

Artigo submetido a 26 de Janeiro 2022; versão final aceite a 16 de Agosto de 2022

Paper submitted on January 26, 2021; final version accepted on August 16, 2022

DOI: <https://doi.org/10.59072/rper.vi67.349>

# National or Regional Recruitment: “Market Area” of University Cities<sup>1</sup>

## Atração Nacional ou Regional de Estudantes: “Áreas de Mercado” de Cidades Universitárias

**Cássio Rolim**

*cassio.rolim@gmail.com*

Departamento de Economia, Universidade Federal do Paraná, Curitiba, Brasil

**Conceição Rego**

*mcpr@uevora.pt*

Departamento de Economia & CEFAGE-U.É., Universidade de Évora, Évora, Portugal

**Andreia Dionísio**

*andreia@uevora.pt*

Departamento de Gestão & CEFAGE-U.É., Universidade de Évora, Évora, Portugal

**Elsa Ferreira**

*elsa.ferreira@estsetubal.ips.pt*

Escola Superior de Tecnologia, Instituto Politécnico de Setúbal, Setúbal, Portugal

### Abstract

The depopulation of eastern Portugal and concentration on the coast has been a longstanding process. Since 1973 new HEIs has been disseminated around the country, and nowadays, every Portuguese region holds a set of HEIs. However, those closer to the coast tend to have a higher demand than those more distant. It is not clear if each one has a kind of restricted market area, such as a spatial monopoly in its neighbouring area or a national recruitment for some of them.

This paper is the first exercise of determination of these areas for the Portuguese public network of HEIs. We use a database from the Ministry of Education (DGEEC-Ministry of Education) with information of the pair, place of residence (municipality level) and respective HEI, for enrolled students. Through techniques of Exploratory Spatial Data Analysis (ESDA), we will look for clusters of municipalities that could be market areas for each “university city” considered. We also use fractional regression models that confirm the positive correlation between the chosen university city and the distance to the family home and belonging to the market area. Despite some overlaps, the market areas of the university cities look to be well defined.

**Keywords:** *higher education; regional development; spatial area of universities; ESDA; fractional regression.*

---

<sup>1</sup> Acknowledgments:

- Conceição Rego and Andreia Dionísio, are pleased to acknowledge financial support from Fundação para a Ciência e a Tecnologia (grant UIDB/04007/2020).
- The authors are grateful for the comments and suggestions of Professor Pedro Amaral (Universidade Federal de Minas Gerais) as well as the anonymous reviewers who made the assessment of the paper in the framework of the submission to this journal. However, any errors and omissions are the authors' responsibility.

*JEL Codes:* C25, I23, R11

## Resumo

O despovoamento do interior de Portugal e a concentração no litoral é um fato. Desde 1973 as novas IES têm sido disseminadas por todo o país e, atualmente, todas as regiões portuguesas possuem um conjunto de instituições de ensino superior (IES). No entanto, as mais próximas do litoral tendem a ter uma procura maior do que as mais distantes. Não está claro se cada uma atrai os seus estudantes com um padrão de monopólio espacial regional ou se há recrutamento nacional para outras.

Este trabalho é o primeiro exercício de determinação destas áreas para a rede pública portuguesa de IES. Utilizamos uma base de dados do Ministério da Educação (DGEEC-Ministério da Educação) com informação do par, local de residência (nível de concelho) e respetiva IES, para os alunos matriculados. Através de técnicas de Análise Exploratória de Dados Espaciais (ESDA), procuramos identificar *clusters* de municípios que possam ser áreas de mercado para cada “cidade universitária” considerada. Também utilizamos modelos de regressão fracionária que confirmam a correlação positiva entre a cidade universitária escolhida e a distância até a residência da família e a pertença à área de mercado. Apesar de algumas sobreposições, as áreas de mercado das cidades universitárias parecem estar bem definidas.

*Palavras-chave:* ensino superior; desenvolvimento regional; áreas de mercado de universidades; ESDA; regressão fracionária.

*Código JEL:* C25, I23, R11

## 1. INTRODUCTION

Higher education institutions’ (HEIs) contribution to improving local / regional qualifications can lie firstly in attracting students, and secondly, in retaining graduates in the local / regional labour market. In Portugal, HEIs are located throughout the country but the number and dimension of HEIs in the different regions vary. The highest concentration of places available coincides with the most densely populated regions.

The geographical origin of higher education students is an important piece of information for institutional and national decision-makers to know their students better and how distant their family home is from the city where they study. This information is particularly important in a context where many HEIs are faced with imbalances between supply and demand and where the sustainability of the higher education (HE) network is fundamental to promote the cohesion of the country. Most of the literature on HE demand considers distance between HEIs and the family home as an important point in choosing an HEI. In general, findings show that distance has a negative relationship with the demand for a particular HEI. However, it is not clear if institutions located in *university cities* (due to data confidentiality, we will replace HEIs by university cities) promote a kind of restricted market area, such as a spatial monopoly, in their area, or if they recruit nationally. The question is crucial because in Portugal access to public HE is through a national system that relates all applicants to all places available (all courses in all HEIs).

This paper aims to determine the “market areas” of some university cities with public HEIs. Taking into account the characteristics of HE supply, the fact that the places available in HEIs are sufficient for the national demand, and where HE students’ families live, it is important to know where Portuguese HEIs recruit their students, or alternatively, where each university city’s students come from. This knowledge is key for HEI leaders and for public policy decision-making, in order to establish the HE supply and define the rules for access to HE. The starting point is an attempt to delimit the market areas of each university city, considering the spatial pattern of recruitment, be it national or regional. The primary hypothesis is that the pattern of recruitment is regional, but that some universities cities, particularly those with more institutions and greater prestige, may have a national pattern.

Through Exploratory Spatial Data Analysis (ESDA) techniques, the paper will look for clusters of municipalities that could be “market areas” for each university city considered. As far as we

know, this study is the first to use municipalities as territorial units (other analyses of the effects of distance in HE demand, in the Portuguese context, use the District, Nuts II or Nuts III as territorial units). In addition, a fractional regression model will test some hypotheses derived from the ESDA exercise.

After this brief introduction, the literature review discusses the effect of distance between home and the chosen HEI. This is followed by some information about the Portuguese HE system and the characteristics of access to this education system. Methods and Data will present the main methodological options. Then, the results will be discussed, highlighting some critical points to be taken into consideration by decision-makers. The paper finishes with some concluding remarks.

## 2. LITERATURE REVIEW

The factors that determine and influence the decision to access HE are very diverse (Vieira & Vieira, 2013; Fonseca & Encarnação, 2012; Oliveira et al., 2012; Sá et al., 2011; Rego & Caleiro, 2004). There are characteristics of a contextual nature: for example, economic development levels and income growth, demographic factors, the number of compulsory schooling years and successful completion of secondary education, the number of higher education institutions in the country and public spending on higher education, as a percentage of GDP. On the other hand, others factors related to individual and family reasons linked to the financial capacity of the student / family, the candidate's previous schooling, the student's family background - in terms of income and level of education -, the expectation of future income associated with an HE diploma, the course they wish to attend, the number of years required to complete the degree, an HEI's proximity to the place of residence of the student / family, emotional and other social and psychological factors, such as the desire to leave the family home to study. Furthermore, reasons linked with HEI characteristics: institutional reputation, fame, the values that HEIs transmit, the HEI's dimension, the diversity of fields of study, the quality of education and research, the communication policy, the existence of financial and other support for students.

However, this research concentrates on the relevance of an HEI's distance from the student's family home, through the concept of “HEI market area”. The literature shows distance to be one of the determinants involved in a decision to leave the family home and move to another city to study. The importance of geographical proximity between the family home and the HEI is shown in several previous studies (among others, Sá et al., 2004; Sá & Tavares, 2018; Azzone & Soncin, 2020; Briggs, 2006; Simões & Soares, 2010; Spiess & Wrohlich, 2010). Many studies show a negative relationship between distance and the choice of HEI. From this point of view, the family's socioeconomic condition restricts students' choices and accessibility influences decision-making in relation to the choice of HEI (Sá et al., 2011). Attending higher education while living at home avoids significant costs with relocation (Lourenço et al., 2020; Spiess & Wrohlich, 2010). The so-called “transaction cost” can include costs with accommodation and transport that affect mainly the decisions of students from lower income families. Financial costs are not the only factor affecting the choice of HEI. In fact, most students prefer the emotional security of remaining close to their network of family and friends (Lourenço et al., 2020; Spiess & Wrohlich, 2010). As well as these reasons, Spiess & Wrohlich (2010) indicate the “neighbourhood effect” (near a university environment, young people may grow up looking at university education as a natural goal). This effect represents inequality in relation to young people living in remote communities, with little or no higher education provision (Lourenço et al., 2020). The desire to attend higher education while living at home is also seen in families belonging to higher income groups and with higher qualifications (Sá et al., 2011).

