DETERMINANTS OF DEMAND FOR CITIES WITH HIGHER EDUCATION INSTITUTIONS: AN APPROACH BASED ON FRACTIONAL REGRESSION

Professor Andreia DIONISIO
Departamento de Gestao & CEFAGE-U.E.,
Universidade de Evora, Portugal
andreia@uevora.pt
ORCID: 0000-0002-4289-9312

Professor Cassio ROLIM
Departamento de Economia,
Universidade Federal do Parana, Curitiba, Brasil
cassio.rolim@gmail.com
ORCID: 0000-0003-4619-5272

Professor Conceicao REGO
Departamento de Economia & CEFAGE-U.E.,
Universidade de Evora, Portugal
mcpr@uevora.pt
ORCID: 0000-0002-1257-412X

Abstract
Higher education institutions are typically situated in urban areas, making them appealing destinations for students seeking advanced education. This paper aims to explore the factors influencing the demand for cities with these institutions, focusing on the Portuguese context. By analysing distance and the quality of life in municipalities, we can better understand what attracts students to these university cities. Our findings, based on a fractional regression model, reveal that proximity to home and the disparity in rental and accommodation expenses play a significant role in the appeal of these cities for students and their families.

Keywords: *higher education institutions, fractional models, market areas, distance, housing costs*

JEL Classification: C21, I23, R12

Citation:
Dionisio A., Rolim C., Rego C., 2025. "Determinants of demand for cities with higher education institutions: An approach based on fractional regression", Sustainable Regional Development Scientific Journal, Vol. II, (1), 2025, pp. 81-93

1. Introduction

The democratization of higher education (HE) and the access to this level of education for a greater number of young people – in addition to the "*numerus clausus*" mechanisms that constraint the supply of vacancies in each higher education institution (HEI) – implies that, in Portugal, many students must leave their parents' home to be able to attend HE in a city far from their usual residence.
Thus, we are dealing with an educational, social and economic phenomenon with specific characteristics: young people who want to pursue their studies in HE and who, to do so, move from their parents' home to live in another city. That said, for this study, several contributions must be considered: on the one hand, the factors that determine the choice of HEI and, among them, aspects related with students' travel behaviour; on the other, the characteristics of cities, particularly those that host HEIs, mainly in terms of access to housing and the labour market.
Research on the factors that promote the demand for HE emphasize personal and family reasons, characteristics of a contextual nature (in particular, demographic, social and economic issues), and motivations linked with the institutional characteristics of HEIs (among others, Azzone & Soncin, 2020; Brigs, 2006; Fonseca et al., 2020; Lourenço and Sá, 2019; Rolim et al., 2024; Sá & Tavares, 2018; Sá et al., 2004; Simões & Soares, 2010; Spiess & Wrohlich, 2010;Vieira & Vieira, 2014). Well-known studies highlight distance as an important determinant in the decision to leave the family and move to another city. The studies related to the Portuguese case refer that most HE students prefer to continue living at their parents' homes; those who leave their parents' home prefer to stay in nearby HEIs. So, there turns out to be a negative relationship regarding the distance between the location of the HEIs and the family residence.
Due to their economic, social and institutional importance, HEIs are often blended in with the cities where they are located. It is generally accepted that "universities are "good" for cities and cities are "good" for universities" (Russo et al., 2003, p.22). The presence of an HEI, even a medium-sized one, is associated with the presence of a few thousand students, many of whom have moved from their family homes and who, for most of the year, come to reside in the city where they study. In the Portuguese case, the availability of accommodation for displaced students is very narrow: about 9%, i.e., 15,000 beds in the residences of public HEIs, to accommodate more than 175,000 displaced students (PNAES, n.d.[1]). So, the greater demand for housing by university students (but also by teachers and staff), from different origins, with a temporary or even permanent nature, puts pressure on the real estate market, causing an increase in the value of house rents as well as the value of properties (Rego et al., 2022). This finding is in line with the results of studies carried out in other European countries. Sá et al. (2012) shows that narrow housing markets reduce the probability of choosing a given university. "Rents play a major role in students' choice of university" (Sá et al, 2012, p.662); in the Netherlands, the choice of a particular university is negatively impacted by house rents. Also in this sense, for Germany, Goehausen and Thomsen (2024) concluded that "housing costs is the largest component of students' expenditures and a key location factor, which have contributed to the slowdown in HE expansion and reduced the skill-binding effect of universities, exacerbating regional inequality" (Goehausen and Thomsen, 2024). These authors quantified the effect of rising housing costs and concluded that "a one standard deviation increase in apartment rents decreased per-capita college enrollment by 1.1 percentage points on average" (Goehausen and Thomsen, 2024). This effect of rising prices in the real estate market is studied in Russo et al. (2003), who highlight also the characteristics of rigidity and segmentation in the housing market for HE students. Berry and Glaeser (2005), highlight the elasticity of housing supply effect for the accommodation of highly qualified workers. Munro et al. (2009), in turn, besides the characteristics of spatial segregation of students' location – in general student accommodation is mainly found in areas of the city close to the buildings of the universities – emphasize the effects of turnover, as well as those associated with disruption and a deterioration in the living conditions for local neighbours.
In brief, increased pressure on demand for housing in cities, for students but also for other permanent or temporary residents, along with the global shortage of supply of new or renovated construction for residential purposes, gives rise to the reduction of the supply of accommodation to the university community. This can constrict the ability of HEIs to evolve and of their students to all have accommodation.
Access to adequate and affordable housing, both for permanent and temporary residence, is a crucial element for well-being and the development of inclusive and resilient societies. This aspect is often considered a key determinant of the quality of life in cities, as highlighted by various studies (Amado et al., 2019; Barreira et al., 2021; OECD, 2022).

