A hierarchical spatial Durbin model (HSDM): An application to regional production efficiency in Europe

Alejandro Almeida Márquez
Universidad Internacional de la Rioja https://orcid.org/0000-0001-6490-4998
Julián Ramajo Hernández (ramajo@unex.es)
Universidad de Extremadura https://orcid.org/0000-0002-3156-8315

Research Article

Keywords: Hierarchical models, Spatial econometrics, Geographical externalities, European regions.

Posted Date: March 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1400376/v1

A hierarchical spatial Durbin model (HSDM): An application to regional production efficiency in Europe (FIRST DRAFT)

Alejandro Almeida (UNIR) and Julián Ramajo (UEX)

1 Introduction

In recent years, the development of different fields in regional analysis has grown thanks, in part, to the availability of disaggregated and nested data (at the local, provincial, regional, national or supranational level), as is the case with the 3000 counties nested in 50 states in the United States or, the different levels of NUTS (Nomenclature of Territorial Units for Statistics) in the European Union (EU).

Two main fields have tried to develop and implement models to analyze the interactions that take place between different geographic areas. On the one hand, the literature of spatial econometric models (see Vanoutrive and Parenti (2009)) has developed different model specifications based on the premise that closest geographical areas will be more related than those that are further away (Elhorst, 2014b). This definition of interaction is important because it is assumed that the relationship between the provinces is given for a particular reason, geographic proximity (Elhorst, 2014b). On the other hand, the literature of multilevel models (also known as hierarchical models, see Finch, Bolin, and Kelley (2019) for a recent review) has also had a great development in recent years, but with a different concept of the relationship between geographical areas. Specifically, the literature on multilevel models understands that the relationship between different geographical areas is produced by having in common a set of characteristics, for example, regions that belong to the same country.

Both types of models have been used empirically several times. For example, some case studies have traditionally been used for the development of spatial econometrics techniques because distance plays an important role in the relationship that exists between different geographical areas of analysis. This has been the case of the analysis of tobacco consumption (see Finch, Bolin, and Kelley (2019),Debarsy (2012) or Debarsy (2012)) or the investigation of inequalities and convergence between the European regions (Geppert and Stephan, 2008; Le Gallo and Dall'Erba, 2008). On the other hand, the use of multilevel models has also had numerous applications. Examples can be found at housing market (Jones, 1991; Dong and Harris, 2015) or

health (Langford et al., 1999).

Empirically, depending on the case study, the presence of these two types of characteristics can be found simultaneously or not. For this reason, in recent years, some works have tried to bring both fields of regional analysis closer together to develop models that can take into account the relationships that occur between different geographical areas due to their proximity and the fact of sharing factors (to be nested) (Lacombe and McIntyre, 2017).

The first work that consider a spatial econometric model in a hierarchical context was, to the best of our knowledge, Anselin and Florax (1995) to backcast school district income tax revenues. From this work, some research continued to bring both fields closer (Langford et al., 1999; Anselin, 2001) and it was not until Anselin and Cho (2002) that the concept of hierarchical spatial econometrics models began to be more discussed in depth. With the work of Smith and LeSage (2004) different hierarchical spatial econometrics models began to be developed. Since then, some works have developed different model specifications, being one of the most recent applications, the model developed by Dong and Harris (2015). This work develops a hierarchical spatial autoregressive model to accommodate a hierarchical data structure to the traditional SAR model of spatial econometrics. Specifically, it allows estimating spatial spillover effects while also controlling and analyzing the existence of group effects.

The objective of this work is to continue with the development and application of hierarchical spatial econometrics models that allow for the existence of interactions between geographic units in data with a hierarchical structure. Specifically, another traditional model of spatial econometrics is developed in a hierarchical structure context, a hierarchical spatial Durbin model (HSDM) based on the work of Dong and Harris (2015).[1].

