ESSAYS ON SPATIAL PANEL ECONOMETRICS

Karen Alejandra Miranda Gualdron

UNIVERSITAT ROVIRA i VIRGILI

Essays on Spatial Panel Econometrics

Karen A. Miranda Gualdrón

DOCTORAL THESIS
2018

Karen A. Miranda Gualdrón

Essays on Spatial Panel Econometrics

PH.D. DISSERTATION

Supervised by:
Miguel Manjón and Oscar Martínez
Deparment of Economics

UNIVERSITAT
ROVIRA i VIRGILI

Reus, 2018

UNIVERSITAT
ROVIRA i VIRGILI

Departament d'Economia

Av. De la Universitat 1,

43204, Reus

Tel: (+34) 977 759 811, 977 759 890

Fax: (+34) 977 758 907

We STATE that the present project, entitled "Essays on Spatial Panel Econometrics", presented by Karen A. Miranda Gualdrón for the award of the degree of Doctor, has been carried out under our supervision at the Department of Economics of this university, and it also fulfils all the requirements to receive the European/International Doctorate Distinction.

Reus, April 03th 2018,

Miguel Manjón Antolín

Oscar Martínez Ibañez

Acknowledgements

First, I would like to express my deepest gratitude to my PhD supervisors: Miguel Manjón and Oscar Martínez. Beyond and behind of this work, their teaching and training are my legacy now. Miguel, thank you for be a great leader, you always leaves behind in me the conviction that I can do great things (despite my doubts). Oscar, you are the funniest supervisor and one of the smartest people I know, thanks for sharing your knowledge, your help and patience. I have been extremely lucky to have these supervisors, their enthusiasm for doing research is a "spillover effect" (I mean, is contagious).

Second, I would like to thank all the members of the Department of Economics at URV. In particular, thanks to Professors Bernd Thielen and Josep Maria Arauzo for their support. I would also like to thank the research groups *QURE-CREIP* and *Agència de Gestió d'Ajuts Universitaris i de Recerca (Agaur)* for providing the funding which allowed me to undertake this research and also to attend conferences in many countries and thus meet so many interesting people.

Finally, thanks to Carlos Acuña, for his continued support, for listening and making me laugh, in particular during my blue mood times. I especially thank my mom (Nohemí), dad (Carlos), sister (Paula) and brother (Sebas), for helping me keep things in perspective. They also deserve thanks for their daily encourage. To end, I would like to express my gratitude to the PhD colleges/friends: Eva Coll, Judit Albiol, Sofia Jiménez and Jilber Urbina. They provided a much needed fun activities (dancing, eating, traveling, drinking, xbox sessions) that were an escape from my studies.

Contents

Chapter 1

Introduction

Externalities play a central role in understanding economic processes. Regions that are surrounded by high productivity regions, for example, may benefit from positive externalities (De Long and Summers, 1991; LeSage and Fischer, 2012). Other examples of spatial externalities include spillovers effects of taxation and expenditures on public services (Brueckner, 2003) and the impact of pollution in environmental economics (Beron et al., 2004). These externalities affect mainly close by neighbours and become less effective for distant regions.

A number of studies over the last decades have benefited from the increasing availability of data with a location component to analyse effects that spread over spatial units (zip codes, municipalities, regions, states, jurisdictions, countries, etc.). However, this growing literature would have probably not developed without the advances witnessed in spatial econometrics (LeSage and Pace, 2009), particularly in relation to spatial panel data models (Elhorst, 2012). This thesis contains three original studies in this area.

Panel data with a location component offer researchers extended modelling specifications, including accounting for effects that cannot be addressed using cross-section data (see e.g. Elhorst, 2014). In this respect, spatial panel data models essentially differ from the standard non-spatial panel data model in that they not only include a “random" or ‘fixed” effect component to control for the unobserved heterogeneity, but deal with the potential spatial dependence. Thus, the most commonly used models in applied research include a spatially lagged dependent variable (Spatial Lag Model), spatial effects in the error term (Spatial Error Model), and spatial effects in both the independent and the dependent variables (Spatial Durbin Model). To date, however, with the notable exception of Beer and Riedl (2012), the proposed model specifications have not considered the inclusion of spatial externalities in the individual effects. This is the main motivation for this thesis.

Distinguishing the individual effects from their spatial spillovers may provide interesting insights into how the unobserved characteristics of the neighbouring territories affect the output of a certain territory and, conversely, how the unobserved characteristics of a territory affect the output of the neighbouring territories. In growth models, for example, a measure of the unobserved productivity of the regions under study can be obtained from the estimated individual effects Islam (1995). But how does the spatial contagion of the unobserved productivity of these regions affects their economic growth? This thesis considers several model specifications to identify and estimate this kind of effects.

Beer and Riedl (2012) advocate using an extension of the Spatial Durbin Model for panel data that includes the spatially weighted individual effects. Ultimately, however, they argue that "it is (...) advisable to remove the spatial lag of the fixed effects from the equation as the inclusion of both, [the individual effects] and [their spatial lags], leads to perfect multicollinearity" (p. 302). Removing the spatial lag of the fixed effects does not generally preclude the consistent estimation of the parameters of the model. However, this practice rules out obtaining an estimate of the individual-specific effects (net of the spatially weighted effects). This raises the question of whether both the individual effects and their spatial spillovers can indeed be identified and estimated in fixed effects panel data models.

This thesis proposes using a correlated-random effects specification (Mundlak, 1978; Chamberlain, 1982) to analyse the spatial externalities of the individual effects. In this vein, the model specifications analysed in this thesis have a spatially weighted error component structure that is closely related to that proposed by Kapoor et al. (2007). In particular, the second chapter considers a correlated random effects specification in a static spatial panel data model and the third does so in a spatial dynamic panel data model. The fourth chapter makes use of the findings of the previous chapters in a growth model with spatial externalities.

More precisely, the second chapter of this thesis, which has already been published in *Spatial Statistics* (Miranda et al., 2017b), analyses the problem of estimating individual effects and their spatial spillovers in linear panel data models. In particular, it considers a spatial-$\boldsymbol{X}$ lag model for panel data (Halleck Vega and Elhorst, 2015). First, it shows that the individual-specific effects and their spatial spillovers are not generally identified in linear panel data models. Under certain conditions, however, it is showed that there is no identification problem if the covariates are correlated with the individual-specific effects (Mundlak, 1978; Chamberlain, 1982). Further, under assumptions of strict and sequential exogeneity on the explanatory variables, this chapter derives appropriate FGLS and IV estimators. Finally, this chapter illustrates the proposed estimators using a Cobb-Douglas production function specification and US state-level data from Munnell (1990). As in previous literature, this study finds no evidence of public capital spillovers. However, public capital does play a role in the positive "outward" spatial contagion of the individual effects.

The third chapter considers a correlated random effects specification of the spatial Durbin dynamic panel model with an error-term containing individual effects and their spatial spillovers. Following Yu et al. (2008) and Su and Yang (2015), this chapter derives the likelihood function of the model and the asymptotic properties of the quasi-maximum likelihood estimator under standard assumptions in the spatial econometrics literature. The model specification corresponds to a restricted version of the dynamic spatial Durbin model of Lee and Yu (2016), since it does not include the spatial lag of the lagged dependent variable among the regressors. This means that, in terms of spatial dependence, the model analysed lies somewhere in between that of Yu et al. (2008), who only consider the spatial lag of the dependent and lagged dependent variables, and that of Su and Yang (2015), in which "spatial dependence is present only in the error term". This chapter also provides illustrative evidence from a growth-initial level equation and data on 26 OECD countries analysed by Lee and Yu (2016). Interestingly, the estimated coefficients and standard errors largely replicate those reported by Lee and Yu (2016). However, results indicate the existence of spatial contagion in the individual effects. In particular, the estimated spill-in/out effects reveal the existence of groups of countries with common spatial patterns in their spillovers.

The fourth chapter presents a growth model with interdependencies in the heterogeneous technological progress, physical capital and stock of knowledge. The basic framework is similar to that of Ertur and Koch (2007), but considers additional sources of externalities across economies. While they assume that the technological progress depends on the own stock of physical capital and the stock of knowledge of the other economies, this study also considers the role of both the physical capital (López-Bazo et al., 2004; Egger and Pfaffermayr, 2006) and the (unobserved) initial level of technology (De Long and Summers, 1991; LeSage and Fischer, 2012) of the other economies. Moreover, it does not assume a common exogenous technological progress but accounts for heterogeneity in the initial level of technology, which here is interpreted as a proxy for total factor productivity (Islam, 1995). To illustrate this point, this chapter uses data on EU-NUTS2 regions and a correlated random effects specification to estimate the resulting spatial Durbin dynamic panel model with spatially weighted individual effects. QML estimates support the proposed model against simpler alternatives that impose a homogeneous technology and limit the sources of spatial externalities. Also, results indicate that rich regions tend to have higher "unobserved productivity" and are likely to stay rich because of the strong time and spatial dependence of the GDP per capita. Poor regions, on the other hand, tend to enjoy "unobserved productivity" spillovers but are likely to stay poor unless they increase their saving rates.

Chapter 2

Estimating individual effects and their spatial spillovers in linear panel data models: Public capital spillovers after all?[1]

2.1 Introduction

Does public capital have an effect on private output? And if it does, does this effect spill over nearby geographical areas? Using US state (and/or county) panel data, production function estimates have consistently concluded that public capital and its spatially weighted counterpart are not statistically significant.[2] In contrast, studies using alternative methodologies (e.g., VAR models), seem to suggest otherwise (Pereira and Andraz, 2013). In this paper we provide production function estimates supporting the existence of public capital spillovers. To be precise, we find no evidence of a direct positive effect of public capital on private output. However, we find evidence of a relation between public capital and the unobserved productivity of the states (i.e., the individual specific effect of the production function) and its spatial spillover.[3]

To obtain these results, this paper introduces a correlated-random effects model (Mundlak, 1978; Chamberlain, 1982) that presents spatial correlation in the individual

[1]This chapter is co-authored with Oscar Marínez-Ibáñez and Miguel Manjón-Antolín. It has already been published in *Spatial Statistics*, 22(Part 1): 1-17.

[2]See e.g. Munnell (1990); Baltagi and Pinnoi (1995); Holtz-Eakin and Schwartz (1995); Garcia-Mila et al. (1996); and Kelejian and Robinson (1997).

[3]As Boarnet (1998, p. 381-382) points out, "[p]ublic capital is provided at a particular place, and if such capital is productive, it enhances the comparative advantage of that location relative to other places". Also, "productive public capital might shift economic activity from one location to another".

effects. To our knowledge, only the random effects model of Kapoor et al. (2007) accounts for this spatial correlation. In the fixed effects case, Beer and Riedl (2012) advocate using an extension of the Spatial Durbin Model for panel data that includes the spatially weighted individual effects. Ultimately, however, they argue that "it is (...) advisable to remove the spatial lag of the fixed effects from the equation as the inclusion of both, [the individual effects] and [their spatial lags], leads to perfect multicollinearity" (p. 302). Removing the spatial lag of the fixed effects does not generally preclude the consistent estimation of the parameters of the model (see e.g. Halleck Vega and Elhorst, 2015). However, this practice rules out obtaining an estimate of the individual-specific effects (net of the spatially weighted effects).[4]

This raises the question of whether both individual effects and their spatial spillovers can indeed be identified in linear panel data models. In this paper we provide identifying conditions in a model specification that spatially weights both the independent variables and the individual effects. In particular, we show that there is no identification problem if the covariates are correlated with the individual-specific effects and the individual effects correspond to deviations from the constant term.

Having proved that the model is identified, we then consider the estimation of its parameters under alternative exogeneity assumptions on the explanatory variables. Under the assumption that all the explanatory variables are strictly exogenous (with respect to the idiosyncratic term), we derive a Feasible Generalised Least Squares (FGLS) estimator. We also prove that, regardless of the structure of the variance-covariance matrix of the correlation functions shocks, this estimator coincides with the within (fixed effects) estimator when all the explanatory variables are used to construct the correlation functions. Under the assumption that the explanatory variables are predetermined, we propose an Instrumental Variables (IV) estimator to address the endogeneity of the means of the predetermined explanatory variables used to approximate the correlation functions. We also advocate using the backward means of these variables (i.e., the means taken, for each period, over only current and past values) as instruments.

Lastly, we use these estimators and a (correlated random effects) production function specification to address the existence of capital spillovers. Using the data and (a spatially weighted variant of) the specification used by Munnell (1990), we find that, under strict exogeneity, our FGLS estimates of a Cobb-Douglas production function for the US states over the period 1970 to 1986 are largely consistent with those reported in related studies

[4] This is a critical issue, for example, in two-step models that use this estimate as the dependent variable (Combes and Gobillon, 2015). Similarly, obtaining an estimate of the spatial spillovers of the individual-specific effects may be of great interest (e.g., for assessing their geographical distribution, which is what we do in our empirical application).

(using this data set, as e.g. Baltagi and Pinnoi 1995; Kelejian and Robinson 1997; and using analogous data sets, as e.g. Holtz-Eakin and Schwartz 1995; Garcia-Mila et al. 1996).[5] However, when we explore the possibility that (some of) the explanatory variables are not exogenous, we find evidence of predeterminedness in the public capital. We then estimate the model by IV to find that, under a sequential exogeneity assumption, states with a larger/smaller estimated individual effect tend to have larger/smaller negative spatial spillovers. In particular, we consider both "spill-in" and "spill-out" effects (LeSage and Chih, 2016), although only the spill-out effects turn out to be statistically significant. Also, while the part of the individual effects associated with the private capital produces negative spatial contagion, the part associated with the public capital produces positive spatial contagion. Consistent with previous literature, however, we find no significant spatial spillovers in the public capital.

The rest of the paper is organised as follows. In Section 2.2 we discuss the identification problem and show that, under mild rank conditions, the correlated random effects model considered is identified. In Section 2.3 we present appropriate (FGLS and IV) estimators. In Section 2.4 we present empirical evidence based on the work of Munnell (1990). Section 2.5 concludes.

2.2 Specification and identification of the model

2.2.1 Spatial spillovers and the identification problem

Let us consider the spatial$-(\boldsymbol{X}, \boldsymbol{\Psi})$ panel data model, that is, the spatial (lag of) $\boldsymbol{X}$ model for panel data with spatially weighted fixed effects:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{\Psi}\boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{2.2.1}$$

where $\boldsymbol{y} = (y_{11}, \ldots, y_{1T}, \ldots, y_{N1}, \ldots, y_{NT})'$ is the dependent variable (as usual, $i = 1, \ldots, N$ denotes cross-sectional, geographical units and $t = 1, \ldots, T$ denotes the time dimension), $\boldsymbol{X} = (\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1T}, \boldsymbol{x}_{21}, \ldots, \boldsymbol{x}_{2T}, \ldots, \boldsymbol{x}_{N1}, \ldots, \boldsymbol{x}_{NT})'$ is the $NT \times K$ matrix of explanatory variables (i.e., $\boldsymbol{x}_{it}$ is a row vector of order K) is the $NT \times K$ matrix of explanatory variables, and $\boldsymbol{\varepsilon}$ is a zero-mean idiosyncratic error term with assumed variance-covariance matrix

[5] The data set we employ is publicly available and can be downloaded, for example, from the **Ecdat** package in R (a standardised binary contiguity spatial weights matrix of the US states is also included in the package).

$\sigma_\varepsilon^2 \boldsymbol{I}_{NT}$, with $\boldsymbol{I}_{NT}$ being the $NT \times NT$ identity matrix.[6]

We assume that neighbourhood relations do not change over time, so the spatial matrix is $\boldsymbol{W} = \mathbf{w} \otimes \boldsymbol{I}_T$, with $\boldsymbol{I}_T$ denoting the $T \times T$ identity matrix and $\mathbf{w} = [w_{ij}]$ being the $N \times N$ spatial weight matrix that describes the spatial arrangement of the units in the sample. Also, unobservable individual-specific effects are collected in $\boldsymbol{\Psi\mu}$, with $\boldsymbol{\Psi} = \boldsymbol{I}_N \otimes \boldsymbol{\iota}_T$, $\boldsymbol{I}_N$ being the $N \times N$ identity matrix and $\boldsymbol{\iota}_T$ a vector of ones of order T. Notice that, in contrast to the so-called SLX model (Halleck Vega and Elhorst, 2015), this model specification accounts for the spatial weights of the individual effects through the term $\boldsymbol{W\Psi\alpha}$. Thus, the parameters of the model are $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ and σ_ε^2. This means that $2(K + N) + 1$ parameters need to be estimated.

The main motivation behind the use of this model specification is the estimation of the individual effects and their spatial spillovers, since these often have a meaningful interpretation (Combes and Gobillon, 2015). In particular, following LeSage and Pace (2009), we define the spatial spillovers of the individual-specific effects in terms of the partial derivative of the (conditional expectation of the) dependent variable,

$$\frac{\partial E\left[\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\Psi}\right]}{\partial \boldsymbol{\Psi}_j} = \left(\boldsymbol{I}_N \mu_j + \mathbf{w}\alpha_j\right) \otimes \boldsymbol{I}_T \tag{2.2.2}$$

where $\boldsymbol{\Psi}_j$ is the j-th column of $\boldsymbol{\Psi}$ and $j = 1, \ldots, N$.

The off-diagonal elements of this matrix of partial derivatives represent the *spillovers* or *indirect effects* of unit j, whereas the diagonal elements of the matrix represent the *direct effects* of unit j. Notice, however, that since these effects are time-invariant, we can, without loss of generality, concentrate on the term in brackets, $\boldsymbol{I}_N \mu_j + \mathbf{w}\alpha_j$. Thus, a generic off-diagonal element of this matrix, $w_{il}\alpha_j$ with $i \neq l$, measures the effect of unit l having the unobservable characteristics of unit j on the dependent variable of unit i. Similarly, a generic element of the diagonal of this matrix, $\mu_j + w_{ii}\alpha_j$, measures the effect of unit i having the unobservable characteristics of unit j on the dependent variable of unit i. This direct effect reduces to μ_j for all i when $w_{ii} = 0$, the standard case.[7]

The vector of parameters $\boldsymbol{\mu}$ has thus a neat interpretation. However, the role the spatial weight matrix $\mathbf{w}$ and the vector of parameters $\boldsymbol{\alpha}$ play in the spill-over effects deserves

[6] Throughout the paper, we assume that a balanced (complete) panel data is available. However, results can easily be extended to incomplete panels.

[7] However, LeSage and Pace (2009) do not recommend reporting unit-level effects but scalar summary measures. Namely, the average of the main diagonal elements (direct effects) and the cumulative sum of the off-diagonal elements from each row, averaged over all rows (indirect effects). In particular, if the spatial weight matrix $\mathbf{w}$ is row-standarised and has zeros in the diagonal, these scalar summary measures correspond to $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$, respectively. We discuss the use of alternative unit-level indirect effects (analogous to the ones proposed by LeSage and Chih 2016) in section 2.4.

some further attention. On the one hand, the structure of $\mathbf{w}$ provides the definition of neighbourhood. That is, which units are affected (the "spill-out" effects) and which are affecting (the "spill-in" effects) by the spillover and, in distance-based matrices, how much will be affected/affecting each unit. Thus, different spatial weight matrices yield different spatial spillovers. On the other hand, the parameter α_j provides a measure of the spatial contagion of the individual effect of unit j irrespective of the number of neighbours it has and how close/distant they are. That is, α_j provides a measure of the "potentiality of the spatial contagion" associated with the individual effect of unit j. We thus refer to $\boldsymbol{\alpha}$ as the "potential" of the spatial spillovers of the individual effects.

We illustrate the calculation and interpretation of these direct and spillover effects in the empirical application of section 2.4. In any case, what is clear is that the estimation of these effects requires that of the parameter vectors $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ (since $\mathbf{w}$ is assumed to be a fixed and/or known matrix). It is therefore critical to determine whether these (and the rest of) parameters of the model are identified. Proposition 1 shows that, in general, this is not the case.

Proposition 2.1. *The Spatial (Lag of) $-(\boldsymbol{X}, \boldsymbol{\Psi})$ model for panel data with spatially weighted fixed effects is not identified for any spatial weight matrix* $\mathbf{w}$.

Proof. See appendix. □

Notice that, as Beer and Riedl (2012) argue, the omission of $\boldsymbol{W\Psi\alpha}$ does not preclude the consistent estimation of the parameters of the model. Thus, if the spatial spillovers of the individual-specific effects are of no interest for the application in hand, their suggestion to remove one of the components – i.e., either $\boldsymbol{\Psi}$ or $\boldsymbol{W\Psi}$ – is perfectly sensible. This is because the model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{WX\gamma} + \boldsymbol{\Psi\mu}^* + \boldsymbol{\varepsilon}, \tag{2.2.3}$$

with $\boldsymbol{\Psi\mu}^* = \boldsymbol{\Psi\mu} + \boldsymbol{W\Psi\alpha}$, is observationally equivalent to (2.2.1).

On the other hand, if the individual-specific effects and/or their spatial spillovers are of some interest, then the identification and estimation of the model need to be discussed. To this end, we go on to propose using a spatial$-(\boldsymbol{X}, \boldsymbol{\Psi})$ panel data model with correlated random effects. In particular, we show that, under mild assumptions, the model is identified. Later we present appropriate estimators under alternative exogeneity assumptions

2.2.2 The Correlated Random Effects Spatial$-(\boldsymbol{X}, \boldsymbol{\Psi})$ Panel Data Model

Fixed effects models implicitly assume that the individual effects are correlated with the covariates. But they somehow ignore this correlation in the estimation procedure. In fact, what the within and analogous transformations do (see e.g. Beer and Riedl, 2012) is to wipe out the individual effects so that this correlation is no longer a concern for the consistent estimation of the model. An alternative procedure for obtaining consistent estimates, however, is to incorporate this correlation into the model (Mundlak, 1978; Chamberlain, 1982). This is the approach followed here. In particular, we make use of the correlation between covariates and the (spatially weighted) individual effects to identify the spatial contagion in the individual effects.

Our modelling approach is related to that of Debarsy (2012), who uses a correlated random effects specification to construct an LR test on "the relevance of the random effects approach" (p. 112).[8] Notice, however, that although we both deal with the correlation between individual effects and covariates, our purposes differ markedly: while he seeks to correctly specify this correlation, we use it as a means to identify the spatial contagion in the individual effects. We also differ in the model specification which, although similar, treats the spatial contagion of the individual effects differently. Debarsy (2012) assumes that the individual effects depend on both the explanatory variables and the explanatory variables in their neighbourhood, but there is no spatial contagion in the individual effects. In contrast, we account for the spatial contagion in the individual effects (i.e., both the individual effects and their spatial spillovers are included in the specification) and assume that the individual effects and the "potential" of their spatial spillovers depend on the (mean of the) explanatory variables, which allows us to identify both the individual effects and their spatial spillovers. These alternative assumptions yield different error component structures: a one-way error in his case, a two-way error in ours (the additional component being a spatially weighted element).[9]

[8] Although this is not the aim of this paper, an analogous Wald test could be developed using our model specification and estimation procedures.

[9] Another important difference with the work of Debarsy (2012) is that whereas he analyses the Spatial Durbin Model (as Beer and Riedl 2012 do), our results are derived for the spatial $-(\boldsymbol{X}, \boldsymbol{\Psi})$ model. This allows us to address the identification and estimation of the model in a linear setting, whereas considering correlated random effects in a Spatial Durbin specification would result in a non-linear model in which identification and estimation are more involved. (Debarsy (2012, p. 115), for example, "assume(s) that all parameters are identified"; see, however, Lee and Yu 2016).

Thus, we assume the following relation between $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ and the explanatory variables:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{T}\boldsymbol{\Psi}'\,\boldsymbol{X}^*\,\boldsymbol{\Pi}_\mu + \boldsymbol{\upsilon}_\mu \\ \boldsymbol{\alpha} &= \frac{1}{T}\boldsymbol{\Psi}'\boldsymbol{X}\boldsymbol{\Pi}_\alpha + \boldsymbol{\upsilon}_\alpha, \end{aligned} \tag{2.2.4}$$

where $\boldsymbol{\Pi}_\mu$ and $\boldsymbol{\Pi}_\alpha$ are $(K+1)\times 1$ and $K\times 1$ parameter vectors to be estimated, respectively, and $\boldsymbol{X}^* = \left(\begin{array}{c:c} \boldsymbol{\iota}_{NT} & \boldsymbol{X} \end{array}\right)$ contains the covariates and a vector of ones. Notice that a constant term needs to be included in one of the equations in (2.2.4) to guarantee identification, since the spatial contagion of any common factor in the individual effects $\boldsymbol{\mu}$ (in particular, a constant term) is not identified. Ultimately, this means that we are implicitly assuming that the individual effects correspond to deviations from the constant term. Also, the error terms $\boldsymbol{\upsilon}_\mu$ and $\boldsymbol{\upsilon}_\alpha$ are assumed to be random vectors of dimension N with $\boldsymbol{\upsilon}_\mu \sim (\mathbf{0}, \sigma^2_\mu \boldsymbol{I}_N)$ and $\boldsymbol{\upsilon}_\alpha \sim (\mathbf{0}, \sigma^2_\alpha \boldsymbol{I}_N)$. However, $\boldsymbol{\upsilon}_\mu$ and $\boldsymbol{\upsilon}_\alpha$ are not assumed to be independent, the covariance parameter, $\sigma_{\mu\alpha}$, being such that $E(\boldsymbol{\upsilon}_\mu \boldsymbol{\upsilon}'_\alpha) = \sigma_{\mu\alpha}\boldsymbol{I}_N$ with E denoting the mathematical expectation.

Plugging equations in (2.2.4) into the model (2.2.1) we obtain

$$\boldsymbol{y} \;=\; \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\gamma} + \frac{1}{T}\boldsymbol{\Psi}\boldsymbol{\Psi}'\,\boldsymbol{X}^*\,\boldsymbol{\Pi}_\mu + \frac{1}{T}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}\boldsymbol{\Pi}_\alpha + \boldsymbol{\eta}, \tag{2.2.5}$$

where $\boldsymbol{\eta} = \boldsymbol{\Psi}\boldsymbol{\upsilon}_\mu + \boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\upsilon}_\alpha + \boldsymbol{\varepsilon}$. Notice that the resulting error component is similar to the one proposed by Kapoor et al. (2007) in that both error components allow for spatial contagion in the individual (random) effects. It is different in that while we assume that the idiosyncratic term is not spatially correlated (and propose an identification strategy that takes into consideration that the individual effects may have spatial effects), Kapoor et al. (2007) assume that the idiosyncratic term is spatially autocorrelated.

It is also interesting to note that our model specification does not impose the existence of spatial contagion in the individual effects. In fact, there is no contagion at all if both $\boldsymbol{\Pi}_\alpha$ and σ_α are zero (while there would still be some "random contagion" if $\boldsymbol{\Pi}_\alpha$ is zero but σ_α is not). Similarly, the model does not impose correlation between the individual effects and the covariates. In fact, the specification becomes that of a pure random effects model if both $\boldsymbol{\Pi}_\mu$ and $\boldsymbol{\Pi}_\alpha$ are zero (still with spatial contagion if σ_α is not zero). Thus, it is the statistical significance of these (sets of) parameters what ultimately determines the existence of spatial contagion in the individual effects and correlation between the individual effects and the covariates.

Finally, in contrast to the spatial $-(\boldsymbol{X}, \boldsymbol{\Psi})$ model for panel data in (2.2.1), Proposition 2 shows that the correlated random effects spatial panel data model in (2.2.5) is generally identified.

Proposition 2.2. *The correlated random effects spatial panel data model in* (2.2.5) *is identified if the matrix* $\widetilde{\mathbf{X}} = \left(\boldsymbol{X} \;\vdots\; \boldsymbol{W}\boldsymbol{X} \;\vdots\; \frac{1}{T}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}^* \;\vdots\; \frac{1}{T}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X} \right)$ *has full column rank.*

Proof. See appendix. □

Notice that since the number of parameters in the model is $4K+1$ (excluding the σ's), $NT \geq 4K+1$ is a necessary identification condition. Notice also that since $\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}^*$ is a $NT \times K+1$ matrix and $\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}$ is an $NT \times K$ matrix, both with row rank equal to N, $N \geq 2K+1$ is an additional necessary identification condition. Further, $\boldsymbol{X}$, $\boldsymbol{W}\boldsymbol{X}$, $\boldsymbol{\Psi}'\boldsymbol{X}^*$ and $\boldsymbol{W}\boldsymbol{\Psi}'\boldsymbol{X}$ must have full rank. Lastly, time-invariant regressors must be included in either the vector of explanatory variables ($\boldsymbol{X}$ and/or $\boldsymbol{W}\boldsymbol{X}$) or in the vector of determinants of the individual effects and their spatial spillovers ($\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}^*$ and/or $\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}$). Otherwise, there is exact multicollinearity between the explanatory variables.

2.3 Estimation

We start by noticing that consistent estimation of the parameters of the model does not depend on what the structure of the error term $\boldsymbol{\eta}$ is. In particular, assuming that the covariates are strictly exogenous (meaning here that $E(\varepsilon_{it}|\boldsymbol{x}_{j1}, \boldsymbol{x}_{j2}, \dots, \boldsymbol{x}_{jT}) = 0$ for all i, j), Ordinary Least Squares (OLS) estimates of (2.2.5) are consistent.[10] Yet a more efficient Generalized Least Squares (GLS) estimator can be derived by accounting for the error components structure of the model.

To be precise, the GLS estimator does not provide efficiency gains in the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters of the correlated random effects model in (2.2.5). In fact, as shown below (see also Mundlak 1978), the GLS estimator of these parameters coincides with the OLS estimator in (2.2.5) and the within or fixed-effects estimator of the (observationally equivalent) model in (2.2.3). This means that, if these parameters are the only ones of interest, efficiency considerations do not justify the use of the GLS estimator. In contrast, the GLS estimator of (2.2.5) may provide efficiency gains (with respect to OLS) in the $\boldsymbol{\Pi}_\mu$ and $\boldsymbol{\Pi}_\alpha$ parameters of the correlation functions. This is why, under the above strict exogeneity assumption, we propose using a FGLS estimator based on the estimates of the parameters of the variance-covariance matrix of $\boldsymbol{\eta}$.

[10] Notice that the spatial structure of our model requires an orthogonality condition involving not only all the time periods (a standard assumption in applied work; see e.g. Wooldridge 2002) but also all the units. Otherwise, we cannot guarantee the exogeneity of $\boldsymbol{W}\boldsymbol{X}$ and $\frac{1}{T}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}$.

On the other hand, none of these estimators are consistent if the strict exogeneity assumption does not hold. In particular, the presence of predetermined variables among the regressors ($\boldsymbol{X}$ and $\boldsymbol{WX}$) makes such variables endogenous when they are included among the variables that compose the correlation functions in (2.2.4). To obtain consistent estimates, we propose an IV estimator and the means of the endogenous variables taken, for each period, over only current and past values (backward means) as instruments.

Next we discuss the derivation of the proposed estimators in detail.

2.3.1 GLS estimation under strict exogeneity

Given our initial assumption of spherical disturbances and the stochastic assumptions about the behaviour of $\boldsymbol{\upsilon}_\mu$ and $\boldsymbol{\upsilon}_\alpha$, the error-component $\boldsymbol{\eta} = \boldsymbol{\Psi}\boldsymbol{\upsilon}_\mu + \boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\upsilon}_\alpha + \boldsymbol{\varepsilon}$ has zero mean and variance-covariance matrix given by

$$\begin{aligned}\boldsymbol{\Omega} &= \sigma_\mu^2\boldsymbol{\Psi}\boldsymbol{\Psi}' + \sigma_\alpha^2\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{W}' + \sigma_{\mu\alpha}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{W}' + \sigma_{\mu\alpha}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}' + \sigma_\varepsilon^2\boldsymbol{I}_{NT} \\ &= \boldsymbol{\Psi}\left(\sigma_\mu^2\boldsymbol{I}_N + \sigma_\alpha^2\mathbf{w}\mathbf{w}' + \sigma_{\mu\alpha}\mathbf{w} + \sigma_{\mu\alpha}\mathbf{w}'\right)\boldsymbol{\Psi}' + \sigma_\varepsilon^2\boldsymbol{I}_{NT} \qquad (2.3.1)\end{aligned}$$

Knowledge of this matrix suffices to derive the GLS estimator (see e.g. Wooldridge 2002) of the parameters of the correlated random effects model in (2.2.5). In particular, Mundlak (1978) proves that, when all the explanatory variables are used to construct the correlation functions and $\boldsymbol{\eta} = \boldsymbol{\Psi}\boldsymbol{\upsilon}_\mu + \boldsymbol{\varepsilon}$, this coincides with the within (fixed effects) estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in (2.2.3). Next we generalise this result to any (non-spherical) variance-covariance matrix of $\boldsymbol{\upsilon}_\mu$.

Proposition 2.3. *Consider the following correlated random effects model:*

$$\boldsymbol{y} = \mathbb{X}\boldsymbol{\lambda} + \frac{1}{T}\boldsymbol{\Psi}\boldsymbol{\Psi}'\mathbb{X}^*\boldsymbol{\Pi} + \boldsymbol{\eta}$$

with $\mathbb{X}^*$ being the matrix $\mathbb{X}$ plus a column of ones, $\boldsymbol{\lambda}$ and $\boldsymbol{\Pi}$ vectors of parameters with the appropriate dimension, and $E(\boldsymbol{\eta}\boldsymbol{\eta}') = \boldsymbol{\Omega} = \boldsymbol{\Psi}\boldsymbol{\Sigma}_\upsilon\boldsymbol{\Psi}' + \sigma_\varepsilon^2\boldsymbol{I}_{NT}$, where $\boldsymbol{\Sigma}_\upsilon$ is any variance-covariance matrix. The GLS, OLS and within (fixed effects) estimators of $\boldsymbol{\lambda}$ are the same.

Proof. See appendix. $\square$

Notice that our model corresponds to $\mathbb{X} = \left(\boldsymbol{X} \;\vdots\; \boldsymbol{WX}\right)$, $\boldsymbol{\lambda} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, $\frac{1}{T}\boldsymbol{\Psi}\boldsymbol{\Psi}'\mathbb{X} = \left(\frac{1}{T}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X} \;\vdots\; \frac{1}{T}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}\right)$ and $\boldsymbol{\Sigma}_\upsilon = \sigma_\mu^2\boldsymbol{I}_N + \sigma_\alpha^2\mathbf{w}\mathbf{w}' + \sigma_{\mu\alpha}\mathbf{w} + \sigma_{\mu\alpha}\mathbf{w}'$. Thus, Proposition 2.3 fully applies. More generally, the previous proof shows that, regardless of the structure of the variance-covariance matrix of $\boldsymbol{v}_\mu$ and $\boldsymbol{v}_\alpha$, the GLS (and OLS) estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ coincides with the within (fixed effects) estimator.

Next we consider the derivation of the feasible version of this GLS estimator. This basically requires a consistent estimate of the vector of parameters $\boldsymbol{\sigma} = (\sigma_\mu^2, \sigma_\alpha^2, \sigma_{\mu\alpha}, \sigma_\varepsilon^2)$. To this end, we notice that each component of $\boldsymbol{\Omega}$ can be written as a linear function of $\boldsymbol{\sigma}$:

$$E\left[\eta_{it}\eta_{ls}\right] = \boldsymbol{\sigma}\boldsymbol{M}_{ilts}, \tag{2.3.2}$$

where $i, l = 1, \ldots, N$ and $t, s = 1, \ldots, T$. Also, $E\left[\eta_{it}\eta_{ls}\right]$ denotes the mathematical expectation of $\eta_{it}\eta_{ls}$ and $\boldsymbol{M}$ is a 4×1 vector whose rows are functions of $\mathbf{w}$. More specifically,

$$E\left[\eta_{it}^2\right] = \sigma_\mu^2 + \sigma_\alpha^2 \sum_{j=1}^{N} w_{ij}^2 + 2\sigma_{\mu\alpha} w_{ii} + \sigma_\varepsilon^2 \tag{2.3.3}$$

$$E\left[\eta_{it}\eta_{is}\right] = \sigma_\mu^2 + \sigma_\alpha^2 \sum_{j=1}^{N} w_{ij}^2 + 2\sigma_{\mu\alpha} w_{ii} \quad \text{for } t \neq s \tag{2.3.4}$$

$$E\left[\eta_{it}\eta_{ls}\right] = \sigma_\alpha^2 \sum_{j=1}^{N} w_{ij} w_{lj} + \sigma_{\mu\alpha}(w_{il} + w_{li}) \quad \text{for } i \neq l \tag{2.3.5}$$

This allows us to consider the following linear regression to estimate $\boldsymbol{\sigma}$:

$$\hat{\eta}_{it}\hat{\eta}_{ls} = \boldsymbol{\sigma}\boldsymbol{M}_{ilts} + u_{ilts} \tag{2.3.6}$$

where $\hat{\boldsymbol{\eta}}$ is obtained as the residual term of a consistent estimation of the model in (2.2.5). Given the assumption of strict exogeneity of the covariates, OLS may be used for this purpose.

Under mild conditions, OLS estimation of (2.3.6) provides consistent estimates of $\boldsymbol{\sigma}$ (denoted by $\hat{\boldsymbol{\sigma}}$) and, with these in hand, we can obtain the FGLS estimates of the model.[11] In particular, $\hat{\boldsymbol{\sigma}}$ allows us to obtain $\widehat{\boldsymbol{\Omega}}_{GLS}$ using (2.3.1) and, by Cholesky Decomposition of its inverse, $\widehat{\boldsymbol{\Omega}}_{GLS}^{-1} = \boldsymbol{D}_{GLS}\boldsymbol{D}'_{GLS}$, the transformation matrix $\boldsymbol{D}'_{GLS}$. Finally, OLS estimation of the transformed model

$$\boldsymbol{D}'_{GLS}\boldsymbol{y} = \boldsymbol{D'}_{GLS}\left(\boldsymbol{X\beta} + \boldsymbol{WX\gamma} + \frac{1}{T}\boldsymbol{\Psi\Psi'X}^*\boldsymbol{\Pi}_\mu + \frac{1}{T}\boldsymbol{W\Psi\Psi'X\Pi}_\alpha + \boldsymbol{\eta}\right), \tag{2.3.7}$$

[11] Alternatively, one may impose the positiveness of the variances (σ_μ^2, σ_α^2 and that σ_ε^2) and that the correlation between $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ lies in the $[-1, 1]$ interval ($-1 \leq \sigma_{\mu\alpha} \times (\sigma_\mu^2 \times \sigma_\alpha^2)^{-1/2} \leq 1$), and use e.g a Non-Linear Least Squares estimator of $\boldsymbol{\sigma}$. Notice that this differs from the approach followed by e.g. Kapoor et al. (2007) in that their estimating equations are non-linear in the parameters of interest and they therefore have to resort to a Generalized Moment estimator (which can also be used here).

provides the FGLS estimates of $\widehat{\boldsymbol{\beta}}$, $\widehat{\gamma}$, $\widehat{\boldsymbol{\Pi}}_{\mu}$ and $\widehat{\boldsymbol{\Pi}}_{\alpha}$.

Interestingly, we can proceed in an analogous way to deal with error structures with idiosyncratic shocks following autoregressive and moving-average processes. In particular, the main difference with respect to the procedure proposed to deal with spherical disturbances is that the presence of a serial correlation matrix Kronecker (post-) multiplying the last term of equation (2.3.1) results in additional (autoregressive and moving-average) parameters in (2.3.2). Alternatively, we can account for the serial correlation in the model by including among the regressors lags of the dependent variable (and possibly of the explanatory variables in $\boldsymbol{X}$ and $\boldsymbol{WX}$). Notice, however, that the presence of the time-invariant components $\boldsymbol{v}_{\mu}$ and $\boldsymbol{v}_{\alpha}$ in the error term makes the lagged dependent variable endogenous. We thus propose instrumenting this variable using lags of the explanatory variables $\boldsymbol{X}$ (and possibly $\boldsymbol{WX}$) to control for the endogeneity of the lagged dependent variable.

2.3.2 Instrumental Variables Estimation Under Sequential Exogeneity

The assumption of strict exogeneity of the covariates is critical to guarantee that the GLS estimators presented in the previous section provide consistent estimates of the parameters of interest. However, in applications the assumption that ε_{it} is uncorrelated with the covariates in all the time periods may not hold. If for example the values of an explanatory variable in period t are related to past values of the dependent variable (e.g., in $t-1$), then future values of these explanatory variables (e.g., in $t+1$) may depend on the values of the idiosyncratic term in t, thus breaking the strict exogeneity assumption (see e.g. Wooldridge 2002).

In such circumstances, a sequential exogeneity assumption, $E(\varepsilon_{it}|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{it}) = 0$, seems more appropriate, since it implies that present values of y_{it} do not affect present and past values of $\boldsymbol{x}_{it}$. However, given the spatial structure of our model and following the strict exogeneity case, we instead propose using an "extended sequential exogeneity assumption" involving all the units in the sample. In mathematical terms, $E(\varepsilon_{it}|\boldsymbol{x}_{js}) = 0$ for all $\forall i, j$ and $s \leq t$. Notice also that if (expected) future values of $\boldsymbol{x}_{it}$ depend on y_{it} (i.e., present values of y_{it} affect the expected value of $\boldsymbol{x}_{it+1}$), then the explanatory variables used to construct the correlation functions in (2.2.4) are endogenous by construction. In other words, the presence of predetermined variables in $\boldsymbol{X}$ and $\boldsymbol{WX}$ means that $\frac{1}{T}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}^*$ and $\frac{1}{T}\boldsymbol{W}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{X}$ are correlated with the idiosyncratic term $\boldsymbol{\varepsilon}$. Therefore, under sequential exogeneity, the GLS estimators presented in the previous section no longer provide consistent estimates of the parameters of interest. Rather, an IV estimator should be considered for this purpose.

The main challenge IV estimators face in practice is that it is often difficult to find good instruments. In this case, however, the structure of the model provides natural candidates. Namely, the means of the exogenous explanatory variables constructed using values up to period t (rather than using all T values).[12] Let

$$\boldsymbol{L}_T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T} & \frac{1}{T} & \frac{1}{T} & \cdots & \frac{1}{T} \end{pmatrix}$$

be the row-standardised lower triangular matrix of ones and $\boldsymbol{\Gamma} = \boldsymbol{I}_N \otimes \boldsymbol{L}_T$ be the transformation matrix that yields the backward-up-to-t mean of the variable (i.e., $\boldsymbol{\Gamma X}$, for example, yields a matrix composed by the means of the exogenous explanatory variables constructed using values up to period t). The matrix of instruments can be thus written as $\boldsymbol{Z}_1 = \left(\boldsymbol{\Gamma X} \,\vdots\, \boldsymbol{\Gamma W X} \right)$.

Notice that these backward means are exogenous variables under the extended sequential exogeneity assumption. But they are also relevant, since by construction they are correlated with the endogenous explanatory variables $\frac{1}{T}\boldsymbol{\Psi\Psi}'\boldsymbol{X}^*$ and $\frac{1}{T}\boldsymbol{W\Psi\Psi}'\boldsymbol{X}$. Notice also that if we use the same explanatory variables to construct both the instruments and the correlation functions (or different variables but the same number), then the model is exactly identified. However, if all the explanatory variables are used to construct the instruments but not all the explanatory variables are used to construct the correlation functions, then the model is overidentified.

To construct the IV estimator, we follow Hausman and Taylor (1981) and Keane and Runkle (1992). Hausman and Taylor (1981) propose a two-step procedure to estimate linear panel data models with endogenous explanatory variables (with respect to the idiosyncratic term, as well as with respect to the individual effect) that boils down to an initial GLS transformation of the model (using a consistent estimate of the variance-covariance matrix of the error term) and then an estimation of the transformed model by IV. However, Keane and Runkle (1992) show that this procedure may not yield consistent estimates when the instruments are predetermined. This is because the GLS-transformation proposed by Hausman and Taylor (1981) results in individual errors that are linear combinations of the errors of the individual in all time periods. To obtain consistent estimates, Keane and

[12] In fact, provided that the number of available periods is long enough, one may use up to lagged periods to construct such means, i.e., one may use values up to period $t-1$, $t-2$, etc..

Runkle (1992) instead propose using the upper-triangular Cholesky decomposition of the serial correlation matrix (forward filtering) to GLS-transform the model. In essence, this is the procedure we follow, except that the complex structure of our error term requires a different GLS transformation and alternative orthogonality conditions between the errors and the explanatory variables.

To be precise, we obtain the IV estimates of our model in the following way. First, we transform the model using the projection matrix onto the column space of the matrix consisting of the exogenous variables and the instruments. That is, we multiply the model by the projection matrix $\boldsymbol{P_Z} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$, with $\boldsymbol{Z} = \left(\boldsymbol{X} \;\vdots\; \boldsymbol{WX} \;\vdots\; \boldsymbol{Z}_1 \right)$. This addresses the endogeneity problem and makes it possible to consistently estimate the transformed model by OLS. However, a more efficient estimation may be obtained if we transform the model to obtain spherical disturbances. To this end, we use these OLS estimates to generate the residuals $\hat{\boldsymbol{\eta}}$ and, after estimating (2.3.6), obtain $\widehat{\boldsymbol{\Omega}}_{IV}$ and $\boldsymbol{D}_{IV}$ in the same way as we did for $\widehat{\boldsymbol{\Omega}}_{GLS}$ and $\boldsymbol{D}_{GLS}$.

In particular, since our instruments are predetermined, we propose using the upper-triangular Cholesky decomposition of the inverse of the variance-covariance matrix (rather than that of the serial correlation matrix used by Keane and Runkle 1992) to obtain $\boldsymbol{D}_{IV}$. However, given the spatial structure of our model, we need to sort the data first by time and then by units within each time period before computing the upper-triangular Cholesky decomposition of $\widehat{\boldsymbol{\Omega}}_{IV}$.[13] This guarantees that the transformed errors in period t contain elements of η_{is} for $s \geq t$ and hence the exogeneity of our instruments in the transformed model.[14]

In the second step of the procedure, we estimate the GLS-transformed model by IV. This means that we again transform the model using the projection matrix $\boldsymbol{P_Z}$, except that now the sorting of the data requires using the matrix $\boldsymbol{\Gamma} = \boldsymbol{L}_T \otimes \boldsymbol{I}_N$ to construct $\boldsymbol{Z}_1$. The

[13]Notice that so far we have followed the standard practice of having the data sorted first by units and then by time within each unit, so e.g. the dependent variable was defined in Section 2.2 as $\boldsymbol{y} = (y_{11}, \ldots, y_{1T}, \ldots, y_{N1}, \ldots, y_{NT})'$. Here we require that $\boldsymbol{y} = (y_{11}, \ldots, y_{N1}, y_{12}, \ldots, y_{N2}, \ldots, y_{1T}, \ldots, y_{NT})'$ and $\boldsymbol{X} = (\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{N1}, \boldsymbol{x}_{12}, \ldots, \boldsymbol{x}_{N2}, \ldots, \boldsymbol{x}_{1T}, \ldots, \boldsymbol{x}_{NT})'$. In particular, notice that this sorting requires using $\boldsymbol{W} = \boldsymbol{I}_T \otimes \mathbf{w}$ and $\boldsymbol{\Psi} = \boldsymbol{\iota}_T \otimes \boldsymbol{I}_N$.

[14]Notice that, in a model without spatial dependence, Keane and Runkle (1992) assume that $E(\eta_{it}|\boldsymbol{z}_{is}) = 0$ for $s \leq t$. However, the presence of spatially weighted covariates in the model means that, for the proposed instruments, this only holds if the extended sequential exogeneity assumptions holds. Notice also that the observation (i,t) of the transformed error term, $\boldsymbol{D}'_{IV}\boldsymbol{\eta}$, contains the original error terms $\eta_{j,t}$ for $j = i, i+1, i+2, ..., N$ as well as $\eta_{j,s}$ for all j and $s > t$. This is why, in the presence of spatial dependence, the orthogonality condition proposed by Keane and Runkle (1992) does not suffice. In contrast, our extended sequential exogeneity assumption guarantees that the proposed instruments are exogenous to the transformed errors, since $E(\eta_{jt}|\boldsymbol{z}_{is}) = 0$ for all j and $s \leq t$.

transformed model,

$$\boldsymbol{P_Z D}'_{IV}\boldsymbol{y} = \boldsymbol{P_Z D}'_{IV}\left(\boldsymbol{X\beta} + \boldsymbol{WX\gamma} + \frac{1}{T}\boldsymbol{\Psi\Psi}'\,\boldsymbol{X}^*\,\boldsymbol{\Pi}_\mu + \frac{1}{T}\boldsymbol{W\Psi\Psi}'\boldsymbol{X\Pi}_\alpha + \boldsymbol{\eta}\right), \qquad (2.3.8)$$

is then estimated by OLS. The IV estimates of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{\Pi}}_\mu$ and $\widehat{\boldsymbol{\Pi}}_\alpha$ we obtain are not only consistent but also more efficient than the initial IV estimates obtained in the first step of the procedure.