[1] Available on https://pnaes.pt/sobre-o-pnaes/ (accessed September, 2024).

Furthermore, research conducted by these authors suggests that the presence of universities plays a significant role in enhancing the quality and prestige of cities. Additionally, the availability of well-paid job opportunities in urban areas has been identified as a major factor driving highly skilled migration and contributing to the growth of cities (Amado et al., 2019; Faggian, 2009; Romão et al., 2018).
In conclusion, access to housing, the presence of educational institutions, and job opportunities are crucial factors that influence the liveability and appeal of cities, ultimately impacting their growth and development. A unique approach in research is to use municipal characteristics to analyse the demand for cities with HEIs. Traditionally, studies have focused solely on HEI characteristics to analyse demand. However, it is the aim of this study to delve deeper by exploring the determinants of the demand for cities with HEIs. This will be achieved by examining variables related to the housing market, purchasing power, employment, and delineation of HEIs market areas, as outlined in Rolim et al. (2024).
In addition to the importance of distance, this study aims to provide a comprehensive understanding of the factors influencing the demand for cities with HEIs. By incorporating a fractional regression model, in this research the aim is to improve the accuracy of the model in predicting the response variable. This innovative approach will provide valuable insights into the field and enhance our understanding of the dynamics that drive the demand for cities with HEIs. Estimating fractional regression models is a crucial advancement in understanding the demand for cities with HEIs. By fully considering the nature of the response variable, this contribution enhances our understanding of urban dynamics and the factors influencing city preferences.
After this introduction, Section 2 presents the method and data used in this study and the main results are identified and discussed in Section 3. A brief section of final remarks concludes the paper.

2. Methods and data

In order to better understand the factors influencing students' choice of city for HEIs, we will be examining the distance between the HEIs and students' family homes, as well as the market areas of cities with HEIs. Additionally, we will be exploring a range of attributes related to the characteristics of cities and the quality of life they offer to residents.
Drawing on the findings of ESDA research conducted by Rolim et al. in 2024, in this article, we seek to identify the key determinants that drive students to choose a particular city with an HEI for their education. By analysing urban characteristics and market areas, we aim to shed light on the factors that influence students' decisions in selecting a city for their HE pursuits.
In his 2009 study, Maier utilized a logit model to identify the factors influencing the market areas of Austrian universities offering business education. Our research also employs a similar modelling approach, albeit with a different focus. Our study examines the proportion of students enrolled in public HEIs in Portugal, considering both the municipality of the HEI and the municipality of the students' permanent residence. Unlike Maier's study, our dependent variable is not binary, but rather a continuous variable that can take on any value within the unit interval.

2.1 The fractional approach

The utilisation of fractional response variables is a prevalent practice within the field of economics. The inherent limitations of these variables, such as the capacity to observe values at the boundaries, give rise to significant concerns regarding functional form and inference. The utilisation of this particular type of data underscores the necessity for novel models and reliable estimation methodologies. Papke and Wooldridge (1996) introduced a set of econometric methods for fractional response variables, which were then developed in several research works, namely Ramalho, Ramalho and Henriques (2010), Ramalho, Ramalho and Coelho (2018), and Ramalho, Ramalho and Murteira (2011).
In order to better illustrate the methodological issues that arise from the use of a fractional dependent variable, consider y_i, $0 \leq y_i \leq 1$, which will be explained by a $1 \times K$ vector of independent variables (x_i);

the global linear model can be described by

$$E(y|x) = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K = \mathrm{x}\beta \quad (1)$$

where β is a $K \times 1$ vector. This specification model rarely provides the best description for $E(y|x)$ because

the dependent variable is bounded between 0 and 1 and the expected value of the dependent value rarely respects such boundaries (Papke and Wooldridge, 1996). An alternative commonly used for modeling this type of problem is utilizing the log-odds ratio as a linear function, known as logit and probit models.

However, despite some advantages, logit/probit modeling has limitations, particularly when the dependent variable has a positive probability of taking on the values 0 or 1. For more detailed information, refer to Papke and Wooldridge (1996) and Ramalho et al. (2011).

To ensure robust and unbiased estimations for determining the market area of Portuguese Higher Education Institutions (HEIs), we employ cross-sectional fractional models. In a cross-sectional context, the standard fractional regression model can be defined as:

$$E(y|x) = G(x_i\theta), \quad (2)$$

where θ 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.

Papke and Wooldridge (1996) introduced the concept of Quasi-Maximum Likelihood (QML) for estimating Eq.(2) using the Bernoulli log-likelihood function. They demonstrated that the estimated parameter, θ̂, is consistent, asymptotically normal, and efficient within a class of estimators that includes linear exponential family-based QML and weighted nonlinear least estimators.