Although most young people prefer to study close to (or at) home, others move away. In this case, students appreciate leisure time (Sá et al., 2011), and move to larger cities and HEIs with a wider range of courses on offer (Fonseca et al, 2020; Van Bouwel & Veugelers, (2009)). Lourenço & Sá (2019) show that “outgoing flows are lower the greater the local supply of higher education and the larger the young population”. In general, moving from smaller cities, with or without HEIs, is more common among students with higher grades, who prefer to attend higher education in a big city with more to offer and greater prestige, as well as wider labour markets. This behaviour, over

time, tends to worsen territorial inequalities, impoverishing less densely occupied areas.

One way of studying geographical proximity is through the concept of market areas. Market areas are, theoretically, a structure of spatial competition between suppliers of a homogeneous product, taking into consideration the cost of transporting the product from the supplier's location (in a limited number of places) to customers, who are scattered all over the territory (Maier, 2009; Simões Lopes, 1984). In the market area approach, producers appear as monopolists, exerting more influence the closer they are to consumers and suppliers. This approach has been applied to identify the areas that higher education (HE) students come from (analyzed as university city market areas), despite HE not being a homogeneous good, and whether university cities have monopoly conditions. There are clearly relevant costs associated with the distance between the family home and the chosen HEI. The assumption is that when students move away from home, they will go to the nearest supplier of higher education.

Exploring the existence of market areas for Austrian HEIs, in the field of economics and management studies, Maier (2009) concluded that spatial monopolies occur: "Most universities offering business education have developed a sizeable area around their location, which they dominate" (Maier, 2009: 265). This means that in these areas the respective HEI is the main supplier, and at the same time, students on these courses come mainly from this region. Also in this connection, Rolim & Garcia (2012) concluded that the market area of the *Universidade Federal do Paraná*, Brazil (UFPR) can be considered primarily local, although it receives students from remote regions of Brazil. "The majority of applicants and successful candidates come from the Metropolitan Region of Curitiba [the capital of Paraná region, where this HEI is located] and 90% of them come from the state of Paraná" (Rolim & Garcia, 2012: 43). The authors also show that distance can be considered one of the main explanatory variables for the spatial influence of UFPR.

## 2.1 Portuguese Higher Education: System and Access

From the 1970s, Portuguese higher education underwent a revolution characterized by institutional diversification and expansion, with the creation of new HEIs and regionalization of the system with the spread of institutions throughout the country. This promoted the massification and democratization of the system and allowed students from different social or regional origins to access higher education. In Portugal, higher education is a binary system that integrates universities and polytechnic institutes, both public and private. The public HEI network covers the whole country, with its intensity depending on population density. "Regional distribution of higher education in Portugal overlaps, essentially, the urban national network" (Fonseca & Encarnação, 2012: 9). The population and economic activity are concentrated in the metropolitan regions (Rego et al., 2021). Private institutions are also located predominantly in the main urban areas. In the academic year 2019/20, 396909 students were enrolled in Portuguese HEIs, distributed among the 284 establishments throughout the country<sup>2</sup> (26,9% of students and 30% of establishments in the Lisbon metropolitan area). In Portugal, the higher education supply has stabilized in recent years (Fonseca & Encarnação, 2012) and it is generally enough to meet the demand; overall, the number of places available is adequate for the number of applications. However, the relationship between demand and supply in the different institutions is unbalanced. The largest institutions, with a more diversified offer and located in the largest cities, receive the vast majority of applications for the first year of HE. To increase demand for HE in the smaller HEIs in inland areas, the Portuguese government has taken some measures to promote the balance and sustainability of the HE network. One of these is the "+Superior scholarship", which intended to promote student mobility, in order to increase applications to inland HEIs experiencing a lower demand, and in this programme, only low-income students are eligible. Another recent public policy measure adopted by the Portuguese government is a 5% reduction in vacancies in public institutions in the two biggest metropolitan areas (Lisbon and Porto), moving these vacancies to inland HEIs, where there is lower demand.

Access to higher education depends on a mechanism of *numerus clausus* that determines the supply in this educational system, in all institutions. The maximum number of places allowed in each course / institution pair is approved by the Government. Thus, access to the first year of a

<sup>2</sup> Data from DGEEC (<http://estatisticas-educacao.dgeec.mec.pt/indicadores/index5.asp>; accessed in February, 2021.)

degree (or integrated master) course is through a centralized process of assigning candidates to places, according to their preferences and taking into account their secondary school classifications (the level below higher education, which is compulsory until 18 years old). This is carried out nationally, and includes all the students who, in a given year, intend to enter higher education and all the places available in public higher education institutions, whether university or polytechnic.

In the decision to apply to higher education, students use the classification obtained in secondary school and the information available from the previous year's access conditions. They choose the course / institution pair with the greatest guarantee of being successful. Students show great pragmatism in their decision-making, opting in many cases for an application they know is likely to be successful, despite not being what they would choose ideally (Fonseca & Encarnaç o, 2012; Rego & Caleiro, 2004).

### 3. METHODS AND DATA

#### 3.1 Methods

##### 3.1.1 Exploratory Spatial Data Analysis (ESDA)

Given the poor representation of the private sector, this study will analyse only the locations and respective HEI market areas of the public sector. To do this, we will use data from the *Registo de Alunos Inscritos e Diplomados do Ensino Superior*, RAIDES (Register of Enrolled Students and Higher Education Graduates). This is an annual national survey, in which all higher education institutions must participate. It focuses on the universe of resident students enrolled in higher education institutions, thus not including those in international mobility.

Through this database, it is possible to establish the link between the student's home address and the location of the HEI chosen. Although the final interest is to detail the destination of the students by HEI, the published data are by municipalities of destination, and so we must work with these aggregate data. These receiving municipalities will be called *university cities*. This information, once geo-referenced, will be initially worked on using *Exploratory Spatial Data Analysis* (ESDA) techniques.

According to Anselin (1994), the techniques used in ESDA are intended to describe spatial distributions, determine patterns of spatial association, suggest different spatial regimes or other forms of spatial (non-stationary) instability, and to identify atypical observations (outliers). ESDA is a preliminary step in the elaboration of spatial econometric models. To a great extent, spatial econometrics was developed considering the importance of spatial contiguity between the territorial units under analysis. This is the environment conducive to space dependence. Specifically, rather than considering that the events occurring in municipality *i* are independent of those occurring in municipality *j*, as in traditional econometrics, here it is considered that there is a dependency and that it is a function of these municipalities' degree of proximity. Spatial dependence, also called spatial autocorrelation, indicates that the value of a variable in territory *i* depends on the value of that variable in territory *j*.

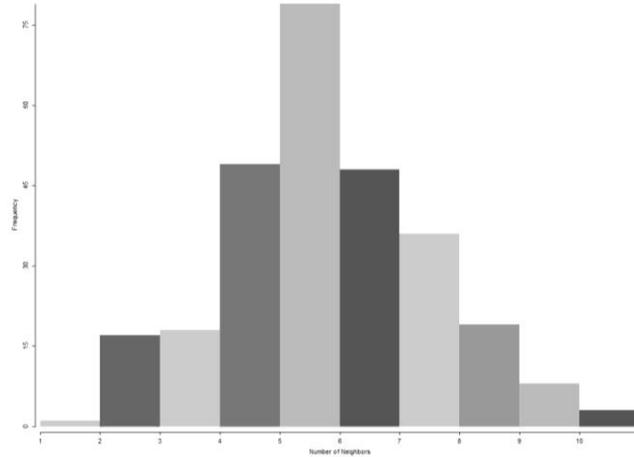
The spatial heterogeneity (non-stationarity) is due to the specificity of each place, i.e., the structural difference between two places. For example, one poor and one prosperous municipality, or one in the plain and the other in the mountains, leads to spatial heterogeneity. Thus, there is the possibility of specifying a behaviour function that will not be valid for the entire sample, as well as having parameters that vary throughout the sample leading to the occurrence of spatial heteroscedasticity.

