TÍTULO

VERTICAL SPILLOVERS IN SPATIAL ECONOMETRICS

AUTOR

Alejandro Almeida Márquez

	Esta edición electrónica ha sido realizada en 2022
Fecha de lectura	29/10/2020
Instituciones	Universidad Internacional de Andalucía ; Universidad de Huelva
Director	Dr. D. Antonio Aníbal Golpe Moya
Programa de Doctorado	Doctorado en Economía, Empresa, Finanzas y Computación
ISBN	978-84-7993-598-6
©	Alejandro Almeida Márquez
©	De esta edición: Universidad Internacional de Andalucía
Fecha documento	2020

International University of Andalucía and
University of Huelva

Doctoral Thesis

Vertical Spillovers in Spatial Econometrics

Author:
Alejandro Almeida Márquez

Supervisor:
Dr. Antonio A. Golpe Moya

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Doctoral Programme: Economics, Business, Finance and Computing Science

July 15, 2020

Declaration of Authorship

I, Alejandro Almeida Márquez, declare that this thesis titled, "Vertical Spillovers in Spatial Econometrics" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ______________________________

Date: ______________________________

“We must understand the Cosmos as it is and not confuse how it is with how we wish it to be.”

Carl Sagan

INTERNATIONAL UNIVERSITY OF ANDALUCÍA AND UNIVERSITY OF HUELVA

Abstract

Doctoral Programme: Economics, Business, Finance and Computing Science

Doctor of Philosophy

Vertical Spillovers in Spatial Econometrics

by Alejandro Almeida Márquez

Spatial econometrics has studied and analyzed the horizontal interactions that take place between different geographic locations. The proximity between two locations makes them behave more similarly than those locations that are further away. The development of this literature has been possible, in part, due to the increase in disaggregated data at the geographical level. This disaggregation also allows us to have data at different geographic scales (i.e., provinces, regions, and countries), ending in nested data sets. This nested nature of the data allows and generates the need to take into account the possible vertical spillovers that occur when a higher scale can influence the lower scales, for example, countries that influence their regions. In recent years, some authors have proposed different models that allow the inclusion of both types of interactions, vertical and horizontal. However, the literature and the empirical applications are still scarce. For this reason, this thesis tries to empirically analyze these models and to develop new models that allow progress in the inclusion of vertical spillovers in the field of spatial econometrics. Through applications in the sensitivity of the regions to the economic cycle, self-employment, cigarette consumption and the productivity of the European countries and regions, different proposed models are analyzed, such the dynamic spatial econometrics model with common factors and hierarchical spatial econometrics models. Chapter 2 analyze which regions are more sensitive to aggregate fluctuations, finding a pattern for Spain where the most sensitive regions are on the Mediterranean coast. Chapter 3 analyzes the spatial dynamics of self-employment in the United States, finding a relationship between high self-employment clusters and sensitivity to the national cycle. In chapter 4 and 5, cigarette consumption in the Spanish provinces is analyzed and the price is modelled as a common national factor, finding heterogeneity in the behaviour of the provinces. Finally, Chapter 6 develops an HSD model of spatial econometrics in a hierarchical context and is applied to analyze the production of European regions and the influence of countries on them.

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Observations of an economic phenomenon are normally made at different scales and levels of geographic disaggregation and are followed over time. Available data sets have a series of characteristics that need to be taken into account in the statistical analysis methods.

The greater availability of geographically disaggregated data means that the analyzes can be carried out at different levels (district, provincial, regional or national, among others). Statistical methods to analyze this type of data structure have been developed in two main fields, the literature of spatial econometric models and the literature of hierarchical models or multilevel models.

The literature on spatial econometric models has had a great development in recent years, extending traditional models (see Elhorst (2014a) for an extensive review of the literature). Specifically, this literature tries to incorporate into the econometric models the interactions that take place between the regions because they are geographically close. Here location becomes meaningful when considering the position of one place relative to another, because as Tobler's First Law of geography indicates: "everything is related to everything else, but near things are more related than distant things" (Anselin, 2002; Miller, 2004). This characteristic in applied economics is useful to analyze phenomena where interactions between different locations can explain part of the results obtained in the analyzes. Many case studies have been re-estimated taking these interactions into account, finding new results as the case of the work presented in chapter 4.

The hierarchical or multilevel literature focuses on analyzing the nested structure that many data sets have. That is observations that are registered at different geographic scales, for example, regions and countries. This type of observations with different hierarchical levels are common in economics and as Langford, Bentham, and McDonald (1998) indicates: "including higher levels of geographical aggregation simultaneously in a model of smaller units is essential to draw useful conclusions from the data analyzed". Some specific characteristics of the data must be measured at the upper hierarchical level by their nature, while others are observed at the lower level. Thus, hierarchical or multilevel models have attempted to incorporate contextual factors into regressions to investigate the role of larger geographic scales. These models have not been widely applied in the applied economic literature, although some examples can be found in health economics (Rice and Jones, 1997) or housing market (Deboosere et al., 2019).

Another literature that works with different geographic scales is the Global VAR (GVAR) literature, see Chudik and Pesaran (2016) for an extensive review of these models. Although this literature focuses on large global macroeconomic models of the world economy, the econometric reasoning that it uses to incorporate vertical interactions has been useful for the development of spatial econometric models as we will see later.

All of the models used in that literatures could be extended taking into account the temporal dimension, that is, observing the phenomena over time and taking into account the possible temporal components of the case. This has been developed by the time-series literature and different models have been extended to incorporate the time dimension. Elhorst (2014a) is a good example in the spatial econometrics literature.

As we have seen previously, the two main literatures that try to model and measure interactions between geographical areas address it from two different perspectives. Specifically, it can be said that the spatial econometrics literature analyzes horizontal relationships, while the hierarchical models analyze vertical relationships between different geographical scales.

Thanks to the increasing availability of regional data, improvements in statistical theory and advances in computation, some researchers have tried to bring the two fields closer together to use models that take into account the vertical and horizontal spillovers that the data present.

This vein of literature is still scarce and more research is necessary for the development of new models and the empirical application of these that allow us to analyze their usefulness.

1.1 Econometric framework

Both fields of literature, spatial econometrics models and hierarchical or multilevel models, have attempted to extend the models that were traditionally used to incorporate vertical and horizontal effects respectively.

From the field of spatial econometrics, the need to include vertical effects arises with the increasing attention to distinguish between strong cross-sectional dependence (also known as common factors) and weak cross-sectional dependence (also known as spatial dependence) (Chudik, Pesaran, and Tosetti, 2011). This idea of distinguishing between different types of spatial interactions arises from the reasoning that spatial zones may correlate as a result of shared factors (for example, belonging to the same country) or as a result of local interactions that generate spillovers.

In this sense, some works have tried to incorporate both types of interactions, Kuersteiner and Prucha (2018) or Bailey, Holly, and Pesaran (2016), taking into account the national averages, have tried incorporate vertical effects in 2-stages models. This strategy is similar to that used by GVAR models. [1] One of the most recent models developed in this line is the spatial econometrics model that accounts simultaneously for serial dynamics, spatial dependence and common factors developed by

[1] See Elhorst, Gross, and Tereanu (2018) to know where spatial econometrics and Global VARs models meet.

Vega and Elhorst (2016).

This model, and its subsequent developments (Ciccarelli and Elhorst, 2018; Elhorst, Madre, and Pirotte, 2020), incorporate common factors that allow controlling for all those factors that geographical areas have in common. The model allows estimating a response parameter to changes that occur at the highest level of aggregation, for example, in a regional model, taking into account changes in national conditions.

From the field of hierarchical or multilevel models, the strategy to account for both types of interactions is carried out trying to incorporate into traditional models the parameters used in the field of spatial econometrics (Fingleton, 2001; Corrado and Fingleton, 2012; Lacombe and McIntyre, 2017).

One of the most recent proposals is that of Dong and Harris (2015), which proposes the extension of a hierarchical model to incorporate the traditional autoregressive spatial model (SAR). This model allows estimating a hierarchical model takes into account the spatial interactions that occur at both the lower and upper levels.

In this field of hierarchical models of spatial econometrics there is still much to do. Following Lacombe and McIntyre (2017) "the existing suite of hierarchical spatial econometric models currently available is quite small" with the need for the development of improvements in model comparison, research on heteroskedasticity and the development of new models.

In the field of spatial econometric models, there is also a need for the development and analysis of new models and empirical applications that allow us to know the real meaning of the results found, and the development of new models that take into account interactions at the higher level, since recent models still treat the upper level as an isolated unit, when, following the reasoning of the hierarchical models of spatial econometrics, the interactions can occur in both lower and upper levels.

1.2 Contribution of this thesis

This thesis contributes to the mentioned literatures through empirical applications, modification of existing models, and analysis and development of new models.

First, in terms of empirical applications, this work has allowed these models to be used in fields of regional analysis to find relevant results in different areas of study such as the field of regional cyclical analysis, cigarette consumption, or self-employment analysis.

Second, this work has shown the usefulness of some of the model proposed in a different way from the original point of view. Specifically, using the traditional common factors way of modelling, it has allowed us to re-estimate the price elasticity of cigarette consumption, allowing the price to be modelled as a common factor.

Third, by applying these techniques and comparing them with other measures of spatial interactions, this work analyzes the meaning of the results found in these models and their relationship with other analysis techniques. Furthermore, it proposes the estimation of a new hierarchical spatial econometrics model and its application to analyze the usefulness of the results founds.

In summary, through an exhaustive analysis of the most recent models of spatial econometrics that include vertical relations and the inclusion of hierarchical models that include typical parameters of spatial econometrics (horizontal interactions), this thesis contributes to the existing literature trying to advance in the knowledge and development of models that take into account the nested structure of the data and the spillovers that occur between different geographic areas, that is, that take into account the vertical and horizontal interactions that the natural structure of economic data presents.

1.3 Chapter overview

This work includes the contributions mentioned above in 5 different chapters. Four chapters are devoted to the literature of spatial econometric models that include common factors, while the last chapter is applied to the literature of hierarchical models of spatial econometric.

Chapter 2 analyze the regional sensitivity of the Spanish regions to the aggregated fluctuations of the country through unemployment rates. Unemployment has been routinely used as a measure of the economic cycle. In addition, regional unemployment rates are characterized by, among other factors, their relation to the national unemployment rate. In this regard, the literature on regional sensitivity to the economic cycle has analyzed how fluctuations in the national unemployment rate affect the regions. In recent years, due to the great impact of past crises, the development of new econometric techniques such as the mentioned before and the possible arrival of new crises, the debate on how sensitive regions are to the economic cycle has reopened. In Spain, this debate is necessary since unemployment rates are very high and display a great deal of heterogeneity. We analyzed regional unemployment rates in Spain between 1978 and 2018 through the recently developed dynamic spatial econometric model with common factors and found that some regions are more sensitive than others to the economic cycle. The results seem to show that in Spain, the sensitivity to the economic cycle displays a geographical pattern where the most sensitive regions are those located on the Mediterranean coast. Specifically, we find that the sensitivity to the economic cycle of unemployment is not determined by the fact that regions have high or low unemployment; it seems that geographical location plays an important role. These results can be useful for the national and regional governments when they implement countercyclical policies. This chapter also, serve as start point for this thesis since it is an interesting application of the model that help to understand its utility.

Chapter 3, focuses on the application of this model to self-employment, a geographical phenomenon influenced by national and regional contexts. However, the study of both contexts combined is scarce in the literature on the formation of regional self-employment clusters. Using panel data from the United States for 1998-2018, we perform different techniques to study both contexts combined, including exploratory spatial data analysis, dynamic spatial estimations and machine learning algorithms. We find evidence of spatial dependency of self-employment rates across the country, although it has decreased over time. Results also suggest that most of the spatial dependency is explained by the clusters of regions with low entrepreneurship activity, and that clusters formed by highly entrepreneurial regions are the most sensitive to fluctuations in the national self-employment rate. This chapter allow us to analyze the relation between traditional measures of spatial interactions such as the hot spot and cold spot analysis with new results provided by these recent models such the sensibility to the national context. Specifically, we find evidence that those regions that represent hot spots are also regions sensitive to the national context while regions representing cold spots are not sensitive to the national context.

Chapter 4 re-estimates the price elasticity of cigarettes in Spain by using the way of introducing common factors to the models proposed in the literature of spatial econometrics, to introduce the price as a common factor since it is stablished by the government and is the same to all of the Spanish provinces. There is an agreement in the literature that tobacco price elasticity is around -0.4 for given location. However, works only focus separately, on the temporal dimension or the spatial dimension, and don't allow to estimate heterogeneous regional price-elasticities. Our work estimates a dynamic spatial econometric model with the price as a common factor to analyze the demand of cigarettes allowing us to estimate short-run, long-run, direct, indirect and total provincial price-elasticities. Results reveals that the consumption of the regions is influenced by the consumption of the neighboring regions in the same period. The price elasticity of cigarettes in the long term exceeds in many cases, in absolute value, unity. This result is novel because tobacco has historically been treated as an inelastic demand good. Finally, we found that the regions that are most sensitive to price are those bordering France and Gibraltar or tourist regions, demonstrating the effect that smuggling has on the behavior of the regions.

Chapter 5 is a short paper that provide an empirical analysis to locate the regions that have distortions in per capita tobacco consumption. The location of these regions and their proximity to other countries allow to detect the need that governments have to harmonize policies. By using panel data from the 47 Spanish provinces from 2002 to 2017 we implement a hot spot and cold spot analysis which allow us to detect areas where low or high per capita tobacco consumption clusters are generated. The results show that areas of Spain bordering countries with high price differentials, such as Gibraltar and France, generate clusters of low and high per capita tobacco consumption, respectively. Indeed, maintaining a low-price differential seems not to generate distortions, as revealed by the Portugal case. This paper could be interpreted as a logical extension of Chapter 4 since findings in provincial price-elasticities could be related to hot or cold spots of cigarette consumption.

Finally, chapter 6 brings to this thesis a new point of view from the perspective of the hierarchical models literature. Specifically, it develops the traditional spatial Durbin model in a hierarchical context of nested data to apply it in the analysis of regional productivity in the European Union. The work intends to continue with the development of applied models of hierarchical spatial econometrics showing its empirical utility for the analysis of European regions (NUTS 2) nested in 28 countries. This model allows us to estimate the random effect that the countries have on the regions, which is interesting in the case study. The results show…

1.4 Conclusions of this thesis

Each of the chapters makes different contributions to expand the literature that tries to unify the vertical and horizontal interactions that the economic data present.

Specifically, Chapter 2 presents one of the latest models developed in the field of spatial econometrics applied to the case of Spain where find which regions are most sensitive to economic fluctuations can prevent the impact of future crises. The work find that regions most sensitive to aggregated fluctuations are those located in the Mediterranean coast. This result shows the usefulness of taking into account the national context when conducting regional analyzes due to the heterogeneity that regions present.

Chapter 3 analyzes the spatial dynamics of self-employment in the United States, looking for similarities between different measures that take into account the role of the states in the national context. From the applied point of view, it seems to show that the importance in the national context of the “hot spots” states has decreased over time while the importance of the “cold spots” has increased. In other words, the states with the highest self-employment rates have lost importance in influencing global spatial dependence in the United States. States with lower self-employment rates appear to be getting influence at the national level. Furthermore, there seems to be a relationship between this dynamic with the fact that "hot spots" are the most sensitive states to national fluctuations. These results are interesting from the applied point of view, but they also show a relationship between different techniques for measuring spatial interactions that will need to be analyzed in greater depth.

Chapter 4 uses the concept of common factor to create a model that allows modelling as a common factor, explanatory variables, that, by their nature needs to be introduced into the model as a national variable. Specifically, in the case of the cigarette demand function in the Spanish provinces, this chapter finds that this modeling allows estimating the provincial price elasticity of cigarettes, finding significant differences between provinces. Furthermore, it seems to find that the traditional value of the elasticity of cigarettes assumed by the literature may not be accurate if the long term, the influences of neighboring provinces and regional differences are taken into account.

Chapter 5, as a continuation of the findings found in Chapter 5, focuses on analyzing the provincial differences in the distribution of cigarette consumption in Spain. This chapter finds that there are significant differences, finding clusters of high and low cigarette consumption. The location of these clusters matches with the location of the most price-sensitive provinces found in chapter 4, which are those close to countries with a high price differential.

Finally, Chapter 6 approaches the field of hierarchical models to propose a traditional model of spatial econometrics in a hierarchical context.

1.5 Publications

As a result of this dissertation, the following works have been developed:

- **Chapter 2**: Almeida, A., Galiano, A., Golpe, A. A., Martín, J. M. (2020). Regional unemployment and cyclical sensitivity in Spain. Letters in Spatial and Resource Sciences, 1-13. https://doi.org/10.1007/s12076-020-00252-3
- **Chapter 3**: Almeida, A., Golpe, A. A., Justo, R. From hot to cold: A spatial analysis of self-employment in the United States. (Under review in Papers in Regional Science).
- **Chapter 4**: Almeida A., Golpe A. A., Iglesias J., Martín J.M. (2020) The price elasticity of cigarettes: new evidence from Spanish regions, 2002-2016, Nicotine Tobacco Research, , ntaa131, https://doi.org/10.1093/ntr/ntaa131
- **Chapter 5**: Almeida A., Golpe A. A., Martín J.M. A spatial analysis of the Spanish tobacco consumption distribution: Is there any consumption clusters? (2020) Public Health (Forthcoming)
- **Chapter 6**: Almeida A., Ramajo J., A hierarquical spatial Durbin model (HSDM): An application to regional production efficiency in Europe. (Working Paper)

Chapter 2

Regional unemployment and cyclical sensitivity in Spain

2.1 Introduction

Many studies since the 1960s have researched the topic of the cyclical sensitivity of the unemployment rate. In the beginning, this literature paid attention to the common component, which is dominant in explaining movements in regional unemployment rates (Martin, 1997). The idea to link the regional to the national unemployment rate and to estimate this relationship for each individual region dates back to Thirlwall (1966) and Brechling (1967), in what is known as the regional cyclical sensitivity literature. Recently, Vega and Elhorst (2016) bring the cyclical sensitivity literature back to the analysis of regional disparities by considering in their methodology serial dynamics, spatial dependence and common factors. Traditionally, dynamic panel data models only accounts for spatial dependence (also called weak spatial dependence) which is an observed correlation across space because of local interactions between regions generating spillover effects. This new model also allows to account for common factors (also known as strong spatial dependence) that is an observed correlation across space as a result of shared factors such a aggregated economic fluctuations, where outcomes change together as these factors change. This modeling opens a new line of interest for analyzing unemployment among regions considering the presence of common factor and ensuring, that way, unbiased results.

Briefly, the literature on regional unemployment disparities identifies four stylized facts that defines regional unemployment disparities: (1) regional unemployment rates are strongly correlated over time, (2) regional unemployment rate behaves in parallel to the national unemployment rate, (3) are correlated across space and (4) display heterogeneity among regions.

Focusing on Spain, there are large differences in the unemployment rate among regions (Bande, Fernández, and Montuenga, 2008; Cuéllar-Martín, Martín-Román, and Moral, 2019) becoming more persistent over time, and bigger after the last economic crisis (Jimeno and Bentolila, 1998; Albulescu and Tiwari, 2018). So that, it is necessary to pay attention to these differences and address them in terms of cyclical sensitivity. Thus, analyzing regional unemployment rate disparities, the dependence between different regions and the relation to the national rate simultaneously is a requirement to implement appropriate policies. The last economic recession in Spain after 2009, reinforced this need, so that many studies had focused on analyzing the regional disparity in unemployment rates since that economic crisis. Nowadays, this necessity reappears due to the dramatic effects produced in the Spanish economy and its labor market because of the COVID-19 pandemic (McKibbin and Fernando, 2020).

In this line, one of the articles that has focused on the analysis of the reactions of regional unemployment to changes in the economic cycle for Spain is Bande, Fernández, and Montuenga (2008). This paper concludes that there is a positive relationship between the regional dispersion of unemployment and the economic cycle, which guarantees that fluctuations in the economic situation of the country directly affect to regional unemployment. Since that study, many papers have focused on analyzing the reason for these disparities, assuming that unemployment is sensitive to the economic cycle. For example, a group of papers make the analysis of regional cyclical sensibility based on dividing regions with high and low unemployment (Bande and Karanassou, 2009; Sala and Trivín, 2014) . Other block of researches focuses on doing spatial analyses of unemployment disparities to find clusters of similar behaviors (Cuéllar-Martín, Martín-Román, and Moral, 2019). Finally, some papers have focused on looking for regional characteristics that motivate regional differences (López-Bazo and Motellón, 2013; Melguizo, 2017).

However, none of the articles have analyzed trends of regional unemployment while simultaneously considering that regional unemployment rates are persistent, heterogeneous, parallel to the national rate and spatially dependent. According to Vega and Elhorst (2016), isolated analysis can potentially lead to biased results, since series dynamics, spatial dependence and common factors are more likely to be interdependent. To the best of our knowledge, this paper is the first to simultaneously analyze, for the Spanish case, the persistence, heterogeneity, spatial dependence and heterogeneity in economic cycle sensitivity of regional unemployment.

The main objective of this work is to provide new evidence about how Spanish regions react to economic fluctuations as a requirement to implement appropriate policies. The economic and labor implication due to public and private containment measures against the COVID-19 pandemic, such as school, shops and factory closures, travel restrictions and quarantines, with the corresponding cut in domestic demand, (Baldwin and Tomiura, 2020) make this knowledge to be crucial. As a result, heterogeneous behaviors will reveal the necessity to introduce regional perspectives against future economic fluctuations.

Our paper contribute to the literature as follow. As the main contribution, we use Vega and Elhorst (2016) methodology to simultaneously consider all components of the stylized fact that defines the disparities in regional unemployment rates. Another contribution is that we show evidence that sensitivity has a geographic pattern since we find that the regions located on the Mediterranean coast are the most sensitive and that as we go inland and northwest, this sensitivity decreases. This pattern reveals, moreover, that regions with higher unemployment rates are not the most sensitive. Our results are in line with the new side of the literature indicating that the regional-specific unemployment rate is not important at all in the unemployment trend but that geographical factors do matter (Cuéllar-Martín, Martín-Román, and Moral, 2019). In this vein, Camacho, Pacce, and Ariza (2018) also find a geographical pattern in the propagation of economic crisis and the way the unemployment rate reacts to recessions.