Lastly, it is interesting to note that, unlike the strict exogeneity case, the treatment of serial correlation under sequential exogeneity does not follow immediately. In the present case, the presence of lags of the idiosyncratic term $\boldsymbol{\varepsilon}$ makes any predetermined variable in the model endogenous (not only those in the correlation functions). This is, however, not a major issue if the number of lags is small (the order of the moving average is low) and the time dimension of the panel is large. If the predetermined variable is among the regressors ($\boldsymbol{X}$ and $\boldsymbol{WX}$), we propose using lags of the variable as instruments; and if the predetermined variable is in the correlation function, we similarly propose adjusting the periods used to compute the backward means. Thus, we lose (at least) one period for each additional lagged term in the idiosyncratic error. The problem, of course, is that if the order of the moving average process driving the idiosyncratic term is not smaller than the number of time periods minus one, then there is no room for using lags and backward means as instruments. In particular, by the Wold representation theorem, this situation arises if the idiosyncratic term follows an AR process (of any order).

Alternatively, in applications in which strict exogeneity does not hold and there is serial correlation in the model, we may include among the regressors lags of the dependent variable (and possibly of the explanatory variables in $\boldsymbol{X}$ and $\boldsymbol{WX}$) and then apply the two-step procedure previously described. Notice, however, that in this specification the lagged dependent variable is endogenous (for the reasons pointed out in the GLS case). We thus proposed extending our matrix of instruments to include lags of the explanatory variables $\boldsymbol{X}$ (and possibly $\boldsymbol{WX}$) to control for the endogeneity of the lagged dependent variable.

2.4 Public capital spillovers in a production function specification: empirical evidence

In this section we use our correlated random effects specification and the proposed FGLS and IV estimators to empirically address the existence of public capital spillovers. To this end, we use a Cobb-Douglas production function specification and yearly data from (Munnell, 1990, p. 77) on 48 US contiguous states over the period 1970 to 1986. The output variable is the gross state product and the inputs include public capital, private capital and labour. "The unemployment rate is also included [in the regressions] to reflect the cyclical nature of productivity". All the variables except unemployment are in logs. This data set has the additional interest of having been partially (e.g. Garcia-Mila et al., 1996) or totally (e.g. Baltagi and Pinnoi, 1995) used in a number of studies on the relation between public capital and private output —see also Boarnet (1998) and Sloboda and Yao (2008). In particular, some of these studies have used spatial econometrics techniques (e.g. Holtz-Eakin and Schwartz, 1995; Kelejian and Robinson, 1997).

This accumulated evidence provides an excellent benchmark for our estimates, which were obtained for a model specification that uses all the inputs and their spatial counterparts to construct the correlation functions ($\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$). Before proceeding with the estimation, however, we considered the identification of the proposed model. We thus computed $det(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})$ to find that it was indeed positive, which means that our identification condition holds.

We report FGLS estimates of the model in Table 2.1 (coefficients and variance components). We also report the joint significance LM-tests of each subset of coefficients ($\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\Pi}_\mu$ and $\boldsymbol{\Pi}_\alpha$). The first thing to notice is that since our model specification uses all regressors ($\boldsymbol{X}$ and $\boldsymbol{WX}$) to construct the correlation functions ($\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$), estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ reported in Table 2.1 correspond to the within estimates of the model. As for the coefficient estimates of the variables that compose the correlation functions, $\boldsymbol{\Pi}_\mu$ and $\boldsymbol{\Pi}_\alpha$, all tend to be statistically significant (both individually and jointly). Lastly, all the variance components $\boldsymbol{\sigma}$ are statistically significant and have reasonable values. This supports our correlated random effects model specification. In particular, given that we reject that $\boldsymbol{\Pi}_\alpha$ is zero, there is evidence of contagion in the individual effects and, given that we reject that $\boldsymbol{\Pi}_\mu$ is zero, there is evidence of correlation between the individual effects and the covariates (fixed effects).

[Insert Table 2.1]

The FGLS estimates of the $\boldsymbol{\beta}$-coefficients associated with the main regressors ($\boldsymbol{X}$) are in line with those reported in previous studies. More precisely, they are close to those reported by Holtz-Eakin and Schwartz (1995) and Kelejian and Robinson (1997). While our estimate of the elasticity of labour is 0.7, for example, they estimated it to be between 0.6 and 0.9; similarly, our estimate of the elasticity of private capital is 0.2, while their estimates range from 0.06 to 0.2.[15] We further concur with the lack of statistical significance of public capital

[15]These estimates tend to be smaller that those reported by Baltagi and Pinnoi (1995) and Garcia-Mila et al. (1996), which may suggest that ignoring spatial dependence results in overestimation of the coefficients.

(see also Baltagi and Pinnoi, 1995; Garcia-Mila et al., 1996) and the statistical significance of the spatially weighted public capital (see the second column of Table 2.1, γ). Lastly, the statistical significance of public capital in the correlation function of the individual effects (see the third column of Table 2.1, $\mathbf{\Pi}_\mu$) is consistent with evidence reported by Baltagi and Pinnoi (1995, p. 396) that rejects the orthogonality between regressors and individual effects "only when the public capital stock (is) included in the production function".

Next we explore the possibility that the explanatory variables are not exogenous but predetermined. Previous related studies have analysed the endogeneity of (some of) the explanatory variables, with mixed results on the endogeneity tests (Baltagi and Pinnoi, 1995; Holtz-Eakin and Schwartz, 1995; Garcia-Mila et al., 1996) and implausible results on the coefficient estimates (Baltagi and Pinnoi, 1995; Holtz-Eakin and Schwartz, 1995). Here we address the predeterminedness of the public capital variable (and its spatially weighted counterpart). This would be the case, for example, if the amount states spend on public capital is related to past values of private output (e.g., because more prosperous states are likely to generate higher tax revenues). Under such circumstances, our previous discussion on the FGLS estimates is flawed, since the variables that compose the correlation functions defining $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ become endogenous and the FGLS estimator is no longer consistent.

We thus report results from an IV estimation in Table 2.2. These were obtained using as instruments backward-up-to-t means of all the explanatory variables and their spatially weighted counterparts. That is, $\boldsymbol{Z}_1$ contains the $\mathbf{\Gamma}-$transformations of public capital, private capital, labour and unemployment as well as of their spatially weighted counterparts.[16] At first sight, the IV estimates of the $\boldsymbol{\beta}$- and $\boldsymbol{\gamma}$-coefficients are not substantially different from those obtained by FGLS (perhaps with the exception of the public capital in $\boldsymbol{\gamma}$). In contrast, IV and FGLS estimates of the coefficients associated with the variables that compose the correlation functions, $\mathbf{\Pi}_\mu$ and $\mathbf{\Pi}_\alpha$, differ substantially. Indeed, a Hausman test between these two estimators strongly rejects the null hypothesis of strict exogeneity (the statistic is 159.42). This supports our tenet that public capital is actually a predetermined variable.

[Insert Table 2.2]

We then obtained an estimate of the direct individual effects (i.e., $\boldsymbol{\mu}$) and the "potential" of their spatial spillovers (i.e., $\boldsymbol{\alpha}$) using these IV estimates. However, we found that while the estimated direct efffects were generally statistically significant (both using FGLS and IV estimates), the "potential" of the spatial spillovers of the invididual effects were only

[16] We experimented with other set of instruments (e.g., without considering unemployment and its spatial weight) and found that coefficients estimates were barely altered.

statistically significant under the strict exogeneity assumption. Under the assumption that capital is a predetermined variable, the estimated α's were not statistically significant.[17]

Seeking for alternative, more efficient specifications that yielded statistically significant α's, we used a variables-selection procedure that resulted in the specification reported in Table 2.3.[18] Notice that the (common) estimated model coefficients are not that different from those reported in Table 2.2. Also, the correlation between the estimated $\boldsymbol{\mu}$'s and $\boldsymbol{\alpha}$'s (obtained from the estimates reported in Table 2.2 and Table 2.3) is 0.99 and 0.90, respectively. However, 14 out of the 48 estimated α's are now statistically significant at standard confidence levels. Alternatively, we considered an ad-hoc model selection procedure in which we explored different specifications in terms of instruments (e.g., using squared terms) and/or variables (e.g., dropping the unemployment, as in Holtz-Eakin and Schwartz 1995). This also produced specifications in which the α's were statistically significant and highly correlated with those obtained using the results reported in Table 2.2, while the coefficients of the model and their statistical significance remained generally unaltered. This was the case, for example, when dropping public capital from $\boldsymbol{WX}$, the unemployment from $\boldsymbol{\Pi}_\mu$, and the unemployment and its spatially weighted counterpart from the set of instruments.

Given the illustrative aim of this empirical exercise, determining which is the best model specification is clearly beyond the scope of this paper. It is important to stress, however, that little differences were observed among the alternative specifications we considered when we plotted on a map of the US states the estimated values of $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ (available from the authors upon request). Bearing this in mind, we have used the IV estimates reported in Table 2.3 to analyse the geographical distribution of the estimated direct individual effects and their spatial spillovers. In particular, our spillover effects are based on LeSage and Chih

[17] We use t-statistics to test the statistical significance of the individual effects and their spatial spillovers. In particular, the standard errors were obtained from $\widehat{Var}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}) = \frac{1}{T^2}\boldsymbol{\Psi}'\boldsymbol{X}^*\widehat{\boldsymbol{\Sigma}}_{\Pi_\mu}\boldsymbol{X}^{*\prime}\boldsymbol{\Psi} + \widehat{\sigma}^2_\mu \boldsymbol{I}_N - 2\frac{1}{T}\boldsymbol{\Psi}'\boldsymbol{X}^* M_\mu \widehat{E\left[\boldsymbol{\eta}\boldsymbol{v}'_\mu\right]}$, with $E\left[\cdot\right]$ denoting the mathematical expectation, $\boldsymbol{\Sigma}_{\Pi_\mu} = Var(\widehat{\boldsymbol{\Pi}}_\mu)$ and $\widehat{\boldsymbol{\Pi}}_\mu - \boldsymbol{\Pi}_\mu = M_\mu \boldsymbol{\eta}$. The expression for $\widehat{Var}(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$ is analogous, only differing in the subindices (i.e., using $\widehat{\boldsymbol{\Sigma}}_{\Pi_\alpha}$ instead of $\widehat{\boldsymbol{\Sigma}}_{\Pi_\mu}$, $\widehat{\sigma}^2_\alpha$ instead of $\widehat{\sigma}^2_\mu$, M_α instead of M_μ, and $\boldsymbol{v}'_\alpha$ instead of $\boldsymbol{v}'_\mu$). Under standard assumptions, these are consistent and (approximately) normally-distributed estimators of $Var(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu})$ and $Var(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$, respectively. Notice that these estimated variances allow us to also test whether the direct effects of two states are statistically equal (rather than whether the direct effect of one state is statistically equal to zero, as we did).

[18] In particular, we proceeded as follows. We dropped the most non-significant variable (i.e., that with the higher p-value) in the original specification (that reported in Table 2.2), reestimated the model and determined which was the most non-significant variable in the new specification. We then tested whether these two variables were jointlty statistically significant in the original specification. If the null hypothesis of this Wald test was not rejected, we dropped the two variables and considered again the most non-significant variable in the resulting specification. Then, we constructed a new Wald test for the null that the three variables were not jointly statistically significant in the original specification. We went on dropping variables and testing their joint significance untill either the jointly statistically non-significant hypothesis was rejected or all the variables in the model were statistically significant.

(2016), who define "spill-in" and "spill-out" effects as the cumulating off-diagonal elements from row- and column-sums of the matrix of marginal effects in (2.2.2), respectively.

[Insert Table 2.3]

To be precise, LeSage and Chih (2016) interpret the row-sums of $\mathbf{w}\alpha_j$ (i.e., $\sum_{l=1}^{N} w_{il}\alpha_j$, assuming that $w_{ii} = 0$) as the "spill-in effects" on the outcome of unit i. Thus, this spill-in effect captures the impact on the outcome of unit i of all the units neighbouring i having the unobserved characteristics of j. However, we find more interesting to report $\sum_{l=1}^{N} w_{il}\alpha_l$, which is the impact on the outcome of unit i of all the units neighbouring i having their unobserved characteristics (i.e., the impact on the outcome of unit i of the individual effects of the units neighbouring i). This is reported in Figure 2.1a.

Similarly, LeSage and Chih (2016) interpret the colum-sums of $\mathbf{w}\alpha_j$ (i.e., $\alpha_j \sum_{l=1}^{N} w_{li}$, assuming that $w_{ii} = 0$) as "spill-out effects" of unit i. In particular, this spill-out captures the impact of unit i having the unobserved characteristics of j on the outcome of the neighbours of unit i. Again, we find more interesting to report $\alpha_i \sum_{l=1}^{N} w_{li}$, which is the impact on the outcome of the units neighbouring i of the individual effect of unit i. Notice, however, that our proposed spill-out effect is the product of α_i (the "potential" of the spatial spillovers of the individual effects) and $\sum_{l=1}^{N} w_{li}$ (which in essence determines the spatial contagion, i.e., which units are affected by the spillover). Since our spill-in effect already reflects the spatial contagion of the individual effects, it seems more interesting to report $\boldsymbol{\alpha}$ rather than $\alpha_i \sum_{l=1}^{N} w_{li}$. This is consequently what we do in Figure 2.1b.

[Insert Figure 2.1]

Results indicate that the geographical distribution of the spatial contagion, as measured by the spill-in (Figure 2.1a) and spill-out (Figure 2.1b) effects, follows very much the same pattern (with some notable exceptions, such as Nebraska and Nevada). This means that most states show spill-in and spill-out effects that are of analogous magnitude. However, the "inwards" spatial contagion of the spill-in effects is generally not statistically significant (Colorado, Montana, Texas and Utah being the exceptions). We thus concentrate on the analysis of the spill-out effects which, as previously pointed out, tend to be statistically significant.

Figure 2.1b shows that, in absolute values, West-Central (from Texas to Kansas, but also North-Dakota and Indiana) and West-Mountain (New Mexico and Wyoming) states stand out as the areas with the highest “outwards" spatial contagion. These are thus states with individual effects that strongly and negatively spill over the neighbouring states. Interestingly, these spill out effects are generally statistically significant at conventional confidence levels. On the other hand, there are two areas of low spill-out effects (in absolute values): the West-Pacific (California and Washington) and the North-East (New York in the mid-Atlantic and Connecticut, Rhode Island, Massachusetts and Vermont in New England). These are thus states with individual effects that negatively spill over the neighbouring states, but for which the relative magnitude of these spillovers is small (and, in fact, not statistically significant).

Further, Figure 2.1c reveals that the states that have the highest estimated values of the direct individual effects are mostly located in the North and West of the country (plus Texas and Louisiana in the South): more precisely, in the East (Illinois, Michigan and Ohio) and West (Nebraska) North Central, the mid-Atlantic (New York and Pennsylvania), the West-Mountain (Montana and Wyoming) and West-Pacific (California and Washington) regions. Figure 2.1c also shows that the states with the lowest estimated values of the direct individual effects concentrate in New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont), although we also find some states in the West-Mountain (i.e., Idaho, Utah and Nevada) and the South-Atlantic (North and South Carolina) regions.

Lastly, it is interesting to note the overlap between Figure 2.1a, Figure 2.1b and Figure 2.1c. In fact, since the estimated values of the individual effects and their spatial spillovers show different sign (see the signs of the $\mathbf{\Pi}$-coefficients in Table 2.1, Table 2.2 and Table 2.3), Figure 2.1 points to a negative relation between $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\alpha}}$ ($\mathbf{w}\widehat{\boldsymbol{\alpha}}$). Importantly, this “proportionality" between the direct and indirect effects is not a feature of the model specification (LeSage and Chih, 2016). Rather, it arises as a genuine characteristic of the data.[19]

Therefore, states with larger/smaller estimated direct individual effects tend to have larger/smaller negative spatial spillovers (both spill-in and spill-out). Notice, however, that this is mostly driven by New England states (with small direct individual effects and small negative spatial spillovers) and the central and southern states (with large direct individual effects and large negative spatial spillovers, except for Nebraska, which shows large direct individual effects and small negative spatial spillovers). The West and North-East states, on the other hand, tend to have large estimated direct individual effects and small negative

[19]A simple regression between $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\alpha}}$, for example, indeed has a negative and statistically significant slope (-0.34, with a p-value of 0.03), whereas a simple regression between $\widehat{\boldsymbol{\alpha}}$ and $\mathbf{w}\widehat{\boldsymbol{\alpha}}$ has a positive and statistically significant slope (0.86 with a p-value of 0.00).

spatial spillovers.

To conclude, it is worth noting that negative spillovers that "might shift economic activity from one location to another" have previously been found by, for example, Boarnet (1998, p. 382) and Sloboda and Yao (2008) with respect to the stock of and the investment in public infrastructure, respectively. However, the source of this "crowding out" effect in our model are the unobservable characteristics of the states or "unobserved productivity" (which these studies cannot identify). In fact, consistent with the work of Holtz-Eakin and Schwartz (1995) and Kelejian and Robinson (1997), results reported in Table 2.2 show that the spatially weighted public capital has a negative coefficient, but is not statistically significant (while private capital is, and has a positive sign). Notice also that both private and public capital are positively related to the states' unobserved productivity (both variables show positive and statistically signs in $\mathbf{\Pi}_{\mu}$). However, the role of these variables in the spatial spillover of the productivity, $\mathbf{\Pi}_{\alpha}$, differs. While the investment in private capital is associated with negative spillovers, the investment in public capital is associated with positive spillovers.

2.5 Conclusions

In this paper we analyse the problem of estimating individual effects and their spatial spillovers in linear panel data models. In particular, we consider models in which the exogenous regressors are spatially weighted and there is no spatially lagged dependent variable (i.e., the so-called spatial-$\boldsymbol{X}$ model). We first show that in this model specification the individual effects and their spatial spillovers are not identified for any spatial weight matrix. Under mild assumptions, however, we show that they are identified in a correlated random effects specification. To be precise, we show that there is no identification problem in a spatial$-(\boldsymbol{X}, \boldsymbol{\Psi})$ panel data model with correlated random effects if certain rank conditions hold and the individual effects correspond to deviations with respect to the constant term.

We then consider the estimation of the parameters of the (identified) model. Under strict exogeneity of the covariates, OLS estimates are consistent. Here, though, we provide more efficient FGLS estimators (at least more efficient with respect to the coefficients of the variables that compose the correlation functions) and propose an IV estimator to tackle situations in which the strict exogeneity assumption may not hold and a sequential exogeneity assumption is upheld. In particular, we suggest using the means of the exogenous explanatory variables constructed using values up to period t as instruments for the endogenous explanatory variables used to construct the correlation function (which, ultimately, are "means-up-to-T" of the exogenous variables). Also, dropping the most recent periods used to construct these instrumental variables (i.e., using "means-up-to-$(t-s)$", with s being a positive integer) may provide further instruments and/or instruments for potentially endogenous regressors.

Lastly, we present results from an empirical application: the estimation of a Cobb-Douglas production function using US state data. We find statistically significant differences between the FGLS and IV estimates, which suggest that the strict exogeneity assumption that sustains the FGLS estimates may not hold because the public capital variable is actually predetermined. Also, IV (and FGLS) estimates show that the variables that compose the correlation functions, as well as the variance components, all tend to be statistically significant. This supports our correlated random effects model specification.

The geographical distribution of the IV-estimated direct individual effects and their spatial spillovers reveals the existence of three major regions: *i)* Central and South states, where both direct individual effects and negative spatial spillovers tend to be large; *ii)* New England states, where both the direct individual effects and negative spatial spillovers tend to be small; and *iii)* West and North-East states, where the estimated direct individual effects tend to be large and the negative spatial spillovers tend to be small. In addition, both the "inwards" and "outwards" spatial contagion of the individual effects (i.e., the spill-in and spill-out effects) involve negative spillovers, although this sign is mostly associated with the private capital (and labour) and is only statistically significant for the spill-out effects. Public capital, on the other hand, is behind the positive spatial contagion of the individual effects. Consistent with previous literature, however, public capital itself does not seem to convey statistically significant spatial spillovers.

2.6 Appendix: Proofs of Propositions

Proof of Proposition 2.1. The model in (2.2.1) is not identified for any spatial weight matrix $\mathbf{w}$ because $\boldsymbol{\Psi}$ and $\boldsymbol{W\Psi}$ are perfectly collinear. We prove this by showing that $\det\left[\left(\boldsymbol{\Psi} \mid \boldsymbol{W\Psi}\right)'\left(\boldsymbol{\Psi} \mid \boldsymbol{W\Psi}\right)\right]$ is zero for any spatial weight matrix $\mathbf{w}$. Let

$$\boldsymbol{A} = \left(\boldsymbol{\Psi} \mid \boldsymbol{W\Psi}\right)'\left(\boldsymbol{\Psi} \mid \boldsymbol{W\Psi}\right) = T\left(\begin{array}{c|c} \boldsymbol{I}_N & \mathbf{w} \\ \hline \mathbf{w}' & \mathbf{w}'\mathbf{w} \end{array}\right) \tag{2.6.1}$$

Then, by Schur complement,

$$\det(\boldsymbol{A}) = T^{2N}\det\left(\boldsymbol{I}_N\right)\det\left(\mathbf{w}'\mathbf{w} - \mathbf{w}'\left(\boldsymbol{I}_N\right)^{-1}\mathbf{w}\right) = T^{2N}\det\left(\mathbf{w}'\mathbf{w} - \mathbf{w}'\mathbf{w}\right) = 0 \tag{2.6.2}$$

$\square$

Proof of Proposition 2.2. Since the correlated random effects spatial$-(\boldsymbol{X}, \boldsymbol{\mu})$ panel data model is linear in parameters, it is identified *iff* $det(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}) \neq 0$. If $\widetilde{\mathbf{X}}$ has full rank, it is easy to show that $det(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}) > 0$ (see e.g. Corollary 14.2.14 and Theorem 14.9.4 in Harville 2008). $\square$

Proof of Proposition 2.3. Let $\widehat{\boldsymbol{\lambda}}_{OLS}$, $\widehat{\boldsymbol{\lambda}}_w$ and $\widehat{\boldsymbol{\lambda}}_{GLS}$ be the OLS, within and GLS estimators of $\boldsymbol{\lambda}$. We prove first that $\widehat{\boldsymbol{\lambda}}_{OLS} = \widehat{\boldsymbol{\lambda}}_w$. To this end, we start by noting that, by the Frisch-Waugh-Lovell theorem and given that $\boldsymbol{\Psi\Psi'\Psi\Psi'} = T\boldsymbol{\Psi\Psi'}$,

$$\widehat{\boldsymbol{\lambda}}_{OLS} = \left(\mathbb{X}'\boldsymbol{M}_1\mathbb{X}\right)^{-1}\left(\mathbb{X}'\boldsymbol{M}_1\boldsymbol{y}\right)$$

with $\boldsymbol{M}_1 = \boldsymbol{I}_{NT} - \frac{1}{T}\boldsymbol{\Psi\Psi'}\mathbb{X}^*\left(\mathbb{X}^{*\prime}\boldsymbol{\Psi\Psi'}\mathbb{X}^*\right)^{-1}\mathbb{X}^{*\prime}\boldsymbol{\Psi\Psi'}$. Also, let $\boldsymbol{Q} = \boldsymbol{I}_{NT} - \frac{1}{T}\boldsymbol{\Psi\Psi'}$, which satisfies that $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{Q}$ and $\boldsymbol{Q\Psi} = 0$. Lastly, since $\mathbb{X}' = \left(\begin{array}{c:c} \mathbf{0}_{K\times 1} & \boldsymbol{I}_{K\times K}\end{array}\right)\mathbb{X}^{*\prime}$, it can be proved that $\mathbb{X}'\boldsymbol{M}_1 = \mathbb{X}'\boldsymbol{Q}$. Consequently,

$$\widehat{\boldsymbol{\lambda}}_{OLS} = \left(\mathbb{X}'\boldsymbol{Q}\mathbb{X}\right)^{-1}\mathbb{X}'\boldsymbol{Q}\boldsymbol{y} = \widehat{\boldsymbol{\lambda}}_w$$

This concludes the first part of the proof. Next we prove that $\widehat{\boldsymbol{\lambda}}_{GLS} = \widehat{\boldsymbol{\lambda}}_{OLS} = \widehat{\boldsymbol{\lambda}}_w$. To this end, we start by noting that $\widehat{\boldsymbol{\lambda}}_{GLS}$ corresponds to the OLS estimator of the considered correlated random effects model transformed using the (upper triangular part of the) Cholesky decomposition of the inverse of the variance covariance matrix of $\boldsymbol{\eta}$, $\boldsymbol{\Omega}^{-1} = \boldsymbol{DD'}$. Therefore, by the Frisch-Waugh-Lovell theorem,

$$\widehat{\boldsymbol{\lambda}}_{GLS} = ((\boldsymbol{D}'\mathbb{X})'\boldsymbol{M}_2\boldsymbol{D}'\mathbb{X})^{-1}((\boldsymbol{D}'\mathbb{X})'\boldsymbol{M}_2\boldsymbol{D}'\boldsymbol{y})$$

with $\boldsymbol{M}_2 = \boldsymbol{I}_{NT} - \boldsymbol{D'\Psi\Psi'}\mathbb{X}^*(\mathbb{X}^{*\prime}\boldsymbol{\Psi\Psi'\Omega}^{-1}\boldsymbol{\Psi\Psi'}\mathbb{X}^*)^{-1}\mathbb{X}^{*\prime}\boldsymbol{\Psi\Psi'D}$. Also, from equation (19) in Henderson and Searle (1981),

$$\begin{aligned}\boldsymbol{\Omega}^{-1} = \left(\sigma_\varepsilon^2\boldsymbol{I}_{NT} + \boldsymbol{\Psi\Sigma}_v\boldsymbol{\Psi}'\right)^{-1} &= \frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_{NT} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{\Psi\Sigma}_v\boldsymbol{\Psi}'\boldsymbol{\Omega}^{-1}\\ &= \frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_{NT} - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi\Sigma}_v\boldsymbol{\Psi}',\end{aligned}$$

which implies that $\boldsymbol{\Psi\Sigma}_v\boldsymbol{\Psi}'\boldsymbol{\Omega}^{-1} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Psi\Sigma}_v\boldsymbol{\Psi}'$ and, using again that $\boldsymbol{\Psi\Psi'\Psi\Psi'} = T\boldsymbol{\Psi\Psi'}$, it can be proved that

$$\boldsymbol{\Psi\Psi'\Omega^{-1}} = \boldsymbol{\Omega^{-1}\Psi\Psi'} \tag{2.6.3a}$$

$$\boldsymbol{Q\Omega}^{-1} = \boldsymbol{\Omega}^{-1}\boldsymbol{Q} = \frac{1}{\sigma_\varepsilon^2}\boldsymbol{Q} \tag{2.6.3b}$$

Lastly, by (2.6.3a) matrix $\boldsymbol{M}_2$ can be rewritten as

$$\boldsymbol{M}_2 = \boldsymbol{I}_{NT} - \frac{1}{T}\boldsymbol{D}'\boldsymbol{\Psi}\boldsymbol{\Psi'}\mathbb{X}^* \left(\mathbb{X}^{*\prime}\boldsymbol{\Omega^{-1}}\boldsymbol{\Psi}\boldsymbol{\Psi'}\mathbb{X}^*\right)^{-1} \mathbb{X}^{*\prime}\boldsymbol{\Psi}\boldsymbol{\Psi'}\boldsymbol{D},$$

and so $\mathbb{X}'\boldsymbol{D}\boldsymbol{M}_2 = \mathbb{X}'\boldsymbol{Q}\boldsymbol{D}$. Consequently,

$$\widehat{\boldsymbol{\lambda}}_{GLS} = \left(\mathbb{X}'\boldsymbol{Q}\boldsymbol{\Omega}^{-1}\mathbb{X}\right)^{-1}\left(\mathbb{X}'\boldsymbol{Q}\boldsymbol{\Omega^{-1}}\boldsymbol{y}\right) = \left(\mathbb{X}'\boldsymbol{Q}\mathbb{X}\right)^{-1}\left(\mathbb{X}'\boldsymbol{Q}\boldsymbol{y}\right) = \widehat{\boldsymbol{\lambda}}_{OLS} = \widehat{\boldsymbol{\lambda}}_w,$$

where the second equality holds by (2.6.3b). □

Table 2.1: **FGLS estimates**.

Coefficients	$\boldsymbol{\beta}$	γ	$\boldsymbol{\Pi}_\mu$	$\boldsymbol{\Pi}_\alpha$
Private capital	0.199***	0.260***	0.197***	−0.477***
	(0.030)	(0.043)	(0.052)	(0.089)
Labour	0.724***	−0.027	−0.212***	0.101
	(0.035)	(0.050)	(0.066)	(0.115)
Unemployment rate	−0.002	−0.007***	−0.013	0.035*
	(0.001)	(0.002)	(0.010)	(0.018)
Public capital	−0.023	−0.129**	0.186***	0.230
	(0.030)	(0.051)	(0.070)	(0.146)
Joint LM-test	250.07***	17.83***	13.10***	8.28***
Variance Components	σ^2_μ	σ^2_α	$\sigma_{\mu\alpha}$	σ^2_ε
	0.0045***	0.0012***	0.0017***	0.0013***
	(0.0001)	(0.0003)	(0.0001)	(0.0002)

Note: *p-value<0.1; **p-value<0.05; ***p-value<0.01. The dependent variable is the gross state product. All the variables are in logs, except for the unemployment. Variance components were estimated by OLS.

Table 2.2: **IV estimates**.

Coefficients	$\boldsymbol{\beta}$	γ	$\boldsymbol{\Pi}_{\mu}$	$\boldsymbol{\Pi}_{\alpha}$
Private capital	0.255^{***}	0.259^{***}	0.351^{***}	-0.601^{***}
	(0.037)	(0.055)	(0.081)	(0.135)
Labour	0.676^{***}	-0.045	-0.666^{***}	-0.100
	(0.059)	(0.078)	(0.132)	(0.217)
Unemployment rate	-0.003	-0.009^{***}	0.009	0.067^{**}
	(0.002)	(0.003)	(0.015)	(0.029)
Public capital	-0.029	-0.100	0.541^{**}	0.661^{*}
	(0.125)	(0.163)	(0.230)	(0.347)
Joint LM-test	168.57^{***}	12.40^{***}	12.77^{***}	6.32^{***}
Variance Components	σ^2_{μ}	σ^2_{α}	$\sigma_{\mu\alpha}$	σ^2_{ε}
	0.0046^{***}	0.0008^{***}	0.0019^{***}	0.0019^{***}
	(0.0001)	(0.0003)	(0.0001)	(0.0002)

Note: *p-value<0.1; **p-value<0.05; ***p-value<0.01. The dependent variable is the gross state product. All the variables are in logs, except for the unemployment. The matrix of instruments consist of backward-up-to-t means of public capital, private capital, labour and unemployment as well as of their spatially weighted counterparts. Variance components were estimated by OLS.

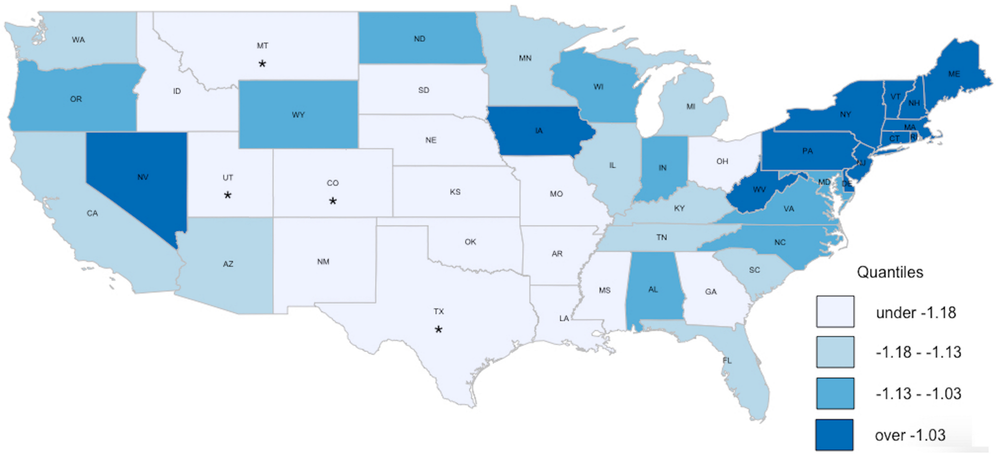
Table 2.3: **IV estimates**.

Coefficients	$\boldsymbol{\beta}$	γ	$\boldsymbol{\Pi}_\mu$	$\boldsymbol{\Pi}_\alpha$
Private Capital	0.252^{***}	0.419^{***}	0.342^{***}	-0.909^{***}
	(0.040)	(0.068)	(0.084)	(0.180)
Labour	0.666^{***}	-0.279^{***}	-0.776^{***}	
	(0.050)	(0.084)	(0.118)	
Unemployment rate	-0.011^{***}	-0.008^{***}		
	(0.002)	(0.003)		
Public Capital			0.660^{***}	0.877^{***}
			(0.132)	(0.193)
Variance Components	σ^2_μ	σ^2_α	$\sigma_{\mu\alpha}$	σ^2_ε
	0.0044^{***}	0.0026^{***}	0.0018^{***}	0.0015^{***}
	(0.0001)	(0.0003)	(0.0001)	(0.0002)

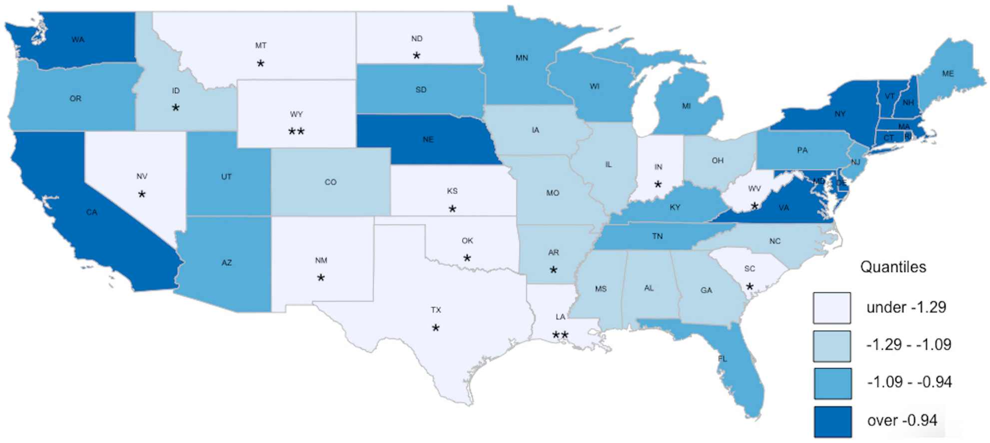
Note: *p-value<0.1; **p-value<0.05; ***p-value<0.01. The dependent variable is the gross state product. All the variables are in logs, except for the unemployment. The matrix of instruments consist of backward-up-to-t means of public capital, private capital, labour and unemployment as well as of their spatially weighted counterparts. Variance components were estimated by OLS.

Figure 2.1: **Estimated individual effects and their spatial spillovers.**

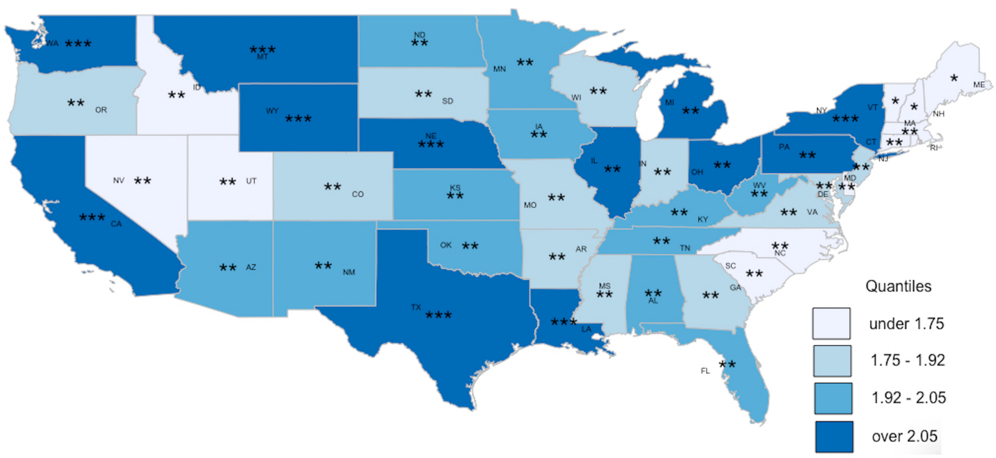
(a) Geographical distribution of $\mathbf{w}\widehat{\boldsymbol{\alpha}}$



(b) Geographical distribution of $\widehat{\boldsymbol{\alpha}}$



(c) Geographical distribution of $\widehat{\boldsymbol{\mu}}$



Note: *p-value<0.1; **p-value<0.05; ***p-value<0.01.

Chapter 3

A correlated random effects spatial Durbin model [1]

3.1 Introduction

The spatial Durbin model is a widely used specification in cross-section studies using georeferenced data (LeSage and Pace, 2009; LeSage, 2014). However, its use appears to be more limited with panel data. Although it has a certain appeal as a general framework to analyse spatial relations, concerns have been raised about its estimation and identification, particularly in its dynamic version (Elhorst et al., 2010; Elhorst, 2012). Despite these concerns, the spatial Durbin dynamic panel model (or, simply, dynamic spatial Durbin model) is expected to gain popularity in applied work, since identification conditions and Monte Carlo evidence for 2-Stage Least Squares (2SLS) and Quasi Maximum Likelihood (QML) estimators have recently been provided by Lee and Yu (2016). It is also interesting to note that Yu et al. (2008) and Su and Yang (2015) have analysed the asymptotic properties of QML estimators in restricted versions of the model specification analysed by Lee and Yu (2016).

In this paper we consider a correlated random effects specification (Mundlak, 1978; Chamberlain, 1982) of the spatial Durbin (dynamic) panel model and, following Yu et al. (2008) and Su and Yang (2015), derive the likelihood function of the model and proof that the QML estimator is consistent and asymptotically normal. To be precise, our model specification corresponds to a restricted version of the dynamic spatial Durbin model of Lee and Yu (2016), since we do not include the spatial lag of the lagged dependent variable among the regressors.[2] This means that, in terms of spatial dependence, our model specification

[1]This chapter is co-authored with Oscar Martínez-Ibáñez and Miguel Manjón-Antolín.

[2]See e.g. Elhorst (2012) for an overview of empirical studies using this model specification. Notice that the inclusion of the spatial lag of the lagged dependent variable would not make a substantial difference in proving the asymptotic properties of the QML estimator (other than complicate it).

lies somewhere in between that of Yu et al. (2008), who only consider the spatial lag of the dependent and lagged-dependent variables, and that of Su and Yang (2015, p. 231), in which "spatial dependence is present only in the error term". A major difference with respect to these papers is that while they consider a rather general variance-covariance matrix of the error term (which may contain individual and/or time effects), we consider an error-components structure with individual effects and their spatial spillovers (time effects can easily be incorporated), which results in a specific albeit involved variance-covariance matrix (see also Kapoor et al., 2007). Our proofs, however, are derived under rather standard assumptions in the spatial econometrics literature.[3]

Our model specification is inspired by the work of Beer and Riedl (2012), who advocate using an extension of the spatial Durbin model for panel data that controls for both the individual effects and the spatially weighted individual effects (see also Miranda et al., 2017b). Ultimately, however, they argue that "it is (...) advisable to remove the spatial lag of the fixed effects from the equation as the inclusion of both, [the individual effects] and [their spatial spillovers], leads to perfect multicollinearity" (p. 302). Removing the spatial lag of the fixed effects does not generally preclude the consistent estimation of the parameters of the model. However, this practice rules out obtaining an estimate of the individual-specific effects (net of the spatially weighted effects), which can be critical in certain applications. This is the case, for example, in growth models, where a measure of the unobserved productivity of the geographical units under study can be obtained from the estimated individual effects (Islam, 1995). Distinguishing the individual effects from their spatial spillovers can thus provide interesting insights into how the unobserved characteristics of the neighbouring territories affect the output of a certain territory and, conversely, how the unobserved characteristics of a territory affect the output of the neighbouring territories.

To illustrate this point, we estimate a growth-initial level equation using OECD data from Lee and Yu (2016). Unlike previous studies (e.g., Yu and Lee, 2012; Ho et al., 2013), however, our model specification not only accounts for observable "technological interdependences" (à la Ertur and Koch 2007) but also for unobserved ones (through the spatial spillovers of the individual effects). Interestingly, our estimated coefficients and standard errors largely replicate those reported by Lee and Yu (2016). This means that, since the spatial autoregressive parameter is not statistically significant, "the role played by technological interdependence on the growth of [OECD] countries" may not be as important as previously thought (Ertur and Koch 2007, p. 1052; see also Elhorst et al. 2010). In contrast, our results point to the existence of "unobservable technological interdependences" (i.e., spatial contagion in the – weakly significant – individual effects). Following Islam

[3]See e.g. Kelejian and Prucha (1998, 2001); Lee (2004); Yu et al. (2008) and Su and Yang (2015).

(1995), this may be interpreted as evidence that the growth of some countries is partially explained by the impact that the (unobserved productivity) of the neighbouring countries have on their economies. Lastly, computation of the "spill-in" and "spill-out" effects of the individual effects indicate that countries that impact less/more on other countries tend to be those that are less/more affected by the spillovers from their neighbours (Debarsy et al., 2012; LeSage and Chih, 2016). Further, they tend to have larger/smaller individual effects.

The rest of the paper is organised as follows. In Section 3.2 we present the model. In Section 3.3 we discuss its estimation by QML and derive the asymptotic properties of the QML estimator. In Section 3.4 we provide illustrative evidence. Section 3.5 concludes.

3.2 Model specification

In this paper we are interested in the following dynamic spatial autoregressive model with spatially weighted regressors and spatially weighted fixed effects:

$$Y_{nt} = \rho_0 Y_{n,t-1} + \lambda_0 W_n Y_{nt} + X_{nt}\beta_{10} + W_n X_{nt}\beta_{20} + \mu_n + W_n\alpha_n + \varepsilon_{nt} \tag{3.2.1}$$

where the subindex 0 denotes the "true" parameters of the model (e.g, ρ_0, λ_0, β_{10} and β_{20}), $Y_{nt} = (y_{1t}, y_{2t}, \cdots, y_{nt})'$ is an $n-$dimensional vector of dependent variables at time t, W_n is the exogenous spatial weight matrix that describes the spatial arrangement of the units in the sample, $X_{nt} = (x'_{1t}, x'_{2t}, \cdots, x'_{nt})'$ is a $n \times K$ matrix of regressors (i.e., x_{it} is a row vector of $1 \times K$), and ε_{nt} is the $n-$dimensional vector of disturbances at time t, with $\varepsilon_{nt} \sim (0, \sigma^2_\varepsilon)$, whose stochastic properties are discussed below. We assume, without loss of generality, that data is available for $i = 1, \ldots, n$ spatial units and $t = 1, \ldots, T$ time periods.[4]

Notice that this model specification critically differs from alternative specifications of the spatial Durbin dynamic panel data model (see e.g. Elhorst 2012) in that it includes both the individual effects (μ_n) and their spatial counterparts (α_n). Although the inclusion of $W_n Y_{nt}$ in the right-hand side of 3.2.1 produces "global" spatial contagion (Anselin, 2003) in the individual effects, our interest here lies in the existence of "local" spatial contagion. In particular, the individual-specific effects and their spatially weighted counterparts need to be estimated in order to determine which units are "locally" affecting and which units are "locally" affected, respectively, by the spatial spillover of the individual effect, and how

[4]Dealing with a "complete panel" is just meant to simplify notation and the burden of some proofs. Our results can easily be extended to incomplete panels. Notice similarly that the model does not contain time effects but these can easily be incorporated into the model (by e.g. including time dummies among the regressors, as we illustrate in the empirical application of Section 3.4).

intense such a “local" spillover is with respect to the total effect (i.e., the partial derivative of the conditional expectation of the dependent variable with respect to the individual effect). We discuss this issue in detail below, but first it is important to notice that this is generally not possible because 3.2.1 is observationally equivalent to a model that only includes individual effects (Beer and Riedl, 2012).

In this paper we follow Miranda et al. (2017b) in using a correlated random effects specification to identify the local spatial contagion in the individual effects. This means making use of the following correlation functions (Mundlak, 1978):

$$\begin{aligned} \mu_n &= l_n c_0 + \overline{X}_n \pi_{\mu_0} + \upsilon_{n\mu} \\ \alpha_n &= \overline{X}_n \pi_{\alpha_0} + \upsilon_{n\alpha}, \end{aligned} \tag{3.2.2}$$

where $\overline{X}_n = (\overline{X}'_{1\cdot}, \overline{X}'_{2\cdot}, \ldots, \overline{X}'_{n\cdot})'$ are composed of the period-means of the regressors, $\overline{X}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, π_{μ_0} and π_{α_0} are $K \times 1$ (“true”) parameter vectors, l_n is the unit vector of dimension $n \times 1$, and c_0 is the constant term to be estimated. The error terms, $\upsilon_{n\mu}$ and $\upsilon_{n\alpha}$, are assumed to be random vectors of dimension n, with $\upsilon_{n\mu} \sim (0, \sigma^2_{\mu_0} I_n)$ and $\upsilon_{n\alpha} \sim (0, \sigma^2_{\alpha_0} I_n)$, uncorrelated with ε_{nt}. Notice, however, that $\upsilon_{n\mu}$ and $\upsilon_{n\alpha}$ are not assumed to be independent, the covariance, $\sigma_{\mu\alpha_0}$, being such that $E(\upsilon_{n\mu}\upsilon'_{n\alpha}) = \sigma_{\mu\alpha_0} I_n$ with E denoting the mathematical expectation. Notice also that although we assume that the correlation functions are linear and have the means of the regressors as their main component, this does not always need to be the case. Non-linear functions, different moments and/or other variables may be employed to construct the correlation functions (Chamberlain, 1984). For the sake of simplicity, however, in this paper we restrict the analysis to the linear-means case.

Plugging equations 3.2.2 into model 3.2.1 we obtain

$$Y_{nt} = l_n c_0 + \rho_0 Y_{n,t-1} + \lambda_0 W_n Y_{nt} + X_{nt}\beta_{10} + W_n X_{nt}\beta_{20} + \overline{X}_n \pi_{\mu 0} + W_n \overline{X}_n \pi_{\alpha 0} + \eta_{nt} \tag{3.2.3}$$

where $\eta_{nt} = \upsilon_{n\mu} + W_n \upsilon_{n\alpha} + \varepsilon_{nt} = V_n + \varepsilon_{nt}$ (see also Kapoor et al., 2007). Notice that the variance-covariance matrix of this error term is given by $E[\eta_{nt}\eta'_{nt}] = E[V_n V'_n] + \sigma^2_{\varepsilon_0} I_n$, where $E[V_n V'_n] = \sigma^2_{\mu_0} I_n + \sigma_{\mu\alpha_0}(W_n + W'_n) + \sigma^2_{\alpha_0} W_n W'_n$ is the variance-covariance matrix of the composed error term of the individual effects and their spatial spillovers, V_n. Thus, if we define $\Sigma_0 = \frac{1}{\sigma^2_{\varepsilon_0}}\left(\sigma^2_{\mu_0} I_n + \sigma_{\mu\alpha_0}(W_n + W'_n) + \sigma^2_{\alpha_0} W_n W'_n\right)$, then the variance-covariance matrix of the error term can be rewritten as $E[\eta_{nt}\eta'_{nt}] = \sigma^2_{\varepsilon_0}(\Sigma_0 + I_n)$.