To check the usefulness of the HSDM model, we estimate this model using a data set from 263 regions nested in 28 countries. This data set contains information on the production (Y) of the European regions and countries as well as two classic inputs, physical capital (K) and employment (L). These data allow us to apply a hierarquical spatial Durbin model to analyze the economic growth of the European regions, since the total productivity factor (TFP) is considered the most important driver behind economic growth (Parente and Prescott, 2005).

Many studies have analyzed the convergence process of the regions in Europe (Cuaresma, Doppelhofer, and Feldkircher, 2014; Piribauer, 2016) even taking into account the presence of spatial correlation (Ramajo and Hewings, 2018). However, to the best of our knowledge, none has taken into account the nested structure that production data presents. Regional data nested in countries. This natural hierarchical structure of the data is used to model the presence of horizontal spillovers (influence between regions or between countries) and vertical spillovers (influence of countries in regions). Specifically, this model allows us to estimate three parameters whose interpretation is of interest. On the one hand, the spatial dependence parameters between regions and between countries that allow us to know at what

[1]See Elhorst (2014b) for an extensive review on the different specifications of spatial econometric models

scale there are greater spillovers in terms of production, and, on the other hand, the random effects that each country has on its regions, where, greater effects may indicate a better productive context in the country. Differences in these random coefficients could show the heterogeneity between countries in Europe.

The rest of the paper is divided as follow. In Methodology section, we explain the econometric strategy used as well as the different weight matrices, in the Results section we present the main results of the application of this model for the specific case study and in Conclusion section we present the main conclusions and implications of this work.

2 Methodology

To apply the hierarchical spatial Durbin model proposed, the empirical analysis focuses on 263 NUTS-2 regions in the 28 European Union countries, excluding the overseas territories of Finland, France, Portugal and Spain. The data used in the empirical application were taken from the Cambridge Econometrics' European Regional Database (ERD) 2016 release that contains complete yearly information for the period 1990-2014 at the regional NUTS-2 classification of the European Union.[2]From the ERD, the following variables were calculated or estimated:

- Regional output (Y), measured as gross value added -GVA- in each region in constant 2005 purchasing power standards -PPS- terms. The original GVA at constant prices time series (measured in €2005m) were adjusted for price differences across countries and over the time with country-specific PPS's.
- Regional labor (L), measured as total employment in each region in 000s of people.
- Gross fixed capital formation (I), measured in €2005m.

To obtain estimations of regional physical capital stocks (K), the perpetual inventory method (PIM) was employed using yearly regional gross fixed capital formation (I) series through the formula $K_{it} = I_{it} + (1 - \delta) K_{i,t-1}$.

The natural hierarchical structure of data brings the necessity of modelling the data taking into account the possible effects that the conditions of each country have on the regions. For this, the multilevel model literature proposes several models to incorporate into the regional modelling the effects of the higher level (national) through fixed or random effects (see Finch, Bolin, and Kelley (2019) for a review of hierarchical models).

In our case study, we will use a hierarchical random intercept model (Raudenbush and Bryk, 2002) following Dong and Harris (2015) procedure.

As a starting point, we focus on the traditional SDM model for the regional production function (Elhorst, 2014b; LeSage and Pace, 2010). This model takes the form:

[2]The primary source of the ERD is the Eurostat's REGIO database, supplemented with the European Commission's AMECO database. The 2016 release of ERD uses the NUTS 2010 regional classification.

$$y = \rho_1 W_1 y + X\beta + W_1 X\theta + \epsilon \tag{1}$$

where y is the vector of observations of the dependent variable (regional production, Y), ρ_1 is the spatial auto-regressive parameter at regional level, W_1 is the regional weight matrix, X is the matrix of explanatory variables (physical capital, K and employment, L), β and θ are the vector of coefficients of response to the explanatory variables and ϵ is the vector of disturbances.

To extend this model to a traditional hierarchical model, we follow the procedure carried out in Dong and Harris (2015), where the hierarchical random intercept model is used and the effects of the countries on the regions are models through random effects. Furthermore, instead of assuming the traditional multilevel model with independent random higher level (national) effects (Jones, 1991), they relax this restriction allowing the random effects to be dependent. This reasoning is applicable in our case study since the countries are also geographically continuous, so it is expected that the effect of a given country is similar to that of its neighbouring countries.