Given the characteristics of our data, a significant portion of limit values in the fractional data, specifically 0s, are likely to be observed. The occurrence of the value 1 is highly improbable due to the nature of the response variable being a proportion of the whole. While Papke and Wooldridge (1996) and Ramalho et al. (2010) suggest that simple fractional regression models can still be utilized, they may not be optimal when dealing with a large number of corner observations.

In light of this, we have also conducted estimations using the two-part fractional regression model proposed by Ramalho et al. (2011). This model involves a binary regression model to predict the probability of specific corner values (0 or 1), followed by a conditional mean model to explain the remaining fractional values.

The two-part fractional regression model can be defined as:

$$E(y|x) = Pr(y_i > 0|x_{ib}).E(y_i|x_{if}, y_i > 0) = G(x_{ib}\theta_b).G_f(x_{if}\theta_f), \quad (3)$$

where x_{ib} and x_{if} are the explanatory variables used in the binary and in the fractional parts of the model,

θ_b and θ_f are vectors of variables coefficients and $G_b(\cdot)$ and $G_f(\cdot)$ are specified in the same way as $G(\cdot)$,

both functions are bounded between 0 and 1. It is commonly understood that the two components of the model in Equation (3) are independent. Therefore, they are estimated separately, with the binary component estimated using maximum likelihood and the fractional component estimated using quasi-maximum likelihood. When we use nonlinear regression models, the magnitude of the regression coefficients cannot be compared across models based on different functional forms. In order to answer this problem, we calculate the partial effects that have comparable interpretation. The partial effects are given by

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \theta_j g(x_i\theta) \quad (4)$$

where $g(x_i\theta)$ is given by $G(x_i\ \theta)[1 - G(x_i\ \theta)]$ for logit; $\phi(x_i\ \theta)$ for probit; $e^{-x_i}\theta G(x_i\ \theta)$ for loglog;

$e^{-x_i}\theta[1 - G(x\_i\ \ \theta)]$ for cloglog; and $\frac{1}{\pi}\frac{1}{(x_i\ \theta)+1}$ for cauchit modelling.

For the two-part models, the partial effects are given by the conjugation of the involved models:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \theta_{bj} g_b(x_i\theta).G_f(x_{if}\theta_f) + \theta_{fj} g_f(x_i\theta).G_b(x_{ib}\theta_b) \quad (5)$$

where $g_b(x_i\theta) = \partial G_b(x_i\theta)/\partial(x_i\theta)$ and $g_f(x_i\theta) = \partial G_f(x_i\theta)/\partial(x_i\theta)$.

Ensuring the accurate specification of the conditional mean of y_i is a critical assumption that must be validated through rigorous testing methods. Two recommended tests for this purpose are: (i) the RESET-type test proposed by Papke and Wooldridge in 1996; and (ii) the P test for general non-nested hypotheses, specifically adapted for fractional modeling by Ramalho et al. in 2011. These tests are essential in ensuring the reliability and validity of the model's predictions.

The RESET test is based on the results of a standard approximation for polynomials, assuming as null: $H_0: \phi = 0$, ϕ being a vector composed by the sum of the polynomials inserted. This test can be utilized to assess the functional form within the individual components of two-part models, even though it does not provide information on alternative specifications. Given these considerations, the P test will be employed in this research study. The P test is capable of evaluating model specifications, both independently and in comparison, to one-part models and other two-part models (or vice versa).

2.2. Data

In this research work, as already defined, the dependent variable, y_i, represents the proportion of higher education students from city j who chose to attend city i. The explanatory variables include the following:

Dij, = distance between city i and city j, in km;

DHH = dummy for market area municipalities of high-high clusters[2];

DLL = dummy for market area municipalities of low-low clusters;

CDHb = difference in bank valuation of properties per m^2 between destination city i and city j, in euros (source: INE 2017);

AC = accommodation cost – rental (difference between the destination city i cost and the city j, in euros (source: INE, 2017);

PP = index of purchasing power (difference between purchasing power of destiny city i and city j (source: INE 2017); and

EMP = employability rate of each city j compared to the employability rate of the destiny city i (source: INE 2018 and 2019).

We expect it to be possible to ascertain whether the market areas for each HEI under analysis are determinants of the proportion of HE students from city j who opted for city i.

The proportion of HE students from city j who opted for city i is in line with the methodology adopted in Rolim et al. (2024, p.89) where it is described in detail.

