ISCTE Business School
Instituto Universitário de Lisboa

HIGHER EDUCATION AND UNEMPLOYMENT - ANALYSIS OF THE DEMAND AND SUPPLY DETERMINANTS AMONGST ALL COURSES

Maria João Toscano Farófia

Dissertation submitted as partial requirement for the conferral of
Master of Science in Business Administration

Supervisor:
Prof. Alexandra Ferreira Lopes, Assistant Professor, PhD, ISCTE Business School, Department of Economics

Co-Supervisor:
Prof. Luís Filipe Martins, Assistant Professor with "Habilitation" of Economics, PhD, ISCTE Business School, Department of Economics

September 2019

Abstract

During the first decade of the 21st Century, social and economic concerns (such as budget deficit and a fragile economy) led the Portuguese Government to publish unemployment rates, by pair of course and Higher Education Institution (HEI) in order to reassess funding to assign to HEIs based on their course demand, and enable candidates to make a more informed choice. Such measure may have influenced the closure of some courses and the increase in vacancies of others, which is why this study emerges.

The strength index is our dependent variable, resulting from the ratio between the number of 1st option candidates and the number of vacancies, for a given course associated with an HEI's organic unit.

We measured the determinants of the demand for higher education courses and the impact of disclosing the unemployment information, for the demand adjustment for courses in Portugal, from 2005 until 2016, using panel data.

The results revealed a negative correlation between the unemployment rate and the strength index. All other determinants (GDP, population density and a dummy variable, built from the percentage of candidates from the district of origin with most candidates per course) had the expected signs. The robustness exercises showed that most signs are robust when dividing by degree, NUTSII region, candidates and graduates' gender, and fields of study. We concluded that the influence of the unemployment rate on course choice and HEI is consensual and that it may have an impact on the demand for each course and on the respective supply by HEIs.

Keywords: "supply"; "demand"; "higher education"; "unemployment".

JEL Classification: C33; l23

Sumário

Na primeira década do século XXI, preocupações sociais e económicas (como o défice orçamental e uma economia frágil) levaram o Governo Português a publicar as taxas de desemprego por curso e Instituição de Ensino Superior (IES), com o intuito de reavaliar o financiamento a atribuir às IES, baseado na procura de cursos, e possibilitar aos candidatos uma escolha mais informada. Esta medida pode ter influenciado alguns cursos, aumentando vagas ou extinguindo-os, razão pela qual surge este estudo.

O Índice de Força é a variável dependente, resultante do rácio entre o número de candidatos em primeira opção e o número de vagas por curso associado a unidade orgânica de uma IES. Mediu-se o impacto da divulgação da taxa no índice de força dos pares cursos/unidades orgânicas e possível ajuste da procura de cursos em Portugal, de 2005 a 2016. Analisámos os principais determinantes da oferta e procura de cursos e da taxa de desemprego por curso e IES, utilizando dados de painel.

Os resultados revelaram uma correlação negativa entre taxa de desemprego e índice de força. As restantes determinantes (PIB, densidade populacional e variável *dummy*, construída a partir da percentagem de candidatos do distrito de origem com maior número de candidatos por curso) apresentaram os sinais esperados. Os exercícios de robustez revelaram maioritariamente os sinais apresentados na estimação original, por grau, região NUTSII, género dos candidatos e diplomados e por áreas de ensino. Concluímos que a influência da taxa de desemprego é consensual e que pode ter um impacto na procura de cada curso e na respetiva oferta pelas IES.

Palavras-chave: “oferta”; “procura”; “ensino superior”; “desemprego”.

Códigos do JEL: C33; l23

Acknowledgements

Firstly, I would like to thank my supervisor Professor Alexandra Ferreira Lopes for the continuous support, patience, and immense exchange of opinions and suggestions during the time of research and writing of this thesis.

I would also like to thank my co-supervisor Professor Luís Martins for bringing in his valuable insight and focus into this study and generously helping me to understand and apply new concepts and methods.

My sincere thanks also go to Professor Ricardo Ribeiro from the Department of Information Science and Technology at the School of Technology and Architecture of the ISCTE-IUL, because he believed in the purpose of this work and decided to also give his precious contribution in the extraction of valuable data despite his busy schedule.

Last but not the least, I would like to thank my parents and my sister for supporting me throughout the writing of this dissertation, and specially to my friend Ágata for being an inspiration of persistence and letting me always know that I could count on her.