Specifically, the extension of the SDM model to the HSDM model takes the form at regional level [3]:

$$y = \rho_1 W_1 y + X\beta + W_1 X\theta + \Delta\alpha + \epsilon \tag{2}$$

where Δ represent a matrix that assigns each region to a country and α is the vector of random intercepts and dependent variable of the national level as follow:

$$\alpha = \rho_2 W_2 \alpha + u \tag{3}$$

where ρ_2 is the spatial auto-regressive parameter at national level, W_2 is the national weight matrix and u is the vector of disturbances.

As observed, the proposed model allows us to model a SDM process at the regional level, where the vertical spillovers that the countries have over the regions, are also taken into account, assuming that these interactions are dependent, that is, that the countries also influence each other, and those who are closer have similar behaviors. Furthermore, the random effects that the upper levels (countries) have on the lower levels (regions) can be estimated and interpreted as the national conditions inherent to each of the countries in our model.

In the proposed formulation, it is necessary to define three matrices, two spatial matrices (W_1 and W_2) and a matrix that assigns each region to the country it belongs

[3] We use the notation proposed in Lacombe and McIntyre (2017)

to (Δ).

Matrix Δ, is a matrix of dummy variables that relates each region to the country to which it belongs.

Matrix W_1 is the lower-level spatial weight matrix (regions) and matrix W_2 is the upper-level spatial weight matrix (countries). To select the type of weight matrix to use, we opted to use the specifications used by Dong and Harris (2015) where the regional weight matrix is a negative exponential matrix of the distance squared and the national weight matrix is a matrix based on the contiguity of the countries. Both matrices have been standardized. [4]

Figure 1 shows the relationship maps generated by the W_1 and W_2 matrices explained above where it can be checked the structure of regional and national relationships used for our model.

3 Results

Results are divided into three parts where the results of the model estimates are found in tables 1, 2 and 3 and the results of the national random effects are found in Figure 2. First, from the econometric point of view, we analyze the results obtained using the HSDM model with respect to the HSAR model proposed by Dong and Harris (2015). Second, we analyze the estimated parameters of the model for each year where the evolution over time of the influence of physical capital (K) and employment (L) on production and the comparison between regional and national spatial dependence can be observed. And third, we investigate the estimated random effects of each country to see the evolution of heterogeneity and to know which country has a better productive context.

Regarding the comparison of the HSAR model with the HSDM in the three years, it is observed that, following the log-likelihood, the HSDM model is slightly better for the applied case although they are very similar. The estimated coefficients do not have significant changes and the interpretations in both models are very similar. However, the inclusion of the regressive spatial parameters of the explanatory variables is significant, which implies that spillovers between regions are produced not only at the production level but also through labour and capital, being in both cases a negative effect that may indicate the existence of competitiveness in employment and capital among the european regions.

Regarding the interpretation of the estimated parameters of the HSDM model, it can be seen that the spatial dependence in terms of production is positive and significant both at the regional and national level, however, the national spatial dependence is significantly higher than the regional one, which implies that at the national level there are greater spillovers than at the regional level. On the other hand, there seems to be a change regarding the influence of capital and labour on production in regions where, in 2000 the influence of labour is greater than that of

[4] We also estimate the models using other regional and national matrix specifications with weights based on distance, the inverse of distance, the k nearest neighbors, and contiguity.

FIGURE 1: Regional (W_1) and national (W_2) weight matrix.

(A) W_1

(B) W_2

Note: W_1: negative exponential matrix of the distance squared. W_2: Queen contiguity matrix.

Table 1: Results 2000

2000	*HSAR MODEL*		*HSDM MODEL*	
	Value	SE	Value	SE
ρ_1	0.108	0.027	0.297	0.063
ρ_2	0.771	0.113	0.817	0.115
L	0.582	0.041	0.588	0.041
K	0.481	0.038	0.472	0.028
W_1L			−0.133	0.068
W_1K			-0.124	0.087
Constant	−0.437	0.395	0.260	0.581
Observations (NUTS2)	263		263	
Countries	28		28	
Pseudo R^2	0.981		0.976	
Log likelihood	-4860.285		-5068.117	

Note:

capital, while in 2007 and 2014 the influence of labour is less than that of capital.