2.2 Data and Method

2.2.1 Data

The data used in this work, extracted from Fuente (2019)[1], are the annual unemployment rates for the 15 autonomous communities of Spain from 1978 to 2018 and the annual Spanish unemployment rate, treated as a common factor following Bailey, Holly, and Pesaran (2016).[2]

Regional unemployment rates tend to have some specific characteristics. In particular, many works have found that regional unemployment is characterized by being correlated in time, in space and with the national rate.

Figure 2.1 shows the evolution over time of regional unemployment rates for the 15 autonomous communities of Spain analyzed together with the national rate. This figure shows how, in general, the regional and national unemployment rates have a similar trend throughout the analyzed period. Some special situations can be observed where, despite behaving similarly to the national rate, the regional rate is always higher than the national unemployment rate, as in the case in Andalucia or Extremadura. On the other hand, La Rioja and Navarra always seem to stay below the national rate.

In addition, this unemployment rate is relatively stable over time, except in different periods of economic growth and downturns. Specifically, three major economic crises can be distinguished in Spain (Cancelo, 2004; Gadea, Gómez-Loscos, and Montañés, 2012; Camacho, Pacce, and Ariza, 2018): the first in 1983-1985, the second in 1991-1994 and the third and most recent in 2008-2014. Several conclusions can be drawn by observing figure 2.1. First, all crises seem to have a strong national component, since regional rates fluctuate similarly to the national rate. Second, there is clear heterogeneity in the trends of regions. For example, during the pronounced crisis of 2008-2014, the highest rate of the regions in Spain was in Andalucia at 35.65%, and the lowest was in Pais Vasco at 9.79%.

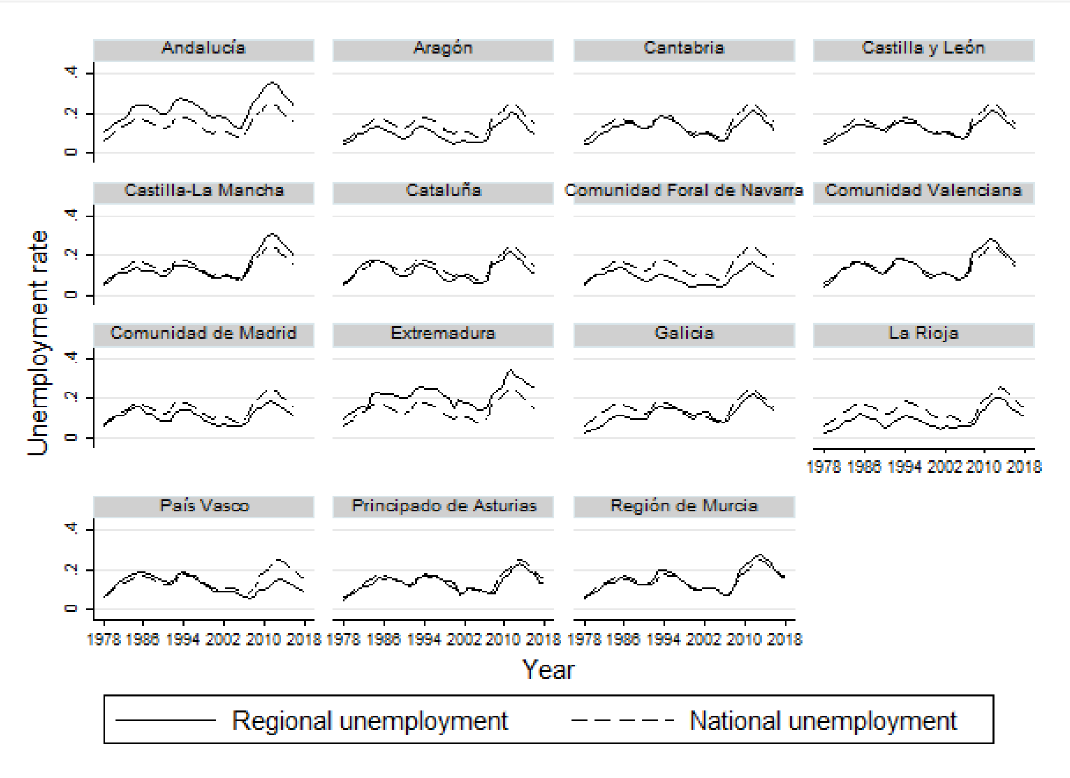
To analyze the spatial distribution of unemployment rates and to determine the possible correlation in space, figure 2.2 maps unemployment rates over the period analyzed every eight years (1978, 1986, 1994, 2002, 2010 and 2018). A clear north/south contrast that has been accentuated in recent years, in addition, shows a pattern of spatial correlation, where neighboring regions have similar rates. To test that space matters in our case study, we have estimated global spatial autocorrelation through the Moran's I statistic from 1978 to 2018. The results of these tests

[1]We replicate the results by using the data set most used by specialists in the Spanish labor market (Spanish Statistical Office). However, in 2002, the Spanish Statistical Office survey changed its methodology to measure unemployment rate which have consequences on the estimates. Nevertheless results are similar and are available upon request.

[2]The Balearic Islands, Canary Islands and Ceuta and Melilla have been extracted from the sample since these areas are treated as islands, as is usual in the spatial econometric literature to avoid complete zero problems in the contiguity matrix (we adopt this matrix since it outperforms other specifications to which we come back later).

show that, for most of the years, the null hypothesis (absence of global spatial autocorrelation) is rejected.[3][4]

FIGURE 2.1: Regional and national unemployment rate.



2.2.2 Method

Our methodological strategy involves the application of the recently developed model by Vega and Elhorst (2016), which simultaneously accounts for serial dynamics, spatial dependence and common factors; their study also shows how not simultaneously including these effects for the regional unemployment rate can produce biased results.

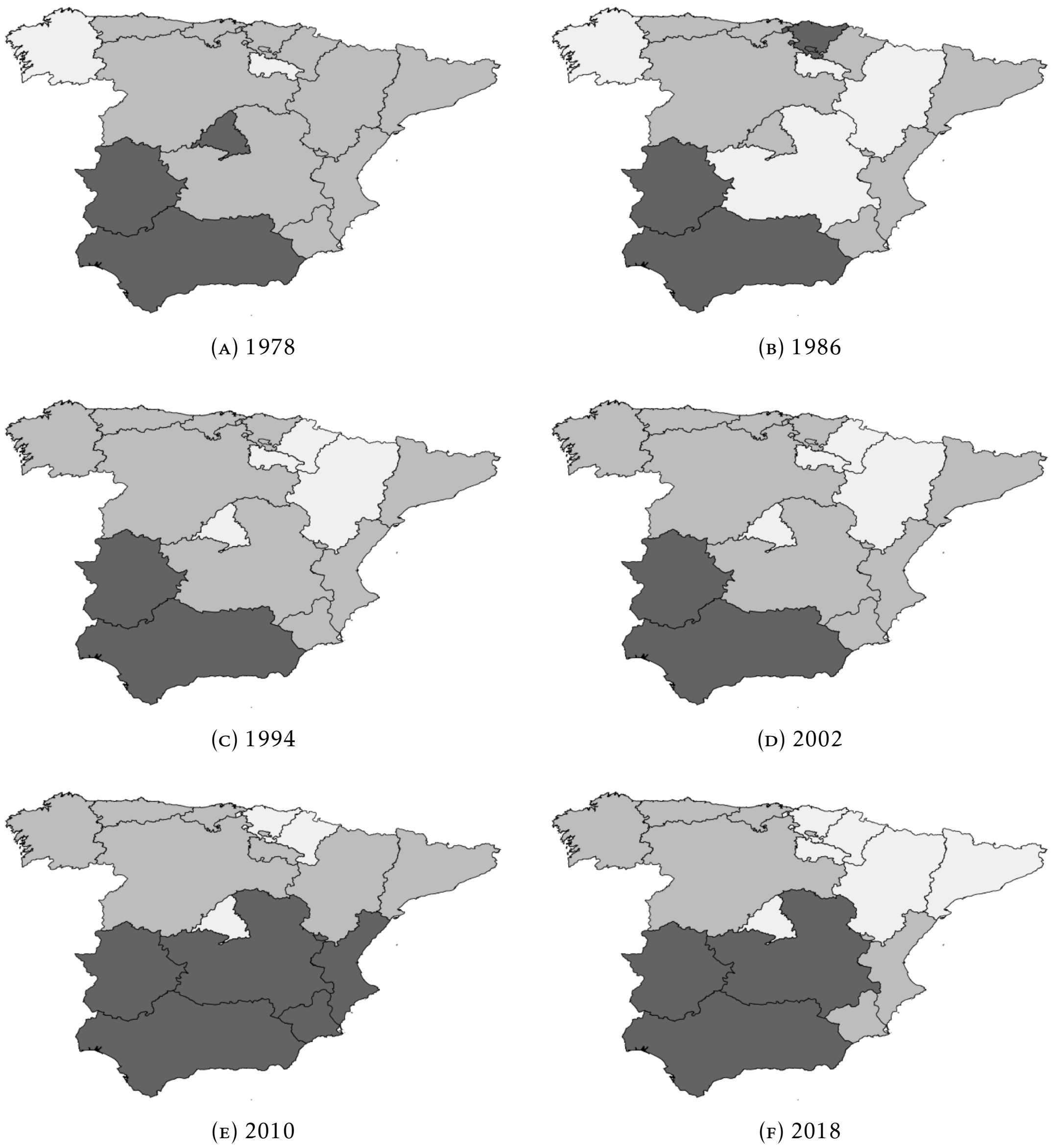
We use the CD test (Pesaran, 2004) in its local version (Moscone and Tosetti, 2009, eq.22) to test for the presence of cross-sectional dependence in our panel data. If the null hypothesis is rejected, we can corroborate the existence of cross-sectional dependence. This test is carried out by specifying the relationship matrix of the 15 Spanish regions (W). The result of this test applied to the data shows the presence of cross-sectional dependence in regional unemployment rates (Z=29.987 with p-value=0.000). This outcome is highly statistically significant, indicating that cross-sectional dependence needs to be accounted for.

For this purpose, we have used a row-normalized binary contiguity matrix, which is an NxN matrix describing the arrangement of the regions in space, with 1 if two

[3]Only for years 1979-1985 and 2009-2016 the null hypothesis cannot be rejected. It is interesting that these years match approximately with two periods of the Spanish economics crisis which have been detected by several authors.

[4]A panel data unit root test have been also estimated finding that regional unemployment rate are stationary.

FIGURE 2.2: Spatial distribution of unemployment over time. Dark colors for high unemployment rates.



(A) 1978

(B) 1986

(C) 1994

(D) 2002

(E) 2010

(F) 2018

regions are neighbors and 0 if not. We use this alternative based on the empirical results explained in Vega and Elhorst (2014) and Elhorst (2017) where they find that this specification outperforms the other alternatives. We tried first and second order binary contiguity matrix. However, second order binary contiguity seems not to reflect well the relations between the Spanish regions since having 15 regions the average number of links in each region is 8.667. This result implies that, for example, Andalusia, a region located in the far south of Spain, would have spillovers with Aragon located in the far north of Spain.

On the other hand, to determine if the nature of this spatial dependence is weak or strong (in other words, if it is due or not to the presence of common factors), we apply the α-test exponent of Bailey, Holly, and Pesaran (2016)[5]. This test can take values between 0 and 1, where values below 0.5 indicate the presence of weak spatial dependence and values equal to 1 indicate the presence of strong spatial dependence. The result of this test applied to the data gives α=1.003[6] and std.err.=0.034 which points to the presence of strong spatial dependence, common factors needs to be accounted for.

Our target model is the one proposed by Vega and Elhorst (2016) that reads as follows:

$$U_t = \tau U_{t-1} + \delta W U_t + \eta W U_{t-1} + \Gamma_1 U_t^n + \Gamma_2 U_{t-1}^n + \mu + \epsilon_t \tag{2.1}$$

where U_t is a column vector with one observation of the dependent variable (unemployment) for every unit (i) at every point at time (t). U_{t-1}, WU_t and WU_{t-1} are vectors of temporal, spatial and spatiotemporal lags, respectively, with τ, δ and η autoregressive coefficients. W is the row-normalized binary contiguity matrix. U_t^n and U_{t-1}^n are the unemployment rates of the whole country at times t and $t-1$, and Γ_1 and Γ_2 column vectors with unit-specified coefficients of response to the common factors. μ represent the spatial fixed effect added to the model and ϵ_t is the Nx1 vector independently and identically distributed error term with zero mean and constant variance σ^2. The parameter of the region's sensitivity to the economic cycle (γ) can be estimated by dividing the elements of Γ_1 by $1-\delta$ or by dividing the elements of Γ_2 by $-\tau-\eta$.

This model allows us to simultaneously measure the four remarkable stylized facts that often arise from analyzing the evolution of regional unemployment rates. First, the presence of time correlation of the regional unemployment rates by incorporating the regional unemployment rate lagged in time as well as in time and space and the common factor (the national unemployment rate) lagged in time. Second, the model allows us to account for the presence of spatial dependence by adding the spatial lag and the spatiotemporal lag. Third this model includes the common factor and its lag in time, which allows for the estimation of an individual sensitivity parameter for each region and fourth we include spatial fixed effects which allow us to account for spatial heterogeneity.

[5]A xtcse2 stata routine have been used (Ditzen, 2019)

[6]Values above the upper bound of the interval (0,1] may occur when not all the asymptotic properties are fully met (Bailey, Holly, and Pesaran, 2016). However, since the hypothesis $\alpha = 1$ cannot be rejected we can conclude that α estimated lies within the interval. Similar results can be also found in Vega and Elhorst (2016).

2.3 Results

We begin by estimating the basic dynamic spatial panel data model with regional fixed effects and without including common factors (model A), and we apply both tests (CD local test and α test) to the residuals of the model. These results are shown in table 2.1, column A, where it can be seen that the residuals of the model continue to point to the presence of strong spatial dependence, with α=0.920 and std.err.=0.048. However, the local CD test points to the fact that weak spatial dependence is no longer present (Z=-0.838 and p-value=0.402), because the null hypothesis of spatial independence between neighboring regions cannot be rejected. Results shows a highly and very significant temporal, spatial and spatiotemporal lag (τ=0.891, δ=0.825 and η=-0.739).

Next, we estimate the model proposed by Vega and Elhorst (2016) incorporating common factors (model B)[7]. Through this model, it seems that both the local spatial dependence (Z=1.196, p-value=0.232) and the presence of common factors (α=0.255, std.err.=0.055) appear to have been effectively covered.[8] As seen in column B, the serially lagged unemployment rate (τ=0.929) is highly significant, reflecting the strong correlation of unemployment rates over time. In addition, the spatially lagged coefficient ($\delta = 0.165$) is positive and significant, reflecting the presence of spatial dependence between regions, and the lagged spatial autoregressive coefficient seems to be significant and negative.

Model B seems to have perfectly covered the presence of spatial dependence and common factors based on the results provided by the test. By introducing common factors, the model correctly cover the difference between spatial dependence and common factors. The δ and η could have been overestimated in model A due to the absence of common factors.[9] Another way to account for the presence of common factors would have been to add a time-period fixed effects to the model too. However, this would be similar to the inclusion of common factors at time t with the unit-specific coefficients replaced by a time dummy with a common coefficient.[10]

In table 2.2 , we show the result of the coefficients of response to the national unemployment rate estimated for each of the regions (Γ). Both coefficients (Γ_1 in t and Γ_2 in t-1) are highly significant. The estimation of the economic cycle sensitivity parameters is shown in the last two columns ($\gamma_1, \gamma 2$). [11]

[7] We have also estimated the two-stage approach developed by Bailey, Holly, and Pesaran (2016). An LR test comparing both models shows that the simultaneous model fits the data better and the Brechling-Thirwall type of cyclical sensitivity estimated seems to show long-lasting problems (Brechling, 1967; Domazlicky, 1980) where regions with estimates greater than one are those with unemployment rates persistently higher that national average.

[8] The model also seems to be stationary and stable since $\tau + \delta + \eta < 1$

[9] Note that model A is measuring both types of cross-sectional dependence (spatial dependence and common factors) through δ and η parameters. When we appropriately include common factors to the model (model B), these parameters reduce their magnitude dramatically as expected.

[10] Results of this model (with log-likelihood function value of -753.298) are quite similar with parameter $\tau = 0.896$ (std.err.=0.018), $\delta = 0.230$ (std.err.=0.046) and $\eta = -0.179$ (std.err.=0.051)

[11] We calculate the parameters according to the procedure proposed in Vega and Elhorst (2016). The standard errors are calculated using formulas for the sum and quotient of random variables (Mood, Graybill, and Boes, 1974, pg. 178-181).

TABLE 2.1: Dynamic Spatial Panel Data Models

Models		
	(A)	(B)
τ	0.891 (0.019)	0.929 (0.021)
δ	0.825 (0.019)	0.165(0.062)
η	-0.739 (0.026)	-0.131(0.054)
Time-period fixed effects	No	No
Regional fixed effects	Yes	Yes
Common Factors	No	Yes
$CorrR^2$	0.834	0.970
Log-Likelihood	-873.419	-714.377
CD_{local} [p-value]	-0.838 [0.402]	1.196 [0.232]
α test (standard error)	0.920 (0.048)	0.255 (0.055)

Note: Standard errors are reported in parentheses.

TABLE 2.2: Common Factors

Regions	Γ_1	Γ_2	γ_1=$\Gamma_1/(1-\delta)$	γ_2=$\Gamma_2/(-\tau-\eta)$
Andalucia	1.051 (0.090)	-0.960 (0.104)	1.258 (0.084)	1.205 (0.085)
Aragon	0.737 (0.084)	-0.717 (0.088)	0.883 (0.084)	0.900 (0.084)
Cantabria	0.778 (0.084)	-0.680 (0.091)	0.932 (0.085)	0.853 (0.085)
Castilla-La Mancha	0.916 (0.086)	-0.834 (0.095)	1.097 (0.084)	1.047 (0.084)
Castilla y Leon	0.655 (0.084)	-0.610 (0.091)	0.785 (0.084)	0.766 (0.085)
Cataluña	0.880 (0.087)	-0.887 (0.090)	1.054 (0.084)	1.114 (0.085)
Madrid	0.640 (0.084)	-0.623 (0.090)	0.767 (0.084)	0.782 (0.086)
Navarra	0.550 (0.080)	-0.556 (0.083)	0.659 (0.084)	0.698 (0.084)
Comunidad Valenciana	0.995 (0.087)	-0.993 (0.092)	1.192 (0.085)	1.246 (0.086)
Extremadura	0.928 (0.091)	-0.811 (0.102)	1.112 (0.085)	1.018 (0.085)
Galicia	0.630 (0.085)	-0.550 (0.095)	0.756 (0.085)	0.690 (0.086)
La Rioja	0.661 (0.079)	-0.654 (0.083)	0.792 (0.084)	0.820 (0.084)
Pais Vasco	0.627 (0.082)	-0.615 (0.085)	0.752 (0.084)	0.772 (0.084)
Asturias	0.743 (0.084)	-0.666 (0.093)	0.890 (0.086)	0.836 (0.085)
Region de Murcia	0.900 (0.090)	-0.891 (0.098)	1.079 (0.084)	1.118 (0.084)

Note: Standard errors are reported in parentheses.

The regions most sensitive to the economic cycle appear to be Comunidad Valenciana, Andalucia, Region de Murcia and Cataluña, each with parameters greater than 1.100. It seems that regions located on the Mediterranean coast share a common pattern. These communities are characterized by being the most touristy in Spain, which may be a possible explanation. Among the least sensitive communities, we find Galicia, Navarra and Castilla y Leon, each with parameters below 0.700.

To show the spatial distribution of the sensitivity of the regions, figure 2.3 shows in four shades of gray the intensity of the economic cycle sensitivity of each region (γ_2).[12] It seems that Spain has a specific sensitivity pattern where the most sensitive regions are those located on the Mediterranean coast and sensitivity decreases as we move inland to the northwest. Particularly, the economic cycle sensitivity of the regions in Spain can be divided into four groups, from the most sensitive (in darker colors) to the least sensitive (in lighter colors). The first includes the four regions most sensitive to the economic cycle of Spain, which are also those located on the Mediterranean coast (Andalucia, Region de Murcia, Comunidad Valenciana and Barcelona). The second group, formed by Extremadura and Castilla-La Mancha, are also regions that are sensitive to the economic cycle, although less so than the previous group; this second group is composed of regions located in the southern interior and neighboring the most sensitive regions. The third group is formed by regions that are not sensitive to the economic cycle and that are located in the northern interior of Spain, away from the most sensitive regions. Finally, the fourth group is formed by the least sensitive regions of Spain, Galicia and Navarra.

Figure 2.3: Sensitivity to economic cycle.

These results seem to point to the fact that when crises and recoveries appear, the regions located on the Mediterranean coast are the ones most affected because they are the most sensitive.

2.4 Conclusion

In recent years, there has been a growing interest among academics, practitioners and policy makers in finding a pattern in Spanish regional unemployment trends. This interest emerges from the need to know how unemployment in each region will react ahead of the increasingly common occurrence of recession phases. In this context, the literature documents that unemployment is persistent, heterogeneous, spatially dependent and parallel to the national rate, but these effects have never

[12] We map parameter γ_2 since it is based on the relative strength of both internal and external habit of persistence (Korniotis, 2010).

been analyzed simultaneously.