Index

List of Figures

List of Tables

List of Abbreviations

1. Introduction

We aim to investigate the determinants of the supply and demand for higher education courses and the impact of the disclosure of information about unemployment by course and by Higher Education Institution, for the demand adjustment for higher education courses in Portugal, from 2005 until 2016.

Several factors play a role in students' decision to choose a certain higher education course. Amongst those factors, does the students' ultimate choice (of higher education course and institution) adjust to existing information about courses' unemployment rates? The total number of employed individuals varies through the years by economic sector and geographic location within each country, which is why prospects of low unemployment may be an incentive to pursue a specific field of study and course, leading to an increase in its demand.

In the middle of the first decade of the 21st century, the Portuguese Government decided to make available the unemployment rates by pairs of course and Higher Education Institution. This decision implied a joint group work from the Ministry of Work and Social Security and from the Ministry of Science and Higher Education, which worked on the matching between the information that unemployed individuals gave to the Ministry of Work and Social Security and information about courses detained by the Ministry of Science and Higher Education. This decision was taken due to several concerns. First, there was an ongoing discussion in Portugal about providing public funds to universities grounded on their course demand. Second, students must have the possibility to choose amongst higher education courses sustained on all information. Finally, growing concerns about the budget deficit and also about meager economic growth, also lead to this reasoning. This measure can have several potential implications in terms of Higher Education policy, namely the closure of some courses and the increase in the vacancies for others.

The study of the supply-demand relationship amongst higher education courses has relevant implications at the economic, social, and political level. For instance, an analysis and possible adjustment of the number of vacancies to the demand level has major economic impact on returns to schooling in two ways. If more vacancies are open to correspond to the demand and the employment does not increase, this will result in additional expenses to the government. On the other hand, if the labor market is in need of specialized people and the number of candidates is lower than the total number of vacancies available, this might result in the forceful recruitment of foreign qualified people. At a social level it is important to consider the personal and professional self-fulfillment of highly skilled people since they play an important role in the modernization and evolution of society.

The empirical results show that the unemployment rate is indeed a major factor that influences candidates' 1st choice in the application process since its increase leads to a decrease of the strength index (the number of first choice applications over vacancies) for a given course. Six different robustness exercises – by type of higher education degree, NUTSII, gender of the candidates, gender of the graduates, fields of study and formation, and by the specific field of health related courses - were performed and most of them show the same results for the signs of the coefficients.

The additional conclusions resulting from a deep analysis of the higher education sector, may contribute to intensify the debate by the Portuguese government and higher education institutions about effective education policies, which can contribute to continuous quality improvement. Ultimately, the design of education policies to respond to students and labor market expectations and to improve their satisfaction levels with the institutions and the higher educational system in general will have a positive impact at a political and economic level.

This study is organized in the following way: Section 2 consists of a literature review of our topic, especially the previous research focused on the Portuguese economy. In Section 3, data is described in detailed; in Section 4 the methodology is defined, and in Section 5 the empirical results are analyzed and we perform several robustness exercises to our original estimation. Finally, in Section 6 conclusions are drawn.

2. Literature Review

In this section, we discuss the literature review on this topic. The first sub-section focus on the international perspective and the second on the Portuguese perspective.

2.1. The International Perspective

For countries in the World other than Portugal, the study of the relationship between unemployment and higher education choices is rare. We have found three studies that relate more directly to our own - Varga (2006), Gajderowicz *et al.* (2014), and Goulas and Megalokonomou

(2019). Varga (2006) analyses the determinants for Hungarian students applications' higher education institutions, finding a significant role for the labour market and probabilities of admission. In this case the labour market variable is the expected wages, not unemployment. They also find that there is a tendency to attribute a higher weight to the expected wage factor on their first choice of application (both institution and field of study) and on the other hand give more importance to the probability of admission on the last choice of application. Gajderowicz *et al.* (2014) analyse social and economic determinants in Poland for the decision to which higher education institution to apply. Amongst the economic determinants, the authors consider wages and probabilities of employment. They found that economic determinants are not significant but social determinants play a very important role in determining the choice of higher education institutions. Goulas and Megalokonomou (2019) study the effect of the business cycle on the choice of students' degree in Greece. To study the impact of employment prospects for each course, the authors use a degree-specific job insecurity index (based on 2006 data) and also youth unemployment, which is not degree-specific. The increase in youth unemployment led to an increase in college applicants but a high job security index (for each course) decreases the number of applicants.