TABLE 2: Results 2007

2007	*HSAR MODEL*		*HSDM MODEL*	
	Value	SE	Value	SE
ρ_1	0.073	0.027	0.275	0.082
ρ_2	0.783	0.129	0.698	0.145
L	0.460	0.047	0.456	0.045
K	0.604	0.045	0.603	0.043
W_1L			–-0.185	0.063
W_1K			-0.075	0.094
Constant	–-0.513	0.456	-0.564	0.468
Observations (NUTS2)	263		263	
Countries	28		28	
Pseudo R^2	0.982		0.979	
Log likelihood	-5931.496		-6040.984	

Note:

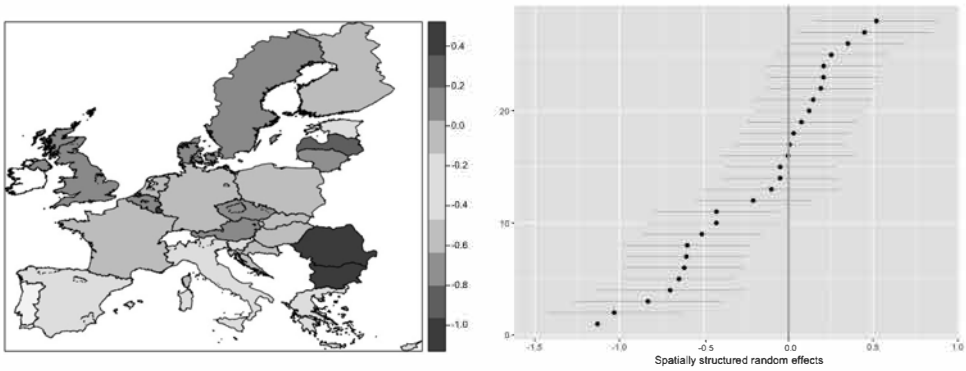
Finally, figure 2 summarizes the estimation of the national random coefficients representing them in maps and caterpillars plots. The maps show the spatial distribution of these random effects that could be interpreted as the national context, with positive values being a favourable context and negative values being an unfavourable context. For our specific case study, a clear geographic pattern is detected in Europe where the countries with a more favourable context are located in the north (with the exception of countries in Eastern Europe), while the most unfavourable context is in eastern Europe. Southern Europe appears to have a neutral context for production. Through the caterpillars, we can observe the dispersion of the estimated coefficients that could be interpreted as homogeneity in the national context of the countries. As it appears, the heterogeneity of the country context seems to decrease in 2007 compared to 2000, however, in 2014, it appears that heterogeneity increases slightly. This could be explained by periods of crisis and expansion in Europe.

Table 3: Results 2014

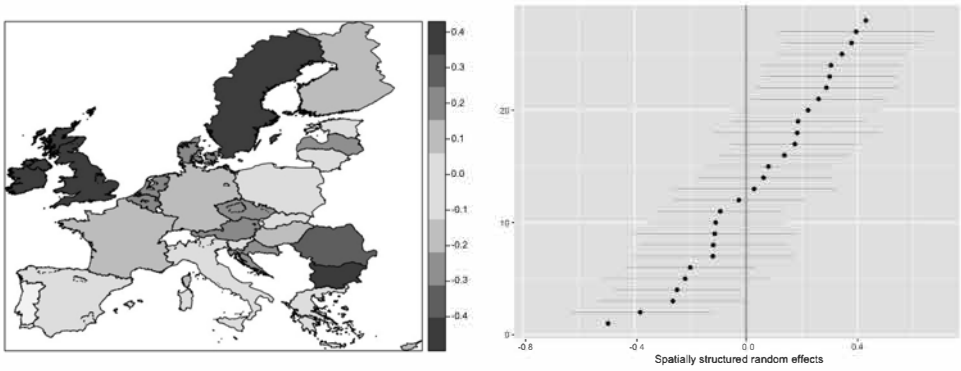
2014	*HSAR MODEL*		*HSDM MODEL*	
	Value	SE	Value	SE
ρ_1	0.071	0.025	0.270	0.083
ρ_2	0.882	0.104	0.692	0.159
L	0.410	0.044	0.403	0.045
K	0.660	0.042	0.666	0.044
W_1L			−0.175	0.059
W_1L			-0.075	0.093
Constant	−1.457	0.462	-0.987	0.471
Observations (NUTS2)	263		263	
Countries	28		28	
Pseudo R^2	0.983		0.980	
Log likelihood	-6320.126		-6370.142	