It is also worth noting the alternative specifications that are nested in our error term structure. The most obvious, perhaps, is the standard “random effects" (without spatial contagion), which is derived from our model by imposing the constraints $\pi_{\mu_0} = \pi_{\alpha_0} = 0$, $\sigma^2_{\alpha_0} = 0$ and $\sigma^2_{\mu 0} \neq 0$ (see e.g. Mundlak 1978 and Chamberlain 1982). Notice, however, that we may alternatively consider a "random effects" specification with spatial contagion by imposing the constraints $\pi_{\mu_0} = \pi_{\alpha_0} = 0$, $\sigma^2_{\alpha_0} \neq 0$ and $\sigma^2_{\mu_0} \neq 0$ and $\sigma_{\mu\alpha_0} \neq 0$ and, as a particular case, a "random effects" specification with proportional spatial contagion by imposing the constraints $\pi_{\mu_0} = \pi_{\alpha_0} = 0$, $\sigma^2_{\mu 0} \neq 0$, $\sigma^2_{\alpha 0} = a^2\sigma^2_{\mu 0}$ and $\sigma_{\mu\alpha 0} = a\sigma^2_{\mu 0}$ (or simply $\pi_{\mu_0} = \pi_{\alpha_0} = 0$ and $\alpha_n = a\mu_n$), with $a \neq 0$ constant. These, in turn, can be seen as a simplified version of the error structure proposed by Kapoor et al. (2007). Interestingly, however, our model also covers "fixed effects" versions of the previously discussed structures ("fixed" in the sense of being correlated with – some of – the regressors). That is, by imposing alternative constraints we may derive: *i)* a "fixed effects" error term ($\pi_{\mu_0} \neq 0$, $\pi_{\alpha_0} = 0$, $\sigma^2_{\alpha_0} = 0$ and $\sigma^2_{\mu_0} \neq 0$) analogous to that discussed by Mundlak (1978) and Chamberlain (1982), $\overline{X}_n\pi_{\mu 0} + \upsilon_{n\mu}$; *ii)* a "fixed effects" error term with spatial contagion ($\pi_{\mu_0} \neq 0$, $\pi_{\alpha_0} \neq 0$, $\sigma^2_{\alpha_0} \neq 0$ and $\sigma^2_{\mu_0} \neq 0$) and, if we impose that $\sigma^2_{\alpha_0} = 0$, a fixed effect error term analogous to that discussed by Debarsy (2012), $\overline{X}_n\pi_{\mu_0} + W_n\overline{X}_n\pi_{\alpha 0} + \upsilon_{n\mu}$, in which we cannot guarantee the existence of spatial contagion in the individual effects[5]; and *iii)* a "fixed effects" error term with proportional spatial contagion ($\pi_{\alpha_0} = a\pi_{\mu_0} \neq 0$, $\sigma^2_{\mu 0} \neq 0$, $\sigma^2_{\alpha_0} = a^2\sigma^2_{\mu_0}$ and $\sigma_{\mu\alpha_0} = a\sigma^2_{\mu_0}$, with $a \neq 0$ constant; or, simply, $\pi_{\mu_0} \neq 0$ and $\alpha_n = a\mu_n$).

3.2.1 Marginal effects: spatial spillovers and diffusion effects

Thus, providing that an estimate of μ_n and α_n is available, our model specification allows us to consider the existence of both "local" and "global" (through λ_0) spatial contagion in the individual effects (Anselin, 2003). However, because of the presence of the dynamic term $Y_{n,t-1}$ in the model, we may also consider the existence of "diffusion effects" in the partial derivative of the (conditional expectation of the) dependent variable with respect to the individual effects (Debarsy et al., 2012). To see this, let us rewrite the model in 3.2.1 as (by repeated substitution):

$$Y_{nt} = \rho_0^t S_0^{-t} Y_{n,0} + \sum_{s=0}^{t-1} \rho_0^s S_0^{-(s+1)} \left[X_{n,t-s}\beta_{10} + W_n X_{n,t-s}\beta_{20} + \mu_n + W_n\alpha_n + \varepsilon_{n,t-s} \right]$$

[5]Except if we impose, as we do, that the direct effect of the individual effects of a unit (see below) only depends on the characteristics of that unit and not on those of the other units.

where $S_0 = I_n - \lambda_0 W_n = S_n(\lambda_0)$.[6] In full matrix form:

$$\mathbf{Y} = \mathbf{G}_0 Y_{n,0} + \mathbf{C}_0 \mathbf{X}\beta_{10} + \mathbf{C}_0 \mathbf{W}\mathbf{X}\beta_{20} + \mathbf{C}_0(l_T \otimes I_n)\mu_n + \mathbf{C}_0 \mathbf{W}(l_T \otimes I_n)\alpha_n + \mathbf{C}_0 \boldsymbol{\varepsilon} \quad (3.2.4)$$

with $\mathbf{Y} = (Y'_{n1}, Y'_{n2}, \ldots, Y'_{nT})'$, $\mathbf{X} = (X'_{n1}, \ldots, X'_{nT})'$, $\boldsymbol{\varepsilon} = (\varepsilon'_{n1}, \ldots, \varepsilon'_{nT})'$, $\mathbf{W} = I_T \otimes W_n$, $\mathbf{G}_0 = \big(\rho_0 (S_0^{-1})', \rho_0^2 (S_0^{-2})', \ldots, \rho_0^T (S_0^{-T})'\big)'$ and

$$\mathbf{C}_0 = \begin{pmatrix} S_0^{-1} & 0 & 0 & \cdots & 0 \\ \rho_0 S_0^{-2} & S_0^{-1} & 0 & \cdots & 0 \\ \rho_0^2 S_0^{-3} & \rho_0 S_0^{-2} & S_0^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_0^{T-1} S_0^{-T} & \rho_0^{T-2} S_0^{-(T-1)} & \rho_0^{T-3} S_0^{-(T-2)} & \cdots & S_0^{-1} \end{pmatrix}$$

Lastly, let e_j be the j-th column of $l_T \otimes I_n$ with $j = 1, \ldots, n$. The marginal effects of the individual-specific effects are:

$$\frac{\partial}{\partial e_j} E\left(\mathbf{Y}|\mathbf{X}\right) = \mathbf{C}_0 \left[I_T \otimes (I_n \mu_j + W_n \alpha_j) \right] \quad (3.2.5)$$

where the diagonal elements of this matrix represent the *direct marginal effects* of unit j and the off-diagonal elements of this matrix represent the *spillovers or indirect marginal effects* of unit j (LeSage and Pace, 2009). Notice, however, that the dynamics of the model make direct and indirect effects stretch over time. That is, although the individual-specific effects are time-invariant, its marginal effects vary over time (to the extent that $\rho_0 \neq 0$). Yet we cannot interpret these variations as the result of "temporary" or "permanent" changes in the individual effects over time (which is the standard interpretation for regressors; see e.g. Debarsy et al. 2012). Bearing this in mind, the impact on the dependent variable in period $t = 1, ..., T$ is

$$\frac{\partial}{\partial e_j} E\left(Y_{nt}|\mathbf{X}\right) = \sum_{s=1}^{t} \rho_0^{s-1} S_0^{-s} (I_n \mu_j + W_n \alpha_j) \quad (3.2.6)$$

This expression can be interpreted as the "global" marginal effect (in period t), to the extent that it involves all the spatial units and not only at those considered to be neighbours

[6] We denote matrices and vectors depending on parameters of the model with the name of the matrix and vector, respectively, followed by the parameter(s) in brackets. For example, $S(\lambda) = S_n(\lambda)$. In particular, in the case of the "true" parameters we simply add the subindex zero to the name of the matrix. Thus, $S_0 = S_n(\lambda_0)$. Notice also that we use bold letters to denote $n \times T$ matrices (and similarly for $nT \times 1$ vectors), i.e., matrices resulting from stacking $n-$dimensional matrices. For example, $\mathbf{Y} = (Y'_{n1}, Y'_{n2}, \ldots, Y'_{nT})'$ and $\mathbf{X} = (X'_{n1}, \ldots, X'_{nT})'$, but also $\mathbf{S}(\lambda) = I_{nT} - \lambda(I_T \otimes W_n)$ and $\mathbf{S}_0 = I_T \otimes S_0$.

by W_n (Anselin, 2003). However, if we rewrite 3.2.6 as

$$\frac{\partial}{\partial e_j} E\left(Y_{nt}|\mathbf{X}\right) = \sum_{s=1}^{t} \rho_0^{s-1}(I_n \mu_j + W_n \alpha_j) + \sum_{r=1}^{\infty} \lambda_0^r W_n^r \sum_{s=1}^{t} \rho_0^{s-1} \sum_{m=0}^{s-1} S_0^{-m}(I_n \mu_j + W_n \alpha_j),$$

we notice that the first term in this expression only involves the neighbouring units (as defined by W_n). Thus, we may interpret $\sum_{s=1}^{t} \rho_0^{s-1}(I_n \mu_j + W_n \alpha_j)$ as the "local" marginal effect (Anselin, 2003). In fact, this is the marginal effect when $\lambda_0 = 0$, since in that case $W_n Y_{nt}$ is missing from the model and there is no "global" spatial contagion.

In particular, the row i and column m elements of $\sum_{s=1}^{t} \rho_0^{s-1} S_0^{-s}(I_n \mu_j + W_n \alpha_j)$ and $\sum_{s=1}^{t} \rho_0^{s-1}(I_n \mu_j + W_n \alpha_j)$ can be interpreted as the global and local impact, respectively, on the outcome of unit i of unit m having the unobserved characteristics of unit j. Following Miranda et al. (2017b), however, we find that is of greater interest to report the impact of unit m having its own unobserved characteristics (i.e., the unobserved characteristics of unit m) on the outcome of unit i. This means using the matrices

$$\sum_{s=1}^{t} \rho_0^{s-1} S_0^{-s} \left[diag(\mu_n) + W_n diag(\alpha_n)\right] \tag{3.2.7}$$

$$\sum_{s=1}^{t} \rho_0^{s-1} \left[diag(\mu_n) + W_n diag(\alpha_n)\right] \tag{3.2.8}$$

to compute the global and local marginal effects of interest, respectively. That is, the global and local marginal effects for each unit of all the other units having their own characteristics.

Thus, the main diagonal elements of these matrices provide, respectively, the direct global and local marginal effects (to reiterate, the impact on each unit of its own characteristics), whereas the off-diagonal elements of these matrices provide, respectively, the indirect global and local marginal effects (for a given time period t). We also obtain the spill-in and spill-out effects of the individual effects by respectively row- and column-summing the off-diagonal elements of these matrices (LeSage and Chih, 2016). In this vein the spill-in effect provides the global and local impact on the outcome of unit i of all the units neighbouring i having their unobserved characteristics, whereas the spill-out effect provides the global and local impact on the outcome of the units neighbouring i of the individual effect of unit i.

3.3 QML estimation: likelihood function and asymptotic properties

In this section we derive the quasi likelihood function of the spatial Durbin dynamic panel model with correlated random effects. We also study the consistency and asymptotic normality of the associated QML estimator. All results are obtained assuming that Y_{n0} is exogenous. The endogenous case, which is more involved (see e.g. Su and Yang, 2015), is left for future research.[7]

3.3.1 The QML estimator

Following the notation introduced in 3.2.4, let us now define $\mathbf{Y}_{-1} = \left(Y'_{n0}, Y'_{n1}, \ldots, Y'_{n(T-1)}\right)'$, $\overline{\mathbf{X}} = l_T \otimes \overline{X}_n$, $\widetilde{\mathbf{X}} = \left(\boldsymbol{l}_{nT} \,\vdots\, \mathbf{Y}_{-1} \,\vdots\, \mathbf{X} \,\vdots\, \mathbf{WX} \,\vdots\, \overline{\mathbf{X}} \,\vdots\, \mathbf{W}\overline{\mathbf{X}} \right)$, and $\boldsymbol{\eta} = (\eta'_{n1}, \ldots, \eta'_{nT})'$. We can then rewrite the model in 3.2.3, evaluated at any parameter value and to include all nT observations, as

$$\mathbf{SY} = \tilde{\mathbf{X}}\theta + \boldsymbol{\eta} \tag{3.3.1}$$

with $\theta = \left(c, \rho, \beta'_1, \beta'_2, \pi'_\mu, \pi'_\alpha\right)'$. Further, let $\psi = \left(\theta', \sigma^2_\varepsilon, \delta'\right)'$, $\delta = (\sigma', \lambda)'$, $\sigma' = (\sigma_1, \sigma_2, \sigma_3)'$, $\boldsymbol{\eta}(\lambda, \theta) = \mathbf{S}(\lambda)\mathbf{Y} - \tilde{\mathbf{X}}\theta$ and $\sigma^2_\varepsilon \boldsymbol{\Omega}(\sigma) = \sigma^2_\varepsilon \left(J_T \otimes \Sigma(\sigma) + I_T \otimes I_n\right)$, with $\Sigma(\sigma) = \sigma_1 I_n + \sigma_2(W_n + W'_n) + \sigma_3 W_n W'_n$. Then, the quasi-loglikelihood function of the model in 3.3.1 can be written as

$$\mathcal{L}(\psi) = \ln|\mathbf{S}(\lambda)| - \frac{nT}{2}\ln(2\pi) - \frac{nT}{2}\ln(\sigma^2_\varepsilon) - \frac{1}{2}\ln|\boldsymbol{\Omega}(\sigma)| - \frac{1}{2\sigma^2_\varepsilon}\boldsymbol{\eta}'(\lambda, \theta)\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}(\lambda, \theta). \tag{3.3.2}$$

where $|\cdot|$ denotes the determinant of a matrix. Notice that, given δ, the values of θ and σ^2_ε that maximize 3.3.2 are given by:

$$\begin{aligned} \hat{\theta}(\delta) &= \left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{S}(\lambda)\mathbf{Y} \\ \hat{\sigma}^2_\varepsilon(\delta) &= \frac{1}{nT}\hat{\boldsymbol{\eta}}'(\delta)\boldsymbol{\Omega}^{-1}(\sigma)\hat{\boldsymbol{\eta}}(\delta), \end{aligned} \tag{3.3.3}$$

[7] In any case, it is interesting to note that Monte Carlo evidence reported by Su and Yang (2015, p. 202-203) shows that, in the random effects case, estimating a model assuming that Y_{n0} is exogenous when it is actually not yields "estimates [that] are in general quite close to the true estimates except [when ρ is] large and positive" whereas, in the fixed effects model, "a wrong treatment on the initial values may lead to misleading results though to a much lesser degree as compared with the case of random effects model".

where $\hat{\boldsymbol{\eta}}(\delta) = \mathbf{S}(\lambda)\mathbf{Y} - \tilde{\mathbf{X}}\hat{\theta}(\delta)$. Thus, substituting 3.3.3 into 3.3.2 we obtain the concentrated quasi-loglikelihood function of δ:

$$\mathcal{L}_c(\delta) = \ln|\mathbf{S}(\lambda)| - \frac{nT}{2}(\ln(2\pi)+1) - \frac{nT}{2}\ln\left(\hat{\sigma}_\varepsilon^2(\delta)\right) - \frac{1}{2}\ln|\boldsymbol{\Omega}(\sigma)| \tag{3.3.4}$$

Maximising 3.3.4 yields the QML estimator of δ, $\hat{\delta} = (\hat{\sigma}', \hat{\lambda})'$, whereas the QMLE estimators of θ and σ_ε^2 are given by $\hat{\theta} \equiv \hat{\theta}(\hat{\delta})$ and $\hat{\sigma}_\varepsilon^2(\hat{\delta}) = \hat{\sigma}_\varepsilon^2$, respectively. Further, the QML estimator of $\left(\sigma_\mu^2, \sigma_{\mu\alpha}, \sigma_\alpha^2\right)$ is given by $\left(\hat{\sigma}_\mu^2, \hat{\sigma}_{\mu\alpha}, \hat{\sigma}_\alpha^2\right) = \hat{\sigma}_\varepsilon^2\left(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3\right) = \hat{\sigma}_\varepsilon^2\hat{\sigma}$. Therefore, $\hat{\psi} = \left(\hat{\theta}', \hat{\sigma}_\varepsilon^2, \hat{\delta}'\right)'$.

3.3.2 Asymptotic Properties

To derive the asymptotic properties of the QML estimator of the model, we must first ensure that $\psi = \left(\theta', \sigma_\varepsilon^2, \delta\right)'$ is identifiable. Notice, however, that given 3.3.3 it suffices to ensure that $\delta = (\sigma', \lambda)'$ is identifiable. To this end, let us define $\mathcal{L}_c^*(\delta) = \max_{\theta, \sigma_\varepsilon^2} E\left[\mathcal{L}(\psi)\right]$. It can be proved that the arguments that maximize $E\left[\mathcal{L}(\psi)\right]$ given δ are:

$$\tilde{\theta}(\delta) = \left[E(\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\widetilde{\mathbf{X}})\right]^{-1} E\left[\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}\right] \tag{3.3.5}$$

$$\tilde{\sigma}_\varepsilon^2(\delta) = \frac{1}{nT}E\left[\tilde{\boldsymbol{\eta}}'(\delta)\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\boldsymbol{\eta}}(\delta)\right] \tag{3.3.6}$$

with $\tilde{\boldsymbol{\eta}}(\delta) \equiv \boldsymbol{\eta}(\tilde{\theta}(\delta), \lambda)$. Consequently:

$$\mathcal{L}_c^*(\delta) = \ln|\mathbf{S}(\lambda)| - \frac{nT}{2}(\ln(2\pi)+1) - \frac{nT}{2}\ln\left(\tilde{\sigma}_\varepsilon^2(\delta)\right) - \frac{1}{2}\ln|\boldsymbol{\Omega}(\sigma)| \tag{3.3.7}$$

Notice also that, by using Lemma 3.3, $\tilde{\theta}(\delta_0) = \theta_0$ and $\tilde{\sigma}_\varepsilon^2(\delta_0) = \sigma_{\varepsilon_0}^2$.

Let us now denote by $\Delta = \Delta_\sigma \times \Delta_\lambda$ the (compact) parameter space of δ, with Δ_σ and Δ_λ being the (compact) parameter spaces of σ and λ, respectively.[8] Further, let us redefine $\widehat{\delta} = \max_{\delta\in\Delta} \mathcal{L}_c(\delta)$. We then require the following assumptions to prove that the QML estimator of the model, $\widehat{\psi} = (\widehat{\theta}', \widehat{\sigma}_\varepsilon^2, \widehat{\delta}')'$, is consistent and asymptotically normally distributed:

Assumption 3.1. *The available observations are* (y_{it}, x_{it}), $i = 1, ..., n$ *and* $t = 1, ..., T$, *with* $T \geq 2$ *fixed and* $n \to \infty$. *Also, all the elements of* x_{it} *are independent across* i, *and have* $4 + \epsilon_0$ *moments for some* $\epsilon_0 > 0$.

Assumption 3.2. *The elements of the disturbance vector* ε_{it} *are i.i.d. for all* i *and* t, *with* $E\left(\varepsilon_{it}\right) = 0$, $Var\left(\varepsilon_{it}\right) = \sigma_{\varepsilon_0}^2$ *and* $E\left|\varepsilon_{it}\right|^{4+\epsilon_0} < \infty$ *for some* $\epsilon_0 > 0$. *Similarly,* $(\upsilon_{i\mu}, \upsilon_{i\alpha})$ *are*

[8] Notice that we do not require specific assumptions about the parametric space of ρ_0. In particular, since we concentrate on the case of T finite and Y_{n0} exogenous, we do not need to assume that $|\rho_0| < 1$ to derive the results obtained in the paper (see Su and Yang, 2015, p. 236).

i.i.d. with $E(\upsilon_{i\mu}) = E(\upsilon_{i\alpha}) = 0$, $Var(\upsilon_{i\mu}) = \sigma^2_{\mu 0}$, $Var(\upsilon_{i\alpha}) = \sigma^2_{\alpha 0}$, $Cov(\upsilon_{i\mu}, \upsilon_{i\alpha}) = \sigma_{\mu\alpha 0}$ *and have* $4+\epsilon_0$ *finite moments for some* $\epsilon_0 > 0$. *Moreover,* ε_{it} *and* $(\upsilon_{j\mu}, \nu_{j\alpha})$ *are* ***i)*** *mutually independent, and* ***ii)*** *independent of* x_{sr} *for all* $i, j, s = 1 \dots n$ *and* $r, t = 1 \dots T$. *Lastly,* $\sigma_0 = (\sigma_{10}, \sigma_{20}, \sigma_{30})'$ *is in the interior of* Δ_σ.

Assumption 3.3. *The elements of* W_n, W_{nij}, *are at most of order* h_n^{-1}, *uniformly in all* i *and* j *with* $h_n/n \to 0$ *as* $n \to \infty$.

Assumption 3.4. *Matrix* $S(\lambda)$ *is nonsingular for all* $\lambda \in \Delta_\lambda$, *with* λ_0 *being in the interior of* Δ_λ.

Assumption 3.5. *The sequence of matrices* W_n *and* $S^{-1}(\lambda)$ *are uniformly bounded in both row and column sums and uniformly in* λ *in the compact parameter space* Δ_λ.[9]

Assumption 3.6. $\lim_{n\to\infty} \frac{1}{nT}\left\{\ln\left|\sigma^2_{\varepsilon 0}\mathbf{S}_0^{-2}\mathbf{\Omega}_0\right| - \ln\left|\tilde{\sigma}^2_\varepsilon(\delta)\mathbf{S}(\lambda)^{-2}\mathbf{\Omega}(\sigma)\right|\right\} \neq 0$ *for any* $\delta \neq \delta_0$. *Also,* $\frac{1}{nT}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}$ *is positive definite almost surely for* n *sufficiently large.*

Assumption 3.7. *Let* $\boldsymbol{H}_n(\psi) = \frac{\partial^2}{\partial\psi\partial\psi'}\mathcal{L}(\psi)$ *be the hessian of the likelihood function and* $\mathcal{G}_n(\psi) = \frac{\partial}{\partial\psi}\mathcal{L}(\psi)\frac{\partial}{\partial\psi'}\mathcal{L}(\psi)$ *be the product of the score vector. Both* $\boldsymbol{H} = \lim_{N\to\infty}\frac{1}{nT}E[\boldsymbol{H}_n(\psi_0)]$ *and* $\mathcal{G} = \lim_{n\to\infty}\frac{1}{nT}E[\mathcal{G}_n(\psi_0)]$ *exist. Also,* $\mathcal{G}$ *and* $-\boldsymbol{H}$ *are positive definite matrices.*

Assumption 3.8. *Matrix* $\mathbf{\Omega}_0^{-1}$ *is uniformly bounded in both row and column sums.*

These assumptions are commonly used in the (spatial) panel data literature. In particular, Assumption 3.1 is standard for (dynamic) linear panel data models with large n and small T where Y_{n0} is exogenous. The first part of Assumption 3.2 is also rather standard in random-effects panel data models. What is not that common is the part that refers to the bivariate random vector $(\upsilon_{i\mu}, \upsilon_{i\alpha})$, which is justified by the existence of spatial spillovers in the individual effects of our model.

As for the next three assumptions, they are widely used in spatial econometrics models. In particular, Assumption 3.3 is a necessary condition for Assumptions 3.6 and 3.7 that can be found in e.g. Lee (2004) and Su and Yang (2015). It is always satisfied if $\{h_n\}$ is a bounded sequence and essentially allows the weight matrices to be rather "general", "cover[ing] spatial weights matrices where elements are not restricted to be nonnegative and those that might

[9] We say that a $k \times m$ matrix A (or a sequence of matrices A_n) is bounded in both row and column sums if there exists a constant $c < \infty$ such that $\max_j \sum_{i=1}^{k} A_{ij} < c$ and $\max_i \sum_{j=1}^{m} A_{ij} < c$.

not be row-normalized" (Lee, 2004, p. 1903). Assumptions 3.4 and 3.5 can be found in e.g. Lee (2004) and parallel Assumptions 3 and 5 of Yu et al. (2008). In particular, Assumption 3.5 was first employed by Kelejian and Prucha (1998, 2001). While Assumption 3.4 guarantees that $\mathbf{Y}$ can be expressed exclusively in terms of the exogenous variables, Assumption 3.5 essentially limits the spatial correlation. Notice also that Assumption 3.4 holds if $\lambda_0 \in \left(\frac{1}{\omega_{\min}}, \frac{1}{\omega_{\max}}\right)$, where $\omega_{\min}$ denotes the smallest and $\omega_{\max}$ denotes the largest characteristic root of the spatial weight matrix W_n $(\omega_{\min} < 0, \omega_{\max} > 0)$.

The last three assumptions have also been previously used to derive the asymptotic properties of a QML estimator in spatial econometrics models for cross-section and panel data (Lee, 2004; Su and Yang, 2015). Firstly, Assumption 3.6 basically provides conditions for the global identification of the estimator. More precisely, the first part is the identification uniqueness condition (White, 1994), while the second part guarantees that the regressors are not asymptotically multicollinear. In particular, in the second part of the assumption we can alternatively assume that $\frac{1}{nT}E(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})$ is positive definite for sufficiently large n. This is a softer condition that only requires some additional proof to be applied. Secondly, Assumption 3.7 guarantees the existence and positive definiteness of the Hessian and the variance covariance matrix of the score vector. It thus plays a basic role in the asymptotic normality results. Thirdly, Assumption 3.8 is necessary for the Central Limit Theorem we use to derive the asymptotic normality of the estimator (Kelejian and Prucha, 2001). In particular, it can be shown that this assumption also holds if $(I_n + T\Sigma(\sigma_0))^{-1}$ is uniformly bounded in both row and column sums.

Theorem 3.1. *Under assumptions 3.1 to 3.6, ψ_0 is globally identified and $\widehat{\psi}$ is a consistent estimator of ψ_0 with $\widehat{\psi} \xrightarrow{p} \psi_0$.*

Theorem 3.2. *Under assumptions 3.1 to 3.8, $\sqrt{nT}\left(\widehat{\psi} - \psi_0\right) \xrightarrow{d} N\left(0, \boldsymbol{H}^{-1}\mathcal{G}\boldsymbol{H}^{-1}\right)$.*

Remark 3.1. *Lee (2004), Yu et al. (2008) and Su and Yang (2015) use analog theorems to prove the consistency and asymptotic normality of their QML estimator in cross-section (Lee, 2004) and panel data (Yu et al., 2008; Su and Yang, 2015) models. In particular, Theorems* 3.1 *and* 3.2 *are similar to Theorems 3.1 and 3.2 of Lee (2004), Theorems 4 and 5 of Yu et al. (2008), and Theorems 4.1 and 4.2 of Su and Yang (2015), respectively. Because of the panel structure, our results are obviously closer to those of Yu et al. (2008), who analyse a spatial dynamic panel data model with fixed effects and no spatial contagion in the error term (and large T and n), and those of Su and Yang (2015), who analyse a dynamic panel data with spatially autocorrelated errors and both fixed and random effects (with small T and large n, as we do). This means that, on the one hand, our set of regressors is similar to that of Yu et al. (2008), except that we do not have the spatial lag of the lagged dependent variable and they do not have the spatially weighted exogenous variables (Su and Yang (2015) do not consider either spatially weighted regressors or the spatial lag of the – lagged – dependent* *variable). But, on the other hand, our error structure does have local spatial contagion, as Su and Yang's does (2015), although ours is in the individual-specific effects and theirs is in the idiosyncratic term (which in turn results in a variance-covariance matrix different from the ones assumed by these papers). Thus, our model specification is different, and so is the variance-covariance matrix, but the approach and the proof of our theorems largely follows their work (see Appendices* 3.6 *and* 3.7 *for details). In particular, the fact that our model specification includes the spatial lag of the endogenous variable makes the proof more involved than that of Su and Yang (2015). On the other hand, the scope of our proof is limited by the fact that we do not cover cases where* Y_{n0} *is endogenous, as they do.*

3.4 Empirical application

In this section we provide empirical evidence on a growth-initial level equation (see e.g. Islam 1995 and Elhorst et al. 2010) using the correlated random effects specification of the spatial Durbin dynamic panel model presented in this paper. The principal aim of this empirical exercise is to show that *i)* we can (largely) replicate the results obtained by Lee and Yu (2016) using a standard spatial dynamic Durbin model (our benchmark); and *ii)* our model specification not only provides an estimate of the individual-specific effects but also of their spatial spillovers.

To this end, we use the data and (basic) model specification of Lee and Yu (2016). The dataset covers 28 OECD countries (see Ho et al. 2013 for details) over the period 1970 to 2005 (in time intervals of 5 years). The dependent variable, Y_{nt}, is the real GDP per capita (units of labour). As for the explanatory variables, $N_{nt} + 0.05$ is the sum of the annual average working-age population growth over the last 5 years (N_{nt}) and an approximation to the sum of the exogenous technical progress rate and the capital depreciation rate (see e.g. Ertur and Koch 2007 for details); S_{nt} is the average investment share in GDP; and $Y_{n,t-1}$ is the real GDP per capita lagged 5 years.

[Insert Table 3.1 about here]

The first column in Table 3.1 reports the results obtained by Lee and Yu (2016) using a weighting matrix W_n defined by the geographical distance between the capital of the countries. Notice that W_n is a row-normalized matrix with zeros in the diagonal. The second column provides the estimates of our model.[10] The parameter ρ measures the effect of the

[10] Estimates were obtained using the optimizing routines of R and the log-likelihood function in 3.3.2.

time-lagged real GDP ($Y_{n,t-1}$) on the dependent variable, whereas λ measures the intensity of its contemporaneous spatial interactions ($W_n Y_{nt}$). Also, the β-parameters measure the effect of the exogenous regressors (β_1 is the coefficient associated with $N_{nt} + 0.05$ and β_2 is the coefficient associated with S_{nt}), whereas the γ-parameters measure the intensity of the spatial contagion between the OECD countries arising from these exogenous regressors (γ_1 and γ_2 are the counterparts of β_1 and β_2). Lastly, the π-parameters are the coefficients associated with the variables included in the correlation functions. In particular, the π_μ-parameters correspond to those employed for the individual effects (π_{μ_1} is the coefficient of the mean of $N_{nt} + 0.05$ and π_{μ_2} is that of the mean of S_{nt}) and the π_α-parameters to those employed for their spatial spillovers (π_{α_1} is the coefficient of the spatially weighted mean of $N_{nt} + 0.05$ and π_{α_2} is that of the spatially weighted mean of S_{nt}).

The first thing to notice is that our results largely concur with those of Lee and Yu (2016). This means that in both cases the coefficients of the working-age population growth rate (β_1) are negative and statistically significant at standard confidence levels, while the coefficients of the savings rate (β_2) are positive and statistically significant. Notice also that while the parameter associated with the time lagged real GDP is positive and statistically significant, the intensity of the contemporaneous spatial interactions of Y_{nt} is not statistically significant. This stands in contrast to the findings of Ertur and Koch (2007) and Elhorst et al. (2010).

It is also worth noting that only the coefficients associated with $N_{nt} + 0.05$ are – weakly – statistically significant in the correlations functions (the $p-$value of π_{μ_1} is 0.14, slightly above the standard 0.10).[11] This contrasts with the clear statistical significance of π_{α_1} (and the joint test for the π_α parameters), which supports the existence of spatial spillovers in the individual effects. However, the estimated variances indicate that the individual effects and their spatial counterparts do not have a significant random component. All in all, these results seem to be consistent with an error term specification analogous to the one proposed by Debarsy (2012).

Thus, if we interpret the estimated individual effects as a proxy for the unobserved productivity of the countries (see Islam 1995), our results suggest that the growth of some countries may be – weakly – related not only to their unobserved productivity, but also to the impact that the unobserved productivity of other countries have on their economies.[12]

[11] We also computed Wald tests for the joint significance of the coefficients in each correlation function. Results show that while the variables included in π_μ are not jointly significant (the p-value was 0.21), the variables included in π_α rejected the null hypothesis (the p-value was 0.01).

[12] Notice that, given the lack of statistical significance of σ_α, our results may also be consistent with the hypothesis (see Debarsy, 2012) that the growth of one country is linked to its unobserved productivity and this, in turn, is related to the (mean) characteristics of the other countries (but not to their unobserved productivities).

More generally, our results point to the importance of unobserved country-specific intrinsic features (economic, social, historical, etc.) in growth.

In order to further explore this idea and following the discussion in Section 3.2, we computed the direct and indirect global and local effects. However, since the λ coefficient is not statistically significant, the global and local effects coincide: the global effects are only of a local nature (Anselin, 2003). Thus, we interpret our results as local effects and, since the weight matrix is defined in terms of geographical distances, closer neighbours will have greater weight than distant neighbours in the indirect effects. In particular, we report the local direct effects for each period in Table 3.2 and the "spill-in" and "spill-out" effects of the estimated individual effects for each period in Tables 3.3 and 3.4, respectively.

The first column in Table 3.2 is the direct local effect in period one, which can be interpreted as the impact on the dependent variable (the log of real GDP per capita) of the estimated individual effects. In other words, these figures provide, for each country, an estimate of the difference in the log of real GDP per capita of having or not the unobserved heterogeneity term (i.e., having a zero value individual effect). As a caveat, notice that, given the weak statistical significance of the π_μ-parameters, these direct effects may not be statistically different than zero.

With this in mind, results indicate the existence of three groups of countries in our sample: those with a large individual effect, with values above the third quartile (Canada, Chile, Israel, Mexico, Netherlands, New Zealand, Turkey and the US); those with a small individual effect, with values below the first quartile (Austria, Belgium, Denmark, Finland, Greece, Italy, Japan, Korea, Norway, Portugal and Switzerland); and those with an intermediate individual effect (Australia, France, Iceland, Ireland, Spain, Sweden and the UK). It is also interesting to note that, for most countries, our ranking does not substantially differ from that of Islam (1995). However, in order to make meaningful comparisons, in the last two columns of Table 3.2 we report his estimated individual effects (obtained from a model without spatial interactions and for a sample of 192 countries over the period 1965 to 1985) and our equivalent estimate, $\hat{\mu} + W_n\hat{\alpha}$. We can see then that fifteen out of the 25 countries commonly analysed barely changed their ranking (Austria, Chile, Denmark, France, Greece, Israel, Italy, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland and the UK) and that, in fact, the most important differences arise from seven countries that dramatically changed their position in the rankings (Japan and Belgium, from the top of his ranking to the bottom of ours, and Finland, Ireland, Korea, Mexico and Turkey, the other way round).

As for spill-in effects reported in Table 3.3, for each 5-year period the columns report the (local) impact on the log of real GDP per capita of each country associated with the unobserved characteristics of the other countries. The most affected countries (above the third quartile) are Austria, Finland, France, Ireland, Italy, Korea, Netherlands, Norway, Sweden and Switzerland, whereas the least affected countries (below the first quartile) are Australia, Canada, Chile, Iceland, Israel, Japan, Mexico, New Zealand, and the US. Notice that most of the countries with a small/large individual effect are among the most/least affected by their neighbours (in terms of geographical distance). Also, as expected figures in the other columns of the table show that, to a large extent, these groups remain stable over time.

Lastly, the columns in Table 3.4 contain, for each 5-year period, the estimated (local) impact on the log of real GDP per capita of the neighbouring countries associated with the unobserved characteristics of each country. However, rather than reporting the spill-out effect as described in Section 3.2, we simply report the estimated $\sum_{s=1}^{t}\rho^{s-1}\alpha_n$, which provides essentially the same picture.[13] Results show that the countries that impact least on their neighbours are Canada, Chile, Iceland, Israel, Korea, Mexico and New Zealand, whereas the countries that impact most on their neighbours are Austria, Belgium, Denmark, Finland, France, Italy, Japan, Sweden, Switzerland and the UK. Notice that countries that impact least/most on other countries tend to be those that are less/more affected by the spillovers from their neighbours (and generally have a larger/smaller individual effect). That is, there is a negative correlation between the estimated individual effects and the estimated spill-in (on average, -0.4) and spill-out (on average, -0.7) effects. Notice also that, as expected, these results largely hold for the seven periods considered.

3.5 Conclusions

In this paper we consider a correlated random effects specification of the spatial Durbin dynamic panel model. We derive the likelihood function of the model and prove the consistency and asymptotic normality of the QML estimator under rather standard assumptions in the spatial econometrics literature. A major difference with respect to previous studies is that our model specification includes individual effects and their spatial spillovers.

Obtaining an estimate of the individual-specific effects (net of the spatially weighted

[13]In particular, following Miranda et al. (2017b, p. 4) we may interpret $\sum_{s=1}^{t}\rho^{s-1}\alpha_n$ "as the "potential" of the spatial spillovers of the individual effects" in each period (i.e., "a measure of the "potentiality of the spatial contagion" associated with the individual effect of [each] unit" in each period).

effects) can be critical in certain applications, such as growth models in which a measure of the unobserved productivity of the geographical units under study can be obtained from the estimated individual effects and hence the existence of spatial spillovers in (unobserved) productivity can be analysed. We illustrate this point by estimating a growth-initial level equation using OECD data and providing evidence of spatial contagion in the individual effects.

Our results point to the importance of unobserved country-specific characteristics and their spatial spillovers in growth. In particular, we find that countries with a small/large estimated individual effect tend to be among the most/least affected by the impact of the estimated individual effects of their neighbours and among those whose individual effects impact most/least on the other countries (in terms of geographical distance). This means that, if we interpret the individual effect as a proxy for the unobserved productivity, more/less productive economies are less/more interrelated with the other economies. According to our estimates, examples of countries that fit into the first pattern include Chile, Israel, Mexico and New Zealand, whereas examples of countries that fit into the second pattern include Austria, Finland, Italy and Switzerland.

Table 3.1: **QML estimates**

Variable	Parameters	Lee and Yu (2016)	Our model
WY_t	λ	-0.040	-0.011
		(0.045)	(0.020)
Y_{t-1}	ρ	0.889^{***}	0.919^{***}
		(0.046)	(0.049)
$N_t + 0.05$	β_1	-0.198^{***}	-0.200^{***}
		(0.04)	(0.042)
S_t	β_2	0.143^{***}	0.141^{***}
		(0.047)	(0.048)
$W(N_t + 0.05)$	γ_1	0.102^{**}	0.108^{**}
		(0.047)	(0.048)
WS_t	γ_2	0.003	-0.001
		(0.057)	(0.057)
$\overline{N_t + 0.05}$	π_{μ_1}		0.115
			(0.079)
$(\overline{S_t})$	π_{μ_2}		-0.061
			(0.057)
$W(\overline{N_t + 0.05})$	π_{α_1}		-0.284^{***}
			(0.091)
$W\overline{S_t}$	π_{α_2}		-0.004
			(0.065)

Variance Components

σ_μ^2	σ_α^2	$\sigma_{\mu\alpha}$	σ_ε^2
0.0001	0.0000	0.0001	0.004^{***}
(0.0002)	(0.0002)	(0.0002)	(0.0005)

Note: *p-value<0.1; **p-value<0.05; ***p-value<0.01. We denote the time-mean of a variable with an upper bar.

Table 3.2: **Local Direct Effects**

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	Islam (1995)	$\hat{\mu} + W_n\hat{\alpha}$
Australia	-0.40	-0.77	-1.11	-1.42	-1.70	-1.97	-2.21	1.69	0.35
Austria	-0.42	-0.81	-1.16	-1.49	-1.79	-2.07	-2.32	1.72	0.38
Belgium	-0.42	-0.82	-1.17	-1.50	-1.81	-2.09	-2.34	1.75	0.35
Canada	-0.39	-0.74	-1.07	-1.37	-1.64	-1.90	-2.13	1.81	0.37
Chile	-0.38	-0.73	-1.05	-1.35	-1.62	-1.87	-2.10	1.49	0.34
Denmark	-0.42	-0.80	-1.15	-1.48	-1.77	-2.05	-2.30	1.74	0.38
Finland	-0.43	-0.82	-1.19	-1.52	-1.83	-2.11	-2.37	1.66	0.39
France	-0.41	-0.79	-1.14	-1.46	-1.75	-2.02	-2.27	1.75	0.39
Greece	-0.42	-0.80	-1.16	-1.48	-1.78	-2.06	-2.31	1.60	0.35
Iceland	-0.41	-0.78	-1.13	-1.45	-1.74	-2.01	-2.26	–	0.36
Ireland	-0.41	-0.78	-1.12	-1.44	-1.73	-1.99	-2.24	1.60	0.40
Israel	-0.39	-0.76	-1.09	-1.40	-1.68	-1.94	-2.18	1.70	0.37
Italy	-0.43	-0.82	-1.18	-1.52	-1.82	-2.10	-2.36	1.69	0.37
Japan	-0.44	-0.84	-1.22	-1.56	-1.87	-2.16	-2.43	1.75	0.29
Korea	-0.42	-0.80	-1.16	-1.48	-1.78	-2.05	-2.31	1.60	0.39
Mexico	-0.38	-0.73	-1.05	-1.35	-1.62	-1.87	-2.10	1.65	0.38
Netherlands	-0.39	-0.75	-1.08	-1.39	-1.67	-1.92	-2.16	1.73	0.41
New Zealand	-0.39	-0.75	-1.08	-1.38	-1.66	-1.91	-2.15	1.69	0.37
Norway	-0.42	-0.81	-1.16	-1.49	-1.79	-2.06	-2.32	1.77	0.39
Portugal	-0.42	-0.81	-1.17	-1.50	-1.81	-2.09	-2.34	1.58	0.35
Spain	-0.41	-0.79	-1.14	-1.46	-1.76	-2.03	-2.28	1.75	0.38
Sweden	-0.41	-0.79	-1.14	-1.46	-1.75	-2.02	-2.27	1.73	0.39
Switzerland	-0.43	-0.82	-1.18	-1.52	-1.82	-2.10	-2.36	1.70	0.37
Turkey	-0.38	-0.73	-1.05	-1.35	-1.62	-1.87	-2.10	1.53	0.39
United Kingdom	-0.40	-0.77	-1.11	-1.42	-1.70	-1.96	-2.21	1.73	0.39
United States	-0.39	-0.75	-1.09	-1.39	-1.67	-1.93	-2.17	1.80	0.36

The last two columns provide the estimated individual effects reported by Islam (1995) and our equivalent estimate, $\hat{\mu} + W_n\hat{\alpha}$.

Table 3.3: **Spill-in Effects**

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
Australia	0.75	1.45	2.09	2.67	3.21	3.71	4.16
Austria	0.80	1.53	2.21	2.83	3.40	3.93	4.41
Belgium	0.78	1.50	2.15	2.76	3.32	3.83	4.30
Canada	0.76	1.46	2.10	2.69	3.24	3.74	4.19
Chile	0.72	1.38	1.99	2.55	3.06	3.54	3.97
Denmark	0.79	1.52	2.19	2.81	3.37	3.89	4.37
Finland	0.82	1.57	2.26	2.89	3.47	4.01	4.50
France	0.80	1.54	2.22	2.84	3.41	3.94	4.42
Greece	0.77	1.47	2.12	2.72	3.26	3.77	4.23
Iceland	0.76	1.47	2.11	2.71	3.26	3.76	4.22
Ireland	0.81	1.55	2.24	2.87	3.45	3.98	4.47
Israel	0.76	1.46	2.11	2.70	3.25	3.75	4.21
Italy	0.80	1.53	2.20	2.82	3.39	3.91	4.40
Japan	0.73	1.40	2.01	2.58	3.10	3.58	4.02
Korea	0.81	1.55	2.23	2.85	3.43	3.96	4.44
Mexico	0.76	1.46	2.10	2.69	3.24	3.74	4.19
Netherlands	0.80	1.54	2.22	2.85	3.43	3.95	4.44
New Zealand	0.76	1.45	2.09	2.68	3.22	3.72	4.17
Norway	0.81	1.56	2.25	2.89	3.47	4.00	4.50
Portugal	0.78	1.49	2.15	2.75	3.30	3.81	4.28
Spain	0.79	1.52	2.19	2.80	3.36	3.88	4.36
Sweden	0.80	1.54	2.22	2.84	3.41	3.94	4.42
Switzerland	0.80	1.53	2.21	2.83	3.40	3.92	4.40
Turkey	0.77	1.48	2.14	2.74	3.29	3.80	4.26
United Kingdom	0.79	1.51	2.18	2.79	3.36	3.87	4.35
United States	0.75	1.44	2.07	2.65	3.19	3.68	4.13

Table 3.4: **Spill-out Effects**

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
Australia	0.76	1.45	2.09	2.68	3.22	3.72	4.17
Austria	0.80	1.53	2.21	2.83	3.40	3.92	4.41
Belgium	0.80	1.54	2.22	2.85	3.42	3.95	4.44
Canada	0.75	1.44	2.07	2.65	3.19	3.68	4.13
Chile	0.73	1.41	2.02	2.59	3.12	3.60	4.04
Denmark	0.81	1.56	2.24	2.88	3.46	3.99	4.48
Finland	0.82	1.57	2.25	2.89	3.47	4.01	4.50
France	0.80	1.53	2.21	2.83	3.40	3.92	4.41
Greece	0.79	1.52	2.19	2.80	3.37	3.89	4.37
Iceland	0.75	1.43	2.07	2.65	3.18	3.67	4.12
Ireland	0.76	1.46	2.11	2.70	3.24	3.75	4.21
Israel	0.72	1.38	1.98	2.54	3.05	3.52	3.95
Italy	0.82	1.57	2.26	2.90	3.49	4.02	4.52
Japan	0.81	1.55	2.23	2.85	3.43	3.96	4.44
Korea	0.73	1.40	2.01	2.58	3.10	3.58	4.02
Mexico	0.72	1.38	1.99	2.55	3.06	3.54	3.97
Netherlands	0.76	1.47	2.11	2.70	3.25	3.75	4.21
New Zealand	0.75	1.45	2.09	2.67	3.21	3.71	4.16
Norway	0.78	1.50	2.16	2.77	3.33	3.84	4.31
Portugal	0.79	1.52	2.18	2.80	3.36	3.88	4.36
Spain	0.78	1.49	2.15	2.75	3.30	3.81	4.28
Sweden	0.82	1.57	2.26	2.89	3.48	4.01	4.51
Switzerland	0.80	1.53	2.20	2.82	3.39	3.91	4.39
Turkey	0.76	1.46	2.11	2.70	3.24	3.74	4.20
United Kingdom	0.81	1.56	2.25	2.88	3.46	3.99	4.48
United States	0.76	1.46	2.10	2.69	3.24	3.74	4.19

3.6 Appendix A: Lemmas

In this section we make extensive use of the following notation: $tr(A)$ denotes the trace of matrix A, $\tau_{\max}(A)$ the largest eigenvalue of matrix A, $\tau_{\min}(A)$ the smallest eigenvalue of matrix A, and $\|A\|_m$ the $m-$norm of matrix A with $m = 1, 2, \infty$ and F ($m = F$ being the Frobenius norm). Further, we use the term u.b.r.c.s. to refer to a matrix or sequence of matrices “uniformly bounded in both row and column sums".

We also make use of the following representation of the model in 3.2.3 and 3.2.4 (obtained by repeated substitution):

$$Y_{nt} = \rho_0^t S_0^{-t} Y_{n,0} + \sum_{j=0}^{t-1} \rho_0^j S_0^{-(j+1)} \left(\mathbb{X}_{n,t-j}\phi_0 + \upsilon_{n\mu} + W_n \upsilon_{n\alpha} + \varepsilon_{n,t-j}\right)$$

where $\phi_0 = (c_0, \beta'_{10}, \beta'_{20}, \pi'_{\mu 0}, \pi'_{\alpha 0})'$, $\mathbb{X}_{nt} = \left(l_n \;\vdots\; X_{nt} \;\vdots\; W_n X_{nt} \;\vdots\; \overline{X}_n \;\vdots\; W_n \overline{X}_n \right)$ is an $n \times (4K + 1)$ matrix, and the other elements are defined in Section 3.2. In full matrix notation:

$$\mathbf{Y} = \mathbf{G}_0 Y_{n,0} + \mathbf{C}_0 \mathbb{X}\phi_0 + \mathbf{L}_0 \left(\upsilon_{n\mu} + W_n \upsilon_{n\alpha}\right) + \mathbf{C}_0 \boldsymbol{\varepsilon} \tag{3.6.1}$$

with $\mathbf{L}_0 = \mathbf{C}_0(l_T \otimes I_n)$.

