Applications

Andrea Nocera

Department of Economics, Mathematics and Statistics

Birkbeck, University of London

A thesis submitted for the degree of

Doctor of Philosophy

June 2017

Declaration

I wish to declare that no part of this doctoral dissertation contains material previously submitted to the University of London or to any other institution for any degree. The fourth chapter of this thesis, titled " House Prices and Monetary Policy in the Euro Area: Evidence from Structural VARs", is a joint work with Mr. Moreno Roma. I certify that I am the lead author of the research, having designed and carried out the empirical analysis, which constitutes the bulk of the chapter.

Andrea Nocera

London, United Kingdom

28 June 2017

Abstract

This thesis offers some theoretical contributions to the literature on large heterogeneous panel data models. It also demonstrates their practical use in empirical research, in the field of housing in macroeconomics, and for the analysis of the determinants of sovereign credit spreads.

The first chapter provides the motivation for the research presented in this thesis.

In the second chapter, we investigate the causes and the finite-sample consequences of negative definite covariance matrices in Swamy type random coefficient models. Monte Carlo experiments reveal that the negative definiteness problem is less severe when the degree of coefficient dispersion is substantial, and the precision of the regression disturbances is high. The sample size also plays a crucial role. We then evaluate the direct consequences of relying on the asymptotic properties of the estimator of the random coefficient covariance for hypothesis tests.

A solution to the aforementioned problem is proposed in the third chapter. In particular, we propose to implement the EM algorithm to compute restricted maximum likelihood estimates of both the average effects and the unit-specific coefficients as well as of the variance components in a wide class of heterogeneous panel data models. Compared to existing methods, our approach leads to unbiased and more efficient estimation of the variance components of the model without running into the problem of negative definite covariance matrices typically encountered in random coefficient models. This in turn leads to more accurate

estimated standard errors and hypothesis tests. Monte Carlo simulations reveal that the proposed estimator has relatively good finite sample properties. In evaluating the merits of our estimator, we also provide an overview of the sampling and Bayesian methods commonly used to estimate heterogeneous panel data. A novel approach to investigate heterogeneity of the sensitivity of sovereign spreads to government debt is presented.

In a final chapter, we use a structural Bayesian (stochastic search variable selection) vector autoregressive model to investigate the heterogeneous impact of housing demand shocks on the macro-economy and the role of house prices in the monetary policy transmission, across euro area countries. A novel set of identification restrictions, which combines zero and sign restrictions, is proposed. By exploiting the cross-sectional dimension of our data, we explore the differences in the propagation channels of house prices and monetary policy and the challenges they pose in the process of real and nominal convergence in the Eurozone. Among the main results, we find a comparatively stronger housing wealth effect on consumption in Ireland and Spain. We provide new evidence in support of the financial accelerator hypothesis, showing that house prices play an important role in the availability of loans. A significant and highly heterogeneous effect of monetary policy on house price dynamics is also documented.

Acknowledgements

My adventure at Birkbeck started in September 2012. Willing to expand my knowledge of econometrics, I enrolled to a postgraduate degree on that topic. It was there that I had the great pleasure of meeting Haris Psaradakis, who was to become my PhD supervisor at the same institution, the year after. I consider myself very fortunate. I am very grateful to him for his encouragements and his constant help, above and beyond the call of duty. I have also been very lucky to have Ron Smith as supervisor. I have truly enjoyed our discussions on panel data and not only; seeing his office door open has always been an irresistible temptation. His guidance has proved extremely helpful. I wish one day to be as helpful and inspiring to the future generations as Haris and Ron were to me.

Ivan Petrella has also played an important role during my years at Birkbeck. I would like to sincerely thank him for his sharing his passion for research, and for his numerous and invaluable advices throughout the years. Completing a PhD thesis can be a long process, during which it is inevitable to encounter many exceptional people. A big thank to the PhD students, Adiya, Angelos, Lasse, Marco, José, Paul, Rubens, and many others, with whom I have shared hope and "despair", and to all members of the staff in our Department.

Recently, I had the pleasure of visiting the USC Dornsife INET, in Los Angeles. I am very grateful to Hashem Pesaran for the kind and warm hospitality received. I feel honoured to have the opportunity to join his centre, as postdoctoral research fellow, and I can't wait to start. Part of the fourth chapter has been written during my traineeship at the European Central Bank, Prices and Costs Division. I would like to thank the staff members for their warm hospitality. I would also like to thank George Kapetanios and Takashi Yamagata for kindly accepting to examine this thesis. Financial support from the Economic and Social Research Council is gratefully acknowledged.

Any occasion is a good one to thank all my dear friends, Anastasia, Andrea, Dimitri, Doriana, Giovanni and Giovanni, Grzegorz, Ido, and Marco, among others.

Finally, I would like to express my strong appreciation to my parents, Alfonso and Ninetta, for their unconditional and constant love and support. Dulcis in fundo, my dearest Natalia. Everything becomes much easier with her by my side.

This thesis is dedicated to Natalia and to my parents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Contributions

Panel data models have become increasingly popular in empirical studies. In fact, by combining a number of observations on a cross-section of units (individuals, countries, or firms, to name a few) over repeated time periods, they provide a number of advantages over a single cross-section, or a single time series. Hsiao (2003) and Baltagi (2005) offer a comprehensive list of benefits from using panel data. Among them, the availability of larger data sets which increases the degrees of freedom, may alleviate multicollinearity and hence improve efficiency, and the ability to study dynamics of adjustment. Another important advantage, to which we pay particular attention in this thesis, is the ability to control for individual heterogeneity. A notable example is Baltagi and Levin (1992). They study cigarette demand across 46 American states by modelling consumption as a function of its own lag, price and income. They note that panel data are able to control for state-invariant (e.g. advertising on TV) and time-invariant variables (such as religion), even though some of them may not be available or measurable. This in turn avoids the omission bias in the resulting estimates.

Traditionally, in panel data with large number of cross-section units (N) and few time periods (T), the effects of unobserved time-invariant heterogeneity and omitted variables have been controlled by allowing for individual-specific intercepts and/or time-specific effects. However, in many economic applications, it is unlikely that the response of a dependent variable to a change in an explanatory variable is the same for all units and/or over time. As T increases, it is possible to test for equality of parameters, and the homogeneity hypothesis is often rejected in practice.[1] Accounting for heterogeneity may help to shed new lights and to better understanding some real economic phenomena. For instance, Eberhardt and Teal (2010) emphasize the importance of parameter heterogeneity in the growth empiric literature. Haque, Pesaran and Sharma (2000) investigate the implications of neglected slope heterogeneity for the fixed effects estimator. 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 rates. At the same time, one can obtain highly significant, but spurious, nonlinear effects for some of the potential determinants, although the country-specific regressions are linear.

The treatment of heterogeneity is particularly important in the context of dynamic models. Pesaran and Smith (1995) show that when the regression coefficients vary across individuals, pooling and aggregating in a dynamic model give inconsistent and misleading estimates of the coefficients. The inconsistency of both fixed and random effects does not disappear even when both T and N go to infinity. Therefore, they argue in favour of heterogeneous estimators and propose the so called Mean Group estimation which yields a consistent estimator of the average effects as both T and N gets large.

Another popular approach which allows for coefficient heterogeneity is the Swamy (1970) random coefficient model. It can be seen as a generalization of the random effects model, since it considers both the intercept and the slope parameters as realizations from a certain probability distribution with common

[1] Different tests for slope homogeneity have been proposed. See for instance, Pesaran and Yamagata (2008).

mean and constant variance-covariance matrix. In view of this assumption, it is quite natural to put the random coefficient model in a Bayesian framework. Bayesian estimation is discussed in Maddala, Trost, Li and Joutz (1997), and Hsiao, Pesaran and Tahmiscioglu (1999), among others.

The above mentioned techniques can be labelled as large heterogeneous panel data models. Excellent surveys are provided by Hsiao and Pesaran (2008), Pesaran (2016), and Smith and Fuertes (2016). This thesis offers some theoretical contributions to this literature. It also contains two novel applications in the context of heterogeneous panels.

First, we study the problem of negative definite covariance matrices in Swamy-type random coefficient models. As in the error-component model, the unbiased estimator of the random coefficient covariance matrix proposed by Swamy (1970) is not necessarily nonnegative definite. This is often the case in empirical applications. Despite being a well acknowledged problem, its causes are not yet fully understood. We perform a Monte Carlo study to address this issue. We show that the problem is particularly severe when the precision of the regression disturbances and the degree of coefficient heterogeneity are low, and/or when the sample size is small. To overcome the negative definiteness problem, Swamy (1971) suggests an alternative estimator of the random coefficient covariance matrix which, although biased, is nonnegative definite and is consistent when the time dimension tends to infinity. We demonstrate that relying on the asymptotic properties of this estimator may lead to poor inference. Unless the time and cross-section dimensions, and/or the degree of coefficient dispersion are high, the estimated standard errors are largely upwards biased. The resulting hypothesis tests may suffer from considerable size distortions. The empirical sizes of the tests are substantially lower than the nominal levels.

A solution to the aforementioned problems is provided in a separate chapter. We show that applying the EM algorithm to obtain restricted maximum likelihood estimates yields an unbiased and more efficient estimator of the random coefficient covariance matrix without running into the problem of negative definiteness. This in turn leads to more accurate standard errors and hypothesis tests. It is also demonstrated that direct maximization of the likelihood which incorporates the prior likelihood of the random coefficients yields an estimator of the coefficients' covariance matrix which does not satisfy the law of total variance. This is not the case when employing the EM algorithm.

Since the seminal work of Dempster, Laird, and Rubin (1977), the EM algorithm has been successfully applied in different contexts, such as linear mixed models (Laird and Ware, 1982), finite mixture models (McLachlanand Peel, 2000), and factor analysis (Engle and Watson, (1983), Quah and Sargent (1993), Doz, Giannone, and Reichlin (2012), Harvey and Liu (2016)), to mention a few. A full-fledged book on the subject is McLachlan and Krishnan (2008). We highlight the relative merits of the EM approach in estimating both average and unit-specific coefficients in heterogeneous panels. In doing so, we also review the existing sampling and Bayesian methods. To extend the applicability of our method, we consider a general framework which incorporates various panel data models as special case, including the random coefficient and the correlated random effects models. Monte Carlo simulations reveal that the obtained (restricted) maximum likelihood estimators have relatively good finite sample properties, in terms of bias, root mean square errors, and power of tests.

An important issue in large panels, which has received particular attention in recent years, is cross-section dependence, i.e. the correlation between errors in different units. The literature is quite vast (see for instance, Holly, Pesaran, and Yamagata (2010), Chudik and Pesaran (2013), Bailey, Kapetanios and Pesaran (2015)) and its analysis is beyond the scope of this thesis. We simply note that our estimation procedure can be adapted to allow for cross-section dependence, following Pesaran (2006), and Chudik and Pesaran (2015).

The methods described above can be quite effective in modelling complex economic relationships. Therefore, part of this thesis is devoted to their application to the analysis of sovereign credit risk and to the role of house price dynamics in the macroeconomy.

In the first application, we show that modelling the random coefficients as a function of selected explanatory variables can be beneficial to the study of the determinants of the sensitivity of sovereign spreads with respect to government debt. It is widely known that macroeconomic fundamentals and volatility are significant drivers of sovereign credit spreads (Akitoby and Stratmann (2008), Hilscher and Nosbusch (2010), among others). On the contrary, there is no study, to the best of our knowledge, which investigates why the response of sovereign spreads to changes in government debt differs significantly across countries. We show that country-specific macroeconomic indicators 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 years in default or restructuring domestic debt is associated with around 0.35% increase in the additional risk premium in response to an increase in debt.

Another important aspect in applied research in economics is the aggregation problem. The implications of aggregation are well known in the econometrics literature (e.g. Granger (1987), Pesaran (2003), and Pesaran and Chudik (2014)). Nevertheless, some of the issues which arise when aggregating time series are sometimes ignored in the applied literature. For example, most of the recent studies which derive insights on the role of the housing market in the Eurozone from multivariate structural models focus on the euro area as a whole. A prominent example is Musso, Neri, and Stracca (2011). However, it is important from a policy perspective to quantify and compare the heterogeneous impacts of house prices across countries as they can amplify the existing economic divergences across Eurozone member. Moreover, as a common monetary policy only reacts to area wide aggregates such as inflation and economic activity, it is crucial to understand what are the consequences in terms of house price dynamics in each country in order to properly address real and financial imbalances at the country level by means of macroprudential policies. This motivates the last part of our research. We use a structural Bayesian stochastic search variable selection vector autoregression for seven euro-area countries (Belgium, France, Germany, Ireland, Italy, the Netherlands, and Spain) for the period 1980:Q1- 2014:Q4 to provide a systematic structural analysis of the effects of housing demand shocks on economic activity and the role of house prices in the monetary policy transmission. A novel set of identification restrictions, which combines zero and sign restrictions, is proposed. We focus on a country by country analysis, given the idiosyncratic characteristics of the housing market in the euro area, which suggest that pooling or aggregating may lead to biased inference and misleading policy recommendations. At the same time, we exploit the cross-sectional dimension of our data, to compare and quantify the degree of heterogeneity of the effects of housing demand and monetary policy shocks across euro area members. In doing so we fill a gap in the literature, largely focused on the US, the UK and the euro area as a whole. Among the main results, we find a comparatively stronger housing wealth effect on consumption in Ireland and Spain, countries having recently experienced a boom-bust pattern in house prices. We provide new evidence in support of the financial accelerator hypothesis, showing that house prices play an important role in the availability of loans. A significant and highly heterogeneous effect of monetary policy on house price dynamics is also documented.

1.2 Outline of the Thesis

This thesis is organised as follows. In Chapter 2, we study the causes of negative definite covariance matrices in Swamy type random coefficient models. We perform Monte Carlo simulations to disentangle the drivers of the problem, and to investigate the finite-sample consequences for hypothesis tests.. A solution is proposed in Chapter 3. We show how to implement the EM algorithm to compute iteratively restricted maximum likelihood (REML) estimates of both fixed and random coefficients, as well as the variance components, in a wide class of heterogeneous panels. We then review some of the existing sampling and Bayesian methods commonly used to estimate heterogeneous panel data, to highlight similarities and differences with the EM-REML approach. Monte-Carlo experiments are employed to examine and compare the finite sample properties of our method, the Swamy random coefficient model, and the Mean Group estimation. Finally, the proposed econometric methodology is used to study the determinants of the sensitivity of sovereign spreads with respect to government debt. In Chapter 4 we use a structural Bayesian stochastic search variable selection VAR model to study the differences in the propagation channels of house prices and monetary policy in the Eurozone. We compare and document significant and highly heterogeneous effects of housing demand shocks on the macro-economy and of monetary policy on house price dynamics across euro area countries. Chapter 5 summarizes and concludes.

Chapter 2

Causes and Effects of Negative Definite Covariance Matrices in Swamy Type Random Coefficient Models

2.1 Introduction

For panel data studies with large N, the number of units, and small T, the time dimension, it is common to assume homogeneity of the slope coefficients. Individual-specific intercepts are the only source of heterogeneity. However, in many economic applications, it is more realistic to allow the response parameters to differ across cross-sectional units. As T increases, it is possible to test for equality of parameters, and the homogeneity hypothesis is very often rejected. Two popular methods which deal with coefficient heterogeneity are the Mean Group estimation, proposed by Pesaran and Smith (1995), and the Swamy (1970) random coefficient model. Both methods require estimating N time series separately. The latter models the regression coefficients as random variables with a certain probability distribution. To reduce the number of parameters to be estimated, it is assumed that the coefficients have constant means and variance-covariances.

Unfortunately, as in the error-component model, the estimator of the random coefficient covariance matrix is not necessarily nonnegative definite. This is often the case in empirical applications. Despite being a well acknowledged problem, its causes are not yet fully understood. In this chapter, we disentangle the drivers of the problem by means of Monte Carlo simulations. Another contribution of this chapter is to examine the finite-sample properties of Swamy's generalized least squares (GLS) estimator in terms of accuracy of inference, when a consistent but biased estimator of the random coefficient covariance is used to overcome the negative definiteness problem.

The Monte Carlo analysis confirms that the negative definiteness problem of this estimator increases with the variance of the regression time-varying disturbances, and it is negatively (and statistically significantly) correlated to the degree of coefficient heterogeneity. The probability of the estimator being negative definite goes much faster to zero following an increase in the level of coefficient dispersion rather than a raise in the precision of the regression disturbances. The problem is also more severe when T and/or N are small, partly due to the fact that the performances of individual OLS and the Mean Group estimators worsen in small samples. As expected, when T goes to infinity, the second term of the estimator goes to zero, and the problem of negative definiteness vanishes.

Whenever the unbiased estimator of the random coefficient covariance is negative definite, Swamy suggests eliminating a term to obtain an estimator which is nonnegative definite and is consistent when T tends to infinity. However, we show that the latter can be severely biased in small samples. We then investigate the finite-sample consequences for hypothesis tests. We find that the resulting estimated standard errors are very often upwards biased. In many cases, this bias can be substantial. This in turn leads to size distorted hypothesis tests, with exact sizes well below the nominal levels.

The remainder of the Chapter is organized as follows. Section 2.2 reviews the random coefficient model. Section 2.3 discusses the derivation of the Swamy estimator of the random coefficient covariance matrix. Monte Carlo experiments are implemented in Section 2.4, where we present the results from regressing the probability of the estimator being negative definite on a number of explanatory variables, and comment on the finite-sample performances of the estimator of interest for inference. The last section concludes.

2.2 The Random Coefficient Model

Consider the following linear regression model

$$y_i = X_i\beta_i + u_i, \qquad i = 1, .., N, \tag{2.1}$$

where $y_i = (y_{i1}, y_{i2}, ..., y_{iT})'$ is a $T \times 1$ vector of observations for the dependent variable, and X_i is a $T \times K$ matrix of strictly exogenous explanatory variables, including a vector of ones to allow for an intercept. The Swamy (1970) random coefficient model treats both intercept and slope coefficients

$$\beta_i = \beta + \delta_i \tag{2.2}$$

as random with common mean β. It is assumed that

$$E(\delta_i) = 0, \qquad E\left(\delta_i\delta_j'\right) = \begin{cases} \triangle & if \quad i = j, \\ 0 & if \quad i \neq j, \end{cases} \tag{2.3}$$

$$E(u_i) = 0, \qquad E\left(u_i u_j'\right) = \begin{cases} \sigma_i^2 I_T & if \quad i = j, \\ 0 & if \quad i \neq j, \end{cases} \tag{2.4}$$

Finally, β_i and u_j are independent for all i and j.

2.2.1 Estimation

Under the above assumptions, the best linear unbiased estimator of β is the generalized least squares (GLS) estimator

$$\begin{aligned}\hat{\beta}_{GLS} &= \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' V_i^{-1} y_i\right) \\ &= \sum_{i=1}^{N} W_i \hat{\beta}_i,\end{aligned} \tag{2.5}$$

where

$$W_i = \left\{\sum_{i=1}^{N} \left[\triangle + \sigma_i^2 (X_i' X_i)^{-1}\right]^{-1}\right\}^{-1} \left[\triangle + \sigma_i^2 (X_i' X_i)^{-1}\right]^{-1},$$

$$\hat{\beta}_i = (X_i' X_i)^{-1} X_i' y_i,$$

and $V_i = X_i \triangle X_i' + \sigma_i^2 I_T$. The GLS estimator is equivalent to the weighted average of the OLS estimates, with weights inversely proportional to their covariance matrices. The variance-covariance matrix of (2.5) is

$$var\left(\hat{\beta}_{GLS}\right) = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i\right)^{-1}. \tag{2.6}$$

As noted by Swamy, if we assume normality of both u_i and β_i, it can be easily shown that the variance of the GLS estimator is equal to the Cramer-Rao lower bound. Therefore, (2.5) is a minimum variance estimator within the class of all unbiased estimators.

However, the GLS estimator for β is infeasible since it depends on the unknown variances σ_i^2 and $\triangle$. Swamy uses the OLS estimators, $\hat{\beta}_i$, and their residuals $\hat{u}_i = y_i - X_i \hat{\beta}_i$, to obtain unbiased estimators of σ_i^2 and $\triangle$,

$$\hat{\sigma}_i^2 = \frac{\hat{u}_i' \hat{u}_i}{T-K}, \tag{2.7}$$

$$\hat{\triangle} = \hat{\triangle}_1 - \hat{\triangle}_2, \tag{2.8}$$

where

$$\begin{aligned} \hat{\triangle}_1 &= \tfrac{1}{N-1} \textstyle\sum_{i=1}^{N} \left(\hat{\beta}_i - N^{-1} \sum_{i=1}^{N} \hat{\beta}_i\right) \left(\hat{\beta}_i - N^{-1} \sum_{i=1}^{N} \hat{\beta}_i\right)', \\ \hat{\triangle}_2 &= N^{-1} \textstyle\sum_{i=1}^{N} \hat{\sigma}_i^2 (X_i' X_i)^{-1}. \end{aligned} \tag{2.9}$$

The second term $\left(-\hat{\triangle}_2\right)$ is necessary for $\hat{\triangle}$ to be an unbiased estimator of $\triangle$. Unfortunately, as in the error-component model, the estimator (2.8) is not necessarily nonnegative definite. As a solution, Swamy suggested using $\hat{\triangle}_1$ as an estimator of $\triangle$. Although biased, this estimator is positive semi-definite and consistent when T tends to infinity. Note that as T gets large, the second term, $\hat{\triangle}_2$, converges in probability to zero.

2.3 The Estimator of the Random Coefficient Covariance Matrix

In this section, we describe the derivation of (2.8) in some detail. We start by noting that the OLS estimator of β_i can be rewritten as

$$\begin{aligned} \hat{\beta}_i &= \beta_i + (X_i' X_i)^{-1} X_i' u_i \\ &= \beta + \delta_i + (X_i' X_i)^{-1} X_i' u_i. \end{aligned} \tag{2.10}$$

Its unconditional and conditional expectations are given by

$$\begin{aligned} E\left(\hat{\beta}_i\right) &= \beta, \\ E\left(\hat{\beta}_i \mid \delta_i\right) &= \beta_i, \end{aligned}$$

respectively. Using equation (2.10), we can compute the variance of the OLS estimator:

$$\begin{aligned} var\left(\hat{\beta}_i\right) &= E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)' \\ &= E\left[(\beta_i-\beta)+\left(\hat{\beta}_i-\beta_i\right)\right]\left[(\beta_i-\beta)+\left(\hat{\beta}_i-\beta_i\right)\right]', \end{aligned} \tag{2.11}$$

where $(\beta_i-\beta)=\delta_i$, and $\left(\hat{\beta}_i-\beta_i\right)=(X_i'X_i)^{-1}X_i'u_i$. Using equations (2.3) and (2.4), and assuming that $E\left(u_i \mid X_i,\delta_i\right)=0$, we get

$$\begin{aligned} var\left(\hat{\beta}_i\right) &= E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)' = E\left(\beta_i-\beta\right)\left(\beta_i-\beta\right)'+E\left(\hat{\beta}_i-\beta_i\right)\left(\hat{\beta}_i-\beta_i\right)' \\ &= \triangle+\sigma_i^2(X_i'X_i)^{-1}. \end{aligned} \tag{2.12}$$

Equation (2.12) states that, for an unbiased estimator where $E\left(\hat{\beta}_i \mid \beta_i\right)=\beta_i$ and $E\left(\hat{\beta}_i\right)=\beta$, the variance of $\hat{\beta}_i$ around β is equal to the variance of β_i around β plus the variance of $\hat{\beta}_i$ around β_i.

The estimator of $\triangle$ given by (2.8), can be obtained by replacing $var\left(\hat{\beta}_i\right)$ with its sample analogue, and $\sigma_i^2(X_i'X_i)^{-1}$ with its estimator averaged across units.

From equation (2.12), it follows that

$$\triangle = E\left(\beta_i-\beta\right)\left(\beta_i-\beta\right)' = \triangle_1-\triangle_2, \tag{2.13}$$

where

$$\begin{aligned} \triangle_1 &= E\left(\hat{\beta}_i-\beta\right)\left(\hat{\beta}_i-\beta\right)', \\ \triangle_2 &= E_{\hat{\beta}_i|\beta_i}\left[\hat{\beta}_i-E\left(\hat{\beta}_i \mid \beta_i\right)\right]\left[\hat{\beta}_i-E\left(\hat{\beta}_i \mid \beta_i\right)\right]'. \end{aligned} \tag{2.14}$$

It can be noted that $\triangle$ is positive semi-definite by definition. Indeed,

$$\left\{\begin{array}{lcl} \hat{\beta}_i - \beta & = & \delta_i + (X_i'X_i)^{-1}X_i'u_i \\ \\ \hat{\beta}_i - \beta_i & = & (X_i'X_i)^{-1}X_i'u_i \end{array}\right. \Longrightarrow E\left(\hat{\beta}_i - \beta\right)\left(\hat{\beta}_i - \beta\right)' \geq E\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_i - \beta_i\right)', \tag{2.15}$$

where $\beta_i = E\left(\hat{\beta}_i \mid \beta_i\right)$, and the inequality sign denotes matrix inequalities. The equality would hold only if $\delta_i = 0$, $\forall i$, which means that the coefficients do not vary across units, i.e. $E\left(\delta_i\delta_i'\right) = 0$, for all i.

It can also be noted that (2.12) satisfies the law of total variance since

$$\begin{aligned} var\left(\hat{\beta}_i\right) &= var\left[E\left(\hat{\beta}_i \mid \beta_i\right)\right] + E\left[var\left(\hat{\beta}_i \mid \beta_i\right)\right] \\ &= var\left(\beta_i\right) + E\left[var\left(\hat{\beta}_i \mid \beta_i\right)\right], \end{aligned}$$

where $var\left(\beta_i\right) = \triangle$, and $var\left(\hat{\beta}_i \mid \beta_i\right) = \sigma_i^2(X_i'X_i)^{-1}$. This implies that $var\left(\hat{\beta}_i\right) \geq \triangle$, and $var\left(\hat{\beta}_i\right) \geq \sigma_i^2(X_i'X_i)^{-1}$, which corroborates (2.15).[1]

2.3.1 Nonspherical Errors

Equation (2.12) has been derived under the assumption that $var\left(u_i\right) = \sigma_i^2 I_T$. If $var\left(u_i\right) = \Omega_i$, where Ω_i is a symmetric and positive definite $T \times T$ matrix, then

$$V_i = X_i \triangle X_i' + \Omega_i,$$

and

$$\begin{aligned} E\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_i - \beta_i\right)' &= E\left((X_i'X_i)^{-1}X_i'u_iu_i'X_i(X_i'X_i)^{-1}\right) \\ &= \left(X_i'X_i\right)^{-1}\left(X_i'\Omega_iX_i\right)\left(X_i'X_i\right)^{-1}. \end{aligned}$$

[1] Henceforth, we use inequality signs to denote matrix inequalities.

Equation (2.11) becomes

$$var\left(\hat{\beta}_i\right) = \triangle + \left(X_i'X_i\right)^{-1}\left(X_i'\Omega_i X_i\right)\left(X_i'X_i\right)^{-1}. \tag{2.16}$$

Therefore, an unbiased estimator of $\triangle$ is

$$\hat{\triangle} = \hat{\triangle}_1 - \hat{\triangle}_3, \tag{2.17}$$

where $\hat{\triangle}_1$ is defined in (2.9), and

$$\hat{\triangle}_3 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)^{-1}\left(X_i'\hat{\Omega}_i X_i\right)\left(X_i'X_i\right)^{-1}. \tag{2.18}$$

In many cases, $\hat{\triangle}_3 \geq \hat{\triangle}_2$, which may exacerbate the negative definiteness problem of $\hat{\triangle}$. When the elements of u_i are negatively autoccorelated, and the K regressors in x_{it} are positively autocorrelated, Goldeberger (1964, pp. 238-42) showed that the diagonal elements of $\hat{\triangle}_3$ can be smaller than the corresponding diagonal elements of $\hat{\triangle}_2$.

Alternatively, as shown in Appendix 2.6.1, one can estimate each time series by applying Aitken's GLS, yielding

$$\tilde{\beta}_i = (X_i'\hat{\Omega}_i^{-1}X_i)^{-1}X_i'\hat{\Omega}_i^{-1}y_i,$$

$$\tilde{\sigma}_i^2 = \frac{\tilde{u}_i'\tilde{u}_i}{T-K},$$

where $\tilde{u}_i$ are the GLS residuals. In such case, the estimator of $\triangle$ becomes

$$\hat{\triangle} = \hat{\triangle}_4 - \hat{\triangle}_5, \tag{2.19}$$

where

$$\begin{aligned}\hat{\triangle}_4 &= \tfrac{1}{N-1}\sum_{i=1}^{N}\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^{N}\tilde{\beta}_i\right)\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^{N}\tilde{\beta}_i\right)', \\ \hat{\triangle}_5 &= \tfrac{1}{N}\sum_{i=1}^{N}\tilde{\sigma}_i^2\left(X_i'\hat{\Omega}_i X_i\right)^{-1},\end{aligned} \tag{2.20}$$

It is reasonable to expect $\hat{\triangle}_5$ to be smaller than $\hat{\triangle}_3$.

One may suspect that taking serial correlation into account, and using (2.19) as an estimator of $\hat{\triangle}$ reduces the probability of $\hat{\triangle}$ being negative definite.[2] For instance, Swamy (1971) indicates mispecification of either the model or the underlying assumptions as possible reasons of the negative definiteness problem. Nevertheless, as shown in the Monte Carlo analysis, $\hat{\triangle}$ can be often negative definite even though the true disturbances are not correlated over time and the model is correctly specified, suggesting that the causes of the problem lie elsewhere.

2.4 Monte Carlo Analysis

Given that $\triangle$ is positive semi-definite by construction, why is (2.8) often negative semi-definite? What goes wrong when replacing the true components with their analogue estimates? In other words, why is $\hat{\triangle}_1$, the estimator of $E\left(\hat{\beta}_i - \beta\right)\left(\hat{\beta}_i - \beta\right)'$, often less than $\hat{\triangle}_2$, the estimator of $E\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_i - \beta_i\right)'$? We address this question by performing a Monte Carlo analysis.

[2]It should be noted that although $\hat{\triangle}_5$ might be smaller than $\hat{\triangle}_3$, $\hat{\triangle}_1$ has to be replaced by $\hat{\triangle}_4$ in estimating $\hat{\triangle}$. As for (2.8), there is no guarantee that (2.19) is nonnegative definite.

2.4.1 The Data Generating Process

The data generating process used to simulate the data is given by

$$\begin{aligned} y_{it} &= c_i + x_{it}\beta_i + \varepsilon_{it}, \\ x_{it} &= c_{x,i}(1-\rho) + \rho x_{it-1} + u_{it}, \end{aligned} \tag{2.21}$$

where

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

We set $\rho = 0.6$, and $x_{i0} = 0$, $\forall i$. Once generated, the x_{it}'s are taken as fixed across different replications.[3] The variances of the time-varying disturbances are generated according to:

$$\begin{aligned} (i) \quad & \sigma_i^2 \sim unif\,[0.1, 0.9], \\ (ii) \quad & \sigma_i^2 \sim unif\,[0.5, 1.5], \\ (iii) \quad & \sigma_i^2 \sim unif\,[1, 3], \\ (iv) \quad & \sigma_i^2 \sim unif\,[3, 5], \end{aligned}$$

such that $E(\sigma_i^2) \in \{0.5, 1, 2, 4\}$. To allow for the presence of outliers, we also consider the following case

$$(v) \quad \sigma_i^2 \sim \varphi \cdot unif\,[0.5, 1.5] + (1-\varphi) \cdot unif\,[4, 6],$$

where φ is binary variable whose distribution is Bernoulli:

$$\varphi = \begin{cases} 1 & p = 0.75 \\ 0 & (1-p). \end{cases}$$

[3] To minimize the effect of initial observations we discard the first 100 observations.

In the latter case, $E(\sigma_i^2) = 2$ as in case (iii) although the variance of most of the units varies between 0.5 and 1.5. In all cases, the σ_i^2 's are sorted so that $\sigma_i^2 > \sigma_j^2$ if $\bar{x}_i > \bar{x}_j$, where $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$. The coefficients differ randomly across units according to

$$\begin{aligned} c_i &= c + \sigma_c \gamma_{1i}, \\ \beta_i &= \beta + \sigma_\beta \gamma_{2i}, \end{aligned}$$

where $\gamma_{ji} \sim i.i.d.N(0,1)$, for $j = 1, 2$. We consider the following options:

Option	1	2	3	4	5	6	7
c	0	0	0	0.5	0.5	0.5	1
β	0.1	0.5	1	0.1	0.5	1	1

For each option, we draw the random effects, γ_{ji}, from a Normal distribution with different degrees of coefficient heterogeneity (from low (1) to high (6)):[4]

Degree of Heterogeneity	1	2	3	4	5
σ_c	0.05	0.1	0.3	0.5	1
σ_β	0.05	0.1	0.3	0.5	1

We generate $G = (n_O \cdot n_H) \cdot n_V = (7 \cdot 5) \cdot 5 = 175$ clusters, where n_O, n_H, and n_V denote the number of options, the number of coefficient heterogeneity cases, and the different specifications for σ_i^2, respectively. Each cluster is of size $S = (n_T \cdot n_N) = (6 \cdot 4) = 24$, where each unit in the cluster consists of the pair (T_j, N_l), with $T_j \in \{10, 20, 30, 50, 70, 140\}$, and $N_l \in \{10, 30, 50, 140\}$. In total, we run $M = (n_T \cdot n_N)(n_O \cdot n_H \cdot n_V) = 24 \cdot 175 = 4200$ different data generating processes (DGP). Within each DGP we run $H = 1500$ iterations.[5]

[4]It should be noted that when generating 1000 observations from $\beta_i \sim N(0.5, 1)$, the range of values that β_i assumed was -3 to 3.4. This is a very high level of dispersion, which we consider for theoretical reasons. If the degree of heterogeneity were so high in real applications, it might be difficult to reconcile the estimates with economic theory.

[5]The time required to run the 1500 iterations is approximately 5 to 30 seconds depending on the sample size. The time necessary to estimate each option is approximately 2 hours, which makes the results replicable.

Degree of Coefficient Heterogeneity. The choice of σ_c and σ_β is in line with Trapani and Urga (2009), and Boyd and Smith (2002). The former review some empirical works which use heterogenous estimators and derive a measure to determine the level of coefficient heterogeneity (the standard deviation of the random coefficients). They find that the levels of heterogeneity obtained using the datasets of Baltagi, Griffin and Xiong (2000) and Baltagi, Bresson, Griffin and Pirotte (2003) are equal to 0.176 and 0.183 respectively. Higher levels are found in Baltagi and Griffin (1997) and Brucker and Silivertovs (2006), where the degree of heterogeneity is equal to 0.323 and 0.428 respectively.

Boyd and Smith (2002) review some econometric issues in estimating models of the transmission mechanism of monetary policy, for 57 developing countries, where $T = 31$. They find a high degree of dispersion of the estimates across countries.[6] For instance, in an inflation persistence equation, the average coefficient on the first lag of inflation is 0.57 with a standard deviation of 0.30. In a static Purchasing Power Parity equation of log spot on log price differential, the mean is 1.13 and the standard deviation of the estimates is 0.52.

2.4.2 Descriptive Statistics

Table 2.1 reports the Monte Carlo estimates of $\varrho = Pr\left(\hat{\triangle} < 0\right)$, the probability that the estimator of $\triangle$ (defined in (2.8), and averaged across the 7 different options) is negative definite, across different sample sizes and different combinations of coefficient and data dispersions, σ_β and $E\left(\sigma_i\right)$ respectively. A few important facts emerge from this simple descriptive analysis:

1. The probability ϱ is a decreasing function of both T and N. However, when T and σ_β are moderate, the probability of $\hat{\triangle}$ being negative definite can still be high even when N is as large as 140.

[6] After estimating the regression coefficients, β_i, for each unit, Boyd and Smith compute the number of standard deviations from the mean as $Z\left(\beta\right) = \left(\hat{\beta}_i - \bar{\beta}\right)/s\left(\hat{\beta}_i\right)$, where $\bar{\beta} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_i$, and $s^2\left(\hat{\beta}_i\right) = \left(N-1\right)^{-1}\sum_{i=1}^{N}\left(\hat{\beta}_i - \bar{\beta}\right)^2$.