Note:

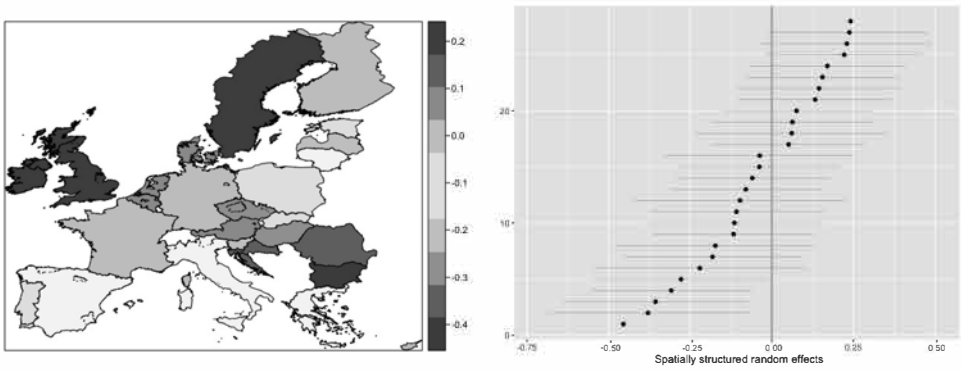
Figure 2: National random effects maps and caterpillar plots.



(a) 2000 - Map (b) 2000 - Caterpillar



(c) 2007 - Map (d) 2007 - Caterpillar



(e) 2014 - Map (f) 2014 - Caterpillar

4 Conclusions

The hierarchical model literature has tried to include in its models the traditional parameters of spatial econometrics to be able to model situations of horizontal and vertical influences simultaneously when we work with nested data.

To continue expanding the recent literature on hierarchical models of spatial econometrics, this work proposes an extension of the autoregressive spatial hierarchical model (HSAR) to a Durbin spatial hierarchical model (HSDM) that allows taking into account the spillovers produced in the independent variables.

For this, the proposed model is estimated to analyze the production function of 263 European regions nested in 28 different countries for years 2000, 2007 and 2014. The results seem to indicate that the proposed model produces results similar to the HSAR model or improves them taking into account the influence that capital and employment levels may have in other regions. Furthermore, this model allows analyzing the process of regional and national spillovers and, country influence on the regions.

Particularly, it seems to show that national spillovers are higher than regional spillovers in terms of production levels, although positive in both cases; however, regional influences in terms of capital and labor are negative, which could show regional competitiveness at European level in employment and capital.

Concerning the national context, there seems to be a heterogeneity between the European countries where the countries of northern Europe, except for those located in the northeast, present favorable contexts, while the eastern countries present unfavorable contexts. The temporal evolution in our analysis seems to suggest that the heterogeneity of the regions decreased in 2007 compared to 2000, however, in 2014, a slight increase in heterogeneity appears to be observed compared to 2007.

In summary, the development of new models such as the HSDM proposed in this work that take into account the vertical and horizontal spillovers that some economic models present, seems to be useful to find new evidences. Future research is necessary from the econometric point of view, for the development or improvement of these models, and, from the empirical point of view, to deepen the meaning and interpretation of some parameters estimated in these models, such as the case of random national effects.