Formalizing,

$$\sum_{i=1}^{n} a_{ij} = \text{distribution of the students of the municipality i between the j university cities} \quad (1.1)$$

However, to perform spatial analysis, it is necessary to relativize these values. An auxiliary indicator could do this. This indicator, r_{ij}, indicates the proportion of students from each municipality in Portugal[3] among the university cities considered. Thus, each *aij* will be divided by the total of the students of the municipality *i* that went to the mainland university cities.

$$r_{ij} = \frac{a_{ij}}{\sum_{i=1}^{k} a_{ij}} \quad (1.2)$$

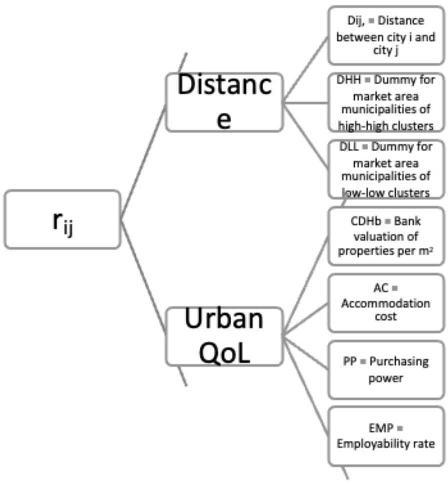
r_{ij} = proportion of students from municipality i to the university cities k.

Figure 1 presents the expected relation between the variables explored in this research work.

Figure 1. Relation between the variables of the model

[2] Dummy variables DLL and DHH were obtained using local indicators of spatial association – LISA (Rolim et al., 2024).

[3] In the present exercise only, the municipalities on the mainland.



Source: Own elaboration.

The dependent variable and the explanatory variables show differences between the cities where the HEIs are located (cf. Table 1). The average of the dependent variable shows that the cities of Lisbon and Porto (the two largest Portuguese cities) have the highest proportion of HE students and the results among the medium-sized cities with a university/HE (Braga, Covilhã, Évora and Vila Real) are similar. In the case of explanatory variables, analysis of distance shows that Bragança (a city located in the far north of the country, close to the border with Spain) stands out as being at the greatest distance from the students' family home. The difference in housing costs (be it bank evaluation of properties per m^2 or accommodation cost) is particularly high in the cases of the cities of Lisbon and Porto, where demand pressure is higher. The medium-sized cities of Bragança and Guarda present values that reflect the existence of cheaper accommodation. The differences in the purchasing power and employability variables are also significant, with the cities of Lisbon and Porto standing out positively in both cases. In the case of differences in the employability rate, the cities of Bragança and Vila Real (medium-sized cities in the north of the country) reveal a less evident dynamic than the average of the cities of origin of the young people who chose them to study.

Table 1. Descriptive of the variables

		Braga	Bragança	Covilhã	Évora	Guarda	Lisbon	Porto	Vila Real
r_{ij} (%)	**Mean**	0.038	0.0332	0.0395	0.0451	0.0174	0.2027	0.0803	0.0344
	St. Dev	0.0947	0.0857	0.0639	0.09	0.0394	0.1924	0.1297	0.076
	Max.	0.6186	0.7128	0.5815	0.5599	0.2632	0.839	0.6946	0.5331
	Min	0	0	0	0	0	0.0215	0	0
Dist (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	143.98	168.128
	Max.	650	774.439	567	561	602.753	506	602	669
	Min	0	0	0	0	0	0	0	0
CDHb (€)	**Mean**	8412.3	-22154.7	9632.3	33961.3	-159.7	183621.3	72632.3	3097.3
	St. Dev	54784.5	54784.5	54784.5	54784.5	54784.5	54784.5	54784.5	54784.5
	Max.	89063	58496	90283	114612	80491	264272	153283	83748
	Min	-529372	-559939	-528152	-503823	-537944	-354163	-465152	-534687
PP	**Mean**	26.48	15.98	6.08	36.78	15.68	139.08	77.28	17.58
	St. Dev	18.68	18.68	18.68	18.68	18.68	18.68	18.68	18.68

	Max.	51.7	41.2	31.3	62	40.9	164.3	102.5	42.8
	Min	-112.6	-123.1	-133	-102.3	-123.4	0	-61.8	-121.5
AC (€)	**Mean**	0.45	-0.68	-0.41	1.43	-0.44	6.28	3.43	0.01
	St. Dev	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
	Max.	2.19	1.06	1.33	3.17	1.3	8.02	5.17	1.75
	Min	-5.83	-6.96	-6.69	-4.85	-6.72	0	-2.85	-6.27
EMP (%)	**Mean**	0.092	-0.002	0.001	0.074	0.048	0.429	0.253	-0.006
	St. Dev	0.076	0.076	0.076	0.076	0.076	0.076	0.076	0.076
	Max.	0.206	0.112	0.116	0.189	0.163	0.543	0.368	0.108
	Min	-0.337	-0.431	-0.427	-0.355	-0.380	0.000	-0.175	-0.435

Source: Own elaboration.

3. Results – Determinants of proportion of students from municipality (market areas)

To gain insight into the factors influencing students' choices of higher education institutions (HEIs), fractional models were utilized in our analysis. Various functional forms including logit, probit, loglog, and cloglog were considered to determine the best specification. The Cauchit functional form was not applicable due to the absence of 0s and 1s in the observations.

In cities where zeros were prevalent, particularly Lisbon which had a mix of both 0s and 1s, two-part models were also estimated using the same functional forms. The complexity of the models required a thorough examination of specifications, starting with the RESET test to ensure accuracy. The results of the test led to the rejection of the null hypothesis in all cases.

Additionally, pseudo-R^2 values for each model and specification were included in Tables 2, 3, 4, and 5 to provide a comprehensive overview of the analysis. To determine the optimal model type and functional form, the P test was conducted. This rigorous approach allowed for a thorough evaluation of the determinants influencing students' choices of HEIs.