Lastly, some of the lemmas make use of the following property:

Property 3.1. *Let $D^{-1}(\sigma)$ be an $r \times r$ symmetric matrix, with $\sigma \in \Delta$ being a $p \times 1$ vector of parameters and Δ a compact parametric space. Then, there exists a matrix $A_k(\sigma, \overline{\sigma})$ such that*

i) $D^{-1}(\sigma) - D^{-1}(\overline{\sigma}) = \sum_{k=1}^{p} (\sigma_k - \overline{\sigma}_k) A_k(\sigma, \overline{\sigma})$ *for all* $\sigma, \overline{\sigma} \in \Delta$

ii) $\sup_{\sigma \in \Delta} \tau_{\max}\left(D^{-2}(\sigma)\right) \le c_\tau < \infty$

iii) $\sup_{\sigma, \overline{\sigma} \in \Delta} \tau_{\max}\left(A_k(\sigma, \overline{\sigma}) A'_k(\sigma, \overline{\sigma})\right) \le c_\tau < \infty$ *for* $k = 1, \ldots, p$

Lemma 3.1. *Let A be a real symmetric $n \times n$ matrix and B a random $n \times m$ matrix. Then,*

$$\tau_{\min}\left(E\left(B'AB\right)\right) \ge \tau_{\min}\left(A\right) \tau_{\min}\left(E\left(B'B\right)\right)$$

Proof. By definition, $\tau_{\min}\left(E\left(B'AB\right)\right) = \min_{z \in R^m} \left\{z' E\left(B'AB\right) z \mid z'z = 1\right\}$. Let $\overline{z}$ be such that $\tau_{\min}\left(E\left(B'AB\right)\right) = \overline{z}' E\left(B'AB\right) \overline{z}$. Let D_A be the diagonal matrix of eigenvalues of A. Since

A is a real symmetric matrix, there exists Q such that $A = QDQ'$ and $QQ' = I_n$. Then,

$$\begin{aligned}
\tau_{\min}\left(E\left(B'AB\right)\right) &= E\left(\overline{z}'B'QD_AQ'B\overline{z}\right) \\
&\geq \tau_{\min}(A)E\left(\overline{z}'B'QQ'B\overline{z}\right) \\
&\geq \tau_{\min}(A)\min_{z\in R^m}\left\{E\left(z'B'Bz\right)\middle|\, z'z=1\right\} \\
&\geq \tau_{\min}(A)\tau_{\min}\left(E\left(B'B\right)\right)
\end{aligned}$$

□

Lemma 3.2. *Let A be a real positive semidefinite $n\times n$ matrix and B a real symmetric $n\times n$ matrix. Then,*

$$tr\left(AB\right) \leq \tau_{\max}\left(B\right) tr\left(A\right).$$

Proof. Since B is a real symmetric matrix, it can be diagonalized. Let P_B be the orthogonal matrix with the eigenvectors of B ($P_BP_B' = I_n$) and let D_B be the diagonal matrix of eigenvalues of B such that $B = P_BD_BP_B'$. Then,

$$tr\left(AB\right) = tr\left(AP_BD_BP_B'\right) = tr\left(P_B'AP_BD_B\right) = tr\left(CD_B\right)$$

where C is a symmetric positive semidefinite matrix (given that A is a positive semidefinite matrix and $y'P_B'AP_By = x'Ax \geq 0$). Using that $tr\left(C\right) = tr\left(A\right)$ and given that $c_{ii} \geq 0$ for $i = 1, ..., n$ (because of the positive definitiveness of C),

$$tr\left(AB\right) = tr\left(CD_B\right) = \sum_{i=1}^{n} c_{ii}\tau_i\left(B\right) \leq \tau_{\max}\left(B\right)\sum_{i=1}^{n}\left|c_{ii}\right| = \tau_{\max}\left(B\right)tr\left(C\right) = \tau_{\max}\left(B\right)tr\left(A\right)$$

□

Lemma 3.3. *Under assumptions 3.1 to 3.6, $E\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] = 0$.*

Proof. We start by noting that, given that $\widetilde{\mathbf{X}} = \left[\begin{array}{c:c:c:c:c:c} \boldsymbol{l}_{nT} & \mathbf{Y}_{-1} & \mathbf{X} & \mathbf{WX} & \overline{\mathbf{X}} & \mathbf{W}\overline{\mathbf{X}}\end{array}\right]$, we only need to prove that $E\left[\mathbf{Y}_{-1}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] = 0$, since $E\left[\mathbf{Z}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] = 0$ for $\mathbf{Z} = \boldsymbol{l}_{nT}, \mathbf{X}, \mathbf{WX}, \overline{\mathbf{X}}$ and $\mathbf{W}\overline{\mathbf{X}}$ by the strict exogeneity of $\mathbf{X}$. Notice also that, by using equation 3.6.1, we have that

$$\mathbf{Y}_{-1} = \mathbf{G}_0^{-}Y_{n,0} + \mathbf{C}_0^{-}\mathbb{X}_{-1}\phi_0 + \mathbf{L}_0^{-}\left(\upsilon_{n\mu} + W_n\upsilon_{n\alpha}\right) + \mathbf{C}_0^{-}\boldsymbol{\varepsilon}, \tag{3.6.2}$$

with $\mathbb{X}_{-1} = \left(0, \mathbb{X}_{n1}', \cdots, \mathbb{X}_{n,T-1}'\right)'$, $\mathbf{G}_0^{-} = \left(I_n, \rho_0 S_0^{-1\prime}, \ldots, \rho_0^{T-2}S_0^{-(T-1)\prime}\right)'$, $\mathbf{L}_0^{-} = \mathbf{C}_0^{-}\left(l_T\otimes I_n\right)$ and

$$\mathbf{C}_0^- = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ S_0^{-1} & 0 & 0 & \cdots & 0 \\ \rho_0 S_0^{-2} & S_0^{-1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_0^{T-2} S_0^{-(T-1)} & \rho_0^{T-3} S_0^{-(T-2)} & \rho_0^{T-4} S_0^{-(T-3)} & \cdots & 0 \end{pmatrix}$$

Thus,

$$\mathbf{Y}_{-1}' \mathbf{\Omega}_0^{-1} \boldsymbol{\eta} = Y_{n0}' \mathbf{G}_0^- \mathbf{\Omega}_0^{-1} \boldsymbol{\eta} + \phi' \mathbb{X}_{-1}' \mathbf{C}_0^{-\prime} \mathbf{\Omega}_0^{-1} \boldsymbol{\eta} + \varepsilon' \mathbf{C}_0^{-\prime} \mathbf{\Omega}_0^{-1} \boldsymbol{\eta} + \left(\upsilon_{n\mu}' + \upsilon_{n\alpha}' W_n' \right) \mathbf{L}_0^{-\prime} \mathbf{\Omega}_0^{-1} \boldsymbol{\eta}$$

First, it is easy to show that $E\left(Y_{n0}' \mathbf{G}_0^- \mathbf{\Omega}_0^{-1} \boldsymbol{\eta}\right) = 0 = E\left(\phi' \mathbb{X}_{-1}' \mathbf{C}_0^{-\prime} \mathbf{\Omega}_0^{-1} \boldsymbol{\eta}\right)$. Second, notice that we can write $E\left(\boldsymbol{\varepsilon}' \mathbf{C}_0^{-\prime} \mathbf{\Omega}_0^{-1} \boldsymbol{\eta}\right) = \sigma_\varepsilon^2 tr\left(\mathbf{\Omega}_0^{-1} \mathbf{C}_0^{-\prime}\right)$ and, given that $\mathbf{L}_0^{-\prime} = \left(l_T' \otimes I_n\right) \boldsymbol{C}_0^{-\prime}$, $J_T = l_T l_T'$ and $E\left[\left(\upsilon_\mu + W_n \upsilon_{n\alpha}\right)\left(\upsilon_\mu + W_n \upsilon_{n\alpha}\right)'\right] = \sigma_{\varepsilon_0}^2 \Sigma_0$,

$$E\left(\left(\upsilon_{n\mu}' + \upsilon_{n\alpha}' W_n'\right) \mathbf{L}_0^{-\prime} \mathbf{\Omega}_0^{-1} \boldsymbol{\eta}\right) = \sigma_{\varepsilon 0}^2 tr\left[\mathbf{\Omega}_0^{-1} \left(J_T \otimes \Sigma_0\right) \mathbf{C}_0^{-\prime}\right]$$

Also, following Magnus (1982), we can rewrite $\mathbf{\Omega}_0^{-1}$ as $\mathbf{\Omega}_0^{-1} = \left(I_T \otimes I_n\right) - \frac{1}{T} J_T \otimes \left[I_n - \left(I_n + T\Sigma_0\right)^{-1}\right]$, which means that

$$\sigma_{\varepsilon_0}^2 tr\left(\mathbf{\Omega}_0^{-1} \mathbf{C}_0^{-\prime}\right) + \sigma_{\varepsilon_0}^2 tr\left[\mathbf{\Omega}_0^{-1} \left(J_T \otimes \Sigma_0\right) \mathbf{C}_0^{-\prime}\right] = \sigma_{\varepsilon_0}^2 tr\left[\mathbf{C}_0^{-\prime}\right] + \sigma_{\varepsilon_0}^2 tr\left[\mathbf{A} \mathbf{C}_0^{-\prime}\right] = 0$$

since

$$\begin{aligned} \mathbf{A} &= -\frac{1}{T}\left(J_T \otimes \left[I_n - \left(I_n + T\Sigma_0\right)^{-1}\right]\right) + \frac{1}{T}\left(J_T \otimes T\Sigma_0\right) \\ &\quad - \left(\frac{1}{T} J_T \otimes \left[I_n - \left(I_n + T\Sigma_0\right)^{-1}\right]\right)\left(J_T \otimes \Sigma_0\right) \\ &= \frac{1}{T} J_T \otimes \left[-I_n + \left(I_n + T\Sigma_0\right)^{-1}\left(I_n + T\Sigma_0\right)\right] = 0 \end{aligned}$$

and $tr\left[\mathbf{C}_0'\right] = 0$ because of the structure of $\mathbf{C}_0'$. □

Lemma 3.4. *Let A, B and C be real constant matrices of order $(n \times r)$, $(r \times r)$ and $(r \times n)$ respectively, with A and C u.b.r.c.s. and B being a symmetric matrix with $\tau_{\max}(B^2) < \infty$. Then, for $Q = ABC$:*

i) $tr\left(QQ'\right) = O(\min(r, n))$

ii) $l_n' QQ' l_n = O\left(n\right)$, *where l_n is a unit vector of dimension $n \times 1$*

iii) $\sum_{i=1}^{n} Q_{ii}^2 = O(min(r, n))$ *and* $tr\left(QQ\right) = O(n)$.

Proof. Firstly, by the Cauchy-Schwarz inequality and Lemma 3.2,

$$\begin{aligned} tr\left(QQ'\right) &= tr\left(ABCC'BA'\right) = tr\left(BCC'BA'A\right) \\ &\leq \left[tr\left(BCC'CC'B\right)\right]^{1/2}\left[tr\left(BAA'AA'B\right)\right]^{1/2} \\ &\leq \tau_{\max}(B^2)\left[tr\left(C'CC'C\right)\right]^{1/2}\left[tr\left(A'AA'A\right)\right]^{1/2} \end{aligned}$$

Then, by using the second part of Lemma B.1 in Su and Yang (2015), we can show that $\tau_{\max}(B^2)\left[tr\left(C'CC'C\right)\right]^{1/2}\left[tr\left(A'AA'A\right)\right]^{1/2} = O(\min(r,n))$.

Secondly,

$$\begin{aligned} l_n'QQ'l_n &= tr\left(l_n'ABCC'BA'l_n\right) = tr\left(BCC'BA'l_nl_n'A\right) \\ &\leq \tau_{\max}\left(BCC'B\right)tr\left(A'l_nl_n'A\right) \leq \tau_{\max}\left(BCC'B\right)tr\left(l_n'AA'l_n\right) = O\left(n\right), \end{aligned}$$

where the last equality holds because

- given that C is u.b.r.c.s., $\|C'\|_2^2 \leq \|C'\|_1^2\|C'\|_\infty^2 \leq c^2$, with $\max_i \sum_{j=1}^n |c_{ij}| \leq c$, $\max_j \sum_{i=1}^n |c_{ij}| \leq c$ and $c < \infty$, and $\|B\|_2^2 = \tau_{\max}\left(B^2\right)$; then, since $\|.\|_2$ is a sub-multiplicative norm[14], $\tau_{\max}\left(BCC'B\right) = \|C'B\|_2^2 \leq \|C'\|_2^2\|B\|_2^2 \leq c^2\tau_{\max}\left(B^2\right)$,
- given that A is u.b.r.c.s., $\max_i \sum_{j=1}^n |a_{ij}| \leq a$, $\max_j \sum_{i=1}^n |a_{ij}| \leq a$ and $a < \infty$; then, $l_n'AA'l_n \leq a^2 l_n'l_n \leq a^2 n$.

Thirdly, $\sum_{i=1}^n Q_{ii}^2 \leq \sum_{i=1}^n\sum_{j=1}^n |Q_{ij}|^2 = \|Q\|_F^2 \leq tr\left(Q'Q\right)$, which, because of result i), is $O(\min(r,n))$. Also using result i), $tr(QQ) \leq tr(QQ')^{1/2}tr(QQ')^{1/2} = tr(QQ') = O(\min(r,n))$. □

Lemma 3.5. *Let $a = (a_1,...,a_n)'$ and $b = (b_1,...,b_n)'$. Also, let $\{(a_i,b_i)\}_{i=1}^n$ be an i.i.d. sequence of random vector variables with $E(a_i) = E(b_i) = 0$ and finite second moments. Lastly, let P be an $n \times n$ constant matrix and let $\Omega = E(ab') = \mu_{ab}I_n$ such that $(a'Pb - tr(P\Omega)) = tr\left(Pba' - P\Omega\right) = \sum_{i=1}^n\sum_{j=1}^n P_{ij}(a_ib_j - \Omega_{ij})$. Then,*

$$E\left[\left(a'Pb - tr(P\Omega)\right)^2\right] = (\sigma_{ab}^2 - \sigma_a^2\sigma_b^2 - \mu_{ab}^2)\sum_{i=1}^n P_{ii}^2 + \sigma_a^2\sigma_b^2 tr\left(PP'\right) + \sigma_{ab}^2 tr\left(PP\right)$$

[14] This means that, for any two matrices A and B, $\|AB\|_2 \leq \|A\|_2\|B\|_2$.

where $\sigma_a^2 = E(a_i^2)$, $\sigma_b^2 = E(b_i^2)$, $E(a_i b_i) = \mu_{ab}$ *and* $E[(a_i b_i - \mu_{ab})^2] = \sigma_{ab}^2$. *Notice that, if* a *and* b *are independent,* $E\left[(a'Pb - tr(P\Omega))^2\right] = \sigma_a^2\sigma_b^2 tr(PP')$. *Notice also that if* $a = b$*, then* $E\left[(a'Pb - tr(P\Omega))^2\right] = (\sigma_a^{(4)} - 2\sigma_a^4)\sum_{i=1}^{n} P_{ii}^2 + \sigma_a^4 tr\left(PP'\right) + \sigma_a^{(4)} tr\left(PP\right)$ *, with* $\sigma_a^{(4)} = E[(a_i^2 - \sigma_a^2)^2]$.

Proof. Notice that $E\left[(a'Pb - tr(P\Omega))^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} P_{ij}P_{kl}E\left[(a_i b_j - \Omega_{ij})(a_k b_l - \Omega_{kl})\right]$. Also, given the independence of (a_i, b_i) and (a_j, b_j) for $i \neq j$, $E\left[(a_i b_j - \Omega_{ij})(a_k b_l - \Omega_{kl})\right] \neq 0$ only for $i = j = k = l$, $i = k \neq j = l$ and $i = l \neq j = k$. Thus,

$$
\begin{aligned}
E\left[(a'Pb - tr(P\Omega))^2\right] &= \sum_{i=1}^{n} P_{ii}^2 E\left[(a_i b_i - \mu_{ab})^2\right] + \sum_{i=1}^{n}\sum_{j\neq i}^{n} P_{ij}^2 E\left[(a_i b_j)^2\right] \\
&+ \sum_{i=1}^{n}\sum_{j\neq i}^{n} P_{ij}P_{ji}E\left[(a_i b_i)(a_j b_j)\right] \\
&= (\sigma_{ab}^2 - \sigma_a^2\sigma_b^2 - \mu_{ab}^2)\sum_{i=1}^{n} P_{ii}^2 + \left(\sigma_a^2\sigma_b^2 tr(PP') + \mu_{ab}^2 tr(PP)\right)
\end{aligned}
$$

□

Lemma 3.6. *Let* $a = (a_1, ..., a_n)'$ *and* $b = (b_1, ..., b_n)'$*, with* $\{(a_i, b_i)\}_{i=1}^{n}$ *i.i.d. sequences of random vector variables with finite second moments. Let* P_n *and* Q_n *be* $n \times r$ *constant matrices u.b.r.c.s.. Lastly, let* $D(\sigma)$ *be an* $r \times r$ *constant symmetric matrix that satisfies Property* 3.1*, with* $\sigma \in \Delta$ *being a* $p \times 1$ *vector of parameters. Then,*

$$
\sup_{\sigma\in\Delta}\left|E\left(a'P_n D^{-1}(\sigma)Q_n b\right)\right| = O(n)
$$

Note that the Lemma still holds if $a = b$ *and* $P_n = Q_n$.

Proof. By the Cauchy-Schwarz inequality and Lemma 3.2,

$$
\begin{aligned}
\sup_{\sigma\in\Delta}\left|E\left(a'P_n D^{-1}(\sigma)Q_n b\right)\right| &\leq \sup_{\sigma\in\Delta} E\left(tr\left(a'P_n D^{-2}(\sigma)P_n' a\right)^{1/2} tr\left(b'Q_n'(\sigma)Q_n b\right)^{1/2}\right) \\
&\leq \left[\sup_{\sigma\in\Delta}\tau_{\max}\left(D^{-2}(\sigma)\right)\right]^{1/2}\left[E\left(tr\left(P_n' aa'P_n\right)^{1/2} tr\left(Q_n bb'Q_n'\right)^{1/2}\right)\right] \\
&\leq \left[\sup_{\sigma\in\Delta}\tau_{\max}\left(D^{-2}(\sigma)\right)\right]^{1/2}\tau_{\max}^{1/2}\left(P_n P_n'\right)\tau_{\max}^{1/2}\left(Q_n' Q_n\right)\left[E\left(tr\left(aa'\right)\right)E\left(tr\left(bb'\right)\right)\right]^{1/2} \\
&\leq C\left[E\left(tr\left(aa'\right)\right)E\left(tr\left(bb'\right)\right)\right]^{1/2}
\end{aligned}
$$

with $C < \infty$ given that $D^{-1}(\sigma)$ satisfies Property 3.1, $\tau_{\max}^{1/2}(A'A) = \|A\|_2 \leq (\|A\|_1 \|A\|_\infty)^{1/2}$, and P_n and Q_n are u.b.r.c.s.. Also, $E(tr(aa')) = E\left[\sum_{i=1}^{n} a_i^2\right] \leq nE(a_i^2)$. Thus, given that a and b have finite second moments, the lemma is proved. □

Lemma 3.7. *Let $a = (a_1, ..., a_n)'$ and $b = (b_1, ..., b_n)'$, with $\{(a_i, b_i)\}_{i=1}^n$ i.i.d. sequences of random vector variables with finite second moments. Let P_n and Q_n be $n \times r$ constant matrices u.b.r.c.s.. Lastly, let $D(\sigma)$ be an $r \times r$ constant symmetric matrix that satisfies Property 3.1, with $\sigma \in \Delta$ being a $p \times 1$ vector of parameters. Then,*

$$\frac{1}{\max(n,r)}\left\{a'P_nD^{-1}(\sigma)Q_nb - E\left[a'P_nD^{-1}(\sigma)Q_nb\right]\right\} \xrightarrow{p} 0 \text{ uniformly in } \sigma \in \Delta$$

Note that the Lemma still holds if $a = b$.

Proof. Let us denote $E(a_i) = \mu_a, E(b_j) = \mu_b$, $E\left[(a_i - \mu_a)^2\right] = \sigma_a^2$, $E\left[(b_j - \mu_b)^2\right] = \sigma_b^2$ for all i and j. We start by proving that

$$\frac{1}{\max(n,r)}\left[\mu_a l_n' P_n D^{-1}(\sigma) Q_n (b - \mu_b)\right] \xrightarrow{p} 0 \text{ uniformly in } \sigma \in \Delta$$

To prove the uniform convergence (see e.g. Theorem 21.9 of Davidson 1994), we prove that $l_n'P_nD^{-1}(\sigma)Q_n(b - \mu_b)$ is stochastically equicontinuous and, for a given σ, satisfies a Law of Large Numbers (LLN hereafter). First we prove the convergence for a given σ. Given that $E\left[l_n'P_nD^{-1}(\sigma)Q_n(b - \mu_b)\right] = 0$, to derive a LLN it is enough to prove that

$$\frac{1}{\max(n,r)^2} Var\left[l_n'P_nD^{-1}(\sigma)Q_n(b - \mu_b)\right] \longrightarrow 0$$

It is straightforward to prove that $Var\left[l_n'P_nD^{-1}(\sigma)Q_n(b - \mu_b)\right] = \sigma_b^2 l_n'P_nD^{-1}(\sigma)Q_nQ_n'D^{-1}(\sigma)P_n'l_n$ and, by Lemma 3.4, $l_n'P_nD^{-1}(\sigma)Q_nQ_n'D^{-1}(\sigma)P_nl_n = O(n)$, so that

$$\begin{aligned}\frac{1}{\max(n,r)^2} Var\left[l_n'P_nD^{-1}(\sigma)Q_n(b - \mu_b)\right] &\leq \frac{1}{\max(n,r)^2} l_n'P_nD^{-1}(\sigma)Q_nQ_n'D^{-1}(\sigma)P_nl_n \\ &\leq \frac{O(n)}{\max(n,r)^2} = o(1)\end{aligned}$$

which proves the LLN. To prove the stochastic equicontinuity, note that, by Property 3.1, the Cauchy-Schwarz inequality and Lemma 3.2,

$$\begin{aligned}
&\left|l_n' P_n D^{-1}(\sigma) Q_n \left(b-\mu_b\right) - l_n' P_n D^{-1}(\overline{\sigma}) Q_n \left(b-\mu_b\right)\right| \leq \left|l_n' P_n \left(D^{-1}(\sigma) - D^{-1}(\overline{\sigma})\right) Q_n \left(b-\mu_b\right)\right| \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| tr^{1/2} \left(l_n' P_n A_k(\sigma,\overline{\sigma}) A_k'(\sigma,\overline{\sigma}) P_n' l_n\right) tr^{1/2} \left(\left(b-\mu_b\right)' Q_n' Q_n \left(b-\mu_b\right)\right) \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| \tau_{\max} \left(A_k(\sigma,\overline{\sigma}) A_k'(\sigma,\overline{\sigma})\right) tr^{1/2} \left(l_n' P_n P_n' l_n\right) tr^{1/2} \left(\left(b-\mu_b\right)' Q_n' Q_n \left(b-\mu_b\right)\right) \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| c_\tau tr^{1/2} \left(l_n' P_n P_n' l_n\right) | tr^{1/2} \left(\left(b-\mu_b\right)' Q_n' Q_n \left(b-\mu_b\right)\right)
\end{aligned}$$

with $c_\tau < \infty$. Also, by Lemma 3.4, $tr\left(l_n' P_n P_n' l_n\right) = O(n)$ and, by Lemma 3.6, $tr\left(\left(b-\mu_b\right)' Q_n' Q_n \left(b-\mu_b\right)\right) = O_p(n)$, so we can apply Theorem 21.10 of Davidson (1994) to prove the stochastic equicontinuity and Theorem 21.9 of Davidson (1994) to prove the uniform convergence.

Next we prove the case $E(a_i) = E(b_i) = 0$. We first prove the convergence in probability given σ. To this end, notice that $E\left\{a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right\} = 0$ and, from Lemmas 3.4 and 3.5, $E\left\{\left(a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right)^2\right\} = O(n)$, so that

$$\lim_{n\to\infty} \frac{1}{\max(n,r)^2} E\left\{\left(a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right)^2\right\} = 0$$

which proves the convergence given σ. To prove the stochastic equicontinuity, note that, by Property 3.1, the Cauchy-Schwarz inequality and Lemma 3.2,

$$\begin{aligned}
\left|a' P_n D^{-1}(\sigma) Q_n b - a' P_n D^{-1}(\overline{\sigma}) Q_n b\right| &\leq \left|a' P_n \left(D^{-1}(\sigma) - D^{-1}(\overline{\sigma})\right) Q_n b\right| \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| \left|a' P_n A_k(\sigma,\overline{\sigma}) Q_n b\right| \\
&\leq \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k| \, tr^{1/2} \left(a' P_n A_k(\sigma,\overline{\sigma}) A_k'(\sigma,\overline{\sigma}) P_n' a\right) tr^{1/2} \left(b' Q_n' Q_n b\right) \\
&\leq c_\tau tr^{1/2} \left(a' P_n P_n' a\right) tr^{1/2} \left(b' Q_n' Q_n b\right) \sum_{k=1}^{p} |\sigma_k - \overline{\sigma}_k|
\end{aligned}$$

Also, by Lemma 3.6, $E\left[tr\left(a' P_n P_n' a\right)\right] = O(n)$ and $E\left[tr\left(b' Q_n' Q_n b\right)\right] = O(n)$. Then,

$$\frac{1}{\max(n,r)} \left|a' P_n D^{-1}(\sigma) Q_n b - a' P_n D^{-1}(\overline{\sigma}) Q_n b\right| = O_p(1)$$
$$\frac{1}{\max(n,r)} \left|E\left(a' P_n D^{-1}(\sigma) Q_n b\right) - E\left(a' P_n D^{-1}(\overline{\sigma}) Q_n b\right)\right| = O(1)$$

and we can apply Theorems 21.9 and 21.10 of Davidson (1994) to prove uniform convergence. Further, the most general case $E(a_i) \neq 0$ and $E(b_i) \neq 0$ follows straightforward by noting that

$$\left\{a' P_n D^{-1}(\sigma) Q_n b - E\left[a' P_n D^{-1}(\sigma) Q_n b\right]\right\} = a^{*\prime} P_n D^{-1}(\sigma) Q_n b^* - E\left[a^{*\prime} P_n D^{-1}(\sigma) Q_n b^*\right] + a^{*\prime} P_n D^{-1}(\sigma) Q_n E(b) + E(a)' P_n D^{-1}(\sigma) Q_n b^*$$

with $a^* = a - \mu_a$ and $b^* = b - \mu_b$. □

Lemma 3.8. *Let $G_{nt} = \rho_0^t S_0^{-t}$, $C_{nt} = G_{nt} S_0^{-1}$ and $L_{nt} = \sum_{j=0}^{t-1} \rho_0^j S_0^{-(j+1)}$. Under Assumption 3.5, $W_n L_{nt}$, $W_n G_{nt}$ and $W_n C_{nt}$ are all u.b.r.c.s. for $t = 1, 2, ..., T$ and $\mathbf{WL}_0$, $\mathbf{WG}_0$ and $\mathbf{WC}_0$ are all u.b.r.c.s..*

Proof. First note that if A and B are two matrices u.b.r.c.s., $A + B$ and AB are also u.b.r.c.s. (see Remark A2 in Kapoor et al. 2007). With this result, under Assumption 3.5 it is easy to prove that G_{nt}, C_{nt} and L_{nt} are u.b.r.c.s.. Further, given that $T < \infty$, it is easy to prove that $\mathbf{WL}_0$, $\mathbf{WG}_0$ and $\mathbf{WC}_0$ are all u.b.r.c.s.. □

Lemma 3.9. *Let $\mathbf{\Omega}(\sigma) = (I_T \otimes I_n) + (J_T \otimes \Sigma(\sigma))$ and $\Sigma(\sigma) = \sum_{k=1}^{3} \sigma_k \Sigma_k = \sigma_1 I_n + \sigma_2(W_n + W_n') + \sigma_3 W_n W_n'$, with W_n u.b.r.c.s. and $(\sigma_1, \sigma_2, \sigma_3) \in \Delta$, being Δ a compact space such that $\Sigma(\sigma)$ is positive semidefinite for any $\sigma \in \Delta$. Then, $\Omega^{-1}(\sigma)$ satisfies Property 3.1 for $A_k(\sigma, \overline{\sigma}) = \Omega^{-1}(\sigma)(J_T \otimes \Sigma_k)\Omega^{-1}(\overline{\sigma})$ and any $\sigma, \overline{\sigma} \in \Delta$. Moreover, $\exists\ c_\tau < \infty$ such that $\sup_{\sigma \in \Delta} \tau_{\max}(\mathbf{\Omega}(\sigma)) < c_\tau$.*

Proof. We start by proving that $\exists\ c_\tau < \infty$ such that $\sup_{\sigma \in \Delta} \tau_{\max}(\mathbf{\Omega}(\sigma)) < c_\tau$. To this end, note that the eigenvalues of the matrix $(I_n + B)$ are $1 + \tau_i(B)$, with $\tau_i(B)$ being the $i = 1, ..., n$ eigenvalues of B. Then, by definition, $\sup_{\sigma \in \Delta} \tau_{\max}(\mathbf{\Omega}(\sigma)) = 1 + \sup_{\sigma \in \Delta} \tau_{\max}((J_T \otimes \Sigma(\sigma))) = 1 + T \sup_{\sigma \in \Delta} \tau_{\max}(\Sigma(\sigma))$. Further, using that $\Sigma(\sigma)$ is a symmetric positive semidefinite matrix, $\sup_{\sigma \in \Delta} \tau_{\max}(\Sigma(\sigma)) = \sup_{\sigma \in \Delta} \|\Sigma(\sigma)\|_2$. Then,

$$\sup_{\sigma \in \Delta} \|\Sigma(\sigma)\|_2 \leq \sum_{k=1}^{3} \sup_{\sigma \in \Delta} |\sigma_k| \|\Sigma_k\|_2 \leq \sup_{\sigma \in \Delta} |\sigma_1| + \sup_{\sigma \in \Delta} |\sigma_2| \|W_n + W_n'\|_2 + \sup_{\sigma \in \Delta} |\sigma_3| \|W_n W_n'\|_2$$

Given that W_n is u.b.r.c.s., $W_n + W_n'$ and $W_n W_n'$ are u.b.r.c.s., too (see Remark A2 in Kapoor et al. 2007). Further, $(\|W_n + W_n'\|_1 \|W_n + W_n'\|_\infty) < \infty$ and $(\|W_n W_n'\|_1 \|W_n W_n'\|_\infty) < \infty$. Then, $\|W_n W_n'\|_2 \leq (\|W_n W_n'\|_1 \|W_n W_n'\|_\infty)^{1/2} < \infty$ and $\|W_n + W_n'\|_2 \leq \|W_n\|_2 + \|W_n'\|_2 \leq 2 (\|W_n\|_1 \|W_n\|_\infty)^{1/2} < \infty$. Finally, given that $\sigma \in \Delta$ and Δ is compact, $\sup_k \sup_{\sigma \in \Delta} \sigma_k < \infty$. Then, $\exists\ c < \infty$ such that $\sup_{\sigma \in \Delta} \tau_{\max}(\Sigma(\sigma)) < c$ and $\sup_{\sigma \in \Delta} \tau_{\max}(\mathbf{\Omega}(\sigma)) < 1 + Tc < \infty$.

Next we prove that $\Omega^{-1}(\sigma)$ satisfies Property 3.1 for $A_k(\sigma,\overline{\sigma}) = \Omega^{-1}(\sigma)(J_T \otimes \Sigma_k)\Omega(\overline{\sigma})$ and any $\sigma, \overline{\sigma} \in \Delta$. To this end, we need to prove that: *i)* $\mathbf{\Omega}^{-1}(\sigma) - \mathbf{\Omega}^{-1}(\overline{\sigma}) = \sum_{k=1}^{3}(\sigma_k - \overline{\sigma}_k)A_k(\sigma,\overline{\sigma})$; *ii)* $\sup_{\sigma\in\Delta} \tau_{\max}(\mathbf{\Omega}^{-2}(\sigma)) < c_\tau < \infty$ and *iii)* $\sup_{\sigma,\overline{\sigma}\in\Delta} \tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) < c_\tau < \infty$ for $k = 1, 2, 3$.

To prove *i)*, note that

$$\begin{aligned}
\mathbf{\Omega}^{-1}(\sigma) - \mathbf{\Omega}^{-1}(\overline{\sigma}) &= \mathbf{\Omega}^{-1}(\sigma)\left[\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}(\sigma)\right]\mathbf{\Omega}^{-1}(\overline{\sigma}) \\
&= \mathbf{\Omega}^{-1}(\sigma)\left[J_T \otimes (\Sigma(\overline{\sigma}) - \Sigma(\sigma))\right]\mathbf{\Omega}^{-1}(\overline{\sigma}) \\
&= \sum_{k=1}^{3}(\overline{\sigma}_k - \sigma_k)\mathbf{\Omega}^{-1}(\sigma)\left[J_T \otimes \Sigma_k\right]\mathbf{\Omega}^{-1}(\overline{\sigma}) = \sum_{k=1}^{3}(\sigma_k - \overline{\sigma}_k)A_k(\sigma,\overline{\sigma}).
\end{aligned}$$

To prove *ii)*, note that, given that $\Sigma(\sigma)$ is a positive semidefinite matrix for all $\sigma \in \Delta$, $\inf_{\sigma\in\Delta} \tau_{\min}(\Sigma(\sigma)) \geq 0$. Then, using that $\mathbf{\Omega}^{-1}(\sigma)$ is positive semidefinite for all $\sigma \in \Delta$ (since all the eigenvalues of J_T and $\Sigma(\sigma)$ are equal to or bigger than zero for all $\sigma \in \Delta$, all the eigenvalues of $\mathbf{\Omega}(\sigma)$ are bigger or equal than 1), $\sup_{\sigma\in\Delta} \tau_{\max}\left(\mathbf{\Omega}^{-2}(\sigma)\right) = \sup_{\sigma\in\Delta}\left[\tau_{\max}\left(\mathbf{\Omega}^{-1}(\sigma)\right)\right]^2 = \left[\inf_{\sigma\in\Delta} \tau_{\min}\left(\mathbf{\Omega}(\sigma)\right)\right]^{-2} \leq \left[1 + T\inf_{\sigma\in\Delta} \tau_{\min}\left(\Sigma(\sigma)\right)\right]^{-2} \leq 1$.

To prove *iii)*, note that $\|.\|_2$ is a sub-multiplicative norm (see footnote 14). Thus,

$$\begin{aligned}
\sup_{\sigma,\overline{\sigma}\in\Delta} \tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) &= \sup_{\sigma,\overline{\sigma}\in\Delta}\left\|\mathbf{\Omega}^{-1}(\sigma)\left(J_T \otimes \Sigma_k\right)\mathbf{\Omega}^{-1}(\overline{\sigma})\right\|_2^2 \\
&\leq \sup_{\sigma\in\Delta}\left\|\mathbf{\Omega}^{-1}(\sigma)\right\|_2^4 \left\|(J_T \otimes \Sigma_k)\right\|_2^2 \\
&\leq \sup_{\sigma\in\Delta}\left[\tau_{\max}\left(\mathbf{\Omega}^{-1}(\sigma)\right)\right]^4 \tau_{\max}\left(J_T \otimes \Sigma_k\right) \\
&\leq \sup_{\sigma\in\Delta}\left[\tau_{\max}\left(\mathbf{\Omega}^{-1}(\sigma)\right)\right]^4 T\tau_{\max}\left(\Sigma_k\right) < c_\tau < \infty
\end{aligned}$$

using *ii)* and $\sup_k \tau_{\max}(\Sigma_k) < c_\tau < \infty$ (from the first part of the proof). □

Lemma 3.10. *Let* $B_n^{-1}(\sigma) = I_n - (I_n + T\Sigma(\sigma))^{-1}$ *and* $\Sigma(\sigma) = \sum_{k=1}^{3}\sigma_k\Sigma_k = \sigma_1 I_n + \sigma_2(W_n + W_n') + \sigma_2 W_n W_n'$*, with* W_n *u.b.r.c.s. and* $(\sigma_1, \sigma_2, \sigma_3) \in \Delta$*, being* Δ *a compact space such that* $\Sigma(\sigma)$ *is positive semidefinite for any* $\sigma \in \Delta$*. Then,* $B_n^{-1}(\sigma)$ *satisfies Property* 3.1 *for* $A_k(\sigma,\overline{\sigma}) = TB_n^*(\overline{\sigma})\left(J_T \otimes \Sigma_k\right)B_n^*(\sigma)$ *and any* $\sigma, \overline{\sigma} \in \Delta$ *with* $B_n^*(\sigma) = (I_n + T\Sigma(\sigma))^{-1}$.

Proof. To prove that $B_n^{-1}(\sigma)$ satisfies Property 3.1 for $A_k(\sigma,\overline{\sigma}) = TB_n^*(\overline{\sigma})\left(J_T \otimes \Sigma_k\right)B_n^*(\sigma)$, we need to prove that: *i)* $B_n(\sigma) - B_n(\overline{\sigma}) = \sum_{k=1}^{3}(\sigma_k - \overline{\sigma}_k)A_k(\sigma,\overline{\sigma})$; *ii)* $\sup_{\sigma\in\Delta} \tau_{\max}(B_n^{-2}(\sigma)) < c_\tau < \infty$ and *iii)* $\sup_{\sigma,\overline{\sigma}\in\Delta} \tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) < c_\tau < \infty$ for $k = 1, 2, 3$.

To prove *i)*, note that

$$\begin{aligned} B_n^{-1}(\sigma) - B_n^{-1}(\overline{\sigma}) &= (I_n + T\Sigma(\overline{\sigma}))^{-1} - (I_n + T\Sigma(\sigma))^{-1} \\ &= T\,(I_n + T\Sigma(\overline{\sigma}))^{-1}\,(\Sigma(\sigma) - \Sigma(\overline{\sigma}))\,(I_n + T\Sigma(\sigma))^{-1} \\ &= T\sum_{k=1}^{3}(\sigma_k - \overline{\sigma})\,(I_n + T\Sigma(\overline{\sigma}))^{-1}\,\Sigma_k\,(I_n + T\Sigma(\sigma))^{-1} \\ &= T\sum_{k=1}^{3}(\sigma_k - \overline{\sigma}) B_n^*(\overline{\sigma})^{-1}\Sigma_k B_n^*(\sigma) \end{aligned}$$

To prove *ii)*, note first that $B_n^{-1}(\sigma)$ is a positive semidefinite matrix for all $\sigma \in \Delta$, since $\inf_{\sigma\in\Delta} \tau_{\min}\left(I_n + T\Sigma(\sigma)\right) \geq 1$, $\sup_{\sigma\in\Delta} \tau_{\max}\left(I_n + T\Sigma(\sigma)\right)^{-1} \leq 1$, and $\inf_{\sigma\in\Delta} \tau_{\min}\left(B_n^{-1}(\sigma)\right) \geq 0$. Note also that $\sup_{\sigma\in\Delta} \tau_{\max}(B_n^{-2}(\sigma)) = \left[\sup_{\sigma\in\Delta} \tau_{\max}(B_n^{-1}(\sigma))\right]^2$, and, since $(I_n + T\Sigma(\sigma))^{-1}$ is a positive semidefinite matrix and $\inf_{\sigma\in\Delta} \tau_{\min}\left[(I_n + T\Sigma(\sigma))^{-1}\right] \geq 0$, then $\sup_{\sigma\in\Delta} \tau_{\max}(B_n^{-1}(\sigma)) \leq 1 - \inf_{\sigma\in\Delta} \tau_{\min}\left[(I_n + T\Sigma(\sigma))^{-1}\right] \leq 1$.

To prove *iii)*, note that, given that $\tau_{\max}\left(\Sigma_k\right) \leq c_\tau < \infty$ (proved in Lemma 3.9) and $\sup_{\sigma\in\Delta} \|B_n^*(\sigma)\|_2 = \sup_{\sigma\in\Delta} \left\|(I_n + T\Sigma(\overline{\sigma}))^{-1}\right\|_2 \leq \left[\inf_{\sigma\in\Delta} \tau_{\min}\left(I_n + T\Sigma(\overline{\sigma})\right)\right]^{-1} \leq 1$, then

$$\sup_{\sigma,\overline{\sigma}\in\Delta} \tau_{\max}(A_k(\sigma,\overline{\sigma})A_k'(\sigma,\overline{\sigma})) = \sup_{\sigma,\overline{\sigma}\in\Delta} \|A_k'(\sigma,\overline{\sigma})\|_2^2 \leq \sup_{\sigma\in\Delta} \|B_n^*(\sigma)\|_2^4 \sup_{\sigma,\overline{\sigma}\in\Delta} \|\Sigma_k\|_2^2 \leq c_\tau < \infty$$

□

Lemma 3.11. *Let* $\mathbf{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma) = \frac{1}{nT}\left\{\mathbf{a}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{b} - E\left[\mathbf{a}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{b}\right]\right\}$, *with* $\boldsymbol{a}, \boldsymbol{b} = \boldsymbol{\eta}, \mathbf{WY}$. *Under Assumptions* 3.1 *to* 3.6*,*

$$\mathbf{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma) \xrightarrow{p} 0 \text{ uniformly in } \sigma$$

Proof. We provide the proof for the most involved case, $\boldsymbol{a}, \boldsymbol{b} = \mathbf{WY}$. The proof of the other cases is similar. From expression 3.6.1 we have that

$$\begin{aligned} \mathbf{Y}'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY} &= Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{G}_0 Y_{n,0} + 2Y_{n,0}'\mathbf{G}_0'\mathbf{W}' + \mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 \\ &+ 2Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu} + W_n\upsilon_{n\alpha}\right) + 2Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} \\ &+ \phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 + 2\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu} + W_n\upsilon_{n\alpha}\right) \\ &+ 2\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} + \left(\upsilon_{n\mu}' + \upsilon_{n\alpha}'W_n'\right)\mathbf{L}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{\mu} + W_n\upsilon_{n\alpha}\right) \\ &+ 2\left(\upsilon_{n\mu}' + \upsilon_{n\alpha}'W_n'\right)\mathbf{L}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} \\ &+ \boldsymbol{\varepsilon}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}'\mathbf{C}_0\boldsymbol{\varepsilon} \end{aligned}$$

The proof of the Lemma follows from proving that each of the previous summands, minus its expected value, converge in probability to 0 uniformly in σ. Following Magnus (1982), we have that $\mathbf{\Omega}^{-1}(\sigma) = I_T \otimes I_n - \frac{1}{T} J_T \otimes B_n^{-1}(\sigma)$ with $B_n^{-1}(\sigma) = I_n - (I_n + T\Sigma(\sigma))^{-1}$. Also, let $G_{nt} = \rho_0^t S_0^{-t}$, $C_{nt} = G_{nt} S_0^{-1}$ and $L_{nt} = \sum_{j=0}^{t-1} \rho_0^j S_0^{-(j+1)}$ (see also Lemma 3.8). Thus, we may rewrite the summands in the previous expression as follows:

$$\begin{aligned}
Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 &= \sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{nt}'W_n'W_nC_{nt-j}\mathbb{X}_{n,j}\phi_0 \\
&\quad -\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{ns}'W_n'B_n^{-1}(\sigma)W_nC_{n,t-j}\mathbb{X}_{n,j}\phi_0 \\
Y_{n,0}'\mathbf{G}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} &= \sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{nt}'W_n'W_nC_{n,t-j}\varepsilon_{n,j} \\
&\quad -\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{t} Y_{n,0}'G_{ns}'W_n'B_n^{-1}(\sigma)W_nC_{n,t-j}\varepsilon_{n,j} \\
\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\mathbb{X}\phi_0 &= \sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{t} \phi_0'\mathbb{X}_{nj}'C_{n,t-j}'W_n'W_nC_{n,t-l}\mathbb{X}_{n,l}\phi_0 \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{s} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nC_{n,s-l}\mathbb{X}_{n,l}\phi_0 \\
\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{C}_0\boldsymbol{\varepsilon} &= \sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{t} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'W_nC_{n,t-l}\varepsilon_{n,l} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\sum_{l=1}^{s} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nC_{n,s-l}\varepsilon_{n,s-l} \\
\phi_0'\mathbb{X}'\mathbf{C}_0'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0(\upsilon_{n\mu} + W_n\upsilon_{n\alpha}) &= \sum_{t=1}^{T}\sum_{j=1}^{t} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}\upsilon_{n\mu} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}\upsilon_{n\mu} \\
&\quad +\sum_{t=1}^{T}\sum_{j=1}^{t}\phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}W_n\upsilon_{n\alpha} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t} \phi_0'\mathbb{X}_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}W_n\upsilon_{n\alpha}
\end{aligned}$$

And, finally,

$$\begin{aligned}
\boldsymbol{\varepsilon}'\mathbf{C}_0'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{W}\mathbf{L}_0\left(\upsilon_{n\mu}+W_n\upsilon_{n\alpha}\right) &= \sum_{t=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}\upsilon_{n\mu} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}\upsilon_{n\mu} \\
&+ \sum_{t=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'W_nL_{nt}W_n\upsilon_{n\alpha} \\
&- \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{t}\varepsilon_{n,j}'C_{n,t-j}'W_n'B_n^{-1}(\sigma)W_nL_{ns}W_n\upsilon_{n\alpha}
\end{aligned}$$

Notice that each of the summands in the previous expressions, minus its expected value, can be written as $\frac{1}{nT}\left[a'P_nD^{-1}(\sigma)Q_nb - E\left(a'P_nD^{-1}(\sigma)Q_nb\right)\right]$ with $a,b = Y_{n,0}, X_{n,j}\beta_{10}, X_{n,j}\beta_{20}, \overline{X}_n\pi_{\mu 0}, \overline{X}_n\pi_{\alpha 0}, \varepsilon_{nt}, \upsilon_{n\mu}, \upsilon_{n\alpha}$; $P_n, Q_n = W_nL_{nt}, W_nG_{nt}, W_nC_{nt}, W_nC_{nt}W_n, \mathbf{WL}_0, \mathbf{WG}_0, \mathbf{WC}_0$; and $D^{-1}(\sigma) = \mathbf{I}_{nT}, \boldsymbol{\Omega}^{-1}(\sigma), B_n^{-1}(\sigma)$. This means that, if we can apply Lemma 3.7, the Lemma is proved. To apply these lemmas, P_n and Q_n must be u.b.r.c.s. which is proved for all the cases in Lemma 3.8. Also, $D(\sigma)^{-1}$ must satisfy Property 3.1, which is proved in Lemma 3.9 for $\boldsymbol{\Omega}^{-1}(\sigma)$ and in Lemma 3.10 for $B_n^{-1}(\sigma)$. Lastly, a and b must be an i.i.d. sequence with finite second moments, which is guaranteed by Assumptions 3.1 and 3.2. $\square$

Lemma 3.12. *Let*

$$\boldsymbol{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma) = \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{a}}'(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}^{-1}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{a}}'(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma)\right]$$

with $\mathbf{a},\mathbf{b} = \boldsymbol{\eta}, \mathbf{WY}$ *and* $\mathbf{Q}_{\mathbf{A},\mathbf{B}}(\delta) = \frac{1}{nT}\mathbf{A}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{B}$. *Under Assumptions* 3.1 *to* 3.6*,*

$$\boldsymbol{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma)\xrightarrow{p}0 \text{ uniformly in } \sigma$$

Proof. The proof of this Lemma is similar to the proof of Lemma 3.11. Thus, we only provide the proof for the case $\mathbf{a} = \mathbf{b} = \mathbf{WY}$ (the others are similar). We start by decomposing $\boldsymbol{\Upsilon}_{\mathbf{WY},\mathbf{WY}}(\sigma)$:

$$\begin{aligned}
&\boldsymbol{\Upsilon}_{\mathbf{WY},\mathbf{WY}}(\sigma) = \\
&\quad \mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}'(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}^{-1}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}'(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right] \\
&\quad = \left\{\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}'(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}'(\sigma)\right]\right\}\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}^{-1}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \\
&\quad + E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}'(\sigma)\right]\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}^{-1}(\sigma)\left\{\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right] - \mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right\}\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \\
&\quad + E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}'(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\left\{\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\}
\end{aligned}$$

First we need to prove that $\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]$ and $\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]$ converges elementwise to 0 uniformly in σ. The proof follows the same steps as that of Lemma 3.11 (note that all the elements of $\tilde{\mathbf{X}}$ are in $\mathbf{WY}$, including, as shown in 3.3.1 and 3.6.1, $\mathbf{Y}_{-1}$), so it is not reproduced here. The elementwise convergence implies, by the Slutsky theorem, that $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ and $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ. Then, by using the properties of the matrix norm, $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]\right\|_2 \leq \left\|\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ and $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_2 \leq \left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_F = o_p(1)$ uniformly in σ.