An important step is the construction of a spatial weighting matrix (W), seeking to reflect the particular type of interaction between the regions studied. The W matrix is fundamental in creating spatial autocorrelation statistics as well as creating explicit spatial variables, such as spatially-lagged variables. It is a nxn matrix in which each element  $w_{ij}$  is the spatial weight. When regions *i* and *j* are neighbours, the weight is a positive number other than zero. When they are not neighbours, the weight is 0. By convention, the weight of the region itself,  $w_{ii}$  is zero, implying that the elements of the main diagonal of the matrix are zeros. Several types of matrices are possible (Almeida, 2012). The simplest is a binary matrix with values of 1 and 0. Each spatial unit is represented by a line *i*

and the neighbours by columns  $j$ , with  $j \neq i$ . The construction of such a matrix is not trivial when a large number of territorial units are considered (278 in mainland Portugal). It will be constructed using specialised software, such as GeoDa.

Once the neighbourhood structure is defined with the spatial weights matrix  $W$ , the spatial lag variable is the weighted sum of the values observed in the neighbouring localities. In our exercise, a queen contiguity weights matrix (in analogy to the movements of the queen in the game of chess) is used. Figure 1 shows the neighbourhood structure created among the 278 municipalities. On average, each municipality has 5.3 neighbours and there is no isolated municipality (without neighbours).

**Figure 1: Spatial Weights Matrix Histogram – Municipalities of Mainland Portugal**



Source: Authors' elaboration

The Moran index measures the spatial autocorrelation. This is a cross product between a variable and its spatial lag, expressed in a centred way (deviation from the mean). In region  $i$ , the variable  $z_i = (x_i - \bar{x})$  where  $\bar{x}$  is the mean of the variable  $x$ .

In this way, Moran's  $I$  can be expressed as,

$$I = \frac{\sum_i \sum_j w_{ij} z_i z_j / S_0}{\sum_i z_i^2 / n} = \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \tag{1}$$

$$Z_I = \frac{I - E(I)}{\sqrt{V(I)}} \tag{2}$$

$$E(I) = \frac{-1}{n-1} \tag{3}$$

$$V(I) = E(I^2) - E(I)^2 \tag{4}$$

In the case of a normalised spatial weights matrix, the range of  $I$  will be:

**Table 1: Moran's I**

Indicator	Range	Correlation	Rule of Thumb
<b>I</b>	$> \frac{-1}{n-1}$	Positive Spatial Correlation	0.25 – 0.50 weak 0.50 – 0.70 moderate 0.70 – 0.90 strong 0.90 – 1.0 marked degree
	$= \frac{-1}{n-1}$	No Spatial Correlation	Asymptotically Zero
	$< \frac{-1}{n-1}$	Negative Spatial Correlation	Rare Occurrence

Source: Getis (1995), Griffith (2009)

In spite of its similarity to the Pearson's correlation coefficient, Moran's  $I$  range depends on complex calculation of eigenvalues (Griffith, 1996; Jong, Sprenger, Veen, 1984). In short, the

expected value of Moran’s I depends on the kind of the weights matrix and the number of observations .  $I = \frac{-1}{n-1}$  indicates a random spatial pattern that means absence of autocorrelation;  $I > \frac{-1}{n-1}$  values will indicate a positive correlation with neighbours; and  $I < \frac{-1}{n-1}$  values, a negative (inverse) correlation. Positive correlations indicate that similar values appear close in space forming clusters; negative correlations indicate that values between neighbours are different; absence of autocorrelation means randomness in the spatial distribution pattern of that variable. The value of I can also show how strong the correlation is.

The statistical significance of Moran's I can be verified in two ways.

The first is to consider that the standardized variable Z(I) has a normal distribution with zero mean and unit variance. In this way,

$$Z(I) = \frac{I - E(I)}{DP(I)} \tag{5}$$

Where E(I) is the expected value of I, and DP(I) is the theoretical standard deviation of I.

The second is called a random permutation. The observed values are randomly permuted; then the I statistic is calculated for each of these permutations, obtaining an empirical reference distribution; the calculated statistic I is compared with the empirical reference distribution, checking whether it is inside or outside the critical rejection region (Almeida, 2012).

The null hypothesis is that the spatial distribution of the variable in question is random against the alternative hypothesis that it is spatially concentrated. The table below summarizes the interpretation of the results.

**Table 2: Statistical Significance of Moran’s I**

The p-value is <i>not</i> statistically significant	Accept the null hypothesis. The spatial distribution is the result of random spatial processes
The p-value <i>is</i> statistically significant, and the z-score is positive.	Reject the null hypothesis. The spatial distribution of high values and/or low values in the dataset is more spatially clustered than expected if underlying spatial processes were random.
The p-value <i>is</i> statistically significant, and the z-score is negative.	Reject the null hypothesis. The spatial distribution of high and low values in the dataset is more spatially dispersed than expected if underlying spatial processes were random. A dispersed spatial pattern often reflects some type of competitive process—a feature with a high value repels other features with high values; similarly, a feature with a low value repels other features with low values.

Source: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm>

The term LISA (Local Indicators of Spatial Association) refers to a set of indicators of which the most famous are the local Moran and the local Geary. However, for simplicity it will refer here to the Local Moran’s I. Spatial heterogeneity is noticed with the help of the Moran local indicator of spatial association. The indicators of local spatial autocorrelation will allow decomposition of the global Moran’s I (considering all observations). They can reveal both local spatial autocorrelation and spatial heterogeneity by the presence of clusters (Anselin, 1995).

The LISA for Moran's I will be:

$$I_i = z_i \sum_j w_{ij} z_j \tag{6}$$

As in the case of global Moran,  $z_i$  and  $z_j$  are variables centred relative to their respective means,  $z_i$  is the variable under observation in the locality  $i$  and  $z_j$  is the value of the variable in the localities neighbouring  $i$ . The spatial weights matrix  $W_{ij}$  is also standardized in the line and  $w_{ii} = 0$  by convention. It is important to note that  $I_i$  is calculated only for the neighbours of  $i$  as defined by the spatial weights’ matrix used.

The sum of the local Moran indicator will be:

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j \tag{7}$$

The statistically significant values calculated are presented in a map with four types of clusters: HH (high values in the core and in the neighbourhood); LL (low values in the core and in the neighbourhood); HL (high in the core and low in the neighbourhood); LH (low in the core and high in the neighbourhood). The presence of different clusters in the same dataset could mean different spatial standards, which indicates spatial heterogeneity.

This procedure, repeated for each university city, will allow delimitation of the respective market areas. The HH clusters will be the market areas.

It is important to point out that through ESDA, spatial dependence can be verified through indicators such as Moran's I. On the other hand, a priori, there is no indicator of spatial heterogeneity in the ESDA. However, its existence can be perceived through local spatial autocorrelation indicators such as the local Moran's I.

The perception of these spatial effects – spatial dependence and heterogeneity – will be of fundamental importance for the econometric modelling phase. It indicates the need to use specific instruments to avoid the problems they entail.

### 3.1.2 The fractional regression model

The second part of the paper tests the adjustment of some outputs from the ESDA analysis to an econometric model.

Maier (2009) used a logit model to find the determinants of the market areas of Austrian universities specializing in business education. To a certain extent, we followed a similar type of modelling although our dependent variable is the proportion of Portuguese students enrolled in public higher education institutions by municipality of higher education institution and municipality of students' permanent residence. This means that we do not have a binary variable, but a variable that takes on all possible values in the unit interval.

The use of fractional response variables is quite common in economics and the bounded nature of such variables, such as the possibility of observing values at the boundaries, raises important functional form and inference issues. Papke & Wooldridge (1996) introduced a set of econometric methods for fractional response variables, which were developed in several studies, namely Ramalho, Ramalho & Henriques (2010), Ramalho, Ramalho & Coelho (2018) and Ramalho, Ramalho & Murteira (2011).