In our paper, by applying the methodology proposed by Vega and Elhorst (2016), the persistence, heterogeneity, spatial dependence and economic cycle sensitivity of Spanish regional unemployment are analyzed simultaneously. Furthermore, this methodology allows us to estimate heterogeneous coefficients of response to aggregate fluctuations. The findings show that there is a high persistence of unemployment in Spain, that there is spatial dependence and that regions show different sensitivities to the economic cycle. Although the first findings are in line with the previous literature, the main contribution of this article is to uncover the geographical pattern of sensitivity to the economic cycle. We find that the economic cycle sensitivity of unemployment is not determined by the fact that regions have high or low unemployment; it seems that geographical location plays an important role. Specifically, the regions that have greater economic cycle sensitivity are located on the Mediterranean coast and include regions with very different unemployment rates. However, what these regions do have in common is that they are areas very focused on the tourism sector, which can be taken as a possible explanation. The findings are related to Melguizo (2017) who shows that regions with a more developed service sector suffer more variations in unemployment rates, and to Camacho, Pacce, and Ariza (2018) who find a similar pattern in how crises and recoveries begin and end in Spain.

Both findings are in line with the regional heterogeneity we observe, since they reveal that some regions suffer more from the effects of business cycle whereas other regions are less affected by economic contingencies. Consequently, the application of national inflexible policies may difficult actions devoted to smooth cyclical swing of regional economic activity. Therefore, that, as main find, we consider that this regional behaviour is indicating the necessity to apply differentiated employment policies when national economy face a crisis. Nowadays, this find become into a crucial due to the employment destruction our regions are suffering because of the economic and labour impact caused by the restrictive measures against the COVID-19 pandemic.

On the one hand, with the pattern in the trend of regional economic cycle sensitivity that this paper finds, the national government can distribute resources in an efficient way to react to future recessions or to recoveries from past crises. On the other hand, regional policies can be adapted to each geographical area depending on its sensitivity, and regional cooperation is needed since spatial dependence points to the presence of possible spillovers.

Chapter 3

From hot to cold: A spatial analysis of self-employment in the United States

3.1 Introduction

Self-employment is an important booster of economic growth and regional prosperity through its positive effects on productivity, innovation, employment and competition (Acs and Armington, 2006; Braunerhjelm et al., 2010; Carree and Thurik, 2003; Carree and Thurik, 2008; Mittelstädt and Cerri, 2008) . Several empirical studies show that self-employment rates differ between and within countries (Cheng and Li, 2011; Kangasharju, 2000). This variation has been studied by considering only national contexts (Edquist and Johnson, 1996; Porter and Stern, 2001; Wennekers et al., 2005) or, more recently, through specific characteristics of regional contexts (Cravo, Becker, and Gourlay, 2015; Hong et al., 2015; Luo and Chong, 2019). Many studies have focused, among these regional characteristics, on regional entrepreneurship culture due to its ability to cultivate the roots of the business tradition and the growth and consequent persistence over time of the self-employment rates (Andersson and Koster, 2010; Fotopoulos and Storey, 2017; Fritsch and Kublina, 2019).

In this context, investigating whether regional clusters exist would confirm if entrepreneurship culture spreads between neighboring regions. In addition, we would suggest that if regional self-employment rates are persistent over time, the entrepreneurship culture is strong and will continue to promote growing self-employment rates. On the other hand, completing this analysis by investigating whether regional cluster formation is also motivated by national factors could be highly relevant for policy makers. This is an issue not considered so far in the literature on self-employment, and we try to shed light on it with our study. Using data from the Regional Information System of the Bureau of Economic Analysis, we apply spatial econometric techniques to analyze the spatial distribution of self-employment rates across US regions, the formation of regional self-employment clusters and the persistence of such clusters over time. Additionally, to consider the national context, we complete this analysis by considering how sensitive each state is to the general evolution of self-employment rates in the country. Finally, we take all the information together to develop a cluster analysis by means of one of the most frequently used unsupervised machine learning algorithms.

A large body of the literature confirms that self-employment is a geographical phenomenon marked by the national context and, especially, by the regional context. Focusing on the national context, some studies examine the macroeconomic factors that influence self-employment rates. These include (i) the evolution of GDP (Bjørnskov and Foss, 2008; Carmona et al., 2016; Klapper et al., 2007), (ii) unemployment levels (Meager, 1992; Parker and Robson, 2004; Reynolds, Miller, and Maki, 1995), (iii) skills at the aggregate level, such as the level of education in a country (Thai and Turkina, 2014), (iv) the institutional framework (Chemin, 2009), (v) topics related to legal regulation, and (vi) the administrative environment (Porter and Stern, 2001), among others. These factors as a whole can be defined as the national self-employment context, as they are vital elements in the business activity of a national economy. In addition to these national-level patterns, a growing body of the literature has focused on regional-level characteristics as the fundamental basis for understanding the development of the business fabric of a given territory (Andersson and Koster, 2010; Fritsch and Wyrwich, 2014; Luo and Chong, 2019; Pijnenburg and Kholodilin, 2014; Fritsch and Wyrwich, 2016). Along these lines, several factors that drive differences in self-employment rates between different regions in a given country have been identified (Cheng and Li, 2011; Okamuro and Kobayashi, 2006). These differences are a reflection of the existence of specific characteristics of each region. Thus, business activity can be defined as a regional affair, that is, the characteristics of the region have a significant influence on the decision to enter self-employment (Ross, Adams, and Crossan, 2015). Some of the regional elements accepted as influential by scholars are the level of competence, the organization of the network, the creation of new firms, the presence of large and small companies, business agglomeration and government policies (Bosma and Sternberg, 2014; Fritsch and Wyrwich, 2016; Luo and Chong, 2019).

However, there is a regional feature that stands out from the previous ones because of its influence on the number of business initiatives undertaken in a region, viz., entrepreneurship culture, which refers to the "set of norms, values and codes of conduct that promote social acceptance and approval of entrepreneurial activities resulting in high self-employment rates which persist over time" (Fritsch and Wyrwich, 2016). Using this definition, several recent studies have explored the relationship between such culture and business behavior in different geographical areas (Fritsch and Wyrwich, 2016; Stuetzer et al., 2018; Fritsch and Wyrwich, 2019). Two of the essential ideas drawn from these works are that (i) high regional rates of self-employment are attributed to regional cultural particularities and (ii) the greater the number of individuals with business-culture values, the higher the self-employment rate. Thus, cultural attributes seem to occupy a prominent position among the drivers of the level and persistence of self-employment rates within a territory (Fritsch and Wyrwich, 2014).

Thus far, the decomposition of this regional self-employment context indicates that to understand self-employment, it is necessary to consider that it is a regional phenomenon marked by the regional framework (Sternberg and Wennekers, 2005). This regional context facilitates the exchange of ideas between individuals, institutions and business, encouraging the transmission of knowledge within the region (Singh, 2005). In turn, this transfer of knowledge might expand, even reaching neighboring regions. Once this occurs, an agglomeration of economic activities in the area may arise, leading to self-employment clusters (Feldman, Francis, and Bercovitz, 2005; Spencer et al., 2010; Zhu et al., 2019). These agglomerations allow companies to take advantage of externalities, including access to information, labor, suppliers, etc., which can lead to a more competitive business environment, inducing the continuity of business activity (Ross, Adams, and Crossan, 2015). In addition, entrepreneurial clusters lead to better productivity and local competitiveness indexes in the region while also favoring the creation of new companies (Armington and Acs, 2002; Huggins, 2008). As a result, the business environment of a given region will be increasingly competitive, and the regions that host such clusters will experience more pronounced economic growth (Wolman and Hincapie, 2015). These benefits can be extended throughout a country due to indirect effects (e.g., lower costs as a result of agglomeration economies, greater dissemination of knowledge, use of technology, etc.) derived from the interaction between conglomerates of nearby regions (Delgado, Porter, and Stern, 2007). Audretsch and Feldman (1996) reinforced this idea by stating that entrepreneurial activity inclines towards spatial grouping.

Furthermore, cluster theory shows that self-employment is linked to spatial dimensions as a consequence of factors that favor geographic interdependence between different locations, including incubation systems, sophisticated network and communication structures, collective learning programs, etc. (Bosma and Sternberg, 2014; Fritsch and Falck, 2007; Lado-Sestayo, Neira-Gómez, and Chasco-Yrigoyen, 2017). Consequently, the spatial dimension is increasingly a topic more used in research on regional economies. Although the previous literature was limited to studying variations between countries, the greater availability of regional variables has allowed the focus to be transferred to regional (Fritsch and Wyrwich, 2016; Stuetzer et al., 2018) and even local (Doms, Lewis, and Robb, 2010; Rupasingha and Goetz, 2013) differences and similarities. These spatial analyses complete the understanding of the elements that determine the level of regional self-employment. In the same way, they help identify the possible transmission of indirect effects between neighboring regions or, in other words, if there is spatial dependence between regions. In general terms, spatial dependence is defined as a process through which the growth of a given region is influenced by what occurs in neighboring regions (LeSage and Fischer, 2009).Audretsch and Keilbach (2007) claim that business creation processes are spatially correlated and, consequently, have an impact on adjacent regions. However, despite the relevance of spatial dependence and its evolution over time in the development of self-employment within regions (Lado-Sestayo, Neira-Gómez, and Chasco-Yrigoyen, 2017), empirical studies on this topic within the self-employment literature are still scarce.

The foregoing shows that the agglomeration of self-employment rates is determined by two types of business contexts: regional and national. First, the regional context has the capacity to influence the volume of business activity in the regions, explain the persistence of regional self-employment rates and favor the spatial agglomeration of business activities, in turn fostering the creation of entrepreneurial regional clusters. Second, the national context offers a vision of the pattern that the national self-employment rate follows at the aggregate level. Thus, at the time of writing of this paper, the literature on self-employment clusters has focused on regional or national contexts. However, the analysis of self-employment cluster formation as a combination of both types of contexts has been a scenario not considered so far. Our study tries to fill this gap by exploring to what extent the phenomenon of self-employment clusters among US regions is explained by the regional and national contexts.

Therefore, to the best of our knowledge, this is the first study that analyzes the formation of self-employment clusters taking into account regional and national influences. To do this, we need a methodology that allows us to analyze i) the creation of self-employment among US counties and its evolution over time and ii) the role played by the regional and national contexts in the counties and the analyzed clusters. To this end, by using data from the Regional Information System of the Bureau of Economic Analysis covering the period 1998 to 2018, we opted for the application of three analysis techniques. The first is called exploratory spatial data analysis (ESDA). It is focused on analyzing the formation and evolution of regional self-employment clusters, having the ability to detect whether high and low self-employment rates interact between neighboring regions. The second technique is based on the recent dynamic spatial econometric panel data model developed by Vega and Elhorst (2016). This model allows us to corroborate the existence of regional interactions—or spatial dependence—and persistence of self-employment rates, as well as estimate the parameters of sensitivity to aggregate fluctuations in the self-employment rate. As a third technique, we use one of the most commonly used unsupervised machine learning algorithms to bring together the results obtained in the previous analyses and to draw final conclusions from the results.

Our results point to the presence of spatial dependence in general terms, although it has decayed over time. Moreover, while the period 1998 to 2008 is characterized by a stronger influence of clusters formed by regions of high entrepreneurial activity surrounded by other highly entrepreneurial regions ("high-high" or HH clusters), the most recent periods have shown a change in that trend. Thus, clusters of regions with low entrepreneurial rates and surrounded by similarly low entrepreneurial regions ("low-low" or LL clusters) tend to cause the overall spatial dependence in the country. Therefore, there has been a destruction of HH clusters, which could imply that agglomeration effects among neighboring US regions are not strong enough to persist over time. Finally, in terms of which regional self-employment rates are sensitive to national self-employment fluctuations or, in other words, which regions follow the national pattern in terms of self-employment rates and are therefore influenced by the national context, the results reveal that these regions are those characterized as participants in HH clusters.

The remainder of this paper is structured as follows. The following section describes the data and the methodology employed in the analysis. Next, the results are presented. Finally, we conclude the paper and discuss the potential implications of our results.

3.2 Data and Method

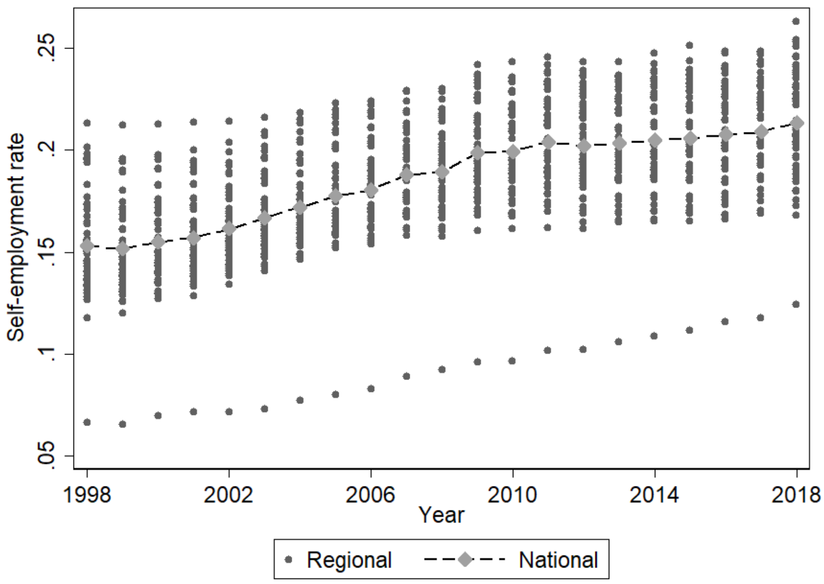
The main source of data that we employ in our analysis stems from the Regional Information System (REIS) of the Bureau of Economic Analysis. This data set provides annual information for the period 1998-2018 on total employment (TE) and its components for the USA at the state level[1]. The REIS database distinguishes between employment and proprietorship employment and has a sectoral composition of total employment. The self-employment rate (SE) is equal to the number of nonfarm proprietors as a ratio of total employment.

[1] Hawaii and Alaska have been removed from the sample as usual in the literature.

Figures 3.1. and 3.2. show the temporal and spatial distribution of self-employment rates from 1998 to 2018. In Figure 3.1., the averages of each year are linked by a line that shows an increase throughout the period analyzed. In 1998, the self-employment rate was 15.32% on average; by 2018, the rate had increased to 21.36%. However, the standard deviation has remained constant since 1998 being 2.65 in 1998 and 2.73 in 2018.

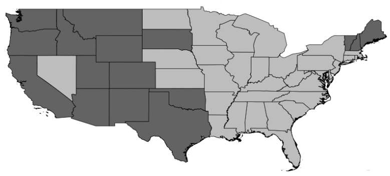
Figure 3.2. illustrates the geographical distribution of self-employment. The maps show that the territorial diffusion of self-employment rates follows a clear West-East gradient and presents some changes throughout the analyzed period. Dark colors appear, in particular, for the Rocky Mountain, Southwest and New England areas, while lighters colors appear in the Great Lakes and its surrounding areas. In 1998, Montana had the highest rate among states, 21.33%, followed by Maine (20.14%), Vermont (19.79%) and Wyoming (19.61%). The lowest rate were in the District of Columbia (6.62%), Delaware (11.64%) and South Carolina (12.65%). In 2018, Florida was the highest rate among states, 26.33%, followed by Wyoming (25.40%), Colorado (25.33%) and Texas (25.08%) while the lowest rate were in the District of Columbia (12.42) followed by West Virginia (16.81%) and Wisconsin (17.26%)

Figure 3.1: Nonfarm self-employment rate over time.

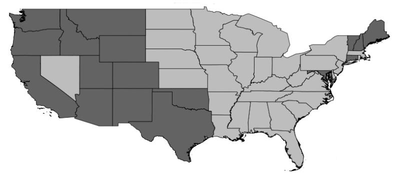


To analyze the spatiotemporal arrangement of nonfarm self-employment in the USA, we follow a sequence of different analyses. As a first step, we go through exploratory data analysis (ESDA), which is a set of techniques to describe and visualize the spatial distribution of the phenomenon to find possible spillovers and clusters between states over the analyzed period. Next, we proceed to estimate a dynamic spatial econometric panel data model proposed by Vega and Elhorst (2016). This methodology allows us to identify two main results: first, to confirm the existence of spatial dependence of self-employment rates between states and, second, to estimate the sensitivity of each state to the national rate of self-employment. Finally, we

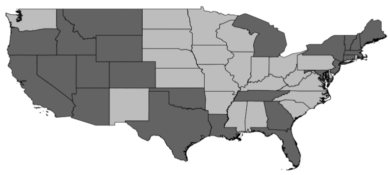
Figure 3.2: Nonfarm self-employment rate over space.



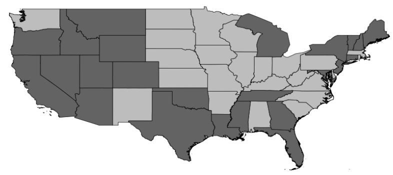
(a) 1998



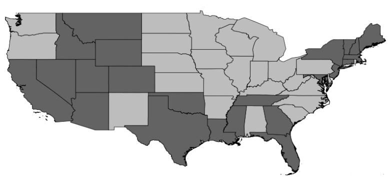
(b) 2003



(c) 2008



(d) 2013



(e) 2018

Note: Dark colors for states above the average.

try to relate the spatial information with the sensitivity to the national rate, clustering states into different groups.

3.2.1 ESDA

The self-employment rates of the neighboring states of a particular state can be good predictors of the self-employment rates of the focal region. To contrast the existence of spatial dependence, we perform a global spatial dependence test with Moran's I statistic (GMI, hereafter) with the aim of corroborating the presence of positive spatial dependence between different states in the whole country:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{3.1}$$

where x_i and x_j are observations of the dependent variable(self-employment) of states i and j, $\bar{X}$ is the average between states and w_{ij} is the ij element of the weight matrix. [2] N is the number of observations and $W = \sum_i^n \sum_j^n w_{ij}$ is the aggregate of all the spatial weights.

Our null hypothesis tests whether the self-employment rate is spatially independent—that is, is spatially autocorrelated—in which case a spatial econometric model must be estimated.

[2] We use the binary contiguity form — which is an NxN matrix with 1 if two regions are neighbors and 0 otherwise — that has been row-standardized. We select this weight matrix following Vega and Elhorst (2016) findings.

Second, we implement the local version of Moran's I statistic (LMI, hereafter) for each region and year to analyze the behavior of each state over the time:

$$I = \frac{x_i - \bar{X}}{S_i^2} \sum_{i=1, j \neq i}^{n} w_{i,j}(x_j - \bar{X}) \tag{3.2}$$

where x_i and x_j are observations of the dependent variable of states i and j, $\bar{x}$ is the average between states, and n is the number of states, w_{ij} is the ij element of the weight matrix and S_i^2:

$$S_i^2 = \frac{\sum_{i=1, j \neq i}^{n} (x_j - \bar{X})^2}{n - 1} \tag{3.3}$$

A positive and significant value of GMI indicates spatial clustering of similar values, while a negative and significant value indicates spatial clustering of dissimilar values. This result may suggest that, in general, states depend spatially on their neighbors. Following the approach of Anselin (1995), this statistic can be used as an indicator of significant local spatial clusters because it measures the similarity between a state and its surrounding neighbors (LMI). These results can be better visualized in a Moran's scatter plot (Anselin et al., 1996), which is a scatter plot divided into four quadrants. The upper right (commonly known as "hot spots", or HH hereafter) and lower left (commonly known as "cold spots", or LL here after) quadrants include the clusters with high (HH) and low (LL) values, respectively. The other two quadrants are taken as outliers because they imply negative spatial autocorrelation—that is, high (low) values surrounded by low (high) values.

3.2.2 Model

Many alternatives have been proposed in the literature to model spatial dependence. The spatial econometric model is an extended linear model to include spatial lags in the dependent variable, error term or independent variables (Elhorst, 2014a).

Our target model is the one proposed by Vega and Elhorst (2016). This model is among the few that simultaneously account for serial dynamics, spatial dependence and common factors. According to the authors of the model, this implication is important because not dealing with these issues simultaneously could produce biased results. Our model can be specified as follows:

$$U_t = \tau U_{t-1} + \delta W U_t + \eta W U_{t-1} + \Gamma_1 U n_t + \Gamma_2 U n_{t-1} + \mu + \epsilon_t \tag{3.4}$$

where U_t is a column vector with one observation of the dependent variable (self-employment) for every unit (i) at every point of time (t) measured in logs, as is typical in the literature. U_{t-1}, WU_t and WU_{t-1} are, respectively, vectors of temporal, spatial and spatiotemporal lags with τ, δ and η being autoregressive coefficients. W is an NxN matrix describing the arrangement of the regions in space [3]. Un_t and

[3] We use the row-standardized binary contiguity matrix.

Un_{t-1} are the national unemployment rates at times t and $t-1$. Γ_1 and Γ_2 are Nx1 column vectors with a unit-specified coefficient of response to the common factors, μ represents the spatial fixed effect added to the model, and ϵ_t is the independently and identically distributed Nx1 vector error term with zero mean and constant variance σ^2.

The parameter of state sensitivity to national rates of self-employment (γ) can be estimated by dividing the elements of Γ_1 by $1-\delta$ or by dividing the elements of Γ_2 by $-\tau-\eta$.[4]

Common Factors

Recently, the need to distinguish between common factors and spatial dependence, also known as strong cross-sectional dependence and weak cross-sectional dependence, has been noted in the field of spatial econometrics (Chudik, Pesaran, and Tosetti, 2011). Specifically, for self-employment, the spatial correlation may be the result of shared factors or local interactions between neighboring regions generating spillover effects. Since the self-employment rate of a state is part of the national self-employment rate, one could use the latter as a predictor of the former; in fact, the national self-employment rate is just the average of the states' rates (Pesaran, 2006), so introducing it in the model is similar to including temporal fixed effects. Otherwise, the introduction of both would cause multicollinearity, as the inclusion of temporal fixed effects implies controlling for those common factors omitted from all the geographical units during the period analyzed. It is important to understand that the introduction of temporal fixed effects can only cover part of what the inclusion of common factors (national average) does because these fixed effects assume that the impact is the same for all regions.