Table 2.1: The probability of $\hat{\triangle}$ being negative definite

		$E(\sigma_i^2) = 2$ (iii)					$E(\sigma_i^2) = 2$ (v)					$E(\sigma_i^2) = 1$				
T	$N \setminus \sigma_\beta$	0.05	0.1	0.3	0.5	1	0.05	0.1	0.3	0.5	1	0.05	0.1	0.3	0.5	1
10	10	88	86	73	57	33	88	86	75	57	41	87	84	66	44	11
	30	**84**	**83**	**63**	**43**	**8**	**85**	**83**	**63**	**54**	**19**	**84**	**80**	**51**	**27**	**1**
	50	84	80	61	36	5	84	80	58	45	17	82	77	43	15	0
	140	81	77	46	24	0	81	76	49	37	3	79	71	32	4	0
20	10	88	85	64	37	7	89	85	61	41	9	87	80	45	18	1
	30	**84**	**79**	**46**	**13**	**0**	**84**	**75**	**51**	**25**	**2**	**80**	**68**	**20**	**2**	**0**
	50	82	72	33	7	0	83	74	44	21	0	79	64	11	0	0
	140	78	64	15	1	0	79	63	33	5	0	73	52	2	0	0
30	10	88	82	49	20	2	89	81	52	33	4	86	77	33	11	0
	30	**84**	**71**	**24**	**2**	**0**	**85**	**70**	**39**	**11**	**0**	**79**	**61**	**6**	**0**	**0**
	50	81	66	18	1	0	79	67	32	5	0	74	54	2	0	0
	140	75	54	3	0	0	75	56	14	0	0	66	39	0	0	0
50	10	87	78	34	9	1	88	78	40	7	1	83	68	17	3	0
	30	**80**	**62**	**10**	**0**	**0**	**80**	**65**	**20**	**1**	**0**	**74**	**49**	**1**	**0**	**0**
	50	75	56	3	0	0	76	57	9	0	0	66	42	0	0	0
	140	67	41	0	0	0	66	49	2	0	0	56	25	0	0	0
140	10	80	62	6	0	0	80	66	12	1	0	73	49	2	0	0
	30	**66**	**39**	**0**	**0**	**0**	**67**	**48**	**1**	**0**	**0**	**54**	**18**	**0**	**0**	**0**
	50	60	28	0	0	0	61	39	0	0	0	46	10	0	0	0
	140	46	11	0	0	0	52	27	0	0	0	31	1	0	0	0

The probability (in percentage) of the estimator of $\triangle$ being negative definite (averaged across the 7 options) across the time dimension (T), the cross-section dimension (N), different degrees of coefficient heterogeneity (σ_β), and the mean of the variance of the time-varying regression disturbances ($E(\sigma_i^2)$), for $i = 1, .., N$. The results shown in columns (iii) and (v) differ as in the former $\sigma_i^2 \sim unif\,[1, 3]$. In the latter, $\sigma_i^2 \sim \varphi \cdot unif\,[0.5, 1.5] + (1 - \varphi) \cdot unif\,[4, 6]$.

2. ϱ can be quite high when σ_β is small or moderate. If $\sigma_\beta = 0.05$, the value of ϱ can be substantial even when $T = 140$ and N is also large. On the contrary, if $\sigma_\beta = 1$, ϱ is almost always equal to zero as soon as T is larger than 20.

3. The variance of the time-varying disturbances also plays an important role. Indeed, for a given degree of coefficient heterogeneity, as σ_i^2 increases, the second term of (2.8) raises. Consequently, the probability that the estimator of the random coefficient covariance matrix is negative definite increases.

4. Whether ϱ is large or small depends on the value of σ_β relative to the σ_i's, the standard deviations of the time-varying regression disturbances. This means that even though σ_β is high, ϱ can be still far from zero if $E\left(\sigma_i^2\right)$ is very large.

2.4.3 Regression Analysis

To corroborates the findings of the theoretical analysis and the insights emerged in the descriptive analysis, we run the following cross-section regression:

$$y_m = \alpha + z_m'\theta + u_m, \qquad m = 1,..,M,$$

where $M = 4200$. The depedent variable y_m measures the probability of $\hat{\triangle}$ being negative definite within each DGP, and it is computed as

$$y_m = Pr\left(\hat{\triangle} < 0\right) = \frac{\sum_{h=1}^{H} \rrbracket\left(\hat{\triangle}^{(h)} < 0\right)}{H},$$

where $\rrbracket$ is a binary indicator that takes the value 1 if $\hat{\triangle}^{(h)} < 0$ (in a matrix sense) and 0 otherwise. The vector z_m may include the following explanatory variables:

- the time dimension, T, and the number of units, N,
- the values of the intercept (c) and slope parameter (β) in (2.21),
- the degree of coefficient heterogeneity, $\sigma_c = \sigma_\beta$,
- the average standard deviation of the regression disturbances, $\bar{\sigma} = N^{-1}\sum_{i=1}^{N}\sigma_i$,
- a measure of the signal-to-noise ratio, $\sigma_\beta/\bar{\sigma}$,

- the bias of the Mean Group estimator of $\psi = (c, \beta)'$,

- the cross-section averages of the absolute value of the biases of the OLS estimators:
$$\frac{1}{N}\sum_{i=1}^{N}\left|\left(\frac{1}{H}\sum_{h=1}^{H}\hat{\psi}_i^{(h)}\right)-\psi\right|,$$

- the trace of the root mean square errors (RMSE) of the Mean Group estimator,

- the trace of the RMSE of the OLS estimators, averaged across units:

$$AvRMSE\left(\hat{\psi}_i\right)=\frac{1}{N}\sum_{i=1}^{N}\left\{\sqrt{\frac{1}{H}\sum_{h=1}^{H}\left(\hat{\psi}_i^{(h)}-\psi_i^{(h)}\right)\left(\hat{\psi}_i^{(h)}-\psi_i^{(h)}\right)'}\right\}.$$

We estimate the model by OLS. Results are shown in Table 2.2. In parenthesis, we report the t-tests computed using White (1980) heteroskedasticity-robust standard errors.[7]

Main Findings. In the simplest specification (1), we regress our dependent variables on a constant, the time dimension (T), the number of units (N), the degree of coefficient heterogeneity (σ_β), and the average of the time-varying regression disturbances' standard deviations ($\bar{\sigma}$). We then include the value of c and β used in equation (2.21) to simulate the data. As expected, the constant, which is approximatively equal to 70%, is statistically significant. One standard deviation increase of σ_β statistically significantly reduces the probability of $\hat{\triangle}$ being negative definite (ϱ) of around 70%. The conditional variability of the data is also a significant predictor: one standard deviation increase of $\bar{\sigma}$ is associated with a statistically significant increase in the dependent variable of 5%.

[7]When calculating the robust standard errors, we make the adjustment for degrees of freedom suggested by MacKinnon and White (1985).

Table 2.2: The drivers of the random coefficient covariance's negative definiteness problem

$Pr\left(\hat{\triangle} < 0\right)$	(1)	(2)	(3)	(4)	(5)	(6)
constant	**0.696**	**0.695**	**0.744**	**0.562**	**0.521**	**0.536**
	(82.77)	(71.27)	(77.33)	(58.52)	(46.54)	45.031
T	**-0.002**	**-0.002**	**-0.002**	**-0.001**	**-0.001**	**-0.001**
	(-32.04)	(-32.03)	(-28.99)	(-12.85)	(-9.38)	-11.596
N	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**	**-0.001**
	(-21.52)	(-21.52)	(-19.09)	(-23.73)	(-8.41)	-10.312
σ_β	**-0.697**	**-0.697**		**-1.160**	**-0.802**	**-1.218**
	(-91.69)	(-91.66)		(-58.12)	(-71.60)	-45.356
$\bar{\sigma}$	**0.052**	**0.052**		**0.018**	**0.015**	**0.020**
	(21.59)	(21.60)		(7.31)	(5.31)	7.318
$\sigma_\beta/\bar{\sigma}$			**-0.639**			
			(-49.98)			
c		0.003	0.002			0.004
		(0.31)	(0.18)			0.477
β		0.001	0.001			-0.002
		(0.11)	(0.152)			-0.333
$bias\left(\hat{c}_{mg}\right)$				-0.198		-0.083
				(-0.32)		-0.132
$bias\left(\hat{\beta}_{mg}\right)$				0.239		0.347
				(0.30)		0.449
$Av\left(\left|bias\left(\hat{c}_{i,ols}\right)\right|\right)$				**13.099**		**12.346**
				(17.26)		12.215
$Av\left(\left|bias\left(\hat{\beta}_{i,ols}\right)\right|\right)$				**14.246**		**13.366**
				(12.09)		11.058
$RMSE\left(\hat{\psi}_{MG}\right)$					**0.373**	**0.301**
					(14.78)	10.777
$RMSE\left(\hat{\psi}_{i,ols}\right)$					**0.163**	**-0.044**
					(16.90)	-2.863
R^2	0.675	0.675	0.576	0.728	0.713	0.734
Theil Adj. R^2	0.675	0.675	0.576	0.728	0.713	0.733

We regress the probability of $\hat{\triangle}$ being negative definite on a number of explanatory variables. The values of the OLS estimators and their corresponding t-ratios (in parentheses) are reported. We use White (1980) heteroskedasticity-robust standard errors with the adjustment for degrees of freedom suggested by MacKinnon and White (1985). Bold values denotes statistical significance at 5% level or lower.

At the same time, an one unit increase in T and N causes a 0.2% and 0.1% decrease of ϱ, respectively. On the contrary, the coefficients associated with the value of the constant and intercept parameters (c and β) are not statistically significant. These findings are consistent across all other specifications.

In a third regression (3), we replace σ_β and $\bar{\sigma}$ with a measure of the signal-to-noise ratio $(\sigma_\beta/\bar{\sigma})$.[8] An one standard deviation increase of the latter statistically significantly decreases the probability of $\hat{\triangle}$ being negative definite by 64%. The R-squared is smaller in the third specification, suggesting that including both σ_β and $\bar{\sigma}$ separately improves the goodness of fits.

Given that $\hat{\triangle}$, described in equation (2.8), is a plug-in estimator, we also test whether the finite sample performances (in terms of bias and RMSE) of both the Mean Group estimator of c and β, and the OLS estimators of the unit-specific regression coefficients affect the probability of $\hat{\triangle}$ being negative definite. The regression analyses (4) to (6) corroborate this hypothesis. For instance, a 1% increase in the cross-section averages of the absolute value of the biases of the OLS estimates raises ϱ of around 12 to 14%.

2.4.4 Finite-Sample Consequences

As shown in Table 2.1, the unbiased estimator of the random coefficient covariance matrix defined in equation (2.8), is likely to be negative definite in many circumstances. This is often the case in many empirical applications. To overcome the problem, Swamy (1971) suggests replacing this estimator by $\hat{\triangle}_1$, defined in equation (2.9). The latter is nonnegative definite and is consistent when T tends to infinity. However, as reported in Table 2.3 , it can be severely biased in small samples.

[8] We have also considered other measures of signal-to-noise ratio: $N^{-1}\sum_{i=1}^{N}\left(\sigma_\beta^2/\sigma_{\varepsilon_i}^2\right)$, $N^{-1}\sum_{i=1}^{N}(\sigma_\beta/\sigma_{\varepsilon_i})$, and $\left(\sigma_\beta^2/\bar{\sigma}^2\right)$, where $\bar{\sigma}^2 = N^{-1}\sum_{i=1}^{N}\sigma_{\varepsilon_i}^2$. They yield very similar result. Therefore, we only report results obtained using $(\sigma_\beta/\bar{\sigma})$, with $\bar{\sigma} = N^{-1}\sum_{i=1}^{N}\sigma_{\varepsilon_i}$ as the corresponding regression coefficient has larger economic value and it is associated with a larger t-ratio. Both the R^2 and the Theil's adjusted R^2 are also relatively larger in the latter case.

Table 2.3: Bias and root mean square errors of $\hat{\triangle}_1$

		$\sigma_\beta = 0.05$				$\sigma_\beta = 0.1$				$\sigma_\beta = 0.3$				$\sigma_\beta = 0.5$			
	$T \backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$bias\{\hat{\sigma}_c\}$	10	0.80	0.81	0.72	0.65	0.70	0.84	0.61	0.67	0.40	0.44	0.45	0.53	0.23	0.51	0.37	0.39
	20	0.39	0.34	0.41	0.38	0.32	0.35	0.32	0.34	0.17	0.19	0.21	0.19	0.13	0.15	0.13	0.14
	30	0.28	0.25	0.30	0.28	0.22	0.27	0.23	0.24	0.14	0.12	0.13	0.15	0.10	0.11	0.08	0.10
	50	0.21	0.22	0.20	0.19	0.18	0.18	0.15	0.17	0.06	0.10	0.08	0.09	0.06	0.05	0.05	0.05
	70	0.17	0.18	0.15	0.16	0.12	0.12	0.12	0.13	0.05	0.07	0.06	0.06	0.02	0.04	0.03	0.04
	140	0.10	0.10	0.11	0.10	0.08	0.07	0.07	0.07	0.01	0.03	0.03	0.03	0.00	0.01	0.01	0.02
$bias\{\hat{\sigma}_\beta\}$	10	0.39	0.33	0.31	0.32	0.35	0.33	0.22	0.27	0.14	0.15	0.18	0.19	0.06	0.13	0.13	0.11
	20	0.20	0.15	0.18	0.17	0.14	0.13	0.15	0.14	0.08	0.07	0.07	0.07	0.02	0.04	0.04	0.04
	30	0.12	0.13	0.12	0.12	0.08	0.10	0.09	0.09	0.03	0.04	0.04	0.04	0.03	0.03	0.02	0.03
	50	0.08	0.08	0.08	0.08	0.05	0.05	0.06	0.06	0.02	0.02	0.02	0.02	0.00	0.01	0.01	0.01
	70	0.07	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.01	0.01	0.01	0.02	0.00	0.00	0.01	0.01
	140	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00
$RMSE\{\hat{\sigma}_c\}$	10	0.84	0.83	0.74	0.66	0.74	0.87	0.62	0.67	0.45	0.46	0.46	0.54	0.29	0.54	0.39	0.40
	20	0.41	0.35	0.42	0.39	0.35	0.36	0.33	0.35	0.21	0.20	0.22	0.20	0.20	0.19	0.15	0.15
	30	0.30	0.26	0.30	0.28	0.24	0.28	0.24	0.24	0.18	0.13	0.14	0.15	0.18	0.14	0.10	0.11
	50	0.23	0.23	0.20	0.19	0.20	0.18	0.16	0.17	0.10	0.12	0.09	0.10	0.15	0.09	0.08	0.06
	70	0.18	0.19	0.15	0.16	0.13	0.13	0.12	0.13	0.10	0.09	0.07	0.06	0.13	0.08	0.06	0.05
	140	0.11	0.11	0.11	0.10	0.09	0.08	0.07	0.07	0.08	0.05	0.05	0.04	0.12	0.07	0.05	0.04
$RMSE\{\hat{\sigma}_\beta\}$	10	0.41	0.34	0.31	0.32	0.37	0.34	0.23	0.28	0.18	0.16	0.19	0.19	0.15	0.16	0.14	0.12
	20	0.21	0.15	0.18	0.17	0.15	0.13	0.15	0.14	0.12	0.09	0.08	0.07	0.13	0.08	0.07	0.05
	30	0.12	0.13	0.12	0.12	0.09	0.11	0.09	0.09	0.09	0.06	0.05	0.05	0.13	0.07	0.06	0.04
	50	0.08	0.08	0.08	0.08	0.06	0.06	0.06	0.06	0.08	0.05	0.04	0.03	0.12	0.07	0.05	0.03
	70	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.07	0.04	0.03	0.02	0.12	0.07	0.05	0.03
	140	0.04	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.07	0.04	0.03	0.02	0.12	0.07	0.05	0.03

The bias and root mean square errors (RMSE) of the square root of the diagonal elements of $\hat{\triangle}_1$, when $E\left(\sigma_i^2\right) = 1$ and $(c, \beta) = (0, 0.5)$ (Option 2), for various degree of coefficient heterogeneity ($\sigma_\beta = \sigma_c$), across different time (T) and cross-section dimensions (N).

Therefore, it is important to assess the finite-sample consequences of using $\hat{\triangle}_1$ as an estimator of $\triangle$. The aim of this subsection is to provide some evidence on whether it is appropriate to rely on the asymptotic properties of this estimator as the basis for inference in finite samples. Without loss of generality, we focus on the results obtained from Option 2, where $(c, \beta) = (0, 0.5)$. We only show results obtained when $E\left(\sigma_i^2\right) = 1$, for various degrees of coefficient heterogeneity. The consequences of using $\hat{\triangle}_1$ as an estimator of $\triangle$ are even more severe when $E\left(\sigma_i^2\right)$

increases.[9] Further analyses are available in an online Appendix.

Notation. Hereafter, we use the following notation to avoid repetition. We let $\psi_0 = (c, \beta)' = (0, 0.5)'$ be the true vector of average effects. The true random coefficient covariance matrix, $\triangle$, is diagonal, where σ_c^2 and σ_β^2 are the $(1,1)$ and $(2,2)$ entries, respectively. We let

$$\hat{\psi}_{GLS} = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' V_i^{-1} y_i \right), \tag{2.22}$$

and

$$\Phi = var\left(\hat{\psi}_{GLS}\right) = \left(\sum_{i=1}^{N} X_i' V_i^{-1} X_i \right)^{-1}, \tag{2.23}$$

where $V_i = X_i \triangle X_i' + \sigma_i^2 I_T$, be the infeasible GLS estimator of ψ, and the infeasible covariance matrix of $\hat{\psi}_{GLS}$, respectively. The feasible GLS estimator, $\hat{\psi}_{FGLS}$, and an estimator of Φ, denoted $\hat{\Phi}$, are obtained by replacing σ_i^2 and $\triangle$ by $\hat{\sigma}_i^2$ and $\hat{\triangle}_1$, as defined in (2.7) and (2.9), respectively.

Accuracy of Estimated Standard Errors

To examine the consequences of overestimating the true random coefficient variances when testing hypotheses, we consider the ratio of the estimated standard errors (of the average effects) to the infeasible standard errors, obtained by taking the square root of the diagonal elements of $\hat{\Phi}$ and Φ respectively. Another measure of interest for inference is the accuracy of the estimated standard errors as approximations to the correct sampling standard deviation of the estimator of ψ.[10] These ratios should ideally be equal to one. Results reported in Table

[9] Case (v) is particularly interesting. Even though the variance of most of the units varies between 0.5 and 1.5, as in case (ii), the presence of some outliers, such that $E\left(\sigma_i^2\right) = 2$, considerably worsen the accuracy of inference.

[10] The accuracy of the estimated standard errors is computed as the ratio of $B^{-1}\sum_{b=1}^{B}\left\{\sqrt{(\hat{\Phi}_b)_{kk}}\right\}$ to the sampling standard deviation of $\hat{\psi}_k$, given by the square root of $(B-1)^{-1}\sum_{b=1}^{B}\left(\hat{\psi}_{k,(b)} - \bar{\hat{\psi}}_k\right)^2$, where $\bar{\hat{\psi}}_k = B^{-1}\sum_{b=1}^{B}\hat{\psi}_{k,(b)}$, for $k = 1, 2$.

2.4, show that relying exclusively on the asymptotic properties of $\hat{\triangle}_1$ may lead to invalid inference in finite samples. The estimated standard errors are upwards biased for the vast majority of cases. These biases can be substantial unless T and N or the degree of coefficient heterogeneity (σ_β) are large. However, if the coefficient dispersion is low, the estimated standard errors can be largely overestimated even when both T and N are equal to 140. These biases can in turn significantly affect inference.

Table 2.4: Accuracy of estimated standard errors

		$\sigma_\beta = 0.05$				$\sigma_\beta = 0.1$				$\sigma_\beta = 0.3$				$\sigma_\beta = 0.5$			
	$T\backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$Accuracy\{se(\hat{c})\}$	10	1.69	1.77	1.84	1.88	1.67	2.17	1.80	1.86	1.47	1.53	1.62	1.69	1.22	1.51	1.46	1.41
	20	1.61	1.65	1.87	1.83	1.53	1.72	1.72	1.75	1.31	1.30	1.33	1.40	1.16	1.21	1.20	1.16
	30	**1.60**	**1.62**	**1.78**	**1.72**	**1.58**	**1.78**	**1.63**	**1.67**	**1.32**	**1.22**	**1.28**	**1.33**	**1.18**	**1.16**	**1.11**	**1.10**
	50	1.54	1.84	1.71	1.72	1.51	1.57	1.58	1.59	1.16	1.20	1.20	1.21	1.11	1.13	1.10	1.05
	70	1.65	1.65	1.61	1.66	1.42	1.46	1.53	1.51	1.12	1.18	1.18	1.17	1.02	1.08	1.07	1.09
	140	1.59	1.57	1.71	1.56	1.38	1.35	1.37	1.38	1.04	1.10	1.07	1.11	0.96	1.00	1.00	1.05
$Accuracy\left\{se\left(\hat{\beta}\right)\right\}$	10	1.66	1.71	1.63	1.71	1.67	1.75	1.62	1.60	1.26	1.30	1.33	1.33	1.10	1.23	1.20	1.19
	20	1.53	1.53	1.62	1.55	1.41	1.49	1.52	1.54	1.17	1.17	1.16	1.15	1.02	1.06	1.10	1.06
	30	**1.50**	**1.53**	**1.57**	**1.53**	**1.36**	**1.46**	**1.40**	**1.45**	**1.09**	**1.12**	**1.13**	**1.11**	**1.06**	**1.06**	**1.05**	**1.03**
	50	1.40	1.54	1.53	1.47	1.23	1.35	1.35	1.30	1.06	1.02	1.09	1.09	1.01	1.02	1.03	1.03
	70	1.45	1.48	1.45	1.49	1.24	1.26	1.27	1.27	1.05	1.03	1.00	1.05	0.99	1.00	1.00	1.04
	140	1.39	1.38	1.37	1.38	1.19	1.16	1.12	1.16	1.02	1.01	1.01	1.00	0.96	1.01	0.99	1.01
$Ratio\{se(\hat{c})\}$	10	2.51	2.60	2.48	2.46	2.52	3.01	2.36	2.55	1.60	1.76	1.78	1.90	1.26	1.63	1.47	1.49
	20	2.12	2.03	2.29	2.20	1.70	2.03	1.95	2.01	1.34	1.37	1.41	1.40	1.18	1.23	1.19	1.21
	30	**1.87**	**1.82**	**2.05**	**1.99**	**1.77**	**1.99**	**1.82**	**1.82**	**1.29**	**1.26**	**1.28**	**1.33**	**1.15**	**1.17**	**1.13**	**1.16**
	50	1.74	2.13	1.96	1.87	1.72	1.74	1.65	1.70	1.15	1.24	1.20	1.23	1.10	1.09	1.09	1.09
	70	1.90	1.91	1.87	1.86	1.52	1.57	1.53	1.60	1.13	1.18	1.16	1.15	1.04	1.07	1.05	1.07
	140	1.74	1.73	1.78	1.72	1.43	1.40	1.41	1.41	1.04	1.08	1.09	1.09	1.00	1.02	1.03	1.04
$Ratio\left\{se\left(\hat{\beta}\right)\right\}$	10	2.39	2.26	2.09	2.16	2.28	2.31	1.94	2.04	1.30	1.35	1.41	1.45	1.11	1.23	1.21	1.18
	20	1.91	1.78	1.99	1.87	1.50	1.63	1.62	1.66	1.20	1.19	1.18	1.18	1.04	1.07	1.07	1.08
	30	**1.60**	**1.65**	**1.75**	**1.72**	**1.49**	**1.57**	**1.51**	**1.50**	**1.10**	**1.11**	**1.11**	**1.13**	**1.05**	**1.05**	**1.05**	**1.05**
	50	1.52	1.63	1.65	1.61	1.33	1.37	1.37	1.40	1.07	1.06	1.07	1.07	1.01	1.02	1.02	1.03
	70	1.60	1.57	1.58	1.57	1.27	1.31	1.29	1.30	1.04	1.05	1.05	1.05	0.99	1.01	1.01	1.02
	140	1.43	1.44	1.43	1.42	1.17	1.15	1.17	1.18	1.00	1.02	1.02	1.02	0.99	1.00	1.00	1.01

$Accuracy\{se(\cdot)\}$ denotes the ratio of the estimated standard errors (of the average effects) to the sampling standard deviations. $Ratio\{se(\cdot)\}$ denotes the ratio of the estimated standard errors to the infeasible standard errors. Results obtained using Option 2, when $E(\sigma_i^2) = 1$.

Hypothesis tests

To test the hypothesis $\psi = \underline{\psi}$, for $\underline{\psi}$ a known $K \times 1$ vector, Swamy (1970) suggests the following criterion:

$$F\left(\hat{\psi}, \underline{\psi}, \hat{\Phi}\right) = \frac{N-K}{K\left(N-1\right)} \left(\hat{\psi} - \underline{\psi}\right)' \hat{\Phi}^{-1} \left(\hat{\psi} - \underline{\psi}\right). \tag{2.24}$$

The asymptotic distribution of the test is F, with K, $N-K$ degrees of freedom.

Empirical Moments of the F-statistic. We now study the finite-sample properties of the distribution of (3.50). In particular, we examine the empirical distributions of $F\left(\hat{\psi}_{k,GLS}, \psi_{0,k}, \Phi_{kk}\right)$, and $F\left(\hat{\psi}_{k,FGLS}, \psi_{0,k}, \hat{\Phi}_{kk}\right)$, computed under the null hypothesis that the estimator (of interest) of ψ_k is equal to the corresponding true value used to generate the data, $\psi_{0,k}$, for $k = 1, 2$.[11] We then compare the mean, standard deviation, skewness, and excess kurtosis of these two empirical distributions with the corresponding population moments of a F-distribution with $1, N-1$ degrees of freedom. Results are reported in Table 2.5. In many cases, the means and standard deviations of the distributions of the F-statistics based on the infeasible GLS estimator are relatively close to the true means and standard deviations. This is not the case when considering the distributions of the F-statistics based on the feasible GLS estimator. The means and standard deviations of the latter can be substantially smaller than the values associated with a F-distribution with 1, $N-1$ degrees of freedom. Results worsen when testing hypothesis about the intercept rather than slope parameters. These results are in line with the fact that $\hat{\triangle}_1$ is often upwards biased. The skewness and excess-kurtosis of the distribution of the F-statistics based on both feasible and infeasible GLS estimators, can be far from the corresponding population moments unless N is large.

[11] Φ_{kk} is the kth diagonal element of Φ. Similarly, ψ_k denotes the kth element of ψ.

Table 2.5: Empirical moments of F-statistics

β	$T\backslash N$	Mean 10	Mean 30	Mean 50	Mean 140	SD 10	SD 30	SD 50	SD 140	Skew 10	Skew 30	Skew 50	Skew 140	Ex-Kurt 10	Ex-Kurt 30	Ex-Kurt 50	Ex-Kurt 140
$F_{1,N-1}$		**1.29**	**1.07**	**1.04**	**1.01**	**2.30**	**1.61**	**1.52**	**1.45**	**6.71**	**3.37**	**3.12**	**2.92**	**214.50**	**19.23**	**15.61**	**13.11**
Infeasible	10	0.96	1.06	0.97	1.04	1.37	1.47	1.47	1.47	2.86	2.40	3.17	2.60	11.67	7.23	14.87	8.78
	20	1.01	1.00	0.94	0.99	1.47	1.56	1.35	1.41	2.92	3.65	2.86	2.95	12.66	21.85	11.26	12.12
	30	**1.02**	**0.98**	**1.03**	**0.97**	**1.38**	**1.42**	**1.40**	**1.41**	**2.44**	**2.65**	**2.35**	**2.92**	**7.69**	**9.43**	**7.06**	**12.41**
	50	1.08	0.96	0.96	1.06	1.58	1.33	1.33	1.53	3.06	2.65	2.52	2.73	13.51	9.55	8.06	9.99
	70	1.00	1.02	0.98	0.98	1.42	1.47	1.46	1.49	3.74	3.41	3.32	3.12	28.41	19.85	17.49	14.11
	140	0.96	0.96	1.07	1.02	1.31	1.33	1.43	1.44	2.64	2.61	2.58	2.51	9.77	8.82	9.14	7.56
Feasible	10	0.37	0.34	0.39	0.39	0.56	0.48	0.55	0.54	3.48	2.95	2.77	2.61	18.91	13.18	11.11	10.25
	20	0.55	0.46	0.44	0.43	0.86	0.72	0.63	0.64	3.86	3.42	2.90	3.24	23.48	17.91	12.25	15.69
	30	**0.58**	**0.48**	**0.51**	**0.48**	**0.85**	**0.70**	**0.69**	**0.69**	**3.25**	**3.59**	**2.49**	**2.97**	**15.13**	**26.34**	**8.97**	**12.46**
	50	0.71	0.56	0.56	0.59	1.14	0.81	0.79	0.85	3.59	3.04	2.74	2.83	18.00	13.08	10.30	11.52
	70	0.70	0.64	0.62	0.62	1.09	0.97	0.94	0.96	4.45	4.48	3.34	3.31	37.08	40.21	17.46	16.76
	140	0.79	0.76	0.82	0.75	1.27	1.11	1.14	1.06	4.36	3.06	2.74	2.56	34.97	13.97	10.30	8.02

c	$T\backslash N$	Mean 10	Mean 30	Mean 50	Mean 140	SD 10	SD 30	SD 50	SD 140	Skew 10	Skew 30	Skew 50	Skew 140	Ex-Kurt 10	Ex-Kurt 30	Ex-Kurt 50	Ex-Kurt 140
$F_{1,N-1}$		**1.29**	**1.07**	**1.04**	**1.01**	**2.30**	**1.61**	**1.52**	**1.45**	**6.71**	**3.37**	**3.12**	**2.92**	**214.50**	**19.23**	**15.61**	**13.11**
Infeasible	10	1.09	1.01	0.96	1.01	1.43	1.37	1.40	1.43	2.34	2.50	2.74	2.67	7.00	8.57	9.76	9.27
	20	0.99	1.00	1.02	0.97	1.40	1.43	1.41	1.44	2.87	2.94	2.72	4.32	12.95	11.53	10.39	38.22
	30	**1.02**	**0.95**	**0.99**	**1.01**	**1.45**	**1.37**	**1.40**	**1.38**	**2.56**	**3.05**	**2.73**	**2.73**	**8.44**	**15.32**	**10.26**	**11.25**
	50	1.02	1.00	0.98	0.98	1.54	1.36	1.37	1.40	2.87	2.64	2.53	3.03	10.30	11.47	8.26	14.45
	70	0.99	1.05	0.95	1.00	1.49	1.50	1.38	1.43	3.17	2.52	3.11	3.07	15.04	7.93	14.63	15.15
	140	0.99	1.01	1.00	0.99	1.35	1.50	1.43	1.34	2.43	2.93	2.78	2.33	7.81	11.53	10.54	6.71
Feasible	10	0.38	0.22	0.31	0.29	0.59	0.30	0.43	0.43	4.76	2.47	2.55	2.77	45.68	8.11	8.78	10.07
	20	0.46	0.35	0.34	0.33	0.77	0.53	0.49	0.47	5.41	3.53	2.79	3.14	51.13	19.04	10.70	14.86
	30	**0.43**	**0.32**	**0.38**	**0.36**	**0.62**	**0.44**	**0.55**	**0.49**	**2.69**	**2.93**	**2.76**	**2.67**	**9.45**	**13.93**	**11.42**	**10.02**
	50	0.47	0.42	0.40	0.40	0.70	0.60	0.57	0.58	3.11	3.05	2.80	3.19	13.65	13.79	11.90	14.99
	70	0.53	0.49	0.44	0.44	0.81	0.71	0.65	0.62	3.72	2.95	3.05	2.90	22.65	12.36	13.24	13.28
	140	0.57	0.57	0.54	0.53	0.86	0.92	0.78	0.73	3.27	4.21	2.85	2.52	16.53	31.53	11.26	8.51

Empirical moments of F-statistics across different sample sizes ($T.N$), when the data are generated from Option 2, with $E(\sigma_i^2) = 1$, and $\sigma_\beta = \sigma_c = 0.1$. In the upper panel, the test statistics are constructed under the null hypothesis H_0: $\beta = 0.5$ against the alternative H_1: $\beta \neq 0.5$. In the lower panel, the null hypothesis is H_0: $c = 0$ against H_1: $c \neq 0$. Row "$F_{1,N-1}$" reports the population moments of a F-distribution with 1, $N-1$ degrees of freedom. The empirical moments reported in "Infeasible" correspond to the F-statistics computed using the infeasible GLS estimator of ψ and the infeasible covariance matrix, Φ. "Feasible" is used to denote the empirical moments of the F-statistics, replacing the unknown components in ψ and Φ by their estimators.

Power Performances. In Table 2.6 we report the empirical sizes of the F-statistic, described in equation (3.50), of the null hypothesis H_0: $\psi_k = \psi_{0,k}$ against the alternative H_1: $\psi_k \neq \psi_{0,k}$. They are computed as the relative rejection frequencies based on the critical regions of nominal size 0.05 of a F-distribution with $1, N-1$ degrees of freedom. This allows us to evaluate the direct consequences of the various results described above for hypothesis tests.

Table 2.6: Empirical sizes based on F-statistics

		$\sigma_\beta = 0.05$				$\sigma_\beta = 0.1$				$\sigma_\beta = 0.3$				$\sigma_\beta = 0.5$			
	$T \backslash N$	10	30	50	140	10	30	50	140	10	30	50	140	10	30	50	140
$size\left(\hat{\beta}_{GLS}\right)$	10	2.47	4.80	4.47	**4.00**	2.13	5.07	4.60	**5.33**	2.13	4.00	4.60	**5.67**	2.27	3.80	4.00	**4.13**
	20	2.67	4.53	4.40	**5.67**	2.93	4.40	3.73	**3.73**	2.80	4.40	4.00	**5.27**	2.47	3.27	3.47	**4.80**
	30	2.20	4.73	4.60	**4.60**	2.33	4.33	5.13	**4.67**	1.60	4.47	3.13	**4.87**	2.00	3.87	4.47	**5.33**
	50	2.27	3.87	3.80	**5.40**	3.07	3.53	4.00	**5.47**	2.60	5.00	3.40	**4.53**	2.20	4.07	4.00	**5.07**
	70	2.33	3.33	3.93	**4.27**	1.87	3.53	4.13	**5.07**	2.60	4.33	5.27	**4.07**	2.33	3.67	5.27	**3.80**
	140	2.53	3.33	4.13	**4.53**	2.13	3.47	4.80	**5.80**	1.93	3.60	4.40	**4.93**	3.33	4.27	4.87	**4.73**
$size\left(\hat{\beta}_{FGLS}\right)$	10	0.00	0.07	0.07	**0.20**	0.13	0.07	0.13	**0.07**	1.07	0.60	0.73	**0.60**	2.20	1.00	1.67	**2.00**
	20	0.07	0.20	0.00	**0.13**	0.40	0.47	0.27	**0.33**	1.80	2.47	2.13	**2.27**	3.47	2.93	2.93	**3.60**
	30	0.20	0.27	0.13	**0.13**	0.67	0.20	0.27	**0.53**	2.20	3.20	2.13	**3.00**	2.73	3.47	3.60	**4.40**
	50	0.33	0.07	0.20	**0.60**	1.20	0.93	0.93	**1.07**	3.07	4.53	3.07	**3.00**	3.87	4.47	3.93	**4.53**
	70	0.20	0.27	0.53	**0.33**	0.87	0.87	1.27	**1.40**	2.93	4.00	4.87	**3.40**	4.60	4.67	4.47	**3.73**
	140	0.40	0.47	0.47	**0.53**	1.33	2.13	2.40	**2.47**	4.33	4.00	4.60	**4.87**	5.33	5.60	5.33	**4.33**
$size\left(\hat{c}_{GLS}\right)$	10	2.33	4.00	4.33	**4.27**	2.53	3.93	3.80	**4.53**	1.93	3.87	3.47	**4.40**	2.33	3.47	3.67	**5.33**
	20	2.47	4.53	3.93	**4.27**	2.13	3.87	3.93	**4.07**	3.20	4.33	5.27	**4.27**	3.00	3.60	4.00	**5.53**
	30	2.60	5.00	3.80	**5.13**	2.73	3.60	4.13	**5.13**	1.67	4.27	4.07	**3.93**	2.27	3.87	4.33	**6.60**
	50	2.60	4.00	4.53	**4.13**	3.13	3.80	4.40	**4.80**	2.13	4.80	3.93	**5.33**	2.33	3.07	3.53	**5.13**
	70	2.00	4.40	5.27	**4.53**	2.53	4.80	3.80	**4.60**	3.00	4.20	3.00	**3.73**	2.27	3.20	4.87	**4.07**
	140	2.07	4.80	3.47	**4.00**	2.07	4.27	4.13	**4.67**	1.93	4.13	4.73	**4.47**	2.87	5.33	5.00	**4.73**
$size\left(\hat{c}_{FGLS}\right)$	10	0.07	0.07	0.00	**0.00**	0.07	0.00	0.00	**0.00**	0.20	0.27	0.47	**0.20**	1.20	0.40	0.53	**0.53**
	20	0.13	0.07	0.13	**0.07**	0.40	0.20	0.07	**0.07**	0.73	0.53	1.27	**0.67**	1.73	1.47	1.80	**2.07**
	30	0.07	0.13	0.13	**0.07**	0.07	0.07	0.07	**0.13**	0.73	1.40	1.20	**0.60**	1.80	2.33	2.60	**3.53**
	50	0.07	0.07	0.13	**0.00**	0.20	0.27	0.20	**0.27**	1.67	1.73	1.80	**1.80**	2.13	2.27	2.60	**3.00**
	70	0.00	0.07	0.13	**0.13**	0.33	0.47	0.33	**0.27**	2.60	1.87	1.93	**1.93**	3.53	2.40	3.80	**3.00**
	140	0.27	0.13	0.27	**0.20**	0.33	1.20	0.73	**0.47**	3.00	3.47	3.00	**3.60**	5.27	5.47	5.00	**3.87**

Rejection frequencies (%) at 5% nominal level obtained computing the F-statistic described in (3.50), under the null hypothesis H_0: $\psi_k = \psi_{0,k}$ against the alternative H_1: $\psi_k \neq \psi_{0,k}$. $\hat{\beta}_{GLS}$ and $\hat{c}_{GLS}$ denote the infeasible GLS estimator of β and c respectively. Similarly, the subscript "FGLS " stands for feasible GLS. The data are generated from Option 2, with $E(\sigma_i^2) = 1$.