In order to obtain robust and unbiased estimations for the determinants of Portuguese higher education institutions' market areas, we use cross-sectional fractional models. The standard fractional regression model used in the cross-sectional context can be defined by:

$$E(y|x) = G(x_i\theta), \quad (8)$$

where  $\theta$  is the vector of parameters of interest and  $G(\cdot)$  is a nonlinear function based on the unit interval. This  $G(\cdot)$  function may assume different forms: (i) logit; (ii) probit; (iii) loglog; (iv) cloglog; and (v) cauchit. The results presented will refer to the partial effect of the models, selected according to the RESET test.

Given the nature of our data, it is probable to observe a substantial proportion of limit values in the fractional data, more precisely 0's. The value 1 is almost impossible to be found, given the fact that the response variable is a proportion of the whole. According to Papke & Wooldridge (1996) and Ramalho, Ramalho & Henriques (2010), it is still possible to use the simple version of the fractional regression models, but this may not be the best option when the number of corner observations is large. In this context, we also estimate the two-part fractional regression model following Ramalho, Ramalho & Murteira (2011). This model uses a binary regression model to explain the probability of a specific corner value (0 or 1) and then uses a conditional mean model (already described before) to explain the remaining fractional values.

The two-part fractional regression model may be defined by:

$$E(y|x) = Pr(y_i > 0|x_{ib}).E(y_i|x_{if}, y_i > 0) = G(x_{ib}\theta_b).G_f(x_{if}\theta_f), \quad (9)$$

where  $x_{ib}$  and  $x_{if}$  are the explanatory variables used in the binary and in the fractional parts of the model,  $\theta_b$  and  $\theta_f$  are vectors of variables coefficients and  $G_b(\cdot)$  and  $G_f(\cdot)$  are specified in the same way as  $G(\cdot)$ , since both functions are bounded between 0 and 1.

The correct specification of the conditional mean of  $y_i$  is an important assumption that should be tested by: (i) RESET-type test (proposed by Papke & Wooldridge, 1996); and (ii) the P test for general nonnested hypothesis adapted for the fractional modelling by Ramalho, Ramalho & Murteira (2011).

The RESET test is based on standard approximation results for polynomials, assuming as null:  $H_0: \phi = 0$ , being  $\phi$  a vector composed of the sum of the polynomials inserted. This test may be used to evaluate the functional form in the separate components of the two-part models, although the information about alternative specifications is not provided. Therefore, in this research work we will use the P test. The P test can also test the specification of the models and test the full specification of the two-part models against one-part models and other two-part models (or vice-versa).

As already defined, the response variable,  $y_{ij}$ , is the proportion of higher education students from city  $i$  who opted for city  $j$ . The explanatory variables are the following:

$D_{ij}$  = Distance between city  $i$  and city  $j$

DHH = Dummy for market area municipalities (High-High clusters)

DLL = Dummy for municipalities of Low-Low clusters

We expect to be able to find the determinants of the market areas of Portuguese university cities. Despite the predominance of the universities of Lisbon and Porto in the national recruitment, there is a territorial delimitation between them, as well as sub-regional markets for the others.

### 3.2 Data

The lines of Table 3 contain the destination of the students from each municipality of mainland Portugal (278) and in the columns the municipalities to which they go. In this exploratory exercise, the columns only show six selected municipalities with public universities (Braga, Porto, Vila Real, Covilhã, Lisbon and Évora) and two with polytechnic institutes (Bragança and Guarda). In fact, this table could have 46 columns. The data refer to the total number of undergraduate and Integrated Masters students for the 2017/2018 academic year<sup>3</sup>.

**Table 3: Basic data matrix structure**

Municipality of ↓ Residence	Municipality of HEI							
	Braga	Bragança	Covilhã	Évora	Guarda	Lisbon	Porto	Vila Real
Abrantes	$a_{11}$	$a_{12}$	$a_{13}$	$a_{14}$	$a_{15}$	$a_{16}$	$a_{17}$	$a_{18}$
Agueda	$a_{21}$	--	--	--	--	--	--	$a_{28}$
-----	--	--	--	--	$a_{ij}$	--	--	--
Viseu	$a_{2771}$	$a_{2772}$	--	--	--	--	--	$a_{2778}$
Vouzela	$a_{2781}$	$a_{2782}$	--	--	--	--	--	$a_{2788}$

Source: Own elaboration.

Formalizing,

$$\sum_{i=1}^n a_{ij} = \text{Distribution of students of municipality } i \text{ between university cities } j \text{ (1.5)}$$

Note that in this specific exercise, the sum in the line will not give the total of the students coming from  $i$  because only eight university cities are present.

Table 3 gives the absolute values for each municipality, but to perform spatial analysis, it is necessary to relativise these values. An auxiliary indicator could do this. This indicator,  $r_{ij}$ , will show the proportion of students from each municipality of mainland Portugal in the 8 university cities considered. Thus, each  $a_{ij}$  will be divided by the total students of the municipality  $i$  who went to the 46 university cities.

<sup>3</sup>These data are available in: [https://www.dgeec.mec.pt/np4/EstatVagasInsc/].

$$r_{ij} = \frac{a_{ij}}{\sum_{i=1}^k a_{ij}} \tag{1.6}$$

$r_{ij}$  = proportion of students from municipality  $i$  to  $j$ -th university city.

**Table 4 - Descriptive statistics of variables**

		Braga	Bragança	Covilhã	Évora	Guarda	Lisboa	Porto	Vila Real
$r_{ij}$ (%)	Mean	0.0380	0.0332	0.0395	0.0451	0.0174	0.2027	0.0803	0.0344
	St.Dev	0.0947	0.0857	0.0639	0.0900	0.0394	0.1924	0.1297	0.0760
	Max.	0.6186	0.7128	0.5815	0.5599	0.2632	0.8390	0.6946	0.5331
	Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0215	0.0000	0.0000
Distance (Km)	Mean	256.373	330.452	233.721	274.037	233.379	247.419	222.551	254.939
	St.Dev	158.745	170.202	108.702	146.147	123.279	120.208	143.980	168.128
	Max.	650	774.439	567	561	602.753	506	602	669
	Min	0	0	0	0	0	0	0	0

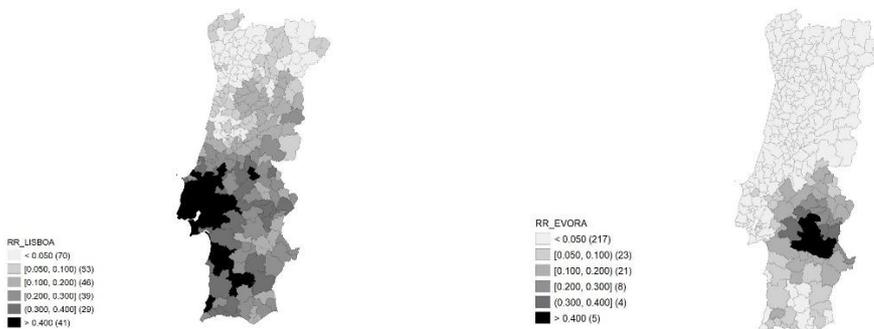
Source: Own elaboration.

## 4. RESULTS

### 4.1 The First Glance

Initially, the proportion of students from each municipality between the university cities is considered, the indicator  $r_{ij}$ . The scales are the same for all cities. The municipalities  $i$  that contribute more than 40% of their students to university city  $j$  compound the market area of  $j$  (Maier, 2009). Maier considers the municipalities that contribute more than 50% of their HE students, but we made a small adjustment for the Portuguese case<sup>4</sup>. Lisbon and Évora, Porto and Braga, Covilhã and Vila Real, Bragança and Guarda were put together. In the first case, the market area of the largest university city, Lisbon, is compared with one of the smaller ones, Évora. However, they are relatively close and, indeed, could potentially compete in the same market area. As well as being a metropolis, Lisbon is home to a wide range of HEIs (more than a dozen public HEIs between universities and polytechnic institutes). In Évora there is only one university. Porto and Braga have several public HEIs, universities and polytechnic institutes, but the most important aspect in this pair of cities is being very close and allowing analysis of a possible market-driven dispute. Covilhã and Vila Real have one public university each and are further away from the coast. Finally, Bragança and Guarda, each having only one HEI (polytechnic institutes) are located in remote areas of the country.