Our proposed model allows us to estimate a parameter for each geographical unit, which will show the heterogeneity of the regions with respect to the national rate, which could be interpreted as a measure of the sensitivity of the regional self-employment rate to the national rate.

Persistence

Since the dependent variable can be nonstationary—in which case it would have a unit root—it is necessary to include temporal dynamics into the model. By incorporating U_{t-1}, WU_{t-1} into the model, the dynamics of the percentage of nonagricultural self-employment in the focal region and its neighbors are modeled. The coefficients τ and δ will display the temporal lag or the persistence of the self-employment rate and the spatiotemporal lag. Together, they could be interpreted as a measure of the relative strength of internal and external habit persistence (Korniotis, 2010).

Additionally, when common factors are added into the model, they may also be nonstationary. In fact, if the dependent variable is nonstationary, the common factor will be nonstationary as well since it is the national rate or the states' average of nonagricultural self-employment.

[4]We will use the second method since it is based on the relative strength of internal and external habit persistence (Korniotis, 2010), which can better reflect the dynamism of the relations.

Cluster Analysis

Once we have estimated the parameter of state sensitivity to the national rate of self-employment, we can find a relationship between these parameters (common factors) and the times that a state has been a hot spot (HH) or a cold spot (LL) (spatial dependence).

To do this we implement one of the most commonly used unsupervised machine learning algorithms for partitioning a data set into k groups, k-means clustering (MacQueen et al., 1967). This algorithm classifies objects into k different groups (prespecified), where objects within the same cluster are as similar as possible and objects from different clusters are as dissimilar as possible. By analyzing the centers of each cluster, we can find which type of state, in terms of sensitivity and spatial dependence, corresponds to each cluster.

We use the Hartigan-Wong algorithm (Hartigan and Wong, 1979), which defines the total within-cluster variation as the sum of squared euclidean distances between states and the corresponding centroid:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{3.5}$$

where x_i is a data observation belonging to cluster C_k and μ_k is the mean value of the points assigned to cluster C_k.

To select the optimal number of clusters to be generated, k, we compute k-means clustering using k different values of clusters, and we estimated the total within-cluster variation defined as follows:

$$\sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{3.6}$$

The location of a bend (knee) in the plot of the total within-cluster variation of each k number of clusters is generally considered an indicator of the appropriate number of clusters.

3.3 Empirical Results

To show spatial dynamics in self-employment, we employ Moran's I statistics for global spatial autocorrelation (GMI) (Table 3.1 and Figure 3.3), a Moran's scatter plot and hot/cold spot analysis for local spatial autocorrelation (LMI) (Figures 3.4 and 3.5), the empirical model for self-employment accounting for serial dynamics, spatial dependence and common factors (Table 3.2 and figure 3.6) and the cluster analysis (Figure 3.7, 3.8 and table 3.3).

The results of global Moran's I are summarized in Table 3.1. The statistical significance of the Moran's I values are tested using both z test and p values. As shown in the table, the high Z scores and low p values suggest that Moran's I values are highly significant statistically and presents positive values. This result provides statistical evidence that the current level of self-employment in a state is correlated with the level of self-employment in the neighboring states as well as evidence of the tendency to cluster of self-employment in the US, which also indicates that regions with high (low) values of self-employment are probably close to regions with high (low) values.

Table 3.1: Global spatial dependence test

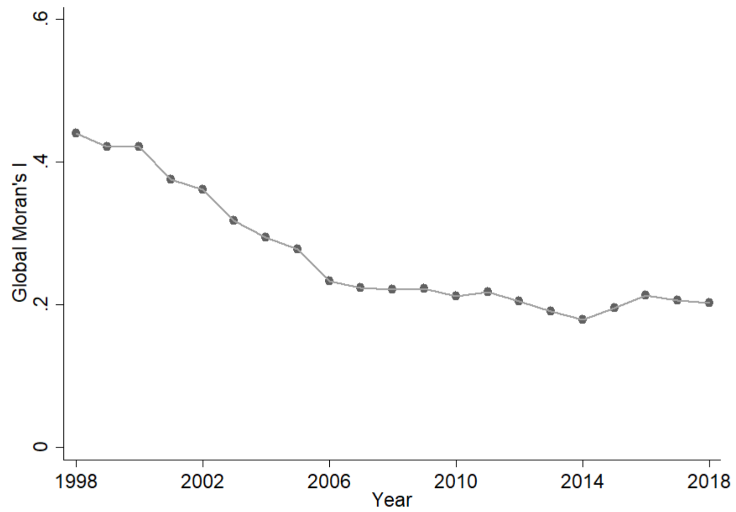
Year	Global Moran I	z	p-value
1998	0.440	4.908	0.000
1999	0.421	4.721	0.000
2000	0.421	4.746	0.000
2001	0.375	4.283	0.000
2002	0.362	4.167	0.000
2003	0.318	3.706	0.000
2004	0.295	3.449	0.000
2005	0.278	3.268	0.001
2006	0.233	2.791	0.003
2007	0.224	2.686	0.004
2008	0.222	2.646	0.004
2009	0.223	2.652	0.004
2010	0.212	2.533	0.006
2011	0.218	2.588	0.005
2012	0.205	2.436	0.007
2013	0.191	2.273	0.011
2014	0.179	2.138	0.016
2015	0.196	2.312	0.010
2016	0.213	2.482	0.007
2017	0.206	2.411	0.008
2018	0.203	2.372	0.009

Figure 3.3 shows the GMI value throughout the analyzed period. In general, the level of spatial autocorrelation between regions in 1998-2018 remains low (less than 0.5). It can be observed that this value is constantly decreasing and that, in 2006, it is half the value of 1998 showing a decline in the global spatial autocorrelation; in other words, the states throughout the US are becoming less dependent on each other.

To visually explore spatial autocorrelation, we create Moran scatter plots that illustrates the relationship between the values of the self-employment rate at a given location (z_t) and the average value of the same attribute at neighboring locations (Wz_t) (Anselin et al., 1996)[5]. This Moran scatter plots presented in figure 3.4 shows the dynamic of the local spatial autocorrelation with four types of local spatial association between a state and its neighbors in the four different quadrants of the scatter plot: high (low) values (self-employment) surrounded by states with high (low) values; HH (LL), which refers to positive spatial autocorrelation, indicating spatial clustering and high (low) values surrounded by low (high) values; HL (LH),

[5]The global spatial autocorrelation can be driven equally by all states or only by a few states; in fact, the sum of all local Moran tests would be almost equivalent to the result of the global Moran test. This simultaneous analysis can give us information about which states or clusters are driving the global dependence.

Figure 3.3: Global Moran's I over the period 1998-2018



which are considered outliers. It can be seen that all significant states (at 5% significance; otherwise the null hypothesis of spatial independence cannot be rejected) are positively spatially autocorrelated.[6]

Specifically, in 1998, 89% of the states that are positively spatially autocorrelated are HH (New Hampshire, Maine, Wyoming, Idaho, Utah, Oregon, Montana and Colorado). Clusters of high values of self-employment were formed. However, in 2018, the percentage of HH states is 12.5%, with the LL states being 87.5% (West Virginia,Indiana, Ohio, Kentucky, Wisconsin, Minnesota and Iowa). This is an important result since HH clusters have disappeared in favor of LL clusters. The finding indicates that global spatial autocorrelation was driven by states with high rates surrounded by states with high rates in 1998, while in 2018 states with low rates of self-employment surrounded by states with low rates were driving the global spatial autocorrelation. Additionally, global spatial dependence decreased from 1998 to 2018, as the slope of the line shows.

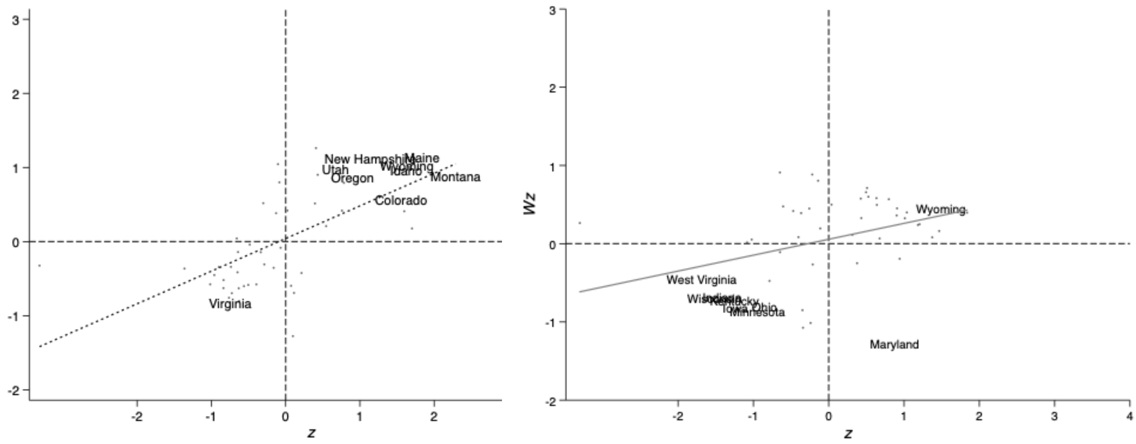
Since spatial arrangement of the Moran scatter plot information can play a relevant role in our set of results, we plot a map known as a hot/cold spot analysis in figure 3.5 for 1998, 2003, 2008, 2013 and 2018 where dark colors represent spatially dependent HH states, whereas spatially dependent LL regions appear in light colors.

The figure clearly shows that cold spots in the Great Lakes region and its surroundings regions (the north of the Southeast region and the east of the Plains region) are beginning to expand; on the other hand, the hot spot located in the Rocky Mountain region is disappearing across the period.

The previous spatial statistical evidence has shown significant spatial spillovers and cluster formation in the USA over time. However, not controlling for serial dynamics and common factors may produce biased results. For this reason, we carry out an estimation of the dynamic spatial panel data model with common factors

[6]With the exception of Maryland, which is an outlier spatially dependent with a high value of self-employment surrounded by states with low values.

FIGURE 3.4: Moran Sacatterplot



(A) 1998 (B) 2018

Note: Printed names are states with significant spatial dependence (at 5%). Small points represent spatially independent states.

proposed by Vega and Elhorst (2016) to confirm the spatiotemporal dynamic of self-employment in USA states and estimate the parameters of sensitivity to the national self-employment rate.

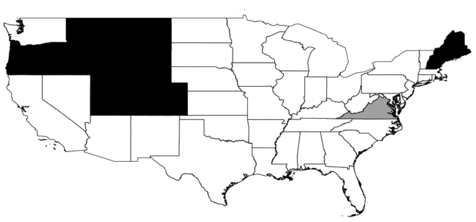
The results are reported in Table 3.2 [7] The temporal lag coefficient ($\tau = 0.876$) turns out to be highly significant, reflecting the strong correlation of self-employment rates over the observed time period. The spatial autoregressive coefficient ($\delta = 0.336$) appears to be positive and highly significant. This result suggests that spatial dependence is crucial in the context of the self-employment rate of US states even when serial dynamics and common factors are accounted for; the result also confirms our previous findings obtained with the GMI approach. The spatiotemporal lag coefficient η is significant and negative. Korniotis (2010) explain this parameter as the external habit of persistence.

The parameter of sensitivity to national rates (γ) could be taken as an indication that a state is cyclically sensitive if the value is greater than 1 (Vega and Elhorst, 2016). These parameters are represented in Table 3.2 where it can be seen that only some states have a parameter greater than one (40%). To understand the spatial distribution of γ parameters, Figure 3.6 maps this information, with dark color indicating numbers greater than 1 and light color numbers smaller than 1. It seems that the states most sensitive to the cycle are those that belong to or are neighbors of hot spots, while nonsensitive states seem to belong to cold spots.

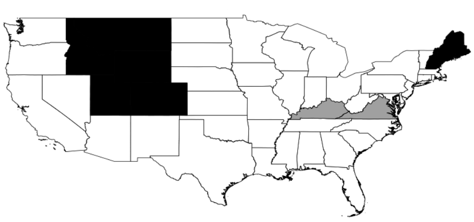
After analyzing the evolution of the spatial dependence (GMI) hot spots and cold spots (LMI) and the sensitivity to the national self-employment rate (dynamic spatial econometric model with common factors), we conduct a cluster analysis to find a relationship between the results obtained. To classify the states we use three variables. The first is the γ sensitivity parameter estimated in the dynamic spatial panel

[7] This model is stable since the $\tau + \delta + \eta - 1 < 0$ condition is satisfied.

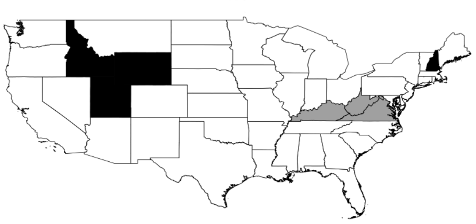
FIGURE 3.5: Hot and Cold spot analysis



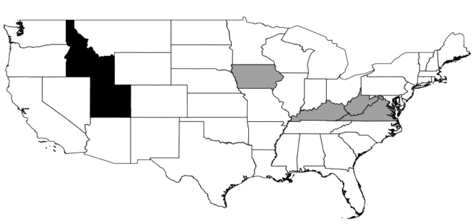
(A) 1998



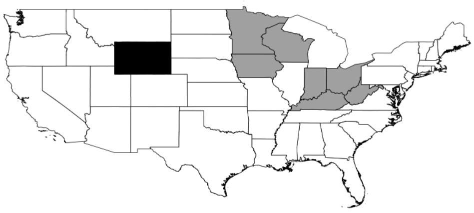
(B) 2003



(C) 2008



(D) 2013



(E) 2018

Note: Dark color for hot spots and grey color for cold spots.

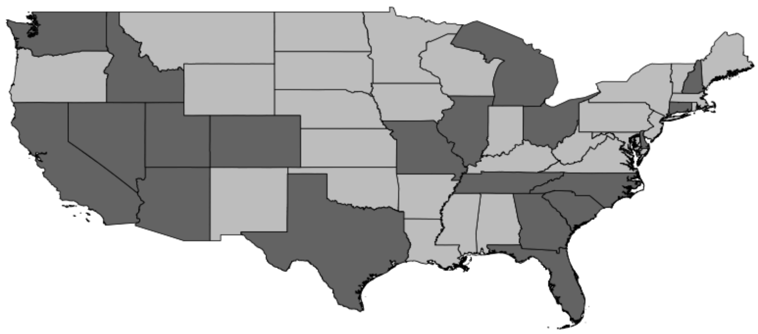
data model, the second variable is the number of times that a state has been hot spot, and the third is the number of times that a state has been a cold spot.[8] Before executing the algorithm to compute the clusters we need to identify the correct number of clusters (k) through the total within-cluster variation analysis.

Figure 3.7 shows the plot of the sum of squares within groups for each number of clusters where the location of what is usually called the "knee" can be determined when we compute the algorithm with 3 groups ($k = 3$), in other words, by partitioning the states into 3 groups, we can optimally reduce the total within-group variation.

Once we compute the results of the clustering by k-means with 3 groups, we can extract the information on the centroids of each group presented in table 3.3 and identify the characterization of the states that belong to each group as well as the possible relationship between the profile of each state in terms of spatial dependence and sensitivity to the national rate. Focusing on this information, cluster 1 includes states that have never been hot spots and, on average, have been cold spots for at least 15 years from 1998 to 2018. The average sensitivity of the first group is 0.586, the lowest of the three groups. Cluster 2 includes the states that have been hot spots for 11.714 years, on average during the analyzed period and have never been cold spots. The average sensitivity of the second group is 0.980, the highest sensitivity of

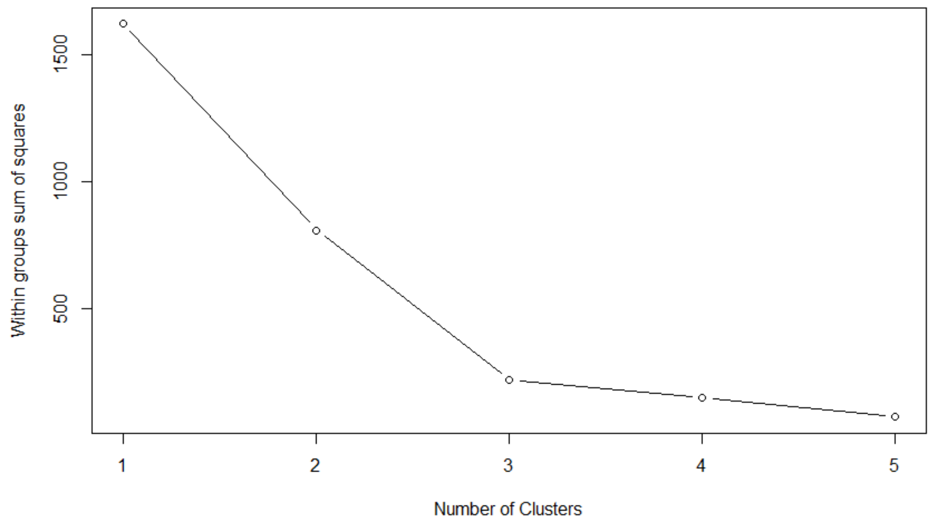
[8]The number of times that a state has been a cold spot is taken in negative.

FIGURE 3.6: Regions sensitive to the national self-employment cycle.



Note: The dark color represents sensitive regions.

FIGURE 3.7: Sum of squares within groups.

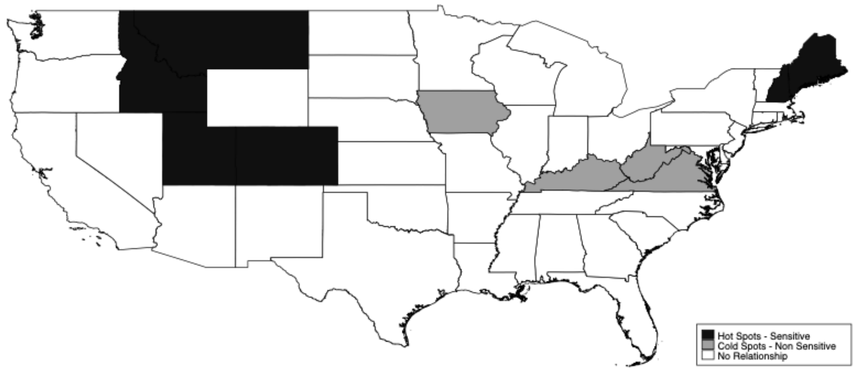


the three groups. Finally, cluster 3 includes states that have been hot spots for 0.205 years and cold spots for 0.462 years, with a sensitivity of 0.877.

In summary, cluster 1 includes states that are hot spots (states with spatial dependence and high self-employment rates) and are sensitive to the national rate, cluster 2 includes states that are cold spots (states with spatial dependence and low rates of self-employment) and are less sensitive to the national self-employment rate, and cluster 3 includes the rest of the states, which have no spatial dependence with a neither high nor low sensitivity. This cluster analysis also shows that there is a relationship between being a hot spot or cold spot and being sensitive or not to the national rate. Specifically, for the United States, states that are cold spots also show a low sensitivity and states that are hot spots a high sensitivity to the national self-employment rate.

To reveal which states belong to each cluster, Figure 3.8 shows each of the clusters found in different colors. The dark color represents cluster 1, which includes the states that are hot spots and are sensitive. These states are those in the Rocky Mountains (Idaho, Montana, Wyoming, Utah and Colorado) and in part of New England (Maine and New Hampshire). The gray color represents cluster 2, which includes cold spots that are nonsensitive and are located in the north of the Southeast region (Kentucky, Virginia and West Virginia). White represents the third cluster.

FIGURE 3.8: Clusters



Note: Dark color for HH and sensitive states, grey color for LL and non-sensitive states and white color for states with no relationship.

3.4 Conclusions

Regional self-employment clusters have become crucial due to their impact on regional economic performance. Consequently, a growing interest in geography-based research on the determinants of self-employment rates has developed among academics (Andersson and Koster, 2010). Although self-employment is currently widely recognized as a geographical phenomenon influenced by national and regional factors, the combination of both sets of factors is largely missing in the literature on the formation of regional self-employment clusters. This paper fills in this significant gap.

This study takes into account three perspectives in the literature that could help explain regional variations in self-employment rates. First, business growth is highly influenced by agglomeration effects (Reynolds, Miller, and Maki, 1995). These agglomeration effects are even more remarkable when we talk about conglomerates located in nearby regions, as companies can benefit from indirect interregional effects (Agarwal, Audretsch, and Sarkar, 2010; Delgado, Porter, and Stern, 2007). Therefore, we study the formation of regional clusters based on self-employment rates and the evolution of such clusters over time. In this way, we are able to detect the interaction between neighboring regions of the US, distinguishing between regions with low and high self-employment rates. Second, spatial proximity facilitates contact between regions, favoring the exchange and dissemination of knowledge (Audretsch and Keilbach, 2007; Singh, 2005) and, consequently, affecting self-employment (Audretsch and Keilbach, 2007). However, there is little empirical evidence as to whether the regions of a given country are spatially dependent (Fossen and Martin, 2018; Hong et al., 2015; Pijnenburg and Kholodilin, 2014). Hence, we investigate the existence of spatial dependence in self-employment rates among the US regions. Third, some characteristics of a particular region may cause self-employment rates to persist over time (Fotopoulos and Storey, 2017; Fritsch and Kublina, 2019). This persistence could influence the development of self-employment clusters. In this sense, we test whether self-employment rates are persistent among the US counties or not. Finally, we complete our analysis through the estimation of the parameters of state sensitivity to the national self-employment rate to test the influence of the national context on the configuration of regional self-employment clusters.