Table 6 summarizes the results obtained for the P tests of the one-part and two-part fractional models. This table should be read as follows: the columns present the null hypothesis of the P test, and the lines present each of the alternative hypothesis evaluated. The values in the table refer to the p-values for each hypothesis.

As an example of interpretation, we can analyse the results presented in Table 3, which indicate that only a few specifications are admissible. For Braga, the P test rejects all the null hypotheses, and for Bragança, the best alternative is logit for a one-part model. Given the fact that the small number of determinants could be the reason for the misspecification, we also use the pseudo-R^2 to select the best model. Given this, the specification for Braga would be a one-part cloglog model (pseudo-R^2 is 0.8390).

Table 2. Specification test for one-part and two-part models (p-values) for Braga and Braganca

Braga

P test	**One-part model**				**Two-part model**			
Braga	**Logit**	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.000***	0.000***	0.0001***	-	0.0002***	0.0002***	0.0010***
H_1: Probit	0.0007***	-	0.0000***	0.0020***	0.0002***	-	0.000***	0.0002***
H_1: Loglog	0.0009***	0.0001***	-	0.0050***	0.0002***	0.0001***	-	0.000***
H_1: Cloglog	0.0007***	0.000***	0.000***	-	0.002***	0.0003***	0.0003***	-
Pseudo-R^2	0.8390	0.7957	0.7547	0.8514	0.8019	0.7541	0.7169	0.8166

Bragança

	One-part model				**Two-part model**			
	Logit	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.0019***	0.0776**	0.0567*	-	0.000***	0.0002***	0.0104**
H_1: Probit	0.0003***	-	0.0083***	0.0071**	0.000***	-	0.0004***	0.0009***
H_1: Loglog	0.0002***	0.000***	-	0.0054***	0.000***	0.0002***	-	0.0002***
H_1: Cloglog	0.0120**	0.0026***	0,0781*	-	0.000**	0.0008***	0.0004***	-
Pseudo-R^2	0.8732	0.8297	0.7666	0.8834	0.8515	0.8021	0.8641	0.8230

Note: ***, ** and * refer to statistic tests which are significant at 1, 5 or 10%, respectively.

For Table 3, using the same approach, we use the one-part cloglog model for Covilhã (pseudo-R^2 is 0.6267) and for Évora (pseudo-R^2 is 0.8273).

Table 3. Specification test for one-part and two-part models (p-values) for Covilhã and Évora

Covilhã

P test	**One-part model**				**Two-part model**			
	Logit	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.0002***	0.0008***	0.0015**	-	0.0001***	0.0004***	0.0007**
H_1: Probit	0.0005***	-	0.0002***	0.0002***	0.0005***	-	0.000***	0.0002***
H_1: Loglog	0.0017***	0.0004***	-	0.0015***	0.0002***	0.0009***	-	0.0002***
H_1: Cloglog	0.0026**	0.0008***	0.0011***	-	0.0002**	0.0005***	0.0002***	-
Pseudo-R^2	0.6163	0.5807	0.5535	0.6267	0.5769	0.5499	0.5769	0.6244

Évora

	One-part model				**Two-part model**			
	Logit	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.0000***	0.000***	0.0001***	-	0.0001***	0.000***	0.0002***
H_1: Probit	0.000***	-	0.000***	0.0003***	0.000***	-	0.0001***	0.0002***
H_1: Loglog	0.000***	0.000***	-	0.000***	0.000***	0.0001***	-	0.0002***
H_1: Cloglog	0.0002***	0.000***	0.0001***	-	0.000***	0.0000***	0.0002***	-
Pseudo-R^2	0.8315	0.8031	0.7770	0.8415	0.7894	0.7690	0.7646	0.8273

Note: ***, ** and * refer to statistic tests which are significant at 1, 5 or 10%, respectively.

According to the results presented for Guarda and Porto (Table 4), using the same approach, we use the one-part cloglog model. Pseudo-R^2 for Guarda is 0.7466 and for Porto 0.8402.

Table 4. Specification test for one-part and two-part models (p-values) for Guarda and Porto

Guarda

P test	**One-part model**				**Two-part model**			
	Logit	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.0026***	0.0017***	0.1052	-	0.0002***	0.0004***	0.0002***
H_1: Probit	0.0187**	-	0.0017***	0.0286*	0.000**	-	0.0002***	0.0002**
H_1: Loglog	0.0411**	0.0071***	-	0.0572**	0.000**	0.0005***	-	0.0001**
H_1: Cloglog	0.0783*	0.0029***	0.0019***	-	0.0002***	0.0090***	0.0001***	-
Pseudo-R^2	0.7432	0.7144	0.6432	0.7466	0.6932	0.6721	0.7278	0.7233

Porto

P test	**One-part model**				**Two-part model**			
	Logit	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.0000***	0.0000***	0.0000***	-	0.0000***	0.0000***	0.0000***
H_1: Probit	0.0000***	-	0.0000***	0.0000***	0.0000***	-	0.0000***	0.0000***
H_1: Loglog	0.0000***	0.0000***	-	0.0000***	0.0000***	0.0000***	-	0.0000***
H_1: Cloglog	0.0000***	0.0000***	0.0000***	-	0.0000***	0.0000***	0.0000***	-
Pseudo-R^2	0.8591	0.8368	0.8001	0.8684	0.8253	0.7864	0.8608	0.8402

Note: ***, ** and * refer to statistic tests which are significant at 1, 5 or 10%, respectively.