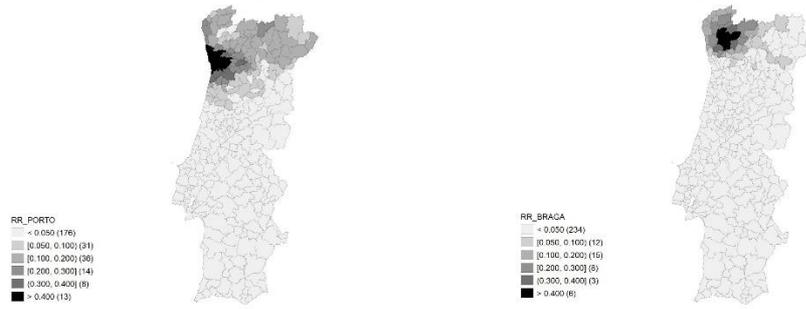
**Figure 2: Market Area of Lisbon and Évora**



Source: Elaboration of the Authors

<sup>4</sup> Due to differences in the structure of the urban network between Portugal and Austria, we chose to use the proportion of 40%.

**Figure 3: Market Area of Porto and Braga**

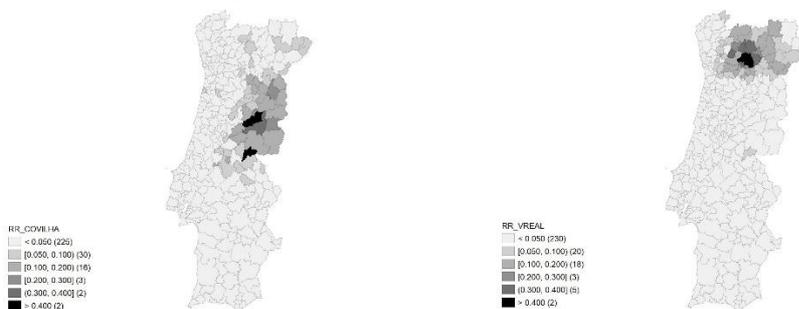


Source: Own elaboration.

In Maier's market area concept (Maier, 2009), 41 municipalities comprise the market area of Lisbon and only 5 correspond to Évora. Porto's market area is composed of 13 municipalities and Braga's of 6. It should be noted that the other university cities' market areas are formed of a small number of municipalities. In the case of Guarda, no municipality reached the value that characterises a market area (more than 40% of university students). Even so, despite some overlap, the market areas remain well delimited. The extent of Lisbon's market area is not surprising since the city receives about 30% of Portuguese university students while the others considered here receive, respectively, Porto 13.5%, Braga 6% and Évora 2%. The large contingent of students received by Lisbon, nevertheless, has a clear territorial connotation, concentrating in the vicinity of the capital and advancing towards the south more than to the east where Évora's market area begins. The extent of Lisbon's market area, however, does not refute what has been shown so far: the existence of reasonably delimited market areas and the importance of the factor of the distance from students' home.

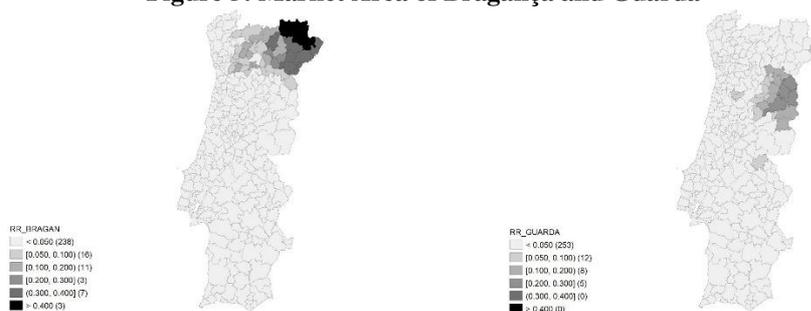
Another finding of this visual analysis is that few municipalities contribute to the influx into university cities. Except for Lisbon and Porto, the number of municipalities sending less than 5% of their students to the university cities in question is more than two hundred. Even for Lisbon and Porto, these numbers are high, respectively 70 and 176 municipalities. This is another element indicating that market areas are concentrated around these cities and that, consequently, distance is a significant factor in choosing the place of study. Figures 2 to 5 illustrate what has been said.

**Figure 4: Market Area of Covilhã and Vila Real**



Source: Own elaboration.

**Figure 5: Market Area of Bragança and Guarda**



Source: Own elaboration.

## 4.2 Analytic - dependence and spatial heterogeneity

The major difference between visual and analytical analysis is that the results of the latter can have their statistical significance tested. In this section, the instruments that can capture the phenomena of dependence (correlation) and spatial heterogeneity are used. As seen in the preceding sections, the most commonly used for spatial correlation is the Moran's I statistic and for the perception of spatial heterogeneity we will use the Local Moran's I.

### a) Global Moran's I - Spatial Dependence.

The Moran scatter plot has the spatially lagged variable on the y-axis and the original variable (variable analysed) on the x-axis. The slope of the linear fit to the scatter plot equals Moran's I (Anselin, 1996). The scatter plot is decomposed into four quadrants. The upper-right quadrant and the lower-left quadrant correspond to *positive* spatial autocorrelation, referred to respectively as *high-high* and *low-low* spatial autocorrelation. In contrast, the lower-right and upper-left quadrant correspond to *negative* spatial autocorrelation, referred to respectively as *high-low* and *low-high* spatial autocorrelation (see Table 5).

Statistical significance can be obtained either by an approximation to the normal curve, since Moran's I has an asymptotically normal distribution, or by a random process. The latter, more common in empirical exercises, is based on random permutations (Anselin, 2014).

**Table 5: Quadrants of the Moran scatter plot (Variable Analysed, x; Spatial Lag, y)**

		Variable Analysed (X)	
		Low	High
Spatial Lag (Y)	High	2nd Quadrant (LH)	1st Quadrant (HH)
	Low	3rd Quadrant (LL)	4th Quadrant (HL)

Source: Own elaboration.

Table 6 presents the result of calculating the global Moran's I for the eight university cities. The first finding is the extremely high degree of correlation of the variable in the municipality *i* (variable analysed) with its value in the neighbourhoods (spatial lag). All Moran's I values were highly significant. Moreover, in the eight cities, most of the observations, as shown in the Moran scatter plots (Figures 6 to 9), are in the quadrants with positive correlation (first and third), regardless of whether they are associations of the High-High or Low-Low type, according to the standards in Table 5. The case of Covilhã is irregular. The value of Moran's I is the lowest among the cities analysed, which is consistent with the fact that its market area is small.

## National or Regional Recruitment: “Market Area” of University Cities

**Table 6: Moran’s I Parameters (999 permutations) for  $r_{ij}$  in the university cities**

	Moran’s I	E(I)	Means	S.D.	Z Value	Pseudo p-value
Lisbon	0.9338	-0.0036	-0.0033	0.0386	24.2975	0.001
Evora	0.8243	-0.0036	-0.0019	0.0384	21.4956	0.001
Porto	0.8974	-0.0036	-0.0038	0.0375	24.0436	0.001
Braga	0.8809	-0.0036	-0.0021	0.0371	23.8153	0.001
Covilha	0.5578	-0.0036	-0.0028	0.0374	15.0019	0.001
Vila Real	0.7422	-0.0036	-0.0033	0.0359	20.7768	0.001
Bragança	0.8305	-0.0036	-0.0023	0.0372	22.3705	0.001
Guarda	0.7629	-0.0036	-0.0035	0.0379	20.2004	0.001

Source: Own elaboration.

**Figure 6 Moran scatter plot. Lisbon and Évora**



Source: Own elaboration.