Next we prove that $\sup_\sigma \left\|\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\right\|_2 = O(1)$ and $\sup_\sigma \left\|E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\|_2 = O(1)$. Let us first consider

$$
\begin{aligned}
\sup_\sigma \left\|\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\right\|_2 &= \sup_\sigma \tau_{\max}\left\{E\left(\frac{1}{nT}\tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)^{-1}\right\} \\
&= \left(\inf_\sigma \tau_{\min}\left\{\frac{1}{nT}E\left(\tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right\}\right)^{-1}
\end{aligned}
$$

Note that, since $\mathbf{\Omega}^{-1}(\sigma)$ is a symmetric definite positive matrix, we can apply Lemma 3.1 to obtain

$$
\begin{aligned}
\inf_\sigma \tau_{\min}\left\{\frac{1}{nT}E\left(\tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right\} &\geq \inf_\sigma \tau_{\min}\left\{\mathbf{\Omega}^{-1}(\sigma)\right\}\tau_{\min}\left\{\frac{1}{nT}E\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\right\} \\
&\geq \left[\sup_\sigma \tau_{\max}\left(\mathbf{\Omega}(\sigma)\right)\right]^{-1}\tau_{\min}\left\{\frac{1}{nT}E\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\right\}
\end{aligned}
$$

From Lemma 3.9, $\sup_\sigma \tau_{\max}\left(\mathbf{\Omega}(\sigma)\right) < c_\tau < \infty$ and, from Assumption 3.6, $\tau_{\min}\left\{\frac{1}{nT}E\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\right\} > 0$ for sufficiently large n. Then, $\sup_\sigma \left\|\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\right\|_2 < C < \infty$ $\Rightarrow \sup_\sigma \left\|\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\right\|_2 = O(1)$.

As for $\sup_\sigma \left\|E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\|_2$, notice that $\left\|E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\|_2 \leq \left\|E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\|_F$ and

$$
\begin{aligned}
\left\|E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\|_F &= tr\left[E\left(\frac{1}{nT}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)E\left(\frac{1}{nT}\tilde{\mathbf{X}}\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY}\right)\right]^{1/2} \\
&= \left(\sum_{k=1}^{K}\left[E\left(\frac{1}{nT}\tilde{\mathbf{X}}'_k\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY}\right)\right]^2\right)^{1/2},
\end{aligned}
$$

where the last expression is $O(1)$ uniformly in σ if $\sup_\sigma \left|E\left(\tilde{\mathbf{X}}'_k\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY}\right)\right| = O(n)$. To prove that $\sup_\sigma \left|E\left(\tilde{\mathbf{X}}'_k\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY}\right)\right| = O(n)$, we follow the same steps as in Lemma

3.11. Thus, we decompose the term $\tilde{\mathbf{X}}_k'\mathbf{\Omega}^{-1}(\sigma)\mathbf{WY}$ in a finite sum of terms that can be written as $a'P_nD^{-1}(\sigma)Q_nb$, with a, b, P_n, Q_n and $D^{-1}(\sigma)$ satisfying the conditions of Lemma 3.6. This provides the proof that $\sup_\sigma \left|E\left(a'P_nD^{-1}(\sigma)Q_nb\right)\right| = O(n)$ and so that of $\sup_\sigma \left|E\left(\tilde{\mathbf{X}}_k'\mathbf{\Omega}(\sigma)^{-1}\mathbf{WY}\right)\right| = O(n)$.

Moreover, given that $\left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - \left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right]\right\|_2 = o_p(1)$ uniformly in σ and $\sup_\sigma \left\|E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right)\right\|_2 = O(1)$, then $\sup_\sigma \left\|\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right\|_2 = O_p(1)$.

Finally, we need to prove that $\sup_\sigma \left\|\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right\|_2 = O_p(1)$. To this end, notice that

$$\begin{aligned}
\sup_\sigma \left\|\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right\|_2 &= \sup_\sigma \tau_{\max}\left(\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right) = \sup_\sigma \tau_{\max}\left(\left[\frac{1}{nT}\tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right]^{-1}\right) \\
&= \left[\inf_\sigma \tau_{\min}\left(\frac{1}{nT}\tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1} \\
&\leq \left[\inf_\sigma \tau_{\min}(\mathbf{\Omega}^{-1}(\sigma))\tau_{\min}\left(\frac{1}{nT}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\right]^{-1} \\
&\leq \sup_\sigma \tau_{\max}(\mathbf{\Omega}(\sigma))\left[\tau_{\min}\left(\frac{1}{nT}\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\right]^{-1},
\end{aligned}$$

which is $O_p(1)$ given that, by Lemma 3.9, $\sup_\sigma \tau_{\max}(\mathbf{\Omega}(\sigma)) < c_\tau < \infty$, and, by Assumption 3.6, $\tau_{min}\left(\frac{1}{nT}\tilde{X}'\tilde{X}\right) > 0$ almost surely for sufficiently large n.

Then,

$$\left\{\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\}\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) = o_p(1)O_p(1)O_p(1)$$

$$\begin{aligned}
&E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\left\{\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right] - \mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right\}\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) \\
&= O_p(1)o_p(1)O_p(1)
\end{aligned}$$

and

$$E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}\left\{\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma) - E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\right\} = O(1)O(1)o_p(1),$$

all the cases uniformly in σ. This proves that $\Upsilon_{\mathbf{WY},\mathbf{WY}}(\sigma) = o_p(1)$ uniformly in σ (and the proof is analogous for the rest of cases). □

Lemma 3.13. *Under Assumptions 3.1 to 3.7,*

$$\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'} - E\left(\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'}\right)\right] = o_p(1)$$

Proof. It can be proved, following the proof of Lemmas 3.11 and 3.12, that each element of the matrix $\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'} - E\left(\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'}\right)$ can be written as the finite sum of $\frac{1}{nT}\left[a'P_nD^{-1}(\sigma_0)Q_nb - E\left(a'P_nD^{-1}(\sigma_0)Q_nb\right)\right]$, with $a,b = Y_{n,0}$, $\mathbb{X}_{n,j}$, ε_{nt}, $\upsilon_{n\mu}$ and $\upsilon_{n\alpha}$; $P_n, Q_n = W_nL_{nt}$, W_nG_{nt}, W_nC_{nt}; and $D^{-1}(\sigma_0) = \boldsymbol{A}_i\left[\boldsymbol{B}_\kappa\boldsymbol{A}_j\boldsymbol{B}_\varrho + \boldsymbol{B}_\varrho\boldsymbol{A}_j\boldsymbol{B}_\kappa\right]\boldsymbol{A}_i$, $\boldsymbol{A}_1\boldsymbol{B}_\kappa\boldsymbol{A}_j\boldsymbol{B}_\varrho + \boldsymbol{B}_\varrho\boldsymbol{A}_j\boldsymbol{B}_\kappa\boldsymbol{A}_1$ for $i,j = 0,1$ and $\kappa,\varrho = 0,1,2,3$ with $\boldsymbol{A}_0 = \boldsymbol{B}_0 = I_n$, $\boldsymbol{A}_1 = B_n^{-1}(\sigma_0)$, $\boldsymbol{B}_\kappa = \Sigma_\kappa$ for $\kappa = 1,2,3$ and $B_n^{-1}(\sigma_0)$ and Σ_κ defined in Lemmas 3.9 and 3.10 (see Appendix 3.8 for details on the elements of $\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'}$). This means that if a, b, P_n, Q_n and $D^{-1}(\sigma_0)$ satisfy the conditions of Lemma 3.7 in all the cases, then $\frac{1}{nT}\left[a'P_nD(\sigma_0)Q_mb - E(a'P_nD(\sigma_0)Q_mb)\right] = o_p(1)$ in all the cases, which proves the Lemma. Notice also that we do not need to prove the uniform convergence because these second derivatives and their expectations are evaluated at the true parameters of the model. It is therefore enough to prove that $D^{-1}(\sigma_0)$ is a symmetric matrix with $\tau_{\max}\left(D^{-2}(\sigma_0)\right) < \infty$.

Firstly, Lemmas 3.11 and 3.12 show that all the possible cases of a, b, P_n and Q_n satisfy the conditions of Lemma 3.7. Secondly, given that $D^{-1}(\sigma_0)$ is by definition symmetric, $\|I_n\|_2 = 1$, $\max_\kappa\|\Sigma_\kappa\|_2 < c_\tau < \infty$ (the bound is provided by Lemma 3.9) and $\left\|B_n^{-1}(\sigma_0)\right\|_2 = \tau_{\max}\left(B_n^{-1}(\sigma_0)\right) < c_\tau < \infty$ (the bound is provided by Lemma 3.10),

$$\tau_{\max}\left(D^{-2}(\sigma_0)\right) = \left\|D^{-1}(\sigma_0)\right\|_2^2 \leq 2\max_i\|\boldsymbol{A}_i\|_2^6\max_\kappa\|\boldsymbol{B}_\kappa\|_2^4 \leq c_\tau < \infty.$$

□

Lemma 3.14. *Let $a_t = \{a_{i,t}\}_{i=1}^n, b_t = \{b_{i,t}\}_{i=1}^n$ be $n\times 1$ zero-mean random vectors independent in i. Let us also define $Q_n = \sum_{t=1}^T a_t'P_{t,n}b_t$ with $P_{t,n}$ $n\times n$ real matrices and $T<\infty$. Lastly, let us denote $\mu_{Q_n} = E(Q_n)$ and $s_{Q_n}^2 = E[(Q_n-\mu_{Q_n})^2]$. If $P_{t,n}$ for $t = 1,...,T$ are u.b.r.c.s. and $\{(a_t,b_t)\}_{t=1}^T$ has $4+\epsilon_1$ finite moments for some $\epsilon_1 > 0$ and $n^{-1}s_{Q_n}^2 \geq c > 0$, then*

$$\frac{Q_n - \mu_{Q_n}}{s_{Q_n}} \xrightarrow{d} N(0,1)$$

Proof. The proof of this Lemma follows the proof of Theorem 1 in Kelejian and Prucha (2001, p. 243). First note that, given the independence in i of a_t and b_t, $\mu_{Q_n} = \sum_{t=1}^T\sum_{i=1}^n P_{t,n}[i,i]E(a_{i,t}b_{i,t})$, where we use the somewhat abusive notation $P_{t,n}[i,j]$ to refer to the row i and column j element of the matrix $P_{t,n}$. Notice also that $Q_n - \mu_{Q_n} = \sum_{t=1}^T\sum_{i=1}^n Y_{i,t}$ with

$$Y_{i,t} = P_{t,n}[i,i]\left(a_{i,t}b_{i,t} - E(a_{i,t}b_{i,t})\right) + a_{i,t}\sum_{j=1}^{i-1}P_{t,n}[j,i]b_{j,t} + b_{i,t}\sum_{j=1}^{i-1}P_{t,n}[i,j]a_{j,t}$$

for $i = 1,2,...,n$.

Let us now consider the $\sigma-$fields $\digamma_{0,n} = \{\emptyset, \Omega\}$ and $\digamma_{i,n} = \sigma(a_i, b_i, a_{i-1}, b_{i-1}, ..., a_1, b_1,)$, with $a_i = \{a_{i,t}\}_{t=1}^T$, $b_i = \{b_{i,t}\}_{t=1}^T$ and $1 \le i \le n$. By construction, $\digamma_{i-1,n} \subset \digamma_{i,n}$ and $Y_{i,t}$ is $\digamma_{i,n}-$measurable. It can also be shown that $E(Y_{i,t}|\digamma_{i-1,n}) = 0$. Therefore, $\{Y_{i,t}, \digamma_{i,n}, 1 \le i \le n, n \ge 1\}$ forms a martingale difference array and so $s_{Q_n}^2 = \sum_{i=1}^{n}\left(\sum_{t=1}^{T} E(Y_{i,t}^2) + 2\sum_{t=2}^{T}\sum_{s=1}^{t-1} E(Y_{i,t}Y_{i,s})\right)$. Thus, the expression for the variance of Q_n follows from

$$E(Y_{i,t}Y_{i,s}) = P_{t,n}[i,i]P_{s,n}[i,i]\sigma_{c,t,s}^{(2)} + \sigma_{a,t,s}^2\sigma_{b,t,s}^2\sum_{j=1}^{i-1}(P_{t,n}[i,j]P_{s,n}[i,j] + P_{t,n}[j,i]P_{s,n}[j,i])$$
$$+\sigma_{c,t,s}\sigma_{c,s,t}\sum_{j=1}^{i-1}(P_{t,n}[i,j]P_{s,n}[j,i] + P_{s,n}[i,j]P_{t,n}[j,i]) \tag{3.6.3}$$

with $\sigma_{c,t,s} = E(a_{i,t}b_{i,s})$, $\sigma_{c,t,s}^{(2)} = E[(a_{i,t}b_{i,t} - \sigma_{c,t,t})(a_{i,s}b_{i,s} - \sigma_{c,s,s})]$, $\sigma_{a,t,s}^2 = E[a_{i,t}a_{i,s}]$ and $\sigma_{b,t,s}^2 = E[b_{i,t}b_{i,s}]$. Also, if we define $X_{i,t} = Y_{i,t}/s_{Q_n}$, then $\{X_{i,t}, \digamma_{i,n}, 1 \le i \le n, n \ge 1\}$ forms a martingale difference array.

In what follows we prove that

$$\frac{Q_n - \mu_{Q_n}}{s_{Q_n}} = \sum_{i=1}^{n}\sum_{t=1}^{T} X_{i,t} \xrightarrow{d} N(0,1)$$

by showing that $X_{i,n} = \sum_{t=1}^{T} X_{i,t}$ satisfies the remaining conditions of the Central Limit Theorem of Gänsler and Stute (1977, p. 365). In particular, we demonstrate that $X_{i,n}$ satisfies the condition:

$$\sum_{i=1}^{k_n} E\left\{E\left[|X_{i,n}|^{2+\delta}\middle|\digamma_{i-1,n}\right]\right\} \longrightarrow 0 \tag{3.6.4}$$

for some $\delta > 0$, which in turn is sufficient for

$$\sum_{i=1}^{k_n} E\left[|X_{i,n}|^2 \mathbf{1}(|X_{i,n} > \varepsilon|)\middle|\digamma_{i-1,n}\right] \xrightarrow{p} 0$$

for all $\varepsilon > 0$ and with $\mathbf{1}(\cdot)$ being an indicator function. Then we prove that $X_{i,n}$ satisfies

$$\sum_{i=1}^{k_n} E\left[X_{i,n}^2\middle|\digamma_{i-1,n}\right] \xrightarrow{p} 1 \tag{3.6.5}$$

Let us take $0 < \delta \le \epsilon_1/2$. We note that, under the maintained moment assumptions on $\{(a_t, b_t)\}_{t=1}^T$, there exists a finite constant, $C_e \ge 1$, such that $E\left(\left|a_{i,t}^{r_1}b_{i,t}^{r_2}a_{i,s}^{r_3}b_{i,s}^{r_4}\right|\right) \le C_e$

for $\sum_{l=1}^{4} r_l \leq 4 + 2\delta$, $r_l \geq 0$, $t = 1, 2, ..., T$ and $i = 1, 2, ..., n$. We further note that, under the maintained assumptions on the matrices $P_{t,n}$, there exists a finite constant, $C_m \geq 1$, such that $\sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|) < C_m$ for $t = 1, 2, ..., T$ and $i = 1, ..., n$. Lastly for $t, s = 1, 2, ..., T$, note that $\sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|)^r \leq C_m^r$ for $r \geq 1$ and

$$\sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|) (|P_{s,n}[k,j]| + |P_{s,n}[j,k]|) \leq \sum_{j=1}^{n} (|P_{t,n}[i,j]| + |P_{t,n}[j,i]|) \sum_{j=1}^{n} (|P_{s,n}[k,j]| + |P_{s,n}[j,k]|) \leq C_m^2$$

Let us now take $q = 2 + \delta$ and let $1/q + 1/p = 1$. We note that $\left|\sum_{t=1}^{T} Y_{i,t}\right|^q \leq T^q \sum_{t=1}^{T} |Y_{i,t}|^q$. Also, using the triangle and Hölder's inequalities, we have that

$$\begin{aligned}
|Y_{i,t}|^q &= \left| P_{t,n}[i,i] \left(a_{i,t}b_{i,t} - \sigma_{c,t,t})\right) + a_{i,t} \sum_{j=1}^{i-1} P_{t,n}[j,i] b_{j,t} + b_{i,t} \sum_{j=1}^{i-1} P_{t,n}[i,j] a_{j,t} \right|^q \\
&\leq 2^q \left| 1/2 P_{t,n}[i,i]^{1/p} P_{t,n}[i,i]^{1/q} \left(a_{i,t}b_{i,t} - \sigma_{c,t,t}\right) + a_{i,t} \sum_{j=1}^{i-1} P_{t,n}[j,i]^{1/p} P_{t,n}[j,i]^{1/q} b_{j,t} \right|^q \\
&+ 2^q \left| 1/2 P_{t,n}[i,i]^{1/p} P_{t,n}[i,i]^{1/q} \left(a_{i,t}b_{i,t} - \sigma_{c,t,t}\right) + b_{i,t} \sum_{j=1}^{i-1} P_{t,n}[i,j]^{1/p} P_{t,n}[i,j]^{1/q} a_{j,t} \right|^q \\
&\leq 2^q \left[\sum_{j=1}^{i} |P_{t,n}[j,i]| \right]^{q/p} \left| 2^{-q} |P_{t,n}[i,i]| \left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |a_{i,t}|^q \sum_{j=1}^{i-1} |P_{t,n}[j,i]| \, |b_{j,t}|^q \right|^{q/q} \\
&+ 2^q \left[\sum_{j=1}^{i} |P_{t,n}[j,i]| \right]^{q/p} \left| 2^{-q} |P_{t,n}[i,i]| \left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |b_{i,t}|^q \sum_{j=1}^{i-1} |P_{t,n}[i,j]| \, |a_{j,t}|^q \right|^{q/q} \\
&\leq 2^q C_m^{q/p} \left(2^{-q} |P_{t,n}[i,i]| \left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |a_{i,t}|^q \sum_{j=1}^{i-1} |P_{t,n}[j,i]| \, |b_{j,t}|^q \right) \\
&+ 2^q C_m^{q/p} \left(2^{-q} |P_{t,n}[i,i]| \left|a_{i,t}b_{i,t} - \sigma_{c,t,t}\right|^q + |b_{i,t}|^q \sum_{j=1}^{i-1} |P_{t,n}[i,j]| \, |a_{j,t}|^q \right).
\end{aligned}$$

Consequently,

$$\sum_{i=1}^{n} E\left\{ E\left[\left| Y1_{i,t} \right|^q \middle| F_{i-1,n} \right] \right\} \leq$$
$$\sum_{i=1}^{n} 2^q C_m^{q/p} \left(\left| P_{t,n}[i,i] \right| E\left[\left| a_{i,t} b_{i,t} - \sigma_{c,t,t} \right|^q \right] + E\left[\left| a_{i,t} \right|^q \right] \sum_{j=1}^{i-1} \left| P_{t,n}[j,i] \right| E\left[\left| b_{j,t} \right|^q \right] \right) +$$
$$\sum_{i=1}^{n} 2^q C_m^{q/p} \left(\left| P_{t,n}[i,i] \right| E\left[\left| a_{i,t} b_{i,t} - \sigma_{c,t,t} \right|^q \right] + E\left[\left| b_{i,t} \right|^q \right] \sum_{j=1}^{i-1} \left| P_{t,n}[i,j] \right| E\left[\left| a_{j,t} \right|^q \right] \right)$$
$$\leq \sum_{i=1}^{n} 2^q C_m^{q/p} C_e \left(2 \left| P_{t,n}[i,i] \right| + \sum_{j=1}^{i-1} \left| P_{t,n}[j,i] \right| + \sum_{j=1}^{i-1} \left| P_{t,n}[i,j] \right| \right) \leq n 2^{q+1} C_m^{q/p+1} C_e$$

Thus,

$$\sum_{i=1}^{n} E\left\{ E\left[\left| X_{i,n} \right|^q \middle| F_{i-1,n} \right] \right\} = \frac{1}{s_{Q_n}^q} \sum_{i=1}^{n} \sum_{t=1}^{T} E\left\{ E\left[\left| Y_{i,t} \right|^q \middle| F_{i-1,n} \right] \right\}$$
$$= \frac{1}{\left[n^{-1} s_{Q_n}^2 \right]^{1+\delta/2}} \frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n} \sum_{t=1}^{T} E\left\{ E\left[\left| Y_{i,t} \right|^q \middle| F_{i-1,n} \right] \right\} \leq \frac{1}{\left[n^{-1} s_{Q_n}^2 \right]^{1+\delta/2}} \left\{ \frac{1}{n^{\delta/2}} 2^{q+1} C_m^{q/p+1} C_e \right\}$$

Since $n^{-1} s_{Q_n}^2 \geq c > 0$, the right-hand side of the last inequality goes to zero as $n \to \infty$, which proves that condition 3.6.4 holds.

Now, using $s_{Q_n}^2 = \sum_{i=1}^{n} \left(\sum_{t=1}^{T} E\left(Y_{i,t}^2 \right) + 2 \sum_{t=2}^{T} \sum_{s=1}^{t-1} E\left(Y_{i,t} Y_{i,s} \right) \right)$ and the definition of $X_{i,n}$ we obtain that

$$\sum_{i=1}^{n} E\left[X_{i,n}^2 \middle| F_{i-1,n} \right] - 1 = \frac{1}{n^{-1} s_{Q_n}^2} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[E\left(Y_{i,t}^2 \middle| F_{i-1,n} \right) - E\left(Y_{i,t}^2 \right) \right]$$
$$+ \frac{2}{n^{-1} s_{Q_n}^2} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{s=1}^{t-1} \left[E\left(Y_{i,t} Y_{i,s} \middle| F_{i-1,n} \right) - E\left(Y_{i,t} Y_{i,s} \right) \right]$$

This means that, since $n^{-1} s_{Q_n}^2 \geq c > 0$, we can prove condition 3.6.5 by proving that

$$\frac{1}{n} \sum_{i=1}^{n} \left[E\left(Y_{i,t} Y_{i,s} \middle| F_{i-1,n} \right) - E\left(Y_{i,t} Y_{i,s} \right) \right] \xrightarrow{p} 0$$

for $t, s = 1, 2, ..., T$. We start the proof by noting that, since $(a_{i,t}, b_{i,t})$ are independent with zero mean, it follows that

$$
\begin{aligned}
&\left[E\left(Y_{i,t}Y_{i,s}\right| F_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] = \\
&\qquad + \sigma_{c,a,t,s}P_{t,n}[i,i]\sum_{j=1}^{i-1} P_{s,n}[j,i]b_{j,s}\sigma_{c,b,t,s}P_{t,n}[i,i]\sum_{j=1}^{i-1} P_{s,n}[i,j]a_{j,s} \\
&\qquad + \sigma_{c,a,s,t}P_{s,n}[i,i]\sum_{j=1}^{i-1} P_{t,n}[j,i]b_{j,t}\sigma_{c,b,s,t}P_{s,n}[i,i]\sum_{j=1}^{i-1} P_{t,n}[i,j]a_{j,t} \\
&\qquad + \sigma_{a,t,s}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1} P_{t,n}[j,i]P_{s,n}[l,i]\left[b_{j,t}b_{l,s} - 1(j=l)\sigma_{b,t,s}\right] \\
&\qquad + \sigma_{c,t,s}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1} P_{t,n}[j,i]P_{s,n}[i,l]\left[b_{j,t}a_{l,s} - 1(j=l)\sigma_{c,s,t}\right] \\
&\qquad + \sigma_{c,s,t}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1} P_{s,n}[j,i]P_{t,n}[i,l]\left[b_{j,s}a_{l,t} - 1(j=l)\sigma_{c,t,s}\right] \\
&\qquad + \sigma_{b,t,s}\sum_{j=1}^{i-1}\sum_{l=1}^{i-1} P_{t,n}[i,j]P_{s,n}[i,l]\left[a_{j,t}a_{l,s} - 1(j=l)\sigma_{a,t,s}\right]
\end{aligned}
$$

with $\sigma_{c,a,t,s} = E(a_{i,t}b_{i,t}a_{i,s})$ and $\sigma_{c,b,t,s} = E(a_{i,t}b_{i,t}b_{i,s})$, and so

$$
\frac{1}{n}\sum_{i=1}^{n}\left[E\left(Y_{i,t}Y_{i,s}\right| F_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] = \sum_{k=1}^{8} H_{k,n}
$$

where the subindex $1,\ldots,8$ indicates, in order of appearance, a summand in the expression above. Thus, to prove that $\frac{1}{s^2_{Q_n}}\sum_{i=1}^{n}\left[E\left(Y_{i,t}Y_{i,s}\right| F_{i-1,n}) - E\left(Y_{i,t}Y_{i,s}\right)\right] \xrightarrow{p} 0$, next we prove that $H_{k,n}\xrightarrow{p}0$ for $k = 1, ..., 8$.

To prove that $H_{1,n} = \sum_{i=1}^{n-1}\varphi_{i,n}b_{i,s}$ with $\varphi_{i,n} = n^{-1}\sigma_{c,a,t,s}P_{t,n}[i,i]\sum_{j=i+1}^{n} P_{s,n}[j,i]$, notice that, given that the $b_{i,s}$ are independent with mean zero, $E\left|b_{i,s}\right|^{1+\delta} \leq C_e$ for $\delta > 0$, $\limsup_{n\to\infty}\sum_{i=1}^{n-1}\varphi_{i,n} = \limsup_{n\to\infty} n^{-1}\sigma_{c,a,t,s}P_{t,n}[i,i]\sum_{j=i+1}^{n} P_{s,n}[j,i] \leq C_e C_m^2 < \infty$, and

$$
\begin{aligned}
\limsup_{n\to\infty}\sum_{i=1}^{n-1}\varphi_{i,n}^2 &= \limsup_{n\to\infty} n^{-2}\sigma^2_{c,a,t,s}\sum_{i=1}^{n-1}P_{t,n}[i,i]^2\left[\sum_{j=i+1}^{n} P_{s,n}[j,i]\right]^2 \\
&\leq n^{-1}C_e^2C_m^2 n^{-1}\sum_{i=1}^{n-1}C_m^2 \leq n^{-1}C_e^2C_m^4 \to 0
\end{aligned}
$$

Then, $H_{1,n}\xrightarrow{p}0$ by Davidson (1994, p. 299). Further, the cases $H_{k,n}$ for $k = 2, 3, 4$ can be proved in the same way.

For $H_{5,n}$, notice that

$$
\begin{aligned}
H_{5,n} &= \sigma_{a,t,s} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{l=1}^{i-1} P_{t,n}[j,i] P_{s,n}[l,i] \left[b_{j,t} b_{l,s} - 1(j=l)\sigma_{b,t,s} \right] \\
&= \sigma_{a,t,s} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{i-1} P_{t,n}[j,i] P_{s,n}[j,i] \left[b_{j,t} b_{j,s} - \sigma_{b,t,s} \right] + \sigma_{a,t,s} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{l=1}^{j-1} P_{t,n}[j,i] P_{s,n}[l,i] b_{j,t} b_{l,s} \\
&+ \sigma_{a,t,s} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{l=1}^{j-1} P_{t,n}[l,i] P_{s,n}[j,i] b_{l,t} b_{j,s} \\
&= H1_{5,n} + H2_{5,n} + H3_{5,n}.
\end{aligned}
$$

To prove that $H1_{5,n} \xrightarrow{p} 0$, we follow the same steps as in $H_{1,n}$. Notice that $H1_{5,n} = \sum_{i=1}^{n-1} \phi_{i,n} \left(b_{i,t} b_{i,s} - \sigma_{b,t,s} \right)$ with $\phi_{i,n} = n^{-1}\sigma_{a,t,s} \sum_{j=i+1}^{n} P_{t,n}[j,i] P_{s,n}[j,i]$. Then, given that $\left(b_{i,t} b_{i,s} - \sigma_{b,t,s} \right)$ are independent with mean zero, $E\left| b_{i,t} b_{i,s} - \sigma_{b,t,s} \right|^{1+\delta} \leq C_e$ for $\delta > 0$, $\limsup_{n\to\infty} \sum_{i=1}^{n-1} \phi_{i,n} = \limsup_{n\to\infty} n^{-1}\sigma_{a,t,s} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P_{t,n}[j,i] P_{s,n}[j,i] \leq \sigma_{a,t,s} C_m^2$ and $\limsup_{n\to\infty} \sum_{i=1}^{n-1} \phi_{i,n}^2 = \limsup_{n\to\infty} n^{-2}\sigma_{a,t,s}^2 \sum_{i=1}^{n-1} \left[\sum_{j=i+1}^{n} P_{t,n}[j,i] P_{s,n}[j,i] \right]^2 \leq \limsup_{n\to\infty} n^{-1}\sigma_{a,t,s}^2 C_m^4 = 0$. Thus, $H1_{5,n} \xrightarrow{p} 0$ by Davidson (1994, p. 299).

Similarly, for $H2_{5,n}$, given that the $b_{i,t} b_{j,s}$ are independent with zero mean, it is not difficult to see that

$$
\begin{aligned}
E\left(H2_{5,n}^2 \right) &\leq n^{-2} C_e \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{l=1}^{j-1} P_{t,n}[j,i]^2 P_{s,n}[l,i]^2 \\
&+ 4n^{-2} C_e \sum_{i=1}^{n} \sum_{j=1}^{i-1} \sum_{r=1}^{j-1} P_{t,n}[j,i] P_{t,n}[r,i] P_{s,n}[j,i] P_{s,n}[r,i] \\
&+ 2n^{-2} C_e \sum_{i=1}^{n} \sum_{k=1}^{i-1} \sum_{j=1}^{k-1} \sum_{r=1}^{k-1} P_{t,n}[j,i] P_{s,n}[r,i] P_{t,n}[j,k] P_{s,n}[r,k] \\
&\leq n^{-2} C_e \sum_{i=1}^{n} C_m^4 + 4n^{-2} C_e \sum_{i=1}^{n} C_m^4 \\
&+ 2n^{-2} C_e \sum_{i=1}^{n} \sum_{j=1}^{k-1} P_{t,n}[j,i] \sum_{k=1}^{i-1} P_{t,n}[j,k] \sum_{r=1}^{k-1} P_{s,n}[r,i] P_{s,n}[r,k] \\
&\leq 7n^{-1} C_e C_m^4 \longrightarrow 0
\end{aligned}
$$

Then, given that $E\left(H2_{5,n} \right) = 0$, $H2_{5,n} \xrightarrow{p} 0$. Also, the proof of $H3_{5,n} \xrightarrow{p} 0$ follows the same steps. This proves that $H_{5,n} \xrightarrow{p} 0$. Lastly, the cases $H_{k,n}$ for $k = 6, 7, 8$ can be proved in the same way. This concludes our proof of 3.6.5, and hence that of the Lemma. □

Lemma 3.15. *Under Assumptions* 3.1 *to* 3.8,

$$\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}(\psi_0)}{\partial\psi}\xrightarrow{d}N\left(0,E\left(\frac{1}{nT}\frac{\partial\mathcal{L}(\psi_0)}{\partial\psi}\frac{\partial\mathcal{L}(\psi_0)'}{\partial\psi}\right)\right)$$

Proof. The key to the proof is to show that $\frac{1}{\sqrt{nT}}\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\xrightarrow{d}N(0,\mathcal{G}_{11})$, with $\mathcal{G}_{11} = \lim_{n\to\infty}\frac{1}{nT}E\left[\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\boldsymbol{\eta}'\boldsymbol{\Omega}_0^{-1}\widetilde{\mathbf{X}}\right]$. In particular, by the Cramér-Wold device, it suffices to show that for any $c=(c_1',c_2',c_3)'\in\mathbb{R}^{4K+2}\times\mathbb{R}$ with $\|c\|=1$, $\frac{1}{\sqrt{nT}}c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\xrightarrow{d}N(0,c'\mathcal{G}_{11}c)$.

Let us define $\mathbb{X}1 = \left[\begin{array}{c:c:c}\boldsymbol{l}_{nT} & \mathbf{X} & \overline{\mathbf{X}}\end{array}\right]$, $\mathbb{X}2 = \left[\begin{array}{c:c}\mathbf{X} & \overline{\mathbf{X}}\end{array}\right]$, $\phi_{10}=(c_0,\beta_{10}',\pi_{\mu0}')'$ and $\phi_{20} = (\beta_{20}',\pi_{\alpha0}')'$. From 3.6.2 we have that:

$$\begin{aligned}
c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} &= c_1'\mathbb{X}1'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}+c_2'\mathbb{X}2'\mathbf{W}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\mathbf{Y}_{-1}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\\
&= c_1'\mathbb{X}1'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}+c_2'\mathbb{X}2'\mathbf{W}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\mathbf{Y}_{n0}'\boldsymbol{G}_{\mathbf{0}}^{-\prime}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}\\
&+c_3\phi_{10}'\mathbb{X}1'\boldsymbol{C}_{\mathbf{0}}^{-\prime}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\phi_{20}'\mathbb{X}2'\boldsymbol{C}_{\mathbf{0}}^{-\prime}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}+c_3\boldsymbol{\eta}'\boldsymbol{C}_{\mathbf{0}}^{-\prime}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}.
\end{aligned}$$

Following the steps of Lemma 3.11, we can write the summands of the previous expression as sums of quadratic forms:

$$\begin{aligned}
c_1'\mathbb{X}1'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} &= \sum_{t=1}^{T}c_1'\mathbb{X}1_{nt}'\xi_{nt}-\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}c_1'\mathbb{X}1_{nt}'B_{n0}^{-1}\xi_{ns}\\
&+\sum_{t=1}^{T}c_1'\mathbb{X}1_{nt}'W_n\upsilon_{n\alpha}-\sum_{t=1}^{T}c_1'\mathbb{X}1_{nt}'B_{n0}^{-1}W_n\upsilon_{n\alpha}\\
c_2'\mathbb{X}2'\mathbf{W}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} &= \sum_{t=1}^{T}c_2'\mathbb{X}2_{nt}'W_n'\xi_{nt}-\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}c_2'\mathbb{X}2_{nt}'W_n'B_{n0}^{-1}\xi_{ns}\\
&+\sum_{t=1}^{T}c_2'\mathbb{X}2_{nt}'W_n'W_n\upsilon_{n\alpha}-\sum_{t=1}^{T}c_2'\mathbb{X}2_{nt}'W_n'B_{n0}^{-1}W_n\upsilon_{n\alpha}\\
\mathbf{Y}_{n0}'\boldsymbol{G}_{\mathbf{0}}^{-\prime}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} &= \sum_{t=1}^{T}\mathbf{Y}_{n0}'G_{nt-1}'\xi_{nt}-\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{Y}_{n0}'G_{nt-1}'B_{n0}^{-1}\xi_{ns}\\
&+\sum_{t=1}^{T}\mathbf{Y}_{n0}'G_{nt-1}'W_n\upsilon_{n\alpha}-\sum_{t=1}^{T}\mathbf{Y}_{n0}'G_{nt-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}\\
\phi_{10}'\mathbb{X}1'\boldsymbol{C}_{\mathbf{0}}^{-\prime}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} &= \sum_{t=1}^{T}\sum_{j=1}^{t}\phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'\xi_{nt}-\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T}\phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns}\\
&+\sum_{t=1}^{T}\sum_{j=1}^{t}\phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'W_n\upsilon_{n\alpha}-\sum_{t=1}^{T}\sum_{j=1}^{T}\phi_{10}'\mathbb{X}1_{nj}'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}
\end{aligned}$$

$$
\begin{aligned}
&\phi_{20}'\mathbb{X}2'W_n'\boldsymbol{C_0^-}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \\
&\qquad \sum_{t=1}^{T}\sum_{j=1}^{t}\phi_{20}\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T}\phi_{20}'\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \\
&\qquad \sum_{t=1}^{T}\sum_{j=1}^{t}\phi_{20}'\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{T}\phi_{20}'\mathbb{X}2_{nj}'W_n'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\eta}'\boldsymbol{C_0^-}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = &\sum_{t=1}^{T}\sum_{j=1}^{t}\xi_{nj}'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T}\xi_{nj}'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \\
&\sum_{t=1}^{T}\sum_{j=1}^{t}\upsilon_{n\alpha}'W_n'C_{n,t-j-1}'\xi_{nt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}\sum_{s=1}^{T}\upsilon_{n\alpha}'W_n'C_{n,t-j-1}'B_{n0}^{-1}\xi_{ns} + \\
&\sum_{t=1}^{T}\sum_{j=1}^{t}\xi_{nj}'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{t}\xi_{nj}'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha} + \\
&\sum_{t=1}^{T}\sum_{j=1}^{t}\upsilon_{n\alpha}'W_n'C_{n,t-j-1}'W_n\upsilon_{n\alpha} - \sum_{t=1}^{T}\sum_{j=1}^{t}\upsilon_{n\alpha}'W_n'C_{n,t-j-1}'B_{n0}^{-1}W_n\upsilon_{n\alpha}
\end{aligned}
$$

with $\xi_{nt} = (\varepsilon_{nt} + \upsilon_{n\mu})$, $C_{n,-1} = 0_{n\times n}$ and $B_{n0}^{-1} = B_n(\sigma_0)^{-1}$ (see Lemma 3.10 for the definition of $B_n(\sigma)^{-1}$).

We can thus write $c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} = \sum_{l=1}^{L} a_l' P_{n,l} b_l$ with $L < \infty$. Then, it is easy to verify that a_l, b_l and $P_{n,l}$ for $l = 1, 2, ..., L$ satisfy the conditions of Lemma 3.14 and, by Assumption 3.7, that $n^{-1}Var(c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}) \geq \bar{c} > 0$, so that $\dfrac{c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta}}{\left[Var(c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta})\right]^{1/2}} \xrightarrow{d} N(0,1)$, which in turn implies that $\dfrac{1}{\sqrt{nT}} c'\widetilde{\mathbf{X}}'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\eta} \xrightarrow{d} N(0, c'\mathcal{G}_{11}c)$ for any $c \in \mathbb{R}^{4K+2} \times \mathbb{R}$ with $\|c\| = 1$. This proves the convergence for the first term of the gradient.

To conclude the proof, we note that each component of $\dfrac{\partial \mathcal{L}(\psi_0)}{\partial \psi}$ (see Appendix 3.8 for details) can be written as a finite sum of quadratic forms, so that the proof for these cases proceeds by closely following the previous steps. We consequently omit the details of these proofs. □

3.7 Appendix B: Proof of Theorems.

We start by proving the consistence of the QML estimator (Theorem 3.1). The proof of normality comes next (Theorem 3.2).

3.7.1 Consistency

Proof of Theorem 3.1. The consistency proof closely follows the proof of Theorem 4.1 of Su and Yang (2015). In particular, by Theorem 3.4 of White (1994), it suffices to show that:

(1.) $\frac{1}{nT}\left[\mathcal{L}_c^*(\delta)-\mathcal{L}_c(\delta)\right]\xrightarrow{p}0$ uniformly in $\delta\in\Delta=\Delta_\sigma\times\Delta_\lambda$
(2.) $\lim\sup_{n\to\infty}\max_{\delta\in N_\epsilon^c(\delta_0)}\frac{1}{nT}\left[\mathcal{L}_c^*(\delta)-\mathcal{L}_c^*(\delta_0)\right]<0$ for any $\epsilon>0$, where $N_\epsilon^c(\delta_0)$ is the complement of an open neighbourhood of δ_0 on Δ of radius ϵ.

To show that **(1.)** holds, it is sufficient to show that the following conditions hold: **(1.a)** $\widehat{\sigma}_\varepsilon^2(\delta)-\widetilde{\sigma}_\varepsilon^2(\delta)\xrightarrow{p}0$ uniformly in $\delta\in\Delta$ and **(1.b)** $\widetilde{\sigma}_\varepsilon^2(\delta)$ is uniformly bounded away from zero on Δ. Since **(1b)** will be checked in the proof of **(2.)**, next we concentrate on the proof of **(1.a)**.

By definition of our model, $\hat{\boldsymbol{\eta}}(\delta)=\mathbf{S}(\lambda)\mathbf{Y}-\tilde{\mathbf{X}}\hat{\theta}(\delta)=\boldsymbol{\Omega}^{1/2}(\sigma)\mathbf{M}(\sigma)\boldsymbol{\Omega}^{-1/2}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}$, where $\mathbf{M}(\sigma)=\mathbf{I}_{nT}-\boldsymbol{\Omega}^{-1/2}(\sigma)\widetilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1/2}(\sigma)$. This means that

$$\begin{aligned}
\widehat{\sigma}_\varepsilon^2(\delta)&=\frac{1}{nT}\left(\mathbf{Y}'\mathbf{S}'(\lambda)\boldsymbol{\Omega}^{-1/2}(\sigma)\mathbf{M}(\sigma)\boldsymbol{\Omega}^{1/2}(\sigma)\right)\boldsymbol{\Omega}^{-1}(\sigma)\left(\boldsymbol{\Omega}^{1/2}(\sigma)\mathbf{M}(\sigma)\boldsymbol{\Omega}^{-1/2}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}\right)\\
&=\frac{1}{nT}\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}-\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\\
&-(\lambda-\lambda_0)\frac{1}{nT}\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}+(\lambda-\lambda_0)\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\\
&-(\lambda-\lambda_0)\frac{1}{nT}\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}+(\lambda-\lambda_0)\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\\
&+(\lambda-\lambda_0)^2\frac{1}{nT}\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}-(\lambda-\lambda_0)^2\frac{1}{nT}\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)
\end{aligned}$$

with $\mathbf{Q}_{\mathbf{A},\mathbf{B}}(\delta)=\mathbf{A}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{B}$.

From $\max_{\theta,\sigma_\varepsilon^2}E\left[\mathcal{L}(\psi)\right]$,

$$\begin{aligned}
\tilde{\theta}(\delta)&=\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{S}(\lambda)\mathbf{Y}\right]\\
&=\theta_0+\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right]\\
&-(\lambda-\lambda_0)\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}\right]
\end{aligned}$$

Then,

$$\begin{aligned}
\eta(\tilde{\theta}(\delta))\equiv\tilde{\eta}(\delta)&=\mathbf{S}(\lambda)\mathbf{Y}-\tilde{\mathbf{X}}\tilde{\theta}(\delta)\\
&=\eta-(\lambda-\lambda_0)\,\mathbf{WY}-\tilde{\mathbf{X}}\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right]\\
&+(\lambda-\lambda_0)\,\tilde{\mathbf{X}}\left[E\left(\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\tilde{\mathbf{X}}\right)\right]^{-1}E\left[\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}\right]
\end{aligned}$$

and, consequently,

$$\begin{aligned}\tilde{\sigma}_{\varepsilon}(\delta) &= \frac{1}{nT}E\left[\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right] - \frac{1}{nT}(\lambda-\lambda_0)E\left[\boldsymbol{\eta}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}\right] \\ &- \frac{1}{nT}E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right] \\ &+ \frac{1}{nT}(\lambda-\lambda_0)E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right] \\ &- \frac{1}{nT}(\lambda-\lambda_0)E\left[\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\boldsymbol{\eta}\right] + \frac{1}{nT}(\lambda-\lambda_0)^2E\left[\mathbf{Y}'\mathbf{W}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{WY}\right] \\ &+ \frac{1}{nT}(\lambda-\lambda_0)E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\boldsymbol{\eta}}(\sigma)\right] \\ &- \frac{1}{nT}(\lambda-\lambda_0)^2E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{WY}}(\sigma)\right]\end{aligned}$$

Let us also define

$$\begin{aligned}\boldsymbol{\Pi}_{\mathbf{a},\mathbf{b}}(\sigma) &= \frac{1}{nT}\left\{\mathbf{a}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{b} - E\left[\mathbf{a}'\boldsymbol{\Omega}^{-1}(\sigma)\mathbf{b}'\right]\right\} \\ \boldsymbol{\Upsilon}_{\mathbf{a},\mathbf{b}}(\sigma) &= \frac{1}{nT}\left\{\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{a}}(\sigma)\mathbf{Q}^{-1}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma) - E\left[\mathbf{Q}'_{\tilde{\mathbf{X}},\mathbf{a}}(\sigma)\right]\left[E\left(\mathbf{Q}_{\tilde{\mathbf{X}},\tilde{\mathbf{X}}}(\sigma)\right)\right]^{-1}E\left[\mathbf{Q}_{\tilde{\mathbf{X}},\mathbf{b}}(\sigma)\right]\right\},\end{aligned}$$

where $\mathbf{a},\mathbf{b} = \boldsymbol{\eta},\mathbf{WY}$. By using these when calculating the difference between $\tilde{\sigma}_{\varepsilon}(\delta)$ and $\hat{\sigma}_{\varepsilon}(\delta)$ we obtain:

$$\begin{aligned}\hat{\sigma}_{\varepsilon}(\delta) - \tilde{\sigma}_{\varepsilon}(\delta) &= \boldsymbol{\Pi}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\sigma) - (\lambda-\lambda_0)\boldsymbol{\Pi}_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) - (\lambda-\lambda_0)\boldsymbol{\Pi}'_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) + (\lambda-\lambda_0)^2\boldsymbol{\Pi}_{\mathbf{WY},\mathbf{WY}}(\sigma) \\ &- \boldsymbol{\Upsilon}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\sigma) + (\lambda-\lambda_0)\boldsymbol{\Upsilon}_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) + (\lambda-\lambda_0)\boldsymbol{\Upsilon}'_{\boldsymbol{\eta},\mathbf{WY}}(\sigma) - (\lambda-\lambda_0)^2\boldsymbol{\Upsilon}_{\mathbf{WY},\mathbf{WY}}(\sigma)\end{aligned}$$

and, therefore, condition **(1.a)** follows by using Lemmas 3.11 and 3.12.

To show that condition **(2.)** holds, we closely follow the literature (Lee, 2004; Yu et al., 2008; Su and Yang, 2015) and use an auxiliary process to show, using Jensen inequality and $\tilde{\sigma}^2_{\varepsilon}(\delta_0) = \frac{\sigma^2_{\varepsilon 0}}{nT}tr(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Omega}_0) = \sigma^2_{\varepsilon 0}$ (which follows from the definition of $\tilde{\sigma}^2_{\varepsilon}(\delta)$ and Lemma 3.3), that

$$\mathcal{L}^*_c(\delta) \leq \mathcal{L}^*_c(\delta_0) \tag{3.7.1}$$

Next we prove that $\frac{1}{nT}\mathcal{L}^*_c(\delta)$ is uniformly equicontinuous on $\delta \in \Delta$ by showing the uniform equicontinuity of $\frac{1}{nT}\ln|\mathbf{S}(\lambda)|$, then that of $\frac{1}{nT}\ln|\boldsymbol{\Omega}(\sigma)|$ and finally that of $\ln(\tilde{\sigma}^2_{\varepsilon}(\delta))$ on $\delta \in \Delta$.