Finally, according to the results in Table 5, we use the one-part cloglog model for Vila Real (pseudo-R^2 is 0.7773) and Lisbon (pseudo-R^2 is 0.8677).

Table 5. Specification test for one-part for Lisbon and two-part models for VilaReal (p-values)

Vila Real

P test	**One-part model**				**Two-part model**			
Vila Real	**Logit**	**Probit**	**Loglog**	**Cloglog**	**Logit**	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.007***	0.0002***	0.0015***	-	0.0005***	0.000***	0.0005***
H_1: Probit	0.0026***	-	0.0003***	0.0027***	0.000***	-	0.000***	0.0001***
H_1: Loglog	0.0077***	0.0006***	-	0.0098***	0.000***	0.0001***	-	0.0002**
H_1: Cloglog	0.0017***	0.0006***	0.0002***	-	0.0007***	0.0002***	0.0021***	-
Pseudo-R^2	0.7661	0.7232	0.6876	0.7773	0.6831	0.6498	0.7391	0.7420

Lisbon

	One-part model			
	Logit	**Probit**	**Loglog**	**Cloglog**
H_1: Logit	-	0.0000***	0.0000***	0.2435
H_1: Probit	0.0000***	-	0.0000***	0.0000***
H_1: Loglog	0.0000***	0.0000***	-	0.0008***
H_1: Cloglog	0.0004***	0.0039***	0.0024***	-
Pseudo-R^2	0.8650	0.8581	0.8406	0.8677

Note: ***, ** and * refer to statistic tests which are significant at 1, 5 or 10%, respectively.

Table 6 reports the regression results for each city. The selection of the presented models was based on the P tests and on the pseudo-R^2, as above.

Table 6. Estimation results for fractional regression models

	Braga		Bragança		Covilhã		Évora		Guarda		Porto		Vila Real		Lisbon	
	One-part Cloglog		One-part Cloglog		One-part Cloglog		One-part Cloglog		One-part Cloglog		One-part Cloglog		One-part Cloglog		One-part Cloglog	
Distance	- 0.0131	***	-0.0160	***	-0.0066	***	-0.0108	***	-0.0114	***	-0.0070	***	-0.0135	***	-0.0041	***
	(0.0015)		(0.0019)		(0.0016)		(0.0011)		(0.0017)		(0.0009)		(0.0022)		(0.0004)	
DHH	1.4363	***	0.5289	**	0.8763	***	0.9400	***	1.058	***	1.0394	***	1.0038	***	0.3866	***
	(0.2003)		(0.2249)		(0.2185)		(0.1569)		(0.2284)		(0.1213)		(0.1723)		(0.0747)	
DLL	-0.4261		-0.4334		-0.5259	***	-0.5214	**	-0.3484		-1.672	***	-0.0225		-0.9926	***
	(0.2938)		(0.3677)		(0.1694)		(0.2641)		(0.2639)		(0.1884)		(0.46134)		(0.0821)	
CDHb	-4.07e-06	**	-3.93e-06	**	-9.27e-07	*	2.76e-07		4.02e-06	*	6.38e-07		9.60e-07		1.04e-07	
	(1.98e-06)		(1.71e-06)		(5.40e-07)		(3.61e-07)		(2.35e-06)		(1.56e-06)		(2.91e-06)		(3.18e-07)	
PP	-0.0025		0.0172	*	0.0143		0.0162	***	0.0073		-0.0133	**	0.0113		-0.0009	
	(0.0123)		(0.0103)		(0.0088)		(0.0052)		(0.0095)		(0.0067)		(0.0127)		(0.0029)	
AC	0.2088		-0.1154		0.1013		-0.0298		0.0329		-0.0488		-0.05123		-0.1423	***
	(0.1409)		(0.1387)		(0.1044)		(0.0810)		(0.1312)		(0.0999)		(0.1417)		(0.0376)	
EMP	2.7586	*	0.5323		-2.8895		-3.1059	***	-1.504		2.5164	***	-0.0545		0.1792	
	(1.4817)		(2.1572)		(1.8403)		(0.9723)		(1.9856)		(0.9448)		(1.7366)		(0.5189)	
CONS	-2.2557	***	-0.8615	*	-2.1581	***	-1.7456	***	-2.6562	***	-1.003	***	-2.1164	***	0.4027	**
	(0.3154)		(0.4771)		(0.3460)		(0.2266)		(0.3443)		(0.1922)		(0.4178)		(0.1556)	
N	278		278		278		278		278		278		278		278	
Log_pseudolikelihood	-22.053		-20.248		-31.315		-27.988		-15.069		-42.829		-23.109		-80.694	
Pseudo-R2	0.8514		0.8834		0.6267		0.8415		0.7466		0.8684		0.7773		0.8677	