Our results are manifold. On the one hand, we confirm the existence of self-employment regional clusters in the United States. This implies that the policies implemented by regional governments can influence spatially dependent neighboring regions. Moreover, while these clusters were composed of regions characterized by high self-employment rates surrounded by other high-self-employment regions (HH clusters), this composition has changed in the most recent periods. In particular, HH clusters have disappeared in favor of LL clusters, which means that the global spatial dependence is now driven by states with low rates of self-employment surrounded by states with low rates of self-employment. This finding suggests that measures taken by governments of high-self-employment regions could have a decreasing impact on the spatially dependent neighboring regions, as the special dependence of these groups is increasingly less significant. In contrast, measures developed by policy makers of the regions that form LL clusters could have a significant impact on the regions of that cluster. Moreover, we find evidence of spatial dependence among US regions, although it has deteriorated gradually. In addition, our results show that regional self-employment rates are persistent. Finally, our results corroborate that regional self-employment rates are sensitive to the national self-employment rate, which means that national circumstances have a relevant impact on regions. Furthermore, we find that the regions sensitive to the national self-employment rate are those characterized as part of HH clusters. A subsequent implication of this finding is that policy makers could predict the effects that self-employment strategies will have to address since national self-employment policies are likely to have a greater impact on regions sensitive to the national self-employment rate.

In sum, knowledge on the persistence of regional self-employment rates, the change in the levels of spatial dependence between regions, and the evolution of regional clusters over time can be of great importance for policy strategies at both the national and the regional levels focused on stimulating the creation of new companies. Therefore, our paper helps to better illuminate the dynamics of regional self-employment across US regions and provides comprehensive results that may be of interest for those involved in the design and implementation of business policy.

Additional research could investigate the factors behind the current decrease in spatial dependence among US regions as well as the change in regional self-employment cluster formation. A better understanding of these phenomena would help in implementing more effective business measures and assessing their potential.

TABLE 3.2: Dynamic Spatial Panel Data Model

τ	δ	η	Nobs	R^2 corrected	log-likelihood
0.876***	0.336***	-0.287***	980	0.993	351.890

States with $\gamma < 1$	γ	States with $\gamma > 1$	γ
Alabama	0.845	Arizona	1.414
Arkansas	0.872	California	1.482
District of Columbia	0.047	Colorado	1.569
Indiana	0.969	Connecticut	1.000
Iowa	0.635	Delaware	1.050
Kansas	0.165	Florida	1.132
Kentucky	0.888	Georgia	1.649
Louisiana	-0.389	Idaho	1.490
Maine	0.831	Illinois	1.009
Maryland	0.862	Michigan	1.533
Massachusetts	0.969	Missouri	1.190
Minnesota	0.912	Nevada	1.481
Mississippi	0.623	New Hampshire	1.084
Montana	0.421	North Carolina	1.007
Nebraska	0.068	Ohio	1.152
New Jersey	0.857	South Carolina	1.149
New Mexico	0.493	Tennessee	1.297
New York	0.73	Texas	1.283
North Dakota	0.406	Utah	1.532
Oklahoma	0.024	Washington	1.129
Oregon	0.939		
Pennsylvania	0.430		
Rhode Island	0.431		
South Dakota	0.730		
Vermont	0.968		
Virginia	0.754		
West Virginia	0.066		
Wisconsin	0.879		
Wyoming	-0.115		

Note: Standard errors in parentheses. *** denotes 1% significance.

TABLE 3.3: Cluster Centroids

	Cluster 1	Cluster 2	Cluster 3
HH years	11.500	0.000	0.154
LL years	0.000	-13.750	-0.385
Sensitivity	1.155	0.585	0.857

Note: numbers represents averages within groups.

Chapter 4

The price elasticity of cigarettes: new evidence from Spanish regions, 2002-2016.

4.1 Introduction

Historically, there has always been a concern among politicians and academics to explain the behaviour of the tobacco market, focusing on finding mechanisms to minimize tobacco consumption, thus reducing the negative effects it generates on public health (Chaloupka, Straif, and Leon, 2011). Indeed, tobacco represents an important part of the budget of high-income nations in which 16.6% of the population over 15 years of age smokes daily and health spending is approximately 11.5% of the national GDP on average (Papanicolas, Woskie, and Jha, 2018). Knowing the details of the tobacco market is crucial for establishing an adequate framework for market regulation and health management. Although tobacco consumption generates addiction, there is a negative relationship between the price of tobacco products and their consumption (Contreary, Chattopadhyay, Hopkins, et al., 2015). This negative relationship has prompted some governments to establish laws that prohibit tobacco manufacturers and retailers from selling below a minimum price to reduce health costs in the population (Escario and Molina, 2004; Pinilla and Abásolo, 2017).

In this context, the tools to manage cigarette consumption are converging with health management, but they are also important for their effects on the illicit tobacco market. In this sense, tax collection and social damages have been analysed through theoretical models to estimate the equilibrium point between them (Saffer and Chaloupka, 1994). From the seminal work of Townsend (1988), based on cross-sectional data from 27 European countries where there was a price elasticity of -0.4, many studies have analysed the relationship between price and tobacco consumption (Gallus, Schiaffino, La Vecchia, et al., 2006; Kostova, Tesche, Perucic, et al., 2013; Fernandez, Gallus, Schiaffino, et al., 2004) for a recent overview see Jawad, Lee, Glantz, et al. (2018). Although the body of literature that analysed this relationship is focused on price elasticity as a basic assumption with the territorial isolation of countries (Fuchs and Meneses, 2018; Liu et al., 2015), some studies have analysed the spatial dependence of the territories (Hoffer, Humphreys, and Ruseski, 2018; Lipton, Banerjee, Levy, et al., 2008; Yu, Peterson, Sheffer, et al., 2010). Furthermore, concerning the spatial dependence of the territories, the idea that tobacco contraband may be playing an important role in this market emerges (Curti, Shang, Chaloupka, et al., 2019). In particular Baltagi and Li (2004), using panel data on cigarette consumption in 46 US states from 1963 to 1992, employed and inspired multiple new models and estimators in spatial econometrics aiming to understand how territorial effects work (Debarsy, Ertur, and LeSage, 2012; Elhorst, 2014a; Kelejian and Piras, 2014). In this way Ciccarelli, De Fraja, and Tiezzi (2020) recently warned that the policies carried out by the government can be distorted by the spatial dependence of the territories. Even more importantly, they showed that the role played by territories, using the pricing method of the Italian government in the 19th century to maximize the benefits generated by the monopoly of the tobacco, implies that the market takes advantage of the spatial dependence of the sub-territories that make up a country. Furthermore, a very recent work by Ciccarelli and Elhorst (2018) reported an analysis of the consumption of tobacco in Italy between 1877 and 1913. They used a dynamic spatial model that allowed estimation of the direct and indirect effects on tobacco consumption and distinguished the behaviour of the regions, finding that less urban regions are more sensitive to national trends than regions with more urban centres.

In this way, our study analysed the two components of elasticity prices, that is, the direct effect of each territory and the indirect effect caused by neighbouring territories. There is no evidence in the literature from a spatial analysis that clarified the territorial effectiveness of the price policies regarding taxation that governments impose to reduce tobacco consumption. To the best of our knowledge, this study is the first that simultaneously analysed, first, a dynamic spatial model used to measure the price elasticity of cigarettes in the short term and long term of the 47 provinces that make up the Spanish territory, detailing the influence of neighbours. Second, given the spatial arrangement of the elasticities observed in the provinces, we can detect behaviours typical of large-scale illicit trade and cross-border purchasing since geographical location can be an important factor in smuggling, and politicians should take this into account when making price policies (Curti, Shang, Chaloupka, et al., 2019).

In addition, this work also analyses the short-run and long-run price elasticity of demand for cigarettes. There are some works that have analyzed the addictive nature of tobacco consumption, discovering that tobacco demand responds more to long-run prices than to short-run (Chaloupka and Tauras, 2011). In the economic literature there are several works that analyze the impact of addiction on the consumption of some addictive goods, such as tobacco. Becker is a pioneering author in analyzing the behavior of the demand for harmful addictive goods. The explanation provided by Becker and Murphy in their Theory of Rational Addiction (Becker and Murphy, 1988) is that the consumers of goods harmful to health take into account the future effects of current consumption when determining the optimal amount of addictive merchandise that they will consume in the present moment. Becker and Murphy assume that an increase in the price of the addictive product causes a decrease in the demand for the addictive product over time. Therefore, in addictive products the price elasticities of demand in the long-run must be greater than those in the short-run. In the case of Spain, Escario and Molina (2000) shows that the price elasticity of demand for cigarettes amounts to -0.7 and -0.84 in the short-run and long-run, respectively. In this work the authors affirm that the resulting estimates support that in Spain the hypothesis of rational addiction of the Becker Murphy theory is fulfilled.

In this paper, we use the following structure. Section 2 describes the data and the empirical application, Section 3 presents the main results, and Section 4 discusses some implications for academics, practitioners and policymakers.

4.2 Empirical Strategy: data and methodology

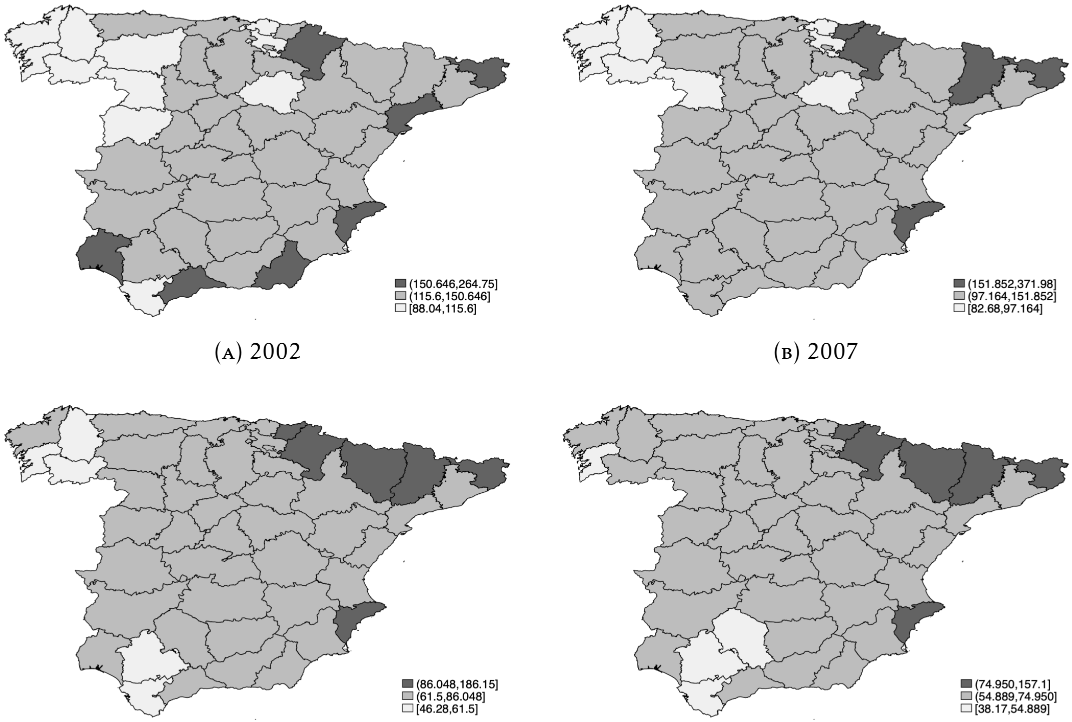
4.2.1 Data

Our empirical analysis was developed using a panel of data from the Spanish provinces from 2002 to 2016 - last data of the provincial GDP published correspond to the year 2016 -. For cigarette consumption, we used the annual tobacco official sales and the average price of a pack of 20 cigarettes in euros, as published by the Commission for Trade of the Tobacco. The real Gross Domestic Product (GDP) is available in the National Institute of Statistics from Spain. All series employed here are per capita (18 years or older), expressed in real terms using the consumer price index (CPI base 2016) provided by the same source using log values as is common in the literature.

As a preview of the selected variables and the configuration of the sample in the Spanish regions, Figure 4.1 presents a geographical analysis of the dynamics plotted for the Spanish Tobacco Market for several periods (2002, 2007, 2012 and 2016). As can be seen, the geographical configuration of the per capita cigarette consumption has changed substantially from 2002 to 2016. Since 2002 (Figure 4.1.A), the consumption of cigarettes in the northwest of the country - An explanation was offered in Lampe (2009), which shows how Galicia has been historically used by criminal organizations from America to smuggle drugs and tobacco to Europe - was lower than average, while normal consumption was concentrated in the centre of the country. This behaviour shows a first approximation of spatial dependence due to the progressive creation of geographical clusters. In 2007 (Figure 4.1.B), it can be seen that the grouping of provinces with consumption lower than the Northwest average becomes smaller. On the other hand, the group of average consumption advances towards the south and begins to create a cluster of high consumption per capita in the north border with France. In 2012 (Figure 4.1.C), the decline of the low-consumption group in the northwest was consolidated and one began to be created in the south - they are border regions with Gibraltar, a territory of the United Kingdom where tobacco is less expensive than in Spain. In addition, it is a route historically used for tobacco smuggling (O'reilly, 1999) -. In addition, the cluster of high consumption of the north grows until covering the entire border with France. Finally, in 2016 (Figure 4.1.D), the northwest group dissolves and the low consumption group in the south and the high consumption group in the north are consolidated (Despite recent tax increases, Spain continues to have the lowest tobacco prices in Western Europe. Thus, A pack of cigarettes in Spain is 40 percent cheaper than in France (Pinilla, Negrín, González-Lopez Valcárcel, et al., 2018)).

This article analyzes the price elasticity of demand for cigarettes. A greater presence of substitute products in the market may be associated with a higher price elasticity of demand for cigarettes. Therefore, it is important to know the composition of tobacco consumption. Specifically, cigarettes, cut tobacco, hand-rolled cigarettes and cigars are sold in Spain. In order to unify units of measurement, cigars have been grouped into packages of twenty units. Furthermore, to convert cut tobacco and hand-rolled tobacco into cigarettes, the equivalence proposed in (Gallus, Lugo, Ghislandi, et al., 2014), 0.75 g per cigarette, has been used. In addition, the resulting cigarettes have also been grouped into packs of twenty units. The results obtained

FIGURE 4.1: Tobacco Consumption per capita in Spain



(A) 2002

(B) 2007

(C) 2012

(D) 2016

are those shown in the following figure:

Figure 4.2 shows that substitutive products consumption has increased over time. However, as can be seen, the dramatic drop in cigarette consumption has not been absorbed by substitute products. Specifically, in the 2008-2016 period, total cigarette consumption decreased by 2191 millions of packs, while the use of substitute products increased by 190.86 millions of packs. This assumes that only 8.71% of the decrease in cigarette consumption has been absorbed by the use of substitute products. The rest of the drop is due to a lower prevalence of smoking or tax evasion.

4.2.2 Methodology

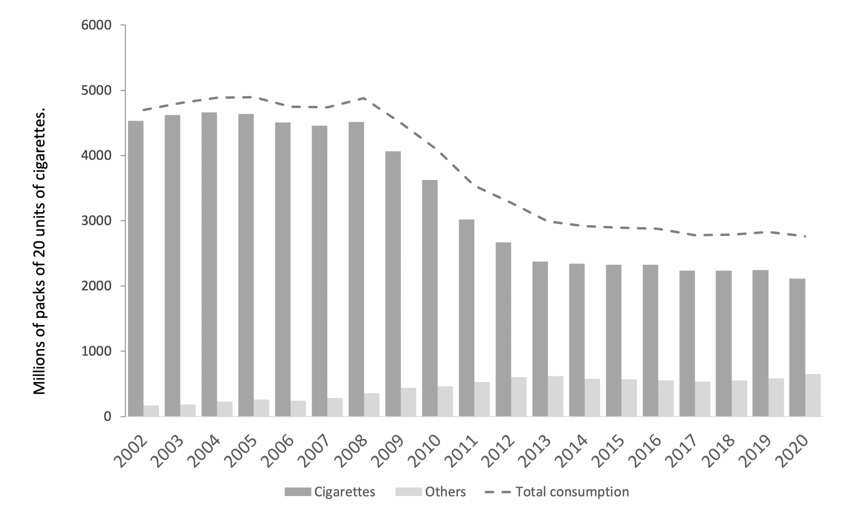
Our empirical analysis used an approximation of the dynamic spatial panel data model with a common factor implemented by Kelejian and Piras (2014).

The proposed model -developed by Lee and Yu (2010) - is the following:

$$C_t = \alpha_1 C_{t-1} + \alpha_2 WC_t + \alpha_3 WC_{t-1} + \beta_1 GDPpc_t + \Gamma_1 Price_t + \mu + \epsilon_t \tag{4.1}$$

where C_t is a column vector with one observation of the dependent variable (tobacco consumption) for every unit i in every point at time t. C_{t-1}, $WC_{i,t}$ and $WC_{i,t-1}$

FIGURE 4.2: Official sales of tobacco products, Spain, 2002–2016.



are vectors of temporal, spatial and spatiotemporal lag with α_1, α_2 and α_3 autoregressive coefficients, respectively. W is an nxn matrix describing the arrangement of the regions in the space with 1 if two regions are neighbours and 0 if not.

GDPpc_t is the per capita income by region i and time t with the β_1 coefficient, Price_t is one observation by region and time of the national price (it is the same for all regions in each year), Γ_1 is an Nx1 column vector with a unit-specified coefficient of response to the common price (which allows measurement of the differences in price sensibility), μ_i represents the spatial fixed effect added to the model – in order to check the robustness we also made estimates using the model with time-period fixed effects, which produced similar results, but we decided to remove them from the model because adding the price as a common factor could produce multicollinearity with time-period fixed effects (Vega and Elhorst, 2016) -, and ϵ_t is the vector independently and identically distributed error term with a mean of zero and constant variance.

We also try several models with more observable or unobservable common factors such as the cross-sectional average of C_t, C_{t-1} or $GDPpc_t$. However, first two common factors proposed would cause multicollinearity problems as the case of time-period fixed effects since national tobacco consumption is strongly correlated with the price in our case study. The inclusion of GDPpct as a common factor produces similar results as the proposed model. Note that each common factor added imply the inclusion of 49 new parameters to estimate and the consequent loss of degrees of freedom.

By estimating a dynamic model, we can check how tobacco consumption of the previous year t-1 explains the present t consumption of tobacco. A high persistence in tobacco consumption during the period could indicate that the policies carried out to date aiming to reduce tobacco consumption have been ineffective, while low persistence rates could indicate a decoupling of the trend to consumption due to past behaviour. Dynamic analysis of tobacco consumption is commonly used in the literature (Baltagi and D., 1992; Galbraith and Kaiserman, 1997).

Also, taking into account the spatial distribution and the possible dependence among regions allows us to verify how tobacco consumption in region j explains the consumption of tobacco in region i if they are neighbours, both in the same period of time t and in the following period t+1 (dynamic model), which is important for several reasons, and regional policies must be elaborated and take into account the possible spillovers they may cause (Anselin and Griffith, 1988). Therefore, detecting these influences demonstrated problems, such as areas sensitive to hiding by proximity to other countries, with price differentials that can affect neighbouring regions, and therefore, the problem was expanded - this has been studied by different authors applying several methodologies (Stoklosa, 2020; Goel and Saunoris, 2018; Calderoni, Dugato, Aglietti, et al., 2017) -.

In addition, two independent variables were incorporated into the model. The first one is Price. By increasing tobacco price by taxes, tobacco consumption is reduced (Chaloupka, Yurekli, and Fong, 2012; Wilson, Avila Tang, Chander, et al., 2012). The second one is GDPpc because many papers have shown that there is a positive relationship between income and cigarette consumption (Ciccarelli, De Fraja, and Tiezzi, 2020; Gallet and List, 2003; Martinez, Mejia, and Perez-Stable, 2015). Following the procedure of the introduction of common factors developed in (Vega and Elhorst, 2016), we introduced the variable price as a common factor since the price is established by the national government and is the same for all regions in the case of Spain. A similar procedure can be found in (Elhorst, Madre, and Pirotte, 2020). By incorporating it into the model as a common factor, we obtained a parameter for each region that measures the price elasticity, which provides identification of the differences between them. This analysis is important when different regions share the same price, as is the case in Spain. The fact that there is a common price means that price policies are fixed commonly for all in that territory, which does not imply that all regions react in the same way to such policies. This methodology helps us to discover differences among regions.

Finally, individual fixed effects have been added to account for possible regional heterogeneity, whose omission could bias the estimates (Baltagi, 2008). The coefficients obtained from the independent variables are not directly interpretable, since it is a model that simultaneously accounts for spatial and temporal dependence. Through the transformation proposed in (Baltagi and Li, 2004) for spatial dynamics panel data models, we obtained the direct effects, which were those with a change in the independent variable of the region itself, i, which causes effects on tobacco consumption of the region i. The indirect effects are variations in the tobacco consumption of a region i due to changes in the independent variables of the neighbouring regions j. These changes can be both short and long term.

In summary, this model allows measurement of the influences between neighbouring regions, the direct and indirect effects of GDPpc and Price, observation of the differences in the sensitivity to price between regions and the effects in the short and long term. In other words, our main objective was to discover whether tobacco consumption in a region is affected by tobacco consumption in neighbouring regions (spatial dependence), which is the true behaviour of regions and their reaction to price.

4.3 Results

Before to present the results of the estimated dynamic panel data model, we performed the Moran's I test whose null hypothesis is the absence of correlation between neighbouring regions in cigarette consumption. Its value in our analysis was between $0.216 < I < 0.308$ over the period analysed, which indicates the presence of spatial dependence and cluster generation in cigarette consumption and leads to the conclusion that regions with similar cigarette consumption tend to cluster.

Results of the dynamic panel data model are shown in Table 4.1. As shown in Panel A, first, the very significant autoregressive coefficient ($\alpha_1 = 0.424$) shows a relatively low persistence of tobacco consumption. This result reveals that the current consumption is slightly explained by previous consumption. In addition, compared to the spatial lag coefficient ($\alpha_2 = 0.553$), it was lower, which led to the first important result of the model. The positive and very significant spatial lag coefficient in period t reveals that the consumption of the regions is influenced by the consumption of the neighbouring regions in the same period. Second, the fact that the spatial lag coefficient is higher than the autoregressive coefficient could reveal that tobacco consumption is more dependent on what happens around a certain region than on the consumption of the region in the previous period. Third, the coefficient of space-time dependence ($\alpha_3 = -0.236$) was very significant, negative and lower in magnitude than the two previous ones. The explanation for the negative sign of this parameter can be found in Tao and Yu (2012), which defines this parameter as an intertemporal budget constraint and suggests that omitting this variable can produce significant biased results and that including an irrelevant spatial time lag causes no loss of efficiency.