The tests based on the feasible GLS estimation severely suffer from size distortions. Unless the degree of coefficient heterogeneity is quite high (e.g. $\sigma_\beta = 0.5$), the sizes are always substantially lower than the nominal levels. They are often close to zero due to the fact that the estimated standard errors are largely biased upward. Once again, the distortions are even more severe when testing about the intercept parameters.

To support these findings, we plot the power functions for the slope and intercept parameters in Figure 2.1 and 2.2 respectively. To save space, we only report results for the case with $E(\sigma_i^2) = 1$ and $\sigma_\beta = 0.1$.[12]

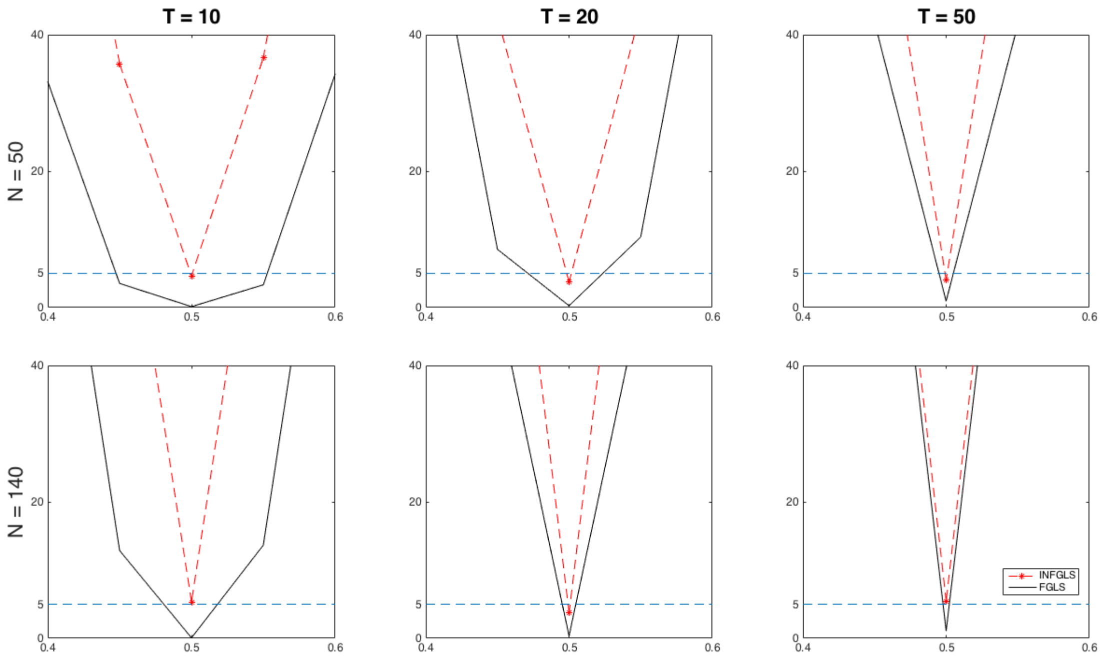


Figure 2.1: Rejection frequency (%) at the 5% nominal level, for the slope parameter (β), in the y-axis. They are computed using the F-statistic described in (3.50), under the null hypothesis H_0: $\beta = \underline{\beta}$ against the alternative $\beta \neq \underline{\beta}$. Different values of $\underline{\beta}$ are reported in the x-axis. The true value of β is 0.5. The black lines and the red dotted lines denote the power performances of feasible and infeasible GLS estimators, respectively. Results obtained using Option 2, with $E(\sigma_i^2) = 1$ and $\sigma_\beta = 0.1$.

[12]Additional results are available in an online Appendix.

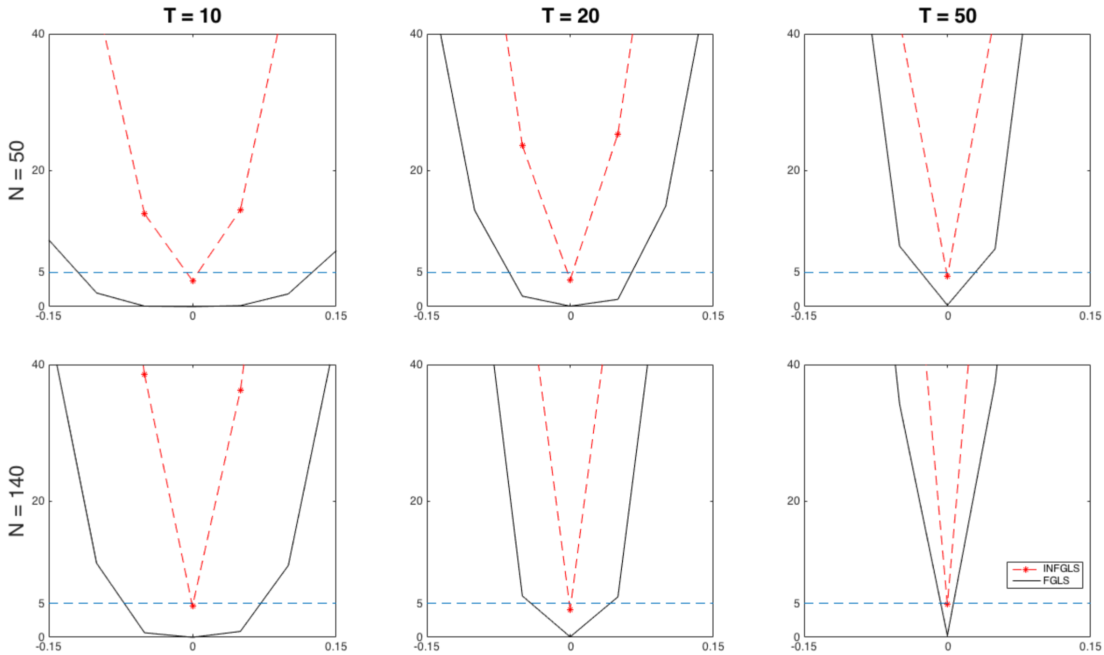


Figure 2.2: Rejection frequency (%) at the 5% nominal level, for the intercept parameter (c), in the y-axis. They are computed using the F-statistic described in (3.50), under the null hypothesis H_0: $c = \underline{c}$ against the alternative $c \neq \underline{c}$. Different values of $\underline{c}$ are reported in the x-axis. The true value of c is 0. The black lines and the red dotted lines denote the power performances of feasible and infeasible GLS estimators, respectively. Results obtained using Option 2, with $E(\sigma_i^2) = 1$ and $\sigma_c = 0.1$.

2.5 Conclusions

As in the error component model, the estimator of the coefficients' covariance matrix in a random coefficient model is often negative definite. The aim of this study is to investigate the causes and effects of the problem. By running some Monte Carlo experiments, we show that the degree of coefficient heterogeneity relative to the (conditional) variability of the dependent variables plays a crucial role. The larger the coefficient dispersion and the precision of the regression disturbances (the inverse of the average variance of the time-varying errors), the lower the probability to observe a negative definite estimator of the random coefficient covariance matrix. An increase in the former has a larger effect than an increase in the latter. Similarly, this probability decreases as the time dimension and the number of units get large, partly due to the fact that the performances (in terms of bias and RMSE) of individual OLS estimates and the Mean Group improves in large samples. It is known that when the time dimension goes to infinity, the negative definiteness problem vanishes.

We then demonstrate that relying on the asymptotic properties of the biased but consistent estimator of the random coefficient covariance matrix may lead to poor inference. Unless the time and cross-section dimensions, and/or the degree of coefficient dispersion are high, the estimated standard errors are largely upwards biased. The resulting hypothesis tests may suffer from considerable size distortions. The empirical sizes of the tests are substantially lower than the nominal levels. Results may worsen when the precision of the regression disturbances decreases. An estimation procedure which yields an unbiased and more efficient estimator of the random coefficient covariance and which performs relatively well in terms of accuracy of inference is proposed in Chapter 3.

2.6 Appendix

2.6.1 Estimation of Parameters in the Presence of Serially Correlated Disturbances

Swamy (1971) considers the estimation problem of Ω_i when the the disturbances follow an AR(1) process:

$$u_{it} = \phi_i u_{i,t-1} + \epsilon_{it}, \qquad 0 <| \phi_i |< 1, \tag{2.25}$$

and $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}\epsilon_{js}) = \sigma_i^2$ if $t = s$ and $i = j$, and 0 otherwise. For $i = j$, $E\left(u_i u_j'\right) = \sigma_i^2 \Omega_i$, where

$$\Omega_i = \frac{1}{1-\phi_i^2}\begin{bmatrix} 1 & \phi_i & \phi_i^2 & \cdots & \phi_i^{T-1} \\ \phi_i & 1 & \phi_i & \cdots & \phi_i^{T-2} \\ \phi_i^2 & \phi & 1 & & \phi_i^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_i^{T-1} & \phi_i^{T-2} & \phi_i^{T-3} & \cdots & 1 \end{bmatrix},$$

and $E\left(u_i u_j'\right) = 0$, if $i \neq j$. A consistent estimator of ϕ_i is given by

$$\hat{\phi}_i = \frac{\sum_{t=2}^{T} \hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{t=2}^{T} \hat{u}_{i,t-1}^2}, \tag{2.26}$$

where $\hat{u}_{it}$ is the t-th element of $\hat{u}_i$, the vector of OLS residuals. An estimator of Ω_i can be obtained by replacing ϕ_i by $\hat{\phi}_i$ in Ω_i. Note also that the inverse of $\hat{\Omega}_i$

can be computed using the fact that $\hat{\Omega}_i^{-1} = \hat{R}_i'\hat{R}_i$, where

$$\hat{R}_i = \begin{bmatrix} \sqrt{1-\hat{\phi}_i^2} & 0 & 0 & \cdots & 0 & 0 \\ -\hat{\phi}_i & 1 & 0 & & & 0 \\ 0 & -\hat{\phi}_i & 1 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ & & & & & \\ 0 & 0 & \cdots & & -\hat{\phi}_i & 1 \end{bmatrix}.$$

Under assumption (2.25), by regressing y_i upon X_i, and applying Aitken's GLS to each time series, we have

$$\begin{aligned} \hat{\beta}_{i,gls} &= (X_i'\Omega_i^{-1}X_i)^{-1}X_i'\Omega_i^{-1}y_i \\ &= \beta + \delta_i + (X_i'\Omega_i^{-1}X_i)^{-1}X_i'\Omega_i^{-1}u_i. \end{aligned} \tag{2.27}$$

The feasible GLS estimator of β_i is given by

$$\tilde{\beta}_i = (X_i'\hat{\Omega}_i^{-1}X_i)^{-1}X_i'\hat{\Omega}_i^{-1}y_i. \tag{2.28}$$

The average effect, β, can be estimated by

$$\hat{\beta}_{FGLS} = \sum_{i=1}^{N} \hat{W}_i\tilde{\beta}_i, \tag{2.29}$$

where

$$\hat{W}_i = \left\{ \sum_{i=1}^{N} \left[\hat{\triangle} + \tilde{\sigma}_i^2 (X_i'\hat{\Omega}_i^{-1}X_i)^{-1} \right]^{-1} \right\}^{-1} \left[\hat{\triangle} + \tilde{\sigma}_i^2 (X_i'\hat{\Omega}_i^{-1}X_i)^{-1} \right]^{-1},$$

$$\tilde{\sigma}_i^2 = \frac{\tilde{u}_i'\tilde{u}_i}{T-K},$$

and $\tilde{u}_i = \hat{R}_i y_i - \hat{R}_i X_i \tilde{\beta}_i$.

Similarly to (2.12), using (2.27), we can compute

$$var\left(\hat{\beta}_{i,gls}\right) = E\left(\hat{\beta}_{i,gls} - \beta\right)\left(\hat{\beta}_{i,gls} - \beta\right)' = \triangle + \sigma_i^2(X_i'\Omega_i^{-1}X_i)^{-1}.$$

The estimator of $\triangle$ becomes

$$\hat{\triangle} = \hat{\triangle}_4 - \hat{\triangle}_5, \tag{2.30}$$

where

$$\begin{aligned} \hat{\triangle}_4 &= \tfrac{1}{N-1}\textstyle\sum_{i=1}^{N}\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^{N}\tilde{\beta}_i\right)\left(\tilde{\beta}_i - N^{-1}\sum_{i=1}^{N}\tilde{\beta}_i\right)', \\ \hat{\triangle}_5 &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tilde{\sigma}_i^2\left(X_i'\hat{\Omega}_i X_i\right)^{-1}. \end{aligned} \tag{2.31}$$

Chapter 3

Estimation and Inference in Mixed Fixed and Random Coefficient Panel Data Models

3.1 Introduction

This chapter considers the problem of statistical inference in random coefficient panel data models, when both N (the number of units) and T (the number of time periods) are quite large. In the presence of heterogeneity, the parameters of interest may be the unit-specific coefficients, their expected values, and their variances over the units. Two main estimators for the expected value of the random coefficients are used in the literature. Pesaran and Smith (1995) suggest estimating N time series separately to then obtain an estimate of the expected value of the unit-specific coefficients by averaging the OLS estimates for each unit. They call this procedure Mean Group estimation. Alternatively, under the assumption that the coefficients are random draws from a common distribution, one can apply Swamy (1970) GLS estimation, which yields a weighted average of the individual OLS estimates.[1] However, as in the error-component model, the Swamy estimator of the random coefficient covariance matrix is not necessarily nonnegative definite. We have investigated the consequences of this drawback in finite samples, in particular when testing hypotheses, in Chapter 2. In this Chapter, we propose a solution to the above mentioned problem by applying the EM algorithm. In particular, following the seminal papers of Dempster et al. (1977), and Patterson and Thompson (1971), we propose to estimate heterogeneous panels by applying the EM algorithm to obtain tractable closed form solutions of restricted maximum likelihood (REML) estimates of both fixed and random components of the regression coefficients as well as the variance parameters. The proposed estimation procedure is quite general, as we consider a broad framework which incorporates various panel data models as special case. Our approach yields an estimator of the average effects which is asymptotically related to both the GLS and the Mean Group estimator, and which performs relatively well in finite sample as shown in our limited Monte Carlo analysis. We also review some of the existing sampling and Bayesian methods commonly used to estimate heterogeneous panel data, 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] The EM algorithm has also recently gained attention in the finance literature. Harvey and Liu (2016) suggest a similar approach to ours to

[1]Swamy focuses on estimating the average effects while the random effects are treated as nuisance parameters 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. 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, Pesaran and Tahmiscioglu (1999). A good survey of the literature is provided by Hsiao and Pesaran (2008) 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.

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 consider a different framework where both the intercept and slope parameters are a function of a set of explanatory variables and are randomly drawn from a common distribution. We derive an expression for the likelihood of the model accordingly. More importantly, differently from Harvey and Liu, our goal is to illustrate the advantages of the EM-REML approach in estimating a general class of heterogeneous panel data models, in relation to the existing methods.

First, estimating heterogeneous panels by EM-REML yields unbiased and more efficient 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. This alternative estimator is consistent when T tends to infinity but it is severely biased in small samples. As shown in the Monte Carlo analysis, this in turn leads to biased estimated standard errors and may affect the power performances of the GLS estimator. Compared to Swamy estimator, the EM-REML method leads to remarkable reduction of the bias and root mean square errors of the estimates of the random coefficient variances. As a results, the estimated standard errors have lower bias, leading to more accurate hypothesis tests. A valid estimator of the random coefficient covariance matrix is also important to correctly detect the degree of coefficient heterogeneity. As noted by Trapani and Urga (2009), the latter plays a crucial role on the forecasting performance of various panel estimators, while other features of the data have a very limited impact. Therefore, our estimator of the covariance matrix may be considered by applied researchers to choose the appropriate estimator for forecast purposes.

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, we demonstrate that their estimate of the random coefficients' variance-covariance matrix does not satisfy the law of total variance. This is not the case when using the EM algorithm. Differently from Lee and Griffiths, 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 effects given the observed data is necessary for the law of total variance to hold.

Another interesting feature of the EM (compared to the papers mentioned in the above paragraph) is that it allows us to make inference on the random effects' population. Indeed, in general, it gives a probability distribution over the missing data.

The random effects are estimated by the mean of their posterior distribution, under the assumption that the regressors are strictly exogenous,. Substituting the unknown variance components by their estimates yields the empirical best linear unbiased predictor. We also note that the EM-REML estimator of the average effects is related to the empirical Bayesian estimator described in Hsiao, Pesaran and Tahmiscioglu (1999). The EM-REML estimators of the variance components are analogous to the Bayes mode of their posterior distribution, derived in Lindley and Smith (1972). In view of the relatively good finite-sample performances, the EM approach should be regarded as a valid alternative to Bayesian estimation in those cases in which the researcher wishes to make inference on the random effects distribution while having little knowledge on what sensible priors might be. At the same time, a drawback of the Bayesian approach is that, when sample sizes is not too large (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). To illustrate, Hsiao, Pesaran and Tahmiscioglu (1999), suggest using the Swamy covariance's estimator as a prior input for the random coefficient covariance matrix. However, they note that the latter affects the empirical and hierarchical Bayes estimates of the regression coefficients adversely, especially when the degree of coefficient heterogeneity decreases. Alternatively, when considering a diffuse prior, their Gibbs sampling algorithm breaks down completely in some experiments. Another merit of our method is to overcome this problem.

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 explain and quantify the cross-sectional difference in the reaction of sovereign spreads to change in government debt.[3] First, we show that financial markets reactions to an increase in government debt are 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 country-specific macroeconomic indicators, commonly found to be significant determinants of sovereign credit risk, 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 years in default or restructuring domestic debt is associated with around 0.35% increase in the additional risk premium in response to an increase in debt.

This chapter is organized as follows. Section 3.2 describes the regression model and its main assumptions. In Section 3.3 an expression for the likelihood of the complete data, which includes both the observed and the missing data, is obtained. The restricted likelihood is also derived. Section 3.4 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 3.5. The problem of inference in finite sample

[3]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.

is addressed in Section 3.6. Results from Monte Carlo experiments are shown in Section 3.7. In Section 3.8, we employ the econometric model to study the determinants of the sensitivity of sovereign spreads. Finally, we conclude.

3.2 A Mixed Fixed and Random Coefficient Panel Data Model

We assume that the dependent variable, y_{it}, is generated according to the following linear panel model with unit-specific coefficients,

$$y_{it} = c_i + x'_{it}\beta_i + \varepsilon_{it}, \tag{3.1}$$

for $i = 1, .., N$ and $t = 1, .., T$, where x_{it} is a $K \times 1$ vector of exogenous regressors. The model can be written in stacked form

$$y_i = Z_i\psi_i + \varepsilon_i, \tag{3.2}$$

where y_i is a $T \times 1$ vector of dependent variables for unit i, and Z_i is a $T \times K^*$ matrix of explanatory variables, including a vector of ones.[4] 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 $T \times k_1^*$ and $\underline{Z}_i$ is $T \times k_2^*$, with $K^* = k_1^* + k_2^*$. The coefficients in ψ_{1i} are assumed to be constant over time but differ randomly across units. Individual-specific characteristics are the main source of heterogeneity in the parameters:

[4] To make notation easier, we assume that $T = T_i$, for all i, although the results are also valid for an unbalanced panel.

$$\psi_{1i} = \Gamma_1 f_{1i} + \gamma_i, \tag{3.3}$$

where γ_i is a $k_1^* \times 1$ vector of random effects, Γ_1 is a $(k_1^* \times l_1)$ matrix of unknown fixed parameters, and 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

$$\psi_{2i} = \Gamma_2 f_{2i}, \tag{3.4}$$

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 (3.3) and (3.4) can be rewritten as

$$\begin{aligned} \psi_{1i} &= \left(f'_{1i} \otimes I_{k_1^*}\right) \bar{\Gamma}_1 + \gamma_i, \\ \psi_{2i} &= \left(f'_{2i} \otimes I_{k_2^*}\right) \bar{\Gamma}_2, \end{aligned} \tag{3.5}$$

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 (3.5) into (3.2) yields

$$y_i = W_i \bar{\Gamma} + \bar{Z}_i \gamma_i + \varepsilon_i, \tag{3.6}$$

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

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

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:

$$\varepsilon_{it} \sim IIN(0, \sigma^2_{\varepsilon_i}). \tag{3.7}$$

(ii) x_{it} and ε_{is} are independently distributed for all t and s (i.e. x_{it} are strictly exogenous). Both set of variables are independently distributed of γ_j, for all i and j.

(iii) f_{1i} and f_{2i} are independent of the ε_{jt}'s and γ_j, for all i and j.

(iv) The vector of unit-specific random effects is independently normally distributed as

$$\gamma_i \sim IIN(0, \triangle), \qquad \forall i. \tag{3.8}$$

Special Cases. Many panel data models can be derived as special cases of the model described in equation (3.6). Among others:

1. Models in which all the coefficients are stochastic and depend on individual-specific characteristics can be obtained from (3.6) 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 fixed coefficients.

3. The correlated random effects (CRE) model proposed by Mundlak (1978b) 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 $\Gamma_1 \equiv c \in \mathbb{R}$.

5. Model with interaction terms (e.g. Friedrich (1982)): $\bar{Z}_i = 0$ and for instance $f_{2i} = 1$, while $\underline{Z}_i$ contains the interaction terms.

6. Common Model for all cross-sectional units: $\bar{Z}_i = 0$, and $f_{2i} = 1$ for all i.[5]

Dynamic Panels. Many economic applications involve behavioural relationships which are dynamic in nature, requiring a model with lagged dependent variables appearing as regressors, such as

$$y_{it} = c_i + \phi_i y_{i,t-1} + x'_{it}\beta_i + \varepsilon_{it}, \tag{3.9}$$

for $i = 1, .., N$, and $t = 1, .., T$. In such cases, when rewriting the model in its stacked form (3.2), Z_i would include lagged values of y_i. However, including lagged dependent variables among the regressors raises a problem of endogeneity since they are a function of the individual random effects. Therefore, although we may maintain the assumptions (i)-(iv), we cannot assume that $E(\gamma_i y_{i,t-1}) = 0$. Consequently, the estimates of the coefficients will be biased and inconsistent when T is fixed, even for large N. Only when T is sufficiently large, one can rely on the consistency properties. It is noteworthy that our estimators are derived under the assumption that the regressors, x_{it}'s, are strictly exogenous and the initial observation y_{i0} are fixed constants. This choice is in line with Maddala et al. (1997) and Hsiao, Pesaran and Tahmiscioglu (1999), and is also motivated by the need of directly comparing our approach with the Swamy method which has been developed for the static case. Anderson and Hsiao (1981, 1982) argue that regarding the first observation as fixed can be a strong assumption for finite T. Nevertheless, as it will be shown in the Monte Carlo analysis (where we allow the initial values of the dependent variables to be random draws from different populations with different means and variances, and correlated with the random effects), the proposed method has relatively good properties when estimating dynamic panels even in moderate samples size (e.g. $N = 30$ and $20 < T \leq 50$). The case where the initial observations are treated as random and correlated with

[5] Models 5 and 6 do not involve any random coefficient and do not require the use of the EM algorithm.

the random effects will be investigated in a separate research.

Estimation of Long-Run Effects. In the dynamic case, one can estimate the vector of long-run effects of a set of regressors on the dependent variables as

$$\hat{\theta}_i = \frac{\hat{\beta}_i}{1 - \hat{\phi}_i}, \tag{3.10}$$

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

Cross-Section Dependence. It may occur that the assumption of independence (across units) of the error terms does 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 ($\bar{f}_t$):

$$\varepsilon_{it} = \tau_i' \bar{f}_t + \epsilon_{it}, \tag{3.11}$$

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. Equation (3.2) should be replaced by

$$y_i = Z_i \psi_i + \epsilon_i, \tag{3.12}$$

where Z_i would now also include cross-section averages of the dependent and right-hand side variables.

3.3 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 containing the non-zero elements of the covariance matrix $\triangle$. We consider the unobserved random effects, $\gamma = (\gamma_1', .., \gamma_N')'$, 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

$$logL(y, \gamma; \theta) = logf(y \mid \gamma; \theta_1) + logf(\gamma; \omega), \tag{3.13}$$

which is the sum of the conditional log-likelihood of the observed data and the log-likelihood of the missing data.[6] Using assumption (3.8), the joint log-likelihood of the vector of missing data can be written as

$$logf(\gamma) = \sum_{i=1}^{N} logf(\gamma_i) = \mu_1 + \frac{N}{2} log \mid \triangle^{-1} \mid - \frac{1}{2} \sum_{i=1}^{N} \gamma_i' \triangle^{-1} \gamma_i. \tag{3.14}$$

We now derive the likelihood of $y = (y_1', .., y_N')'$ given γ.[7] From (3.6) we can easily obtain 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 $var(y_i \mid \gamma_i) = var\,(\varepsilon_i) = R_i = \sigma^2_{\varepsilon_i} I_T$, respectively. Under the assumption that both the regression error terms, ε_i, and the random effects, 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-

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

[7] If the model is dynamic, we derive the joint likelihood of y_i given γ_i, for $i = 1, .., N$, under the assumption that the first observations of the dependent variables are deterministic. 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 maximum likelihood estimator (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 $\mid \phi \mid > 1$.

likelihood of the observed data is given by

$$log f(y \mid \gamma) = \sum_{i=1}^{N} log f(y_i \mid \gamma_i) = \mu_2 - \frac{1}{2}\sum_{i=1}^{N} log \mid R_i \mid - \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i' R_i^{-1}\varepsilon_i, \quad (3.15)$$

where

$$\varepsilon_i = y_i - W_i\bar{\Gamma} - \bar{Z}_i\gamma_i. \quad (3.16)$$

Having found an explicit formulation for $log f(y \mid \gamma; \theta_1)$ and $log f(\gamma; \omega)$, we can derive an expression for the log-likelihood of the complete data by substituting (3.15) and (3.14) into (3.13). 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 $log f(\gamma; \omega)$ does not contain any information about θ_1 and similarly $log f(y \mid \gamma; \theta_1)$ does not contain any information about ω. Second, as stated in Harville (1977), the maximum likelihood estimation takes no account of the loss in degrees of freedom that results from estimating the fixed coefficients, leading to a biased estimator of σ_ε^2. In the next subsection, we eliminate this problem by using the restricted maximum likelihood (REML) approach, described formally by Patterson and Thompson (1971).

3.3.1 Restricted Likelihood

Following Patterson and Thompson (1971), we can separate $log f(y_i \mid \gamma_i; \theta_1)$ in two parts: L_{1i} and L_{2i}. By maximizing the former, we can estimate $\sigma_{\varepsilon_i}^2$. An estimate of $\bar{\Gamma}$ is obtained after maximizing L_{2i}. The two parts can be obtained by defining two matrices S_i and Q_i such that the likelihood of $(y_i \mid \gamma_i)$ (for $i = 1, .., N$) can be decomposed as the product of the likelihoods of $S_i y_i$ and $Q_i y_i$, i.e.

$$log f(y_i \mid \gamma_i; \theta_1) = L_{1i} + L_{2i}. \quad (3.17)$$

Such matrices must satisfy the following properties: (i) the rank of S_i is not greater than $T - \underline{K}$, while Q_i is a matrix of rank $\underline{K}$, (ii) L_{1i} and L_{2i} are statistically independent, i.e. $cov\left(S_i y_i, Q_i y_i\right) = 0$, (iii) the matrix S_i is chosen so that $E\left(S_i y_i\right) = 0$, i.e. $S_i W_i = 0$, and (iv) the matrix $Q_i W_i$ has rank $\underline{K}$.[8]

Finding an expression for L_{1i}. Premutiplying both sides of (3.6) by S_i, we have $E\left(S_i y_i \mid \gamma_i\right) = S_i \bar{Z}_i \gamma_i$, since $S_i W_i = 0$ and $var\left(S_i y_i \mid \gamma_i\right) = S_i R_i S_i'$. Therefore, the conditional log-likelihood of $S_i y_i$ is given by

$$L_{1i} = \mu_3 - \frac{1}{2} log \mid S_i R_i S_i' \mid - \frac{1}{2}\left(y_i - \bar{Z}_i \gamma_i\right)' S_i' \left(S_i R_i S_i'\right)^{-1} S_i \left(y_i - \bar{Z}_i \gamma_i\right). \quad (3.18)$$

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.[9] Indeed, equation (3.18) can be rewritten as

$$L_{1i} = \mu_3 - \frac{1}{2} log|R_i| - \frac{1}{2} log \mid W_i' R_i^{-1} W_i| - \frac{1}{2} \bar{\varepsilon}_i' R_i^{-1} \bar{\varepsilon}_i, \quad (3.19)$$

where $\bar{\varepsilon}_i = y_i - W_i \hat{\bar{\Gamma}} - \bar{Z}_i \gamma_i$, and $\hat{\bar{\Gamma}}$ denotes the generalized least squares (GLS) estimator of $\bar{\Gamma}$, which we describe in Subsection 3.4.4.

Finding an expression for L_{2i}. Following Patterson and Thompson (1971), we can set $Q_i = W_i' R_i^{-1}$ since it satisfies $cov\left(S_i y_i, Q_i y_i\right) = 0$.[10] After premutiplying both sides of (3.6) by Q_i, we have $E\left(Q_i y_i \mid \gamma_i\right) = W_i' R_i^{-1}\left(W_i \bar{\Gamma} + Z_i \gamma_i\right)$ and $var\left(Q_i y_i \mid \gamma_i\right) = W_i' R_i^{-1} W_i$. The log-likelihood of $Q_i y_i \mid \gamma_i$ is given by

$$L_{2i} = \mu_4 - \frac{1}{2} log \mid W_i' R_i^{-1} W_i \mid - \frac{1}{2} \varepsilon_i' H_i \varepsilon_i, \quad (3.20)$$

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 errors defined in (3.16).

[8] $\underline{K} = rank(W_i)$.
[9] Detailed derivations of L_{1i} are described in Appendix 3.10.1.
[10] See Appendix 3.10.1 for a proof.

3.4 EM-Algorithm

3.4.1 Generalities

Using equations (3.13), (3.14) and (3.15), the log-likelihood of the complete data can be rewritten as

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

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 $var\{E(\gamma_i \mid y_i)\}$ as an estimator of $var(\gamma_i)$, the unconditional variance of the γ_i, the EM algorithm yields an estimator of the latter satisfying the law of total variance. Finally, the EM allows us to make inference on the random effects' population.

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

$$logL(y_i,\gamma_i;\theta) = L_{1i} + L_{2i} + logf(\gamma_i;\omega_i), \tag{3.21}$$

for $i = 1,..,N$, where $logf(y_i \mid \gamma_i;\theta_1)$ has been decomposed as shown in equation (3.17).

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

$$\begin{aligned} Q = Q(\theta; \theta^{(b-1)}) &= E_{\theta^{(b-1)}} \{logL(y, \gamma; \theta) \mid y\} \\ &= \textstyle\sum_{i=1}^{N} E_{\theta^{(b-1)}} \{logL(y_i, \gamma_i; \theta) \mid y_i\} = \textstyle\sum_{i=1}^{N} Q_i, \end{aligned} \tag{3.22}$$

where

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

and

$$\begin{aligned} Q_{1i} &= E_{\theta^{(b-1)}} \{L_{1i} \mid y_i\}, \\ Q_{2i} &= E_{\theta^{(b-1)}} \{L_{2i} \mid y_i\}, \\ Q_{3i} &= E_{\theta^{(b-1)}} \{logf(\gamma_i; \omega) \mid y_i\}. \end{aligned} \tag{3.23}$$

In practice, we replace the missing variables, i.e. the random effects (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

$$\theta^{(b)} = \underset{\theta}{argmax}\, Q(\theta; \theta^{(b-1)}).$$

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.

3.4.2 Best Linear Unbiased Prediction

Within the EM algorithm, the random effects, γ_i, are estimated by best linear unbiased prediction (BLUP).[11] 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. The conditional expectation of γ_i given the data is

$$\begin{aligned} \hat{\gamma}_i = E\left(\gamma_i \mid y_i\right) &= \triangle \bar{Z}_i' \left(\bar{Z}_i \triangle \bar{Z}_i' + R_i\right)^{-1} \left(y_i - W_i \bar{\Gamma}\right) \\ &= \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)^{-1} \bar{Z}_i' R_i^{-1} \left(y_i - W_i \bar{\Gamma}\right), \end{aligned} \tag{3.24}$$

which is also the argument that maximizes the complete data likelihood, as defined in (3.13), with respect to γ_i. It can be noted from the first equality of (3.24) that this expression is related to the predictor of the random effects derived in Lee and Griffiths (1979), Lindley and Smith (1972) and Smith (1973). The main difference concerns the way the regression coefficients and the variances components are estimated.

The conditional variance of γ_i is given by

$$V_{\gamma_i} = var\left(\gamma_i \mid y_i\right) = \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)^{-1}, \tag{3.25}$$

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

These two formulae have an empirical Bayesian interpretation. Given that γ is random, the likelihood $f(\gamma)$ can be thought as the "prior" density of γ. The posterior distribution of the latter is Normal with mean and variance given by (3.24) and (3.25), respectively.

[11] Further details are provided in Appendix 3.10.2.

3.4.3 E-step

At each iteration, the E-step requires the calculation of the conditional expectation of (3.21) given the observed data and the current fit for the parameters, to obtain an expression for $Q_i(\theta)$, for $i = 1, .., N$.[12]

To obtain Q_{1i}, we take conditional expectation of both sides of (3.19). 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

$$\begin{aligned} Q_{1i} = E_{\theta^{(b-1)}}\left(L_{1i} \mid y_i\right) &= \mu_3 - \tfrac{1}{2} log|R_i| - \tfrac{1}{2} log \mid W_i' R_i^{-1} W_i| \\ &\quad -\tfrac{1}{2} Tr\left(\bar{Z}_i' R_i^{-1} \bar{Z}_i V_{\gamma_i}^{(b)}\right) - \tfrac{1}{2} \hat{\bar{\varepsilon}}_i' R_i^{-1} \hat{\bar{\varepsilon}}_i. \end{aligned} \tag{3.26}$$

where $\hat{\gamma}_i^{(b)}$ and $V_{\gamma_i}^{(b)}$ are given by (3.24) and (3.25) respectively, after substituting the current fit for θ at each iteration $b = 1, 2, ...$

To obtain Q_{2i}, we take the conditional expectation of (3.20). 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

$$\begin{aligned} Q_{2i} = E_{\theta^{(b-1)}}\left(L_{2i} \mid y_i\right) &= \mu_4 - \tfrac{1}{2} log \mid W_i' R_i^{-1} W_i \mid \\ &\quad -\tfrac{1}{2} Tr\left(\bar{Z}_i' H_i \bar{Z}_i V_{\gamma_i}^{(b)}\right) - \tfrac{1}{2} \hat{\varepsilon}_i' H_i \hat{\varepsilon}_i. \end{aligned} \tag{3.27}$$

Finally, substituting

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

[12] Detailed computations are shown in Appendix 3.10.3.

into $E_{\theta^{(b-1)}}\left\{logf(\gamma_i) \mid y_i\right\}$, yields

$$Q_{3i} = E_{\theta^{(b-1)}}\left(logf(\gamma_i) \mid y\right) = -\frac{K^*}{2}log2\pi + \frac{1}{2}log \mid \triangle^{-1} \mid -\frac{1}{2}Tr\left(\triangle^{-1}V_{\gamma_i}^{(b)}\right) - \frac{1}{2}\hat{\gamma}_i^{(b)\prime}\triangle^{-1}\hat{\gamma}_i^{(b)}. \tag{3.28}$$

3.4.4 M-step

The M-Step consists in maximizing (3.22) with respect to the parameters of interest, contained in θ.

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

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

The solution is

$$\bar{\Gamma}^{(b)} = \left(\sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1} W_i\right)^{-1} \sum_{i=1}^{N} W_i' R_{i(b-1)}^{-1}\left(y_i - \bar{Z}_i\hat{\gamma}_i^{(b)}\right). \tag{3.29}$$

which corresponds 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.

Estimation of the Variances of the Error Terms. An estimate of $\sigma_{\varepsilon_i}^2$ can be derived by maximizing (3.22). Because Q_{3i} is not a function of $\sigma_{\varepsilon_i}^2$ and given that no information is lost by neglecting Q_{2i} (as noted by Patterson and Thompson (1971), and Harville (1977)), we base inference for $\sigma_{\varepsilon_i}^2$ only on Q_{1i}, which is defined in (3.26).