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

For each explanatory variable, we provide the coefficient along with its standard deviation in brackets. Additionally, the pseudo-R2 for each model is included, demonstrating that the chosen models effectively capture the data compared to alternative models. Regarding the significance of the explanatory variables, it is evident that most models share a similar interpretation. Distance consistently exerts a significant and negative impact on the response variable across all cities examined, while DHH exhibits a positive and significant effect on student distribution by municipality. Conversely, DLL demonstrates statistical significance in Covilhã, Évora, Porto, and Lisbon, consistently yielding a negative influence.
As regards the value of properties per m^2, a negative influence can be seen for Braga, Bragança and Covilhã and slightly positive for Guarda. If we add the results for AC (accommodation cost) we find a negative impact only in Lisbon. These results may show that Lisbon is penalized in terms of demand by the high rental prices for accommodation. Cities such as Braga, Bragança and Covilhã, on the other hand, are essentially affected by the sale prices of properties, which may be due to their geographic location and the fact that rental values do differ much from the country average.
Another issue that may be a determinant of the proportion of HE students from city j who opted for city i is the difference in the level of employability from the destination city i from the origin city j. We found a positive and significant impact in Braga and Porto as destination cities and a negative impact in Évora. These results seem to suggest differences in the dynamism of labour markets in these cities. While Porto and Braga are cities where there is a predominance of employment in private companies, many of which are medium-sized and large, in the case of the city of Évora, public employment prevails, meaning that the dynamics of creating new jobs are less significant. This result is in line with what Fonseca (2023) had already concluded: the existence of a movement of graduates from small and medium-sized cities to larger urban areas, due to the fragility of the job market in these cities, where there is a temporary effect of over-education.
Given the fact that we have non-linear models, the direct interpretation of the coefficients it is not possible, with the calculus of the partial effects being necessary.
In Table 7, for each model we report the respective partial effects, which were calculated as the mean of partial effects for each municipality in the sample.

Table 7. Sample averages of partial effects

	Braga	**Bragança**	**Covilhã**	**Évora**	**Guarda**	**Porto**	**Vila Real**	**Lisbon**
Distance	-0.0004	-0.0005	-0.0002	-0.0004	-0.0002	-0.0005	-0.0004	-0.0006
DHH	0.0479	0.0154	0.0329	0.0381	0.0176	0.0714	0.0318	0.0610
DLL	-0.0142	-0.0127	-0.0198	-0.0212	-0.0058	-0.1148	-0.0007	-0.1567
CDHb	-1.36e-07	-1.15e-07	-3.48e-08	1.12e-08	6.70e-08	4.38e-08	3.04e-08	1.64e-08
PP	-0.00008	0.0005	0.0005	0.0007	0.0001	-0.0009	0.0004	-0.0001
AC	0.007	-0.0034	0.0038	-0.0012	0.0005	-0.0034	-0.0016	-0.0225
EMP	0.0919	0.0155	-0.1085	-0.1260	-0.0251	0.1728	-0.0017	0.0283

The analysis presented in Table 7 clearly indicates that distance plays a significant role in determining the distribution of students across municipalities. This impact is particularly noticeable in cities like Lisbon and Porto. Conversely, the availability of dormitory housing (DHH) has a positive effect on student distribution, with larger cities such as Lisbon, Porto, and Braga experiencing the greatest influence.

Additionally, the variable DLL consistently shows a negative impact across all models, with Lisbon and Porto being the most affected cities. Other factors such as CDHb, AC, PP, and EMP exhibit varying effects on student distribution. However, as demonstrated in Table 7, many of these coefficients are statistically insignificant, suggesting that these variables do not significantly impact the demand for a particular university city.
In conclusion, while distance and dormitory housing availability are key determinants in student distribution, other factors may not play as significant a role in shaping demand for specific university cities.

4. Final remarks

Understanding the motivations behind HE students choosing to study at institutions located in cities other than their hometowns is crucial for both these academic institutions and public policy decision-makers, especially in the fields of HE and urban planning. The aim of this study was to identify the determinant factors in the demand for cities where HEIs are located, presented as the proportion of HE students from city *j* who opted to study in city *i*. In this sense, the model is estimated using a fractional response variable. The econometric methodology relies on fractional models, which enable regression models to be applied to continuous variables within a specific range. One key benefit of this approach is that it eliminates the need for researchers to make strong assumptions about all data aspects when developing a model to achieve reliable parameter estimates. Therefore, we recognize the importance of the response variable and strive to achieve consistency in our results, leading to significant gains. To achieve the objective of the study, two sets of potential determinants were used as explanatory variables: one related to the distance between the city where the HEI is located and the city where the student's family lives, and another associated with quality of life. The results lead to the conclusion that, on one hand, the distance between the city of family residence and the cities where the HEIs are located is the main factor that explains the choice made by students: in general, the greater the distance between the two cities, the lower the demand. These results are aligned with others already obtained, which identified the negative relationship associated with the distance between the family's city of residence and the city where HEI is located. This arises from the fact that many students prefer to continue their HE studies close to their family home. On the other hand, the difference in housing costs also proved to be decisive, as expected, particularly in the case of Lisbon, and diminishes demand. The greater difference in rental prices between this city and other locations may be affecting the demand for HE in institutions located there. High housing costs increase the total cost of attending HE and therefore also discourage people from leaving their parents' home and choosing larger cities, where these prices are higher. The cost of accommodation is one of the main costs for those studying away from home and the increase in this has led many young people to adapt their choice of cities for HE studies to the possibility of paying for housing. In the most extreme cases, the cost of accommodation for displaced students can make it impossible to stay and continue studying in HE. These effects penalize equity among students.