The direct and indirect effects in the short and long term of the independent variables are reported in Table 4.1, Panel B - incorporating the Price as a common factor allows obtaining a coefficient for each province-.

The direct effects show how a change in the independent variables (GDPpc or Price) of region i influences the tobacco consumption of region i. The indirect effects show how a change in the independent variables of the neighbouring region j influence the consumption of region i. The sum of both would be the effect of a change in the dependent variables in both the region itself and the neighbouring regions.

To date, most studies have only been concerned with calculating effects within their own region (Rodrıguez-Iglesias, Schoj, Chaloupka, et al., 2017; Stoklosa, Goma, Nargis, et al., 2019; Yeh, Schafferer, Lee, et al., 2017), obviating, in this case, the spillovers that the neighbouring regions generate, not only directly by consumption clusters but also through changes in their variables. When we look at the short-term direct coefficient of (GDPpc = 0.192), we see that a 1% increase in the GDPpc of a certain region increases the tobacco consumption of the region by 0.192%, but taking into account that an increase of 1% in the GDPpc of the neighbouring regions would increase tobacco consumption of the region itself by 0.199%, the total effect of an increase of 1% of GDPpc in all regions of Spain would cause an increase in tobacco consumption of approximately 0.40% in the short-term.

Table 4.1: Estimation results

PANEL A						
Serial dynamics		Spatial dependence coefficients				Corr-R2
α_1	0.424***(0.035)	α_2	0.553***(0.041)	α_3	-0.236***(0.048)	0.772
PANEL B						
	Direct	Indirect	Total	Direct	Indirect	Total
GDPpc						
	0.192***(0.025)	0.199***(0.028)	0.392***(0.044)	0.334***(0.005)	0.345***(0.005)	0.679***(0.007)
Price						
Almería	-0.300***(0.038)	-0.312***(0.043)	-0.612***(0.069)	-0.521***(0.007)	-0.538***(0.008)	-1.060***(0.011)
Cádiz	-0.450***(0.043)	-0.468***(0.058)	-0.918***(0.077)	-0.781***(0.008)	-0.808***(0.011)	-1.589***(0.011)
Córdoba	-0.369***(0.041)	-0.384***(0.049)	-0.754***(0.071)	-0.642***(0.007)	-0.663***(0.009)	-1.306***(0.011)
Granada	-0.277***(0.043)	-0.286***(0.039)	-0.563***(0.072)	-0.480***(0.008)	-0.493***(0.060)	-0.974***(0.011)
Huelva	-0.295***(0.044)	-0.305***(0.040)	-0.600***(0.073)	-0.512***(0.008)	-0.526***(0.007)	-1.038***(0.011)
Jaén	-0.248***(0.031)	-0.257***(0.039)	-0.506***(0.068)	-0.432***(0.007)	-0.444***(0.007)	-0.876***(0.011)
Málaga	-0.452***(0.044)	-0.471***(0.059)	-0.923***(0.079)	-0.785***(0.008)	-0.814***(0.011)	-1.599***(0.012)
Sevilla	-0.497***(0.044)	-0.518***(0.064)	-1.015***(0.081)	-0.863***(0.008)	-0.895***(0.013)	-1.759***(0.013)
Huesca	-0.290***(0.043)	-0.300***(0.041)	-0.590***(0.072)	-0.503***(0.008)	-0.518***(0.007)	-1.021***(0.011)
Teruel	-0.184***(0.041)	-0.190***(0.036)	-0.375***(0.072)	-0.321***(0.007)	-0.327***(0.006)	-0.648***(0.012)
Zaragoza	-0.257***(0.040)	-0.266***(0.042)	-0.524***(0.072)	-0.446***(0.007)	-0.460***(0.007)	-0.906***(0.011)
Cantabria	-0.337***(0.037)	-0.351***(0.050)	-0.688***(0.072)	-0.584***(0.006)	-0.607***(0.010)	-1.191***(0.012)
Albacete	-0.246***(0.041)	-0.254***(0.039)	-0.501***(0.073)	-0.427***(0.008)	-0.438***(0.007)	-0.866***(0.011)
Ciudad Real	-0.249***0.038)	-0.258***(0.037)	-0.508***(0.067)	-0.433***(0.007)	-0.445***(0.007)	-0.878***(0.010)
Cuenca	-0.252***(0.039)	-0.261***(0.039)	-0.514***(0.068)	-0.438***(0.007)	-0.451***(0.007)	-0.890***(0.011)
Guadalajara	-0.310***(0.036)	-0.323***(0.047)	-0.633***(0.071)	-0.539***(0.007)	-0.558***(0.009)	-1.097***(0.011)
Toledo	-0.318***(0.037)	-0.331***(0.048)	-0.649***(0.071)	-0.552***(0.006)	-0.572***(0.009)	-1.124***(0.012)
Ávila	-0.260***(0.040)	-0.269***(0.040)	-0.530***(0.071)	-0.452***(0.007)	-0.465***(0.007)	-0.917***(0.011)
Burgos	-0.278***(0.038)	-0.288***(0.042)	-0.566***(0.068)	-0.482***(0.007)	-0.498***(0.008)	-0.981***(0.011)
León	-0.254***(0.037)	-0.263***(0.040)	-0.518***(0.067)	-0.441***(0.007)	-0.455***(0.007)	-0.897***(0.011)
Palencia	-0.231***(0.040)	-0.239***(0.038)	-0.470***(0.070)	-0.401***(0.007)	-0.411***(0.006)	-0.812***(0.011)
Salamanca	-0.293***(0.038)	-0.305***(0.043)	-0.599***(0.070)	-0.510***(0.007)	-0.527***(0.008)	-1.037***(0.012)
Segovia	-0.283***(0.038)	-0.294***(0.044)	-0.578***(0.070)	-0.492***(0.007)	-0.508***(0.008)	-1.000***(0.011)
Soria	-0.202***(0.039)	-0.208***(0.037)	-0.410***(0.070)	-0.351***(0.007)	-0.359***(0.006)	-0.710***(0.011)
Valladolid	-0.301***(0.038)	-0.313***(0.044)	-0.614***(0.069)	-0.523***(0.007)	-0.541***(0.008)	-1.064***(0.012)
Zamora	-0.246***(0.038)	-0.254***(0.038)	-0.500***(0.067)	-0.427***(0.007)	-0.439***(0.007)	-0.866***(0.011)
Barcelona	-0.218***(0.047)	-0.224***(0.037)	-0.442***(0.077)	-0.378***(0.008)	-0.385***(0.006)	-0.764***(0.012)
Girona	-0.395***(0.041)	-0.411***(0.052)	-0.807***(0.073)	-0.687***(0.008)	-0.711***(0.010)	-1.398***(0.012)
Lleida	-0.457***(0.040)	-0.477***(0.062)	-0.934***(0.079)	-0.794***(0.008)	-0.824***(0.012)	-1.619***(0.013)
Tarragona	-0.369***(0.040)	-0.384***(0.051)	-0.753***(0.073)	-0.640***(0.007)	-0.664***(0.010)	-1.304***(0.012)
Madrid	-0.284***(0.039)	-0.295***(0.042)	-0.580***(0.070)	-0.494***(0.007)	-0.510***(0.008)	-1.004***(0.011)
Navarra	-0.311***(0.039)	-0.323***(0.042)	-0.634***(0.070)	-0.540***(0.007)	-0.558***(0.008)	-1.098***(0.011)
Alicante	-0.445***(0.039)	-0.465***(0.042)	-0.910***(0.067)	-0.773***(0.006)	-0.805***(0.013)	-1.578***(0.014)
Castellón	-0.316***(0.037)	-0.329***(0.064)	-0.646***(0.080)	-0.550***(0.007)	-0.569***(0.009)	-1.119***(0.011)
Valencia	-0.296***(0.039)	-0.307***(0.045)	-0.603***(0.071)	-0.513***(0.007)	-0.530***(0.008)	-1.044***(0.011)
Badajoz	-0.231***(0.040)	-0.239***(0.044)	-0.470***(0.072)	-0.402***(0.008)	-0.411***(0.006)	-0.813***(0.012)
Cáceres	-0.245***(0.043)	-0.254***(0.039)	-0.499***(0.074)	-0.426***(0.007)	-0.437***(0.006)	-0.864***(0.011)
A Coruña	-0.273***(0.041)	-0.284***(0.038)	-0.557***(0.071)	-0.475***(0.007)	-0.490***(0.008)	-0.965***(0.011)
Lugo	-0.203***(0.039)	-0.210***(0.043)	-0.414***(0.071)	-0.353***(0.007)	-0.362***(0.006)	-0.716***(0.011)
Ourense	-0.192***(0.039)	-0.198***(0.038)	-0.391***(0.071)	-0.334***(0.007)	-0.342***(0.006)	-0.676***(0.012)
Pontevedra	-0.342***(0.040)	-0.356***(0.037)	-0.698***(0.072)	-0.593***(0.007)	-0.615***(0.010)	-1.209***(0.012)
La Rioja	-0.221***(0.038)	-0.228***(0.050)	-0.449***(0.073)	-0.383***(0.007)	-0.393***(0.006)	-0.777***(0.011)
Álava	-0.259***(0.040)	-0.268***(0.037)	-0.527***(0.070)	-0.449***(0.007)	-0.463***(0.007)	-0.913***(0.011)
Guipúzcoa	-0.489***(0.037)	-0.511***(0.070)	-1.001***(0.085)	-0.850***(0.007)	-0.885***(0.014)	-1.735***(0.014)
Vizcaya	-0.165***(0.041)	-0.170***(0.035)	-0.336***(0.073)	-0.288***(0.007)	-0.292***(0.006)	-0.580***(0.012)
Asturias	-0.239***(0.039)	-0.248***(0.039)	-0.487***(0.071)	-0.416***(0.007)	-0.427***(0.007)	-0.843***(0.011)
Murcia	-0.265***(0.042)	-0.274***(0.040)	-0.540***(0.073)	-0.461***(0.008)	-0.473***(0.006)	-0.934***(0.012)
Average	-0.296	-0.307	-0.603	-0.514	-0.53	-1.044
Std. Dev.	0.08	0.084	0.165	0.139	0.147	0.286
Min	-0.498	-0.518	-1.016	-0.863	-0.896	-1.759
Max	-0.166	-0.17	-0.336	-0.288	-0.292	-0.581

Note: Standard errors are reported in parentheses; Panel A shows the main results of the dynamic spatial panel data model, while Panel B shows the coefficients of direct and indirect effects in the short and long term of the independent variables.
*** Significance at 1

Focusing on Price, the short-term direct effect to an increase of 1% of the Price in the region itself decreases consumption by -0.28% on average; however, the fact that there is spatial dependence and the Price increase also affects the neighbouring regions (because of common price), the increase becomes 1% (in both the neighbouring regions and in the region itself), reducing tobacco consumption by approximately twice what was estimated without taking into account the spillovers among regions.

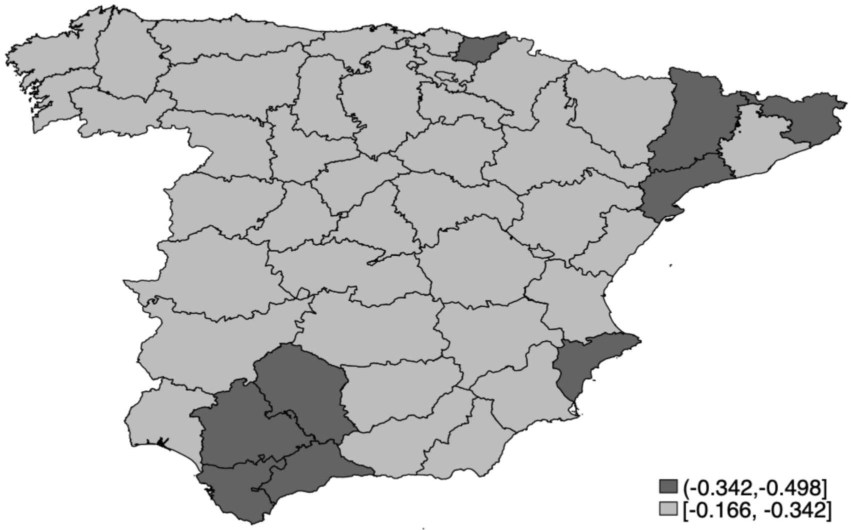
In the long term, all the effects are greater than the short term, which reveals the persistence of the effect of the changes in the GDPpc and Price on tobacco consumption. Furthermore, the total long-term price elasticity exceeds the unit in absolute terms in approximately 50% of the provinces studied. Elasticities greater than traditional values may lead to the conclusion that price reduces the prevalence of smoking more than is traditionally accepted. However, according to previous works, a price elasticity of demand for cigarettes higher (in absolute value) than traditional values is due to the increase in tax evasion (Chaloupka and Tauras, 2011). In most of the works that analyse the price elasticity of demand for cigarettes, official data is used, that is, legal sales data. Therefore, a greater sensitivity of the population to changes in the price of cigarettes may be associated with a higher consumption of illegal cigarettes. Another reason that the literature attributes to a high price elasticity of demand for cigarettes is the existence of more affordable substitute products. If substitute products are a cheaper alternative to cigarettes, then the price increases, with the aim of reducing overall demand; however, it may lead some smokers to replace the cigarette with a cheaper substitute (Cornelsen and Normand, 2013).

Following the statistical summary in the end of Table 1, a clear variance between provinces can be found. For example, the direct short-term price elasticity of tobacco consumption (typical estimated parameter in the literature), takes values between -0.166 and -0.498 with an average value of -0.296. These results show that in Spain there are provinces with a very different behaviour in tobacco consumption when price change.

These differences can be found in Figure 4.4 where provinces are divided into 2 groups. The dark colour represents provinces with the highest price-elasticities in Spain. A k-means cluster analysis applied to price-elasticity parameters reveals that this province forms a group statistically different from the rest of the country. These differences are important because the price in Spain is set by the central government and fiscal policies regarding the price of tobacco can have different impacts in different provinces.
Results seem to be robust since the estimation of other models such as the one with the inclusion of GDPpct as a common factor or the one proposed by (Ciccarelli and Elhorst, 2018) leads to very similar results that strength our findings.

Focusing on the most sensitive provinces, we can distinguish several behaviours. First, in the south, Cádiz borders Gibraltar and the neighbouring regions that have a low average price (consumption of lower-quality tobacco). The same situation occurs in the north where the Guipúzcoa, Lleida and Girona regions bordering France also have low average prices (low quality) and high sensitivity to price. Finally, Malaga, Alicante and Tarragona, which are touristic provinces in Spain, are sensitive regions but with high prices (consumption of higher-quality tobacco). In summary, it seems that the geographical distribution of the price elasticity of demand for cigarettes is

Figure 4.3: Price elasticity of cigarettes in Spain



Note: Dark colours represent regions with higher price elasticity, and light colours represent regions with lower price elasticity.

in accordance with the explanation provided by the literature when a price elasticity of demand for cigarettes is found that is higher than traditional values. As can be seen, the most sensitive areas are those bordering Gibraltar and France (countries with cigarette price differentials that make tax evasion and consumption of illegal cigarettes more attractive). In addition, Alicante and Tarragona are added to these areas, two usual destinations for foreign tourists who can take advantage of their tourist trips to acquire and transport cigarettes cheaper than in their country of origin. A recent study (Stoklosa, 2020) shows that a convergence of cigarette prices across EU Member States would reduce cross-border cigarette purchasing.

In this scenario, following (Stoklosa, 2020), few studies that have analysed smuggling distinguish between large-scale illicit trade and cross-border purchasing. The results presented in the previous paragraph show pieces of evidence. The first two situations describe regions bordering Gibraltar and France that can be classified as large-scale illicit trade because they are regions with border influence and have low average prices that can be explained by an attempt by smugglers to achieve the greatest price differential. Third, the group formed by tourist regions seem to fit into the so-called cross-border purchasing of high-quality tobacco that tourists do for domestic consumption.

Overall, in an environment of spatial dependence among regions, an analysis of the sensitivity to the price and quality of tobacco is a useful tool for detecting and classifying smuggling. In general, we found that the regions that are most sensitive to price are those bordering France and Gibraltar or tourist regions, demonstrating the effect that smuggling has on the behaviour of the regions - we did not find that border regions with Portugal had different behaviours, which can be explained by the fact that the price differential with this country is low -.

4.4 Conclusion

In recent years, there has been a growing interest to understand the mechanisms that can control cigarette consumption due to the large impact of cigarettes not only for health management but also for their effects on the illicit marketing of tobacco. In this context, the empirical literature devoted to the analysis of the effectiveness of economic policy tools concludes that the most effective policies are sustained in the intervention of cigarette prices through taxes. This study has shown that the consumption of cigarettes is influenced by the neighbouring regions and also measured different sensitivities for each region. Considering this influence, the effects of Price and GDP on cigarette consumption have been estimated through a dynamic spatial panel data model to measure these effects in the short and long term. Furthermore, as the price of cigarettes is common for the entire territory studied, the proposed model allows estimation of the price elasticity for each region and verifies the existing differences.

In particular, the results found on the income elasticity of cigarette consumption are similar to those found in the previous literature, i.e., the generally accepted value of 0.4. Regarding the standard price elasticity of cigarette consumption usually described in the empirical literature, we only observed it in the short term for certain provinces (-0.4). This may be because the influence of neighbours has traditionally not been considered. This influence is measured by the indirect effect and, in most cases, causes the price elasticity of cigarettes to be twice what is usually accepted. Moreover, the price elasticity of cigarettes in the long term exceeds in many cases, in absolute value, unity. This second result is novel because tobacco has historically been treated as an inelastic demand good. Therefore, when governments develop policies to control tobacco consumption, considering cigarettes as a product of inelastic demand, the policies implemented will be more effective than expected in terms of health. Furthermore, the results also suggested that the previous tobacco consumption of a certain region is a weaker indicator than the consumption of the neighbours in the same period. For this reason, regional cooperation in tobacco control policies may have better effects than the elaborated policies based on historical information.

To the best of our knowledge, this study is the first to quantify the price elasticity among regions. This allows detection of regions where policies for the control of tobacco consumption by prices are less effective than desired. In this sense, we found that the most sensitive regions are the border and tourist zones, evidencing the existence of large-scale illicit trade and cross-border purchasing. It should be noted that there are no results that show smuggling of tobacco in the border areas with Portugal. This result highlights the effectiveness of the common policies implemented by Portugal, which consists of maintaining a low-price differential with Spain.

This set of results reveals several recipes for the agendas of the agents involved. Academics should account for spatial dependence to measure tobacco consumption instead of temporal analysis. For their part, policymakers should consider that tobacco could be an elastic good in the long term and that cooperation between countries in terms of price differential should be taken to avoid tobacco smuggling. The allocation of resources to control smoking should consider the special dependence shown in this report. If this is not the case, practitioners in provinces where per capita consumption is medium will be harmed because they will have fewer resources than those in provinces where consumption is distorted due to this dependence.

Finally, the results seem to show that price increases are having the desired effect on public health, as they have a negative impact in cigarette consumption. However, tax evasion (both from Gibraltar to Spain and from Spain to France) suggest that the Spanish government is not realizing the full public health benefit that the increase in the price of cigarettes generates. In this way, this result seems to recommend that the fight against tax evasion must accompany the increase in the price of cigarettes, so that the decrease in cigarette consumption is real and is not influenced by the effect of tax evasion. Lastly, the increase in the price of cigarettes must be accompanied by a review of the behaviour of substitute products.

Chapter 5

A spatial analysis of the Spanish tobacco consumption distribution: Is there any consumption clusters?

5.1 Introduction

Economic convergence in Europe has been an important issue since the establishment of the trade and customs unions has led to many efforts being made to reduce the policy inequalities (Borsi and Metiu, 2015).

Indeed, analyze some sectors that, due to failures in cooperation between countries legislation, may have cross-border activity with relevant consequences on policy effectiveness. In particular, the most regulated products can suffer the consequences of the lack of cooperation between governments. For example, the tobacco market and its firm government regulation to reduce health consequences, have been analyzed over time by several academics focusing on the European Union (West et al., 2008) and Spain (Martín-Sánchez et al., 2018; Rana et al., 2016)c due to the existence of cross-border cigarette purchases between countries in those locations (Joossens and Raw, 2012). Tobacco price differential across the European Union creates a favorable environment for cross-border cigarette purchases, where the spatial distribution of tobacco consumption plays an important role in detecting it (Agaku et al., 2016).

Although there are studies that have uncovered the cross-border purchases of cigarettes (Blecher, Gilmore, and Ross, 2012), few studies have analyzed the tobacco consumption distortions that it generates. A very recent study (Stoklosa, 2020) finds that price differentials constitute the main issue that generates cross-border purchases, and policymakers should utilize tax harmonization between countries to discourage it. In this body of literature, our study focuses on Spain where, through spatial analysis we can detect distortions in the spatial distribution of per capita tobacco consumption. It is important to note that this analysis is possible because in Spain the price is established by the national government and is the same for all regions.

The objective of this short communication is to provide a previous empirical analysis to locate the regions that have distortions in per capita tobacco consumption. The location of these regions and their proximity to other countries allow to detect the need that governments have to harmonize policies.

5.2 Method

To develop our empirical analysis, we employed panel data from the 47 Spanish provinces from 2002 to 2017 (Canary Island, Balearic Island, Ceuta and Melilla are treated as island as usual in the literature of spatial analysis). We used the annual official tobacco sales as published by the Tobacco Market Commission of Spain, and we have divided them among the population over 18 years of age to calculate per capita tobacco consumption. The population over 18 years old is available in the National Institute of Statistics in Spain.