**Figure 7: Moran scatter plot. Porto and Braga**



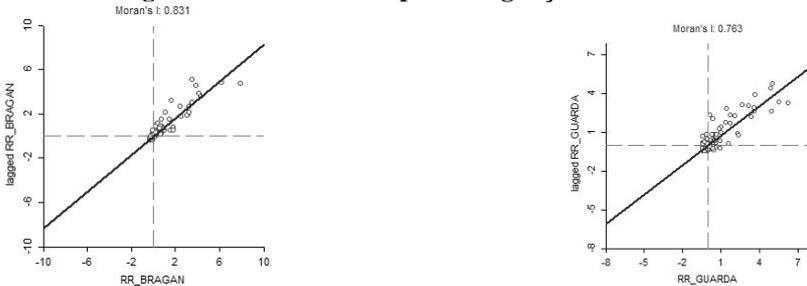
Source: Own elaboration.

**Figure 8: Moran scatter plot. Covilhã and Vila Real**



Source: Own elaboration.

**Figure 9: Moran scatter plot. Bragança and Guarda**



Source: Own elaboration.

The results of this section show there is a huge spatial dependence (autocorrelation) on the proportion of students from municipality  $i$  that are addressed, respectively, to each of the eight university cities, with the same distribution verified in the neighbouring municipalities. This autocorrelation is known, but it is not yet known where and how it occurs. Consequently, it is necessary to decompose the Moran's  $I$ .

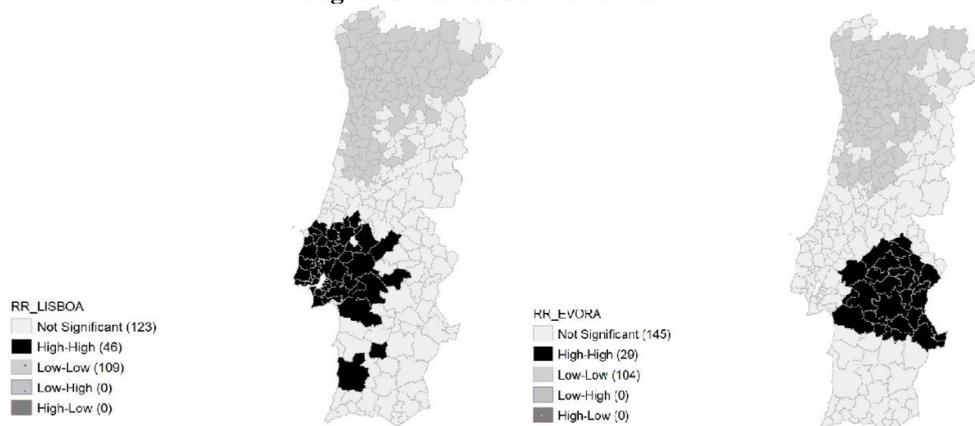
### b) Spatial Heterogeneity Through Local Indicators of Spatial Association—LISA

The results of the Moran local indicator calculations (LISA) are presented in a map of clusters which, like in the global Moran scatter plot, will be HH (high values in the core and in the neighbourhood); LL (low values in the core and in the neighbourhood); HL (high in the core and low in the neighbourhood); LH (low in the core and high in the neighbourhood). The difference is that here, only the values with statistical significance will be considered. The map also shows locations considered non-significant, which tend to be the majority. The interpretation for these locations is that the variable under analysis is not statistically different from the average in the other regions (Almeida, 2012).

The significance of  $I_i$  will also be obtained through a permutation process like that of the global Moran's  $I$ , and will now be performed for each observation. The result will be a  $p$ -value for each locality which will be used to obtain the significance of the estimator.

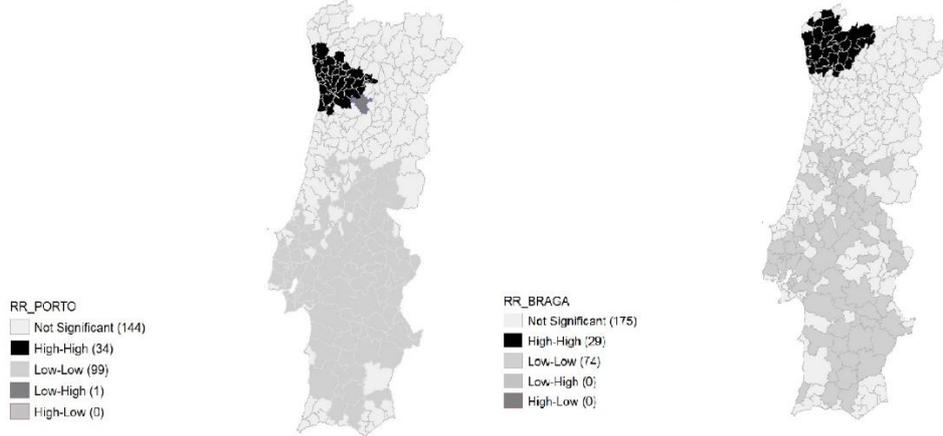
The LISA result for each university city appears in Figures 10 to 13. The dark areas are the clusters of municipalities in which the percentage of students that go to each university city is high (High-High), both in municipality  $i$  and in its neighbours. The areas in grey represent the clusters of municipalities in which this phenomenon is reduced (Low-Low). The very light grey areas, not significant, are those where the value of the variable does not statistically differ from the average of the set of municipalities in the whole country. The light grey and dark grey areas indicate outliers. In the eight analyses carried out, there are outliers in only 3. In Porto there is an outlier, characterised by a municipality with a low value of the variable and with neighbouring municipalities with a high value (Low-High); in Covilhã, two low-high and one high-low outliers; in Vila Real, only 1 High-Low. The two distinct clusters of municipalities, High-High and Low-Low, for all university cities, point to the presence of spatial heterogeneity.

Figure 10: LISA Lisbon and Évora



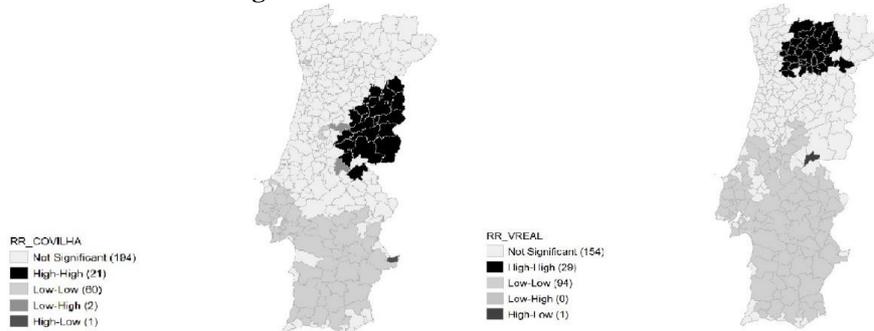
Source: Own elaboration.

**Figure 11: LISA Porto and Braga**



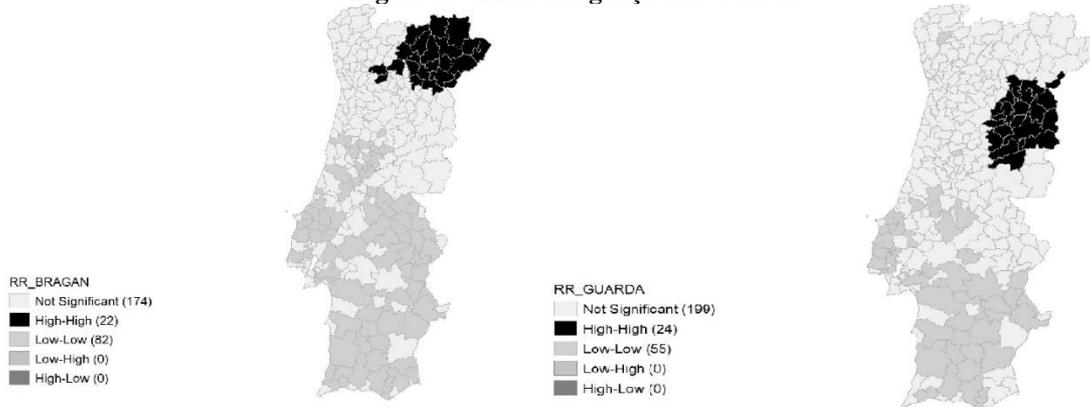
Source: Own elaboration.

**Figure 12: LISA Covilhã and Vila Real**



Source: Own elaboration.

**Figure 13: LISA Bragança and Guarda**



Source: Own elaboration.