Substituting $R_i = var\left(\varepsilon_i\right) = \sigma_{\varepsilon_i}^2 I_T$ into (3.26) and equating the first derivative of the latter with respect to $\sigma^2_{\varepsilon_i}$ to zero, yields

$$\sigma^{2(b)}_{\varepsilon_i} = \frac{\hat{\hat{\varepsilon}}_i'\hat{\hat{\varepsilon}}_i + Tr\left(\bar{Z}_i'\bar{Z}_i V^{(b)}_{\gamma_i}\right)}{T - r(W_i)}, \tag{3.30}$$

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

Estimation of the Random Coefficient Variance-Covariance Matrix. Under the law of total variance, the unconditional variance of γ_i can be written as

$$\begin{aligned} \triangle = var\left(\gamma_i\right) &= var\left[E\left(\gamma_i \mid y_i\right)\right] + E\left[var\left(\gamma_i \mid y_i\right)\right] \\ &= var\left(\hat{\gamma}_i\right) + E\left(V_{\gamma_i}\right). \end{aligned} \tag{3.31}$$

Therefore, it can be shown that

$$\hat{\triangle} = \frac{1}{N}\sum_{i=1}^{N}\left\{\hat{\gamma}_i\hat{\gamma}_i' + V_{\gamma_i}\right\} \tag{3.32}$$

is an unbiased estimator of $\triangle$. Indeed, taking expectation of both sides of (3.32) and using (3.31), we get

$$E\left(\hat{\triangle}\right) = \frac{1}{N}\sum_{i=1}^{N}\left\{E\left(\hat{\gamma}_i\hat{\gamma}_i'\right) + E\left(V_{\gamma_i}\right)\right\} = \frac{1}{N}\sum_{i=1}^{N}\left\{var\left(\hat{\gamma}_i\right) + E\left(V_{\gamma_i}\right)\right\} = \triangle.$$

Notably, the EM estimator of the variance-covariance matrix of the random effects (which is the argument which maximizes (3.28) with respect to $\triangle$) is equal to

$$\triangle^{(b)} = \frac{1}{N}\sum_{i=1}^{N}\left\{\hat{\gamma}^{(b)}_i\hat{\gamma}^{(b)\prime}_i + V^{(b)}_{\gamma_i}\right\}, \tag{3.33}$$

which is equivalent to (3.32) after substituting the unknown parameters with their current fit in the EM algorithm.[13]

3.4.5 EM-REML Algorithm - Complete Iterations

The EM algorithm steps can be summarised as follows. We start with some initial guess: $\psi^{(0)}$, $\triangle_{(0)}$ and $R_{i(0)} = \sigma^{2(0)}_{\varepsilon_i} I_{T-p}$. We suggest using Swamy (1970) estimates, which are reported in the next Section, since they are consistent estimators of the average effects and the variance components. 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 $E_{\theta^{(b-1)}}\left(\gamma_i \mid y_i\right)$, which are given by

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

   respectively.

2. The average coefficients are given by

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

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

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

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

$$\triangle^{(b)} = \frac{1}{N} \sum_{i=1}^{N} \left\{V^{(b)}_{\gamma_i} + \hat{\gamma}^{(b)}_i \hat{\gamma}^{(b)\prime}_i\right\}.$$

[13]See Appendix 3.10.4 for computations.

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.

3.5 Comparison between EM-REML Estimation and Alternative Methods

In this section, we review some of the existing sampling and Bayesian methods commonly used to estimate heterogeneous panel data, to highlight similarities and differences with the EM-REML approach.

3.5.1 Average Effects

Following Searle (1978, eq. 3.17), representing (3.29) and (3.24) as a system of two equations, we can rewrite these two formulae as

$$\hat{\hat{\Gamma}} = \left(\sum_{i=1}^{N} W_i' V_i^{-1} W_i \right)^{-1} \sum_{i=1}^{N} W_i' V_i^{-1} y_i, \tag{3.34}$$

$$\hat{\gamma}_i = \triangle \bar{Z}_i' V_i^{-1} \left(y_i - W_i \hat{\hat{\Gamma}} \right), \tag{3.35}$$

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 (3.34) is related to the Swamy GLS estimator. The latter can be rewritten as a weighted average of the least squares estimates of the individual units:

$$\hat{\hat{\Gamma}} = \sum_{i=1}^{N} \Psi_i \hat{\psi}_{i,ols}, \tag{3.36}$$

where

$$\begin{aligned}\Psi_i &= \left\{\sum_{i=1}^{N}[\triangle + \sigma^2_{\varepsilon_i}(\bar{Z}_i'\bar{Z}_i)^{-1}]^{-1}\right\}^{-1}[\triangle + \sigma^2_{\varepsilon_i}(\bar{Z}_i'\bar{Z}_i)^{-1}]^{-1}, \\ \hat{\psi}_{i,ols} &= (\bar{Z}_i'\bar{Z}_i)^{-1}\bar{Z}_i'y_i.\end{aligned} \tag{3.37}$$

Swamy's estimator is a two-step procedure, which requires first to estimate N time series separately as if the individual coefficients were fixed (in the sense that they are not realizations from a common distribution) and all different in each cross-section. Instead, the EM-REML is an iterative method which shrinks the unit-specific parameters towards a common mean. 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, Pesaran and Tahmiscioglu (1999) show that $\hat{\bar{\Gamma}}$ is equivalent to 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 Ω, with $\Omega^{-1} = 0$. Another important contribution of the aforementioned paper is to establish that the Bayes estimator $\hat{\bar{\Gamma}}$ is asymptotically equivalent to the mean group estimator proposed by Pesaran and Smith (1995), as $T \to \infty$, $N \to \infty$, and $\sqrt{N}/T \to 0$.

3.5.2 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 (3.34) and (3.24) into (3.5) yields the best linear unbiased predictor of ψ_i which following Lee and Griffiths (1979), can be rewritten as

$$\begin{aligned} \hat{\psi}_i &= \hat{\bar{\Gamma}} + \hat{\gamma}_i \\ &= \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)^{-1} \left(\left(\bar{Z}_i' R_i^{-1} \bar{Z}_i\right) \hat{\psi}_{i,ols} + \triangle^{-1} \hat{\bar{\Gamma}}\right). \end{aligned} \tag{3.38}$$

The latter expression is also related to the empirical Bayes estimator of ψ_i, described in Maddala et al. (1997). The EM-REML predictor of ψ_i is thus a weighted average between the OLS estimator of ψ_i and the estimator of the overall mean, $\bar{\Gamma}$, given by (3.34). Interestingly, as shown in Smith (1973), the latter can be rewritten as a simple average of the $\hat{\psi}_i$:

$$\hat{\bar{\Gamma}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\psi}_i. \tag{3.39}$$

Mean Group and Shrinkage Estimators. When the time dimension is large enough (relative to the number of parameters to be estimated), it is sensible to estimate a different time-series model for each unit, as proposed by Pesaran and Smith (1995). Besides its simplicity, one strong advantage of their Mean Group (MG) estimator is that it does not require to impose any assumption on the distribution of the unit-specific coefficients. However, a drawback of the MG estimation is that it may perform rather poorly when either N or T are small (Hsiao, Pesaran and Tahmiscioglu, 1999). Moreover, as noted in Smith and Fuertes (2016), the MG estimator is very sensitive to outliers. Boyd and Smith (2002) find that the weighting which the Swamy estimator applies, may not suffice to reduce this problem. To overcome the latter, one could either consider robust versions which trim the outliers to minimize their effect, or shrinkage methods. 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. 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. Our estimation method belongs to the class of shrinkage estimators. In fact, the unobserved idiosyncratic components of the random coefficients, γ_i, are estimated by BLUP. This choice arises naturally in the EM algorithm, and in some applications may be advantageous compared to estimating N time series separately since BLUP estimates 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. Shrinkage approaches can be seen as an intermediate strategy between heterogeneous models (which avoid bias) and pooled methods (which allow for efficiency gains), and therefore might help reducing the trade-off between bias and efficiency discussed in Baltagi, Bresson and Pirotte (2008). As shown in the Monte Carlo analysis, as $T \to \infty$ (for fixed N) the difference between the Swamy, the MG, and the EM-REML estimators goes to zero. Finally, our approach can be advantageous (i) when individual-specific characteristics which do not vary over time enter the regression equation, and (ii) when the interest lies in explaining the drivers of coefficients heterogeneity. In the first case, computing the OLS estimates for each unit is not feasible. In the second case, if N is large one could first estimate N time series separately and in a second step regress the OLS estimates on a set of unit-specific characteristics. Instead, our likelihood approach does not require N to be very large.

3.5.3 Variance Components

We now compare the EM-REML estimator of the random coefficient variance-covariance matrix, given by (3.33), with the Swamy (1970) and Lee and Griffiths (1979) estimators. Swamy suggested estimating $var\left(\gamma_i\right)$ as

$$\hat{\triangle}_S = \hat{\triangle}_{S_1} - N^{-1} \sum_{i=1}^{N} v\hat{a}r\left(\hat{\psi}_{i,ols}\right), \tag{3.40}$$

where

$$\hat{\triangle}_{S_1} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\hat{\psi}_{i,ols} - N^{-1} \sum_{i=1}^{N} \hat{\psi}_{i,ols} \right) \left(\hat{\psi}_{i,ols} - N^{-1} \sum_{i=1}^{N} \hat{\psi}_{i,ols} \right)', \quad (3.41)$$

$\hat{\psi}_{i,ols}$ are obtained by estimating N time series separately by OLS, $v\hat{a}r\left(\hat{\psi}_{i,ols}\right) = \hat{\sigma}^2_{\varepsilon_i}(\bar{Z}_i'\bar{Z}_i)^{-1}$, and

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{T-K^*}\left(y_i - \bar{Z}_i\hat{\psi}_{i,ols}\right)'\left(y_i - \bar{Z}_i\hat{\psi}_{i,ols}\right) \quad (3.42)$$

are the OLS estimated variances of the error terms. However, (3.40) is not necessarily nonnegative definite. Therefore, if that is the case the author suggests considering only (3.41). 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. Unfortunately, this estimator can be severely biased in finite sample. Another drawback of (3.40) is that it is subject to large discontinuities.[14] As shown in the Monte Carlo analysis, the root mean square errors of this estimator can be quite large. To understand, note that the estimator to be used in practical applications can be rewritten as

$$\hat{\hat{\triangle}} =]\!\!] \left(\hat{\triangle} > 0\right)\hat{\triangle} +]\!\!] \left(\hat{\triangle} \leq 0\right)\hat{\triangle}_{S_1},$$

where $]\!\!]\,(A) = 1$ if event A occurs. Focusing on the kth diagonal element, and assuming for illustrative purposes that

$$\hat{\triangle}_{S_1,k} = 2, \qquad v\bar{\hat{a}}r\left(\hat{\psi}_{ik}\right) = \left\{N^{-1}\textstyle\sum_{i=1}^{N} v\hat{a}r\left(\hat{\psi}_{ik}\right)\right\} \in \{1,2,3,4\},$$

[14] I am grateful to Ron Smith who pointed out this issue in a meeting.

we have

$$\hat{\hat{\triangle}}_k = \begin{cases} 2 & if \quad v\bar{\hat{a}}r\left(\hat{\psi}_{ik}\right) \in \{2,3,4\} \\ 1 & if \quad v\bar{\hat{a}}r\left(\hat{\psi}_{ik}\right) = 1 \end{cases}$$

When the variances are unknown, Lee and Griffiths (1979) suggest maximizing the joint likelihood of the random coefficients and the observed data given in (3.13) with respect to the unknown parameters of the model, to get the following iterative solutions of the variance components:[15]

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{T}\left(y_i - \bar{Z}_i\hat{\psi}_i\right)'\left(y_i - \bar{Z}_i\hat{\psi}_i\right), \tag{3.43}$$

where $\hat{\psi}_i$ is given by (3.38), and

$$\hat{\triangle}_{LG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_i\hat{\gamma}_i'. \tag{3.44}$$

Within the EM algorithm, the random effects, 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 effects, given in (3.13). 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 3.4, the estimator given by (3.44) 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, since

$$E\left(\hat{\triangle}_{LG}\right) \leq E\left(\hat{\triangle}_{EM}\right) \equiv \triangle \leq E\left(\hat{\triangle}_{S_1}\right), \tag{3.45}$$

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

where

$$\hat{\triangle}_{EM} = \frac{1}{N}\sum_{i=1}^{N}\{V_{\gamma_i} + \hat{\gamma}_i\hat{\gamma}_i'\} \tag{3.46}$$

is the maximum likelihood estimator obtained by applying the EM algorithm. Result (3.45) is of relevance because $\triangle$ appears not only in both the formula for the average effect and the predicted random effects but also in their standard errors. Testing hypothesis crucially depends on correctly estimating the random coefficient 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

$$\hat{\sigma}^2_{\varepsilon_i} = \frac{1}{T + \upsilon_i + 2}\left\{\upsilon_i\lambda_i + \left(y_i - \bar{Z}_i\hat{\psi}_i\right)'\left(y_i - \bar{Z}_i\hat{\psi}_i\right)\right\}, \tag{3.47}$$

$$\bar{\triangle} = \frac{1}{N + \rho - K^* - 2}\left\{\Upsilon + \sum_{i=1}^{N}\hat{\gamma}_i\hat{\gamma}_i'\right\}, \tag{3.48}$$

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 χ^2 with prior parameters υ_i and λ_i, and is independent of $\triangle$. Note from (3.38) that $\hat{\gamma}_i = \hat{\psi}_i - \hat{\bar{\Gamma}}$. Smith (1973) suggests vague priors by setting $\rho = 1$ and Υ to be a diagonal matrix with small positive entries (such as .001). We note that, by setting $\rho = K^* + 2$, $\upsilon_i = -r\left(W_i\right) - 2$ and $\upsilon_i\lambda_i = Tr\left(\bar{Z}_i'\bar{Z}_i\Upsilon\right)$, we can draw an analogy between the EM-REML estimates, given by (3.46) and (3.30), and the modes of the posterior distributions of $\triangle$ and $\sigma^2_{\varepsilon_i}$, given by (3.48) and (3.47), respectively.

3.5.4 Comparison between EM and a Full Bayesian Implementation

We can now compare the EM approach to the Bayesian estimation. The EM algorithm gives a probability distribution over the random effects, 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. The advantage of the EM compared to the iterative Bayesian approach developed by Lindley and Smith (1992) and the Gibbs sampling-based approach suggested in Hsiao, Pesaran and Tahmiscioglu (1999), would be that there is no need to specify prior means and variances, the choice of which may not be always obvious. At the same time, 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. 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, in the context of random coefficient models the choice of Swamy (1970) estimates as starting values is rather natural, as they are consistent parameter estimates.

Moreover, 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 (e.g. Hobert and Casella (1996)). For instance, assuming that $\triangle^{-1}$ has a Wishart distribution with scale matrix $(\rho\Upsilon)$ and ρ degrees of freedom, Hsiao, Pesaran and Tahmiscioglu (1999) note that the bias of both the empirical and hierarchical Bayes estimators of the regression coefficients is sensitive to the specification of the prior scale matrix. Being unable to use a diffuse prior for the covariance matrix, which would cause their Gibbs algorithm to break down, they set $\Upsilon = \hat{\triangle}_S$, the Swamy estimator of the random coefficient covariance matrix. If the latter is negative definite, the consistent (but biased) version (3.41) must be used, affecting the Bayes estimates of the regression coefficients adversely.

Finally, it is known that the EM algorithm may converge slowly. However, in the context of random coefficient models, convergence is usually achieved almost as quickly as in the Gibbs sampler.[16]

3.6 Hypothesis Testing

3.6.1 Inference for Fixed Coefficients

Covariance Matrix of the Estimator of the Fixed Coefficients. Unlike the Newton-Raphson and related methods, the EM algorithm does not automatically provide an estimate of the covariance matrix of the maximum likelihood estimates. However, in the context of the random coefficient type models here considered, 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 $(log f(y;\theta))$ 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 the inverse of the Fisher information matrix, given by

$$\hat{\Phi} = \left(\sum_{i=1}^{N} W_i' V_{i(B)}^{-1} W_i \right)^{-1}, \tag{3.49}$$

where $V_i = var(y_i) = \bar{Z}_i \triangle \bar{Z}_i' + R_i$, while B denotes the last iteration of the EM algorithm.

[16]For instance, in the panel model used in the application, with $N = 38$ and $60 \leq T_i \leq 87$, and $K = 8$ regressors, including the constant, the EM algorithm converges after around 17 seconds. The Gibbs sampling algorithm is quicker, requiring around 10 seconds to run 5000 iterations. If we increase the number of regressors to 20, the difference slightly increases, with the EM algorithm and the Gibbs sampler requiring around 40 and 15 seconds, respectively. Despite being slower than its Bayesian counterpart, the EM algorithm converges rather quickly.

Adjusted Estimator of the Covariance Matrix of Fixed Coefficients. Let $\tilde{\bar{\Gamma}} = \bar{\Gamma}^{(B)}$ be the "feasible" estimator of $\bar{\Gamma}$ obtained by substituting the unknown parameters with their estimates into the "infeasible" estimator $\hat{\bar{\Gamma}}$, given by equation (3.34). We define $\Phi = var\left(\hat{\bar{\Gamma}}\right)$, which is a function of $\vartheta = \left(\omega', \sigma_{\varepsilon}^{2'}\right)'$, the $\bar{r} \times 1$ vector of variance-covariance parameters of the model.

We note that $\hat{\Phi} = \Phi\left(\hat{\vartheta}\right)$ is a biased estimator of $var\left(\tilde{\bar{\Gamma}}\right)$. The literature on linear mixed models offers good insights into the two main sources of this bias. First, $\Phi\left(\vartheta\right)$ takes no account of the variability of $\hat{\vartheta}$ in $\tilde{\bar{\Gamma}}$. This problem was addressed by Kackar and Harville (1984). Second, $\hat{\Phi}$ underestimates Φ, as shown by Kenward and Roger (1997). The solution provided by the latter can be easily applied into our setting to obtain an estimator of $var\left(\tilde{\bar{\Gamma}}\right)$, $\hat{\Phi}_A$, which incorporates the necessary adjustments to correct both form of bias.[17]

Hypothesis Testing of Average Effects. To test the hypothesis $\bar{\Gamma} = \bar{\Gamma}_0$, for $\bar{\Gamma}_0$ a known $\bar{K} \times 1$ vector, we use the following criterion suggested by Swamy (1970):

$$\frac{N - \bar{K}}{\bar{K}\left(N-1\right)}\left(\tilde{\bar{\Gamma}} - \bar{\Gamma}\right)' \hat{\Phi}_A^{-1}\left(\tilde{\bar{\Gamma}} - \bar{\Gamma}\right), \tag{3.50}$$

whose asymptotic distribution is F, with $\bar{K}$, $N - \bar{K}$ degrees of freedom.

3.6.2 Assessing the Precision for the Unit-Specific Coefficients

In the general case, the standard errors of the predictor of ψ_{1i} can be computed as the square root of the diagonal elements of

$$var\left(\hat{\psi}_{1i} - \psi_{1i}\right) = F_{1i}\Phi F_{1i}' + var\left(\hat{\gamma}_i - \gamma_i\right) - F_{1i}\Lambda - \Lambda' F_{1i}', \tag{3.51}$$

[17] Details to compute $\hat{\Phi}_A$ are given in Appendix 3.10.5. A Matlab code to obtain the latter is also provided.

where

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

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

and $\Phi = var(\hat{\bar{\Gamma}})$ as defined in (3.49).[18]

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

$$\gamma_i^{(s)} \mid y_i \sim N\left(\hat{\gamma}_i, V_{\gamma_i}\right), \tag{3.52}$$

where $\hat{\gamma}_i$ and V_{γ_i} are given by (3.24) and (3.25) 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 the empirical Bayesian methods, $\bar{\Gamma}$ can be drawn from its posterior distribution given by

$$\bar{\Gamma}^{(s)} \mid y \sim N\left(\hat{\bar{\Gamma}}, \hat{\Phi}\right), \tag{3.53}$$

where $\hat{\bar{\Gamma}}$ and Φ are given by (3.34) and (3.49) respectively. It follows that the individual coefficients, as defined in (3.5) can be drawn from the following Gaussian distribution:

$$\psi_{1i}^{(s)} \mid y_i \sim N\left(F_{1i}\hat{\bar{\Gamma}} + \hat{\gamma}_i, F_{1i}\Phi F_{1i}' + V_{\gamma_i}\right), \tag{3.54}$$

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

3.7 Monte Carlo Simulations

In this section, we employ Monte-Carlo experiments to examine and compare the finite sample properties of the proposed EM-REML method, the Swamy's

[18]Expression (3.51) is equivalent to the one proposed by Lee and Griffiths (1979). See Appendix 3.10.5 for further details. The estimator of ψ_i derived in equation (3.38) has been obtained under the assumption that $F_{1i} = I$, $\forall i$. Therefore, its standard errors can also be obtained from (3.51) after substituting $F_{1i} = I$, $\forall i$.

random coefficient model, and the Mean Group (MG) estimation. We report results on the bias and root mean square error (RMSE) of the average effects and of the variance components of the model. Particular attention is also paid to the accuracy of the estimated standard errors and to the power performances of the estimators.

3.7.1 Data Generating Process

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

$$\begin{aligned} y_{it} &= c_i + \beta_i x_{it} + \phi_i y_{it-1} + \varepsilon_{it}, \\ x_{it} &= c_{x,i}(1-\rho) + \rho x_{it-1} + u_{it}, \end{aligned} \tag{3.55}$$

where

$$\begin{aligned} \varepsilon_{it} &\sim i.i.d.N(0, \sigma^2_{\varepsilon_i}), \\ u_{it} &\sim i.i.d.N(0,1), \\ c_{x,i} &\sim i.i.d.N\left(1,1\right). \end{aligned} \tag{3.56}$$

The sample sizes considered are $N = \{30, 50\}$ and $T = \{10, 20, 30, 40, 50, 60, 80, 100\}$. We set $\rho = 0.6$. Once generated, the x_{it} are taken as fixed across different replications. The variances of the time-varying disturbances are generated from $\sigma^2_{\varepsilon_i} = (\zeta \bar{x}_i)^2$, where $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$, and $\zeta = 0.5$. The coefficients differ randomly across units according to

$$\begin{aligned} c_i &= c + \gamma_{1i}, \\ \beta_i &= \beta + \gamma_{2i}, \\ \phi_i &= \phi + \gamma_{3i}, \end{aligned} \tag{3.57}$$

where $\psi = (c, \beta, \phi) = (0, 0.1, 0.5)$. Moreover, we assume that $\gamma_{ji} \sim i.i.d.N(0, \sigma^2_{\gamma_j})$, for $j = 1, 2, 3$. We set $\sigma_{\gamma_1} = 0.1$, and $\sigma_{\gamma_2} = 0.224$. We choose $\sigma_{\gamma_3} = 0.07$ in order to avoid explosive behaviour. Under these settings the median signal-to-noise ratio corresponding to the slope parameters $(\sigma^2_{\gamma_2}/\sigma^2_i)$ for $N = 30$ and averaged across the different T cases, is equal to 0.1950.[19]

The initial values for the dependent variables are generated from

$$y_{i0} = \bar{\theta}_{i0} + \upsilon_{i0},$$

for $i = 1, .., N$, where $\upsilon_{i0} \sim N(0, \sigma^2_\upsilon)$, and

$$\begin{aligned} \bar{\theta}_{i0} &= E(y_{i0} \mid \gamma_i) = \textstyle\sum_{s=0}^{\infty} \phi_i^s x_{i,-s}\beta_i + \frac{c_i}{1-\phi_i}, \\ \sigma^2_\upsilon &= var(y_{i0} \mid \gamma_i) = var\left\{\textstyle\sum_{s=0}^{\infty} \phi_i^s \varepsilon_{i,-s}\right\} \\ &= \frac{\sigma^2_{\varepsilon_i}}{1-\phi_i^2}. \end{aligned}$$

In practice, we consider only a finite number of $x_{i,-s}$. For each i, we generate 10 observations $(x_{i0}, .., x_{i,-9})$ given that when $\mid \phi_i \mid < 1$, the contribution of earlier observations is quite low. The vector $(x_{i0}, .., x_{i,-9})$ is not used for estimation and inference.

3.7.2 Monte Carlo Results

In this subsection, we describe the results based on 500 replications. Table 3.3 reports the bias and the root mean square errors (RMSE) of the EM-REML estimators of the average effects and of the variances of the random coefficients, as well as the standard errors of such biases, for $N = 30$ and $T = \{10, 20, 30, 40, 50, 60, 80, 100\}$.[20] An overall measure of the bias of the estimated average coefficients (which is chosen to be the Euclidean norm of the bias of ψ), and two measures of the accuracy of the estimated standard errors are also given. Table 3.4, and 3.5 describe the results for Swamy (1970), and the MG estimator, respectively.

[19] The cross-section average, computed as $N^{-1}\sum_{i=1}^{N} \sigma^2_{\gamma_2}/\sigma^2_i$, and averaged across the different T cases, is higher and equal to 10.63, partly due to the fact that some of the draws of $\sigma^2_{\varepsilon_i}$ are smaller than $\sigma^2_{\gamma_2}$.

[20] Similar results hold for $N = 50$, which we do not report here.

Using the data simulated from the DGP described in the previous subsection, we find that the EM-REML approach does quite well even when the sample size is not too large. In many cases, it outperforms both Swamy and the MG estimator in term of bias of both the average effects and the variance components. For any time dimension, the REML estimators of the average coefficients and the variance components obtained applying the EM algorithm have smaller RMSE than the MG one. The RMSE of the EM-REML estimators are also smaller than the Swamy one, unless T is quite large, in which case they almost coincide.

The bias of the EM-REML estimator of the common intercept is equal to 0.0015 when $T = 10$, and to 0.0005 when $T = 20$. When $T = 100$, the bias amounts to -0.0007. In most of the cases, it is smaller than the bias of Swamy and the MG estimators, and it has lower RMSE.

Regarding the slope coefficient associated to x_{it}, the bias of the EM-REML estimator is equal to -0.0033 when $T = 10$, which amounts to -3.3 percent of the true value. When $T = 20$, the bias reduces to 1.7 percent of the true value till becoming equal to 0.1 percent when $T = 100$. In some cases, the EM-REML estimator may have a slightly larger bias than the Swamy one but in all cases it has a smaller or at most equal RMSE.

The advantages in term of bias of the EM-REML approach are even more notable when considering the autoregressive coefficient. For instance, when $T = 10$, the bias of the EM-REML estimator is equal to 0.0408, which is equivalent to 8.16% of the true value. The biases of Swamy GLS and the MG estimators of the autoregressive coefficient, when $T = 10$, are larger and equal to -22.6% and -41.44% of the true value, respectively. As expected, the bias reduces as T increases. Monte Carlo experiments corroborate Maddala et al. (1997) argument in favor of iterative procedures to two-step estimators when the model is dynamic and confirm Hsiao, Pesaran and Tahmiscioglu (1999) finding that the MG estimator is unlikely to be an appropriate estimator when either N or T are small.

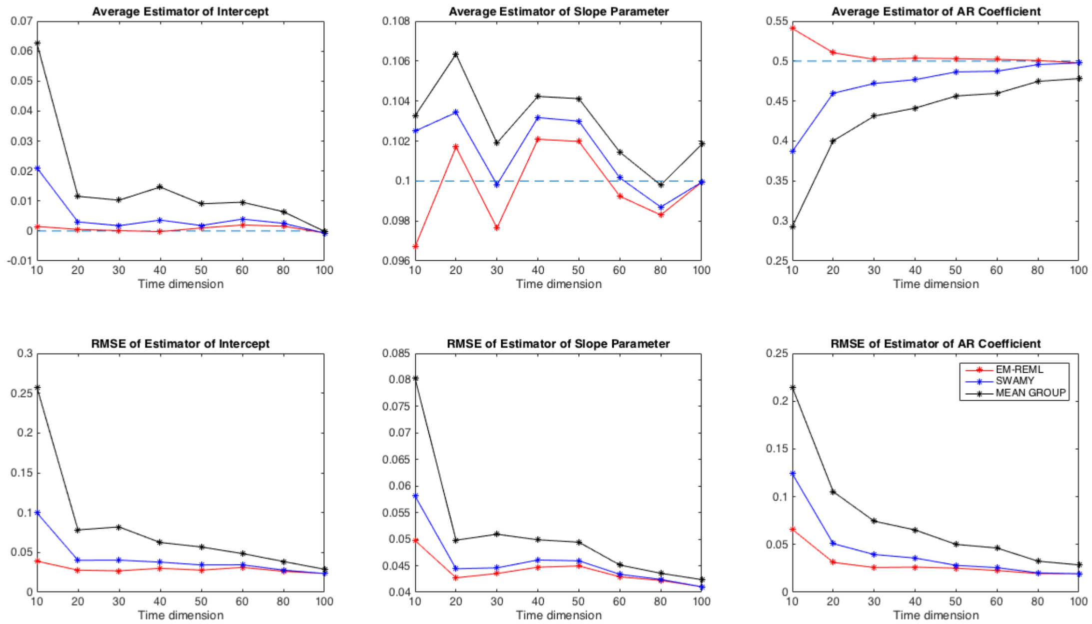


Figure 3.1: **Upper panel**: The estimators of the intercept (left), slope (middle) and autoregressive parameter (right panel), averaged across the 500 replications, are plotted for $N = 30$ and $T = \{10, 20, 30, 40, 50, 60, 80, 100\}$. The dashed blue lines indicate the true values (used to simulate the data). The red, blue, and black solid lines correspond to the EM-REML, Swamy, and Mean Group estimator, respectively. The distances between those lines and the one corresponding to the true value measure the bias of the estimators. **Lower panel**: the root mean square errors (RMSE) of the estimators are reported.

A graphical summary of these results is provided in Figure 3.1. The upper panels show the average values (across 500 Monte Carlo replications) of the EM-REML, Swamy, and MG estimators of the average effects. The differences between the latter and the corresponding true values measure the bias of the estimates. The RMSE of the estimators are depicted in the lower panels.

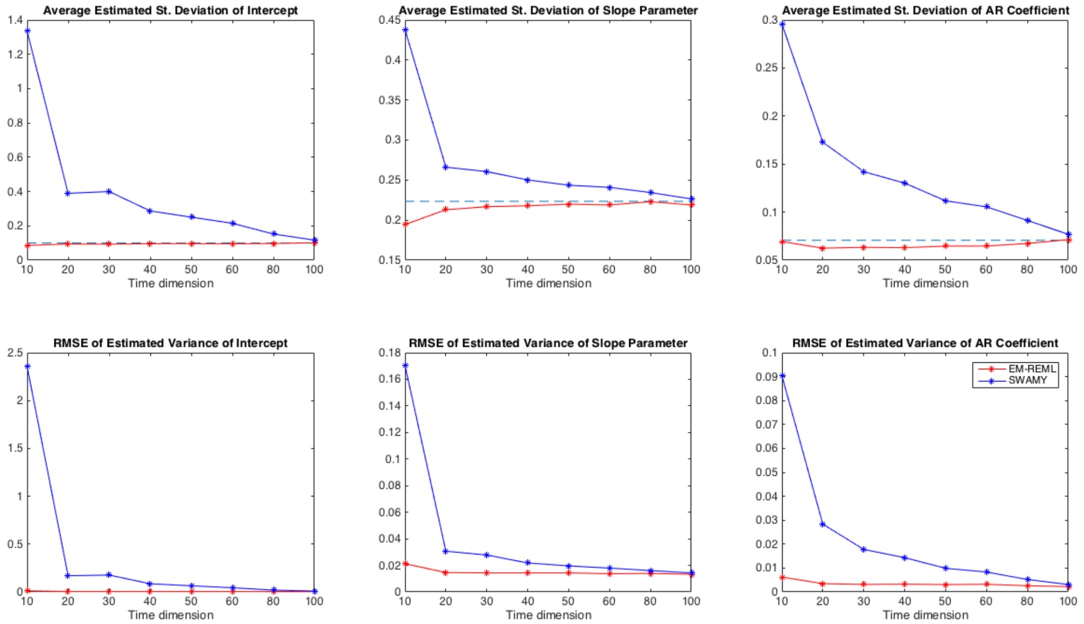


Figure 3.2: **Upper panel**: The estimated standard deviation of the intercept (left), slope (middle) and autoregressive parameter (right panel), averaged across the 500 replications, are plotted for $N = 30$ and $T = \{10, 20, 30, 40, 50, 60, 80, 100\}$. The dashed blue line indicates the true value (used to simulate the data). The red, and blue lines correspond to the EM-REML, and Swamy estimator, respectively. The distances between those lines and the one corresponding to the true value measure the bias of the estimators. **Lower panel**: the root mean square errors (RMSE) of the estimated variances are reported.

Figure 3.2 illustrates the performance of the EM-REML and Swamy estimators of the random coefficients' variances. As expected, the latter largely overestimates the true variance components of the model. The size of the bias can be substantial unless the time dimension is quite large. For example, when $T = 10$, the probability that the Swamy unbiased estimator of the covariance matrix is negative definite is equal to 81 percent. This means that in most of the cases it has to be replaced by its consistent but biased version. At the same time, given that in some replications we are able to use the unbiased estimator and in others only the consistent one, the RMSE of the estimated variance components obtained using Swamy procedure can be quite substantial although it reduces as T increases.

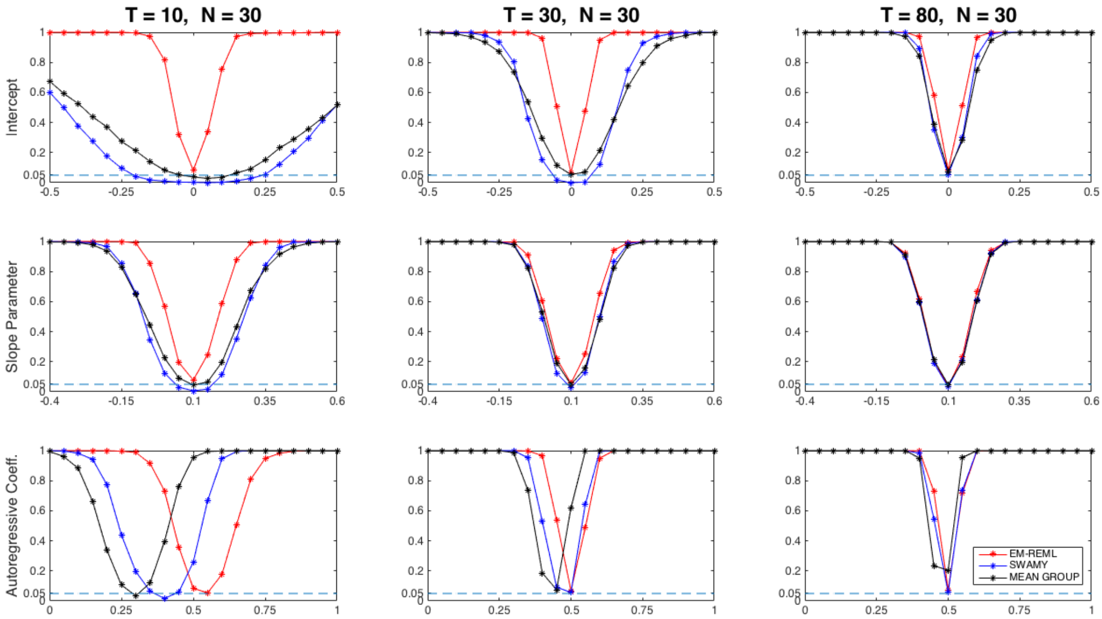


Figure 3.3: Rejection frequency at 5% nominal size, for the intercept (upper panels), the slope (middle panels), and autoregressive parameters (lower panels), when $(c, \beta, \phi) = (0, 0.1, 0.5)$. The panels on the left show results for $(T, N) = (10, 30)$, the panels on the middle for $(T, N) = (30, 30)$, and those on the right for $(T, N) = (80, 30)$. The red, blue, and black lines denote the power performances of the EM-REML, Swamy, and Mean Group estimators respectively.

To examine the consequences of overestimating or underestimating the true random coefficient variances when testing hypotheses, we consider the ratio between the "infeasible" standard errors (which are obtained substituting the true values used to generate the DGP into equation (3.49)) and the estimated standard errors of the average effects. Another important measure for inference is the accuracy of the estimated standard errors as approximations to the correct sampling standard deviation of the estimator of interest.[21] These ratios should ideally be equal to one. Results are reported in Tables 3.3, 3.4, and 3.5. We find that the standard errors obtained estimating the parameters of the model using Swamy GLS approach, are in many cases largely overestimated unless T is quite large. In the

[21] In particular, the accuracy of the estimated standard errors is computed as the ratio of the latter averaged across $B = 500$ replications, $B^{-1}\sum_{b=1}^{B}\left\{\sqrt{v\hat{a}r\left(\hat{\psi}_{k,(b)}\right)}\right\}$, and the sampling standard deviation of the estimator of interest, given by the square root of $(B-1)^{-1}\sum_{b=1}^{B}\left(\hat{\psi}_{k,(b)} - \bar{\hat{\psi}}_k\right)^2$, where $\bar{\hat{\psi}}_k = B^{-1}\sum_{b=1}^{B}\hat{\psi}_{k,(b)}$, for $k = 1, 2, 3$.

latter case, the percentage of replications in which the Swamy estimator of the random coefficient covariance matrix is negative definite diminishes, and the ratio of standard errors approaches one.

Our Monte Carlo experiments reveal that the biases of the Swamy estimator of the variance components and of the resulting standard errors can be too large to be neglected. This in turn affects hypothesis tests adversely. To demonstrate the latter point, we consider the power performances of the various estimators. We plot the power functions in Figure 3.3. They are computed using the Swamy type test described in equation (3.50) for $N = 30$, and various T.[22] It is shown that the EM-REML approach performs comparatively well even when the sample size is small. When T is small the power functions of the Swamy and MG estimators of the autoregressive coefficients are not centred at the true value of $\phi = 0.5$. As the time dimension increases the differences in the power performances of the various estimators reduce.

The Sensitivity of the Bayesian Estimator to the Choice of the Prior

As discussed in Subsection 3.5.4, the choice of the prior may affect the performance of the Bayesian estimation. For instance, assuming that $\triangle^{-1}$ has a Wishart distribution with scale matrix $(\rho\Upsilon)$ and ρ degrees of freedom, Hsiao Pesaran and Tahmiscioglu (1999) note that the bias of both the empirical and hierarchical Bayes estimators of the regression coefficients can be sensitive to the specification of the prior scale matrix. Therefore, it is interesting to compare the performance of the hierarchical Bayes estimator with different prior choices. In a first specification, we use the same prior structure as in Hsiao, Pesaran and Tahmiscioglu (1999).[23] In a second specification, we set Υ equal to the EM-REML estimator

[22] For the Mean Group estimation, the t-ratios are also appropriate. To facilitate comparison we only report the power functions computed using the Swamy type test, noting that in both cases the results are very similar.