2.2. The Portuguese Context

The literature about the state and evolution of the Higher Education sector in Portugal through the years has been growing. One of the most explored subjects is returns to schooling[1]. Different authors have contributed to the notion that even though returns to schooling have fluctuated, particularly on the last 30 years, they have been high in Portugal when compared to other European Union countries (Vieira, 1999; Hartog *et al.*, 2001; Portugal, 2004; Alves *et al*., 2010; Sousa *et al.*, 2015; Campos, 2017).

A less explored topic is the analysis of supply and demand in tertiary education in order to understand its determinants and what measures can be taken by policymakers in order to promote improvements.

Correia *et al.* (2002) analyse the implications of the emergence of Private Higher Education institutions in Portugal and if its' powerful initial purpose of generating more regional and

[1] Not only for Portugal, but for other countries as well.

disciplinary diversity was indeed achieved. The periods from 1980 to 1999 and from 1992 to 1999 were considered for the analysis and the variables that the authors use are the number of vacancies in each district and each region and the number of vacancies for each course, having as point of reference the number of vacancies of the private (45875) and public (42898) sector for the academic year of 1998/1999. In their study, these authors use some variables that include three of our variables: vacancies, territorial units (NUTS), and field of study.

To evaluate the contribution of the public and private Higher Education institutions to improve their regional distribution, the authors started by analysing the total distribution of vacancies across the country and noticed a large asymmetry between the region of Lisbon and Tagus Valley (43.7%) and regions like Alentejo, Algarve, and the islands of Madeira and Azores (7.7%). Also, the districts of Lisbon and Oporto gather more than 50% of all vacancies and the number of vacancies of Lisbon more than double of the ones from Oporto. The authors verified that public Higher Education institutions have a distribution that extends to every district, while the private ones do not exist in three, and that the majority of vacancies of the private Higher Education institutions are offered in the most populated districts like Lisbon and Oporto, which turn out to exceed the offer of vacancies by the public institutions in the same districts. This led Correia *et al.* (2002) to conclude that there is a lower discrepancy in the distribution of the vacancies offered by the public Higher Education institutions than the ones from the private sector. Regarding the evolution of the number of vacancies available in Higher Education institutions from each sector, a higher increase is noticed (14770) in the private sector contrasting with the public one (14082), for the academic year of 1998/99 when compared with the academic year of 1992/93. Nevertheless, the authors once again notice that the increase in vacancies from the public sector shows a greater balance amongst the different regions, contrary to the observed in the same increase in the private sector. Additionally, Correia *et al.* (2002) also compared the number of vacancies in public Higher Education institutions in different disciplinary areas (Teacher training, Arts, Social Sciences, Commerce and Law, Sciences, Architecture, Health and Social Security, Services, Others) between the academic year of 1992/93 and 1998/99. Once more, the authors could see that the diversity of offer is higher among public Higher Education institutions and that in the private sector the priority goes to areas such as Social Sciences, Commerce, and Law, while areas like Architecture and Engineering are not given the same relevance despite of the political will of the government to give them priority. For the aforementioned period, general stability is shown in the variation of the number of vacancies across all disciplinary areas in the public sector (e.g. Social Sciences, Commerce, and Law varied from 25.71% to 27.16%), whilst in the private sector there are bigger differences in values (e.g. Social Sciences, Commerce, and Law varied from 60.66% to 47.80%). Correia *et al.* (2002) state that these discrepancies in values are due to the private sector's higher dependency on the students' fees leading those institutions to attribute more value to their patterns of demand.

In this matter, Correia *et al.* (2002) projected that a surplus of vacancies in both sectors relative to the demand would continue, since the birth rate had been decreasing on the last two decades before the year of their study.[2] Overall, the authors concluded that neither market demands nor higher institution's responsiveness to external demands were the reasons behind the expansion and diversification of both sectors in the Portuguese Higher Education system, but rather the strategies created by the institutions and the creation and development of new areas of study. After the revolution of 1974, an excess in demand emerged that the public sector could not meet, thus the private sector took advantage of this opportunity but focusing particularly in low cost areas. Given this scenario, there were candidates willing to take more chances regardless of the cost or study programme, leading to very difficult market regulation. The private sector developed with very little control which compromised its quality and correspondence to market needs. It is the authors' belief that the perceptions of the outside information addressed to influence and anticipated demand were the major influence that led to the increase of Higher Education's offer and not the expectations and needs of the industrial, service, and educational organisations.