We analyze the behavior of per capita tobacco consumption in the various regions, implementing the local version of Moran´s I statistic for each region and year. This method allows us to detect spatial cluster formation with significant high or low per capita tobacco consumption:

$$I = \frac{x_i - \bar{X}}{S_i^2} \sum_{i=1, j \neq i}^{n} w_{i,j} (x_j - \bar{X}) \tag{5.1}$$

where x_i and x_j are observations of the per capita tobacco consumption of regions i and j, $\bar{X}$ is the average between regions, n is the number of regions, $w_{i,j}$ is the ij element of the weight matrix and S_i^2:

$$S_i^2 = \frac{\sum_{i=1, j \neq i}^{n} (x_j - \bar{X})^2}{n-1} \tag{5.2}$$

We estimate this test with 20 different weight matrices including contiguity matrices (rook and queen of order 1 and 2), distance matrix (150km, 200km, 250km, 300km and inverse matrices) and k-nearest neighbors matrices (with k=2, k= 3, k = 4 and k = 5). All matrices testes produce similar results, so we decided that the first-order contiguity matrix is the one that better reflects these spillovers between provinces due to its simplicity and not having to make arbitrary considerations about whether there are relationships beyond the provinces with which borders are shared.

A positive and significant value allow us to find spatial clusters of similar per capita tobacco consumption. In sum, we estimate 752 tests for each region and each year. We use this information to implement a Hot and Cold Spot Analysis. This analysis present on a map the location of the clusters found. Significant clusters of high sales are called “hot spot” and significant clusters of low sales are called “cold spot”. Through this analysis we can detect where, the clusters of regions with high or low per capita tobacco consumption are located.

Before performing the Hot and Cold Spot analysis, we estimated whether there is spatial dependence between the provinces throughout the period analyzed (2002-2017) using the CD-test developed by Pesaran (2015), whose null hypothesis is the absence of strong spatial dependence between regions. The result of this test is CD

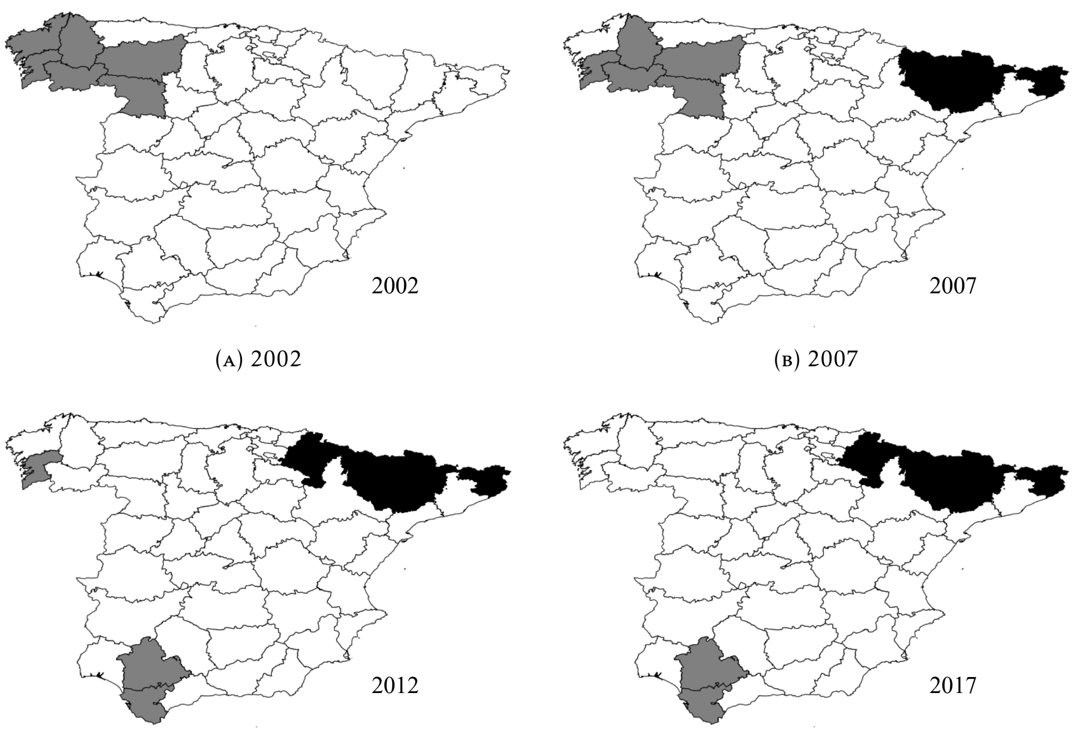
= 130.882 and p-value = 0.000 so the null hypothesis of weakly cross-sectional dependent is rejected showing that tobacco consumption of the provinces in Spain is correlated with tobacco consumption of other provinces.

Finally, an approximation is made by comparing price differentials between neighboring countries.

5.3 Results

Results of the spatial analysis and a comparison of prices between countries are presented in this section. Figure 5.1 shows the results of the hot and cold spot analysis represented in four maps for years 2002, 2007, 2012 and 2017. These graphs include the presence of three significant clusters, two of them cold spots (low per capita tobacco consumption in grey color) and one a hot spot (high per capita tobacco consumption in black color).

FIGURE 5.1: Hot spot, cold spot maps



(A) 2002 (B) 2007 (C) 2012 (D) 2017

Note: Grey color for clusters of low tobacco consumption and black color for clusters of high tobacco consumption.

The first cluster, in grey color, can be observed in the northwest area, mainly in the regions of Galicia. The presence of this cluster indicates that the per capita tobacco consumption in these regions was noticeably lower than the rest of the regions. Over time, it decreases to be null in the present, where there is no cluster formation. The second clusters, in dark color, are, in this case, areas of high per capita tobacco consumption (hot spot); they are located on the border with France, and these regions have a per capita tobacco consumption significantly higher than the rest of regions.

The third cluster, in grey color, which shows low per capita tobacco consumption (cold spot), appeared on the Gibraltar border and its surrounding areas for the first time in 2012 and remains there today. Being a cluster of low consumption tells us that the per capita tobacco consumption in these regions is substantially lower than the rest of regions. As price can play an important role in this situation, we analyze the price differential (measured in percentage of the Spanish price) between Spain and its bordering countries (Portugal and France) for the period 2004-2017. Data has been taken from a recent study (Stoklosa, 2020).

This data show how the price differential between Spain and Portugal maintained throughout the analyzed years has always been low (between 0.82% and 22.22%), while the price differential between Spain and France has always been quite high (between 50.59% and 156.41%) with a decreasing trend in recent years. No tobacco price data have been found regarding Gibraltar, but a recent memorandum by the Government of Gibraltar indicates the need to harmonize the price differential with Spain by up to 32% because the current price differential is generating illicit tobacco trade in Spain (HM Government of Gibraltar, 2018).

These results, together with the cluster analysis, show clear evidence that the existence of high price differentials (Gibraltar greater than 32% and France greater than 50.59%) generate distortions in the per capita tobacco consumption, and this occurs not only in the border regions but also in regions close to these due to the existence of spillovers and the generation of clusters. On the other hand, maintaining a low-price differential between countries does not generate distortion, as evidenced by the Portuguese case.

5.4 Discussion

Tobacco, which is strongly regulated by the government due to its negative effects on public health, has been a product susceptible to cross-border purchases between countries over the years. The novelty of this short communication is measuring the per capita tobacco consumption distortions through a spatial analysis.

By analyzing the consumption clusters that are generated in Spain we find that, first, in 2002, the lowest per capita consumption of legal tobacco in Spain was concentrated in Galicia. This coastal region has been one of the most utilized by organizations dedicated to the smuggling of tobacco and drugs from America11. However, in 2017, its behaviour is similar to that of the Spanish average.

Second, we observe a cluster of high per capita consumption in the border area between Spain and France, which can be explained by price differential and the possible cross-border purchases of tobacco that distort the consumption in this area.

Third, a cluster of low consumption is detected in the area bordering Gibraltar, which can be explained by the price differential that can generate cross-border purchases, too. However, we do not find clusters of consumption in the Portuguese border regions. This result can be explaining by the tobacco low-price differential between Spain and Portugal.

In sum, this short communication can serve as basis for governments to detect areas where the lack of price harmonization for tobacco products between countries, can produce consumption distortions with health consequences. Future research may investigate the consequences of the lack of harmonization in health policies by focusing in the causal relationship between per capita tobacco consumption over time and variations in price differential with nearest countries to find an optimal price differential that does not generate distortions.

Chapter 6

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

6.1 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.

6.2 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.

$$y = \rho_1 W_1 y + X\beta + W_1 X\theta + \epsilon \tag{6.1}$$

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]:

$$y = \rho_1 W_1 y + X\beta + W_1 X\theta + \Delta\alpha + \epsilon \tag{6.2}$$

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:

$$\alpha = \rho_2 W_2 \alpha + u \tag{6.3}$$

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 6.1 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.

6.3 Results

Results are divided into three parts where the results of the model estimates are found in tables 6.1, 6.2 and 6.3 and the results of the national random effects are found in Figure 6.2. 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.

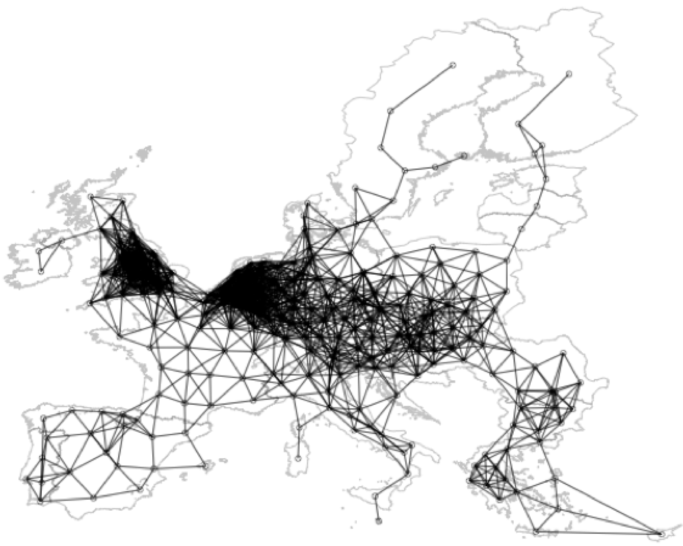
Regarding the comparison of the HSAR model with the HSDM in the three years, it is observed that, following the log-likelihood, the HSDM model is slightly better for the applied case although they are very similar. The estimated coefficients do not have significant changes and the interpretations in 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 6.1: 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 6.1: 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 6.2: 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:

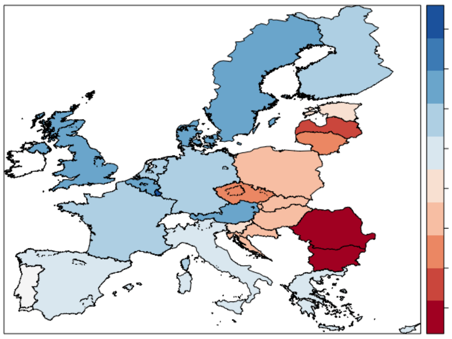
Finally, figure 6.2 summarizes the estimation of the national random coefficients representing them in maps and caterpillars plots. The maps show the spatial distribution of these random effects that could be interpreted as the national context, with positive values being a favourable context and negative 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 6.3: Results 2014

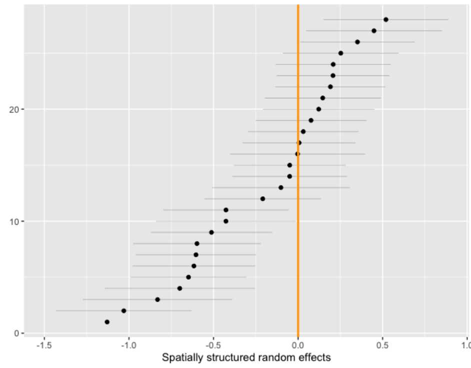
2014	*HSAR MODEL*		*HSDM MODEL*	
	Value	SE	Value	SE
ρ_1	0.071	0.025	0.270	0.083
ρ_2	0.882	0.104	0.692	0.159
L	0.410	0.044	0.403	0.045
K	0.660	0.042	0.666	0.044
W_1L			−0.175	0.059
W_1L			-0.075	0.093
Constant	−1.457	0.462	-0.987	0.471
Observations (NUTS2)	263		263	
Countries	28		28	
Pseudo R^2	0.983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

Note:

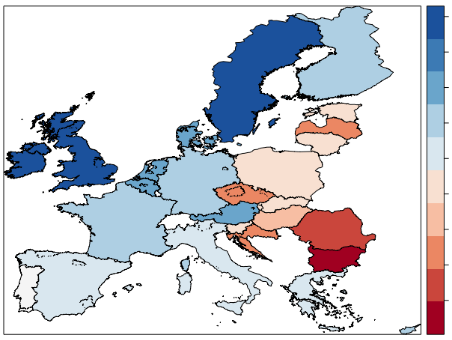
Figure 6.2: National random effects maps and caterpillar plots.



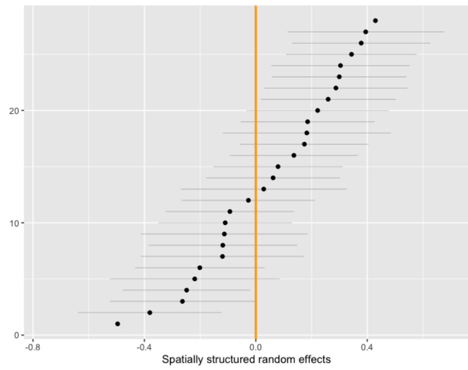
(a) 2000 - Map



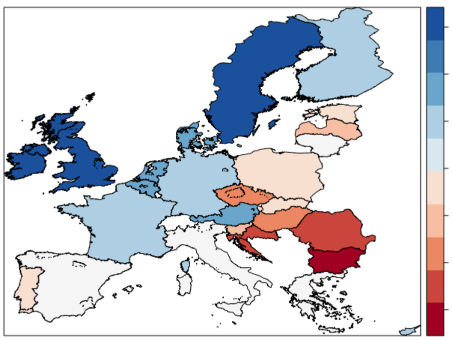
(b) 2000 - Caterpillar



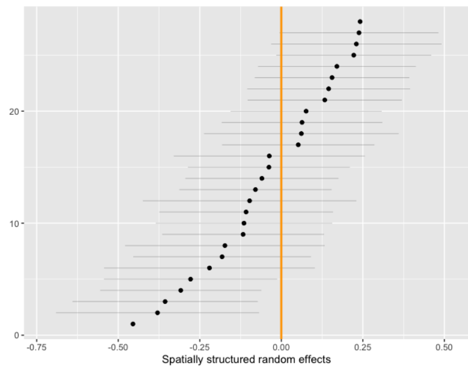
(c) 2007 - Map



(d) 2007 - Caterpillar



(e) 2014 - Map



(f) 2014 - Caterpillar

6.4 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.

Bibliography

Acs, Zoltan J and Catherine Armington (2006). *Entrepreneurship, geography, and American economic growth*. Cambridge University Press.
Agaku, Israel T et al. (2016). "Impact of cigarette price differences across the entire European Union on cross-border purchase of tobacco products among adult cigarette smokers". In: *Tobacco control* 25.3, pp. 333–340.
Agarwal, Rajshree, David Audretsch, and MB Sarkar (2010). "Knowledge spillovers and strategic entrepreneurship". In: *Strategic Entrepreneurship Journal* 4.4, pp. 271–283.
Albulescu, Claudiu Tiberiu and Aviral Kumar Tiwari (2018). "Unemployment persistence in EU countries: new evidence using bounded unit root tests". In: *Applied Economics Letters* 25.12, pp. 807–810.
Andersson, Martin and Sierdjan Koster (2010). "Sources of persistence in regional start-up rates—evidence from Sweden". In: *Journal of Economic Geography* 11.1, pp. 179–201.
Anselin, L. and DA. Griffith (1988). "Do spatial effects really matter in regression analysis?" In: *Pap Reg Sci* 65, p. 1.
Anselin, Luc (1995). "Local indicators of spatial association—LISA". In: *Geographical analysis* 27.2, pp. 93–115.
Anselin, Luc (2001). "Spatial effects in econometric practice in environmental and resource economics". In: *American Journal of Agricultural Economics* 83.3, pp. 705–710.
Anselin, Luc (2002). "Under the hood issues in the specification and interpretation of spatial regression models". In: *Agricultural economics* 27.3, pp. 247–267.
Anselin, Luc and Wendy K Tam Cho (2002). "Spatial effects and ecological inference". In: *Political analysis*, pp. 276–297.
Anselin, Luc and Raymond JGM Florax (1995). "Small sample properties of tests for spatial dependence in regression models: Some further results". In: *New directions in spatial econometrics*. Springer, pp. 21–74.
Anselin, Luc et al. (1996). "Simple diagnostic tests for spatial dependence". In: *Regional science and urban economics* 26.1, pp. 77–104.
Armington, Catherine and Zoltan J Acs (2002). "The determinants of regional variation in new firm formation". In: *Regional studies* 36.1, pp. 33–45.
Audretsch, David B and Maryann P Feldman (1996). "R&D spillovers and the geography of innovation and production". In: *The American economic review* 86.3, pp. 630–640.
Audretsch, David B and Max Keilbach (2007). "The theory of knowledge spillover entrepreneurship". In: *Journal of Management Studies* 44.7, pp. 1242–1254.
Bailey, Natalia, Sean Holly, and M Hashem Pesaran (2016). "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence". In: *Journal of Applied Econometrics* 31.1, pp. 249–280.
Baldwin, Richard and Eiichi Tomiura (2020). "Thinking ahead about the trade impact of COVID-19". In: *Economics in the Time of COVID-19*, p. 59.
Baltagi, B. (2008). *Econometric Analysis of Panel Data*. John Sons; Wiley.

Baltagi, B. H. and Levin D. (1992). “Cigarette taxation: raising revenues and and reducing consumption”. In: *Struct. Chang* 3, p. 2.

Baltagi, B. H. and D. Li (2004). “Prediction in the panel data model with spatial correlation”. In: *Advances in spatial econometrics*. ; p. 283–295: Springer.

Bande, Roberto, Melchor Fernández, and Víctor Montuenga (2008). “Regional unemployment in Spain: Disparities, business cycle and wage setting”. In: *Labour Economics* 15.5, pp. 885–914.

Bande, Roberto and Marika Karanassou (2009). “Labour market flexibility and regional unemployment rate dynamics: Spain 1980–1995”. In: *Papers in Regional Science* 88.1, pp. 181–207.

Becker, G. S. and KM. Murphy (1988). “A theory of rational addiction”. In: *Polit Econ* 96, pp. 675–700.

Bjørnskov, Christian and Nicolai J Foss (2008). “Economic freedom and entrepreneurial activity: Some cross-country evidence”. In: *Public Choice* 134.3-4, pp. 307–328.

Blecher, Evan, Anna Gilmore, and Hana Ross (2012). “Price and Tax Measures and Illicit Trade in the Framework Convention on Tobacco Control: What We Know and What Research Is Required”. In:

Borsi, Mihály Tamás and Norbert Metiu (2015). “The evolution of economic convergence in the European Union”. In: *Empirical Economics* 48.2, pp. 657–681.

Bosma, Niels and Rolf Sternberg (2014). “Entrepreneurship as an urban event? Empirical evidence from European cities”. In: *Regional studies* 48.6, pp. 1016–1033.

Braunerhjelm, Pontus et al. (2010). “The missing link: knowledge diffusion and entrepreneurship in endogenous growth”. In: *Small Business Economics* 34.2, pp. 105–125.

Brechling, Frank (1967). “Trends and cycles in British regional unemployment”. In: *Oxford Economic Papers* 19.1, pp. 1–21.

Calderoni, F., M. Dugato, V. Aglietti, et al. (2017). “Price and Non-price Determinants of the Illicit Cigarette Trade: Analysis at the Sub-national Level in the EU”. In: *Dual Markets:* 267.

Camacho, Máximo, Matias Pacce, and Camilo Ulloa Ariza (2018). “Regional Business Cycle Phases in Spain.” In: *Estudios de Economía Aplicada* 36.3.

Cancelo, José Ramón (2004). “El ciclo del empleo en España: un análisis regional”. In: *Clm. economía*, pp. 107–138.

Carmona, Mónica et al. (2016). “Self-employment and business cycles: searching for asymmetries in a panel of 23 OECD countries”. In: *Journal of Business Economics and Management* 17.6, pp. 1155–1171.

Carree, Martin A and A Roy Thurik (2003). “The impact of entrepreneurship on economic growth”. In: *Handbook of entrepreneurship research*. Springer, pp. 437–471.

Carree, Martin Anthony and Adriaan Roy Thurik (2008). “The lag structure of the impact of business ownership on economic performance in OECD countries”. In: *Small Business Economics* 30.1, pp. 101–110.

Chaloupka, F. J. and J. A. Tauras (2011). *The demand for cigarettes in Ireland*. Health Service Executive.

Chaloupka, F. J., A. Yurekli, and G. T. Fong (2012). “Tobacco taxes as a tobacco control strategy”. Tobacco control, 21(2), 172-180.

Chaloupka, FJ, K. Straif, and ME. Leon (2011). “Effectiveness of tax and price policies in tobacco control”. In: *Tob Control* 20.3, pp. 235–238.

Chemin, Matthieu (2009). “The impact of the judiciary on entrepreneurship: Evaluation of Pakistan’s “Access to Justice Programme””. In: *Journal of Public Economics* 93.1-2, pp. 114–125.

Cheng, Shaoming and Huaqun Li (2011). "Spatially varying relationships of new firm formation in the United States". In: *Regional Studies* 45.6, pp. 773–789.

Chudik, Alexander and M Hashem Pesaran (2016). "Theory and practice of GVAR modelling". In: *Journal of Economic Surveys* 30.1, pp. 165–197.

Chudik, Alexander, M Hashem Pesaran, and Elisa Tosetti (2011). "Weak and strong cross-section dependence and estimation of large panels". In: *The Econometrics Journal* 14.1, pp. C45–C90.