[23] Under the assumption that the vector of average effects, ψ, has a prior distribution which is Normal with mean μ and variance Ω, the authors set $\Omega^{-1} = 0$, $\rho = 2$, and choose Υ equal to the Swamy estimate of $\triangle$.

of the covariance matrix $\triangle$ instead of the Swamy estimator.[24] Results are shown in Figure 3.4 and 3.5. By simply replacing the prior for Υ with a more precise estimate of $\triangle$, obtained employing the EM-REML approach, the performances of the posterior mean of both the average effects and the variance components (in terms of bias and RMSE), notably improve, especially in small samples. This evidence confirms Kass and Wasserman (1996) argument that the prior choice can have heavy weight on the posterior when sample sizes are small.

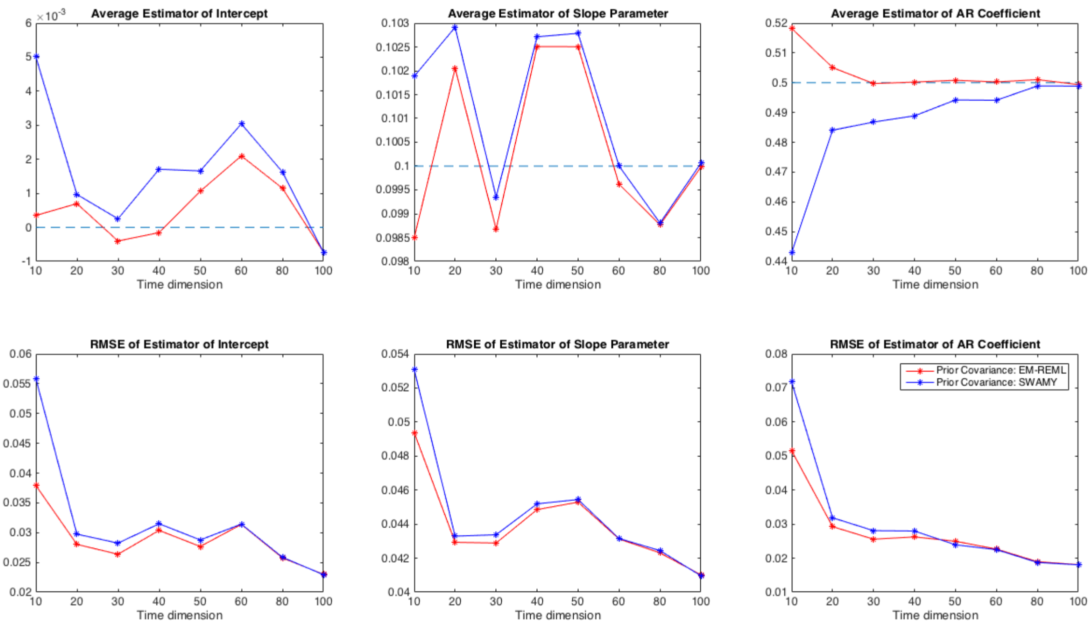


Figure 3.4: **Upper panel**: Posterior means for the intercept (left), slope (middle) and autoregressive parameter (right panel), averaged across 500 replications, are plotted for $N = 30$ and $T = \{10, 20, 30, 40, 50, 60, 80, 100\}$. The dashed blue lines indicate the true values (used to simulate the data). Results in blue are obtained using priors as in Hsiao et al. (1999). Results in red are obtained using the EM-REML estimate of the random coefficient covariance as prior input. The distances between those lines and the one corresponding to the true value measure the bias of the estimators. **Lower panel**: the root mean square errors (RMSE) of the estimators are reported.

[24]In both cases, we simulate a Markov chain of 6000 cycles, and discard the initial 1000 burn-in replications. Hsiao, Pesaran and Tahmiscioglu (1999) note that convergence is quickly achieved, and suggest using 3000 iterations.

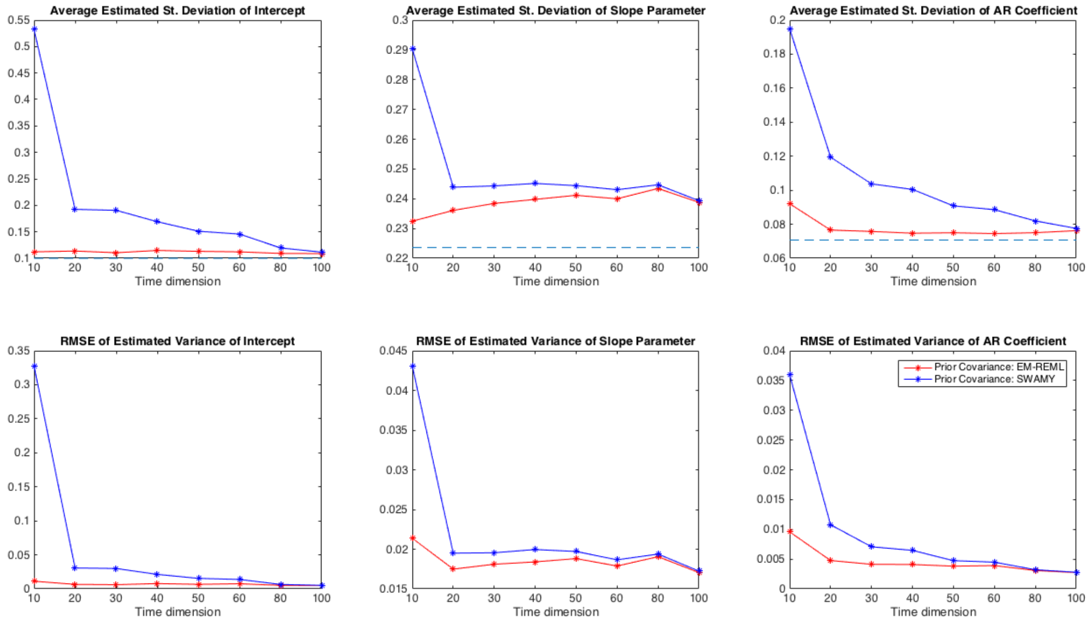


Figure 3.5: **Upper panel**: Posterior means for the variance of the intercept (left), slope (middle) and autoregressive parameter (right panel), averaged across 500 replications, are plotted for $N = 30$ and $T = \{10, 20, 30, 40, 50, 60, 80, 100\}$. The dashed blue line indicates the true value (used to generate the data). Results in blue are obtained using priors as in Hsiao et al. (1999). Results in red are obtained using the EM-REML estimate of the random coefficient covariance. The distances between those lines and the one corresponding to the true value measure the bias of the estimators as prior input. **Lower panel**: the root mean square errors (RMSE) of the estimators are reported.

3.7.3 Limitations and Future Directions

We shall note that the results described in this Section hold under the specific DGP used to run the simulations. Possible topics for future investigation are to evaluate the performances of the various estimators in the following cases: (i) the random coefficients do not follow a Normal distribution; (ii) the exogenous variables are correlated with the intercept; (iii) the true DGP is static (i.e. the linear regression model does not include lagged values of the dependent variables among the regressors). The scope of the comparative analysis could be also extended. First, in the static case, when the regressors are exogenous, Pesaran and Smith (1995) note that the fixed effect estimator of the average effects is unbiased and consistent even in the presence of coefficient heterogeneity. Therefore, it would be interesting to include the fixed effects estimation in the comparison, to evaluate its efficiency, the accuracy of the standard errors, and the power performance, relative to other methods. Second, given that the Mean Group estimator may be sensitive to outliers, we would like to investigate whether trimming them is a viable solution. Finally, the literature on heterogeneous panels largely centers on the behaviour of the estimators of the average effects. A Monte Carlo analysis focusing on the performance of the estimators of the unit-specific coefficients seems very much needed, given that such parameters can be of particular interest in many economic applications.

3.8 Application

Reinhart, Rogoff and Savastano (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 macroeconomic fundamentals and a set of explanatory variables which reflect the history of government debt and economic crises of various forms. We find that 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.[25] 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).[26] Second, we consider a dynamic model. Third, given that we are comparing countries with very different characteristics, even within group, we allow for heterogeneity rather than pooling. The implications of neglected heterogeneity and dynamics can be severe. Pesaran and Smith (1995) show that if the DGP includes lagged values of the dependent variables among the explanatory variables, pooling give inconsistent and potentially highly misleading estimates of the coefficients when the latter differ across units. Haque, Pesaran and Sharma (2000) find that ignoring differences across countries can lead to overestimating the influence of certain factors. They argue that one can obtain highly significant, but spurious, nonlinear effects for some of the potential determinants, even though the country-specific regressions are linear.

Finally, the focus of this application is on understanding which factors determine the additional risk premium to charge during episodes of debt growth.

Assume that sovereign spreads are a function of debt-to-GDP ratio, a proxy for history of default and other macroeconomic fundamentals. Rather than looking at how spreads change with respect to one variable while debt-to-GDP and the remaining covariates are held constant (i.e. partial effect), we investigate which country characteristics significantly affect the magnitude of sovereign spreads' reaction to changes in debt. Studying the sensitivity of financial markets during episode of debt growth may help understand why emerging markets cannot borrow at level comparable to more developed economies without having to pay relatively high interest rates.

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

[26]The panel is slightly unbalanced. The individual time observations vary between $60 \leq T_i \leq 87$. The choice of countries is dictated by the availability of data. The list of countries is reported in Appendix 3.11.

3.8.1 The Empirical Model

Following Edwards (1984), we assume that the spreads over U.S. (or Germany) Treasuries can be explained by a set of macroeconomic indicators. We focus on real GDP growth, inflation, and the growth rates of 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.[27]

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 assume that the economy possesses a recursive structure where spreads are ordered last. The last equation of the recursive system can be written as

$$y_{it} = \phi_i y_{it-1} + x'_{it}\beta_{0i} + x'_{i,t-1}\beta_{1i} + \mu_i + \varepsilon_{it}, \tag{3.58}$$

for $i = 1,..,N$ and $t = 1,..,T$; y_{it} includes the first difference of sovereign spreads. The number of lags has been selected using the BIC criterion (averaged across units) since it results in more parsimonious model than the AIC. The panel data model in matrix notation can be written as in equation (3.2) where all the coefficients are random and follow (3.3). When doing parameter equality tests 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.

3.8.2 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. To test the null hypothesis $H_0: \ \psi_1 = .. = \psi_N = \psi$ (i.e. to test whether the coefficient vectors $\psi_i = (\mu_i, \beta'_{0i}, \phi_i, \beta'_{1i})'$ are constant across units), we can use the

[27] A description of the data is provided in Appendix 3.11.

following test proposed by Swamy (1970):

$$F = \frac{1}{(N-1)} \sum_{i=1}^{N} F_i \sim F\left(K^*(N-1), (\sum_{i=1}^{N} T_i - NK^*)\right), \tag{3.59}$$

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 are 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 296 and 2708 degrees of freedom, the F-value that leaves exactly 0.01 of the area under the F curve in the right tail of the distribution is smaller than 1.32.[28] Because our test has a value of 2.58, we are able to reject the null of homogenous slope and intercept parameters.

3.8.3 The Sensitivity of Spreads to Debt

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

Using Reinhart and Rogoff (2011) historical time series on countries credit-worthiness and financial turmoil, we calculate the percentage of years (between 1980 and 2010) each country has been in default or restructuring on its domestic and external debt, the percentage of years with annual inflation of 20% or higher,

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

and the percentage of years with annual depreciation vs US dollar of 15% or more. We then estimate equation (3.58) while allowing the coefficients to be a function of a common constant, and the percentage of years in default or restructuring domestic and external debt. Results are shown in Table 3.1.

Table 3.1: Determinants of sensitivity of spreads: EM-REML Estimates.

	const.	**% y-DomDef**	**% y-ExtDef**
c_i	-0.017	0.596	-0.180
	(1.572)	(0.647)	(0.897)
$\beta_0^{(gdp)}$	**-0.016***	-0.252	-0.082
	(3.727)	(0.364)	(0.597)
$\beta_0^{(cpi)}$	0.008	0.624	-0.226
	(0.143)	(1.005)	(1.502)
$\beta_0^{(debt)}$	-0.006	**0.344****	**0.068***
	(1.311)	(5.998)	(3.264)
ϕ	**0.112*****	-0.326	-0.175
	(8.035)	(0.265)	(0.501)
$\beta_1^{(gdp)}$	0.010	**-0.752***	**0.314*****
	(1.060)	(2.900)	(8.096)
$\beta_1^{(cpi)}$	**0.037***	-0.874	0.062
	(3.527)	(2.201)	(0.128)
$\beta_1^{(debt)}$	0.003	-0.101	0.004
	(0.751)	(0.616)	(0.014)

Swamy F-statistic (described in equation (3.50)) between parentheses. The critical values for a F distribution with 1 degree of freedom for the numerator, and $N-1$ for the denominator, associated with a significance level equal to 0.1, 0.05, and 0.01, are 2.84, 4.08, and 7.31 respectively. Simbols ***, **, and * denote significance (at least) at 1%, 5% and 10% respectively. Estimated standard errors are corrected for finite-sample bias, following Kenward and Roger (1997). "% y DomDef" ("% y ExtDef") denotes the percentage of years in default or restructuring domestic (external) debt; ϕ is the autoregressive coefficient; $\beta^{(k)}$ is the sensitivity of spread to the kth variable.

Our results seem to suggest that history of repayment plays an important role: "bad" reputation leads to higher sensitivity of spreads to debt. A 1% increase in the percentage of years in default or restructuring domestic debt is associated with a 0.34% increase in the sensitivity of spread. As a consequence, relatively small increase in debt-to-GDP may lead to level of interest rates which can be difficult to tolerate. Although significant, the impact of our proxy for history of repayment of external debt is rather low, around 0.07 percent.

The above analysis is robust when augmenting the regression equation for the coefficients with additional explanatory variables. In particular, we include the percentage of years in which a country has faced an annual inflation rate of 20 percent or higher and the percentage of years in which an annual depreciation versus the US dollar (or another relevant anchor currency) of 15 percent or more occurs.[29] We also consider measures of macroeconomic fundamentals such as the average and standard deviation of real GDP growth, of rate of currency depreciation, of inflation and current account to GDP growth. Standard deviations over the sample period under considerations are used as measure of volatility. The standard deviation of the average growth rate of general gross government debt to GDP can be considered as a proxy for sudden increases in debt's level.

In Table 3.2, we focus on the coefficients equation corresponding to the sensitivity of spreads to debt and report results from using different specifications. 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 matters. The percentage of years in defaults or restructuring on domestic debt have a statistically and economically significant effect on the sensitivity of spreads across all the different specifications. Interestingly, domestic defaults have a larger impact than external ones. Our finding that domestic defaults play a significant role in explaining changes in the sensitivity of spreads is in line with 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. Contrary to the literature which emphasizes the role of volatility of macroeconomic aggregates in explaining sovereign credit risks, we do not find strong evidence that such variables affect

[29] A detailed description of the data is provided by Reinhart and Rogoff (2009).

markets when calculating the additional risk premium to charge in response to an increase in debt.[30] This seems to suggest that markets decisions during episodes of debt growth may also be driven by sentiments (as defined by Eichengreen and Mody, 2000). At the same time, we have seen that a bigger reaction is usually associated with countries with a weak history of repayment.

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.

The analysis could be extended in the future by including other important factors, such as measures of political instability and of the composition of debt, which could shed further light on why the sensitivity of spreads to debt differs across countries. At the same time, an interesting avenue for future research would be to allow the coefficients to vary over time, by modelling the latter as function of observable time-varying characteristics or possibly unobservable common factors.

[30] Selected studies on the role of volatility in explaining sovereign defaults are: Eaton and Gersovitz (1981), Catao and Kapur (2006), and Hilscher and Nosbuch (2010).

Table 3.2: Determinants of sensitivity of spreads to government debt: EM-REML Estimates.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Constant	-0.006	-0.008	-0.018	-0.016	0.001	-0.007	-0.019
	(1.311)	(0.944)	(1.719)	(1.280)	(0.005)	(0.181)	(1.172)
% y Dom Def	**0.344**	**0.328**	**0.306**	**0.308**	**0.350**	**0.340**	
	(5.998)	(3.637)	(4.592)	(4.535)	(6.085)	(5.383)	
% y Ext Def	**0.068**	0.026	0.058	**0.066**	0.010	0.014	
	(3.264)	(0.290)	(2.507)	(2.990)	(0.074)	(0.122)	
% y Curr Crisis		-0.005					
		(0.006)					
% y Infl Crisis		0.066					
		(1.035)					
Volatility FX			0.003	0.003	0.001	0.003	0.006
			(1.116)	(1.156)	(0.045)	(0.486)	(1.802)
Volatility Debt/GDP				-0.001	0.006	0.004	0.005
				(0.313)	(2.652)	(1.023)	(1.889)
Volatility Inflation					0.011	0.005	0.008
					(1.525)	(0.196)	(0.468)
Volatility RGDP					**-0.030**	-0.016	-0.016
					(5.798)	(0.553)	(0.532)
Volatility CA/GDP						-0.004	-0.008
						(0.392)	(1.209)

Swamy F-statistic (described in equation (3.50)) between parentheses. The critical values for a F distribution with 1 degree of freedom for the numerator, and $N-1$ for the denominator, associated with a significance level equal to 0.1, 0.05, and 0.01, are 2.84, 4.08, and 7.31 respectively. Estimated standard errors are corrected for finite-sample bias, following Kenward and Roger (1997). Bold values denotes statistical significance at 10% level or lower. "% y Curr Crisis" and "% 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 years in default or restructuring domestic (external) debt.

3.9 Conclusion

This chapter shows how to implement the EM algorithm to compute iteratively restricted maximum likelihood (REML) estimates of both fixed and random coefficients, as well as the variance components, in a wide class of heterogeneous panels. The proposed method has several benefits. First, the EM-REML approach yields an unbiased and more efficient estimator of the random coefficient covariance without running into the problem of negative definite matrices typically encountered in the Swamy type random coefficient models. This is turn leads to more accurate estimated standard errors and hypothesis tests. We also demonstrate that Lee and Griffiths (1979) approach to jointly estimate the random components and constant underlying parameters, yield an estimator of the coefficients' covariance matrix which does not satisfy the law of total variance. This is not the case when employing the EM algorithm. Second, the latter allows us to make inference on the random effects' population. The EM approach should be considered as a valid alternative to Bayesian estimation in those cases in which the researcher wishes to make inference on the random effects' distribution while having little knowledge on what sensible priors might be. At the same time, it helps overcome one drawback of the Bayesian inference: 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.

Monte Carlo experiments confirm that our approach performs relatively well in finite sample, in term of bias, root mean square errors and power of tests.

Another contribution of this chapter is to review in a coherent manner, some of the existing sampling and Bayesian methods commonly used to estimate random coefficient panel data models.

An empirical application is also presented. We investigate what causes the sensitivity of spreads to differ significantly across countries by modelling the latter as a function of macroeconomics fundamentals and a set of explanatory variables which reflect the history of government debt and economic crises of various forms. 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 (including underlying volatility) 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. An 1% increase in the percentage of years in default or restructuring domestic debt is associated with around 0.35% increase in the additional risk premium in response to an increase in debt. Our findings indicate that countries who have defaulted in the past may find it hard 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. This helps explain why 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 3.3: Properties of EM-REML estimator as T gets large, for fixed $N = 30$

	EM-REML							
$N = 30/T$	10	20	30	40	50	60	80	100
$Bias\left(\hat{c}\right)$	**0.0015**	**0.0005**	**0.0001**	**-0.0002**	**0.0010**	**0.0020**	**0.0016**	**-0.0007**
$se\left\{Bias\left(\hat{c}\right)\right\}$	0.0017	0.0012	0.0012	0.0013	0.0012	0.0014	0.0012	0.0011
$Bias\left(\hat{\beta}\right)$	**-0.0033**	**0.0017**	**-0.0024**	**0.0021**	**0.0020**	**-0.0008**	**-0.0017**	**-0.0001**
$se\left\{Bias\left(\hat{\beta}\right)\right\}$	0.0022	0.0019	0.0019	0.0020	0.0020	0.0019	0.0019	0.0018
$Bias\left(\hat{\phi}\right)$	**0.0408**	**0.0104**	**0.0022**	**0.0037**	**0.0030**	**0.0022**	**0.0007**	**-0.0022**
$se\left\{Bias\left(\hat{\phi}\right)\right\}$	0.0023	0.0013	0.0012	0.0012	0.0011	0.0010	0.0009	0.0009
$\parallel Bias\left(\hat{\psi}\right) \parallel$	**0.0410**	**0.0106**	**0.0032**	**0.0042**	**0.0037**	**0.0030**	**0.0024**	**0.0023**
$RMSE\left(\hat{c}\right)$	0.0391	0.0278	0.0269	0.0301	0.0278	0.0313	0.0263	0.0235
$RMSE\left(\hat{\beta}\right)$	0.0497	0.0427	0.0435	0.0447	0.0450	0.0429	0.0422	0.0410
$RMSE\left(\hat{\phi}\right)$	0.0659	0.0313	0.0259	0.0263	0.0252	0.0227	0.0196	0.0193
$Bias\left(\hat{var}\left(\gamma_1\right)\right)$	**-0.0026**	**-0.0011**	**-0.0013**	**-0.0007**	**-0.0007**	**-0.0009**	**-0.0005**	**0.0003**
$se\left\{Bias\left(\hat{var}\left(\gamma_1\right)\right)\right\}$	0.0005	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
$Bias\left(\hat{var}\left(\gamma_2\right)\right)$	**-0.0122**	**-0.0046**	**-0.0030**	**-0.0025**	**-0.0016**	**-0.0020**	**-0.0002**	**-0.0021**
$se\left\{Bias\left(\hat{var}\left(\gamma_2\right)\right)\right\}$	0.0008	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
$Bias\left(\hat{var}\left(\gamma_3\right)\right)$	**-0.0002**	**-0.0011**	**-0.0010**	**-0.0010**	**-0.0008**	**-0.0008**	**-0.0004**	**0.0001**
$se\left\{Bias\left(\hat{var}\left(\gamma_3\right)\right)\right\}$	0.0003	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
$RMSE\left\{\hat{var}\left(\gamma_1\right)\right\}$	0.0118	0.0045	0.0046	0.0053	0.0047	0.0055	0.0040	0.0041
$RMSE\left\{\hat{var}\left(\gamma_2\right)\right\}$	0.0213	0.0146	0.0144	0.0143	0.0143	0.0138	0.0140	0.0133
$RMSE\left\{\hat{var}\left(\gamma_3\right)\right\}$	0.0062	0.0035	0.0032	0.0033	0.0031	0.0032	0.0026	0.0023
$Ratio\left\{se\left(\hat{c}_i\right)\right\}$	1.01	0.98	0.96	1.01	0.98	1.00	0.98	1.01
$Ratio\left\{se\left(\hat{\beta}_i\right)\right\}$	0.95	0.96	0.97	0.97	0.98	0.98	0.99	0.97
$Ratio\left\{se\left(\hat{\phi}_i\right)\right\}$	1.56	1.12	1.08	1.15	1.06	1.11	1.01	1.03
$Accuracy\left\{se\left(\hat{c}_i\right)\right\}$	0.87	0.94	0.93	0.95	0.93	0.90	0.89	0.97
$Accuracy\left\{se\left(\hat{\beta}_i\right)\right\}$	0.89	0.97	0.95	0.93	0.92	0.96	0.98	0.98
$Accuracy\left\{se\left(\hat{\phi}_i\right)\right\}$	1.06	0.95	0.95	1.04	0.88	1.03	0.95	0.94

The data generating process is described in equation (3.55) and (3.56), in Section 3.7. We assume that $c_i \sim N(0, \sigma_c^2)$, $\beta_i \sim N(0.1, \sigma_\beta^2)$, and $\phi_i \sim N(0.5, \sigma_\phi^2)$, where $(\sigma_c, \sigma_\beta, \sigma_\phi) = (0.1, 0.224, 0.07)$. "se" stands for standard errors; RMSE indicates the root mean square errors. The Euclidean norm ($\|\cdot\|$) is used as an overall measure of the bias. $Ratio(se(\cdot))$ measures the ratio between the "infeasible" standard errors (which are obtained substituting the true values used to generate the DGP into equation (3.49)) and the estimated standard errors of the average effects. $Accuracy(se(\cdot))$ denotes the accuracy of the estimated standard errors as approximations to the correct sampling standard deviation of the EM-REML estimator.

Table 3.4: Properties of Swamy estimator as T gets large, for fixed $N = 30$

	Swamy							
$N = 30/T$	10	20	30	40	50	60	80	100
$Bias\left(\hat{c}\right)$	**0.0211**	**0.0030**	**0.0018**	**0.0036**	**0.0018**	**0.0040**	**0.0025**	**-0.0007**
$se\left\{Bias\left(\hat{c}\right)\right\}$	0.0044	0.0018	0.0018	0.0017	0.0015	0.0015	0.0012	0.0011
$Bias\left(\hat{\beta}\right)$	**0.0025**	**0.0034**	**-0.0002**	**0.0032**	**0.0030**	**0.0002**	**-0.0013**	**-0.0001**
$se\left\{Bias\left(\hat{\beta}\right)\right\}$	0.0026	0.0020	0.0020	0.0021	0.0021	0.0019	0.0019	0.0018
$Bias\left(\hat{\phi}\right)$	**-0.1130**	**-0.0402**	**-0.0281**	**-0.0233**	**-0.0136**	**-0.0127**	**-0.0044**	**-0.0022**
$se\left\{Bias\left(\hat{\phi}\right)\right\}$	0.0023	0.0014	0.0012	0.0012	0.0011	0.0010	0.0009	0.0009
$\parallel Bias\left(\hat{\psi}\right)\parallel$	**0.1150**	**0.0405**	**0.0281**	**0.0237**	**0.0140**	**0.0133**	**0.0052**	**0.0023**
$RMSE\left(\hat{c}\right)$	0.0998	0.0401	0.0403	0.0378	0.0343	0.0345	0.0278	0.0235
$RMSE\left(\hat{\beta}\right)$	0.0581	0.0444	0.0446	0.0461	0.0459	0.0434	0.0424	0.0410
$RMSE\left(\hat{\phi}\right)$	0.1240	0.0509	0.0395	0.0358	0.0282	0.0257	0.0203	0.0194
$Bias\left(v\hat{a}r\left(\gamma_1\right)\right)$	**1.7745**	**0.1419**	**0.1501**	**0.0724**	**0.0530**	**0.0362**	**0.0133**	**0.0037**
$se\left\{Bias\left(v\hat{a}r\left(\gamma_1\right)\right)\right\}$	0.0695	0.0042	0.0042	0.0020	0.0017	0.0011	0.0007	0.0004
$Bias\left(v\hat{a}r\left(\gamma_2\right)\right)$	**0.1420**	**0.0209**	**0.0179**	**0.0126**	**0.0093**	**0.0081**	**0.0050**	**0.0013**
$se\left\{Bias\left(v\hat{a}r\left(\gamma_2\right)\right)\right\}$	0.0042	0.0010	0.0010	0.0008	0.0008	0.0007	0.0007	0.0006
$Bias\left(v\hat{a}r\left(\gamma_3\right)\right)$	**0.0825**	**0.0248**	**0.0152**	**0.0120**	**0.0075**	**0.0061**	**0.0033**	**0.0009**
$se\left\{Bias\left(v\hat{a}r\left(\gamma_3\right)\right)\right\}$	0.0017	0.0006	0.0004	0.0004	0.0003	0.0003	0.0002	0.0001
$RMSE\left\{v\hat{a}r\left(\gamma_1\right)\right\}$	2.3571	0.1701	0.1769	0.0855	0.0648	0.0441	0.0199	0.0090
$RMSE\left\{v\hat{a}r\left(\gamma_2\right)\right\}$	0.1705	0.0307	0.0278	0.0218	0.0196	0.0180	0.0161	0.0143
$RMSE\left\{v\hat{a}r\left(\gamma_3\right)\right\}$	0.0904	0.0284	0.0178	0.0144	0.0099	0.0083	0.0052	0.0031
$Ratio\left\{se\left(\hat{c}_i\right)\right\}$	7.34	2.90	3.06	2.13	1.96	1.65	1.33	1.08
$Ratio\left\{se\left(\hat{\beta}_i\right)\right\}$	1.85	1.19	1.17	1.12	1.09	1.07	1.04	1.00
$Ratio\left\{se\left(\hat{\phi}_i\right)\right\}$	2.00	1.61	1.46	1.33	1.25	1.20	1.14	1.02
$Accuracy\left\{se\left(\hat{c}_i\right)\right\}$	2.53	1.93	1.98	1.60	1.51	1.35	1.13	1.03
$Accuracy\left\{se\left(\hat{\beta}_i\right)\right\}$	1.49	1.16	1.11	1.04	1.00	1.05	1.02	1.01
$Accuracy\left\{se\left(\hat{\phi}_i\right)\right\}$	1.35	1.30	1.19	1.15	1.06	1.13	1.06	0.93
% Negative Definite	0.81	0.71	0.72	0.72	0.60	0.59	0.47	0.16

The data generating process is described in equation (3.55) and (3.56), in Section 3.7. We assume that $c_i \sim N(0, \sigma_c^2)$, $\beta_i \sim N(0.1, \sigma_\beta^2)$, and $\phi_i \sim N(0.5, \sigma_\phi^2)$, where $(\sigma_c, \sigma_\beta, \sigma_\phi) = (0.1, 0.224, 0.07)$. "se" stands for standard errors; RMSE indicates the root mean square errors. The Euclidean norm ($\| \cdot \|$) is used as an overall measure of the bias. $Ratio(se(\cdot))$ measures the ratio between the "infeasible" standard errors (which are obtained substituting the true values used to generate the DGP into equation (3.49)) and the estimated standard errors of the average effects. "% Negative Definite" measures the number of times (in percentage) the estimated random coefficients' covariance matrix is negative definite. $Accuracy(se(\cdot))$ denotes the accuracy of the estimated standard errors as approximations to the correct sampling standard deviation of the Swamy GLS estimator.

Table 3.5: Properties of Mean Group estimator as T gets large, for fixed $N = 30$

	Mean Group							
$N = 30/T$	10	20	30	40	50	60	80	100
$Bias\left(\hat{c}\right)$	**0.0626**	**0.0116**	**0.0103**	**0.0146**	**0.0091**	**0.0096**	**0.0064**	**-0.0001**
$se\left\{Bias\left(\hat{c}\right)\right\}$	0.0112	0.0035	0.0036	0.0027	0.0025	0.0021	0.0017	0.0013
$Bias\left(\hat{\beta}\right)$	**0.0033**	**0.0063**	**0.0019**	**0.0042**	**0.0041**	**0.0014**	**-0.0002**	**0.0018**
$se\left\{Bias\left(\hat{\beta}\right)\right\}$	0.0036	0.0022	0.0023	0.0022	0.0022	0.0020	0.0020	0.0019
$Bias\left(\hat{\phi}\right)$	**-0.2072**	**-0.0998**	**-0.0688**	**-0.0589**	**-0.0437**	**-0.0404**	**-0.0254**	**-0.0220**
$se\left\{Bias\left(\hat{\phi}\right)\right\}$	0.0025	0.0015	0.0013	0.0013	0.0011	0.0010	0.0009	0.0008
$\parallel Bias\left(\hat{\psi}\right)\parallel$	**0.2165**	**0.1006**	**0.0696**	**0.0608**	**0.0448**	**0.0415**	**0.0262**	**0.0220**
$RMSE\left(\hat{c}\right)$	0.2575	0.0783	0.0820	0.0625	0.0569	0.0486	0.0387	0.0289
$RMSE\left(\hat{\beta}\right)$	0.0803	0.0498	0.0509	0.0499	0.0494	0.0451	0.0436	0.0424
$RMSE\left(\hat{\phi}\right)$	0.2146	0.1053	0.0744	0.0652	0.0501	0.0463	0.0325	0.0287
$Accuracy\left\{se\left(\hat{c}_i\right)\right\}$	0.98	0.99	0.98	0.95	0.95	0.97	0.89	0.97
$Accuracy\left\{se\left(\hat{\beta}_i\right)\right\}$	1.03	1.01	0.96	0.93	0.92	0.99	0.99	0.99
$Accuracy\left\{se\left(\hat{\phi}_i\right)\right\}$	1.05	1.06	1.02	0.95	0.97	0.99	0.98	1.00

The data generating process is described in equation (3.55) and (3.56), in Section 3.7. We assume that $c_i \sim N(0, \sigma_c^2)$, $\beta_i \sim N(0.1, \sigma_\beta^2)$, and $\phi_i \sim N(0.5, \sigma_\phi^2)$, where $(\sigma_c, \sigma_\beta, \sigma_\phi) = (0.1, 0.224, 0.07)$. "se" stands for standard errors; RMSE indicates the root mean square errors. The Euclidean norm ($\| \cdot \|$) is used as an overall measure of the bias. $Accuracy(se(\cdot))$ denotes the accuracy of the estimated standard errors as approximations to the correct sampling standard deviation of the Mean Group estimator.

3.10 Appendix

3.10.1 Restricted Likelihood

The Choice of S_i.

The projection matrix M_i. One plausible choice for S_i, is the projection matrix:

$$M_i = I - W_i \left(W_i' W_i\right)^{-} W_i', \tag{3.60}$$

where $(W_i'W_i)^{-}$ denotes the generalized inverse of $W_i'W_i$. The matrix M_i is of rank $T - \underline{K}$, with $\underline{K} \leq \bar{K} < T$, and 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

$$U_i M_i U_i' = \begin{bmatrix} I_{T-\underline{K}} & 0 \\ 0 & 0 \end{bmatrix},$$

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

$$M_i A_i = A_i, \qquad A_i' M_i = A_i' \ . \tag{3.61}$$

Since U_i' is orthogonal and non-singular, A_i' has full rank and $A_i' W_i = 0$. Using (3.61), Searle and Quaas (1978) show that $A_i \left(A_i' R_i A_i\right)^{-1} A_i'$ is the Moore-Penrose inverse of $M_i R_i M_i$:

$$\left(M_i R_i M_i\right)^{+} = A_i \left(A_i' R_i A_i\right)^{-1} A_i'. \tag{3.62}$$

Given that 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, $S_i = JM_i$, satisfies $S_iW_i = 0$. A generalization of M_i is

$$P_i = R_i^{-1} - R_i^{-1}W_i\left(W_i'R_i^{-1}W_i\right)^{-}W_i'R_i^{-1}, \tag{3.63}$$

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

$$\begin{aligned} R_iP_i &= I - W_i\left(W_i'R_i^{-1}W_i\right)^{-}W_i'R_i^{-1}, \\ P_iR_i &= I - R_i^{-1}W_i\left(W_i'R_i^{-1}W_i\right)^{-}W_i'. \end{aligned} \tag{3.64}$$

Therefore,

$$P_iR_iP_i = P_i, \tag{3.65}$$

and also $(P_iR_i)^2 = P_iR_i$. It follows that $tr(P_iR_i) = rank(P_iR_i) = rank(P_i) = T - \underline{K}$. Since P_i also satisfies $P_iW_i = 0$, we can choose $S_i = P_i$.

Relationship between M_i and P_i. Using (3.60) and the fact that $P_iW_i = 0$, it can be seen that

$$P_iM_i = P_i = M_iP_i. \tag{3.66}$$

Furthermore, post-multiplying (3.64) 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 (3.66) by R_iM_i

$$P_iM_iR_iM_i = P_iR_iM_i = M_iP_iR_iM_i = M_i^2 = M_i. \tag{3.67}$$

From (3.66) and (3.67), Searle and Quaas (1978) establish P_i as the Moore-Penrose inverse of $M_iR_iM_i$:

$$P_i = (M_iR_iM_i)^{+}. \tag{3.68}$$

Since $(M_i R_i M_i)^+$ is unique, equations (3.62) and (3.68) imply that

$$P_i = (M_i R_i M_i)^+ = A_i \left(A_i' R_i A_i\right)^{-1} A_i'. \tag{3.69}$$

Some Lemmas from Searle and Quaas (1978).

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

$$\begin{aligned} S_i' \left(S_i R_i S_i'\right)^{-1} S_i &= A_i F_i \left(F_i' A_i' R_i A_i F_i\right)^{-1} F_i' A_i' \\ &= A_i \left(A_i' R_i A_i\right)^{-1} A_i \qquad = \quad P_i. \end{aligned} \tag{3.70}$$

where the last equality follows from (3.69).

Lemma 2. As shown in Lutkepohl (1996, pag. 50, eq. 6), if A, B, C, and D are $(m \times m)$, $(m \times n)$, $(n \times m)$, and $(n \times n)$ matrices, respectively, then

$$\begin{aligned} det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \mid D \mid \cdot \mid A - BD^{-1}C \mid \quad if \quad D\, nonsingular \\ &= \mid A \mid \cdot \mid D - CA^{-1}B \mid \quad if \quad A\, nonsingular \end{aligned} . \tag{3.71}$$

Using this property of the determinant, we can show that

$$\mid A_i R_i A_i' \mid = \frac{\mid R_i \mid \cdot \mid W_i' R_i^{-1} W_i \mid}{\mid W_i' W_i \mid}. \tag{3.72}$$

To prove the latter, let

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

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 (3.69), 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 (3.64) into the latter equation and then using the following property of determinants, $det(AB) = det(A) \cdot det(B)$, yields (3.72).