Cardoso *et al*. (2007) perform an evaluation of the publics' perception and confidence on the changes promoted by the Bologna process by analysing its impact on the demand for Higher Education courses. The authors focus on the academic years of 2003/2004 to 2006/2007, which only includes two of the 12 academic years we will explore, from 2004/2005 to 2015/2016. The authors use count data regression analysis with very similar variables to some of the ones we will use, such as: "number of applicants who placed that program in that institution among their choices (irrespective of its ranking, from first to sixth)"; "number of applicants who place that institution and program as their first choice", and "number of vacancies available at each program in each of the two stages of the application process". Notice that in our work we also consider the total of

[2] In a previous study by Amaral and Teixeira (2000) a decrease of the number of candidates applying to Higher Education is equally noticed.

candidates for all six options and that the number of vacancies will only be relative to the first stage of the application process. Ultimately, the authors concluded that those courses that underwent the restructuring of the Bologna process revealed a higher demand, when compared with other courses that did not, and also an increase in demand in courses that were converted into integrated master degrees since they didn’t suffer changes in its total duration.

Tavares *et al.* (2008) based their study on more than 59000 questionnaires directed to students enrolling for the first time in the academic year of 2004/05, which inquiries about their preferential institutions and study programmes as well as the reasons behind it. The number of questions included in the questionnaire was 32 under the binary, multiple choice, nominal and ordinal scales format. Optical recognition was the method by which the collected data was read, with the resulting production of a text ASCII file for each higher education institution, which was consequently organized by school year and a conversion was made to the Statistical Package for Social Sciences (SPSS) format.

Overall, their study uncovers results for course characteristics, which may potentially help students to make their choice regarding courses. Characteristics such as ‘academic efficiency’, ‘characteristics of the syllabus’ and ‘innovative character’ are not very relevant to the decision. The second mentioned determinant (practical strength of the syllabus) obtains a higher result than theoretical strength, and the last determinant (‘innovative character”) obtained the lowest result of all (1.6%). It became clear to the authors that the main sources of information consulted by the students were not the ones provided by official entities, like the Ministry of Education or Higher Education institutions, but rather the opinions of relatives, friends, and colleagues. “The influence of gender in students’ preferences determinants” was also taken into consideration, revealing a slight discrepancy in the valorisation of ‘vocation’ (higher results for women), of ‘employment’ (higher results for men), but also in the influences that have a major impact on the choice of the course, where the authors found that men are more likely to value more the opinions of friends rather than their families’ opinions, contrary to women. The concepts of masculinity and femininity are typically associated to a certain job by students, which the authors believe might explain the predominance of women in educational sciences, social, and human sciences and in health, contrary to men, which show higher predominance in sciences and engineering. Another variant analysed by Tavares *et al.* (2008) was the impact of the family cultural background and its economic capital, which undeniably play a role in the consideration of entering higher education, the choice of course, and the higher education institution. In the course selection, students with higher cultural and economic capital were more likely to choose highly reputed courses, like medicine and law, as well as prestigious institutions, in contrast with students with lower cultural and economic backgrounds, which opted for "education and teacher training programs" and polytechnics as their educational institution of choice. When it comes to the Portuguese student's selection process of course and/or higher education institution, Tavares *et al.* (2008) found 'Vocation' (39% of all answers) to have a greater weight on course choice than 'employment prospects' (25.2% of all answers), bearing in mind the subjectivity of the concept of vocation and the lack of certainty involved with some existing data (or lack of), which conditions the understanding of employment prospects.