Firstly, by the mean value theorem, $\ln|\mathbf{S}(\lambda^*)| - \ln|\mathbf{S}(\lambda^{**})| = \left(\frac{\partial}{\partial\lambda}\ln|\mathbf{S}(\overline{\lambda})|\right)(\lambda^* - \lambda^{**})$ with $\overline{\lambda} \in (\lambda^*, \lambda^{**})$. Also,

$$\frac{1}{nT}\frac{\partial}{\partial\lambda}\ln|\mathbf{S}(\overline{\lambda})| = \frac{1}{nT}tr\left[\mathbf{S}^{-1}(\overline{\lambda})\mathbf{W}\right] = O(1),$$

since $\mathbf{S}^{-1}(\lambda)\mathbf{W}$ is u.b.r.c.s. uniformly in λ and hence $tr\left[\mathbf{S}^{-1}(\overline{\lambda})\mathbf{W}\right] = O(nT)$. Thus, $\ln|\mathbf{S}(\lambda)|$

is uniformly equicontinuous in λ on Δ_λ.

Secondly, by the mean value theorem, $\ln|\mathbf{\Omega}(\sigma^*)| - \ln|\mathbf{\Omega}(\sigma^{**})| = \sum_{k=1}^{3}\left(\frac{\partial}{\partial\sigma_k}\ln|\mathbf{\Omega}(\overline{\sigma})|\right)(\sigma_k^* - \sigma_k^{**})$, with $\overline{\sigma}$ lying elementwise between σ^* and σ^{**}. Also,

$$\begin{aligned}
\frac{1}{nT}\frac{\partial}{\partial\sigma_1}\ln|\mathbf{\Omega}(\overline{\sigma})| &= \frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})(J_T \otimes I_n)\right] \\
\frac{1}{nT}\frac{\partial}{\partial\sigma_2}\ln|\mathbf{\Omega}(\overline{\sigma})| &= \frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})\left(J_T \otimes (W_n + W_n')\right)\right] \\
\frac{1}{nT}\frac{\partial}{\partial\sigma_3}\ln|\mathbf{\Omega}(\overline{\sigma})| &= \frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})(J_T \otimes W_n W_n')\right]
\end{aligned}$$

Notice that, by Lemma 3.2, and, given that $tr(W_n + W_n') = O(n)$ (see Remark A2 in Kapoor et al. 2007) and $\sup_\sigma \tau_{\max}(\Omega^{-1}(\sigma)) < c_\tau < \infty$ (by Lemma 3.9), we can show that $\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}(\overline{\sigma})\left(J_T \otimes (W_n + W_n')\right)\right] \leq \frac{1}{nT}\left[\sup_\sigma \tau_{\max}(\Omega^{-1}(\sigma))\right]^{-1} tr\left(J_T \otimes (W_n + W_n')\right) \leq \frac{1}{nT}c_\tau tr(J_T)tr(W_n + W_n') = O(1)$ uniformly on Δ, and similarly for the other two cases (since, by Remark A2 in Kapoor et al. 2007, $tr(W_n W_n') = O(n)$). Thus, $\ln|\mathbf{\Omega}(\sigma)|$ s uniformly equicontinuous in σ on Δ_λ.

Thirdly, to show that $\ln[\tilde{\sigma}_\varepsilon^2(\delta)]$ is uniformly equicontinuous on Δ it suffices to show that $\tilde{\sigma}_\varepsilon^2(\delta)$ is uniformly equicontinuous and uniformly bounded away from zero on Δ. Thus, we start by noting from the definition of $\tilde{\sigma}_\varepsilon^2(\delta)$ used in the proof of **(1.a)** that all its elements appear in $\mathbf{\Pi_{a,b}}(\sigma)$ and $\mathbf{\Upsilon_{a,b}}(\sigma)$, which, using the same arguments as in Lemmas 3.11 and 3.12, and the results in Lemma 3.7, proves the uniform equicontinuity of $\tilde{\sigma}_\varepsilon(\delta)$. Next, to show that $\tilde{\sigma}_\varepsilon^2(\delta)$ is uniformly bounded away from zero, we follow Su and Yang (2015) and establish the claim by a counter argument based on making its dependence on n explicit. To this end, we include the subindex n in $\tilde{\sigma}_\varepsilon^2(\delta)$, so that it then becomes $\tilde{\sigma}_{\varepsilon,n}^2(\delta)$.

If $\tilde{\sigma}_{\varepsilon,n}^2(\delta)$ is not uniformly bounded away from zero on Δ, then there exists a sequence $\{\delta_n\}$ in Δ such that $\lim_{n\to\infty}\tilde{\sigma}_{\varepsilon,n}^2(\delta) = 0$. Now, by 3.7.1 we have that, for all δ,

$$-\ln\left[\tilde{\sigma}_\varepsilon^2(\delta)\right] \leq -\ln\left[\tilde{\sigma}_\varepsilon^2(\delta_0)\right] + \frac{1}{nT}\left[\ln|\mathbf{S}_0| - \ln|\mathbf{S}(\lambda)|\right] + \frac{2}{nT}\left[\ln|\mathbf{\Omega}(\sigma)| - \ln|\mathbf{\Omega}_0|\right]$$

Using the mean value theorem, as we previously did, it can be proved that $\frac{1}{nT}\left[\ln|\mathbf{S}_0| - \ln|\mathbf{S}(\lambda)|\right] = O(1)$ and $\frac{2}{nT}\left[\ln|\mathbf{\Omega}(\sigma)| - \ln|\mathbf{\Omega}_0|\right] = O(1)$ uniformly in Δ. This implies that $-\ln\left[\tilde{\sigma}_\varepsilon^2(\delta)\right]$ is bounded above, which is a contradiction, so we conclude that $\tilde{\sigma}_{\varepsilon,n}^2(\delta)$ is uniformly bounded away from zero on Δ.

Finally, the identification uniqueness also follows by contradiction. Using $\tilde{\sigma}_\varepsilon^2(\delta_0) = \sigma_{\varepsilon 0}^2$ (see Lemma 3.3) we have that

$$\begin{aligned}\frac{1}{nT}\left[\mathcal{L}_c^*(\delta)-\mathcal{L}_c^*(\delta_0)\right] &= \frac{1}{2nT}\left\{\ln|\mathbf{\Omega}_0|-\ln|\mathbf{\Omega}(\sigma)|\right\}+\frac{1}{2}\left\{\ln\left[\sigma_{\varepsilon 0}^2\right]-\ln\left[\tilde{\sigma}_\varepsilon^2(\delta)\right]\right\}\\ &+\frac{1}{nT}\left\{\ln|\mathbf{S}(\lambda)|-\ln|\mathbf{S}_0|\right\}\\ &= \frac{1}{2nT}\left\{\ln\left|\sigma_{\varepsilon 0}^2\mathbf{S}_0^{-2}\mathbf{\Omega}_0\right]-\ln\left|\tilde{\sigma}_\varepsilon^2(\delta)\mathbf{S}^{-2}(\lambda)\mathbf{\Omega}(\sigma)\right|\right\}\end{aligned}$$

If the identification uniqueness condition does not hold, then there exists an $\epsilon > 0$ and a sequence $\{\delta_n\}$ in $N_\epsilon^c(\delta_0)$ such that

$$\lim_{n\to\infty}\frac{1}{nT}\left[\mathcal{L}_{c,n}^*(\delta)-\mathcal{L}_{c,n}^*(\delta_0)\right]=0,$$

where we have written $\mathcal{L}_{c,n}^*(.)$ for $\mathcal{L}_c^*(.)$ to stress its dependence on n. However, by the compactness of $N_\epsilon^c(\delta_0)$, there exists a convergent subsequence $\{\delta_{n_k}\}$ of $\{\delta_n\}$ with the limit δ_+ of δ_{n_k} being in $N_\epsilon^c(\delta_0)$. This implies that $\delta_+ \neq \delta_0$. Furthermore, by the uniform equicontinuity of $\frac{1}{nT}\mathcal{L}_{c,n}^*(\delta)$, $\lim_{n\to\infty}\frac{1}{n_kT}\left[\mathcal{L}_{c,n}^*(\delta_+)-\mathcal{L}_{c,n}^*(\delta_0)\right]=0$. Yet this contradicts Assumption 3.6, since it amounts to $\lim_{n\to\infty}\frac{1}{nT}\left[\mathcal{L}_{c,n}^*(\delta)-\mathcal{L}_{c,n}^*(\delta_0)\right]\neq 0$ for any $\delta\neq\delta_0$. This completes the proof of the theorem. □

3.7.2 Asymptotic normality

Proof of Theorem 3.2. By Taylor series expansion,

$$0=\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}(\psi)}{\partial\psi}\bigg|_{\widehat{\psi}}=\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}(\psi)}{\partial\psi}\bigg|_{\psi_0}+\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi)}{\partial\psi\partial\psi'}\bigg|_{\overline{\psi}}\sqrt{nT}\left(\widehat{\psi}-\psi_0\right)$$

where the elements of $\overline{\psi}=\left(\overline{\theta}',\overline{\sigma}_\varepsilon^2,\overline{\lambda},\overline{\sigma}'\right)'$ lie in the segment joining the corresponding elements of $\widehat{\psi}$ and ψ_0. Thus,

$$\sqrt{nT}\left(\widehat{\psi}-\psi_0\right)=\left[-\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi)}{\partial\psi\partial\psi'}\bigg|_{\overline{\psi}}\right]^{-1}\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}(\psi)}{\partial\psi}\bigg|_{\psi_0}=\left[-\frac{1}{nT}\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\psi\partial\psi'}\right]^{-1}\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}(\psi_0)}{\partial\psi}$$

By Theorem 3.1, $\widehat{\psi}\xrightarrow{p}\psi_0$, and so $\overline{\psi}\xrightarrow{p}\psi_0$. Therefore, it suffices to show that:

i) $\frac{1}{nT}\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\psi\partial\psi'}-\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'}=o_p(1)$,

ii) $\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'}\xrightarrow{p}E\left(\frac{1}{nT}\frac{\partial^2\mathcal{L}(\psi_0)}{\partial\psi\partial\psi'}\right)$, and

iii) $\frac{1}{\sqrt{nT}}\frac{\partial\mathcal{L}(\psi_0)}{\partial\psi}\xrightarrow{d}N\left(0,E\left(\frac{1}{nT}\frac{\partial\mathcal{L}(\psi_0)}{\partial\psi}\frac{\partial\mathcal{L}(\psi_0)'}{\partial\psi}\right)\right)$.

Since ii) and (iii) follow from Lemmas 3.13 and 3.15, respectively, only (i) is left to be shown. In particular, given the expression of $\dfrac{\partial^2 \mathcal{L}(\psi)}{\partial\psi\partial\psi'}$ provided in Appendix 3.8, it suffices to show that $\dfrac{1}{nT}\dfrac{\partial^2\mathcal{L}(\overline{\psi})}{\partial\omega\partial\varpi'} - \dfrac{1}{nT}\dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\omega\partial\varpi'} = o_p(1)$ for $\omega, \varpi = \theta, \sigma^2_\varepsilon, \lambda$ and σ. However, we only show this for $(\omega,\varpi) = (\theta,\theta), (\theta,\sigma^2_\varepsilon), (\lambda,\lambda)$ and $(\sigma_\kappa, \sigma_\varrho)$, with $\kappa, \varrho = 1,2,3$, for the other cases can be shown in an analogous way.

For the (θ,θ) case, notice that

$$\begin{aligned}\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}(\overline{\psi})}{\partial\theta\partial\theta'} - \frac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\theta'}\right] &= -\frac{1}{nT}\frac{1}{\overline{\sigma}^2_\varepsilon}\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\overline{\sigma})\widetilde{\mathbf{X}} + \frac{1}{nT}\frac{1}{\sigma^2_{\varepsilon 0}}\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}} \\ &= \left(\frac{\overline{\sigma}^2_\varepsilon - \sigma^2_{\varepsilon 0}}{\sigma^2_{\varepsilon 0}\overline{\sigma}^2_\varepsilon}\right)\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}} + \frac{1}{\overline{\sigma}^2_\varepsilon}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)\mathbf{\Omega}^{-1}(\overline{\sigma})\widetilde{\mathbf{X}}\right] \qquad (3.7.2)\end{aligned}$$

Given that $\sigma^2_{\varepsilon 0} > 0$ and $\overline{\sigma}^2_\varepsilon \xrightarrow{p} \sigma^2_{\varepsilon 0}$, $\left(\dfrac{\overline{\sigma}^2_\varepsilon - \sigma^2_{\varepsilon 0}}{\sigma^2_{\varepsilon 0}\overline{\sigma}^2_\varepsilon}\right) = o_p(1)$, from the proof of Lemma 3.12 we can show that $\dfrac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}} = O_p(1)$. As for the second term in the r.h.s. of 3.7.2, note that $\tau^{1/2}_{\max}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right) = O_p(\|\overline{\sigma} - \sigma\|) = o_p(1)$. To prove this, notice that, since $\tau_{\max}(A\otimes B) \le \tau_{\max}(A)\,\tau_{\max}(B)$, then $\tau^{1/2}_{\max}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right) = T\tau^{1/2}_{\max}\left((\mathbf{\Sigma}(\overline{\sigma}) - \mathbf{\Sigma}_0)^2\right)$. Further,

$$\begin{aligned}\tau^{1/2}_{\max}\left((\mathbf{\Sigma}(\overline{\sigma}) - \mathbf{\Sigma}_0)^2\right) &= \|(\mathbf{\Sigma}(\overline{\sigma}) - \mathbf{\Sigma}_0)\|_2 \\ &= \|(\overline{\sigma}_1 - \sigma_{10})\,I_n + (\overline{\sigma}_2 - \sigma_{20})(W_n + W_n') + (\overline{\sigma}_3 - \sigma_{30})\,W_nW_n'\|_2 \\ &\le |\overline{\sigma}_1 - \sigma_{10}|\,\|I_n\|_2 + |\overline{\sigma}_2 - \sigma_{20})\,\|\|(W_n + W_n')\|_2 + |\overline{\sigma}_3 - \sigma_{30}|\,\|W_nW_n'\|_2\end{aligned}$$

Then, given that W_n is u.b.r.c.s., $\|W_n + W_n'\|_2 \le \left(\|W_n + W_n'\|_1\|W_n + W_n'\|_\infty\right)^{1/2} < \infty$ and $\|W_nW_n'\|_2 \le \left(\|W_nW_n'\|_1\|W_nW_n'\|_\infty\right)^{1/2} < \infty$. Thus, $\tau^{1/2}_{\max}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right) \le [|\overline{\sigma}_1 - \sigma_{10}| + |\overline{\sigma}_2 - \sigma_{20}| + |\overline{\sigma}_2 - \sigma_{20}|]\,Tc_\tau$ with $c_\tau < \infty$.

Let c be an arbitrary column vector in $\mathbb{R}^{4K+2}$. Then, by the Cauchy-Schwarz inequality, Lemmas 3.9 and 3.2, and $\dfrac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right| = O_p(1)$ (which can be proved following the same steps as in Lemma 3.12), we have that

$$\begin{aligned}&\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\ \mathbf{\Omega}_0^{-1}(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)\mathbf{\Omega}^{-1}(\overline{\sigma})\widetilde{\mathbf{X}}c\right| \\ &\qquad\le \frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}}c\right|^{1/2}\left|c'\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\overline{\sigma})(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)(\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)\mathbf{\Omega}^{-1}(\overline{\sigma})\widetilde{\mathbf{X}}c\right|^{1/2} \\ &\qquad\le \tau_{\max}\left(\mathbf{\Omega}_0^{-1}\right)\tau_{\max}\left(\mathbf{\Omega}^{-1}(\overline{\sigma})\right)\tau^{1/2}_{\max}\left((\mathbf{\Omega}(\overline{\sigma}) - \mathbf{\Omega}_0)^2\right)\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right| \\ &\qquad\le O_p(\|\overline{\sigma} - \sigma\|)\,O_p(1) = o_p(1)\end{aligned}$$

Since $\dfrac{1}{\overline{\sigma}^2_\varepsilon} = O_p(1)$, it follows that the second term in the r.h.s. of 3.7.2 is $o_p(1)$.

For the $\left(\theta, \sigma_{\varepsilon}^{2}\right)$ case, notice that

$$\begin{aligned}
\frac{1}{nT}\left[\frac{\partial^{2}\mathcal{L}\left(\overline{\psi}\right)}{\partial\theta\partial\sigma_{\varepsilon}^{2}}-\frac{\partial^{2}\mathcal{L}\left(\psi_{0}\right)}{\partial\theta\partial\sigma_{\varepsilon}^{2}}\right] =& \frac{1}{nT}\left[\frac{1}{\overline{\sigma}_{\varepsilon}^{4}}\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)-\frac{1}{\sigma_{\varepsilon}^{4}}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\boldsymbol{\eta}\right] \\
=& \left(\frac{1}{\overline{\sigma}_{\varepsilon}^{4}}-\frac{1}{\sigma_{\varepsilon}^{4}}\right)\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\boldsymbol{\eta}+\frac{1}{\overline{\sigma}_{\varepsilon}^{4}}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\Omega_{0}^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right] \\
&+\frac{1}{\overline{\sigma}_{\varepsilon}^{4}}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\left(\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)-\boldsymbol{\eta}\right)\right]
\end{aligned}$$

Following the same steps as in Lemmas 3.12 and 3.13, it can be proved that $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\boldsymbol{\eta}=o_{p}\left(1\right)$, $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\mathbf{Y}=O_{p}\left(1\right)$ and $\frac{1}{nT}\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\widetilde{\mathbf{X}}=O_{p}\left(1\right)$. Thus, by using $\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)=\left(\lambda_{0}-\overline{\lambda}\right)\mathbf{Y}+\boldsymbol{\eta}+\widetilde{\mathbf{X}}\left(\overline{\theta}-\theta_{0}\right)$, it can be proved that the three summands in the previous expression are $o_{p}\left(1\right)$.

For the $\left(\lambda, \lambda\right)$ case, notice that

$$\begin{aligned}
&\frac{1}{nT}\left[\frac{\partial^{2}\mathcal{L}\left(\overline{\psi}\right)}{\partial\lambda\partial\lambda}-\frac{\partial^{2}\mathcal{L}\left(\psi_{0}\right)}{\partial\lambda\partial\lambda}\right] \\
&=\frac{1}{nT}\left[tr\left(\left(\mathbf{S}_{0}^{-1}\mathbf{W}\right)^{2}\right)-tr\left(\left(\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right)^{2}\right)+\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\mathbf{W}\mathbf{Y}-\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{W}\mathbf{Y}\right] \\
&=\frac{1}{nT}\left[tr\left(\left(\mathbf{S}_{0}^{-1}\mathbf{W}\right)^{2}-\left(\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right)^{2}\right)\right]+\frac{1}{nT}\left[\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}\left(\overline{\sigma}\right)-\mathbf{\Omega}_{0}\right)\mathbf{W}\mathbf{Y}\right]
\end{aligned}$$

Let us now consider the first term of the previous expression. Given that $\mathbf{S}^{-1}\left(\lambda\right)$ and $\mathbf{W}$ are u.b.r.c.s. uniformly in λ, then

$$\begin{aligned}
\frac{1}{nT}tr\left(\left(\mathbf{S}_{0}^{-1}\mathbf{W}\right)^{2}-\left(\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right)^{2}\right) \leq& \left|\overline{\lambda}-\lambda_{0}\right|\frac{1}{nT}tr\left(\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right) \\
&+\left|\overline{\lambda}-\lambda_{0}\right|\frac{1}{nT}tr\left(\mathbf{S}_{0}^{-1}W\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right) \\
\leq& \, o_{p}\left(1\right)O\left(1\right),
\end{aligned}$$

where the second inequality holds because $tr\left(\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right)=O\left(nT\right)$ and $tr\left(\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}_{0}^{-1}\mathbf{W}\mathbf{S}^{-1}\left(\overline{\lambda}\right)\mathbf{W}\right)=O\left(nT\right)$. As for the second term, $\frac{1}{nT}\left[\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}\left(\overline{\sigma}\right)-\mathbf{\Omega}_{0}\right)\mathbf{W}\mathbf{Y}\right]=\frac{1}{nT}\left[\theta_{0}'\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}\left(\overline{\sigma}\right)-\mathbf{\Omega}_{0}\right)\widetilde{\mathbf{X}}\theta_{0}\right]+\frac{1}{nT}\left[\boldsymbol{\eta}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}\left(\overline{\sigma}\right)-\mathbf{\Omega}_{0}\right)\boldsymbol{\eta}\right]+2\frac{1}{nT}\left[\theta_{0}'\widetilde{\mathbf{X}}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}\left(\overline{\sigma}\right)-\mathbf{\Omega}_{0}\right)\boldsymbol{\eta}\right]$, so that, using arguments analogous to the ones used in previous cases, it can be proved that $\frac{1}{nT}\left[\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_{0}^{-1}\left(\mathbf{\Omega}\left(\overline{\sigma}\right)-\mathbf{\Omega}_{0}\right)\mathbf{W}\mathbf{Y}\right]=o_{p}\left(1\right)$.

For the $(\sigma_\kappa, \sigma_\kappa)$ case, notice that

$$\begin{aligned}
\frac{1}{nT}\left[\frac{\partial^2 \mathcal{L}(\overline{\psi})}{\partial \sigma_\kappa \partial \sigma_\kappa} - \frac{\partial^2 \mathcal{L}(\psi_0)}{\partial \sigma_\kappa \partial \sigma_\kappa}\right] &= \frac{1}{nT}\left[\frac{1}{\overline{\sigma}_\varepsilon^2}\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\overline{\sigma})\mathbf{\Sigma}_\kappa\mathbf{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\eta}(\overline{\theta}) - \frac{1}{\sigma_\varepsilon^2}\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] \\
&= \left(\frac{1}{\overline{\sigma}_\varepsilon^2} - \frac{1}{\sigma_\varepsilon^2}\right)\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] \\
&+ \frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}^{-1}(\overline{\sigma})\mathbf{\Sigma}_\kappa\mathbf{\Omega}^{-1}(\overline{\sigma})\boldsymbol{\eta}(\overline{\theta},\overline{\lambda}) - \widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] \\
&= \left(\frac{1}{\overline{\sigma}_\varepsilon^2} - \frac{1}{\sigma_\varepsilon^2}\right)\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\right] \\
&+ \frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\left(\mathbf{\Omega}^{-1}(\overline{\sigma})\mathbf{\Sigma}_\kappa\mathbf{\Omega}^{-1}(\overline{\sigma}) - \mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\right)\boldsymbol{\eta}(\overline{\theta},\overline{\lambda})\right] \\
&+ \frac{1}{\overline{\sigma}_\varepsilon^2}\frac{1}{nT}\left[\widetilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\widetilde{\mathbf{X}}\left(\overline{\theta} - \theta_0\right)\right]
\end{aligned}$$

The first and third summands can be proved to be $o_p(1)$ using arguments analogous to the ones used in previous cases. For the second one, note that

$$\tau_{\max}^{1/2}\left(\left(\mathbf{\Omega}^{-1}(\overline{\sigma})\mathbf{\Sigma}_\kappa\mathbf{\Omega}^{-1}(\overline{\sigma}) - \mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\right)^2\right) = T^{3/2}\tau_{\max}^{1/2}\left(\left(\Sigma^{-1}(\overline{\sigma})\Sigma_\kappa\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\Sigma_\kappa\Sigma_0^{-1}\right)^2\right)$$

Also, since $\left\|\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\right\|_2 = o_p(1)$ and previous results show that $\exists c_\tau < \infty$ such that $\tau_{\max}\left(\Sigma_0^{-1}\right) \leq c_\tau$, $\tau_{\max}\left(\Sigma^{-1}(\overline{\sigma})\right) \leq c_\tau$ and $\tau_{\max}(\Sigma_\kappa) \leq c_\tau$ for $\kappa = 1, 2, 3$,

$$\begin{aligned}
\tau_{\max}^{1/2}\left(\left(\Sigma^{-1}(\overline{\sigma})\Sigma_\kappa\Sigma^{-1}(\overline{\sigma})\right.\right. &\left.\left. - \Sigma_0^{-1}\Sigma_\kappa\Sigma_0^{-1}\right)^2\right) \\
&= \left\|\left(\Sigma^{-1}(\overline{\sigma})\Sigma_\kappa\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\Sigma_\kappa\Sigma_0^{-1}\right)\right\|_2 \\
&\leq \left\|\left(\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\right)\Sigma_\kappa\Sigma^{-1}(\overline{\sigma})\right\|_2 + \left\|\Sigma_0^{-1}\Sigma_\kappa\left(\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\right)\right\|_2 \\
&\leq \left\|\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\right\|_2\left\|\Sigma_\kappa\right\|_2\left(\left\|\Sigma^{-1}(\overline{\sigma})\right\|_2 + \left\|\Sigma_0^{-1}\right\|_2\right) \\
&\leq \left\|\Sigma^{-1}(\overline{\sigma}) - \Sigma_0^{-1}\right\|_2 c_\tau = o_p(1)
\end{aligned}$$

Thus, $\tau_{\max}^{1/2}\left(\left(\mathbf{\Omega}^{-1}(\overline{\sigma})\mathbf{\Sigma}_\kappa\mathbf{\Omega}^{-1}(\overline{\sigma}) - \mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\right)^2\right) = o_p(1)$.

Further, let c be an arbitrary column vector in $\mathbb{R}^{4K+2}$. Then, by the Cauchy-Schwarz inequality, Lemma 3.2, the fact that $\frac{1}{nT}\left|c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right| = O_p(1)$ and $\frac{1}{nT}\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)'\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right) = O_p(1)$

(which can be proved following the same steps as in Lemmas 3.12 and 3.13),

$$
\begin{aligned}
&\frac{1}{nT}\left[c'\widetilde{\mathbf{X}}'\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right] \\
&\leq\left[\frac{1}{nT}c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right]^{1/2}\left[\frac{1}{nT}\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)'\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right)\right. \\
&\qquad\qquad\left.\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\Omega_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]^{1/2} \\
&\leq\tau_{\max}^{1/2}\left(\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right)^2\right)\left[\frac{1}{nT}c'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}c\right]^{1/2}\left[\frac{1}{nT}\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)'\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]^{1/2},
\end{aligned}
$$

so that, given the previous result showing that $\tau_{\max}^{1/2}\left(\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\Sigma_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right)^2\right)=o_p\left(1\right)$, $\frac{1}{\overline{\sigma}_{\varepsilon}^2}\frac{1}{nT}\left[\frac{1}{\overline{\sigma}_{\varepsilon}^2}\widetilde{\mathbf{X}}'\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right)\boldsymbol{\eta}\left(\overline{\theta},\overline{\lambda}\right)\right]=o_p\left(1\right)$.

Finally, to prove the $(\sigma_{\kappa},\sigma_{\varrho})$ case notice that $\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\sigma_{\kappa}\partial\sigma_{\varrho}}-\frac{\partial^2\mathcal{L}\left(\psi_0\right)}{\partial\sigma_{\kappa}\partial\sigma_{\varrho}}\right]$ can be expressed as

$$
\begin{aligned}
&\frac{1}{2}\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\varrho}-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\varrho}\right]+ \\
&\qquad\frac{1}{nT}\left[\frac{1}{\sigma_{\varepsilon0}^2}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\varrho}\Omega_0^{-1}\boldsymbol{\eta}-\frac{1}{\overline{\sigma}_{\varepsilon}^2}\boldsymbol{\eta}\left(\overline{\lambda},\overline{\theta}\right)'\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\varrho}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\boldsymbol{\eta}\left(\overline{\lambda},\overline{\theta}\right)\right]
\end{aligned}
$$

Note also that $\tau_{\max}^{1/2}\left(\left(\left[\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right]\right)^2\right)=o_p\left(1\right)$ and, given that $\mathbf{\Sigma}_{\kappa}$ is u.b.r.c.s., $tr\left(\mathbf{\Sigma}_{\kappa}^2\right)=O\left(nT\right)$. Then, $\frac{1}{nT}tr\left[\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\varrho}-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\varrho}\right]\leq\frac{1}{nT}\tau_{\max}^{1/2}\left(\left(\left[\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)-\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\right]\right)^2\right)(nT)^{1/2}tr^{1/2}\left(\mathbf{\Sigma}_{\varrho}^2\right)=o_p\left(1\right)O\left(1\right)=o_p\left(1\right)$ and

$$
\begin{aligned}
&\tau_{\max}^{1/2}\left(\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\varrho}\mathbf{\Omega}_0^{-1}-\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\varrho}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\right)^2\right) \\
&\quad\leq\left\|\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}_0^{-1}-\Omega^{-1}\left(\overline{\sigma}\right)\mathbf{\Sigma}_{\kappa}\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\right\|_2\tau_{\max}\left(\mathbf{\Sigma}_{\varrho}\right)\tau_{\max}\left(\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\right) \\
&\quad+\left\|\mathbf{\Omega}_0^{-1}-\mathbf{\Omega}^{-1}\left(\overline{\sigma}\right)\right\|_2\tau_{\max}\left(\mathbf{\Omega}_0^{-1}\right)^2\tau_{\max}\left(\mathbf{\Sigma}_{\kappa}\right)\tau_{\max}\left(\mathbf{\Sigma}_{\varrho}\right)=o_p\left(1\right)O(1).
\end{aligned}
$$

Therefore, using arguments analogous to the ones used in previous cases, $\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}\left(\overline{\psi}\right)}{\partial\sigma_{\kappa}\partial\sigma_{\varrho}}-\frac{\partial^2\mathcal{L}\left(\psi_0\right)}{\partial\sigma_{\kappa}\partial\sigma_{\varrho}}\right]=o_p(1)$.

We conclude the proof by noting that the $o_p\left(1\right)$ of the other components of $\frac{1}{nT}\left[\frac{\partial^2\mathcal{L}\left(\psi\right)}{\partial\psi\partial\psi'}-\frac{\partial^2\mathcal{L}\left(\psi\right)}{\partial\psi\partial\psi'}\right]$ can be proved using previous results and arguments analogous to the ones used in the cases considered here. We consequently omit the details of these proofs. $\square$

3.8 Appendix C: Gradient and Hessian of the QML function

3.8.1 Gradient

The gradient function $\bigtriangledown\mathcal{L}(\psi_0)=\dfrac{\partial\mathcal{L}(\psi_0)}{\partial\psi}$ has the following elements:

$$\begin{aligned}
\frac{\partial\mathcal{L}(\psi_0)}{\partial\theta} &= \frac{1}{\sigma^2_{\varepsilon_0}}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\\
\frac{\partial\mathcal{L}(\psi_0)}{\partial\sigma^2_{\varepsilon_0}} &= -\frac{nT}{2\sigma^2_{\varepsilon_0}}+\frac{1}{2\sigma^4_{\varepsilon_0}}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\\
\frac{\partial\mathcal{L}(\psi_0)}{\partial\sigma_\kappa} &= -\frac{1}{2}tr\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right)+\frac{1}{2\sigma^2_{\varepsilon_0}}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}\\
\frac{\partial\mathcal{L}(\psi_0)}{\partial\lambda} &= -tr\left(\mathbf{S}^{-1}\mathbf{W}\right)+\frac{1}{\sigma^2_{\varepsilon}}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}
\end{aligned}$$

where $\kappa=1,2,3$ and $\mathbf{\Sigma}_\kappa=\dfrac{\partial\mathbf{\Omega}(\sigma)}{\partial\sigma_\kappa}$. Also, $\mathbf{\Sigma}_1=J_T\otimes\Sigma_1=J_T\otimes I_n$, $\mathbf{\Sigma}_2=J_T\otimes\Sigma_2=J_T\otimes(W_n+W_n')$ and $\mathbf{\Sigma}_3=J_T\otimes\Sigma_3=J_T\otimes W_nW_n'$.

3.8.2 Hessian matrix

The Hessian of the likelihood function in 3.3.2 is:

$$\boldsymbol{H}_n(\psi_0)=\begin{pmatrix}
\dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\theta'} & \dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\sigma^2_\varepsilon} & \dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\theta\partial\delta'}\\
 & \dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\sigma^2_\varepsilon\partial\sigma^2_\varepsilon} & \dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\sigma^2_\varepsilon\partial\delta'}\\
 & & \dfrac{\partial^2\mathcal{L}(\psi_0)}{\partial\delta\partial\delta'}
\end{pmatrix}$$

Next we provide detailed results for each row of the Hessian matrix. Thus, the first row of the Hessian matrix is

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\theta'} &= -\frac{1}{\sigma_\varepsilon^2}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\tilde{\mathbf{X}} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\sigma_\varepsilon^2} &= -\frac{1}{\sigma_\varepsilon^4}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\sigma_\kappa} &= -\frac{1}{\sigma_\varepsilon^2}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\theta\partial\lambda} &= -\frac{1}{\sigma_\varepsilon^2}\tilde{\mathbf{X}}'\mathbf{\Omega}_0^{-1}\mathbf{W}\mathbf{Y},
\end{aligned}$$

while the second row of the Hessian matrix is

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\varepsilon^2\partial\sigma_\varepsilon^2} &= \frac{nT}{2\sigma_\varepsilon^4} - \frac{1}{\sigma_\varepsilon^6}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\varepsilon^2\partial\lambda} &= -\frac{1}{\sigma_\varepsilon^4}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\varepsilon^2\partial\sigma_\kappa} &= -\frac{1}{2\sigma_\varepsilon^4}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta}
\end{aligned}$$

and the third row of the Hessian matrix is

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\lambda\partial\lambda} &= -tr\left(\left(\mathbf{S}_0^{-1}\mathbf{W}\right)^2\right) - \mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\mathbf{W}\mathbf{Y} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\lambda\partial\sigma_\kappa} &= -\frac{1}{\sigma_\varepsilon^2}\mathbf{Y}'\mathbf{W}'\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\kappa\partial\sigma_\kappa} &= \frac{1}{2}tr\left[\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right)^2\right] - \frac{1}{\sigma_\varepsilon^2}\boldsymbol{\eta}'\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right)^2\mathbf{\Omega}_0^{-1}\boldsymbol{\eta} \\
\frac{\partial^2 \mathcal{L}(\psi_0)}{\partial\sigma_\kappa\partial\sigma_\varrho} &= \frac{1}{2}tr\left(\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\varrho\right) \\
&\quad - \frac{1}{2\sigma_\varepsilon^2}\boldsymbol{\eta}'\mathbf{\Omega}_0^{-1}\left[\mathbf{\Sigma}_\kappa\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\varrho + \mathbf{\Sigma}_\varrho\mathbf{\Omega}_0^{-1}\mathbf{\Sigma}_\kappa\right]\mathbf{\Omega}_0^{-1}\boldsymbol{\eta},
\end{aligned}$$

with $\kappa \neq \varrho$ and $\varrho = 1, 2, 3$.

Chapter 4

Growth, heterogeneous technological interdependence, and spatial externalities: Theory and Evidence [1]

4.1 Introduction

Historically, the empirical economic growth literature consisted mostly of "aspatial empirical analyses that have ignored the influence of spatial location on the process of growth" (De Long and Summers, 1991; Fingleton and López-Bazo, 2006, p. 178). In the last two decades, however, a number of studies seek to incorporate "spatial effects" in the standard (i.e., non-spatial) economic growth models. In particular, the idea that the spatial location of an economy may drive its economic growth has been developed using models of absolute location, which account for the location of one economy in the geographical space, and models of relative location, which account for the location of one economy with respect to the others. Econometrically, these two types of models are closely related to the concepts of spatial heterogeneity and spatial dependence (Abreu et al., 2005).

Although spatial heterogeneity is usually associated with parameter heterogeneity (see e.g. Ertur and Koch, 2007; Basile, 2008), the most common approach in the literature is to allow for unobserved differences using panel data (Islam, 1995; Elhorst et al., 2010). Also, knowledge spillovers are the main mechanism employed to incorporate interactions between economies into the Solow-Swan neoclassical growth model (López-Bazo et al., 2004; Egger and Pfaffermayr, 2006; Pfaffermayr, 2009, 2012). It is interesting to note, however, that these two streams of the literature have developed rather separately. Notable exceptions include Elhorst et al. (2010), who consider the extension of the model proposed by Ertur and Koch

[1] This chapter is co-authored with Miguel Manjón-Antolín and Oscar Martínez-Ibáñez.

(2007) to panel data; Ho et al. (2013), who consider an ad-hoc extension of the model proposed by Mankiw et al. (1992) that includes a spatial autoregressive term and a spatial time lag term; and Yu and Lee (2012), who, using a simplified version of the technology assumed by Ertur and Koch (2007), derive a growth model with spatial externalities based on the model of Mankiw et al. (1992). This paper aims to contribute to this limited literature by considering a growth model with spatial heterogeneity and spatial externalities that nests the models introduced by Islam (1995), López-Bazo et al. (2004) and Ertur and Koch (2007).

To be precise, we present a growth model with interdependencies in the (heterogeneous) technological progress, physical capital and stock of knowledge.[2] The basic framework is similar to that of Ertur and Koch (2007), but we consider additional sources of externalities across economies. While they assume that the technological progress depends on the own stock of physical capital and the stock of knowledge of the other economies, we also consider the role of both the physical capital (López-Bazo et al., 2004; Egger and Pfaffermayr, 2006) and the (unobserved) initial level of technology (De Long and Summers, 1991; LeSage and Fischer, 2012) of the other economies. Moreover, we do not assume a common exogenous technological progress but account for heterogeneity in the initial level of technology, which here is interpreted as a proxy for total factor productivity (Islam, 1995).

Having presented our model, we then derive the steady-state equation and a growth-initial equation that can be taken to the data. This is where the generality of our model comes at a cost, since not all the parameters of interest are identified (a limitation that also arises in the benchmark model of Ertur and Koch 2007). To be precise, the model contains a set of theoretical constraints that, if valid, allows us to identify most of the implied parameters (in contrast, the constrained version of Ertur and Koch's (2007), which is derived from a subset of our constraints, is fully identified). This means that, provided that the constrained model is valid, finding evidence of global technological interdependence between the economies would lead us to reject the models of Islam (1995) and López-Bazo et al. (2004), whereas, if no such evidence was found, we would reject the model of Ertur and Koch (2007). In any case, finding evidence of heterogeneity in the initial level of technology would support our model against these alternatives.

Ultimately, the identification problem arises because, even in the constrained model, we cannot separate: *i)* the (direct) effect that, as an input of the production function, the stock of physical capital has on the output from the (indirect) effect that it has as a driver of the technology; and *ii)* the local) effect that the stock of physical capital of the neighbouring economies has on the technology and, subsequently, the output from the

[2]It is worth noting that the model can easily be extended to incorporate the role of human capital (López-Bazo et al., 2004; Fingleton and López-Bazo, 2006). We leave this issue for future research.

(global) effect that, due to the presence of technological interdependencies, the stock of physical capital of the neighbouring economies has on the output.[3] Still, we argue that simple changes in the unconstrained model specification (e.g., introducing the stock of physical capital lagged one period in the technological progress, rather than using its current value) and/or additional restrictions on the parameters of the constrained model (e.g., the equality of the effects that the stock of physical capital has on the output as a production input and as a technology input) may address this limitation. To illustrate our argument, we consider imposing the additional constraint(s) that the remaining unidentified technological parameters are consistent with the model of López-Bazo et al. (2004) and/or that of Ertur and Koch (2007).

The econometric specification of the resulting growth-initial equation corresponds to the spatial Durbin dynamic panel model (see also Elhorst et al., 2010; Yu and Lee, 2012; Ho et al., 2013), but with spatially weighted individual-specific effects. Thus, given the obvious interest in distinguishing the individual effects from their spatial spillovers, we resort to a correlated random effects specification (Miranda et al., 2017a,b). In particular, we estimate our growth-initial equation by Quasi-Maximum Likelihood (see also Lee and Yu, 2016) using EU-NUTS2 regional data from Cambridge Econometrics. We use regional data because, as López-Bazo et al. (2004, p. 43) argue, once it is taken on board that "[e]conomies interact with each other (...), linkages are [likely] to be stronger [between close-by regions] than across heterogeneous countries". We provide results for both the constrained and unconstrained versions of our model.

We find evidence of technological interdependence in the output per capita of the EU regions, that is, a positive and significant impact of the level of technology of the neighbouring regions. However, there is also evidence of "unobserved" technological interdependence in the EU regions (i.e., local spatial contagion of the "unobserved productivity"), which supports our assumed technology. Also, the constrained model specification produces estimates of the implied parameters that statistically reject the models of Islam (1995) and López-Bazo et al. (2004). Lastly, results from our identification strategy do not support the role that López-Bazo et al. (2004) and Ertur and Koch (2007) assume capital plays in shaping the technological progress.

The rest of the paper is organised as follows. In Section 4.2 we present the model.[4] In

[3] Anselin (2003, p. 154) is generally credited for "distinguishing between a *global* and a *local* range of dependence and the way in which this translates into the incorporation in a regression specification of spatially lagged dependent variables (Wy) [and] spatially lagged explanatory variables (WX)".

[4] To facilitate the reading, we moved the derivation of some results (the balanced growth rate, a Taylor approximation to the marginal productivity of capital, the speed of convergence, and the solution to the steady state differential equation) to the appendix.

Section 4.3 we discuss the data and the estimation results. Section 4.4 concludes.

4.2 The Model

4.2.1 Technological interdependencies in growth

Our starting point is the Solow growth model originally proposed by Mankiw et al. (1992) using cross-section data and extended later by Islam (1995) to panel data (see also Barro and Sala-i-Martin, 2003). Let us then consider a Cobb-Douglas production function for region $i = 1, \ldots, N$ in time $t = 1, \ldots, T$:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{1-\alpha}, \tag{4.2.1}$$

where Y_{it} denotes output, K_{it} physical capital (α is thus the capital share or output elasticity parameter), L_{it} labour, and A_{it} technology. All the variales are in levels and there are constant returns to scale in production. Also, while output, capital and labour are typically assumed to be observable, technology is assumed to be (partially) unobservable. Mankiw et al. (1992), for example, assume that $\ln A = a + \varepsilon$, where a is a constant term and ε is the standard i.i.d error.

For the purposes of this paper, a major feature of this model is that technology is assumed to grow exogenously and at the same rate in all regions. This rules out the existence of knowledge spillovers arising from technological interdependence between the regional economies. However, accounting for technological interdependence and knowledge spillovers is critical when analysing how "the relative location of an economy affects economic growth" (Elhorst et al., 2010, p. 338). In the literature, depending on whether knowledge spillovers turn out to be "local" or "global" (Anselin, 2003), we find two main approaches to the introduction of spatial externalities in the Solow growth model.

On the one hand, López-Bazo et al. (2004) and Egger and Pfaffermayr (2006) consider growth models where the knowledge spillovers are local in nature, in the sense that they are limited to the neighbouring regions (at least initially).[5] To be precise, in López-Bazo et al. (2004) technology is assumed to depend on both the physical and human capital of the neighbouring regions, whereas in Egger and Pfaffermayr (2006) is assumed to grow exogenously and at the same rate in all regions (as in Mankiw et al. 1992 and Islam 1995), so that the externalities arise from the assumption that total factor productivity depends

[5] See also Fingleton and López-Bazo (2006), Pfaffermayr (2009) and Pfaffermayr (2012).

on the capital-labour ratio of the region and the spatially weighted capital-labour of the other regions. Ertur and Koch (2007), on the other hand, assume that the technological progress of an economy depends on the stock of physical capital per worker in that economy as well as the stock of knowledge of the other economies. More specifically, they assume that the technology of an economy is a geometrically weighted average of the technology of the other economies, thus making knowledge spillovers to spread over all the regions (and hence become "global"). However, it is still assumed that "some proportion of technological progress is exogenous and identical in all countries" [p. 1036].

In this paper, we extend the model of Ertur and Koch (2007) by introducing spatial dependence in the stock of capital, as well as heterogeneity and spatial dependence in the exogenous technological progress (while holding the assumption that the technological progress of an economy depends on the stock of knowledge of the other economies). In this vein, our assumed technology combines the alternative sources of spatial externalities considered in models of relative location with the unobserved heterogeneity that characterises the models of absolute location (Abreu et al., 2005). In particular, our model shares with that of Ertur and Koch (2007) the main source of parameter heterogeneity. Namely, the speed of convergence to the steady state, as discussed below. Yet we eventually estimate a constrained version in which the speed of convergence is identical for all economies (Elhorst et al., 2010; Yu and Lee, 2012). To be precise, the estimated econometric specification corresponds to a variant of the spatial Durbin dynamic panel model recently considered by Lee and Yu (2016) that includes not only individual-specific effects but also their spatial spillovers (Miranda et al., 2017a).[6]

Next we derive our empirical specification, which adopts the form of a growth-initial equation. To a large extent, our approach follows the steps of Ertur and Koch (2007). Thus, we first discuss and motivate the assumed technology, then we obtain the output per worker equation at the steady state, and finally the growth-initial equation.

[6] As Basile (2008, p. 532-533) points out, "the local Spatial Durbin Model (...) proposed by Ertur and Koch (2007) is a general and flexible specification, since it allows identification of both spatial-interaction effects and parameter heterogeneity (...). In essence, this is the model considered here. The global Spatial Durbin Model (...) represents a less general specification, because it imposes the restriction of parameter homogeneity". In essence, this is the model we estimate. Lastly, "[t]he model proposed by López-Bazo et al. (2004) (..) imposes a further restriction on the parameters since the spatial lags of the structural characteristics of the regions are not included" (this also applies to the model proposed by Egger and Pfaffermayr 2006). In essence, this model is nested in ours.

4.2.2 Technology

Let us denote by Ω_{it} the exogenous technological progress and by $k_{it} = \frac{K_{it}}{L_{it}}$ the level of physical capital per worker (of region i in period t). Ertur and Koch (2007, p. 1036) assume that the technology of region i in period t is given by

$$A_{it} = \Omega_{it} k_{it}^{\phi} \prod_{j \neq i}^{N} A_{jt}^{\gamma w_{ij}}, \tag{4.2.2}$$

where "[t]he parameter ϕ describes the strength of home externalities generated by physical capital accumulation" and "the degree of [regional] technological interdependence generated by the level of spatial externalities is described by γ". Notice that the spatial relation between region i and its neighbouring regions is represented by a set of spatial weights or "exogenous friction terms" w_{ij}, with $j = 1, \ldots, N$, that are assumed to satisfy the following properties: $w_{ij} = 0$ if $i = j$, $0 \leq w_{ij} \leq 1$, and $\sum_j w_{ij} = 1$. Lastly, Ertur and Koch (2007) assume that $\Omega_{it} = \Omega_t = \Omega_0 \exp(\mu t)$, where μ is the constant rate of growth of the exogenous technological progress. Therefore, the technology eventually assumed is $A_{it} = \Omega_0 \exp(\mu t) k_{it}^{\phi} \prod_{j \neq i}^{N} A_{jt}^{\gamma w_{ij}}$.

However, as previously pointed out, there are alternative approaches to the inclusion of knowledge spillovers in the Solow model. In a series of papers, López-Bazo et al. (2004, p. 46), Egger and Pfaffermayr (2006), Fingleton and López-Bazo (2006) and Pfaffermayr (2009, 2012) argue that the physical (and human) capital may be an alternative source of externalities, "[t]he reasoning behind such spillovers [being] basically the diffusion of technology from other regions caused by investments in physical (...) capital". In mathematical terms, such a technology may adopt the following functional form:

$$A_{it} = \Omega_0 \exp(\mu t) \prod_{j \neq i}^{N} k_{jt}^{\gamma w_{ij}}, \tag{4.2.3}$$

where, for the sake of comparability, we have used the same notation as in 4.2.2. However, the interpretation of the parameter γ is different here, for it now "measures the [strength of the] externality across economies" originated from variations in physical capital (López-Bazo et al., 2004; Fingleton and López-Bazo, 2006, p. 46). It is also important to stress that these papers maintain the assumption of an homogeneous exogenous technological progress growing at a constant rate, i.e., $\Omega_{it} = \Omega_0 \exp(\mu t)$.

Our assumed technology features those displayed in 4.2.2 and 4.2.3. However, we depart from these studies in the assumptions they made with respect to the exogenous technological progress. First, they assume that it is homogeneous across regions. However, as Mankiw et al.