The existence of a High-High cluster (in dark) can be considered, with more precision than Maier's criterion, as the market area of the university city in question. This greater precision comes from the fact that the cluster has statistical significance. It should be noted that, although they have some overlap, they clearly differ territorially. The university cities of the South have as their market area the municipalities closest to them and have a weak power of attraction in a vast area of the North and Centre of the country. The same happens with the university cities of the North but in the opposite direction.

### 4.3 Market Areas According to Different Criteria and Shared Areas

The market areas for each university city were delimited considering two criteria. The first one follows Maier (Maier, 2009) and the second uses LISA (Anselin, 1995). Except for Lisbon, in all other university cities, the market area obtained with the first criterion is contained in that obtained with the second. In the case of Guarda, the first criterion failed to define a market area.

The next step is to calculate the percentage of students from the municipalities that make up the respective market areas in each university city’s student population (total in the column). As shown in Table 7, the percentage of the second criterion was higher for all cities. The only case in which this difference was not significant is Lisbon. On the other hand, the importance of the market area for most cities exceeds 50%, especially for Lisbon, Porto and Braga. The explanation for these cities is the vast surrounding population and the diversity of courses they offer. In the case of Covilhã and Guarda, the explanation could come from their areas’ low population density. In the case of Évora, it is necessary to analyse Table 8 more carefully.

**Table 7: Percentage of students in the respective market area for each university city according to different concepts of market area**

Percentage of students in the respective market area for each university city according to different concepts of market area								
	Lisbon	Évora	Porto	Braga	Covilhã	V.Real	Bragança	Guarda
<b>Maier (destin for more than 40%)</b>	73.9	27.9	61.9	51.6	16.3	20.7	36.0	0.0
<b>Cluster HH</b>	76.2	44.5	80.9	85.3	34.2	51.8	57.6	48.5

Source: Own elaboration.

Table 8 deals with the market area segment shared by the pairs of university cities presented in the text. The choice of pairs considers the proximity between them and the fact that they are predominantly home to universities or, in the case of Bragança and Guarda, polytechnic institutes. We also rearranged two new pairs composed of universities and polytechnic institutes.

The market areas of Lisbon and Évora have six municipalities in common. Lisbon captures 529 students from this common segment, representing only 1.21% of its market area’s total students. Évora captures 304 students who represent almost 17% of the total students of its market area. In other words, in this shared area, Évora loses out to Lisbon. The case of Porto and Braga is more balanced, in absolute numbers, but with proportionately different results. Their respective market areas share ten municipalities. Attracting students in this common area is about twice as important for Braga. Covilhã and Vila Real have no common area, and between Bragança and Guarda, it is irrelevant.

The picture changes a lot when the pairs are Vila Real and Bragança, and Covilhã and Guarda. In the first case, there are nine municipalities in common. The competition between them is much more critical for Vila Real, which has about 30% of its students from this shared area. In turn, the 14 municipalities shared by Covilhã and Guarda are an important part of their respective market areas, especially for Guarda.

**Table 8: Market areas shared by pairs of university cities**

	Lisbon	Évora	Porto	Braga	Covilhã	V.Real	Bragança	Guarda	V.Real	Bragança	Covilhã	Guarda
<b>nº municipalities shared</b>	6		10		0		3		9			14
<b>nº students from shared area</b>	529	304	3 629	3 475	0	0	30	32	1 423	518	1 389	654
<b>% studentes from the respective market area (cluster HH)</b>	1,21	16,9	13,7	29,1	0	0	0,72	2,1	29,9	12,5	29,3	42,2

Source: Own elaboration.

Tables 7 and 8 reinforce what has been said throughout the text: students’ attraction to the HEIs of the respective university cities occurs much more on a regional basis than on a national basis. Furthermore, contrary to common sense, the most significant weight of the market areas in large agglomerations such as Lisbon, Porto and Braga is essentially regional. However, those supposedly with regional predominance, such as Évora, Covilhã and Guarda, recruit more students out of their market areas.

#### 4.4 Determinants of Proportion of Students from Municipality (Market Areas)

In order to understand and identify the determinants of students' choice of HEI, fractional models were estimated. To select the best specification, the logit, probit, loglog and cloglog functional forms were used (the cauchit functional form cannot be used here because it requires the presence of 0's and 1's in the observations) and the P test was performed. Given the existence of zeros in a significant portion in some cities (Lisbon is the only city that does not have a zero proportion in the dependent variable because it is the only one that receives students from all municipalities), we also estimate the two-part models, assuming the same functional forms already mentioned.

Given the large number of models that can be estimated, for one-part and two-part models, we start our empirical analysis by testing the models' specification. We perform the RESET test for all specifications, and the null hypothesis was rejected for all (results are available upon request).

Table 9 reports the regression results for each city. Selection of the models presented was based on the P tests and on the Pseudo R<sup>2</sup>

**Table 9: Estimation results for fractional regression models**

	Braga One-part Cloglog	Bragança One-part Cloglog	Covilha Two-part Cloglog	Évora One-part Cloglog	Guarda One-part Logit	Porto One-part Cloglog	Vila Real One-part Cloglog	Lisboa One-part Cloglog
Distance	-0.0119 *** (0.0015)	-0.0154 *** (0.0013)	-0.0072 *** (0.0016)	-0.0097 *** (0.001)	-0.0117 *** (0.0016)	-0.0074 *** (0.0011)	-0.0133 *** (0.0021)	-0.0049 *** (0.0004)
DHH	1.5125 *** (0.1846)	0.5608 ** (0.2425)	0.7576 *** (0.2415)	1.0491 *** (0.1439)	1.1227 *** (0.2483)	1.0732 *** (0.1253)	1.0976 *** (0.1741)	0.436 *** (0.0783)
DLL	-0.5195 *** (0.3173)	-0.5213 *** (0.3774)	-0.5728 *** (0.1595)	-0.5904 ** (0.2583)	-0.4569 * (0.2673)	-1.4531 *** (0.1919)	-0.0642 (0.4459)	-0.9433 *** (0.0899)
Constant	-2.1475 *** (0.2283)	-0.4668 (0.3062)	-1.8631 *** (0.3418)	-1.6578 *** (0.1860)	-2.4753 *** (0.2968)	-1.5284 *** (0.1554)	-1.888 *** (0.2495)	-0.3767 *** (0.0916)
N	278	278	278	278	278	278	278	278
Log_pseudolikelihood	-22.3554	-20.4577	-31.0469	-28.1521	-15.1272	-43.2932	23.2002	-81.4483
Pseudo R <sup>2</sup>	0.8180	0.8684	0.6251	0.8351	0.7405	0.8167	0.7758	0.8381

Note: \*\*\*, \*\* and \* refer to statistic tests which are significant at 1, 5 or 10%, respectively. All models were estimated with robust standard deviations (into brackets)

Source: Own elaboration.

For each explanatory variable, we report the coefficient and its standard deviation in brackets. For each model, the Pseudo R<sup>2</sup> is presented, showing that the selected models fit the data at least as well as the competing models. In terms of significance of explanatory variables, we may conclude that, qualitatively, most of the models have a similar interpretation. Distance has a significant and negative influence on the response variable for all the cities studied, and DHH shows a positive and significant influence on the distribution of students by municipality. On the other hand, DLL presents statistical significance for Covilhã, Évora, Guarda, Porto and Lisbon, always with a negative influence. Given that we have nonlinear models, direct interpretation of the coefficients it is not possible, with it being necessary to calculate the partial effects.

In Table 10 we report for each model, the respective partial effects, which were calculated as the mean of partial effects for each municipality in the sample.

**Table 10: Sample averages of partial effects**

	Braga	Bragança	Covilha	Évora	Guarda	Porto	Vila Real	Lisboa
Distance	-0.0004	-0.0004	-0.0003	-0.0004	-0.0002	-0.0005	-0.0004	-0.0008
DHH	0.0507	0.0162	0.0308	0.0426	0.0179	0.0747	0.03474	0.0697
DLL	-0.0174	-0.0151	-0.0274	-0.0239	-0.0073	-0.1011	-0.002	-0.1508

Source: Own elaboration.

According to the results presented in Table 10, we easily conclude that Distance is always a negative determinant of the distribution of students by municipality, its influence being more pronounced for Lisbon and Porto. On the other hand, DHH has a positive impact on the response variable, and again, it is the biggest municipalities that are most affected (Lisbon, Porto and Braga). Finally, DLL is negative for all models, with Lisbon and Porto being the cities with the greatest impact from this variable.