This study highlights the importance of knowledge in the field as it confirms previous findings in both Portugal and other regions. It shows that the demand for HE and the likelihood of students staying in this level of education is positively influenced by proximity to HEIs and the availability of affordable accommodation for displaced students.

Public policy must consider that a more widespread network of HEIs can help democratize access to HE and improve territorial cohesion. This can also increase the availability of public university accommodation, which in turn can boost demand and retention rates in HE.

Additionally, the methodology used in this study is more robust in econometric terms, as the response variable is not unbounded and the consistency of estimators is a significant achievement in this research.

Acknowledgements

Andreia Dionísio and Conceição Rego are pleased to acknowledge financial support from Fundaçao para a Ciencia e a Tecnologia [(grant UIDB/04007)].

References

Amado C, Barreira AP, Santos S, Guimarães MH. (2019). Comparing the quality of life of cities that gained and lost population: An assessment with DEA and the Malmquist index. *Papers in Regional Science*, 98, 2075–2097. DOI: https://doi.org/10.1111/pirs.12448

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

Barreira, A. P., Amado, C., Santos, S., Andraz, J., & Guimarães, M. H. (2021). Assessment and Determinants of the Quality of Life in Portuguese Cities. *International Regional Science Review*, *44*:6, 647–683. DOI: 10.1177/0160017620979611

Berry, C. R., Edward L. Glaeser, E. L. (2005). The divergence of human capital levels across cities. *Papers in Regional Science*, 84: 3, 407–444.

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

Faggian, A., McCann, P. (2009). Universities, Agglomerations and Graduate Human Capital Mobility. *Journal of Economic and Human Geography*, 100:2, 210–223, DOI: 10.1111/j.1467-9663.2009.00530.x.

Fonseca, M. (2023). Innovation in the peripheries: Counter-flows of students to second tier cities in Portugal. *Geoforum*, 141, 103732. https://doi.org/10.1016/j.geoforum.2023.103732.

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

Goehausen, J., Thomsen, S. L. (2024). Housing Costs, College Enrollment, and Student Mobility. *Discussion Papers Series*, IZA – Institute of Labor Economics, January. IZA DP No. 16726.

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

Maier, G. (2009). Product differentiation or spatial monopoly? The market areas of Austrian universities in business education. In Varga, A. (2009). *Universities, Knowledge Transfer and Regional Development: Geography, Entrepreneurship and Policy*. Edward Elgar, Cheltenham, U.K.

Munro, M., Turok, I., Livingston, M. (2009). Students in Cities: A Preliminary Analysis of Their Patterns and Effects. *Environment and Planning A: Economy and Space*, *41:*8, 1805–1825. DOI: 10.1068/a41133

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*, 7 (1), https://doi.org/10.1515/jem-2015-0019

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., Freire, M., Ramos, I.J., Lucas, M.R. (2022). Universidades e Desenvolvimento Local: discussão em torno dos efeitos das instituições de ensino superior nas cidades. In Tatiane Salete Mattei, Cínthia Santos Silva e Lucir Reinaldo Alves (Orgs.). *Economia e desenvolvimento local*. Toledo, PR: Núcleo de Desenvolvimento Regional; ISBN: 978-65-00-44814-6. pp. 38–48. Available at: https://www.unioeste.br/portal/nucleos-toledo/ndr.

Rolim, C., Rego, C., Dionísio, A., Ferreira, E. (2024). Atração Nacional ou Regional de Estudantes: "áreas de mercado" de cidades universitárias. *Revista Portuguesa de Estudos Regionais*, 67, 95–113. DOI:10.59072/rper.vi67.349.

Romão, J., Kourtit, K., Neuts, B., Nijkamp, P. (2018). The smart city as a common place for tourists and residents: A structural analysis of the determinants of urban attractiveness. *Cities*, 78, 67–75, DOI: 10.1016/j.cities.2017.11.007.

Russo, A., Van den Berg. L., Lavanga, M. (2003). The Student City. Strategic Planning for Student Communities in EU Cities, ERSA conference papers ersa03p485, European Regional Science Association. Available at https://ideas.repec.org/p/wiw/wiwrsa/ersa03p485.html.

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

Sá, C., Florax, R. J.G.M., Rietveld, P. (2012). Living Arrangement and University Choice of Dutch Prospective Students. *Regional Studies*, 46:5, 651–667, DOI: 10.1080/00343404.2010.529119

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.

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

Vieira, C., Vieira, I. (2014). What drives university applications? An attempt to explain aggregate demand for higher education. *Journal of Higher Education Policy and Management*, 36:6, 616–631, DOI: 10.1080/1360080X.2014.957894