(1992, p. 6) point out, the Ω_0 "term reflects not just technology but resource endowments, climate, institutions, and so on; it may therefore differ across countries". In line with this argument, we introduce regions' heterogeneity into the definition of the exogenous technological progress by assuming that $\Omega_{it} = \Omega_{i0}\exp(\mu t)$.[7]

Second, as Islam (1995, p. 1149) points out, Ω_{i0} "is an important source of parametric difference in the aggregate production function across [regions]". Econometrically, it can be interpreted as an individual-specific effect (possibly correlated with some of the covariates in the initial-growth specification eventually derived). Economically, it is "a measure of efficiency with which the [regions] are transforming their capital and labor resources into output and hence is very close to the conventional concept of total factor productivity" [p. 1155-1156]. These arguments are behind the second twist we introduce with respect to the models of López-Bazo et al. (2004), Ertur and Koch (2007) and others, since it opens the door to considering productivity spillovers as an additional source of spatial externalities (LeSage and Fischer, 2012; Miranda et al., 2017b). As De Long and Summers (1991, p. 487) point out, "it is difficult to believe that Belgian and Dutch or US and Canadian economic growth would ever significantly diverge, or that substantial productivity gaps would appear within Scandinavia".

All in all, a production technology that may account for these alternative sources of spatial dependence is the following:

$$A_{it} = \Omega_{it} \prod_{j \neq i}^{N} \Omega_{jt}^{\gamma_1 w_{ij}} k_{it}^{\phi} \prod_{j \neq i}^{N} k_{jt}^{\gamma_2 w_{ij}} \prod_{j \neq i}^{N} A_{jt}^{\gamma_3 w_{ij}} \tag{4.2.4}$$

with $\Omega_{it} = \Omega_{i0}\exp(\mu t)$ and Ω_{i0} non-observable (which is why Ω_{it} does not have a coefficient in 4.2.4). Notice that γ_3 and γ_2 play the same role as γ in 4.2.2 and 4.2.3, respectively, whereas γ_1 can be interpreted as the degree of technological interdependence generated from the (unobserved) productivity spillovers. In particular, $\gamma_1 = \phi = \gamma_2 = \gamma_3 = 0$ would lead us to the model proposed by Islam (1995), $\gamma_1 = \phi = \gamma_3 = 0$ (with $\gamma_2 \neq 0$) to the model proposed by López-Bazo et al. (2004), and $\gamma_1 = \gamma_2 = 0$ (with $\phi \neq 0$ and $\gamma_3 \neq 0$) to the model proposed by Ertur and Koch (2007). Notice also that, in contrast to the local contagion models of López-Bazo et al. (2004) and Egger and Pfaffermayr (2006), both ours and that of Ertur and Koch (2007) are models of global contagion (Anselin, 2003). We differ, however, in that whereas in their case there are no (global) spatial externalities in the stock of knowledge unless $\gamma_3 \neq 0$, there still are here if either $\gamma_1 \neq 0$ or $\gamma_2 \neq 0$ (albeit of a local nature). This

[7] Alternative ways of modelling the exogenous technological progress are $\Omega_{it} = \Omega_0\exp(\mu_i t)$ and $\Omega_{it} = \Omega_{i0}\exp(\mu_i t)$. However, these proposals would considerably increase the number of parameters of the model (by more than N, since it can be shown that the balanced growth rate becomes heterogeneous too) and make identification difficult, if not impossible (Lee and Yu, 2016).

is because our model accounts for both global and local contagion. Lastly, it is interesting to note that there are no capital externalities in our model if $\phi = \gamma_2 = 0$ (neither global nor local). This is because our model accounts for both the role of the own physical capital (Ertur and Koch, 2007) and that of the other economies (López-Bazo et al., 2004) in the technological progress.

4.2.3 The production function

In order to obtain the explicit form of the Cobb-Douglas production function in 4.2.1 given our assumed technology, let us consider 4.2.4 expressed in logs and matrix form:

$$\begin{aligned} A &= \Omega + \gamma_1 W\Omega + \phi k + \gamma_2 Wk + \gamma_3 WA \\ &= (I-\gamma_3 W)^{-1}\Omega + \gamma_1 (I-\gamma_3 W)^{-1} W\Omega + \phi (I-\gamma_3 W)^{-1} k + \gamma_2 (I-\gamma_3 W)^{-1} Wk \end{aligned} \tag{4.2.5}$$

where the parameters γ_1, γ_2 and γ_3 have been previously described (in particular, it is now assumed that $1/\gamma_3$ is not an eigenvalue of W when $\gamma_3 \neq 0$), A is the $N \times 1$ vector of logarithms of the technology, I is the $N \times N$ identity matrix, $\Omega = \Omega_0 + \iota_N \mu t$ is the $N \times 1$ vector of logarithms of the exogenous technological progress with $\Omega_0 = (\ln \Omega_{10}, \dots, \ln \Omega_{N0})'$ and ι_N being a $N \times 1$ vector of ones, k is the $N \times 1$ vector of of logarithms of the capital per worker, and W is the $N \times N$ spatial weight matrix that describes the spatial arrangement of the regions.

Le us now denote by $w_{ij}^{(r)}$ the row i and column j element of matrix W^r. Notice that, since W is assumed to be row-normalized, if all the eigenvalues of W lie in the interval $(-1, 1)$ and $|\gamma_3| < 1$, then $(I-\gamma_3 W)^{-1} = \sum_{r=0}^{\infty} \gamma_3^r W^r$. Thus,

$$\begin{aligned} \ln A_{it} &= \sum_{j=1}^{N} \sum_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r)} \ln \Omega_{jt} + \gamma_1 \sum_{j=1}^{N} \sum_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r+1)} \ln \Omega_{jt} + \phi \sum_{j=1}^{N} \sum_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r)} \ln k_{jt} \\ &\quad + \gamma_2 \sum_{j=1}^{N} \sum_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r+1)} \ln k_{jt} \\ &= \sum_{j=1}^{N} \ln \Omega_{jt}^{\sum_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r)}} + \sum_{j=1}^{N} \ln \Omega_{jt}^{\frac{\gamma_1}{\gamma_3} \sum_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}} + \sum_{j=1}^{N} \ln k_{jt}^{\phi \sum_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r)}} + \sum_{j=1}^{N} \ln k_{jt}^{\frac{\gamma_2}{\gamma_3} \sum_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}}, \end{aligned}$$

so that we may rewrite 4.2.5 as

$$A_{it} = \prod_{j=1}^{N} \Omega_{jt}^{\sum\limits_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r)}} \prod_{j=1}^{N} \Omega_{jt}^{\frac{\gamma_1}{\gamma_3} \sum\limits_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}} \prod_{j=1}^{N} k_{jt}^{\phi \sum\limits_{r=0}^{\infty} \gamma_3^r w_{ij}^{(r)}} \prod_{j=1}^{N} k_{jt}^{\frac{\gamma_2}{\gamma_3} \sum\limits_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}}$$
$$= \Omega_{it}^{1+\left(\frac{\gamma_3+\gamma_1}{\gamma_3}\right) \sum\limits_{r=1}^{\infty} \gamma_3^r w_{ii}^{(r)}} \prod_{j\neq i}^{N} \Omega_{jt}^{\left(\frac{\gamma_3+\gamma_1}{\gamma_3}\right) \sum_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}} k_{it}^{\phi+\left(\frac{\phi\gamma_3+\gamma_2}{\gamma_3}\right) \sum_{r=1}^{\infty} \gamma_3^r w_{ii}^{(r)}} \prod_{j\neq i}^{N} k_{jt}^{\left(\frac{\phi\gamma_3+\gamma_2}{\gamma_3}\right) \sum_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}}$$

by using $\prod_{j=1}^{N} \Omega_{jt}^{w_{ij}^{(0)}} = \Omega_{it}$ and $\prod_{j=1}^{N} k_{jt}^{\phi w_{ij}^{(0)}} = k_{it}^{\phi}$.

Also, let us now define $u_{ii} = \alpha + \phi + \left(\frac{\phi\gamma_3+\gamma_2}{\gamma_3}\right) \sum_{r=1}^{\infty} \gamma_3^r w_{ii}^{(r)}$ and $u_{ij} = \left(\frac{\phi\gamma_3+\gamma_2}{\gamma_3}\right) \sum_{r=1}^{\infty} \gamma_3^r w_{ij}^{(r)}$, with $u_{ii} + \sum_{j\neq i}^{N} u_{ij} = \sum_{j=1}^{N} u_{ij} = \alpha + \phi + \frac{\phi\gamma_3+\gamma_2}{1-\gamma_3} = \alpha + \frac{\phi+\gamma_2}{1-\gamma_3}$.
Then, given that $y_{it} = A_{it} k_{it}^{\alpha}$,

$$y_{it} = \Omega_{it}^{1+\left(\frac{(\gamma_3+\gamma_1)(u_{ii}-\alpha-\phi)}{(\phi\gamma_3+\gamma_2)}\right)} \prod_{j\neq i}^{N} \Omega_{jt}^{\frac{(\gamma_3+\gamma_1)u_{ij}}{\phi\gamma_3+\gamma_2}} k_{it}^{u_{ii}} \prod_{j\neq i}^{N} k_{jt}^{u_{ij}} \tag{4.2.6}$$

Notice that "this model implies spatial heterogeneity in the parameters of the production function", a feature shared with that of Ertur and Koch (2007, p. 1037). We differ, however, in that it is no longer the case that "if there are no physical capital externalities, i.e., $\phi = 0$, we have $u_{ii} = \alpha$ and $u_{ij} = 0$, (...) then the production function is written in the usual form" (as in e.g. Mankiw et al. 1992 and Islam 1995). As previously pointed out, here we further require that $\gamma_1 = \gamma_2 = \gamma_3 = 0$. Put it differently, there are no physical capital externalities in the model of Ertur and Koch (2007) if $\phi = 0$, *regardless of* γ_3. In our model, however, we further require that $\gamma_2 = 0$. That is, there are capital externalities to the extent that $\gamma_3 \neq 0$ and either $\phi \neq 0$ or $\gamma_2 \neq 0$. This is because, following López-Bazo et al. (2004), we account for the local role of the capital in the technology through the parameter γ_2.

4.2.4 The Steady State equation

To derive the equation describing the output per worker of region i at the steady state, we proceed in the following way. First we rewrite the production function in matrix form, $y = A + \alpha k$, and substitute the technology by its expression in 4.2.5. We then pre-multiply both sides of the resulting equation by $I - \gamma_3 W$ to obtain

$$y = \Omega + \gamma_1 W\Omega + (\alpha + \phi)k + (\gamma_2 - \alpha\gamma_3)Wk + \gamma_3 Wy \tag{4.2.7}$$

Lastly, we replace in 4.2.7 the log of the capital per worker in region i by its log value at the steady state, $\ln k^*_{it}$. To this end, we start by noting that the evolution of capital is governed by the following dynamic equation:

$$\dot{k}_{it} = s_i y_{it} - (n_i + \delta) k_{it} \tag{4.2.8}$$

where the dot over a variable denotes its derivative with respect to time, s_i is the fraction of output saved, n_i is the growth rate of labour, and δ is the annual rate of depreciation of capital (common to all regions). Given that production shows decreasing returns to scale, equation 4.2.8 implies that the capital-output ratio is constant and converges to a balanced growth rate g defined by $\frac{\dot{k}_{it}}{k_{it}} = ln\,\dot{y}_{it} = ln\,\dot{k}_{it} = g = \frac{\mu(1+\gamma_1)}{(1-\gamma_3)(1-\alpha)-\phi-\gamma_2}$ (see appendix 4.5). Also, it can be shown that, given a balanced growth rate g and 4.2.8 (see e.g. Barro and Sala-i-Martin, 2003), $\frac{k^*_{it}}{y^*_{it}} = \frac{s_i}{n_i+\delta+g}$ and $\ln k^*_{it} = \ln y^*_{it} + \ln\left(\frac{s_i}{n_i+\delta+g}\right)$.[8]

What is thus left is to introduce in 4.2.7 (rewritten for economy i rather than in matrix form) the expression obtained for the log of the capital per worker in region i at the steady state. In doing so, we obtain the equation describing the output per worker of region i at the steady state:

$$\begin{aligned}\ln y^*_{it} = {} & \frac{\ln \Omega_{it}}{1-\alpha-\phi} + \frac{\gamma_1}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln \Omega_{jt} + \frac{\alpha+\phi}{1-\alpha-\phi}\ln\left(\frac{s_i}{n_i+\delta+g}\right) \\ & + \frac{\gamma_2-\alpha\gamma_3}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\left(\frac{s_j}{n_j+\delta+g}\right) + \frac{(1-\alpha)\gamma_3+\gamma_2}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln y^*_{jt}\end{aligned} \tag{4.2.9}$$

Notice that this equation differs from that obtained by Ertur and Koch (2007) in two main features, reflecting ultimately differences in the assumed technology. First, the heterogeneous exogenous technological progress, since Ω_{it} is assumed to be Ω_t in Ertur and Koch (2007) and, consequently, no exogenous technological interdependences are considered. In particular, the term $\frac{\gamma_1}{(1-\alpha-\phi)}\sum_{j=1}^{N} w_{ij}\ln\Omega_{jt}$ is missing in their steady state equation. Second, the relation between the output per worker of an economy at the steady state and that of its neighbours,

[8] It is also interesting to note that, if we compute the marginal productivity of capital, $\frac{\dot{k}_{it}}{k_{it}} = s_i\frac{y_{it}}{k_{it}} - (n_i+\delta)$, using the expression defining y_{it} in 4.2.6, we obtain $\frac{\dot{k}_{it}}{k_{it}} = s_i \Omega_{it}^{1+\left(\frac{(\gamma_3+\gamma_1)(u_{ii}-\alpha-\phi)}{(\phi\gamma_3+\gamma_2)}\right)} \prod_{j\neq i}^{N} \Omega_{jt}^{\frac{(\gamma_3+\gamma_1)u_{ij}}{\phi\gamma_3+\gamma_2}} k_{it}^{u_{ii}-1} \prod_{j\neq i}^{N} k_{jt}^{u_{ij}} - (n_i+g)$. Therefore, provided that $\alpha + \frac{\phi+\gamma_2}{1-\gamma_3} < 1$, there are diminishing returns to the capital, as in the model of Ertur and Koch (2007).

$\frac{(1-\alpha)\gamma_3+\gamma_2}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln y_{jt}^*$. Whereas in the model of Ertur and Koch (2007) there is no global contagion in the output unless $\gamma_3 \neq 0$ (provided of course that $\alpha \neq 1$), here we may still have such contagion to the extent that $\gamma_2 \neq 0$ (as in López-Bazo et al. 2004), *even if* $\gamma_3 = 0$. More generally, these features of our model are also absent in the above mentioned growth studies (López-Bazo et al., 2004; Egger and Pfaffermayr, 2006; Fingleton and López-Bazo, 2006; Pfaffermayr, 2009, 2012).

4.2.5 The growth-initial equation

In the standard, non-spatial growth models (see e.g. Barro and Sala-i-Martin, 2003), the analog of equation 4.2.9 gives an expression for the output per worker in the steady state that does not depend on the output per worker in the steady state of the other economies (i.e., the term $\frac{(1-\alpha)\gamma_3+\gamma_2}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln y_{jt}^*$ is missing). Thus, a log-linear approximation to the dynamics around the steady state using a Taylor expansion produces a growth-initial regression equation that can be estimated using the appropriate method. In our case, however, this approach would produce a rather complex system of first-order differential linear equations whose solution is not directly estimable due to the presence of variables at the steady state (Egger and Pfaffermayr, 2006, for example, approximate them using a set of exogenous variables). In particular, a log linearisation of the marginal productivity of capital, $\frac{\dot{k}_{it}}{k_{it}}$, around the steady state yields (see appendix 4.6)

$$\frac{\dot{k}_{it}}{k_{it}} = g + (u_{ii}-1)(n_i+\delta+g)\left(\ln k_{it} - \ln k_{it}^*\right) + \sum_{j\neq i}^{N} u_{ij}(n_i+\delta+g)\left(\ln k_{jt} - \ln k_{jt}^*\right) \quad (4.2.10)$$

Notice that this result coincides with the one obtained by Ertur and Koch (2007).

To tackle this issue, Ertur and Koch (2007) hypothesise that the differences between the observed and the steady state values of the capital and output per worker across regions correspond to the following expressions:

$$\begin{aligned} \ln y_{it} - \ln y_{it}^* &= \Theta_j \left(\ln y_{jt} - \ln y_{jt}^*\right) \\ \ln k_{it} - \ln k_{it}^* &= \Phi_j \left(\ln k_{jt} - \ln k_{jt}^*\right) \end{aligned} \quad (4.2.11)$$

This yields the following speed of convergence (see appendix 4.7):

$$\frac{d\ln y_{it}}{dt} = g - \lambda_i \left(\ln y_{it} - \ln y_{it}^*\right) \quad (4.2.12)$$

with

$$\lambda_i = \frac{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}(n_j+g+\delta)}{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}} - \sum_{j=1}^{N} u_{ij}(n_j+\delta+g)\frac{1}{\Theta_j} \tag{4.2.13}$$

Solving the differential equation in 4.2.12 for $\ln y_{it}$ (see appendix 4.8) and evaluating the solution at $t = t_2$:

$$\ln y_{it_2} = g\left(t_2 - t_1 e^{-\lambda_i \tau}\right) - e^{-\lambda_i \tau}\ln y_{it_1} + (1-e^{-\lambda_i \tau})\ln y^*_{i0} \tag{4.2.14}$$

with $\tau = t_2 - t_1$. In particular, under the assumption that the speed of convergence is homogeneous across regions ($\lambda_i = \lambda$ for $i = 1, \cdots, N$):

$$\ln y_{it_2} = g\left(t_2 - t_1 e^{-\lambda \tau}\right) + e^{-\lambda \tau}\ln y_{it_1} + (1-e^{-\lambda \tau})\ln y^*_{i0} \tag{4.2.15}$$

At this point it is convenient to write the previous expression in matrix form:

$$y(t_2) = g\left(t_2 - t_1 e^{-\lambda \tau}\right)\iota_N + e^{-\lambda \tau} y(t_1) + \left(1-e^{-\lambda \tau}\right) y^*(0) \tag{4.2.16}$$

where $y(t_2)$ is a $N \times 1$ vector containing the log of the outcome per worker at t_2, ι_N is a $N \times 1$ vector of ones, $y(t_1)$ is a $N \times 1$ vector containing the log of the outcome per worker at t_1, and $y^*(0)$ is a $N \times 1$ vector containing the log of the initial level of output per worker at the steady state. The reason for this is that facilitates replacing $y^*(0)$ by 4.2.9 at $t = 0$, which, in matrix form, is:

$$y^*(0) = (I-\rho W)^{-1}\left[\frac{1}{1-\alpha-\phi}\Omega(0) + \frac{\gamma_1}{1-\alpha-\phi}W\Omega(0) + \frac{\alpha+\phi}{1-\alpha-\phi}S + \frac{\gamma_2-\alpha\gamma_3}{1-\alpha-\phi}WS\right] \tag{4.2.17}$$

where $\rho = \dfrac{(1-\alpha)\gamma_3+\gamma_2}{1-\alpha-\phi}$ (it is assumed that $1/\rho$ is not an eigenvalue of W when $\rho \neq 0$) and $S = \left\{\ln\left(\dfrac{s_i}{n_i+\delta+g}\right)\right\}_{i=1,\ldots,N}$.

Thus, we introduce 4.2.17 in 4.2.16 and pre-multiply both sides of the resulting equation by $I-\rho W$ to obtain:

$$\begin{aligned} y(t_2) = {} & g(1-\rho)\left(t_2 - t_1 e^{-\lambda\tau}\right)\iota_N + e^{-\lambda\tau}\left(I-\rho W\right)y(t_1) + \rho W y(t_2) \\ & + \left(1-e^{-\lambda\tau}\right)\left[\frac{1}{1-\alpha-\phi}\Omega(0) + \frac{\gamma_1}{1-\alpha-\phi}W\Omega(0) + \frac{\alpha+\phi}{1-\alpha-\phi}S + \frac{\gamma_2-\alpha\gamma_3}{1-\alpha-\phi}WS\right] \end{aligned} \tag{4.2.18}$$

Alternatively, we can rewrite this equation for country i as

$$\begin{aligned}
\ln y_{it_2} &= e^{-\lambda\tau}\ln y_{it_1} - \rho e^{-\lambda\tau}\sum_{j=1}^{N} w_{ij}\ln y_{jt_1} + \rho\sum_{j=1}^{N} w_{ij}\ln y_{jt_2} \\
&+ \frac{\left(1-e^{-\lambda\tau}\right)(\alpha+\phi)}{1-\alpha-\phi}\ln s_i - \frac{\left(1-e^{-\lambda\tau}\right)(\alpha+\phi)}{1-\alpha-\phi}\ln(n_i+\delta+g) \\
&+ \frac{\left(1-e^{-\lambda\tau}\right)(\gamma_2-\alpha\gamma_3)}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln s_j - \frac{\left(1-e^{-\lambda\tau}\right)(\gamma_2-\alpha\gamma_3)}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln(n_j+\delta+g) \\
&+ \left(\frac{\left(1-e^{-\lambda\tau}\right)}{1-\alpha-\phi}\ln\Omega_{i0}\right) + \left(\frac{\left(1-e^{-\lambda\tau}\right)\gamma_1}{1-\alpha-\phi}\sum_{j=1}^{N} w_{ij}\ln\Omega_{j0}\right) \\
&+ g(1-\rho)\left(t_2 - t_1 e^{-\lambda\tau}\right)
\end{aligned} \tag{4.2.19}$$

4.3 Empirical results

4.3.1 Model specification and identification strategies

To derive our econometric specification, notice that equation 4.2.19 (plus an i.i.d. shock ε) corresponds to the spatial Durbin dynamic panel model with individual-specific effects and their spatial spillovers:

$$\begin{aligned}
z_{it} &= \overline{\gamma}_1 z_{i,t-1} + \overline{\gamma}_2\sum_{j=1}^{N} w_{ij}z_{j,t-1} + \rho\sum_{j=1}^{N} w_{ij}z_{jt} + \beta_1 x_{1_{it}} + \beta_2 x_{2_{it}} + \theta_1\sum_{j=1}^{N} w_{ij}x_{1_{jt}} + \theta_2\sum_{j=1}^{N} w_{ij}x_{2_{jt}} \\
&+ \mu_i + \sum_{j=1}^{N} w_{ij}\alpha_j + f_t + \varepsilon_{it}
\end{aligned} \tag{4.3.1}$$

where $z_{it} = \ln y_{it_2}$, $z_{i,t-1} = \ln y_{it_1}$, $x_{1_{it}} = \ln s_{it}$, $x_{2_{it}} = \ln(n_{it}+\delta+g)$, $\overline{\gamma}_1 = e^{-\lambda\tau}$, $\overline{\gamma}_2 = -\rho e^{-\lambda\tau}$, $\beta_1 = \dfrac{\left(1-e^{-\lambda\tau}\right)(\alpha+\phi)}{1-\alpha-\phi}$, $\beta_2 = -\dfrac{\left(1-e^{-\lambda\tau}\right)(\alpha+\phi)}{1-\alpha-\phi}$, $\theta_1 = \dfrac{\left(1-e^{-\lambda\tau}\right)(\gamma_2-\alpha\gamma_3)}{1-\alpha-\phi}$, $\theta_2 = -\dfrac{\left(1-e^{-\lambda\tau}\right)(\gamma_2-\alpha\gamma_3)}{1-\alpha-\phi}$, $\mu_i = \dfrac{\left(1-e^{-\lambda\tau}\right)}{1-\alpha-\phi}\ln\Omega_{i0}$, $\alpha_i = \dfrac{\left(1-e^{-\lambda\tau}\right)\gamma_1}{1-\alpha-\phi}\ln\Omega_{i0}$ and $f_t = g(1-\rho)\left(t_2-t_1e^{-\lambda\tau}\right)$.

This means that equation 4.3.1 corresponds to the model specification discussed by Lee and Yu (2016), except that their model does not distinguishes the spatial counterparts of the individual effects $(\sum_{j=1}^{N} w_{ij}\alpha_j)$. In other words, their individual effects correspond to $\mu_i + \sum_{j=1}^{N} w_{ij}\alpha_j$ in 4.3.1. In fact, in our model the individual effects and their spatial counterparts are proportional (by a rate γ_1). This is therefore a particular case of the more general specification proposed by Miranda et al. (2017a).

To distinguish the individual effects from their spatial spillovers, we assume a correlated random effects specification for the individual effects (μ_i) and their spatial spillovers (α_i). This means making use of the following correlation functions (Mundlak, 1978; Chamberlain, 1982):

$$\begin{aligned} \mu_i &= \pi_{\mu_1}\left(\frac{1}{T}\sum_{t=1}^{T} x_{1_{it}}\right) + \pi_{\mu_2}\left(\frac{1}{T}\sum_{t=1}^{T} x_{2_{it}}\right) + \upsilon_{\mu i} \\ \alpha_i &= \pi_{\alpha_1}\left(\frac{1}{T}\sum_{t=1}^{T} x_{1_{it}}\right) + \pi_{\alpha_2}\left(\frac{1}{T}\sum_{t=1}^{T} x_{2_{it}}\right) + \upsilon_{\alpha i}, \end{aligned} \tag{4.3.2}$$

where π_{μ_1}, π_{μ_2}, π_{α_1} and π_{α_2} are the parameters associated with the period-means of the regressors, and $\upsilon_{\mu i}$ and $\upsilon_{\alpha i}$ are random error terms with $E(\upsilon_{\mu i}) = 0 = E(\upsilon_{\alpha i})$, $Var(\upsilon_{\mu i}) = \sigma^2_\mu$, $Var(\upsilon_{\alpha i}) = \sigma^2_\alpha$ and $Cov(\upsilon_{\mu i}, \upsilon_{\alpha i}) = \sigma_{\mu\alpha}$.

The last thing to notice about our econometric specification is that the implied parameters (α, ϕ and γ_2, on the one hand; γ_1, λ, γ_3, and $\ln \Omega_{i0}$, on the other) are not identified. In particular, we cannot obtain a single estimate of γ_1 (since this can be obtained from each (α_i, μ_i) pair, but also from either π_{μ_1} and π_{α_1} or π_{μ_2} and π_{α_2}), λ (since this can be obtained from $\overline{\gamma}_1$, but also from $\overline{\gamma}_2$ and ρ), γ_3 (since this requires ρ, $\overline{\gamma}_1$, either β_1 or β_2, and either θ_1 or θ_2, respectively) and $\ln \Omega_{i0}$ (since this requires either μ_i, $\overline{\gamma}_1$ and either β_1 or β_2, or α_i, $\overline{\gamma}_1$, γ_1 and either β_1 or β_2) because in principle these parameters are overidentified. However, it is easy to see that equations 4.3.1 and 4.3.2 three sets of constraints on the parameters: *i)* $\beta_1 = -\beta_2$ and $\theta_1 = -\theta_2$ (arising from the assumption that the production function is homogeneous of degree one, thus making the output per capita to depend only on the stock of physical capital); *ii)* $\overline{\gamma}_2 = -\rho\overline{\gamma}_1$ (arising from the assumed spatial-time dynamics of the technology); and *iii)* $\alpha_i = \gamma_1\mu_i$ (i.e., $\pi_\alpha = \gamma_1\pi_\mu$, $\sigma^2_\alpha = \gamma_1^2\sigma^2_\mu$ and $\sigma_{\mu,\alpha} = \gamma_1\sigma^2_\mu$, which arise from the assumed spatial contagion in the heterogeneous exogenous technology and unobserved productivity).[9] By imposing these six constraints on 4.3.1 and 4.3.2 (i.e., the "unconstrained model"), we obtain a constrained version of our model in which γ_1, λ, γ_3, and $\ln \Omega_{i0}$ are identified.

[9]While *i)* also arises in the model of Ertur and Koch (2007), *ii)* and *iii)* are specific to our model specification. In this respect, notice that Elhorst et al. (2010, p. 343) also consider the constraint $\overline{\gamma}_2 = -\rho\overline{\gamma}_1$. However, while in our case it arises directly from the derivation of our model specification, they argue that this "constraint is unnecessarily restrictive because no theoretical or empirical reason exists to impose it".

To this end, we start by replacing 4.3.2 into 4.3.1, which in matrix form yields:

$$Z_t = \overline{\gamma}_1 Z_{t-1} + \overline{\gamma}_2 W Z_{t-1} + \rho W Z_t + X_t \beta + W_t X \theta + \overline{X}\Pi_\mu + W\overline{X}\Pi_\alpha + f_t + \eta_t \tag{4.3.3}$$

where $X_t = \left(x_{1_t} \,\vdots\, x_{2_t} \right)$, $\overline{X}$ denote period-means of X_t, $\beta = (\beta_1, \beta_2)'$, $\theta = (\theta_1, \theta_2)'$, $\Pi_\mu = (\pi_{\mu_1}, \pi_{\mu_2})'$, $\Pi_\alpha = (\pi_{\alpha_1}, \pi_{\alpha_2})'$, and the error term is $\eta_t = \upsilon_\mu + W\upsilon_\alpha + \varepsilon_t$, with variance-covariance matrix given by $J_T \otimes (\sigma_\mu^2 I + \sigma_{\mu\alpha}(W + W') + \sigma_\alpha^2 WW') + \sigma_\varepsilon^2 I_{NT}$, J_T being a $T \times T$ matrix of ones and I_{NT} being the $NT \times NT$ identity matrix. This is the unconstrained version of our econometric model.

Let us now define $S_1 = I - \rho W$, $S_2 = I + \gamma_1 W$ and $X_{it}^* = \ln\left(\dfrac{s_{it}}{n_{it} + \delta + g}\right) = \ln(S_{it})$. Then, the constrained model is given by

$$S_1 Z_t = \overline{\gamma}_1^c S_1 Z_{t-1} + \beta^c X^* + \theta^c W X^* + S_2 \overline{X}\Pi_\mu^c + f_t + \eta_t^c \tag{4.3.4}$$

with $\overline{\gamma}_2 = -\rho\overline{\gamma}_1^c$, $\beta_1 = -\beta_2 = \beta^c$, $\theta_2 = -\theta_1 = \theta^c$ and $\Pi_\alpha = \gamma_1 \Pi_\mu^c$ and $\eta_t^c = \varepsilon_t + S_2 \upsilon_\mu$, with variance-covariance matrix given by $J_T \otimes (\sigma_\mu^2 S_2 S_2') + \sigma_\varepsilon^2 I_{NT}$. Notice that, in contrast to 4.3.3, the estimation of the constrained version of our econometric model in 4.3.4 (see e.g. Lee and Yu, 2016; Miranda et al., 2017a) allows us to obtain an estimate of: *i)* the degree of technological interdependence between the unobserved productivity, γ_1 (directly from S_2); *ii)* the speed of convergence, λ (from $\overline{\gamma}_1^c$); *iii)* the degree of technological interdependence between the economies, γ_3 (from $\overline{\gamma}_1^c$, β^c and θ^c); and *iv)* the unobserved productivity, $\ln \Omega_{i0}$ (from μ_i, β^c and $\overline{\gamma}_1^c$). In particular, obtaining a statistically significant estimate of γ_1 should be interpreted as supportive evidence for our model. Also, obtaining a statistically significant estimate of γ_3 would lead us to reject the models of Islam (1995) and López-Bazo et al. (2004).

The problem, of course, is that we still cannot identify α, ϕ and γ_2 (only $\alpha + \phi$ and $\gamma_2 - \alpha\gamma_3$), although in this case it is because these parameters are under-identified. Since both the own stock of physical capital and that of the neighbouring economies are arguments of the technology, we cannot separate the effect that, as an input of the production function, the stock of physical capital has on the output (i.e., α) from the effect that it has as a driver of the technology (i.e., ϕ). Neither can we separate the local effect that the stock of physical capital of the neighbouring economies has on the technology and, subsequently, the output (i.e., γ_2), from the global effect that the stock of physical capital of the neighbouring economies has on the technology and, subsequently, the output (i.e., $\alpha\gamma_3$). Still, there are ways to circumvent this identification problem.

One way is to modify the specification of the model. There are no identification problems, for example, if we are willing to assume that the stock of physical capital enters the technological progress lagged one period. That is, if we are willing to assume that

$A_{it} = \Omega_{it} \prod_{j \neq i}^{N} \Omega_{jt}^{\gamma_1 w_{ij}} k_{it-1}^{\phi} \prod_{j \neq i}^{N} k_{jt-1}^{\gamma_2 w_{ij}} \prod_{j \neq i}^{N} A_{jt}^{\gamma_3 w_{ij}}$ (see, in contrast, equation 4.2.4). Neither there are if we argue that different arguments of the technology require different weight matrices. In mathematical terms, this means assuming that $A_{it} = \Omega_{it} \prod_{j \neq i}^{N} \Omega_{jt}^{\gamma_1 w_{ij}^{\Omega}} k_{it}^{\phi} \prod_{j \neq i}^{N} k_{jt1}^{\gamma_2 w_{ij}^{k}} \prod_{j \neq i}^{N} A_{jt}^{\gamma_3 w_{ij}^{A}}$, where, in obvious notation, w_{ij}^{Ω}, w_{ij}^{k} and w_{ij}^{A} denote different weight matrices (see e.g. Lee and Yu, 2016). These approaches, however, involve the derivation of a new model (the steady state equation and the speed of convergence, for example, would surely be altered) and/or require additional data to construct the weight matrices (in our empirical application, we may for example need data on bilateral trade flows and geographical distances between the EU regions). We thus leave these approaches for future research.

In this paper, we simply notice that some of the remaining implied parameters would be identified if an additional appropriate constrain was available. If we were willing to assume, for example, that the impact of the own physical stock and that of the other economies in the level of technology is the same (i.e., $\phi = \gamma_2$), then we may obtain an estimate of α and $\phi = \gamma_2$ from γ_3, β^c and θ^c. However, since our assumed technology encompasses that of López-Bazo et al. (2004) and Ertur and Koch (2007), we find that it is of greater interest to constrain the under-identified implied parameters of the technology (i.e., ϕ and γ_2) to be consistent with the technology these papers assume. Thus, under the assumption that the technology of López-Bazo et al. (2004) is the appropriate (i.e., $\phi = 0$), we can obtain an estimate of α and γ_2 from 4.3.4, whereas under the assumption that the technology of Ertur and Koch (2007) is the appropriate (i.e., $\gamma_2 = 0$), we can obtain an estimate of α and ϕ from 4.3.4. Further, under the assumption that neither the own capital nor that of the neighbouring economies have a role in shaping the technology (i.e., $\phi = \gamma_2 = 0$), we can obtain an estimate of α from the following constrained model:[10]

$$S_1 Z_t = \overline{\gamma}_1^c S_1 Z_{t-1} + \beta^c S_1 X^* + S_2 \overline{X} \Pi_{\mu}^c + f_t + \eta_t^c \tag{4.3.5}$$

However, these estimates should be interpreted with care. If $\gamma_3 = 0$, then we can interpret an statistically significant estimate of γ_2 that is obtained under $\phi = 0$ as supportive evidence for the model of López-Bazo et al. (2004). That is, there exist local spatial externalities in the technology associated with the stock of capital (but not global, since the data supports that $\gamma_3 = 0$). If $\gamma_3 \neq 0$, however, the evidence is consistent with the model of Ertur and Koch (2007, p. 1036), and then the question that we can address by imposing the constraint

[10]Notice that, while imposing the assumptions that either $\phi = 0$ or $\gamma_2 = 0$ in 4.3.4 does not yield a new model specification, imposing both assumptions simultaneously does yield a new model specification (4.3.5) in which $\theta^c = \beta^c \gamma_3 = \beta^c \rho$ (see also Ertur and Koch 2007 and Elhorst et al. 2010).

$\gamma_2 = 0$ (but not that $\phi = 0$) is whether there are indeed "home externalities generated by physical capital accumulation". In particular, only if $\phi \neq 0$ we may conclude that there exist global spatial externalities associated with the stock of capital (although we remain uncertain about whether there are local externalities in capital because we have imposed that $\gamma_2 = 0$). Lastly, if $\phi = \gamma_2 = 0$, we may conclude that, regardless of the value taken by γ_3, there is no spatial contagion of the stock of capital (neither in the technology nor in the output). However, if the data rejects the validity of these constraints, then we remain uncertain about what is the nature of these spatial externalities: local (i.e., $\phi \neq 0$), global (i.e., $\gamma_2 \neq 0$), or both local and global (i.e., $\phi \neq 0$ and $\gamma_2 \neq 0$).

4.3.2 Estimates from EU-NUTS2 regions

We estimate the model given by 4.3.3 using the approach and model specifications of Lee and Yu (2016) and Miranda et al. (2017a). We use the first as a benchmark for our basic parameters ($\overline{\gamma}_1$, $\overline{\gamma}_2$, ρ, β_1, β_2, θ_1 and θ_2, which, since all the variables are in logs, can be interpreted as elasticities) and the second to obtain the whole set of estimates (i.e., the basic ones plus those appearing in the correlation functions: π_{μ_1}, π_{μ_2}, π_{α_1} and π_{α_2}), test the validity of the constrained version of the model (using a Likelihood Ratio test), and estimate the implied parameters (using the constrained version of the model). We also follow this scheme in the discussion of the results. This means that we will start with an analysis of the estimates of the basic and correlation functions parameters in the unconstrained and constrained models, then will go on with the estimates of the implied parameters, and we will conclude with a description of the geographical distribution of the estimated "unobserved productivity" of the EU regions ($\ln \hat{\Omega}_{i0}$) and its estimated spatial spillover ($\hat{\gamma}_1 \sum_{j=1}^{N} w_{ij} \ln \hat{\Omega}_{j0}$).

First, however, a word about the data. We use EU NUTS2 regional data from Cambridge Econometrics to estimate our model. In particular, our initial sample is analogous to the one analysed by Elhorst et al. (2010), so that we can use their results as a benchmark to which ours will be compared. Thus, we initially consider 189 regions across 14 EU countries (Austria, Belgium, Germany, Denmark, Greece, Finland, France, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom) using time intervals of five years (see also Ho et al., 2013; Lee and Yu, 2016) over the period 1982 to 2002. This results in a balanced panel dataset with 4 time periods (1982-1987, 1987-1992, 1992-1997, 1997-2002).[11]

[11] To be precise, the (small) differences between our sample and that of Elhorst et al. (2010) are the following. First, they have data on Luxembourg and the period 1977-1982. Second, in their sample "the islands (such as those associated with southern European countries) are assumed to be connected to the

It is worth noting, however, that we have explored alternative samples to check the robustness of our results. First, we extended our initial sample to cover the years of the recent global crisis (the time intervals 2002-2007 and 2007-2012). Second, we considered different time intervals in a wider time period (1980 to 2015, with observations for 1980-1985, 1985-1990, and up to 2010-2015, which was the last available period at the moment of writing this paper). Third, we considered alternative groups of countries (e.g., including Norway, which is a non-EU country, and/or dropping Portugal, Ireland, Italy, Spain and/or Greece, which are countries that have faced –severe– economic growth problems over the last decade). In all those cases, the estimates we obtained for the (un)constrained model remained largely unaltered. We illustrate this by reporting results from these alternative sampling schemes: the period 2002 to 2012, the period 1980 to 2015, the period 1982 to 2002 without including Portugal, Ireland, Italy, Spain and Greece (the so-called "PIIGS") and the period 1982 to 2002 without including Greece (since in all these cases results when including Norway were not substantially different).

All these estimates were obtained using real GDP per capita as the dependent variable (i.e., y_{it} is real GDP at 2005 constant prices over total population, in thousands of people). As for the explanatory variables, s_{it} is the ratio between investment expenditures and gross value-added (at 2005 constant prices and as a percentage) and n_{it} is the growth rate of the population over time (computed as in Islam 1995). As it is common in the literature (see e.g. Mankiw et al., 1992; Islam, 1995; Ertur and Koch, 2007), we assume that $\delta + g = 0.05$ to compute the depreciation rate. Note also that time dummies and a constant term were included in the set of explanatory variables to account for f_t.

[Insert Table 4.1 about here]

Table 1 provides descriptive statistics for the dependent and main explanatory variables (i.e., y_{it}, s_{it} and n_{it}). In particular, we report the statistics for the five samples considered and the periods effectively used in estimation in each case (notice that we lose one observation due to the inclusion of the lagged dependent variable in the model). The differences in the values of the statistics across the samples considered are of small magnitude, particularly between the original sample and the same sample without Greece. In fact, the observed differences arise in the GDP and the saving rate, whereas the distribution of the depreciation rate remains almost unaltered across samples. It is also interesting to note that the recent economic crisis seems to have increased the levels of GDP and savings, but mostly for those

mainland, so that each region has at least one neighbour" (p. 353). Here we only consider continental regions, which means that our sample does not include the Spanish cities of Ceuta and Melilla, the French's "Départements d'outre mer", and the Greek, Finish, French, Italian and Spanish islands.

regions that were already in the top of the distribution (i.e., the centre of the distribution of these variables has shifted to the right and the upper tail has increased, thus making differences between the extremes larger). The effect is similar when dropping the PIIGS from the original sample, except that now it it is the lower tail of the distribution the one that increases (i.e., we are dropping regions with levels of GDP and savings that are lower than those of the rest of the sample).

[Insert Table 4.2 about here]

We move now to the analysis of the estimates of the model and, as previously pointed out, start by considering the estimates of the unconstrained version of the model. These are reported in Table 4.2. In particular, the first reported estimates (in column two) were obtained using the approach and model specification of Lee and Yu (2016), whereas the rest (columns three to seven) were obtained using that of Miranda et al. (2017a). We report results for the initial sample (period 1982 to 2002) in columns two and three and, subsequently, for the other samples considered (periods 1982 to 2012, 1980 to 2015, 1982 to 2002 without the PIIGS, and 1982 to 2002 without Greece).

We find a remarkable regularity in both the values and the statistical significance of the coefficients across the samples and estimation approaches considered. Perhaps the only differences worth mentioning are: *i)* the slightly lower value of the coefficient associated with the time-lagged dependent variable ($\overline{\gamma}_1$) when estimating the model using the approach of Lee and Yu (2016); and *ii)* the lack of statistical significance of the coefficient associated with the saving rate (β_1) when considering the years of the recent crisis (i.e., the samples covering the periods 2002 to 2012 and 1980 to 2015). This caveat aside, all sets of estimates provide essentially the same picture.

In particular, the basic parameters are all statistically significant (except for θ_2) and have the predicted signs (see Ertur and Koch, 2007).[12] Consistent with the constraint $\overline{\gamma}_2 = -\rho\overline{\gamma}_1$, the spatial and time lagged dependent variables have a high and positive coefficient, whereas the spatially weighted lagged dependent variable has a negative and smaller coefficient in absolute value (see also Ho et al., 2013; Lee and Yu, 2016). Thus, the level of GDP per capita of the European regions is largely determined by its past GDP per capita and the current and past GDP per capita of their neighbours. Further, the saving rate of an economy

[12]Our estimates of the basic parameters are largely consistent with those reported by Basile (2008) using an analogous sample of regions and the period 1988 to 2000. They also concur with those reported in panel data studies analysing countries rather than regions (see e.g. Ho et al., 2013; Lee and Yu, 2016). In contrast, we find some differences with those reported by Pfaffermayr (2009), who consider an analogous period of analysis but whose sample includes Norway's and Switzerland's regions.

contributes positively to its GDP per capita, but its depreciation rate and the saving rate of the neighbouring regions both contribute negatively. All in all, these results indicate that richest areas are likely to stay rich (more so they if are geographically close to rich areas, like e.g. in the so-called "blue banana") while poorer areas can only (partially) catch up if they increase their saving rates and/or are geographically close to rich areas.

As for the correlation functions parameters, there is evidence of *i)* correlation between the individual effects and the covariates (since both the –mean– saving and depreciation rates are statistically significant) and *ii)* spatial contagion in the individual effects (since the spatially weighted –mean– saving rate is generally statistically significant). In addition, two of the variance components, σ_{μ}^{2} and σ_{ε}^{2}, are statistically significant. This supports our correlated random effects model specification. In particular, results are consistent with the constraint $\alpha_i = \gamma_1\mu_i$, which implies a "fixed effects" error term model with proportional spatial contagion (Miranda et al., 2017a).

[Insert Table 4.3 about here]

Next we consider the results for the constrained version of the model, which are reported in Table 4.3. Before discussing the estimates, however, it is important to assess the validity of equation 4.3.4 in the different samples considered. To this end, we used a Likelihood Ratio test. We found that the "fully" constrained version of the model (i.e., the model resulting from imposing the constraints $\beta_1 = -\beta_2$, $\theta_1 = -\theta_2$, $\overline{\gamma}_2 = -\rho\overline{\gamma}_1$ and $\alpha_i = \gamma_1\mu_i$) was statistically supported only in the last two samples (i.e., the period 1982 to 2002 without the PIIGS and without Greece).[13] Estimates from this fully constrained version of the model are reported in Table 4.3*b*. Still, after testing the validity of each constraint individually, we found that a "partially" constrained version of the model in which only the constraint $\alpha_i = \gamma_1\mu_i$ was imposed was not rejected in the first three samples (periods 1982 to 2002, 1982 to 2012, and 1980 to 2015). Estimates from this partially constrained version of the model are reported in Table 4.3*a*.[14]

At first sight, there is very little to comment on the results reported in Table 4.3*a* since, as expected, they are very similar to the ones obtained from the unconstrained model (see Table

[13] In particular, the Likelihood Ratio test statistics we obtained in the first three samples were 18.42 (period 1982 to 2002), 42.06 (period 1982 to 2012) and 27.26 (period 1980 to 2015), all statistically significant at standard levels. The Likelihood Ratio test statistics of the other samples (the period 1982 to 2002 without the PIIGS and without Greece) are reported in the last row of Table 4.3.

[14] Notice that, since γ_1 is an implied parameter, it is reported in Table 4.4 (along with the rest of the implied parameters obtained from the fully constrained model). However, since γ_1 is identified in both the fully and partially constrained models, for the sake of comparability we have also included its estimates among the results reported in Table 4.3.

4.2). Yet two things are worth mentioning. First, the correlation functions parameters and the variance components parameters are all statistically significant. This again supports our correlated random effects specification. Second, the coefficient reflecting the degree of technological interdependence generated from the productivity spillovers, γ_1, shows a negative and (at least in two of the samples considered) statistically significant value. Also, the estimates we obtain for γ_1 are similar across the samples considered. This indicates, given the imposed constraint $\alpha_i = \gamma_1 \mu_i$, that there exists a negatively proportional relation between the individual effects of the EU regions and their spatial spillovers. We will return to this point when we analyse the geographical distribution of $\ln \hat{\Omega}_{i0}$ and $\hat{\gamma}_1 \sum_{j=1}^{N} w_{ij} \ln \hat{\Omega}_{j0}$.

As for the estimates of the "fully" constrained version of the model, the first thing to notice is that they are similar in the two samples considered (except for the lack of statistical significance of θ^c in the sample without the PIIGS). In particular, the basic parameters are all statistically significant and have the predicted signs (see Ertur and Koch, 2007). Also, if we compare our results with those obtained by Elhorst et al. (2010), our estimates of the difference in the logs of the saving and depreciation rates, as well as that of its spatial counterpart, are both larger (and statistically significant, whereas only the former is in their case). The estimated coefficient of the spatially lagged dependent variable, on the other hand, is analogous to the one reported by Elhorst et al. (2010). Lastly, the rest of parameters have estimated values in line with those obtained for the "partially" constrained version of the model (including γ_1, as previously pointed out).

We then use these "fully constrained" estimates to obtain the implied parameters of the theoretical model. These are reported in Table 4.4. In particular, the first block of Table 4.4 contains the estimates of the parameters that are directly identified (γ_1, λ and γ_3), the second block the estimates of α and γ_2 obtained under the assumption that the technology considered by López-Bazo et al. (2004) is the appropriate (i.e., imposing the additional constraint $\phi = 0$), and the third block the estimates of α and ϕ obtained under the assumption that the technology considered by Ertur and Koch (2007) is the appropriate (i.e., imposing the additional constraint $\gamma_2 = 0$).