## 5. FINAL REMARKS

With the knowledge available until now, it is possible to say that, despite some overlap, university cities have different market areas - regardless of the market area definition - and these areas are strongly affected by the distance from students' homes. Even Lisbon, which receives students from all municipalities on the mainland, is no exception. At this point, the answer to university cities' spatial monopoly versus national attraction tends to confirm a kind of spatial monopoly, although higher education is not necessarily a homogeneous good.

This result is particularly important since the population, young people, are unequally distributed between the Portuguese coast (highly populated) and inland regions (sparsely occupied) and the volume of applications sent to the various institutions is very unequal. Recent policies to encourage higher education candidates to move from the most populated centres to inland regions, through special support to stay in higher education, seems to be a good measure in order to improve the balance in the Portuguese HE network and the possibility for smaller, younger institutions to contribute effectively to their regions' development. Furthermore, information on the geographical origin (municipalities) of students attending an HEI, together with demographic projections, makes it possible to predict future demand and take the most appropriate measures for the sustainability of the Portuguese higher education network.

Although the focus of the case study is Portugal, the methodology used is quite differentiating and can be applied in several contexts. In the future, this research may continue along two main lines: firstly, through applying fractional models, exploring the role of the determinants of HE demand (besides distance), and secondly, evaluating the existence of market areas, in certain courses/institutions, using microdata, which need specific authorization to be used.

## REFERENCES

- Almeida, E. (2012). *Econometria Espacial Aplicada (1a)*. Campinas: Alínea.
- Anselin, L. (1994). Exploratory spatial data analysis and geographic information systems. In M. Painho (Org.), *New tools for spatial analysis*, 45–54. Eurostat.
- Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27(2), 93–115.
- Anselin, L. (1996). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In Fischer, Manfred, Scholten, Henk, & Unwin, David (Eds.), *Spatial Analytical Perspectives on Gis in Environmental and Socio-Economic Sciences*, 111–125, London, UK: Taylor, Francis.
- Anselin, L. (2014). Geoda Workbook 1.10. A: <https://geodacenter.github.io/documentation.html>
- Azzone, G. & Soncin, M. (2020) Factors driving university choice: a principal component analysis on Italian institutions, *Studies in Higher Education*, 45:12, 2426-2438, DOI:10.1080/03075079.2019.1612354
- Briggs, S. (2006). An Exploratory Study of the Factors Influencing Undergraduate Student Choice: The Case of Higher Education in Scotland. *Studies in Higher Education* 31 (6): 705–22, DOI: 10.1080/03075079.2019.1612354
- De Jong, P., Sprenger, C., & Veen, F. (1984). On Extreme Values of Moran's I and Geary's c. *Geographical Analysis*, 16 (1), 17-24.
- Fonseca, M. P., Encarnação, S. (2012). *O sistema de ensino superior em Portugal em mapas e em números*, A3ESReadins nº 4, Agência de Avaliação e Acreditação do Ensino Superior, Lisboa.
- Fonseca, M., Justino, E., Amaral, A. (2020). Students' migration in a Portuguese hinterland public university. *Studies in Higher Education*, 45 (6), 1160-1182, DOI: 10.1080/03075079.2018.1553155
- Getis, A. (1995). Cliff, A.D. and Ord, J.K. 1973: Spatial autocorrelation. London: Pion. *Progress in Human Geography*, 19, 245 - 249.
- Griffith, D. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *The Canadian Geographer/ Le Geographe Canadien*, 40 (4), 351–367.

- Griffith, D. (2009). Spatial Autocorrelation. In *International Encyclopedia of Human Geography*, 396–402, R. Kitchin and N. Thrift: Elsevier,
- Lourenço, D., Sá, C. (2019). Spatial competition for students: What does (not) matter?. *The Annals of Regional Science*, 63 (1), 147-162, DOI: 10.1007/s00168-019-00930-1
- Lourenço, D., Sá, C., Tavares, O., Cardoso, S. (2020). Enrolling in Higher Education: The Impact of Regional Mobility and Public-Private Substitution Effects. *Journal of Economic Issues*, 54 (1), 183-197, DOI: 10.1080/00213624.2020.1720574
- Maier, G. (2009). Product differentiation or spatial monopoly? The market areas of Austrian universities in business education. In Varga, A. (Eds). *Universities, Knowledge Transfer and Regional Development: Geography, Entrepreneurship and Policy*, Edward Elgar, Cheltenham, U.K., 243-266.
- Oliveira, M., Vieira, C., Vieira, I. (2012). Explaining demand for higher education, Proceedings of ICERI2012 Conference, 19th-21st November 2012, Madrid, Spain ISBN: 978-84-616-0763-1
- Papke, L., Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11 (6), 619–632.
- Ramalho, E.A., Ramalho, J.J.S., Coelho L.M.S. (2018). Exponential Regression of Fractional-Response Fixed-Effects Models with an Application to Firm Capital Structure. *Journal of Econometric Methods*, De Gruyter, 7(1), 1-18.
- Ramalho, E.A., Ramalho, J.J.S., Henriques, P.D. (2010). Fractional regression models for second-stage DEA efficiency analyses. *Journal of Productivity Analysis*, 34 (3), 239–255.
- Ramalho, E.A., Ramalho, J.J.S., Murteira, J.M.R. (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys*, 25 (1), 19–68.
- Rego, C., Caleiro, A. (2004). A atracção das Universidades em Regiões Economicamente Deprimidas: o caso da Universidade de Évora. *Revista Portuguesa de Estudos Regionais*, nº 7, 3º quadrimestre, 19-40, APDR, Coimbra,
- Rego, C., Ramos, I.J., Lucas, M.R., Baltazar, M.S., Dionísio, A. (2021), New geography in old territories: a multivariate approach based on Portuguese regions. *Regional Studies, Regional Science*, 8 (1), 25-50. DOI: 10.1080/21681376.2020.1860807
- Rolim, C., Garcia, J. R. (2012). Área de influência territorial da UFPR. *Revista Economia & Tecnologia (RET)*, 8 (4), 33-44, Out/Dez.
- Sá C., Florax, R. J. G. M., Rietveld, P. (2004). Determinants of the regional demand for higher education in The Netherlands: a gravity model approach. *Regional Studies*, 38, 375–392 DOI: 10.1080/03434002000213905.
- Sá, C., Amado Tavares, D., Justino E., Amaral, A (2011). Higher education (related) choices in Portugal: joint decisions on institution type and leaving home. *Studies in Higher Education*, 36:6, 689-703, DOI: 10.1080/03075071003725343.
- Sá, C., Tavares, O. (2018) How student choice consistency affects the success of applications in Portuguese higher education. *Studies in Higher Education*, 43:12, 2148-2160, DOI: 10.1080/03075079.2017.1313219
- Simões Lopes, A. (1984). *Desenvolvimento Regional – problemática, teoria e modelos*, 2ª edição, Fundação Calouste Gulbenkian, Lisboa.
- Simões, C., Soares A. M. 2010. Applying to Higher Education: Information Sources and Choice Factors. *Studies in Higher Education* 35 (4), 371–89. DOI: /10.1080/03075070903096490
- Spiess, K, Wrolich, K. (2010). Does distance determine who attends a university in Germany?. *Economics of Education Review*, 29 (3), 470-479, DOI: 10.1016/j.econedurev.2009.10.009
- Van Bouwel, L., Veugelers, R. (2009). The Determinants of Student Mobility in Europe: The Quality Dimension, K.U.Leuven - Faculty of Business and Economics; Leuven (Belgium). FBE Research Report MSI\_0912FBE, 1 – 39.
- Vieira, C., Veira, I. (2013). Procura de Ensino Superior em Portugal: determinantes e perspectivas, in Rego, C., Caleiro, A., Vieira, C., Vieira, I. e Baltazar, S. (Coord.) *Redes de Ensino Superior: Contributos Perante os Desafios do Desenvolvimento*, Edição: Centro de Estudos e Formação Avançada em Gestão e Economia da Universidade de Évora, Évora, ISBN: 978-989-20-4215-2, 213-220.