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

$$| S_iR_iS_i' |=| F_i |^2| A_i'R_iA_i | . \tag{3.73}$$

Finding an expression for L_{1i}

Using (3.72) and (3.73), we have

$$log \mid S_iR_iS_i' \mid= \mu + log|R_i| + log \mid W_i'R_i^{-1}W_i|, \tag{3.74}$$

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

Furthermore, using (3.70), we get

$$\begin{aligned} \left(y_i - Z_i\gamma_i\right)' S_i' \left(S_iR_iS_i'\right)^{-1} S_i \left(y_i - Z_i\gamma_i\right) &= \left(y_i - Z_i\gamma_i\right)' P_i \left(y_i - Z_i\gamma_i\right) \\ &= \left(y_i - W_i\hat{\hat{\Gamma}} - Z_i\gamma_i\right)' R_i^{-1} \left(y_i - W_i\hat{\hat{\Gamma}} - Z_i\gamma_i\right). \end{aligned} \tag{3.75}$$

Substituting (3.74) and (3.75) into (3.18) yields (3.19).

Proof of Equation (3.75). Let $\hat{\bar{\Gamma}}$ 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 $R_i = var(\varepsilon_i)$.[31] The solution to the problem is given by

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

It follows that

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

Therefore, using (3.65) and after a few computations, we get

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

Finding an expression for L_{2i}.

The Choice of Q_i. It can be shown that $Q_i = W_i' R_i^{-1}$ satisfies $cov(S_i y_i, Q_i y_i) = 0$, and therefore is a plausible choice to obtain L_{2i}. We first compute the covariance conditional on γ_i, to then show that the unconditional covariance is equal to zero.

$$\begin{aligned} cov(S_i y_i, Q_i y_i \mid \gamma_i) &= E(S_i y_i y_i' Q_i' \mid \gamma_i) - E(S_i y_i \mid \gamma_i) E(y_i' Q_i' \mid \gamma_i) \\ &= S_i E(y_i y_i' \mid \gamma_i) Q_i' - \left(S_i \bar{Z}_i \gamma_i \right) \left(\bar{\Gamma}' W_i' + \gamma_i' \bar{Z}_i' \right) R_i^{-1} W_i, \end{aligned} \tag{3.76}$$

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

Substituting

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

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

and

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

into (3.76), we get

$$\begin{aligned} cov\left(S_iy_i, Q_iy_i \mid \gamma_i\right) &= -S_i\bar{Z}_i\gamma_i\bar{\Gamma}'W_i'R_i^{-1}W_i \\ &\quad -S_i\bar{Z}_i\gamma_i\gamma_i'\bar{Z}_i'R_i^{-1}W_i. \end{aligned} \tag{3.77}$$

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

$$\begin{aligned} cov\left(S_iy_i, Q_iy_i\right) &= E\left[cov\left(S_iy_i, Q_iy_i \mid \gamma_i\right)\right] \\ &\quad +cov\left(E\left(S_iy_i \mid \gamma_i\right), E\left(Q_iy_i \mid \gamma_i\right)\right). \end{aligned} \tag{3.78}$$

Taking expectation of both sides of (3.77), we get

$$E\left[cov\left(S_iy_i, Q_iy_i \mid \gamma_i\right)\right] = -S_i\bar{Z}_i\triangle\bar{Z}_i'R_i^{-1}W_i, \tag{3.79}$$

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

$$\begin{aligned} cov\left(E\left(S_iy_i \mid \gamma_i\right), E\left(Q_iy_i \mid \gamma_i\right)\right) &= E\left[S_i\bar{Z}_i\gamma_i\left(W_i'R_i^{-1}W_i\bar{\Gamma} + W_i'R_i^{-1}\bar{Z}_i\gamma_i\right)'\right] \\ &\quad -E\left[E\left(S_iy_i \mid \gamma_i\right)\right]E\left[E\left(Q_iy_i \mid \gamma_i\right)'\right] \\ &= S_i\bar{Z}_i\triangle\bar{Z}_i'R_i^{-1}W_i. \end{aligned} \tag{3.80}$$

Therefore, substituting (3.79) and (3.80) into (3.78) we can show that $cov\left(S_iy_i, Q_iy_i\right) = 0$.

3.10.2 Best Linear Unbiased Prediction

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

$$\begin{aligned} \hat{\gamma}_i = E(\gamma_i \mid y_i) &= E(\gamma_i) + cov(\gamma_i, y_i)\,[var(y_i)]^{-1}\,[y_i - E(y_i)] \\ &= \kappa' V_i^{-1}\left(y_i - W_i\bar{\Gamma}\right), \end{aligned} \tag{3.81}$$

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

$$\begin{aligned} var(\gamma_i \mid y_i) &= var(\gamma_i) - cov(\gamma_i, y_i)\,[var(y_i)]^{-1} \cdot cov(y_i, \gamma_i) \\ &= var(\gamma_i) - \kappa' V_i^{-1} \kappa. \end{aligned} \tag{3.82}$$

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

$$\begin{aligned} \triangle \bar{Z}_i' \left[\bar{Z}_i \triangle \bar{Z}_i' + R_i\right]^{-1} &= \left\{\left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)^{-1} \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)\right\} \cdot \\ &\quad \cdot \triangle \bar{Z}_i' \left[\bar{Z}_i \triangle \bar{Z}_i' + R_i\right]^{-1} \\ &= \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)^{-1} \cdot \bar{Z}_i' R_i^{-1}. \end{aligned} \tag{3.83}$$

This result is used to derive the second equality in equation (3.24) and to obtain equation (3.25).

Properties. Following Henderson (1984, Chap. 5), it can be shown that:

(i) $\hat{\gamma}_i$ is an unbiased predictor of γ_i:

$$\begin{aligned} E(\hat{\gamma}_i) &= E\left[\kappa' V_y^{-1}(y_i - W_i\bar{\Gamma})\right] \\ &= \kappa' V_y^{-1}\left[E(y_i) - W_i\bar{\Gamma}\right] \;=\; E(\gamma_i), \end{aligned} \tag{3.84}$$

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

(ii) $cov(\hat{\gamma}_i, \gamma_i) = var(\hat{\gamma}_i)$, from which it follows that $var(\hat{\gamma}_i - \gamma_i) = var(\gamma_i) - var(\hat{\gamma}_i)$.

(iii) the BLUP maximizes the correlation between $\hat{\gamma}_i$ and γ_i.

Finally, note that

$$\begin{aligned} var\left(\hat{\gamma}_i\right) = var\left[\kappa' V_i^{-1}(y_i - W_i\bar{\Gamma})\right] &= \kappa' V_i^{-1}\kappa \\ &= \triangle_i \bar{Z}_i' \left(\bar{Z}_i \triangle \bar{Z}_i' + R_i\right)^{-1} \bar{Z}_i \triangle \\ &= \left(\bar{Z}_i' R_i^{-1} \bar{Z}_i + \triangle^{-1}\right)^{-1} \bar{Z}_i' R_i^{-1} \bar{Z}_i \triangle. \end{aligned} \tag{3.85}$$

3.10.3 Expectation Step

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

$$E_{\theta^{(b-1)}}\left(\varepsilon_i' H_i \varepsilon_i \mid y_i\right) = Tr\left[H_i E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right)\right]. \tag{3.86}$$

To find $E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right)$, recall that for a random vector x, with mean μ_x and variance V_x, $var(x) = E(xx') - E(x)E(x')$, from which it follows $E(xx') = V_x + \mu_x \mu_x'$. Therefore,

$$E_{\theta^{(b-1)}}\left(\varepsilon_i \varepsilon_i' \mid y_i\right) = V_{\varepsilon_i} + \hat{\varepsilon}_i \hat{\varepsilon}_i', \tag{3.87}$$

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

$$\begin{aligned} V_{\varepsilon_i} = var\left(\varepsilon_i \mid y_i; \theta^{(b-1)}\right) &= var\left(y_i - W_i\bar{\Gamma} - \bar{Z}_i\gamma_i \mid y_i, \theta^{(b-1)}\right) \\ &= \bar{Z}_i V_{\gamma_i}^{(b)} \bar{Z}_i', \end{aligned} \tag{3.88}$$

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 (3.87) into (3.86) yields

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

We can now write

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

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

3.10.4 Estimation of the Coefficient Covariance Matrix

An estimator of $\triangle$ can be obtained by maximizing $\sum_{i=1}^{N} Q_{3i}$, where Q_{3i} is defined in (3.28), with respect to $\triangle$. Before proceeding, we report a few results of matrices differentiation shown in Lutkepohl (1996).

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

$$\frac{\partial a' X^{-1} b}{\partial X} = -(X^{-1})' ab' (X^{-1})'. \tag{3.89}$$

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

$$\frac{\partial tr\left(AX^{-1}B\right)}{\partial X} = -\left(X^{-1}BAX^{-1}\right)'. \tag{3.90}$$

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

$$\frac{\partial ln \mid X \mid}{\partial X} = (X')^{-1}. \tag{3.91}$$

Therefore,

$$\frac{\partial Q_{3i}}{\partial \triangle} = \underset{(3.91)}{-\triangle^{-1}} + \underset{(3.90)}{\triangle^{-1} V_{\gamma_i}^{(b)} \triangle^{-1}} + \underset{(3.89)}{\triangle^{-1} \hat{\gamma}_i^{(b)} \hat{\gamma}_i^{(b)\prime} \triangle^{-1}} = 0,$$

The solution to $\left(\partial \sum_{i=1}^{N} Q_{3i}/\partial\triangle\right) = 0$ is given by

$$\hat{\triangle} = \frac{1}{N}\sum_{i=1}^{N}\left\{V_{\gamma_i} + \hat{\gamma}_i\hat{\gamma}_i'\right\}. \tag{3.92}$$

Unbiased Estimator. It can be shown that

$$\hat{\triangle} = \frac{1}{N}\sum_{i=1}^{N}\left\{V_{\gamma_i} + \hat{\gamma}_i\hat{\gamma}_i'\right\} \tag{3.93}$$

is an unbiased estimator of $\triangle$ since

$$\begin{aligned} E\left(\hat{\triangle}\right) &= N^{-1}\sum_{i=1}^{N}\left\{E\left(\hat{\gamma}_i\hat{\gamma}_i'\right) + E\left(V_{\gamma_i}\right)\right\} \\ &= N^{-1}\sum_{i=1}^{N}\left\{E\left[\kappa' V_i^{-1}\left(y_i - W_i\bar{\Gamma}\right)\left(y_i - W_i\bar{\Gamma}\right)' V_i^{-1}\kappa\right] + \triangle - \kappa' V_i^{-1}\kappa\right\} \\ &= N^{-1}\sum_{i=1}^{N}\left\{\kappa' V_i^{-1}\kappa - \kappa' V_i^{-1}\kappa\right\} + \triangle = \triangle. \end{aligned}$$

3.10.5 Hypothesis Testing

Covariance of Estimator of Fixed Coefficients.

Noting that $V = var\left(y\right)$ has the linear form

$$V = \sum_{s=1}^{\bar{r}} \vartheta_s \Pi_s,$$

the adjusted estimator of the small sample variance-covariance matrix of $\tilde{\bar{\Gamma}}$ is given by

$$\hat{\Phi}_A = \hat{\Phi} + 2\hat{\Lambda}, \tag{3.94}$$

where

$$\begin{aligned} \hat{\Lambda} &= \hat{\Phi}\left\{\sum_{s=1}^{r}\sum_{j=1}^{r}\Upsilon_{sj}\left(\Xi_{sj} - \varSigma_s\hat{\Phi}\varSigma_j\right)\right\}\hat{\Phi}, \\ \varSigma_s &= -\sum_{i=1}^{N} W_i' V_i^{-1}\Pi_{s,i}V_i^{-1}W_i, \\ \Xi_{sj} &= \sum_{i=1}^{N} W_i' V_i^{-1}\Pi_{s,i}V_i^{-1}\Pi_{j,i}V_i^{-1}W_i. \end{aligned}$$

An approximation to Υ, the variance-covariance matrix of $\hat{\vartheta}$, can be obtained from the inverse of the expected information matrix I_E, where

$$2\left\{I_E\right\}_{sj} = tr\left(\sum_{i=1}^{N} V_i^{-1}\Pi_{s,i}V_i^{-1}\Pi_{j,i}\right) - 2tr\left(\Phi\Xi_{sj}\right) + tr\left(\Phi\Sigma_s\Phi\Sigma_j\right).$$

Detailed derivations are provided by Alnosaier (2007).

Assessing the Errors of Estimation for the Unit-Specific Coefficients.

The variance-covariance matrix of the predictor of (3.5), is given by

$$\begin{aligned} var\left(\hat{\psi}_{1i} - \psi_{1i}\right) &= F_{1i}var\left(\hat{\bar{\Gamma}}\right)F_{1i}' + var\left(\hat{\gamma}_i - \gamma_i\right) + F_{1i}cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \hat{\gamma}_i - \gamma_i\right) \\ &\quad + \left[F_{1i}cov\left(\hat{\bar{\Gamma}} - \bar{\Gamma}, \hat{\gamma}_i - \gamma_i\right)\right]', \end{aligned} \tag{3.95}$$

where

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

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

3.11 Data

3.11.1 List of Countries

Advanced Economies: Australia (AU), Austria (OE), Belgium (BG), Canada (CN), Denmark (DK), Finland (FN), France (FR), Greece (GR), Iceland (IC), Ireland (IR), Italy (IT), Japan (JP), Netherlands (NL), New Zealand (NZ), Norway (NW), Portugal (PT), Singapore (SP), Spain (ES), Sweden (SD), Taiwan (TW), United Kingdom (UK).
Emerging Market and Developing Economies: Argentina (AG), Brazil (BR), Chile (CL), China (CH), Croatia (CT), Hungary (HN), India (IN), Malaysia (MY), Mexico (MX), Peru (PE), Philippines (PH), Poland (PO), Russia (RS), South Africa (SA), Thailand (TH), Turkey (TK), Venezuela (VE).

The classification of countries follows from IMF, World Economic Outlook, October 2015 (pag.187-188).

3.11.2 Data Sources

Bond Yields: J.P. Morgan EMBI Global, OECD Main Economic Indicators, Eurostat (for DK, GR, LX, and PT), and national authorities (OE, IN, IT, SP, SD, TW).
Bond Spreads: for all European countries but Iceland, the bond spread is measured against German long-term government bond yields. For the remaining countries, the bond spread is measured against US long-term government bond yields.
Current Accounts: OECD Main Economic Indicators, Oxford Economics, and national Central Banks (LV, PE).
Government Debt: Oxford Economics, Eurostat (LV, LX, SJ), and Bank for International Settlements (IR, IS, NZ, PE).
CPI inflation: IMF - International Financial Statistics, OECD Main Economic

Indicators (AG, CL, CH, SX), and Oxford Economics (SP, TW, TH).

Real GDP: Oxford Economics, national authorities (IS, LV, LX, NZ, PE), and OECD Main Economic Indicators (SJ).

Exchange Rates: IMF - International Financial Statistics, OECD Main Economic Indicators, and Oxford Economics (TW).

Financial History: Historical time series on countries creditworthiness and financial turmoil are obtained from Reinhart and Rogoff (2009, 2011).

Chapter 4

House Prices and Monetary Policy in the Euro Area: Evidence from Structural VARs

4.1 Introduction

In the light of the recent global financial crisis, it is crucially important to understand the role that house prices played in the past and the linkages between housing, monetary policy and macroeconomic activity in general, in order to detect future housing imbalances and to improve financial stability. As a result, the literature on housing in macroeconomics has grown very rapidly in recent years.[1] Nevertheless, most of the current studies focus on the aggregate euro area, the UK and the US.[2] Yet, little is known about the effects of house prices in each euro area member states. A notable exception is Giuliodori (2005) although this work only covers the pre-EMU period.[3] The first contribution of this chapter is

[1] An excellent survey is provided by Piazzesi and Schneider (2016).

[2] Notably, Iacoviello (2005), Iacoviello and Neri (2010), Jarocinski and Smets (2008), and Mishkin (2007) for the US. Musso et al. (2011) compare the US with the (aggregated) euro area.

[3] Using a recursively identified VAR, the author focuses on the role of house prices in the monetary transmission and provides some evidence on their effects on household consumption

to fill this gap. We use a structural Bayesian vector autoregression model for seven euro-area countries (Belgium, France, Germany, Ireland, Italy, the Netherlands, and Spain) for the period 1980:Q1-2014:Q4. We focus on a country by country analysis, given the idiosyncratic characteristics of the housing market in the euro area (ECB, 2003) which suggest that pooling or aggregating may lead to biased inference (Pesaran and Smith, 1995) and misleading policy recommendations. Adopting a novel set of identification restrictions which combines zero and sign restrictions, based on the algorithm developed by Arias et al. (2014), we provide a systematic structural analysis of the effects of housing demand shocks on economic activity and the role of house prices in the monetary transmission mechanism across euro area countries. The combination of zero and sign restrictions allows us to distinguish between a housing demand shock and an aggregate demand shock. Disentangling these two shocks would be less than obvious if we were to use only sign restrictions. The priors are selected using the Bayesian stochastic search variable selection (SSVS) approach developed by George, Sun and Ni (2008). This method allows for shrinkage of the VAR coefficients (to overcome the over-parameterization problem) while selecting restrictions that are supported by the data itself. This in turn allows appropriate finite sample inference and exploits in full the intrinsic cross-country heterogeneity typical of the housing market.

Second, using to the extent possible a dataset composed of comparable data sources, sample periods, and mostly by employing the same econometric methodology for each country, we exploit the cross-sectional dimension of our data to quantify the degree of heterogeneity of the impact of housing demand shocks on the macro-economy and the role of house prices in the transmission of monetary policy across Eurozone members.[4] In fact, the current literature lacks of such comparative studies, especially with respect to the role of housing wealth

while leaving aside their role in driving other important variables such as GDP, inflation, lending rates and most importantly the supply of credit.

[4]A description of the data sources is presented in Appendix 4.8.1.

on economic activity. The investigation of heterogeneity in the euro area housing markets is clearly relevant from a policy perspective. Given the ongoing recovery in house prices, it is fundamental to ask what are the implications for the broader macro-economy and to investigate how the heterogenous impacts of house prices across countries can amplify the existing economic divergences across Eurozone members.[5] House prices in the euro area appear to be currently supported by several factors: favourable financing conditions, the enfolding recovery in growth and employment and a low yield environment which makes housing investment relative attractive compared to alternative asset classes (ECB, 2015). A protracted increase in house prices is therefore foreseeable and its macroeconomic implications need to be carefully assessed by policy makers. Moreover, considering the numerous interactions - of a real and financial nature - characterizing the housing market (Wachter, 2015), it is inevitable that the two aforementioned questions - how housing demand shocks affect the macro-economy and how monetary policy affects house prices - are intrinsically interrelated. In this context, house prices, like other asset prices, represent a potentially important component in monetary policy transmission, to the extent that changes in interest rates and other (non-standard) monetary policy measures affect house prices, thereby influencing economic activity and private consumption.

With regard to the transmission of monetary policy, Calza et al. (2013) conduct an analysis similar to our investigation. Accounting for heterogeneity in the estimation, the authors classify 19 advanced economies into two groups according to the degree of development of mortgage markets and the type of interest rate structure, to examine whether the national mortgage markets' institutional characteristics influence the effects of monetary shocks. Identification of the latter is achieved via Cholesky decomposition. Differently from them, we ask whether a common monetary policy could amplify divergences in house prices fluctuations

[5] The focus is on the implications in terms of real GDP, real private consumption, inflation and credit developments.

among the Eurozone members when reacting to area wide aggregates such as inflation and economic activity. As noted in Bini Smaghi (2011), given its primary objective of maintaining price stability in the euro area, the ECB has 'no choice but to take a euro area perspective'. Therefore, since its policy decisions aim at price stability at the area-wide level and 'cannot be tailored to the specific needs of a single Member State', it is important to quantify and compare the different effects of monetary shocks on house price dynamics across euro area countries, regardless of the degree of development of national mortgage markets, which according to Cardarelli et al. (2008) is rather low in all countries under study, with the exception of the Netherlands.[6] Being aware of such heterogeneity is essential when addressing real and financial imbalances at the country level by means of macroprudential policies.[7]

Estimation results confirm that the effects of housing demand and monetary policy shocks differ widely across the countries under investigation. We find evidence of the existence of a housing wealth effect in the euro area, although with a certain degree of heterogeneity in the response of household consumption to house price increases. While there is a broad consensus on the housing wealth effect in the U.S. (e.g. Iacoviello and Neri, 2007), and the U.K. (e.g. Campbell and Cocco, 2007), it is argued that such an effect is relatively modest in the euro area (ECB, 2009). Although this is true for many countries under investigation, the same cannot be said for Ireland and Spain, where we show that an increase in real house prices has a positive and statistically significant impact on real private consumption. Both countries have recently experienced a boom-bust pattern in house prices. This finding supports the view (e.g. Shiller, 2005) that house price booms play an important role in boosting confidence, which in turn

[6]For instance, if we were to classify the countries examined according to their Loan-to-Value ratio following Calza et al. (2013), all countries would belong to the "low development" group but Belgium and the Netherlands. If we were to use mortgage equity withdrawal, only the Netherlands would belong to the 'high development' group. More heterogeneity is found in terms of the type of interest rate structure, with Italy and Spain being characterized by variable rates, although this is a recent development.

[7]On this topic, see for example Schoenmaker (2014) and Hartmann (2015).

stimulate consumption. The historical decomposition analysis corroborates the importance of housing demand shocks in driving consumption, especially in Ireland and Spain. In both countries, in the absence of housing demand shocks, the growth rates of real private consumption would have been lower than the actual rates between 2002 and 2007, and larger between 2008 and 2013. To illustrate, in Ireland, the cumulative effect of housing demand shocks on consumption is equal to 0.79% up to 2006 and to -1.16% at the end of 2011. In Spain, it is equal to 0.5% up to 1995 and 2004, and to -0.66% in 2012. Housing demand shocks also play an important role in explaining variation in the supply of credit, confirming the 'financial accelerator' hypothesis, according to which changes in the collateral affect borrowing capacity (Bernanke et al. (1996) and Almeida et al. (2006)). The impact of housing demand shocks on loans is less heterogeneous than the effect on consumption. Furthermore, we corroborate the strong role of house prices in the monetary policy transmission for the euro area while documenting high heterogeneity in the impact of monetary policy shocks on euro area countries' house price fluctuations. Historical decompositions highlight a strong contribution of monetary shocks to real house price growth. A substantial increase in house prices would have occurred in Ireland and Spain between 2001 and 2006 even if all the other structural shocks but monetary policy had been turned off.

The remainder of the chapter is organised as follows. Section 4.2 presents the data used and some stylized facts. In Section 4.3 we provide a brief review of the literature on housing in macroeconomics. Section 4.4 describes the methodology used. Section 4.5 presents results of the structural VAR analysis. In particular, we highlight the strategy used to identify the structural shocks and describe the main findings from impulse response analysis, the forecast error variance decomposition, and historical decomposition. Finally, we conclude.

4.2 Data and Stylized Facts

Some Stylized Facts. The house price cycle has turned the corner in the euro area. The annual rate of change in euro area house prices started to increase in mid-2013 and turned mildly positive since the second half of 2014, subsequently reaching a post crisis high at the beginning of 2016. This aggregate trend follows heterogeneous developments across euro area countries. Large downward adjustments in real house prices took place in Spain and Ireland since the beginning of 2008, where prices declined around 40% from the peaks reached before the financial crisis. Sizable declines in excess of 20% were also experienced in the same period in the Netherlands and Italy, while real house prices were broadly stable in Belgium and increased notably in Germany by 27%. Indeed, the building and subsequent correction of house price imbalances – typical of a boom-bust pattern characterising the housing market - renders the observed aggregate recovery relatively muted and characterised by a differentiated pace across countries. In this context, exploring how house price dynamics affect the macroeconomy and how monetary policy influence house prices appear of particular interest. To further grasp the importance of housing for the macro-economy and put things into perspective, housing wealth in the euro area represents, on average, 37% of households' net worth. In turn, at the end of 2014, real estate-related loans to households and non-financial firms in the euro area accounted for nearly 57% of euro area banks' total loans to the non-financial private sector and more than half of euro area GDP (ECB, (2015)).

Data. A detailed description of the database used in the descriptive and econometric analysis is provided in Appendix 4.8.1. In Appendix 4.8.2, we provide charts which depict the variables of interest for the seven euro area countries examined (Belgium (BE), France (FR), Germany (DE), Ireland (IE), Italy (IT), Spain (ES), and the Netherlands (NL)) and, for illustrative purposes, the euro area. The variables are: real house prices, consumer price inflation, real GDP, real loans to households, lending rates and monetary policy rates. All variables are shown as an index level, except for the interest rates which are in percentage terms. From a first graphical inspection of the data some interesting points emerge. First, the heterogeneity in real house price dynamics across the panel of euro area countries is broadly matched by qualitatively similar dynamics in real loans to households, generally suggesting a high degree of co-movement between the two variables. In particular, countries experiencing boom-bust episodes in house prices (such as Ireland and Spain) have also undergone sharp increases in credit to households before the financial crisis followed by reversals after the crisis. This also holds to some extent for Italy and the Netherlands. At the same time, sustained house price dynamics in France and Belgium, especially in the latter part of the sample period, have been accompanied by continued growth in loans to households. On the contrary, in the case of Germany, declining or subdued house price dynamics for a large part of the sample period have been matched by a modest increase in loans to households. Second, a certain degree of co-movement between real house prices and economic activity in terms of real GDP is also evident, in particular in the case of Ireland, Spain and the Netherlands. The findings described above are confirmed when looking at cross correlations. The alignment between real house prices and the business cycle (in terms of real GDP) is highest in terms of maximum correlation in Spain, Ireland and the Netherlands and it is found at broadly coincident level. The maximum correlation between annual real house prices and real loans to households is found in the case of France, Ireland, Spain (around 70%) and Belgium and the Netherlands (around 50-60%). In the case of the Netherlands, Spain and Belgium real house price growth tends to slightly lead annual growth in real loans to households. Third, consumer prices were characterised by a much lower degree of cross-country heterogeneity compared to house prices, as well by more moderate increases or less pronounced falls. Finally, the well-known downward trend in lending rates is evident, notwithstanding some volatility in the initial part of the sample – before 1999 – characterised by different monetary policy regimes.

4.3 Literature

In this section, we briefly review the main theories on the role of house prices in the business cycle and in the transmission of monetary policy. A non exhaustive list of works which use multivariate structural models to quantify the impact of housing demand and monetary policy shocks is also reported.

4.3.1 The Interaction between House Prices and the Business Cycle

How do house price fluctuations affect households' consumption decisions? House price changes may have significant effects on aggregate consumption through different channels. First, an increase in house prices leads to a rise in homeowners' financial wealth which is the sum of liquid financial assets and real estate's value minus outstanding debt. However, as noted in Campbell and Cocco (2007), such an increase does not necessarily correspond to a raise in real wealth and therefore may have no effect on consumption. In fact, a house price increase does not affect the consumption behavior of a homeowner who is not planning to sell his house. It is just a compensation for a higher implicit rental cost of living in the house as pointed out in Sinai and Souleles (2005). The age structure of the population may play a role. While Campbell and Cocco (2007) find that house price increases benefit mostly old owners, rather than young renters, confirming the so-called 'wealth hypothesis', Attanasio et al. (2009) findings support the so called 'common factor hypothesis': the impact of house prices on consumption is the same across different age groups.

Second, even in the absence of wealth effect, an increase in house prices may lead to an increase in consumption, since housing can be used as collateral in a loan. It therefore allows borrowing constrained homeowners to smooth consumption over the life cycle, as shown in Ortalo-Magne and Rady (2006) and Lustig and Van Nieuwerburgh (2006). As argued in Almeida et al. (2006), if the collateral-based accelerator theory were to hold, one should expect a larger increase in consumption (following an increase in house prices) in high loan-to-value (LTV) ratio countries. In fact, as stated by the authors 'the procyclical increase in borrowing capacity may allow households to further increase housing spending, amplifying the collateral-based spending cycle'. Countries with a high LTV ratio are characterized by higher marginal opportunity to borrow. This argument is also made in Muellbauer (2015), where the author argues that in countries with low first-time buyer FTB-LTV ratio, higher house prices may have a negative impact on aggregate consumption if they are not accompanied by higher income or income growth expectations. The main reason is that those who want to become owner-occupiers need to save more while renters anticipate higher rents in the future which therefore negatively affects their spending decisions.

House prices and economic activity. House price shocks may have a positive impact on GDP through higher consumption since, as discussed above, an increase in house prices implies a higher value of colletaral which can be used by a borrowing-constrained households to obtain more credit. Furthermore, due to the 'Tobin's q' effect, a rise in house prices encourages companies to invest more in housing construction (because their market value is higher than their construction costs) which in turn affects real growth. At the same time, housing demand shocks may exercise upward pressure on inflation directly trough higher rents (which are a component of CPI services inflation) and indirectly through consumption. The impact on inflation through higher rents should be larger in those countries where home-ownership is lower. According to the 'financial accelerator' theory, the (indirect) impact on CPI (through consumption) should be bigger in countries with higher LTV ratio.

House prices and the market for loans. As discussed in Basten and Koch (2016), different mechanisms are at play in the relationship between house prices and mortgage volumes. First, an increase in house prices which is not accompanied by a contemporaneous increase in households wealth may induce those who are seeking to buying to resort to more loans when purchasing a new housing. This increase in demand will result in higher equilibrium mortgage amounts, even in the absence of an outward shift in the mortgage supply curve. At the same time, given that the value of the collateral has increased, banks may be more willing to extend loans, especially if they expect future house prices to grow further. In such a case, an increase in house prices can also cause a shift of the credit supply curve. The subsequent consequences on lending rates depend on many factors among which, the availability of credit and regulatory capital ratio requirements, the risk perception of potential borrowers or the degree of competition in the banking sector.

4.3.2 The Role of Monetary Policy Shock for House Price Fluctuations

Bernanke and Gertler (1995) argue that credit market frictions can have a relevant impact on households' borrowing and spending decisions on durable items such as houses which in turn affect residential investment and therefore aggregate economic activity. Monetary policy can affect residential investment through the balance sheet channel. In fact, the authors note a direct link between housing demand and consumer balance sheets, due to features such as 'down-payment requirements, up-front transaction costs, like closing costs and "points" and minimum income-to-interest-payment ratios'. The lending channel also plays a role. According to Iacoviello and Minetti (2008), in the occurrence of a liquidity shock, banks may tend to shift from less to more liquid loans or to securities. Therefore, the relative illiquidity of mortgages becomes crucial especially in those countries where mortgage standardisation and securitisation are not common. At the same time, a fall in bank mortgages will result in a shortage of funds for house purchases, especially in those countries in which the supply of loans from specialist mortgage lenders or from the state is not enough to satisfy the demand for housing purchases.

4.3.3 Selected Empirical Evidence

In this subsection, we present a non-exhaustive overview of the empirical literature on housing and the monetary policy transmission.The studies presented in Table 4.1 differ in terms of methodology, country coverage, and sample periods. Therefore, their comparability is inevitably limited. We focus only on the literature which derives insights on the quantitative importance of different mechanisms from multivariate structural models. Most of the works reviewed here use VAR models estimated using classical inference, with a few exceptions. Jarocinski and Smets (2008) use two Bayesian VAR specifications: a VAR in levels which uses standard Minnesota priors and one in differences with priors about the steady state, as in Villani, (2008). Goodhart and Hofmann (2008) use a Fixed-effects panel VAR. Iacoviello and Minetti (2008) estimate both VAR and VEC models.

4.4 The Bayesian SSVS-VAR Model

We run a Bayesian VAR model for each country, namely Belgium, France, Germany, Ireland, Italy, the Netherlands, and Spain. VAR models have been widely used in the study of house prices and monetary policy, given that linear interdependencies may exist among the time series under study, and because of their ability to forecast and quantify impulse responses to macroeconomic shocks, among other reasons. The choice of countries is dictated by three reasons. First, we focus on a country dimension given the intrinsic idiosyncratic nature of housing markets across the euro area members. Second, the country coverage is influenced by the

Table 4.1: Selected Empirical Evidence

Study	Country Coverage	Sample Period	Identification	Shocks	Confidence Bands
Aoki et al. (2002)	UK	1975:Q1 - 1999:Q4	(a)	MP	2 Standard Errors
Bjornland and Jacobsen (2010)	NO, SE, UK	1983:Q1 - 2006:Q4	(b)	MP, HP	68%
Calza et al. (2013)	19 advanced countries	1980:Q1 - 2007:Q4	(a)	MP	2 Standard Errors
Elbourne (2008)	UK	1987:Jan - 2003:May	(d)	MP, HP	90%
Giuliodori (2005)	9 European countries	1979:Q3 - 1998:Q4	(a)	MP, HP	90%
Goodhart and Hofmann (2008)	17 industrialized countries	1973:Q1 - 2006:Q4	(e)	-	-
Iacoviello and Minetti (2008)	FI, UK, DE & NO	1974:Q2 -1999:Q4	(a) & (b)	MP, Mix	1 Standard Error
Jarocinski and Smets (2008)	US	1987:Q1 - 2007:Q2	(c)	HP, MP & TS	68%
Musso et al. (2011)	US & (aggregated) EA	1986:Q1 - 2009:Q2	(a)	MP, HP & CS	68%

Identification Strategy: (a) stands for recursively identified system (Choleski decomposition); (b) for mix of short and long-run restrictions; (c) for mix of zero and sign restrictions; (d) Kim & Roubini (2000) approach; and (e) Reduced-Form analysis only. MP, HP and CS stands fo monetary policy, housing demand and credit supply shocks, respectively. TS is the term spread computed as the difference between long-term interest rates and federal funds rates. Mix denotes the external finance mix, which is the fraction of housing loans by 'non-banks'. Iacoviello and Minetti (2008) use (b) to identifying MP shocks, and (a) for Mix. EA, FI, DE, NO, and SE are an abbreviation of Euro Area, Finland, Germany, Norway and Sweden, respectively.

need of sufficiently long time series and reliable house price data. Finally, differently from the current literature, we are interested on a cross-country comparison for euro area countries rather than focusing on the euro area as a whole.

For each country, the reduced form VAR(p) model van be written as

$$y_t = \mu + A_1 y_{t-1} + ... + A_p y_{t-p} + u_t, \tag{4.1}$$

for $t = 1, .., T$, where $u_t \sim N(0, \Sigma_u)$ and y_t is a $m \times 1$ vector of endogenous variables.[8]

The vector of endogenous variables in our baseline VAR model includes lending rates to households (for house purchase), national banks' official rates (starting from 1999, we use the ECB rate on the marginal lending facility) and (annualized) growth rates of real house prices, real consumption (or alternatively real GDP), the consumer price index (CPI), and real loans to households. The choice of variables is in line with Giuliodori (2005) and Musso et al. (2011), among others.

[8]The subscript i, denoting the particular country of interest, is omitted for clarity of exposition

Equation (4.1) can be rewritten in compact form as

$$\mathbf{Y} = XA + U, \tag{4.2}$$

where $\mathbf{Y}$ and U are two $T \times m$ matrices, X is of dimension $T \times K$ and $A = (\mu, A_1, .., A_p)'$ is a $K \times m$ matrix of coefficients, with $K = (mp+1)$. The estimation sample is 1980Q1:2014Q4.[9] The lag order of the model for each country has been chosen using the Akaike information criterion.[10]

4.4.1 The Choice of the Prior

When the number of observations is limited, the number of parameters to be estimated may be too large relative to the available data. In the absence of restrictions in the regression coefficients and the covariance matrix, the model is over-parameterized. Consequently, the precision of inference and the reliability of prediction are negatively affected.

To overcome this problem, the Bayesian approach has become widely used for VAR modelling, as it incorporates prior knowledge about parameter values. Various priors for unrestricted and restricted VARs which allow for shrinkage of the coefficients have been proposed. Prior elicitation is typically based on the ground of formal or informal economic theory or using information about pattern on macroeconomic data. For instance, Doan et al. (1984) suggested a Minnesota prior that shrinks the VAR parameters towards a random walk model. However, as noted by George, Sun and Ni (2008), such approaches "are based on an implicit assumption that the relevant restrictions are known" even though "at least for some economic problems, current theoretical knowledge does not warrant such confidence". Moreover, Koop and Korobilis (2010) note that they require "substantial prior input from the researcher (although this prior input can

[9] The sample periods vary across countries depending on data availability.

[10] We choose to estimate the optimal lag order using Aikake rather than Schwartz criterion, as the former always yields a larger order.

be of an automatic form such as in the Minnesota prior)". In view of the above reasons, following George, Sun and Ni (2008), we use a Bayesian stochastic search approach (SSVS) to selecting restrictions for VAR models that are supported by the data itself. It does so in an automatic fashion by using a hierarchical model, where the prior for a parameter is a function of a hyperparameter which in turn has its own prior. Therefore, it allows us to impose plausible restrictions on both the covariance matrix and the VAR regression coefficients while requiring "minimal prior input from researcher" (Koop and Korobilis, 2010).

In particular, let $\alpha = vec(A)$ be the $Km \times 1$ vector of regression coefficients. The SSVS assumes that the prior distribution of α_j (the jth element of α) is a mixture of two Normal distributions:

$$\alpha_j \mid \gamma_j \sim (1-\gamma_j) \cdot N\left(0, \tau_{0j}^2\right) + \gamma_j \cdot N\left(0, \tau_{1j}^2\right), \tag{4.3}$$

where γ_j is a dummy variable; τ_{0j} is set to be small and τ_{1j} large (for $j = 1, .., Km$) so that α_j is restricted to be very close to zero when $\gamma_j = 0$ and unrestricted when $\gamma_j = 1$. The dummy γ_j is unknown and it has to be estimated in a data-based fashion. In particular, it is assumed that the γ_j's are independent Bernoulli random variables so that

$$P(\gamma_j = 1) = p_j, \quad P(\gamma_j = 0) = 1 - p_j, \quad j = 1, .., Km.$$

As noted by George, Sun and Ni (2008), for each j, p_j reflects the prior belief that α_j should be unrestricted. In the absence of such prior information, one could set $p_j \equiv .5$ and let the data decide whether to shrink or not the coefficient to zero, as we do in this chapter. A similar prior for Σ_u is assumed, in order to impose restriction on the covariance matrix. We refer the reader to Appendix 4.7.1 and to George, Sun and Ni (2008) for details.