[Insert Table 4.4 about here]

Firstly, the statistical significance of the degree of technological interdependence generated from the (unobserved) productivity spillovers, γ_1, supports our assumed technology (against the related alternatives of Islam 1995, López-Bazo et al. 2004 and Ertur and Koch 2007). Secondly, the estimated speed of convergence, as measured by λ, is around 2% and statistically significant, which is a standard result in the literature (Barro and Sala-i-Martin, 2003; López-Bazo et al., 2004; Ertur and Koch, 2007; Lee and Yu, 2016). Thirdly, the statistical significance of the degree of technological interdependence, as measured by γ_3, supports the model of Ertur and Koch (2007) and contradicts the models of Islam (1995) and López-Bazo et al. (2004). Moreover, its value is similar to the one found by Ertur and Koch (2007) and Elhorst et al. (2010), somewhere in between them. Fourthly, the estimates of the capital share, as measured by α, obtained when imposing the additional constraint(s) that $\phi = 0$ (López-Bazo et al., 2004) and/or $\gamma_2 = 0$ (Ertur and Koch, 2007) are in line with those obtained in the literature (Barro and Sala-i-Martin, 2003; Ertur and Koch, 2007; Elhorst et al., 2010). Fifthly, the parameter capturing capital externalities at the local level (γ_2) and that allowing for capital externalities at the global level (ϕ, through γ_3), obtained when imposing the additional constraint that either $\phi = 0$ (López-Bazo et al., 2004) or $\gamma_2 = 0$ (Ertur and Koch, 2007), respectively, are not statistically significant. In fact, we cannot reject the null hypothesis that both parameters are zero. The LR test statistic obtained from the models 4.3.4 and 4.3.5 is 0.53 for the sample 1982 to 2002 without the PIIGS and 0.53 for the sample 1982 to 2002 without Greece, none of them being statistically significant at standard levels.[15]

All in all, these results point to the the existence of spatial spillovers in the unobserved productivity and the level of technology. That is, we find evidence supporting the existence of both local and global spillovers in the stock of knowledge. In contrast, there is no sign of the capital externalities in technology found by either López-Bazo et al. (2004) or Ertur and Koch (2007). That is, we do not find evidence of spatial externalities in the stock of capital. Lastly, our estimates support our model specification against that of Islam (1995), López-Bazo et al. (2004) and Ertur and Koch (2007).

[Insert Figure 4.1 about here]

To conclude our empirical analysis, we report the geographical distribution of the estimated "unobserved productivity" and its spatial spillover (to reiterate, obtained from the constrained model in 4.3.4) in Figure 4.1. More precisely, Figure 4.1 presents a map of the European regions considered and the values of these statistics grouped by quantiles: Figure 4.1a reports $\ln \hat{\Omega}_{i0}$ (the "unobserved productivity") whereas Figure 4.1b reports $\hat{\gamma}_1 \sum_{j=1}^{N} w_{ij} \ln \hat{\Omega}_{j0}$ (the spatial spillover of the "unobserved productivity", that is, the impact on the technology of unit i of all the units neighbouring i having their "unobserved productivity"). Notice that we have opted for using the estimates from the 1982-2002 sample without Greece to construct Figure 4.1 because this allows us to analyse a larger number

[15] More generally, the other four findings largely hold when imposing on the "fully constrained" model in 4.3.4 the additional constraint that $\phi = \gamma_2 = 0$, that is, when estimating the constrained model in 4.3.5.

of regions. It is important to stress, however, that results were not substantially different when using the 1982-2002 sample without the PIIGS. Notice also that, given the negative and statistically significant value found for γ_1, there is a negatively proportional relation between the unobserved productivity of each EU region and the spatial contagion of this unobserved productivity on its neighbouring regions.[16]

With this in mind, we start by noting the considerable heterogeneity that Figure 4.1*a* displays, which contradicts the standard assumption of homogeneous exogenous technological progress. In particular, results indicate that the regions with the lowest estimated "unobserved productivity" are mostly located in Scandinavia (Finland and Sweden), Scotland and North of England, Northern Ireland, Central-South of France, South-Est of Germany, Austria, Central and North-West of Spain, and North-West and South of Italy. Figure 4.1*a* also shows that the geographical distribution of the higher estimated "unobserved productivity" covers the so-called "blue banana" (from the South of the UK to the South-West of Germany, thus including the North of France, Belgium and the Netherlands), plus Denmark and the Mediterranean regions of the South-West of France and Central Italy.

What is also interesting to note is that about half of the regions in the high productivity group can be qualified as "rich", meaning here that their average GDP per capita over the periods considered is in the upper quantile of the distribution. On the other hand, the same criterion would lead us to qualify about half of the regions with low estimated productivities as "poor". Thus, it seems that many of the richer/poorer regions tend to have higher/lower (unobserved) productivities. In fact, the Spearman rank correlation between $\ln \hat{\Omega}_{i0}$ and the average GDP per capita is 0.36 and statistically significant.

As for the spillovers associated with the "unobserved productivity", Figure 4.1*b* reveals that the pattern tends to be opposite to the one found for the estimated "unobserved productivity". In particular, the largest values are found in the Northern regions (i.e., Scandinavia, East of Ireland, the UK Midlands and South of Scotland), but also in the East (i.e., Austria) and South (South-West of France, North and West of Spain, and South of Italy) of Europe. This means that these are (often poor) regions whose "unobserved productivity" is more impacted by the "unobserved productivity" of its neighbours. South of England and Ireland, Belgium, the Netherlands, and West Germany, on the other hand, stand as the areas with the lowest spillovers. This means that these are (mostly rich) regions whose output per capita is barely affected by the "unobserved productivity" of its neighbours.

[16] These spillovers correspond to the (local) spill-in effects proposed by Miranda et al. (2017b). We do not report the spill-out effects because, given the proportional relation that imposes the constraint $\alpha_i = \gamma_1 \mu_i$, its geographical distribution is no more informative than that of $\ln \hat{\Omega}_{i0}$ (in fact, since both $\ln \hat{\Omega}_{i0}$ and γ_1 take negative values, the spill-out effects take positive values and are larger/smaller the smaller/larger $\ln \hat{\Omega}_{i0}$ is).

4.4 Conclusions

We present a growth model that extends previous knowledge-spillovers models in several directions. First, we do not assume a common exogenous technological progress but account for heterogeneity in the initial level of technology. Second, we assume that the technological progress depends not only on the stock of physical capital and the stock of knowledge of the other economies, but also on the physical capital and the (unobserved) initial level of technology of the other economies. Thus, our assumed technology combines the alternative sources of (global and local) spatial externalities considered in previous models of relative location with the unobserved heterogeneity that characterises previous models of absolute location.

We use EU-NUTS2 regional information from Cambridge Econometrics to test whether the data supports the main features of our growth model. In particular, our econometric specification is derived from the growth-initial equation of the model and takes the form of a spatial Durbin dynamic panel model with spatially weighted individual effects. As a downside, some of the implied parameters of the model are not identified. We discuss alternative ways to circumvent this limitation.

We estimate the model by QML using a correlated random effects specification for the individual effects and their spatial spillovers. Our results support our model specification. Also, they are largely *i)* consistent with other studies using analogous data; and *ii)* robust to the use of alternative specifications, samples and estimation approaches. In particular, we find evidence of the existence of (global) spatial spillovers arising from the level of technology, but not from the investment in capital (neither global nor local). Also, our estimates indicate that the level of GDP per capita of the European regions is largely determined by their past GDP per capita and the current and past GDP per capita of their neighbours, their saving rate and that of their neighbours, and their depreciation rate. However, the role of unobservable characteristics is worth noting: richest areas (e.g., the “blue banana") are so partially because of their higher “unobserved productivity" and a number of poor regions benefit from “unobserved productivity" spillovers.

Table 4.1: **Descriptive statistics**

(a) *Sample I: 1982-2002*

Variable	Mean	St. Dev.	Min	P25	Median	P75	Max
GDP	23,393	9,961	6,321	18,554	22,307	26,227	133,452
s	23.39	4.50	9.98	20.65	23.08	25.77	46.08
$n+\delta+g$	0.05	0.00	0.04	0.05	0.05	0.06	0.07

(b) *Sample II: 1982-2012*

Variable	Mean	St. Dev.	Min	P25	Median	P75	Max
GDP	25,355	11,536	6,321	19,698	23,997	28,934	176,529
s	23.69	4.76	9.98	20.64	23.47	26.17	48.84
$n+\delta+g$	0.05	0.01	0.04	0.05	0.05	0.06	0.08

(c) *Sample III: 1980-2015*

Variable	Mean	St. Dev.	Min	P25	Median	P75	Max
GDP	25,322	11,842	5,798	19,567	24,020	28,829	191,016
s	23.51	4.81	9.39	20.50	23.26	25.83	46.31
$n+\delta+g$	0.05	0.01	0.04	0.05	0.05	0.06	0.07

(d) *Sample IV: 1982-2002 w/o PIIGS (Portugal, Ireland, Italy, Spain and Greece)*

Variable	Mean	St. Dev.	Min	P25	Median	P75	Max
GDP	25,317	10,247	12,208	20,464	23,397	27,307	133,452
s	23.28	4.44	10.82	20.65	23.00	25.41	46.08
$n+\delta+g$	0.05	0.00	0.04	0.05	0.05	0.06	0.07

(e) *Sample V: 1982-2002 w/o EL (Greece)*

Variable	Mean	St. Dev.	Min	P25	Median	P75	Max
GDP	23,936	9,881	6,321	19,188	22,620	26,525	133,452
s	23.34	4.41	9.98	20.65	23.08	25.70	46.08
$n+\delta+g$	0.05	0.00	0.04	0.05	0.05	0.06	0.07

Note: Number of observations: $189\times4 = 756$ (Sample I), $189\times6 = 1,134$ (Sample II), $189 \times 7 = 1,323$ (Sample III), $139 \times 4 = 556$ (Sample IV), and $180 \times 4 = 720$ (Sample V). GDP is real GDP (at 2005 constant prices, in Euros) per capita (using total population, in thousands of people). s is the ratio between investment expenditures and gross value-added (as a percentage and at 2005 constant prices, in Euros). n is is the working-age population growth rate (computed as in Islam 1995) and $\delta+g = 0.05$ (as in e.g. Mankiw et al., 1992; Islam, 1995; Ertur and Koch, 2007).

Table 4.2: **QML estimates (unconstrained model)**

	Sample I (1982-2002)	Sample I (1982-2002)	Sample II (1982-2012)	Sample III (1980-2015)	Sample I (w/o PIIGS)	Sample I (w/o EL)
$\overline{\gamma}_1$	0.6291***	0.9049***	0.9177***	0.8520***	0.8681***	0.8980***
	(0.0304)	(0.0145)	(0.0160)	(0.0294)	(0.0221)	(0.0157)
$\overline{\gamma}_2$	−0.3202***	−0.4317***	−0.4746***	−0.3934***	−0.4757***	−0.4706***
	(0.0556)	(0.0366)	(0.0290)	(0.0338)	(0.0412)	(0.0362)
ρ	0.5281***	0.5047***	0.5603***	0.5463***	0.5587***	0.5357***
	(0.0432)	(0.0380)	(0.0277)	(0.0273)	(0.0513)	(0.0383)
β_1	0.1149***	0.0774**	0.0124	−0.0053	0.0604	0.1031***
	(0.0283)	(0.0354)	(0.0187)	(0.0149)	(0.0405)	(0.0349)
β_2	−0.1624***	−0.1952***	−0.1742***	−0.1045***	−0.1564***	−0.1536***
	(0.0434)	(0.0542)	(0.0370)	(0.0320)	(0.0506)	(0.0529)
θ_1	−0.0944***	−0.0907**	−0.0526**	0.0018	−0.1090***	−0.1154***
	(0.0339)	(0.0419)	(0.0259)	(0.0187)	(0.0506)	(0.0410)
θ_2	0.0553	0.0528	0.0337	0.0317	0.1085	0.0446
	(0.0577)	(0.0714)	(0.0482)	(0.0404)	(0.0703)	(0.0697)
π_{μ_1}		−0.1131***	−0.0526**	−0.0606**	−0.1185**	−0.1432***
		(0.0397)	(0.0259)	(0.0306)	(0.0482)	(0.0403)
π_{μ_2}		0.3486***	0.3321***	0.3310***	0.3358***	0.3037***
		(0.0728)	(0.0596)	(0.0752)	(0.0888)	(0.0737)
π_{α_1}		0.0954*	0.0829**	0.0613	0.1223*	0.1189*
		(0.0502)	(0.0337)	(0.0393)	(0.0637)	(0.0508)
π_{α_2}		−0.1637	−0.1244	−0.1453	−0.2721	−0.0975
		(0.1112)	(0.0846)	(0.1011)	(0.1360)	(0.1137)
σ^2_μ		0.0006***	0.0004**	0.0010**	0.0013***	0.0007***
		(0.0002)	(0.0001)	(0.0003)	(0.0003)	(0.0002)
σ^2_α		1.7×10^{-5}	0.0002	0.0008	0.0001	1.2×10^{-5}
		(0.0004)	(0.0003)	(0.0005)	(0.0008)	(0.0004)
$\sigma_{\mu\alpha}$		0.0001	−0.0002	−0.0007*	−0.0003	−0.0001
		(0.0002)	(0.0002)	(0.0004)	(0.0006)	(0.0003)
σ^2_ε		0.0035***	0.0030***	0.0028***	0.0019***	0.0028***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)

Note: All estimates obtained using the approach proposed by Miranda et al. (2017a), except for those in column two, which were obtained using the approach proposed by Lee and Yu (2016). Time dummies included but not reported. The symbol * indicates statistically significant at the 10% level, ** at the 5% level and *** at the 1% level.

Table 4.3: **QML estimates (constrained model)**

(a) Partially constrained model

	Sample I (1982-2002)	Sample II (1982-2012)	Sample III (1980-2015)
$\overline{\gamma}_1$	0.9028*** (0.0144)	0.9217*** (0.0147)	0.8674*** (0.0271)
$\overline{\gamma}_2$	−0.4455*** (0.0373)	−0.4853*** (0.0281)	−0.4099*** (0.0319)
ρ	0.5253*** (0.0384)	0.5608*** (0.0271)	0.5434*** (0.0273)
β_1	0.0595 (0.0394)	0.0019 (0.0196)	−0.0086 (0.0151)
β_2	−0.1891*** (0.0562)	−0.1717*** (0.0372)	−0.1044*** (0.0321)
θ_1	−0.0453 (0.0377)	−0.0116 (0.0174)	0.0142 (0.0153)
θ_2	0.0323 (0.0790)	0.0283 (0.0487)	0.0361 (0.0391)
$\pi_{\mu_1^c}$	−0.0903*** (0.0446)	−0.0336 (0.0261)	−0.0467* (0.0270)
$\pi_{\mu_2^c}$	0.3423*** (0.0781)	0.3309*** (0.0597)	0.3320*** (0.0707)
$\sigma_\mu^{2^c}$	0.0005** (0.0002)	0.0004*** (0.0001)	0.0009*** (0.0003)
$\sigma_\varepsilon^{2^c}$	0.0035*** (0.0002)	0.0031*** (0.0002)	0.0029*** (0.0001)
γ_1	−0.3803 (0.3303)	−0.3773* (0.2267)	−0.4644** (0.1903)
LR-test	4.82	5.35	3.02

(b) Fully constrained model

	Sample I (w/o PIIGS)	Sample I (w/o EL)
$\overline{\gamma_1^c}$	0.8700*** (0.0195)	0.9026*** (0.0131)
ρ^c	0.5747*** (0.0405)	0.5496*** (0.0361)
β^c	0.0797** (0.0323)	0.1083*** (0.0333)
θ^c	−0.0575 (0.0318)	−0.0780** (0.0383)
$\pi_{\mu_1^c}$	−0.1257*** (0.0412)	−0.1445*** (0.0392)
$\pi_{\mu_2^c}$	0.2195*** (0.07530)	0.2507*** (0.0617)
σ_μ^c	0.0013*** (0.0003)	0.0006*** (0.0002)
$\sigma_\varepsilon^{2^c}$	0.0020*** (0.0002)	0.0031*** (0.0002)
γ_1	−0.3854* (0.2108)	−0.4432* (0.2459)
LR-test	9.88	7.65

Note: The upper letter c denotes constrained parameters (see section 3.2 for details). All estimates obtained using the approach proposed by Miranda et al. (2017a). Time dummies included but not reported. The symbol * indicates statistically significant at the 10% level, ** at the 5% level and *** at the 1% level. LR-test is the Likelihood Ratio test statistic of the hypothesis that the constraint $\alpha_i = \gamma_1\mu_i$ is valid (Table 4.3a) and the constraints $\beta_1 = -\beta_2$, $\theta_1 = -\theta_2$, $\overline{\gamma}_2 = -\rho\overline{\gamma}_1$ and $\alpha_i = \gamma_1\mu_i$ are valid (Table 4.3b).

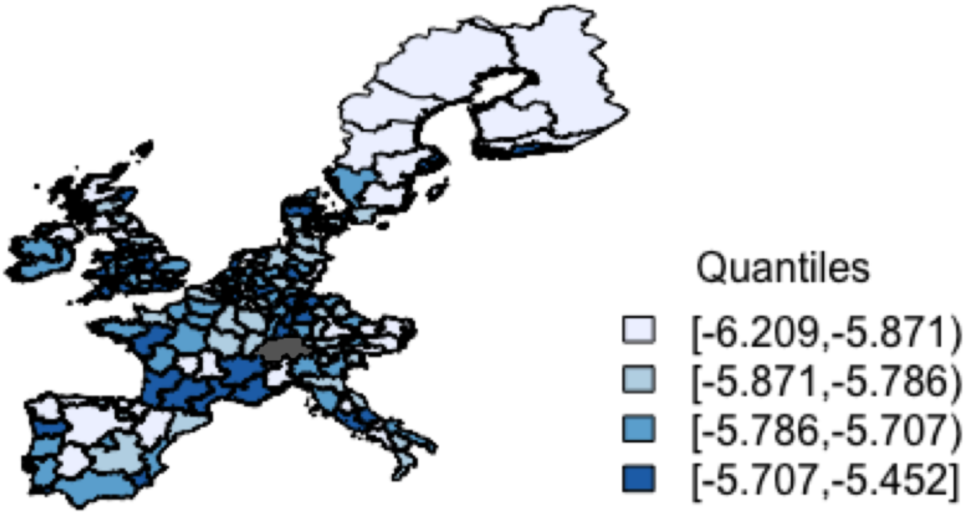
Table 4.4: **Implied Parameters**

	Sample I (w/o PIIGS)	Sample I (w/o EL)
γ_1	-0.3854^{*}	-0.4432^{*}
	(0.2108)	(0.2459)
λ	0.0205^{***}	0.0279^{***}
	(0.0029)	(0.0045)
γ_3	0.6394^{***}	0.6305^{***}
	(0.1346)	(0.1162)
Assumed technology: López-Bazo et al. (2004)		
α	0.5265^{***}	0.3801^{***}
	(0.0843)	(0.1021)
γ_2	-0.0425	-0.0346
	(0.0561)	(0.0679)
Assumed technology: Ertur and Koch (2007)		
α	0.5930^{***}	0.4349^{***}
	(0.1216)	(0.1374)
ϕ	-0.0665	-0.0548
	(0.0744)	(0.0983)

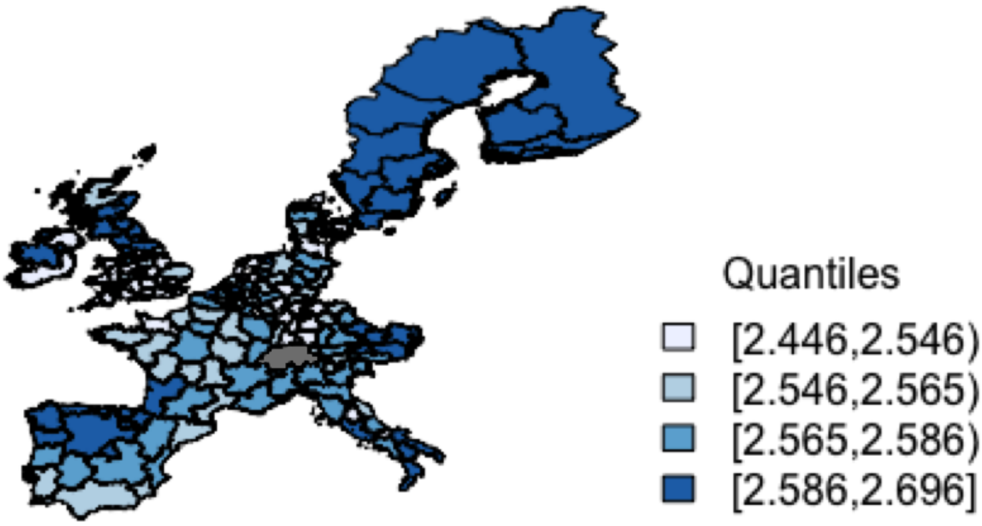
Note: * indicates statistically significant at the 10% level, ** at the 5% level and *** at the 1% level. Except for γ_1, standard errors were obtained using the delta method.

Figure 4.1: **Estimated individual effects and their spatial spillovers**

(a) Geographical distribution of $ln\hat{\Omega}_{i0}$



(b) Geographical distribution of $\hat{\gamma}_1 \sum_{j=1}^{N} w_{ij} \ln \hat{\Omega}_{j0}$



4.5 Appendix A: The balanced growth rate

From equation 4.2.6:

$$\ln y_{it} = \left[1 + \left(\frac{(\gamma_3+\gamma_1)(u_{ii}-\alpha-\phi)}{(\phi\gamma_3+\gamma_2)}\right)\right]\ln\Omega_{it} + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\ln\Omega_{jt}$$
$$+ u_{ii}\ln k_{it} + \sum_{j\neq i}^{N} u_{ij}\ln k_{jt}$$

Since $\ln\Omega_{it} = \ln\Omega_{i0} + \mu t$, then:

$$\frac{d\ln y_{it}}{dt} = \left[1 + \left(\frac{(\gamma_3+\gamma_1)(u_{ii}-\alpha-\phi)}{(\phi\gamma_3+\gamma_2)}\right)\right]\mu + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\mu + u_{ii}g + \sum_{j\neq i}^{N} u_{ij}g$$

Also, using $u_{ii} + \sum_{j\neq i}^{N} u_{ij} = \sum_{j=1}^{N} u_{ij} = \alpha + \frac{\phi+\gamma_2}{1-\gamma_3}$,

$$\frac{d\ln y_{it}}{dt} = \left(1 - \frac{(\gamma_3+\gamma_1)(\alpha+\phi)}{(\phi\gamma_3+\gamma_2)} + \frac{(\gamma_3+\gamma_1)}{(\phi\gamma_3+\gamma_2)}\left(\frac{\alpha(1-\gamma_3)+\phi+\gamma_2}{1-\gamma_3}\right)\right)\mu + \sum_{j=1}^{N} u_{ij}g = g,$$

which after some algebra becomes:

$$\left(\frac{1+\gamma_1}{1-\gamma_3}\right)\mu + \left(\frac{\alpha(1-\gamma_3)+\phi+\gamma_2}{1-\gamma_3}\right)g = g$$

Therefore,

$$g = \frac{\mu\,(1+\gamma_1)}{(1-\gamma_3)(1-\alpha)-\phi-\gamma_2}$$

4.6 Appendix B: Taylor approximation to the marginal productivity of capital

The Taylor approximation of $\frac{\dot{k}_{it}}{k_{it}}$ around the steady state $(k^*_{1t}, \cdots, k^*_{Nt})$ is

$$\frac{\dot{k}_{it}}{k_{it}} = \frac{\dot{k}^*_{it}}{k^*_{it}} + \sum_{j=1}^{N} \left\{ \left. \frac{\partial \left(\dot{k}_{it}/k_{it} \right)}{\partial \ln k_{jt}} \right|_{k^*_{jt}} (\ln k_{jt} - \ln k^*_{jt}) \right\}$$

$$= g + \left. \frac{\partial \left(\dot{k}_{it}/k_{it} \right)}{\partial \ln k_{it}} \right|_{k^*_{it}} (\ln k_{it} - \ln k^*_{it}) + \sum_{j \neq i}^{N} \left\{ \left. \frac{\partial \left(\dot{k}_{it}/k_{it} \right)}{\partial \ln k_{jt}} \right|_{k^*_{jt}} (\ln k_{jt} - \ln k^*_{jt}) \right\}$$

Next we calculate the two derivatives involved. First, let us rewrite the marginal productivity of capital (see footnote 8) as

$$\frac{\dot{k}_{it}}{k_{it}} = s_i \Omega_{it}^{c_{ii}} \prod_{j \neq i}^{N} \Omega_{jt}^{c_{ij}} e^{(u_{ii}-1)\ln k_{it}} \prod_{j \neq i}^{N} e^{u_{ij} \ln k_{jt}} - (n_i + \delta)$$

with $k_{it}^{u_{ii}-1} = e^{(u_{ii}-1)\ln k_{it}}$, $c_{ii} = 1 + \left(\frac{(\gamma_3 + \gamma_1)(u_{ii} - \alpha - \phi)}{(\phi\gamma_3 + \gamma_2)} \right)$ and $c_{ij} = \frac{(\gamma_3 + \gamma_1)u_{ij}}{\phi\gamma_3 + \gamma_2}$. Thus,

$$\left. \frac{\partial \left(\dot{k}_{it}/k_{it} \right)}{\partial \ln k_{it}} \right|_{k^*_{it}} = s_i \Omega_{it}^{c_{ii}} \prod_{j \neq i}^{N} \Omega_{jt}^{c_{ij}} (u_{ii} - 1) e^{(u_{ii}-1)\ln k^*_{it}} \prod_{j \neq i}^{N} e^{u_{ij} \ln k^*_{jt}}$$

Also, given that $s_i \left[\frac{y^*_{it}}{k^*_{it}} \right] - (n_i + \delta) - g = 0$, replacing y^*_{it} by 4.2.6 at the steady state we obtain

$$s_i \Omega_{it}^{c_{ii}} \prod_{j \neq i}^{N} \Omega_{jt}^{c_{ij}} \prod_{j \neq i}^{N} {k^*_{jt}}^{u_{ij}} = (n_i + \delta + g)\, {k^*_{it}}^{1-u_{ii}} \tag{4.6.1}$$

Consequently,

$$\left.\frac{\partial\left(\dot{k}_{it}/k_{it}\right)}{\partial \ln k_{it}}\right|_{k_{it}^*} = (u_{ii}-1)(n_i+\delta+g)$$

Lastly, bearing in mind that $\prod_{j\neq i}^{N} e^{u_{ij}\ln k_{jt}^*} = e^{\sum_{j\neq i}^{N} u_{ij}\ln k_{jt}^*}$,

$$\left.\frac{\partial\left(\dot{k}_{it}/k_{it}\right)}{\partial \ln k_{jt}}\right|_{k_{jt}^*} = s_i\Omega_{it}^{c_{ii}}\prod_{j\neq i}^{N}\Omega_{jt}^{c_{ij}}e^{u_{ij}\ln k_{jt}^*}u_{ij} = u_{ij}(n_i+\delta+g)$$

Therefore:

$$\frac{\dot{k}_{it}}{k_{it}} = \frac{d\ln k_i(t)}{dt} = g + (u_{ii}-1)(n_i+\delta+g)\left(\ln k_{it}-\ln k_i^*\right) + \sum_{j\neq i}^{N} u_{ij}(n_i+\delta+g)\left(\ln k_{jt}-\ln k_{jt}^*\right)$$

4.7 Appendix C: Speed of convergence

Let us take the total derivative of 4.2.6:

$$\begin{aligned}\frac{d\ln y_{it}}{dt} &= \left[1+\left(\frac{(\gamma_1+\gamma_2)(u_{ii}-\alpha-\phi)}{\phi\gamma_3+\gamma_2}\right)\right]\frac{d\ln\Omega_{it}}{dt} + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\frac{d\ln\Omega_{jt}}{dt}\\ &+ u_{ii}\frac{d\ln k_{it}}{dt} + \sum_{j\neq i}^{N} u_{ij}\frac{d\ln k_{jt}}{dt}\end{aligned}$$

Given that $\frac{d\ln\Omega_{it}}{dt} = \frac{d\ln\Omega_{jt}}{dt} = \mu$, we concentrate on the derivatives with respect to k. To this end, let us consider the final result of appendix 4.6:

$$\begin{aligned}\frac{d\ln k_{it}}{dt} &= g + (u_{ii}-1)(n_i+\delta+g)\left(\ln k_{it}-\ln k_{it}^*\right) + \sum_{j\neq i}^{N} u_{ij}(n_i+\delta+g)\left(\ln k_{jt}-\ln k_{jt}^*\right)\\ &= g - (n_i+\delta+g)\left(\ln k_{it}-\ln k_{it}^*\right) + u_{ii}(n_i+\delta+g)\left(\ln k_{it}-\ln k_{it}^*\right)\\ &+ \sum_{j\neq i}^{N} u_{ij}(n_i+\delta+g)\left(\ln k_{jt}-\ln k_{jt}^*\right)\end{aligned}$$

Then, using equation 4.2.6

$$\ln y_{it} = \left[1 + \left(\frac{(\gamma_1+\gamma_2)(u_{ii}-\alpha-\phi)}{\phi\gamma_3+\gamma_2}\right)\right]\ln\Omega_{it} + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\ln\Omega_{jt}$$
$$+ u_{ii}\ln k_{it} + \sum_{j\neq i}^{N} u_{ij}\ln k_{jt}$$

and its value at the steady state

$$\ln y^*_{it} = \left[1 + \left(\frac{(\gamma_1+\gamma_2)(u_{ii}-\alpha-\phi)}{\phi\gamma_3+\gamma_2}\right)\right]\ln\Omega_{it} + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\ln\Omega_{jt}$$
$$+ u_{ii}\ln k^*_{it} + \sum_{j\neq i}^{N} u_{ij}\ln k^*_{jt}$$

we obtain

$$\ln y_{it} - \ln y^*_{it} = u_{ii}(\ln k_{it} - \ln k^*_{it}) + \sum_{j\neq i}^{N} u_{ij}(\ln k_{jt} - \ln k^*_{jt}) \tag{4.7.1}$$

Therefore,

$$\frac{d\ln k_{it}}{dt} = g - (n_i+\delta+g)\left(\ln k_{it} - \ln k^*_{it}\right) + (n_i+\delta+g)(\ln y_{it} - \ln y^*_{it})$$

Plugging the previous result into the total derivative of 4.2.6:

$$\begin{aligned}
\frac{d\ln y_{it}}{dt} &= \left[1 + \left(\frac{(\gamma_1+\gamma_2)(u_{ii}-\alpha-\phi)}{\phi\gamma_3+\gamma_2}\right)\right]\mu + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\mu \\
&+ u_{ii}\left(g - (n_i+\delta+g)\left(\ln k_i(t) - \ln k^*_i\right) + (n_i+\delta+g)(\ln y_i - \ln y^*_i)\right) \\
&+ \sum_{j\neq i}^{N} u_{ij}\left(g - (n_j+\delta+g)\left(\ln k_j(t) - \ln k^*_j\right) + (n_j+\delta+g)(\ln y_j - \ln y^*_j)\right) \\
&= \left[1 + \left(\frac{(\gamma_1+\gamma_2)(u_{ii}-\alpha-\phi)}{\phi\gamma_3+\gamma_2}\right)\right]\mu + \frac{(\gamma_3+\gamma_1)}{\phi\gamma_3+\gamma_2}\sum_{j\neq i}^{N} u_{ij}\mu + u_{ii}g + \sum_{j\neq i}^{N} u_{ij}g \\
&- \left(u_{ii}(n_i+\delta+g)\left(\ln k_{it} - \ln k^*_{it}\right) + \sum_{j\neq i}^{N} u_{ij}(n_j+\delta+g))\left(\ln k_{jt} - \ln k^*_{jt}\right)\right) \\
&+ \left(u_{ii}(n_i+\delta+g)\left(\ln y_{it} - \ln y^*_{it}\right) + \sum_{j\neq i}^{N} u_{ij}(n_j+\delta+g))\left(\ln y_{jt} - \ln y^*_{jt}\right)\right)
\end{aligned}$$

The first term in the previous expression corresponds to the balanced growth rate g (see appendix 4.5). As for the second term, let us assume that, for each economy i, there exists

Λ_i such that:

$$\sum_{j=1}^{N} u_{ij}(n_j+g+\delta)(\ln k_{jt}-\ln k_{jt}^*) = \Lambda_i\left(u_{ii}(\ln k_{it}-\ln k_{it}^*)+\sum_{j\neq i}^{N} u_{ij}(\ln k_{jt}-\ln k_{jt}^*)\right)$$

Thus,

$$\begin{aligned}
\frac{d\ln y_{it}}{dt} &= g-\Lambda_i\left(u_{ii}(\ln k_{it}-\ln k_{it}^*)+\sum_{j\neq i}^{N} u_{ij}(\ln k_{jt}-\ln k_{jt}^*)\right) \\
&+ u_{ii}(n_i+\delta+g)\left(\ln y_{it}-\ln y_{it}^*\right)+\sum_{j\neq i}^{N} u_{ij}(n_j+\delta+g)\left(\ln y_{jt}-\ln y_{jt}^*\right) \\
&= g-\Lambda_i(\ln y_{it}-\ln y_{it}^*)+u_{ii}(n_i+\delta+g)\left(\ln y_{it}-\ln y_{it}^*\right)+\sum_{j\neq i}^{N} u_{ij}(n_j+\delta+g)\left(\ln y_{jt}-\ln y_{jt}^*\right)
\end{aligned}$$

where the second expression is obtained by using 4.7.1.

Finally, from the first hypothesis in 4.2.11 we have that $\left(\ln y_{it}-\ln y_{it}^*\right)\Theta_j^{-1}=\ln y_{jt}-\ln y_{jt}^*$. This allows us to obtain the speed of convergence to the steady state:

$$\begin{aligned}
\frac{d\ln y_{it}}{dt} &= g-\left(\Lambda_i-u_{ii}(n_i+\delta+g)-\sum_{j\neq i}^{N} u_{ij}(n_j+\delta+g)\Theta_j^{-1}\right)\left(\ln y_{it}-\ln y_{it}^*\right) \\
&= g-\lambda_i\left(\ln y_{it}-\ln y_{it}^*\right)
\end{aligned}$$

What is left is to derive the expressions defining Λ_i and λ_i. First, by plugging the second hypothesis in 4.2.11, $\left(\ln k_{it}-\ln k_{it}^*\right)\Phi_j^{-1}=\ln k_{jt}-\ln k_{jt}^*$, into our assumption on the existence of Λ_i:

$$\begin{aligned}
\sum_{j=1}^{N} u_{ij}(n_j+g+\delta)(\ln k_{jt}-\ln k_{jt}^*) &= \Lambda_i\left(u_{ii}(\ln k_{it}-\ln k_{it}^*)+\sum_{j\neq i}^{N} u_{ij}(\ln k_{jt}-\ln k_{jt}^*)\right) \\
\sum_{j=1}^{N} u_{ij}(n_j+g+\delta)(\ln k_j-\ln k_j^*) &= \Lambda_i\sum_{j=1}^{N} u_{ij}(\ln k_j(t)-\ln k_j^*) \\
\sum_{j=1}^{N} u_{ij}(n_j+g+\delta)\left(\ln k_i(t)-\ln k_i^*\right)\Phi_j^{-1} &= \Lambda_i\sum_{j=1}^{N} u_{ij}\left(\ln k_i(t)-\ln k_i^*\right)\Phi_j^{-1} \\
\Lambda_i &= \frac{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}(n_j+g+\delta)}{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}}
\end{aligned} \tag{4.7.2}$$

Second, plugging the previous result into $\lambda_i=\Lambda_i-u_{ii}(n_i+\delta+g)-\sum_{j\neq i}^{N} u_{ij}(n_j+\delta+g)\Theta_j^{-1}$

and assuming that $\Theta_i^{-1} = 1$:

$$\lambda_i = \Lambda_i - u_{ii}(n_i + \delta + g)\Theta_i^{-1} - \sum_{j\neq i}^{N} u_{ij}(n_j + \delta + g)\Theta_j^{-1}$$

$$\lambda_i = \Lambda_i - \sum_{j=1}^{N} u_{ij}(n_j + \delta + g)\Theta_j^{-1}$$

$$\lambda_i = \frac{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}(n_j + g + \delta)}{\sum_{j=1}^{N} u_{ij}\frac{1}{\Phi_j}} - \sum_{j=1}^{N} u_{ij}(n_j + \delta + g)\frac{1}{\Theta_j}$$

4.8 Appendix D: Differential equation solution

We start by noticing that the steady state in 4.2.9 can be written as

$$\ln y_{it}^* = \frac{1}{1-\alpha-\phi}\sum_{j=1}^{N}\sum_{r=0}^{\infty}\rho^r w_{ij}^{(r)}\ln\Omega_{jt} + \frac{\gamma_1}{1-\alpha-\phi}\sum_{j=1}^{N}\sum_{r=0}^{\infty}\rho^r w_{ij}^{(r+1)}\ln\Omega_{jt}$$
$$+ \left(\frac{\alpha+\phi}{1-\alpha-\phi}\right)\sum_{j=1}^{N}\sum_{r=0}^{\infty}\rho^r w_{ij}^{(r)}\ln\left(\frac{s_j}{n_j+\delta+g}\right) + \frac{\gamma_2-\alpha\gamma_3}{1-\alpha-\phi}\sum_{j=1}^{N}\sum_{r=0}^{\infty}\rho^r w_{ij}^{(r+1)}\ln\left(\frac{s_j}{n_j+\delta+g}\right)$$

with $\rho = \frac{\gamma_2-\alpha\gamma_3}{1-\alpha-\phi}$. Using this, we can see that $\frac{d\ln y_{it}^*}{dt} = \frac{(1+\gamma_1)\mu}{1-\alpha-\phi}\left(\frac{1}{1-\rho}\right)$. In fact, since $\frac{1}{1-\rho} = \frac{1-\alpha-\phi}{(1-\alpha)(1-\gamma_3)-\phi-\gamma_2}$,

$$\frac{d\ln y_{it}^*}{dt} = g \tag{4.8.1}$$

Notice that 4.8.1 can be seen as another differential equation, which have a particular solution on $\ln y_{i0}^*$:

$$\ln y_{it}^* = gt + \ln y_{i0}^* \tag{4.8.2}$$

Plugging equation 4.8.2 and 4.2.12 we obtain:

$$\frac{d\ln y_{it}}{dt} = g - \lambda_i\left(\ln y_{it} - gt - \ln y_{i0}^*\right) \tag{4.8.3}$$

We use the integrating factor method to solve the differential equation in 4.8.3. We first reorder terms and then multiply the equation by the integrating factor $e^{\int \lambda_i dt} = e^{\lambda_i t}$ to obtain

$$\frac{d}{dt}\left(e^{\lambda_i t}\ln y_{it}\right) = e^{\lambda_i t}g + \lambda_i e^{\lambda_i t}\left(gt + \ln y_{i0}^*\right)$$

By integrating on both sides, we obtain the general solution:

$$\ln y_{it} = gt + \ln y_{i0}^{*} + Ce^{-\lambda_i t}$$

The particular solution for $t = t_1$ implies that $C = \left(\ln y_{it_1} - gt_1 - \ln y_{i0}^{*}\right) e^{\lambda_i t_1}$. Thus, for any t we have:

$$\ln y_{it} = g\left(t - t_1 e^{-\lambda_i (t-t_1)}\right) + \ln y_{it_1} e^{-\lambda_i (t-t_1)} + \left(1 - e^{-\lambda_i (t-t_1)}\right) \ln y_{i0}^{*}$$

Chapter 5

Conclusions and Future Research

This thesis contributes to the spatial econometrics literature by presenting a suitable framework for modelling the spatial spillovers of the individual effects in panel data models. The key to this new approach is the use of a correlated random effects specification. In essence, each chapter of this thesis considers this approach in a different model specification and provides illustrative evidence. Next I summarise the main findings of each chapter and discuss future research extensions.

The second chapter analyses the spatial $X-$lag model for panel data. It is first showed that the individual effects and their spatial spillover are not generally identified. However, they are identified in a correlated random effects specification provided that some mild rank conditions on the covariates hold. Further, this chapter proposes using FGLS and IV estimators under strict and sequential exogeneity assumptions on the covariates. Lastly, results from an empirical application based on a Cobb-Douglas production function and US state data are presented. While the main findings are largely consistent with previous literature, there is also evidence of "inward" and "outward" spatial effects of the individual effects. FGLS and IV estimates are found to differ substantially, which indicates that the strict exogeneity assumption may not hold. The small samples properties of these estimators may be addressed in future research. Another extension of this study is to analyse the spatial contagion of individual effects for the case of time varying spatial weights matrices.

The third chapter considers a correlated random effects specification of a spatial Durbin dynamic model. It presents the likelihood function of the model and prove that the QML estimator is consistent and asymptotically normal when the initial period ($t = 0$) of the dependent variable is exogenous, the number of spatial units (N) is large, and the time periods (T) are fixed. The case where the initial value of the dependent variable is endogenously given will be handle it in future research. In addition, this chapter presents an illustrative empirical application examining the existence of spillovers effects in economic growth. Estimates from a sample of 26 OECD countries over the period 1971 to 2005 show that countries with a small/large estimated individual effect tend to be among the most/least affected by the impact of the estimated individual effects of their neighbours and among those whose individual effects impact most/least on the other countries (in terms of geographical distance).

Finally, the fourth chapter extends previous knowledge-spillovers models of growth in several directions. First, it does not assume a common exogenous technological progress but accounts for heterogeneity in the initial level of technology. Second, it assumes that the technological progress depends not only on the stock of physical capital and the stock of knowledge of the other economies, but also on the physical capital and the (unobserved) initial level of technology of the other economies. Using EU-NUTS2 regional data from Cambridge Econometrics, this chapter also tests whether the data supports the main features of the proposed growth model. In particular, the econometric specification derived from the growth-initial equation of the model takes the form of a spatial Durbin dynamic panel model with spatially weighted individual effects. Results support the proposed model specification. Also, they are largely *i)* consistent with other studies using analogous data; and *ii)* robust to the use of alternative specifications, samples and estimation approaches. Future research should investigate the use of alternative assumptions on the technology to achieve identification.

Bibliography

Abreu, M., de Groot, H., and Florax, R. (2005). Space and growth: A survey of empirical evidence and methods. *Région et Développement*, 21:13–44.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):153–166.

Baltagi, B. and Pinnoi, N. (1995). Public capital stock and state productivity growth: Further evidence from an error components model. *Empirical Economics*, 20(2):351–59.

Barro, R. J. and Sala-i-Martin, X. (2003). *Economic Growth, 2nd Edition*. The MIT Press.

Basile, R. (2008). Regional economic growth in Europe: A semiparametric spatial dependence approach. *Papers in Regional Science*, 87(4):527–544.

Beer, C. and Riedl, A. (2012). Modelling spatial externalities in panel data: The spatial durbin model revisited. *Papers in Regional Science*, 91(2):299–318.

Beron, K. J., Hanson, Y., Murdoch, J. C., and Thayer, M. A. (2004). *Hedonic Price Functions and Spatial Dependence: Implications for the Demand for Urban Air Quality*, pages 267–281. Springer Berlin Heidelberg, Berlin, Heidelberg.

Boarnet, M. G. (1998). Spillovers and the locational effects of public infrastructure. *Journal of Regional Science*, 38(3):381–400.

Brueckner, J. K. (2003). Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*, 26(2):175–188.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46.

Chamberlain, G. (1984). Panel data. In Griliches, Z. and Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, chapter 22, pages 1247–1318. Elsevier, 1 edition.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, V., and Strange, W., editors, *Handbook of Urban and Regional Economics, vol. 5A*, pages 247–348. Elsevier.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

De Long, J. B. and Summers, L. (1991). Equipment investment and economic growth. *The Quarterly Journal of Economics*, 106(2):445–502.

Debarsy, N. (2012). The Mundlak Approach in the Spatial Durbin Panel Data Model. *Spatial Economic Analysis*, 7(1):109–131.

Debarsy, N., Ertur, C., and Lesage, J. P. (2012). Interpreting dynamic space-time panel data models. *Statistical Methodology*, 9:158–171.

Egger, P. and Pfaffermayr, M. (2006). Spatial convergence. *Papers in Regional Science*, 85(2):199–215.

Elhorst, J. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Springer.

Elhorst, J. P. (2012). Dynamic spatial panels: models, methods, and inferences. *Journal of Geographical Systems*, 14(1):5–28.

Elhorst, P., Piras, G., and Arbia, G. (2010). Growth and convergence in a multiregional model with spaceï£¡time dynamics. *Geographical Analysis*, 42(3):338–355.

Ertur, C. and Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22(6):1033–1062.

Fingleton, B. and López-Bazo, E. (2006). Empirical growth models with spatial effects. *Papers in Regional Science*, 85(2):177–198.

Gänsler, P. and Stute, W. (1977). *Wahrscheinlichkeitstheorie*. Springer Verlag.

Garcia-Mila, T., McGuire, T. J., and Porter, R. H. (1996). The Effect of Public Capital in State-Level Production Functions Reconsidered. *The Review of Economics and Statistics*, 78(1):177–80.

Halleck Vega, S. and Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, 55(3):339–363.

Harville, D. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer.

Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, 49(6):1377 – 1398.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60.

Ho, C.-Y., Wang, W., and Yu, J. (2013). Growth spillover through trade: A spatial dynamic panel data approach. *Economics Letters*, 120(3):450–453.

Holtz-Eakin, D. and Schwartz, A. (1995). Spatial productivity spillovers from public infrastructure: Evidence from state highways. *International Tax and Public Finance*, 2(3):459–468.

Islam, N. (1995). Growth empirics: A panel data approach. *The Quarterly Journal of Economics*, 110(4):1127–1170.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1):97 – 130. Analysis of spatially dependent data.

Keane, M. P. and Runkle, D. E. (1992). On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *Journal of Business & Economic Statistics*, 10(1):1–9.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2):219 – 257.

Kelejian, H. H. and Robinson, D. P. (1997). Infrastructure productivity estimation and its underlying econometric specifications: A sensitivity analysis. *Papers in Regional Science*, 76(1):115–131.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925.

Lee, L.-F. and Yu, J. (2016). Identification of spatial durbin panel models. *Journal of Applied Econometrics*, 31(1):133–162.

LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *The Review of Regional Studies*, 44(1):13–323.

LeSage, J. and Chih, Y.-Y. (2016). Interpreting heterogeneous coefficient spatial autoregressive panel models. *Economics Letters*, 142:1 – 5.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. CRC Press.

LeSage, J. P. and Fischer, M. M. (2012). Estimates of the impact of static and dynamic knowledge spillovers on regional factor productivity. *International Regional Science Review*, 35(1):103–127.

López-Bazo, E., Vayá, E., and Artís, M. (2004). Regional externalities and growth: Evidence from European regions. *Journal of Regional Science*, 44(1):43–73.

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

Mankiw, N. G., Romer, D., and Weil, D. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107(2):407–437.

Miranda, K., Martínez-Ibañez, O., and Manjón-Antolín, M. (2017a). A correlated random effects spatial durbin model. Department of Economics, Rovira i Virgili Universiy.

Miranda, K., Martínez-Ibañez, O., and Manjón-Antolín, M. (2017b). Estimating individual effects and their spatial spillovers in linear panel data models: Public capital spillovers after all? *Spatial Statistics*, 22(Part 1):1–17.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Munnell, A. H. (1990). How does public infrastructure affect regional economic performance? In Munnell, A. H., editor, *Is There a Shortfall in Public Capital Investment?*, pages 69–103. Federal Reserve Bank of Boston.

Pereira, A. and Andraz, J. (2013). On the economic effects of public infrastructure investment: A survey of the international evidence. *Journal of Economic Development*, 4(38):1–37.

Pfaffermayr, M. (2009). Conditional [beta]- and [sigma]-convergence in space: A maximum likelihood approach. *Regional Science and Urban Economics*, 39(1):63–78.

Pfaffermayr, M. (2012). Spatial Convergence Of Regions Revisited: A Spatial Maximum Likelihood Panel Approach. *Journal of Regional Science*, 52(5):857–873.

Sloboda, B. W. and Yao, V. W. (2008). Interstate spillovers of private capital and public spending. *The Annals of Regional Science*, 42(3):505–518.

Su, L. and Yang, Z. (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics*, 185(1):230 – 258.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Econometric Analysis of Cross Section and Panel Data. MIT Press.

Yu, J., de Jong, R., and Lee, L.-F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1):118 – 134.

Yu, J. and Lee, L.-F. (2012). Convergence: A spatial dynamic panel data approach. *Global Journal of Economics*, 1(1):1–36.