Ciccarelli, C., G. De Fraja, and S. Tiezzi (2020). "How hard is it to maximize profit? Evidence from a 19th-century Italian state monopoly". In: *Oxf Econ Pap.* (forthcoming).

Ciccarelli, C. and JP. Elhorst (2018). "A dynamic spatial econometric diffusion model with common factors: The rise and spread of cigarette consumption in Italy". In: *Reg Sci Urban Econ* 72, p. 131.

Contreary, KA., S. K. Chattopadhyay, D. P. Hopkins, et al. (2015). "Economic impact of tobacco price increases through taxation: a community guide systematic review". In: *Am J Prev Med* 49.5, pp. 800–808.

Cornelsen, S. and C. Normand (2013). "Is roll-your-own tobacco substitute for manufactured cigarettes: evidence from Ireland?" In: *J Public Health (Oxf).* 2013.36, p. 1.

Corrado, Luisa and Bernard Fingleton (2012). "Where is the economics in spatial econometrics?" In: *Journal of Regional Science* 52.2, pp. 210–239.

Cravo, Tulio A, Bettina Becker, and Adrian Gourlay (2015). "Regional growth and SMEs in Brazil: a spatial panel approach". In: *Regional Studies* 49.12, pp. 1995–2016.

Cuaresma, Jesús Crespo, Gernot Doppelhofer, and Martin Feldkircher (2014). "The determinants of economic growth in European regions". In: *Regional Studies* 48.1, pp. 44–67.

Cuéllar-Martín, Jaime, Ángel L Martín-Román, and Alfonso Moral (2019). "An Empirical Analysis of Natural and Cyclical Unemployment at the Provincial Level in Spain". In: *Applied Spatial Analysis and Policy* 12.3, pp. 647–696.

Curti, D., C. Shang, F. J. Chaloupka, et al. (2019). "Tobacco taxation, illegal cigarette supply and geography: findings from the ITC Uruguay Surveys". In: *Tob Control* 28.

Debarsy, N., C. Ertur, and JP. LeSage (2012). "Interpreting dynamic space–time panel data models". In: *Stat* 9, pp. 1–2.

Debarsy, Nicolas (2012). "The Mundlak approach in the spatial Durbin panel data model". In: *Spatial Economic Analysis* 7.1, pp. 109–131.

Deboosere, Robbin et al. (2019). "Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue". In: *Regional Studies, Regional Science* 6.1, pp. 143–156.

Delgado, M, ME Porter, and S Stern (2007). "Clusters, convergence and economic performance (mimeo)". In: *NBER*.

Ditzen, Jan (2019). "XTCSE2: Stata module to estimate the exponent of cross-sectional dependence in large panels". In: *Boston College Department of Economics*.

Domazlicky, Bruce R (1980). "Regional business cycles: a survey". In: *Journal of Regional Analysis and Policy* 10.1100-2016-89839, pp. 15–34.

Doms, Mark, Ethan Lewis, and Alicia Robb (2010). "Local labor force education, new business characteristics, and firm performance". In: *Journal of Urban Economics* 67.1, pp. 61–77.

Dong, Guanpeng and Richard Harris (2015). “Spatial autoregressive models for geographically hierarchical data structures”. In: *Geographical Analysis* 47.2, pp. 173–191.

Edquist, Charles and Björn Johnson (1996). *Institutions and organizations in systems of innovation*. Univ.

Elhorst, J. P., J. L. Madre, and A. Pirotte (2020). “Car traffic habit persistence, cross-sectional dependence, and spatial heterogeneity: New insights using French departmental data”. In: *Transp Res Part A Policy Pract* 132, pp. 614–632.

Elhorst, J Paul (2014a). “Dynamic spatial panels: models, methods and inferences”. In: *Spatial Econometrics*. Springer, pp. 95–119.

Elhorst, J Paul (2014b). “Spatial panel data models”. In: *Spatial econometrics*. Springer, pp. 37–93.

Elhorst, J Paul (2017). “Spatial Panel Data Analysis.” In: *Encyclopedia of GIS* 2, pp. 2050–2058.

Elhorst, J Paul, Marco Gross, and Eugen Tereanu (2018). “Spillovers in space and time: Where spatial econometrics and global VAR models meet”. In:

Escario, J. J. and J. A. Molina (2000). “Estimating anticipated and nonanticipated demand elasticities for cigarettes in Spain”. In: *Int Adv Econ Res* 6.4, pp. 782–793.

Escario, J. J. and JA. Molina (2004). “Modeling the optimal fiscal policy on tobacco consumption”. In: *J Policy Model* 26.1, pp. 81–93.

Feldman, Maryann, Johanna Francis, and Janet Bercovitz (2005). “Creating a cluster while building a firm: Entrepreneurs and the formation of industrial clusters”. In: *Regional studies* 39.1, pp. 129–141.

Fernandez, E., S. Gallus, A. Schiaffino, et al. (2004). “Price and consumption of tobacco in Spain over the period 1965– 2000”. In: *Eur J Cancer Prev* 13.3, pp. 207–211.

Finch, W Holmes, Jocelyn E Bolin, and Ken Kelley (2019). *Multilevel modeling using R*. Crc Press.

Fingleton, Bernard (2001). “Theoretical economic geography and spatial econometrics: dynamic perspectives”. In: *Journal of Economic Geography* 1.2, pp. 201–225.

Fossen, Frank M and Thorsten Martin (2018). “Entrepreneurial dynamics over space and time”. In: *Regional Science and Urban Economics* 70, pp. 204–214.

Fotopoulos, Georgios and David J Storey (2017). “Persistence and change in interregional differences in entrepreneurship: England and Wales, 1921–2011”. In: *Environment and Planning A: Economy and Space* 49.3, pp. 670–702.

Fritsch, Michael and Oliver Falck (2007). “New business formation by industry over space and time: a multidimensional analysis”. In: *Regional studies* 41.2, pp. 157–172.

Fritsch, Michael and Sandra Kublina (2019). “Persistence and change of regional new business formation in the national league table”. In: *Journal of Evolutionary Economics* 29.3, pp. 891–917.

Fritsch, Michael and Michael Wyrwich (2014). “The long persistence of regional levels of entrepreneurship: Germany, 1925–2005”. In: *Regional Studies* 48.6, pp. 955–973.

Fritsch, Michael and Michael Wyrwich (2016). “The effect of entrepreneurship on economic development—an empirical analysis using regional entrepreneurship culture”. In: *Journal of Economic Geography* 17.1, pp. 157–189.

Fritsch, Michael and Michael Wyrwich (2019). “Regional Entrepreneurship Culture and Growth”. In: *Regional Trajectories of Entrepreneurship, Knowledge, and Growth*. Springer, pp. 69–94.

Fuchs, A. and F. Meneses (2018). *Tobacco price elasticity and tax progressivity in Moldova*. World Bank;

Fuente, Angel de la (2019). "Estudios sobre la Economía Española-2019/19". In: *FEDEA*.

Gadea, Maria Dolores, Ana Gómez-Loscos, and Antonio Montañés (2012). "Cycles inside cycles: Spanish regional aggregation". In: *SERIEs* 3.4, pp. 423–456.

Galbraith, J. W. and M. Kaiserman (1997). "Taxation, smuggling and demand for cigarettes in Canada: evidence from time-series data". In: *J Health Econ* 16, p. 3.

Gallet, C. A. and JA. List (2003). "Cigarette demand a meta-analysis of elasticities". In: *Health Econ* 12, p. 10.

Gallus, S., A. Lugo, S. Ghislandi, et al. (2014). "Roll-your-own cigarettes in Europe: use, weight and implications for fiscal policies". In: *Eur. J. Cancer Prev* 23, p. 186.

Gallus, S., A. Schiaffino, C. La Vecchia, et al. (2006). "Price and cigarette consumption in Europe". In: *Tob Control* 15.2, pp. 114–119.

Geppert, Kurt and Andreas Stephan (2008). "Regional disparities in the European Union: Convergence and agglomeration". In: *Papers in Regional Science* 87.2, pp. 193–217.

Goel, R. K. and JW. Saunoris (2018). "Cigarette smuggling using the shadow economy or creating its own?" In: *J Econ Fin*, pp. 1–12.

Hartigan, John A and Manchek A Wong (1979). "Algorithm AS 136: A k-means clustering algorithm". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1, pp. 100–108.

HM Government of Gibraltar (2018). "Memorandum of undertanding on tobacco and other products". Memorandum. hjgf. Gibraltar.

Hoffer, A., B. R. Humphreys, and JE. Ruseski (2018). "State cigarette taxes and health expenditures: Evidence from dynamic spatial lag panel models". In: *Pap Reg Sci* 98.2, pp. 925–950.

Hong, Eunsuk et al. (2015). "Entrepreneurship across time and space: empirical evidence from Korea". In: *Small Business Economics* 44.3, pp. 705–719.

Huggins, Robert (2008). "The evolution of knowledge clusters: Progress and policy". In: *Economic development quarterly* 22.4, pp. 277–289.

Jawad, M., J. T. Lee, S. Glantz, et al. (2018). "Price elasticity of demand of non-cigarette tobacco products: a systematic review and meta-analysis". In: *Tob Control* 27.6, pp. 689–695.

Jimeno, Juan F and Samuel Bentolila (1998). "Regional unemployment persistence (Spain, 1976–1994)". In: *Labour economics* 5.1, pp. 25–51.

Jones, Kelvyn (1991). "Specifying and estimating multi-level models for geographical research". In: *Transactions of the institute of British geographers*, pp. 148–159.

Joossens, Luk and Martin Raw (2012). "From cigarette smuggling to illicit tobacco trade". In: *Tobacco Control* 21.2, pp. 230–234.

Kangasharju, Aki (2000). "Regional variations in firm formation: Panel and cross-section data evidence from Finland". In: *Papers in Regional Science* 79.4, pp. 355–373.

Kelejian, H. H. and G. Piras (2014). "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes". In: *Reg Sci Urban Econ* 46, p. 140.

Klapper, Leora et al. (2007). *Entrepreneurship and firm formation across countries*. The World Bank.

Korniotis, George M (2010). "Estimating panel models with internal and external habit formation". In: *Journal of Business & Economic Statistics* 28.1, pp. 145–158.

Kostova, D., J. Tesche, A. M. Perucic, et al. (2013). “Exploring the relationship between cigarette prices and smoking among adults: a cross-country study of low- and middle-income nations”. In: *Nicotine Tob Res* 16, S10–S15.

Kuersteiner, Guido M and Ingmar R Prucha (2018). “Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity”. In: *arXiv preprint arXiv:1802.01755*.

Lacombe, Donald J and Stuart G McIntyre (2017). “Hierarchical spatial econometric models in regional science”. In: *Regional Research Frontiers-Vol. 2*. Springer, pp. 151–167.

Lado-Sestayo, Rubén, Isabel Neira-Gómez, and Coro Chasco-Yrigoyen (2017). “Entrepreneurship at regional level: Temporary and neighborhood effects”. In: *Entrepreneurship Research Journal* 7.4.

Lampe, K. von (2009). “Transnational organised crime connecting Eastern and Western Europe: Three case studies. Crime, money and criminal mobility in Europe,” in: *Nijmegen, NL: Wolf Legal Publishers* 19.

Langford, Ian H, Graham Bentham, and Anné-Lise McDonald (1998). “Multi-level modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European Community”. In: *Statistics in medicine* 17.1, pp. 41–57.

Langford, Ian H et al. (1999). “Multilevel modelling of the geographical distributions of diseases”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 48.2, pp. 253–268.

Le Gallo, Julie and Sandy Dall’Erba (2008). “Spatial and sectoral productivity convergence between European regions, 1975–2000”. In: *Papers in Regional Science* 87.4, pp. 505–525.

Lee, Lf and J. Yu (2010). “A spatial dynamic panel data model with both time and individual fixed effects”. In: *Econ Theory* 26, p. 2.

LeSage, James P and Manfred M Fischer (2009). “The impact of knowledge capital on regional total factor productivity”. In: *Available at SSRN 1088301*.

LeSage, James P and R Kelley Pace (2010). “Spatial econometric models”. In: *Handbook of applied spatial analysis*. Springer, pp. 355–376.

Lipton, R., A. Banerjee, D. Levy, et al. (2008). “The spatial distribution of underage tobacco sales in Los Angeles”. In: *Subst Use Misuse* 43, p. 11.

Liu, H. et al. (2015). “How do smokers respond to cigarette taxes? evidence from China’s cigarette industry”. Health economics, 24(10), 1314-1330.

López-Bazo, Enrique and Elisabet Motellón (2013). “The regional distribution of unemployment: What do micro-data tell us?” In: *Papers in Regional Science* 92.2, pp. 383–405.

Luo, Bei and Terence Tai-Leung Chong (2019). “Regional differences in self-employment in China”. In: *Small Business Economics* 53.3, pp. 813–837.

MacQueen, James et al. (1967). “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA, pp. 281–297.

Martin, Ron (1997). “Regional unemployment disparities and their dynamics”. In: *Regional studies* 31.3, pp. 237–252.

Martín-Sánchez, Juan C et al. (2018). “Sex and age specific projections of smoking prevalence in Spain: a Bayesian approach”. In: *Nicotine and Tobacco Research* 20.6, pp. 725–730.

Martinez, E., R. Mejia, and EJ. Perez-Stable (2015). “An empirical analysis of cigarette demand in Argentina”. In: *Tob Control* 24, p. 1.

McKibbin, Warwick J and Roshen Fernando (2020). "The global macroeconomic impacts of COVID-19: Seven scenarios". In: *CAMA Working Paper*.

Meager, Nigel (1992). "Does unemployment lead to self-employment?" In: *Small Business Economics* 4.2, pp. 87–103.

Melguizo, Celia (2017). "An analysis of Okun's law for the Spanish provinces". In: *Review of Regional Research* 37.1, pp. 59–90.

Miller, Harvey J (2004). "Tobler's first law and spatial analysis". In: *Annals of the Association of American Geographers* 94.2, pp. 284–289.

Mittelstädt, Axel and Fabienne Cerri (2008). "Fostering entrepreneurship for innovation". In: *OECD*.

Mood, A, FA Graybill, and DC Boes (1974). "(1974), Introduction to the theory of statistics". In: *McGraw&Hill*.

Moscone, Francesco and Elisa Tosetti (2009). "A review and comparison of tests of cross-section independence in panels". In: *Journal of Economic Surveys* 23.3, pp. 528–561.

Okamuro, Hiroyuki and Nobuo Kobayashi (2006). "The Impact of Regional Factors on the Start-Up Ratio in Japan". In: *Journal of Small Business Management* 44.2, pp. 310–313.

O'reilly, G. (1999). "Gibraltar: sovereignty disputes and territorial waters". In: *IBRU Boundary and Security Bulletin* 67.

Papanicolas, I., LR. Woskie, and AK. Jha (2018). "Health care spending in the United States and other high-income countries". In: *Jama* 319.10, pp. 1024–1039.

Parente, Stephen L and Edward C Prescott (2005). "A unified theory of the evolution of international income levels". In: *Handbook of economic growth*. Vol. 1. Elsevier, pp. 1371–1416.

Parker, Simon C and Martin T Robson (2004). "Explaining international variations in self-employment: evidence from a panel of OECD countries." In: *Southern economic journal.* 71.2, pp. 287–301.

Pesaran, M Hashem (2004). "General diagnostic tests for cross section dependence in panels". In: *CEsifo*.

Pesaran, M Hashem (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure". In: *Econometrica* 74.4, pp. 967–1012.

Pesaran, M Hashem (2015). "Testing weak cross-sectional dependence in large panels". In: *Econometric reviews* 34.6-10, pp. 1089–1117.

Pijnenburg, Katharina and Konstantin A Kholodilin (2014). "Do regions with entrepreneurial neighbours perform better? A spatial econometric approach for German regions". In: *Regional studies* 48.5, pp. 866–882.

Pinilla, J. and I. Abásolo (2017). "The effect of policies regulating tobacco consumption on smoking initiation and cessation in Spain: is it equal across socioeconomic groups". In: *Tob* 15, p. 1.

Pinilla, J., M. Negrín, B. González-Lopez Valcárcel, et al. (2018). "Using a Bayesian Structural Time–Series Model to Infer the Causal Impact on Cigarette Sales of Partial and Total Bans on Public Smoking". In: *Jahrb Natl Okon Stat* 238, p. 5.

Piribauer, Philipp (2016). "Heterogeneity in spatial growth clusters". In: *Empirical Economics* 51.2, pp. 659–680.

Porter, Michael E and Scott Stern (2001). "Innovation: location matters". In: *MIT Sloan management review* 42.4, p. 28.

Ramajo, Julián and Geoffrey JD Hewings (2018). "Modelling regional productivity performance across Western Europe". In: *Regional Studies* 52.10, pp. 1372–1387.

Rana, P et al. (2016). "Impact of a comprehensive law on the prevalence of tobacco consumption in Spain: evaluation of different scenarios". In: *Public Health* 138, pp. 41–49.

Raudenbush, Stephen W and Anthony S Bryk (2002). *Hierarchical linear models: Applications and data analysis methods*. Vol. 1. sage.

Reynolds, Paul D, Brenda Miller, and Wilbur R Maki (1995). "Explaining regional variation in business births and deaths: US 1976–88". In: *Small business economics* 7.5, pp. 389–407.

Rice, Nigel and Andrew Jones (1997). "Multilevel models and health economics". In: *Health economics* 6.6, pp. 561–575.

Rodrıguez-Iglesias, G., V. Schoj, F. Chaloupka, et al. (2017). "Analysis of cigarette demand in Argentina: the impact of price changes on consumption and government revenues". In: *Salud publica MexFeb;* 59, p. 95.

Ross, Andrew G, John Adams, and Kenny Crossan (2015). "Entrepreneurship and the spatial context: A panel data study into regional determinants of small growing firms in Scotland". In: *Local Economy* 30.6, pp. 672–688.

Rupasingha, Anil and Stephan J Goetz (2013). "Self-employment and local economic performance: Evidence from US counties". In: *Papers in Regional Science* 92.1, pp. 141–161.

Saffer, H. and F. Chaloupka (1994). "Alcohol tax equalization and social costs". In: *East. Econ* 20.1, pp. 33–43.

Sala, Hector and Pedro Trivín (2014). "Labour market dynamics in Spanish regions: evaluating asymmetries in troublesome times". In: *SERIEs* 5.2-3, pp. 197–221.

Singh, Jasjit (2005). "Collaborative networks as determinants of knowledge diffusion patterns". In: *Management science* 51.5, pp. 756–770.

Smith, Tony E and James P LeSage (2004). "A Bayesian probit model with spatial dependencies". In: *Advances in econometrics* 18.18, pp. 127–160.

Spencer, Gregory M et al. (2010). "Do clusters make a difference? Defining and assessing their economic performance". In: *Regional studies* 44.6, pp. 697–715.

Sternberg, Rolf and Sander Wennekers (2005). "Determinants and effects of new business creation using global entrepreneurship monitor data". In: *Small business economics* 24.3, pp. 193–203.

Stoklosa, M., F. Goma, N. Nargis, et al. (2019). "Price, tax and tobacco product substitution in Zambia: findings from the ITC Zambia Surveys". In: *Tob ControlMay;* 28.

Stoklosa, Michal (2020). "Prices and cross-border cigarette purchases in the EU: evidence from demand modelling". In: *Tobacco control* 29.1, pp. 55–60.

Stuetzer, Michael et al. (2018). "Entrepreneurship culture, knowledge spillovers and the growth of regions". In: *Regional Studies* 52.5, pp. 608–618.

Tao, J. and J. Yu (2012). "The spatial time lag in panel data models". In: *Econ LettDec;* 117, p. 3.

Thai, Mai Thi Thanh and Ekaterina Turkina (2014). "Macro-level determinants of formal entrepreneurship versus informal entrepreneurship". In: *Journal of Business Venturing* 29.4, pp. 490–510.

Thirlwall, Anthony Philip (1966). "REGIONAL UNEMPLOYMENT AS A CYCLICAL PHENOMENON 1". In: *Scottish Journal of Political Economy* 13.2, pp. 205–219.

Townsend, J. (1988). *Price tax and smoking in Europe*. World Health Organisation.

Vanoutrive, Thomas and Angela Parenti (2009). "On proximity and hierarchy: exploring and modelling space using multilevel modelling and spatial econometrics". In: *49th European congress of the Regional Science Association International (ERSA Congress 2009)*. European Regional Science Association (ERSA).

Vega, Solmaria Halleck and J Paul Elhorst (2014). "Modelling regional labour market dynamics in space and time". In: *Papers in Regional Science* 93.4, pp. 819–841.

Vega, Solmaria Halleck and J Paul Elhorst (2016). "A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors". In: *Regional Science and Urban Economics* 60, pp. 85–95.

Wennekers, Sander et al. (2005). "Nascent entrepreneurship and the level of economic development". In: *Small business economics* 24.3, pp. 293–309.

West, Robert et al. (2008). "Why combating tobacco smuggling is a priority". In: *Bmj* 337, a1933.

Wilson, L. M., E. Avila Tang, G. Chander, et al. (2012). "Impact of tobacco control interventions on smoking initiation, cessation, and prevalence: a systematic review". In: *J Environ Public Health* 2012.

Wolman, Harold and Diana Hincapie (2015). "Clusters and cluster-based development policy". In: *Economic Development Quarterly* 29.2, pp. 135–149.

Yeh, C. Y., C. Schafferer, J. M. Lee, et al. (2017). "The effects of a rise in cigarette price on cigarette consumption, tobacco taxation revenues, and of smoking-related deaths in 28 EU countries applying threshold regression modelling". In: *BMC Public HealthSep;* 17, p. 1.

Yu, D., N. Peterson, M. Sheffer, et al. (2010). "Tobacco outlet density and demographics: analysing the relationships with a spatial regression approach". In: *Public health* 124, p. 7.

Zhu, Xiwei et al. (2019). "Entrepreneurship and industrial clusters: evidence from China industrial census". In: *Small Business Economics* 52.3, pp. 595–616.