Bayesian stochastic search approach is advantageous to select restrictions, shrinking many coefficients to zero, while providing relative probabilities of the selected models. Therefore, it helps researchers to focus on the more realistic submodels and in turn to make adequate finite sample inference. It differs from previous VAR modeling approaches as "it does not a priori rule out submodels of the VAR under consideration. Instead, it allows for the comparison of submodels based on the data". In fact, as noted in Koop and Korobilis (2010), the result of the SSVS-MCMC algorithm will be Bayesian model averaging (BMA). At the same time, using simulated numerical examples, George, Sun and Ni (2008) find that their model performs well in selecting a satisfactory model and lead to improvements in forecasting in terms of Mean Squared Errors.

4.5 Structural Analysis

4.5.1 Identification Strategy

To investigate the heterogeneous effects of house prices on the macro-economy, and their role in the monetary policy transmission across euro area countries, we identify both hous- ing demand and monetary policy shocks. Intuitively, a housing demand shock is mainly attributable to households' preferences. Such a shock would increase the relative attractiveness of housing vis-a-vis other goods/services, for example via a more favourable tax treatment or deductibility of mortgage expenditures, or in terms of improved location due to enhanced services and amenities (think of a new underground project connecting the suburb of a city with its centre). Such a shock could also be interpreted as a preference shock resulting in changes in the political and social environments that encourage an increase in home-ownership as in Baldi (2014). In the case of Spain, Aspachs-Bracons and Rabanal (2011) interpret such a shock as driven by population changes: increased immigration, the baby boom generation, and social changes that reduce the number of persons per households and increase the number of household units.

Since the works of Faust (1998), Canova and De Nicolo (2002), and Uhlig

Table 4.2: Short Run Responses to Housing Demand and Monetary Policy Shocks

	Housing Demand	Mon Pol	Loans Supply	Lend Rates	A. Supply	A. Demand
House Prices	+	−				
Monet. Rate	0	+	0	0		
Loans	+	−	+	−		
Lending Rates	+	+	−	+		
Consump. (GDP)	0	0			+	+
Inflation	0	0			−	+

The first column lists the endogenous variables of the VAR, which react to the shocks reported in the first row: housing demand shocks, monetary policy innovations, shocks to the credit supply (third and fourth column), aggregate supply and demand shocks. Our interest is in identifying housing demand and monetary policy shocks.

(2005), identification via sign restrictions has become increasingly popular (see Fry and Pagan (2011) for a review). We identify housing demand and monetary policy shocks by using a combination of zero and sign restrictions, using the algorithm proposed by Arias et al. (2014). The matrix of contemporaneous impacts of the shocks on the endogenous variables is defined in Table 4.2.[11]

Identification of housing demand shocks. Among other reasons, combining zero with sign restrictions allows us to specify enough information to discriminate between a housing demand shock and an aggregate demand shock. This distinction would have not been possible by simply using an identification strategy which only imposes sign restrictions.

The assumption that real consumption (or real GDP) and inflation do not react on impact to a house price shock, captures the idea of stickiness in the transmission of the shock due, for example, to the transaction time required to buy/sell a property and/or to a lagged or muted reaction of rents affecting inflation with some delay. In fact, Muellbauer (2015) note that rents (both private and commercial) adjust relatively slowly to an increase in house prices. A slug-

[11]Identification of housing demand shocks does not require imposing any restrictions in rows 3 to 6 of the second column, as well as no restrictions in the (2,3) and (2,4) elements of the matrix shown in Table 4.2. Similarly, when identifying monetary policy shocks, we do not impose any restriction in rows 3 to 6 of the first column and in the (4,3) and (3,5) elements of the matrix. We do so to facilitate replicability of results. In fact, imposing all the restrictions is computationally costly as identification of each country's VAR shocks would require around two days.

gish response of inflation to house prices is also found in Bjornland and Jacobsen (2010). This assumption is also used in Jarocinski and Smets (2008) and it is in line with Giuliodori (2005) and Musso et al. (2011), who imposes a recursive structure in which house prices are ordered after GDP and inflation. The patterns used to distinguish aggregate demand and supply shocks are commonly used in the literature (e.g. Fry and Pagan (2011)). The zero contemporaneous impact of a house demand shock on monetary policy is consistent with a Taylor Rule. Furthermore, by imposing sign restriction rather than simply assuming a recursive causal structure of the system (e.g Sims, 1980), we are able to discriminate house prices shocks from loans supply and lending rates shocks on the ground of economic theory. Instead, it would be more difficult to find an appropriate theoretical justification in what order those variables are recursive. We assume that a house demand shock causes a contemporaneous increase in both real loans and lending rates. To understand the latter hypothesis, we look at the market for loans, and suppose demand and supply are in equilibrium. An exogenous housing demand shock may shift up the demand curve and down the supply curve (given that the value of housing collateral increases). We assume that the shift in the demand curve will be higher than the shift in supply because of regulatory requirements and balance sheets conditions that banks have to satifsy (e.g. the availability of mortgage credit is limited to a maximum LTV ratio, Almeida et al. (2006)) and given that some borrowers may be highly leveraged (e.g. their debt-to-income ratio is quite high even before applying for a new loan). As a result, the new equilibrium will be characterized by higher market lending rates and an increase in the volume of loans.[12] This is in line with Jarocinski and Smets (2008) and Iacoviello and Neri (2010), who assume that an increase in real house prices is not associated with a fall in nominal short-term interest rate to rule out an expansionary monetary poliy shock. This assumption allows us to distinguish

[12]For example, Basten and Koch (2016) studying the causal effect of house prices on the mortgage market in Switzerland, find that higher house prices lead to an increase in mortgage demand which is not accompanied by an expansion in mortgage supply.

between housing demand shocks and a (positive) loans supply shock. The latter may be associated to various events, such as changes in regulatory capital ratio requirements which increase the amount of banks capital available for loans.[13]

Identification of a monetary policy shock. When recoverying the monetary policy (MP) shock, combining zero and sign restrictions allow us to be consistent with both the literature which studies systematic changes in monetary policy rules (e.g. Christiano et al., 1999) and the one which focuses on the role of asset prices in the transmission of MP shocks (e.g. Zettelmeyer (2004), Rigobon and Sack (2004) and Kuttner (2005)). In line with the former, we assume that central banks react endogenously to conemporaneous movement in current prices and output, among other things. In other words, we assume that output and prices respond only with a lag to a policy instrument shock. The latter branch of literature argues that house prices, and asset prices in general, react almost instantaneously to news and therefore are important transmission mechasnism of monetary policy shocks. Therefore, we assume that both interest rates and house prices react simultaneously to news. To consider house prices as forward looking variables which respond immediately to monetary policy news is consistent with economic theory, see Iacoviello (2005). A similar assumption is made in Bjornland and Jacobsen (2010). Instead, using a recursive structure, Goodhart and Hofmann (2001) and Giuliodori (2005) impose that house prices do not respond immediately to monetary policy shocks. Assuming a decrease (increase) in loans after a contractionary (expansionary) monetary policy shock is in line with Gerali et al. (2010) and Gertler and Karadi (2011). We also assume that central banks do not react contemporaneously to a loans supply shock. In fact, as central banks target inflation, they intervene only if inflationary pressure from supply shocks realizes.[14]

[13] See Gambetti and Musso (2016) and Eickmeier and Ng (2015) for a comprehensive list of possible events which may trigger a shock in the supply of credit.

[14] As discussed in Hristov et al. (2012) and Gambetti and Musso (2016), it is not always clear that an increase in the supply of credit loans causes a contemporaneous rise in inflation.

Model Identification Problem. Although sign restrictions provide sufficient information to identify the structural parameters, they do not lead to a unique set of impulse responses.[15] As noted by Fry and Pagan (2011), there is a variety of models which are consistent with the imposed sign restrictions and which provide the same fit to the data. In other words, sign restrictions do not overcome the "model identification problem". We adopt Fry and Pagan (2007)'s "Median Target" (MT) strategy which consists in finding a single model whose impulse responses are closest to the median responses across all the qualifying models. By devising a criterion to do this, we solve the "multiple models" problem since it ensures that the impulses come from the same model and that the corresponding shocks are orthogonal. As noted in Fry and Pagan (2011), the MT criterion selects the median responses when these are uncorrelated. Finally, we note that although the choice of the MT criterion, rather than other magnitude of impulses, may be arbitrary, it is a popular choice as it captures the central tendency of all the plausible models found (Eickemeier and Ng, 2015). Another advantage of using the MT strategy is that when employing this criterion, the sum of the contributions of each error to the forecast error variance of each endogenous variable is equal to one.

4.5.2 The Impact of Housing Demand Shocks

In this section we report the impulse responses of selected variables to a housing demand shock - in terms of a 1% increase in real house prices - for each country under investigation. Comparison across countries is facilitated by the fact that all the structural shocks have unit variances. In addition, we apply the Mean Group (MG) estimation procedure proposed by Pesaran and Smith (1995) to obtain cross-sectional average responses. In particular, let $\zeta_{kl}^{(i)}$ be a $h \times 1$ vector containing the MT responses of variable l to an impulse in variable k over h

[15] In our setting, at each Gibbs sampling iteration, we generate at most 100 structural matrices satisfying the imposed zero and sign restrictions.

periods, for country i. The MG responses of variable l to an impulse in variable k (over h periods), can be computed as the cross-sectional average

$$MG_{kl} = \frac{1}{N}\sum_{i=1}^{N} \zeta_{kl}^{(i)}. \tag{4.4}$$

Similarly, the credible intervals for the MG responses can be computed by taking the cross-sectional average of the impulse responses associated with the percentile of interest.

For the vast majority of countries, a positive housing demand shock has a significant positive impact on inflation, economic activity and real loans to households. Figure 4.1 shows the effects of housing demand shocks on house prices, real GDP, and inflation. The magnitude of the maximum impact on inflation varies between 0.02% and 0.2%, averaging around 0.05% across the countries examined. The impact on real GDP varies between 0.04% and 0.12%, averaging 0.09% and is significant for France, Ireland, Spain, and partially Italy. The maximum impact of the shock is achieved after three quarters on average for activity and inflation. Three country specific observations can be made. First, the impact of a housing demand shock on real GDP is highest for Ireland and Spain, countries having experienced a boom-bust pattern in house prices.[16] Second, the impact of the shock on inflation are larger and significant in the case of Germany, France, and Spain. Muellbauer (2015) argues that in Germany, having a system of comparatively flexible rent controls, an increase in house prices may be followed by a rise in rents affecting in turn inflation developments.

[16]When using 95% confidence bands, the impact is significant in Ireland and Spain, as well as in Italy and France. For Belgium, the Netherlands and Germany significance holds only when considering 80% credible intervals.

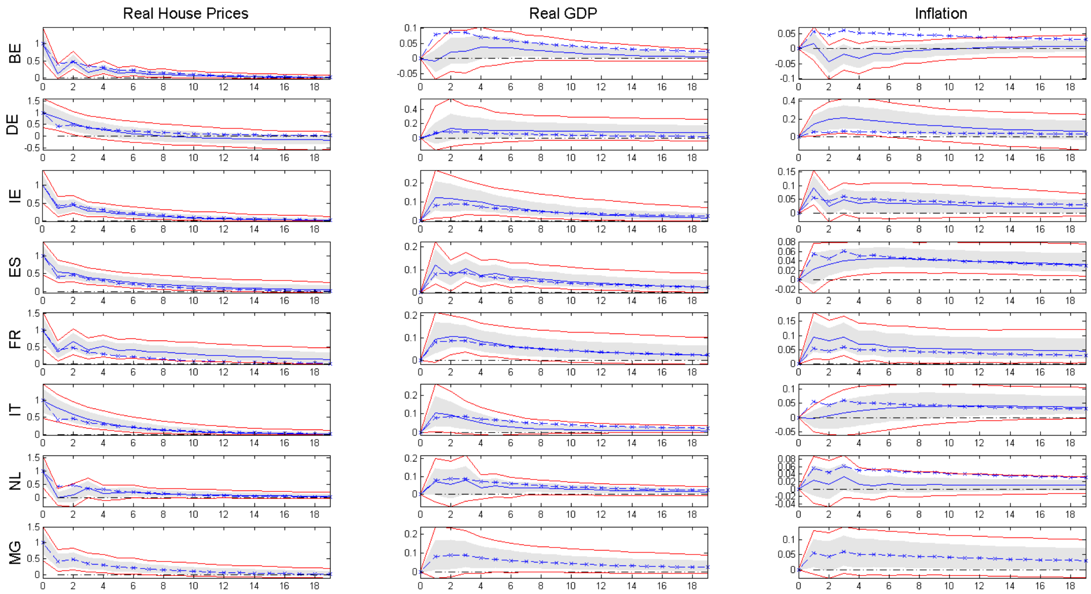


Figure 4.1: Impulse response functions to a housing demand shock, for real house prices growth, real GDP growth, and inflation, across countries. The red lines delimit the 95 per cent credible interval. The grey shaded area delimits the space between 10th and 90th percentiles. The blue line is the median impulse response, while the crossed blue line is the median Mean Group (MG) response, both obtained using the MT approach.

House prices and the market for loans. The impact on real loans to households, shown in Figure 4.2, varies between 0.10% and 0.5%, averaging 0.35%. The positive effect of house price shocks on loans is highest on impact for all countries and reflects the tight links between the credit and the housing markets. The impulse responses of real loans exhibits a fairly similar pattern across countries in terms of expected sign, size and statistical significance.

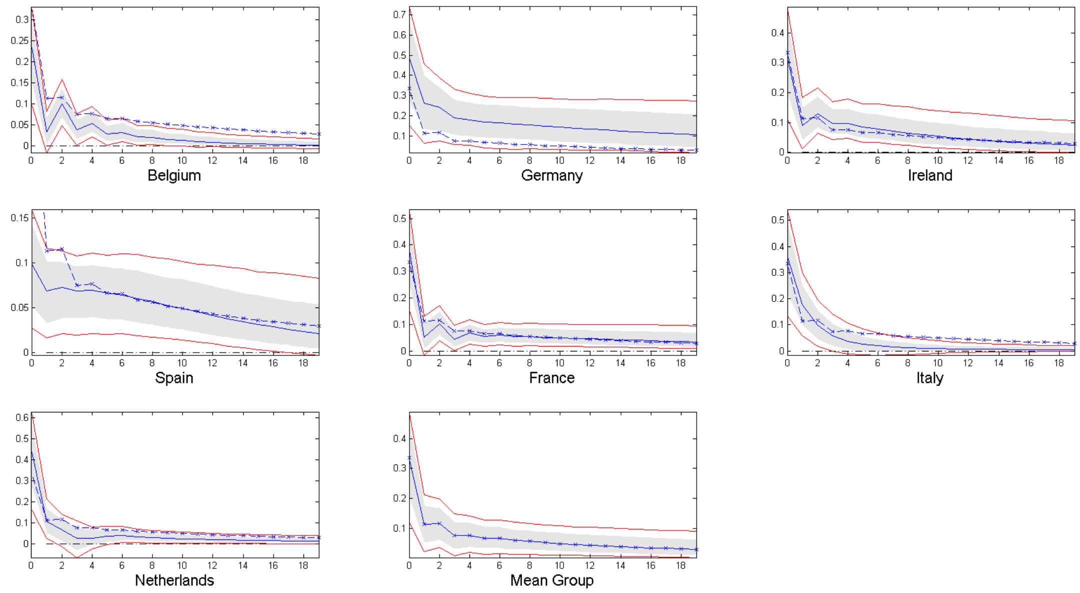


Figure 4.2: Impulse response of real loans to a housing demand shock. Cross-country comparisons. The red lines delimit the 95 per cent credible interval. The grey shaded area delimits the space between 10th and 90th percentiles. The blue line is the median impulse response while the crossed blue line is the median Mean Group response, both obtained using the MT approach.

Wealth Effects. Results shown in Figure 4.3 indicate that the magnitude of the maximum impact on real private consumption varies between 0.02% and 0.16% across the countries examined, averaging 0.1%, with the maximum impact occurring after three quarters on average. They confirm the intuition that liquidity constrained households can expand their consumption capabilities using housing wealth as collateral to obtain higher borrowing (Iacoviello, 2004). Moreover, the ordering of countries is confirmed. Ireland and Spain exhibit the largest wealth effects on consumption (around 0.15%) followed by Italy. Results for the first two countries are highly significant, while those for the others exhibit a lower magnitude of the impact and lower statistical significance. Muellbauer (2015) finds a negative effect of real house prices on consumption in both France and Germany. Instead, our analysis reveals that a housing demand shock has a positive impact on consumption in both countries, even though, as noted above, this effect is

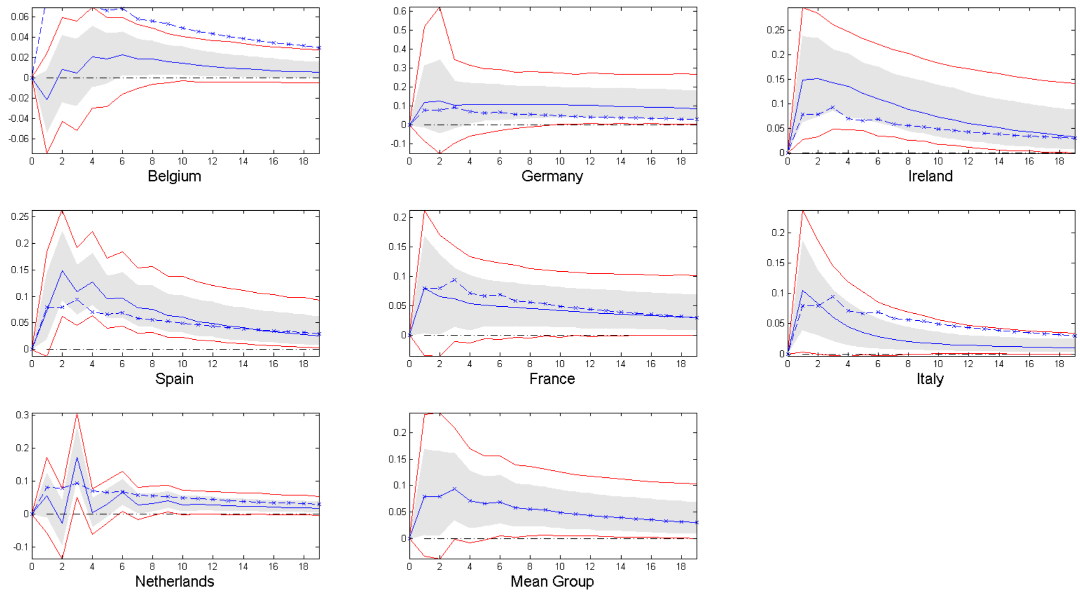


Figure 4.3: Impulse response of real private consumption to a housing demand shock. Cross-country comparisons. The red lines delimit the 95 per cent credible interval. The grey shaded area delimits the space between 10th and 90th percentiles. The blue line is the median impulse response while the crossed blue line is the median Mean Group response, both obtained using the MT approach.

rather muted.

Variance Decomposition

Forecast error variance (FEV) decomposition is crucial to understand how important a housing shock is for consumption, credit supply and other variables of interest. In fact, this 'innovation accounting' analysis allows us to answer to what extent the variability in the aforementioned variables can be explained by a housing demand shock. Figure 4.4 shows the proportion of 1, 3, 5, and 20 quarters ahead forecast error variance of each endogenous variable of the VAR accounted for by innovations in real house prices.

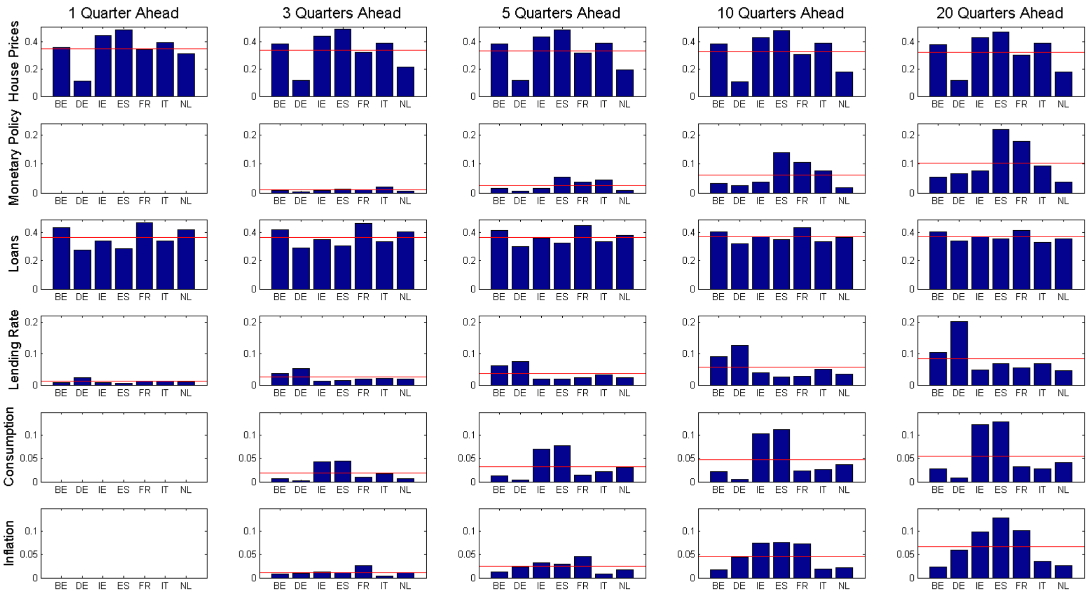


Figure 4.4: Variance Decomposition. Proportion of 1, 3, 5, 10, and 20 quarters ahead FEV of each variable accounted for by innovations in house prices. Cross-country comparisons. Red line indicates the average contribution across countries.

The contribution of the housing demand shock to fluctuations of aggregate consumption, inflation, loans and lending rates is highly heterogeneous across countries. About 40% of the FEV of house prices is accounted for by own innovations in Ireland and Spain, followed by Belgium and Italy. The identified structural shock seems to play a minor role in explaining the variation of consumption across countries, with the exception of Spain and Ireland, where house prices innovations contribute slightly less than 15% to the forecast error variance of consumption. For all the other countries, the contribution is less than 5% and almost zero in Germany. In France, Ireland, and Spain, only 10% of the FEV of inflation can be explained by a housing shock. On the other hand, housing demand shocks play an important role in explaining loans FEV across all countries under investigation (the average contribution is approximately 40%), confirming the 'financial accelerator' hypothesis, according to which changes in the collateral affect borrowing capacity (Bernanke et al. (1996) and Almeida et al. (2006)).

4.5.3 Monetary Policy Shocks and the Role of House Prices

Regarding the role of house prices in the transmission of monetary policy, we find that monetary policy shocks have a significant, strong and lasting impact on house prices, corroborating the existence of a credit channel in the euro area housing market.

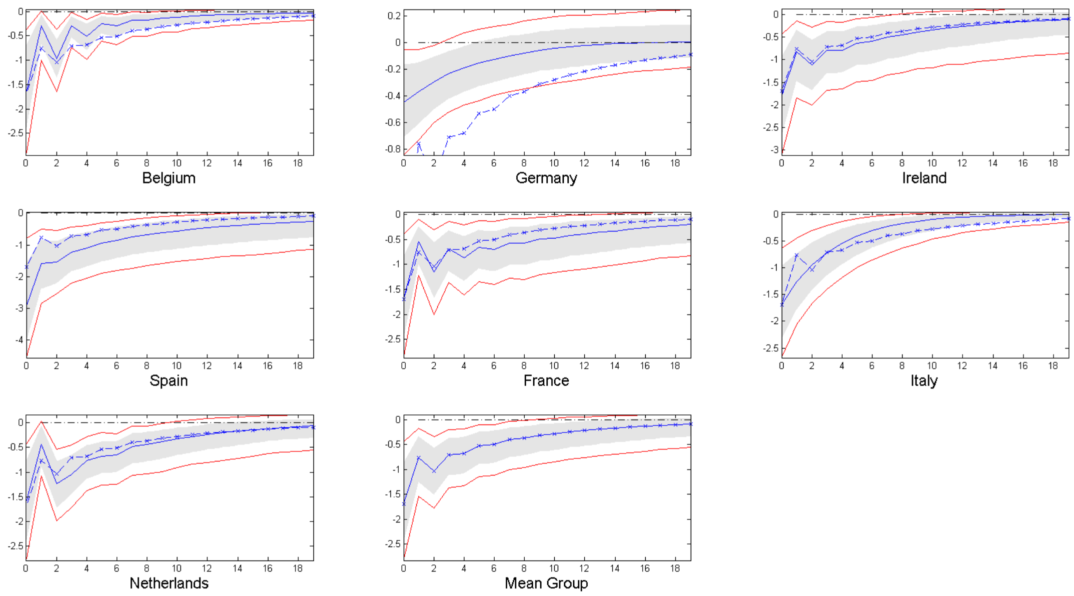


Figure 4.5: Impulse response of house prices to a monetary policy shock. Cross-country comparisons. The red lines delimit the 95 per cent credible interval. The grey shaded area delimits the space between 10th and 90th percentiles. The blue line is the median impulse response while the crossed blue line is the median Mean Group response, both obtained using the MT approach.

Figure 4.5 depicts the impulse responses of real house price growth to a monetary policy shock in terms of 25 basis points increase in the policy rate. The responses are highly heterogeneous across the countries examined. A 25 basis points contractionary monetary policy shock significantly reduces real house prices between 0.4% (in Germany) and 3% (in Spain), on average by 1.6%, with the maximum impact generally occurring contemporaneously.[17] Being aware of such heterogeneity is crucial to address possible imbalances across countries and

[17]Comparisons across countries is further facilitated by the fact that all the structural shocks have unit variances.

to design policies to mitigate risks deriving from residential property markets. The result for Spain seems to suggest that lower interest rates may have played a role in stimulating the demand for housing by easing financing conditions. In Ireland, the response of house prices to a monetary shock are in line with the average responses.

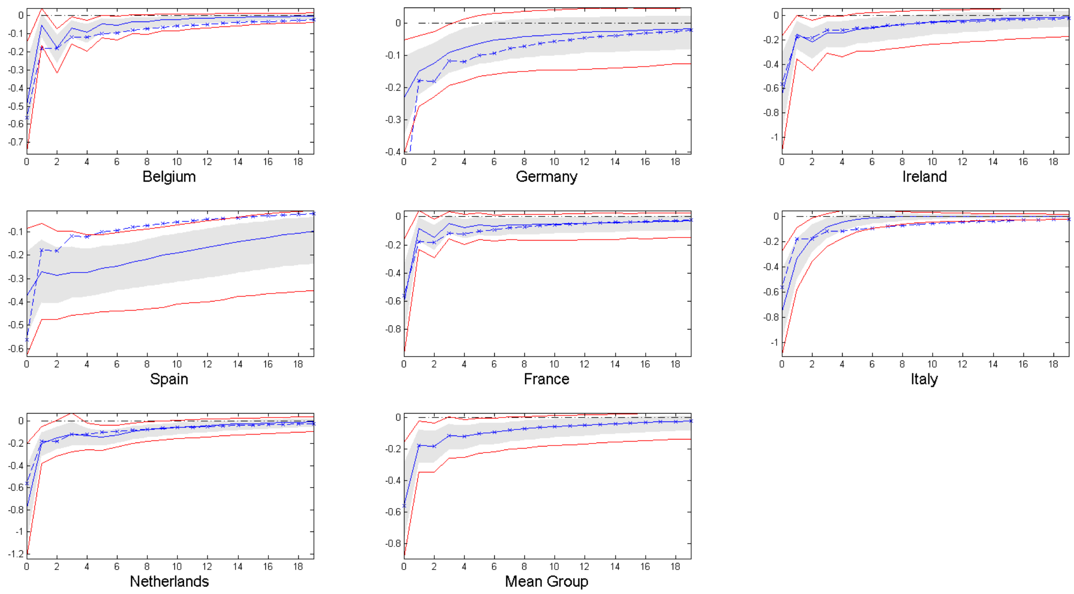


Figure 4.6: Impulse response of real loans to a monetary policy shock, in terms of a 25 basis points increase in the monetary policy rate. Cross-country comparisons. The grey area delimits the space between 10th and 90th percentiles. The red lines delimits the 95 per cent credible interval. The blue line is the median impulse response while the crossed blue line is the median Mean Group response, both obtained using the MT approach.

As shown in Figure 4.6, a contractionary monetary policy shock also causes a significant decline of real loans to households, between 0.2% (Germany) and 0.8% (the Netherlands), on average by 0.6%, with the maximum impact occurring contemporaneously. The heterogeneity across countries of the impact of the shock on real loans does not appear to be related to the tenure of the mortgage rate structure. The magnitude of the impact is also quite dispersed within countries characterised by prevalence of variable rates – such as Italy, Spain and Ireland – or fixed rates – such as France, Belgium, Germany and the Netherlands (ESRB

(2015)).

Variance Decomposition

To quantify the importance of monetary policy shocks we compute the forecast error variance decomposition. A monetary policy shock accounts on average for around 25% of the FEV of real house price growth (Figure 4.7). The contribution is above the average for Spain, followed by Italy and the Netherlands and is the lowest in Germany.

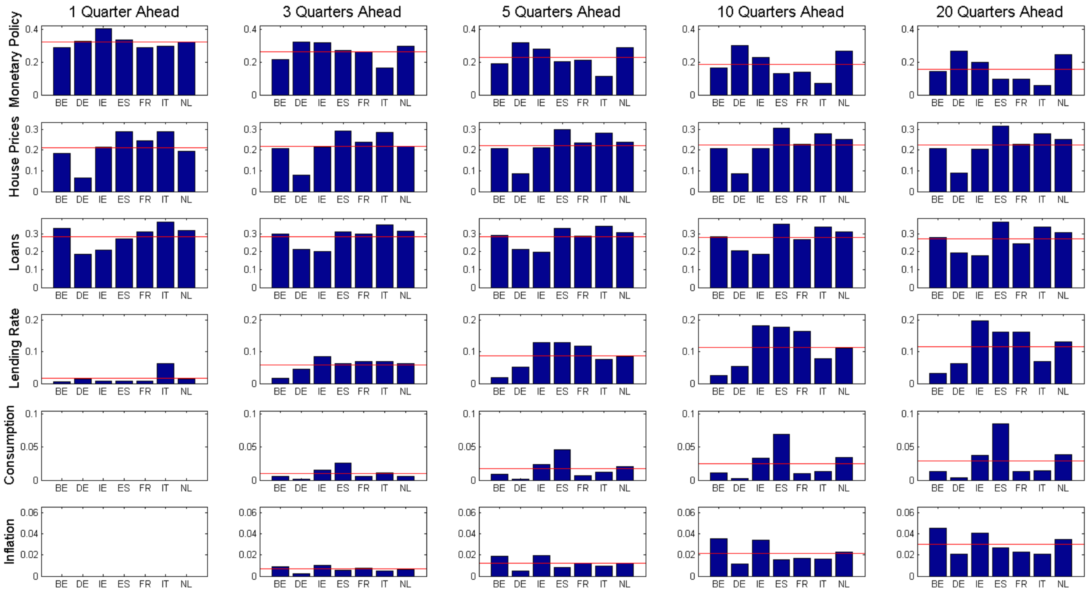


Figure 4.7: Variance Decomposition. Proportion of 1, 3, 5, 10, and 20 quarters ahead FEV of each variable accounted for by innovations in monetary rate. Cross-country comparisons. Red line indicates the average contribution across countries.

In the latter country, only less than 10% of the FEV of house price growth is accounted for by a monetary policy innovation. These results corroborate the evidence that the housing market plays an important role in the monetary policy transmission mechanism, and provide new evidence on the heterogeneous impact of monetary policy on house prices fluctuations. Monetary policy shocks also contribute on average around 30% of the variability in real loan growth, which is less than the average contribution of housing demand shocks.

4.5.4 Historical Decompositions

So far we have studied how the structural shocks of interest affect average movements in the data, by means of forecast error variance decompositions and impulse response functions. In this section, we use historical decompositions in order to assess the cumulative effects of housing demand and monetary policy shocks on the business cycle and their relative importance in explaining the observed fluctuations in the endogenous variables of the VAR, at each point in time.

To compute the historical decompositions, we rewrite the VAR model described in equation (4.1), in its moving average representation:

$$y_t - \mu = \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \sum_{s=0}^{\infty} \Theta_s w_{t-s} \tag{4.5}$$

where $\Theta_s = \Phi_s \tilde{A}_o$, and $w_{t-s} = \tilde{A}_o^{-1} u_{t-s}$ are the orthogonal shocks. $\tilde{A}_o$ is the contemporaneous structural matrix satisfying the imposed zero and sign restrictions. As we cannot estimate all the 'infinite' shocks, w_{t-s}, for $s = 0, .., \infty$, we truncate the series. We denote such an approximation as

$$\hat{y}_t \approx \sum_{s=0}^{t-1} \hat{\Theta}_s \hat{w}_{t-s}$$

where $\hat{y}_t = (\hat{y}_t - \hat{\mu})$. The unknown values in the right-hand side are replaced by their estimates. Each endogenous variable, $\hat{y}_{kt}$, for $k = 1, .., m$ can be written as

$$\hat{y}_{kt} = \sum_{j=1}^{m} \hat{y}_{kt}^{(j)} = \sum_{j=1}^{m} \sum_{s=0}^{t-1} \hat{\theta}_s^{(k,j)} \hat{w}_{j,t-s}$$

where $\hat{y}_{kt}^{(j)}$ is the cumulative effect of shock j to the kth variable of the VAR process and $\hat{\theta}_s^{(k,j)}$ is the (k, j) element of $\hat{\Theta}_s$.

As suggested in Kilian and Lütkepohl (2017), we demean both $\hat{y}_t$ and y_t to remove any discrepancy among them. We discard the initial (transients) observations (in particular, we remove the first 20 observations) so that the two observations coincide with minimal approximation errors (which could arise from the truncation of the infinite sum).

Following Kilian and Lütkepohl (2017), we construct counterfactuals as an alternative way of assessing the cumulative effect of a structural shock to the observed data y_{kt}, for $k = 1, .., m$. They are defined as

$$c_{kt}^{(j)} = y_{kt} - \hat{y}_{kt}^{(j)} \tag{4.6}$$

where the counterfactual, $c_{kt}^{(j)}$, represents the evolution of y_{kt} in the absence of the jth structural shock.

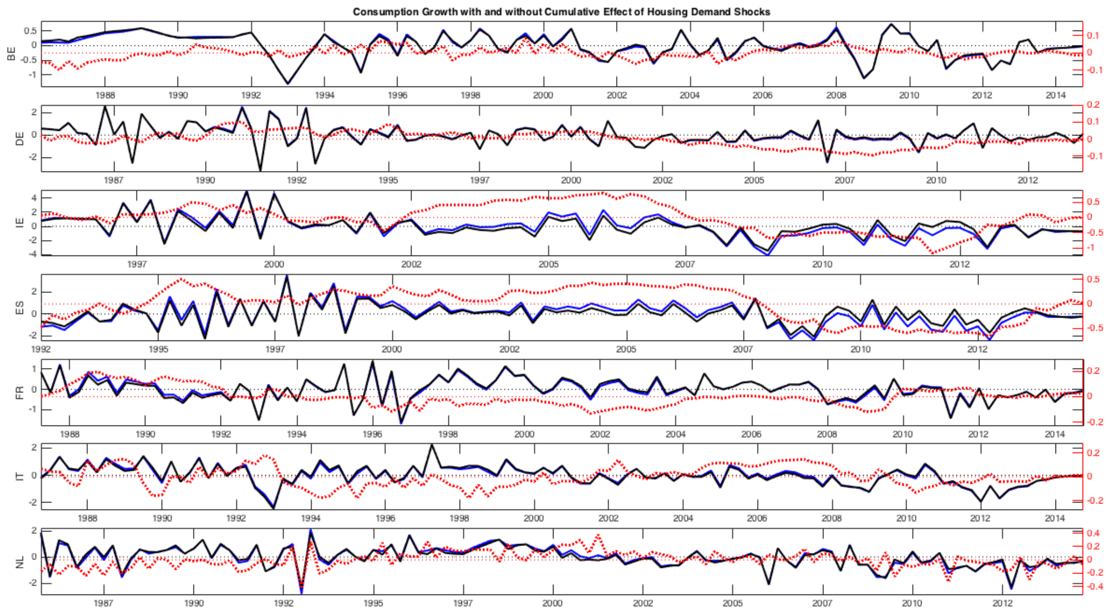


Figure 4.8: Historical counterfactuals for real (private) consumption growth. The counterfactuals (black line) indicate the evolution of real consumption growth in the absence of housing demand shocks. The difference between actual data (blue line) and counterfactuals corresponds to the cumulative effects of housing shocks to real consumption growth over time (dotted red line). If the red line lies above zero, it means that the shocks positively contributed to the growth rates of real consumption. The left y-axis measures quarterly changes (from the sample mean) in real consumption growth with and without the cumulative effects of housing shocks. The right y-axis reports quarterly changes (from their sample mean) in the cumulative effects of housing demand shocks.

Housing Demand Shocks and Business Cycles. Figure 4.8 illustrates the evolution of (quarterly) real private consumption growth (y_{kt}) in deviations from the sample mean against its counterfactual ($c_{kt}^{(j)}$). The latter indicates how real consumption growth would have evolved if all the realizations of housing demand shocks had been equal to zero, while maintaining the remaining structural shocks in the model. The difference between the two ($\hat{y}_{kt}^{(j)}$) represents the cumulative effects of housing demand shocks on consumption up to a certain point in time. It measures how growth rates of real consumption would have evolved if the economy had been hit only by housing demand shocks, in the absence of all other concurrent shocks. A line above zero reveals that the structural shock exerted upward pressure on consumption.

The historical decompositions analysis suggests an important role of housing demand shocks in driving consumption in Ireland and Spain, confirming previous findings for the impulse response analysis. It is noteworthy that in both countries, in the absence of housing demand shocks, the growth rates of real private consumption would have been lower than the actual rates between 2002 and 2007, and larger between 2008 and 2013. In particular, in Ireland, the cumulative effect of housing demand shocks on consumption is equal to 0.79% up to 2006 and equal to -1.16% at the end of 2011. Similarly, in Spain, in the absence of other shocks, if the growth rates of real consumption (in deviations from the sample mean) had been driven exclusively by housing demand shocks, they would have been largest around 1995 and 2004 (0.5%), and lowest in 2012 (-0.66%).[18] The cumulative effect of housing demand shocks is rather muted in the remaining countries.

[18] Excluding monetary policy from the VAR, these differences would have been even higher.

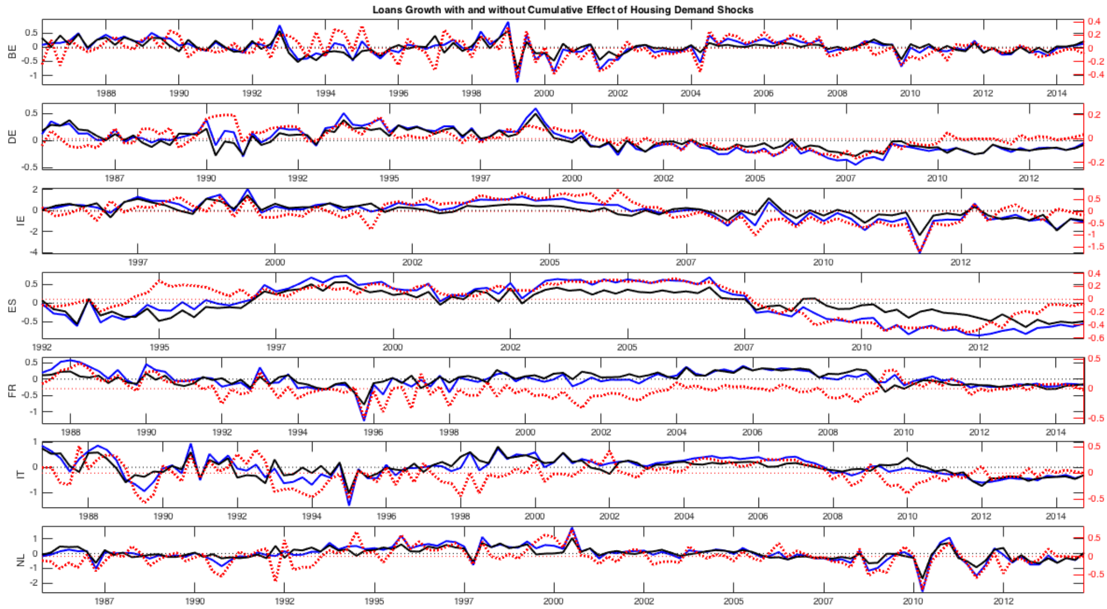


Figure 4.9: Historical counterfactuals for real loans growth. The counterfactuals (black line) indicate the evolution of real loans growth in the absence of housing demand shocks. The difference between actual data (blue line) and counterfactuals corresponds to the cumulative effects of housing shocks to real loans growth over time (dotted red line). If the red line lies above zero, it means that the shocks positively contributed to the growth rates of real loans. The left y-axis measures quarterly changes (from the sample mean) in real loans growth with and without the cumulative effects of housing demand shocks. The right y-axis reports quarterly changes (from the sample mean) in the cumulative effects of housing shocks.

As shown in Figure 4.9, the historical decomposition analysis corroborates the "financial accelerator" hypothesis according to which increases in the collateral improve households borrowing capacity. The cumulative effects of housing demand shocks to real loans growth are sizeable across most of the countries under investigation. As for consumption, the contribution of housing shocks to real loans growth was larger in Spain and especially in Ireland. In both countries, exogenous housing price increases significantly contributed to the raise of real loans growth during the specific housing boom episode observed in the sample. The subsequent bursting of the housing bubble and the consequent decline in house prices drastically reduced the availability of loans, which in turn may also have had negative consequences on real consumption growth.

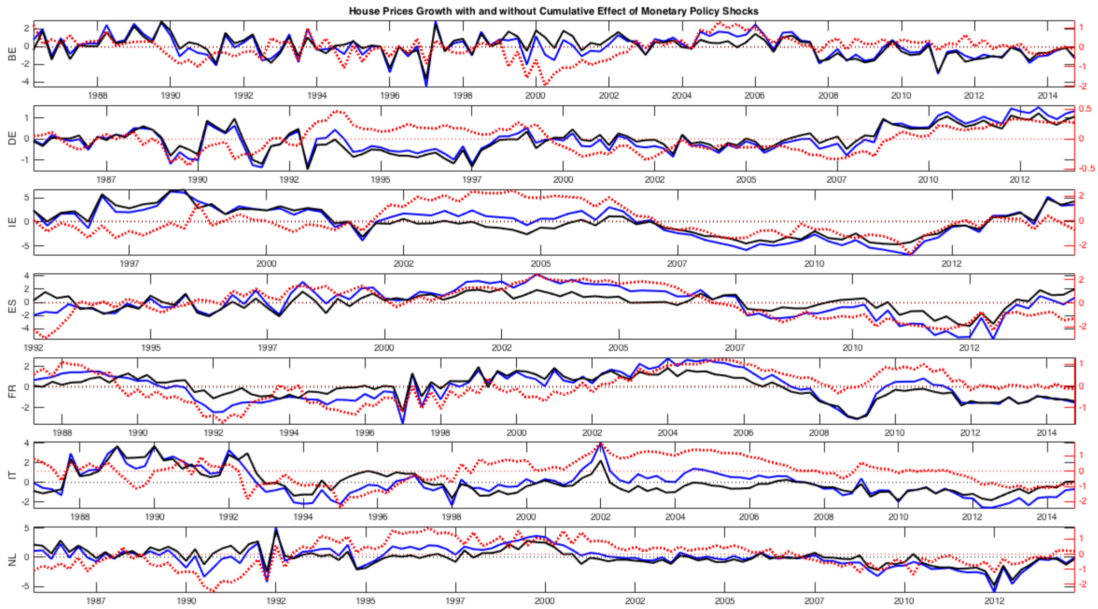


Figure 4.10: Historical counterfactuals for real house prices growth. The counterfactuals (black line) indicate the evolution of real house prices growth in the absence of monetary policy shocks. The difference between actual data (blue line) and counterfactuals corresponds to the cumulative effects of monetary policy shocks to real house prices growth over time (dotted red line). If the red line lies above zero, it means that the shocks positively contributed to the growth rates of real house prices. The left y-axis measures quarterly changes (from the sample mean) in real house prices growth with and without the cumulative effects of monetary shocks. The right y-axis reports quarterly changes (from the sample mean) in the cumulative effects of monetary policy shocks.

Monetary Policy Shocks and House Price Dynamics. Figure 4.10 plots the historical evolution of (quarterly) real house prices growth (in terms of deviations from the sample mean), the dynamics of the counterfactual, as well as the difference between the two (i.e. the hypothetical movements of real house prices growth if all structural shocks but monetary policy shocks had been turned off). The cumulative contribution of monetary policy shocks to house price growth differ widely across countries: from a peak of 0.4% in Germany to a maximum cumulative effect of above 2% in Ireland, followed by the Netherlands and Spain. The maximum contribution is also relevant in Italy (slightly above 1%), while in Belgium and France it lies in a middle ground. It is worthy of note that a substantial increase in house prices would have occurred in Ireland and Spain between 2001 and 2006 even in the absence of all other structural shocks but monetary policy. Finally, we note that the cumulative contributions of monetary policy shocks vary not only in size but also they do not appear to be correlated over time across countries.

4.6 Conclusions

In this chapter, we use a structural Bayesian VAR model to provide a systematic structural analysis of the effects of housing demand shocks on the macro-economy across selected euro area countries, and the role of house prices in the monetary transmission mechanism. A novel identification strategy which combines zero and sign restrictions is proposed. Among other things, by doing so, we are able to distinguish between a house price and an aggregate demand shock, which would be difficult otherwise. To overcome the over-parameterization problem, the priors are selected using a Stochastic Search Variables Selection method, which allows for shrinkage of the VAR coefficients while selecting restrictions that are supported by the data itself. This in turn makes adequate finite sample inference and exploits in full the intrinsic cross-country heterogeneity typical of the housing market.

Furthermore, given the lack in the literature of comparative studies that try to quantify the degree of heterogeneity of the impact of house prices and their role in the transmission of monetary policy across euro zone members, we exploit the cross-sectional dimension of our data to quantify and compare the different dynamics of house prices, their heterogeneous effect on the macro-economy and the diverse impact of monetary policy in driving house price cycles across euro-zone member states. Quantifying such diverse effects is important from a policy perspective, in particular when addressing real and financial imbalances at the country level.

The structural analysis confirms that the effects of housing demand and monetary policy shocks differ widely across the countries under investigation. We document the existence of a housing 'wealth effect' in Ireland and Spain, where a one percent increase in real house prices is associated with a 0.15% rise in real private consumption. The fact that both countries experienced a housing bubble corroborates the view that house price booms play an important role in boosting confidence, which in turn stimulate consumption. The historical decomposition analysis supports these findings. The cumulative effects of housing demand shocks on consumption are larger in Ireland and Spain. In both countries, housing demand shocks significantly contributed to the surge of consumption growth during the specific housing boom episode observed in the sample and to the subsequent decline during the recession started around 2007.

The impact of housing demand shocks on real loans to household exhibits a less heterogeneous and a fairly similar pattern across countries in terms of sign, size and statistical significance. On average, a housing demand shock, in terms of a 1% increase in house prices, causes a 0.35% increase in real loans. Housing demand shocks play an important role in explaining loans forecast error variance across all countries under investigation. This clearly suggests that changes in the value of collateral affect borrowing capacity. The historical decomposition analysis provides further evidence in support of the 'financial accelerator' theory. We then show that monetary policy has a strong and lasting impact on house prices, corroborating the existence of a credit channel in the euro area housing market and an important role of house prices in the monetary transmission mechanism. The impact is highly heterogeneous, varying between 0.4% (in Germany) and 3% (in Spain). A monetary policy shock accounts on average for around 25 to 30% of the forecast error variance of real house price growth. The historical decomposition analysis documents a highly heterogenous contribution of monetary policy shocks to house price dynamics.

4.7 Technical Appendix

4.7.1 The VAR with SSVS Prior

Prior on the VAR coefficients. Let $\alpha = vec(A)$ be the $Km \times 1$ vector of regression coefficients. The SSVS assumes that the prior distribution of α_j (the jth element of α) is a mixture of two Normal distributions:

$$\alpha_j \mid \gamma_j \sim (1-\gamma_j) \cdot N\left(0, \tau_{0j}^2\right) + \gamma_j \cdot N\left(0, \tau_{1j}^2\right)$$

To select τ_{0i}^2 and τ_{1i}^2 , we follow George, Sun and Ni (2008). They propose a "default semi-automatic approach" which involves setting $\tau_{0j} = c_0\hat{\sigma}_{\alpha_j}$ and $\tau_{1j} = c_1\hat{\sigma}_{\alpha_j}$, where $c_0 \ll c_1$, i.e. $c_0 = 0.1$ and $c_1 = 10$; $\hat{\sigma}_{\alpha_j}$ is the standard error associated with the unconstrained least squares estimate of α_j. The dummy γ_j is unknown and it has to be estimated in a data-based fashion. In particular, it is assumed that the γ_j's are independent Bernoulli random variables so that

$$P\left(\gamma_j = 1\right) = p_j, \quad P\left(\gamma_j = 0\right) = 1 - p_j, \quad j = 1, .., Km.$$

As noted by George, Sun and Ni (2008), for each j, p_j reflects the prior belief that α_j should be unrestricted. In the absence of such prior information, one could set $p_j \equiv .5$ and let the data decide whether to shrink or not the coefficient to zero, as we do in this chapter.

Prior on the covariance parameters. The covariance matrix can be decomposed as $\Sigma_u^{-1} = \Psi\Psi'$, where Ψ is upper triangular. Let ψ_{ij} be the (i, j)th entry of Ψ. Each off-diagonal element has the prior distribution

$$\psi_{ij} \mid \omega_{ij} \sim (1-\omega_{ij}) \cdot N\left(0, \kappa_{0ij}^2\right) + \omega_{ij} \cdot N\left(0, \kappa_{1ij}^2\right)$$

We arbitrarily set $\kappa_{0ij} = 0.1$, and $\kappa_{1ij} = 6$. Alternatively, κ_{0ij} and κ_{1ij} can be chosen using similar consideration for setting τ_{0j} and τ_{1j}.

We assume that the ω_{ij} are independent Bernoulli random variables such that

$$P(\omega_{ij} = 1) = q_{ij}, \qquad P(\omega_{ij} = 0) = 1 - q_{ij}$$

for $i = 1, ..., m$ and $j = 2, ..., m - 1$ and where $q_{ij} \in (0, 1)$.

Given the absence of prior information on whether ω_{ij} should be unrestricted, we follow George, Sun and Ni (2008) suggestion, by setting $q_{ij} = 0.5$.

For the diagonal elements, it is assumed prior independence with $\omega_{ii}^2 \sim gamma\,(a_i, b_i)$. The hyperparameters (a_i, b_i) are set equal to $(0.01, 0.01)$ to render this prior non-influential.

Posterior Distribution. Posterior computation is carried out using the Gibbs sampling algorithm described in George, Sun and Ni (2008).

Following the latter, we simulate a Markov chain of 20.000 cycles and discard the initial 10.000 burn-in replications. In their simulated numerical examples, the authors note that simulation results using a larger number of cycles (50.000) change little, suggesting that the Markov chains converge rather quickly.

Estimation of the reduced form VAR requires approximately one minute (for each country).

4.8 Data

4.8.1 Data Sources

All data cover the period 1980Q1-2014Q4, unless otherwise specified. All variables were transformed in annualised quarter on quarter changes except for interest rates which are in levels. When seasonally adjusted data are not directly available, we make the necessary adjustments, using the X-12 Census method. Nominal house prices and nominal loans were deflated using CPI indices.

GDP, Private Consumption and Consumer Price Indices. Sources: OECD – Main Economic indicators. Data on private consumption for Ireland starts on 1990Q1. Data on private consumption for Germany are obtained from the European Central Bank's Multi Country Model Dataset.

House Prices. For Belgium and Italy, we use 'Residential property prices, New and existing dwellings; Residential property in good & poor condition; Whole country'. For France, Ireland, the Netherlands, and Spain, we use 'Residential property prices, New dwellings; Residential property in good & poor condition; Whole country'. For Germany, an annual series covering new dwellings (apartments and houses) in 50 West German cities were used given its long time span (starting in 1975) and due to the absence of structural breaks related to the German reunification compared to other available series (the annual series was linearly interpolated at a quarterly frequency). House price data for Spain starts in 1987Q1.

Sources: ECB and national sources, and BIS (Germany).

Loans. We use data on "Credit to Households and NPISHs from All sectors".

Source: BIS http://www.bis.org/statistics/totcredit.htm

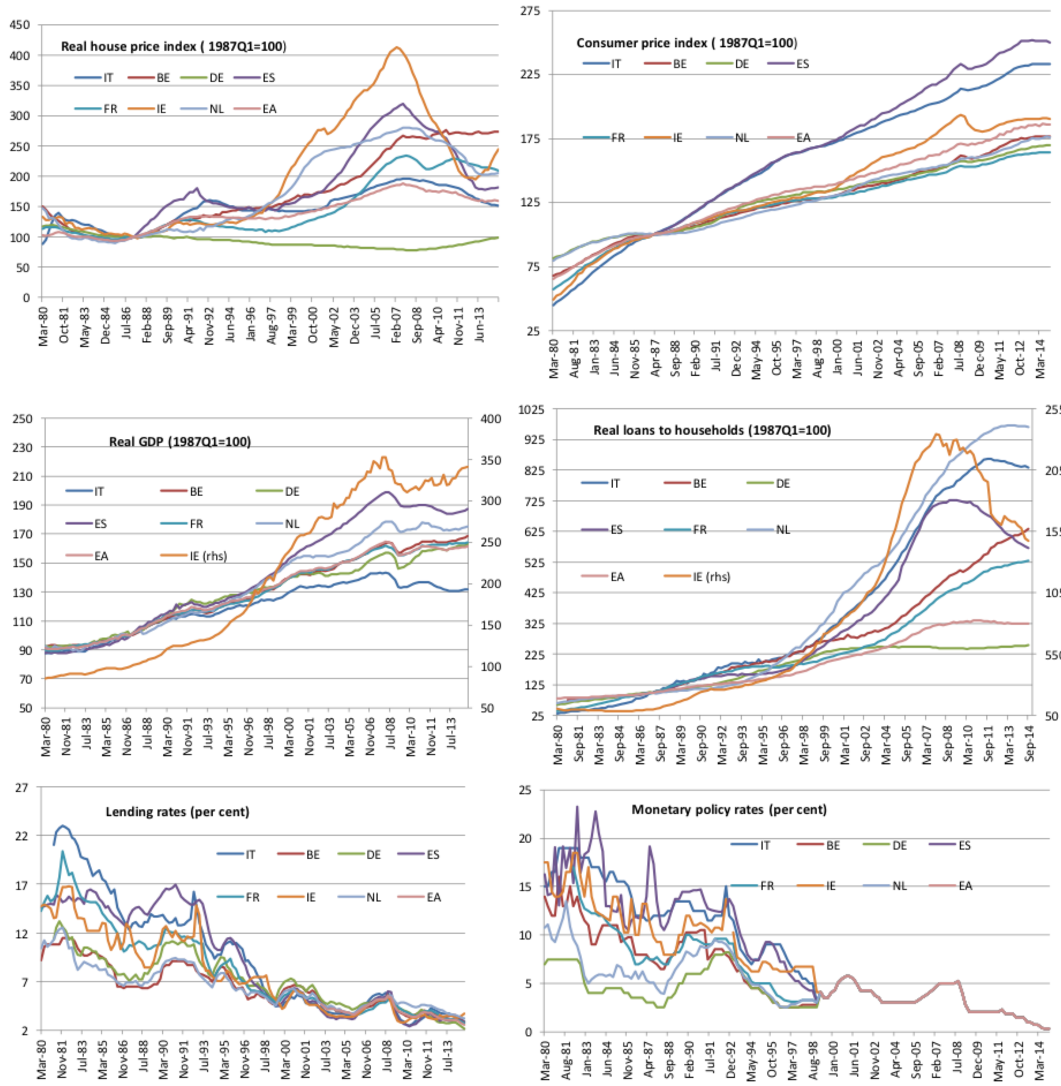
Lending Rates. The dataset consists of "Lending to households for house purchase excluding revolving loans and overdrafts, convenience and extended credit card debt".

Sources: ECB - MFI Interest Rate Statistics.

Monetary Policy Rates. For the pre-EMU period we use national official discount rates from BISM Dataset: "BIS Macro-economic series". From 1999, we use the ECB Marginal Lending Facility.

4.8.2 Supplementary Charts

Figure 4.11: Dynamics across the Euro Area



Chapter 5

Conclusion

This thesis combines some theoretical contributions to the literature on large heterogeneous panel data models with some applications in international macroeconomics.

The first part considers the problem of statistical inference in random coefficient panel data models when both the number of units and the number of time periods are quite large.

In Chapter 2, we examine, by Monte Carlo methods, the causes and effects of negative definite covariance matrices in Swamy (1970) type random coefficient models. First, we show that the degree of coefficient heterogeneity relative to the (conditional) variability of the observed data plays a crucial role. The sample size and the finite sample performances (in terms of bias and root mean square errors) of the individual ordinary least squares and the Mean Group estimators may also have an impact on the probability of observing a negative definite estimator of the random coefficient covariance matrix. We then investigate the finite sample consequences for hypothesis tests. Whenever the unbiased estimator of the random coefficient covariance is negative definite, Swamy suggests eliminating a term to obtain an estimator which is nonnegative definite and is consistent when T tends to infinity. However, we show that the latter can be severely biased in finite samples. The resulting estimated standard errors are very often upwards biased, and in many cases, this bias can be substantial. This in turn leads to size distorted hypothesis tests, with exact sizes well below the nominal levels.

A solution is proposed in the third Chapter. We show that the maximum likelihood estimator of the random coefficient covariance matrix, obtained by applying the EM algorithm, satisfies the law of total variance without running into the problem of negative definiteness. This in turn leads to more accurate estimated standard errors and hypothesis tests. To extend the applicability of our method, we consider a general framework which incorporates various panel data models as special case. The use of the EM algorithm has other benefits. It allows us to estimate both the average effects and the unit-specific components. It also gives a probability distribution over the random effects. Monte Carlo simulations reveal that the (restricted) maximum likelihood estimators, obtained by applying the EM algorithm, have relatively good finite sample properties, in terms of bias and root mean square errors. In evaluating the merits of our approach, we also provide an overview of the sampling and Bayesian methods commonly used to estimate heterogeneous panel data. Our method represents a valid alternative to Bayesian estimation in those cases in which the researcher wishes to make inference on the random effects distribution while having little knowledge on what a sensible prior might be. At the same time, it helps overcome one drawback of the Bayesian inference: 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. To demonstrate the usefulness of the EM approach in empirical research, we apply our method to the analysis of the determinants of sovereign credit risk. In particular, we investigate what causes the sensitivity of sovereign spreads to debt to differ significantly across countries by modelling the latter as a function of macroeconomic fundamentals and a set of explanatory variables which reflect the history of government debt and economic crises of various forms. We argue that history of repayment in financial market is an important explanatory factor of the cross-country differences in the magnitude of sovereign spreads' reaction to changes in government debt.

In Chapter 4, we use a structural Bayesian vector autoregressive model to provide a systematic structural analysis of the effects of housing demand shocks on the macro-economy and the role of house prices in the monetary transmission mechanism, across selected euro area countries. A novel identification strategy which combines zero and sign restrictions is proposed. We focus on a country by country analysis, given the idiosyncratic characteristics of the housing market in the euro area, which suggest that pooling or aggregating may lead to biased inference and misleading policy recommendations. At the same time, given the lack in the literature of comparative studies that try to quantify the degree of heterogeneity of the impact of house prices and their role in the transmission of monetary policy across euro area countries, we exploit the cross-sectional dimension of our data to quantify and compare the different dynamics of house prices, their heterogeneous effects on the macro-economy and the diverse impact of monetary policy in driving house price cycles across Eurozone member states. Quantifying such diverse effects is important from a policy perspective, in particular when addressing real and financial imbalances at the country level.

Bibliography

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

[2] H. Almeida, M. Campello, and C. Liu. The financial accelerator: evidence from international housing markets. *Review of Finance*, 10(3):321–352, 2006.

[3] W. S. Alnosaier. *Kenward-Roger approximate F test for fixed effects in mixed linear models*. PhD thesis, 2007.

[4] 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.

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

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

[7] K. Aoki, J. Proudman, and G. Vlieghe. Houses as collateral: Has the link between house prices and consumption in the UK changed? *Federal Reserve Bank of New York Economic Policy Review*, 8(1):163–78, 2002.

[8] J. Arias, J. F. Rubio-Ramirez, and D. F. Waggoner. Inference based on SVARs identified with sign and zero restrictions: Theory and applications. *FRB Atlanta Working Paper*, 2014.

[9] O. Aspachs-Bracons and P. Rabanal. The effects of housing prices and monetary policy in a currency union. *IMF Working Papers*, pages 1–52, January 2011.

[10] O. P. Attanasio, L. Blow, R. Hamilton, and A. Leicester. Booms and busts: Consumption, house prices and expectations. *Economica*, 76(301):20–50, 2009.

[11] N. Bailey, G. Kapetanios, and M. H. Pesaran. Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31:929–960, 2015.

[12] G. Baldi. The economic effects of a central bank reacting to house price inflation. *Journal of Housing Economics*, 26:119–125, 2014.

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

[14] B. Baltagi, G. Bresson, and A. Pirotte. To pool or not to pool. In Matyas and Sevestre, editors, *The Econometrics of Panel data*, volume 3rd edition, pages 517–546. Springer-Verlag, Berlin, 2008.

[15] B. H. Baltagi, G. Bresson, J. M. Griffin, and A. Pirotte. Homogeneous, heterogeneous or shrinkage estimators? Some empirical evidence from french regional gasoline consumption. *Empirical Economics*, 28(4):795–811, 2003.

[16] B. H. Baltagi and J. M. Griffin. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, 77(2):303–327, 1997.

[17] B. H. Baltagi, J. M. Griffin, and W. Xiong. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics*, 82(1):117–126, 2000.

[18] B. H. Baltagi and D. Levin. Cigarette taxation: raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, 3(2):321–335, 1992.

[19] C. Basten and C. Koch. The causal effect of house prices on mortgage demand and mortgage supply: Evidence from Switzerland. *Journal of Housing Economics*, 30:1–22, 2015.

[20] B. Bernanke, M. Gertler, and S. Gilchrist. The financial accelerator and the flight to quality. *Review of Economics and Statistics*, 78:1–15, 1996.

[21] B. S. Bernanke and K. N. Kuttner. What explains the stock market's reaction to Federal Reserve policy? *The Journal of Finance*, 60(3):1221–1257, 2005.

[22] L. Bini Smaghi. The European Central Bank in a global perspective - central banking and the challenge of rising inflation. *Speech at the 16th Annual Conference of the German-British Forum*, May 2011.

[23] H. C. Bjørnland and D. H. Jacobsen. The role of house prices in the monetary policy transmission mechanism in small open economies. *Journal of Financial Stability*, 6(4):218–229, 2010.

[24] D. Boyd and R. P. Smith. Some econometric issues in measuring the monetary transmission mechanism, with an application to developing countries. In L. Mahadeva and P. Sinclair, editors, *Monetary transmission in diverse economies*, pages 68–99. Cambridge University Press, 2002.

[25] H. Brücker and B. Siliverstovs. On the estimation and forecasting of international migration: how relevant is heterogeneity across countries? *Empirical Economics*, 31(3):735–754, 2006.

[26] A. Calza, T. Monacelli, and L. Stracca. Housing finance and monetary policy. *Journal of the European Economic Association*, 11(s1):101–122, 2013.

[27] J. Y. Campbell and J. F. Cocco. How do house prices affect consumption? evidence from micro data. *Journal of Monetary Economics*, 54(3):591–621, 2007.

[28] F. Canova and G. De Nicolo. Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics*, 49(6):1131–1159, 2002.

[29] R. Cardarelli, D. Igan, and A. Rebucci. The changing housing cycle and the implications for monetary policy. *World Economic Outlook: Housing and the Business Cycle*, 19:52, 2008.

[30] L. J. Christiano, M. Eichenbaum, and C. L. Evans. Monetary policy shocks: What have we learned and to what end? *Handbook of Macroeconomics*, 1:65–148, 1999.

[31] A. Chudik and M. H. Pesaran. Large panel data models with cross-sectional dependence: A survey. *CESifo Working Paper Series*, 4371, 2013.

[32] 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.

[33] 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*, pages 1–38, 1977.

[34] T. Doan, R. Litterman, and C. Sims. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1):1–100, 1984.

[35] C. Doz, D. Giannone, and L. Reichlin. A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of economics and statistics*, 94(4):1014–1024, 2012.

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

[37] M. Eberhardt and F. Teal. Econometrics for grumblers: a new look at the literature on cross-country growth empirics. *Journal of Economic Surveys*, 25(1):109–155, 2011.

[38] S. Edwards. LDC's foreign borrowing and default risk: An empirical investigation. *The American Economic Review*, 74(4):726–734, 1984.

[39] B. Eichengreen and A. Mody. What explains changing spreads on emerging-market debt: Fundamentals or market sentiment? *NBER Working Paper Series*, page 6408, 1998.

[40] S. Eickmeier and T. Ng. How do US credit supply shocks propagate internationally? a GVAR approach. *European Economic Review*, 74:128–145, 2015.

[41] A. Elbourne. The UK housing market and the monetary policy transmission mechanism: An SVAR approach. *Journal of Housing Economics*, 17(1):65–87, 2008.

[42] European Central Bank. Structural factors in the EU housing markets. March 2003.

[43] European Central Bank. Housing wealth and private consumption in the euro area. *ECB Monthly Bulletin*, pages 59–71, January 2009.

[44] European Central Bank. The state of the house price cycle in the euro area. *ECB Economic Bulletin*, (6), 2015.

[45] European Central Bank. Financial Stability Review. November 2016.

[46] European Systemic Risk Board. Report on residential real estate and financial stability in the EU. December 2015.

[47] J. Faust. The robustness of identified VAR conclusions about money. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 207–244. Elsevier, 1998.

[48] A. Ferrero. House price booms, current account deficits, and low interest rates. *Journal of Money, Credit and Banking*, 47(S1):261–293, 2015.

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

[50] R. Fry and A. Pagan. Some issues in using sign restrictions for identifying structural vars. *National Centre for Econometric Research Working Paper*, 14, 2007.

[51] R. Fry and A. Pagan. Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, 49(4):938–960, 2011.

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

[53] L. Gambetti and A. Musso. Loan supply shocks and the business cycle. *Journal of Applied Econometrics*, 2016.

[54] J. Geanakoplos. The leverage cycle. In *NBER Macroeconomics Annual 2009, Volume 24*, pages 1–65. University of Chicago Press, 2010.

[55] E. I. George, D. Sun, and S. Ni. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580, 2008.

[56] A. Gerali, S. Neri, L. Sessa, and F. M. Signoretti. Credit and banking in a DSGE model of the euro area. *Journal of Money, Credit and Banking*, 42(s1):107–141, 2010.

[57] M. Gertler and P. Karadi. A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34, 2011.

[58] M. Giuliodori. The role of house prices in the monetary transmission mechanism across european countries. *Scottish Journal of Political Economy*, 52(4):519–543, 2005.

[59] A. S. Goldberger. *Econometric theory.* New York: John Wiley & Sons, 1964.

[60] C. Goodhart and B. Hofmann. Asset prices, financial conditions, and the transmission of monetary policy. In *Conference on Asset Prices, Exchange Rates, and Monetary Policy, Stanford University*, pages 2–3, February 2001.

[61] C. W. Granger. Implications of aggregation with common factors. *Econometric Theory*, 3(02):208–222, 1987.

[62] C. W. J. Granger and P. Newbold. *Forecasting economic time series.* Academic Press, 2014.

[63] J. D. Hamilton. *Time series analysis.* Princeton University Press, 1994.

[64] U. N. Haque, M. H. Pesaran, and S. Sharma. Neglected heterogeneity and dynamics in cross-country savings regressions. In J. Krishnakumar and E. Rouchetti, editors, *Panel Data Econometrics - Future Directions: Papers in Honour of Prof. Balestra*, pages 53–82. Elsevier Science, 2000.

[65] P. Hartmann. Real estate markets and macroprudential policy in Europe. *Journal of Money, Credit and Banking*, 47(S1):69–80, 2015.

[66] C. R. Harvey and Y. Liu. Rethinking performance evaluation. *National Bureau of Economic Research*, (No. w22134), 2016.

[67] 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.

[68] F. Hayashi. Econometrics. *Princeton University Press*, 2000.

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

[70] J. Hilscher and Y. Nosbusch. Determinants of sovereign risk: Macroeconomic fundamentals and the pricing of sovereign debt. *Review of Finance*, pages 235–262, 2010.

[71] J. P. Hobert and G. Casella. The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473, 1996.

[72] S. Holly, M. H. Pesaran, and T. Yamagata. A spatio-temporal model of house prices in the usa. *Journal of Econometrics*, 158(1):160–173, 2010.

[73] N. Hristov, O. Hülsewig, and T. Wollmershäuser. Loan supply shocks during the financial crisis: Evidence for the euro area. *Journal of International Money and Finance*, 31(3):569–592, 2012.

[74] C. Hsiao. *Analysis of panel data*. Cambridge University Press, 2003.

[75] 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.

[76] C. Hsiao and M. H. Pesaran. Random coefficient panel data models. In Matyas and Sevestre, editors, *The Econometrics of Panel Data*, 2008.

[77] C. Hsiao, M. H. Pesaran, and A. K. Tahmiscioglu. Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, Lahiri, L. K., and M. Pesaran, editors, *Analysis of Panels and Limited Dependent Variable Models*, 1999.

[78] M. Iacoviello. House prices, borrowing constraints, and monetary policy in the business cycle. *The American Economic Review*, 95(3):739–764, 2005.

[79] M. Iacoviello and R. Minetti. The credit channel of monetary policy: Evidence from the housing market. *Journal of Macroeconomics*, 30(1):69–96, 2008.

[80] M. Iacoviello and S. Neri. The role of housing collateral in an estimated two-sector model of the US economy. *Boston College Working Papers in Economics*, (412), 2007.

[81] M. Iacoviello and S. Neri. Housing market spillovers: evidence from an estimated DSGE model. *American Economic Journal: Macroeconomics*, 2(2):125–164, 2010.

[82] M. Jarocinski and F. Smets. House prices and the stance of monetary policy. *Federal Reserve Bank of St. Louis Review*, 90(4):339–365, 2008.

[83] R. N. Kackar and D. A. Harville. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79(388):853–862, 1984.

[84] 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.

[85] M. G. Kenward and J. H. Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pages 983–997, 1997.

[86] L. Kilian and H. Lütkepohl. *Structural vector autoregressive analysis*. Cambridge University Press, 2017.

[87] G. Koop and D. Korobilis. Bayesian multivariate time series methods for empirical macroeconomics. In *Foundations and Trends in Econometrics*, volume 3, pages 267–358. Now Publishers Inc, 2010.

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

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

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

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

[92] H. Lutkepohl. *Handbook of matrices*. John Wiley & Sons, 1997.

[93] J. G. MacKinnon and H. White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325, 1985.

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

[95] 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.

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

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

[98] G. McLachlan and D. Peel. *Finite mixture models.* John Wiley & Sons, 2004.

[99] 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.

[100] J. Muellbauer. Housing and the macroeconomy: Inflation and the financial accelerator. *Journal of Money, Credit and Banking*, 47(S1):51–58, 2015.

[101] Y. Mundlak. Models with variable coefficients: integration and extension. In *Annales de l'INSEE*, pages 483–509, 1978(a).

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

[103] A. Musso, S. Neri, and L. Stracca. Housing, consumption and monetary policy: How different are the us and the euro area? *Journal of Banking & Finance*, 35(11):3019–3041, 2011.

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

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

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

[107] 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.

[108] M. H. Pesaran. Aggregation of linear dynamic models: an application to life-cycle consumption models under habit formation. *Economic Modelling*, 20(2):383–415, 2003.

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

[110] M. H. Pesaran and A. Chudik. Aggregation in large dynamic panels. *Journal of Econometrics*, 178:273–285, 2014.

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

[112] M. H. Pesaran and T. Yamagata. Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1):50–93, 2008.

[113] 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.

[114] M. Piazzesi and M. Schneider. Housing and macroeconomics. In J. Taylor and H. Uhlig, editors, *Handbook of Macroeconomics*. Elsevier, 2016.

[115] D. Quah and T. J. Sargent. A dynamic index model for large cross sections. In J. Stock and M. Watson, editors, *Business cycles, indicators and forecasting*, pages 285–310. University of Chicago Press, 1993.

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

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

[118] C. M. Reinhart and K. S. Rogoff. *This time is different: eight centuries of financial folly*. Princeton University Press, 2009.

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

[120] K. Reinhart, Carmen Rogoff, , and M. Savastano. Debt intolerance. *Brookings Papers on Economic Activity*, 1:1–74, 2003.

[121] R. Rigobon and B. Sack. The impact of monetary policy on asset prices. *Journal of Monetary Economics*, 51(8):1553–1575, 2004.

[122] 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.

[123] D. Schoenmaker. Macroprudentialism. *VoxEU.org eBook*, 2014.

[124] 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.

[125] R. J. Shiller. *Irrational exuberance*. Princeton University Press, 2005.

[126] C. A. Sims. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.

[127] T. Sinai and N. S. Souleles. Owner-occupied housing as a hedge against rent risk. *The Quarterly Journal of Economics*, 120(2):763–789, 2005.

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

[129] P. Swamy. *Statistical inference in random coefficient regression models*, volume 55. Springe-Verlag, 1971.

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

[131] L. Trapani and G. Urga. Optimal forecasting with heterogeneous panels: A monte carlo study. *International Journal of Forecasting*, 25(3):567–586, 2009.

[132] H. Uhlig. What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419, 2005.

[133] S. Wachter. The housing and credit bubbles in the United States and Europe: A comparison. *Journal of Money, Credit and Banking*, 47(S1):37–42, 2015.

[134] M. W. Watson and R. F. Engle. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400, 1983.

[135] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838, 1980.

[136] J. Zettelmeyer. The impact of monetary policy on the exchange rate: evidence from three small open economies. *Journal of Monetary Economics*, 51(3):635–652, 2004.