Simões and Soares (2010) apply a quantitative design technique with data collected from the University of Aveiro, mentioned in the article as ABC University. The authors developed a questionnaire with the focus of understanding the information sources and the choice factors that students rely on when applying to higher education institutions. Important issues are the level of both relevance and useful content of information that applicants search for and in what ways it influences their higher education institutions' preferences for application. Three sections were included in this questionnaire: 1) Individual background - age, gender residence; final secondary grade, and field of study such as: arts and humanities, engineering and computer sciences, health studies, sciences, and social studies; 2) Information sources - ranking a list of the three sources most used: interpersonal, marketer controlled, the consumers' direct inspection of the good or service, third party independent; and a selection of three more that could improve ABC University's appeal to candidates such as improvements on the University's website, organization of on-campus visits, organization of events of cultural, scientific or sportive nature aimed at secondary schools, and several other promotional events at secondary schools or at the university itself; and 3) Choice factors – academic reputation, geographical proximity, guidance from vocational advisors or teachers, and personal influences. The gathered data was analysed through descriptive statistics and test statistics in order to assess existing differences among the groups. The majority of respondents were female (54.8%) and the most common age goes between 17 and 19 (86.3%). The area of study that gathered the higher percentage of students was engineering and computer sciences (33.3%), while the one with the least percentage of students 'vote is arts and humanities (10.7%). In their work geographical proximity, related to our territorial unit variable, was the choice factor that revealed to be the most important to students, as it showed the highest percentage of students that ranked it first in terms of relevance (45.5%), followed by academic reputation (24.5%). In addition, the majority of the students come from the northern area of Portugal (91.7%) and chose ABC University as their choice in first place.

In the Portuguese higher educational context, Machado *et al.* (2011) conducted a survey for continental Portugal and the islands of Madeira and Açores for 13000 undergraduates students with the goal of providing data to analyse the level of satisfaction and its's corresponding aspects with their higher education experience. A comprehensive student success model was used to interface the collected data, aiming to make the identification of the strengths and weaknesses of the educational system. Also, the data was analysed through methods such as descriptive univariate, bivariate, and multivariate analysis. As already metioned, the authors focused on students' satisfaction with the country's education institutions by measuring the perceptions of importance and satisfaction with academics (includes relevance of courses and academic advising), academic support (quality of the facilities and technological resources available), personal growth (expectations at the personal and academic level), and processes and services, finding that "the importance attributed to each of the surveyed factors is always higher than the level of students' satisfaction with those same factors." Overall, the level of student satisfaction with the aforementioned factors will indeed influence the "course of studies, employability and social prestige of the course" (Machado *et al.*, 2011).

After analysing all of these studies we realised that none of them have established a relationship between course demand and graduates' unemployment. This leads us to the research question of our work: what are the determinants of course choices by candidates and in what way labour market expectations (specifically, the information about the unemployment rate of each course) is influential in these choices? An analysis of the supply and demand of Higher Education courses seems to be currently inexistent for the Portuguese Higher Education system, which is why we decided to consider a variety of data, concerning each course and for every Higher Education institution. In addition to the listing of several courses from different degrees and respective organic unities within a period of 12 years, and their corresponding number of vacancies and field of study, several other variables will be studied in this work. Regarding graduates, we consider total number, number by gender and the number of unemployed. For candidates, the gender, the district with highest number applications, the number of candidates for all six options of application, and finally the number of candidates for the course and institution pair chosen as the first option of application, will also be taken into consideration. Also, the territorial unit to each the Higher Education institution belongs to are included as well as their corresponding Gross Domestic Product (GDP).

3. Data

3.1 Data Sources and Treatment

The data collected in this study includes a final selection of 4303 course codes (in the database this variable is designated as **Course_Cod**) and names (in the database defined as **Course_Nam**), and respective degree name (**Deg_Nam**) and abbreviation (**Deg**), which includes six categories: Undergraduate degrees (before and after the Bologna Process), Masters degrees (integrated Master and stand alone Master), and PHDs. These courses were also classified by: (1) 9 major groups - Agriculture; Arts and Humanities; Services; Engineering, Manufacturing, and Construction; Health and Social Protection; Education; Social Sciences, Commerce and Law; Sciences, Mathematics and Informatics; and Unknown or not specified -, which corresponds to the CNAEF (National Classification of Fields of Study and Formation) at one-digit level; (2) 23 fields of study, which corresponds to the CNAEF at the two-digit level, and (3) 85 areas of education and formation, which correspond to the CNAEF at three-digit level. In the database these classifications are designated as **CNAEF_1D** (1-digit CNAEF area code), **CNAEF_1D_Nam** (name of the 1-digit CNAEF area), **CNAEF_2D** (2-digit CNAEF area code), **CNAEF_2D_Nam** (name of the 2-digit CNAEF area), **CNAEF_3D** (3-digit CNAEF area code), **CNAEF_3D_Nam** (name of the 3-digit CNAEF area). See Table A1 in the Appendix. Moreover, we have 134 Higher Education Institutions (HEI) and 362 organic unities codes (designated as **Org_Unit_Cod**) and names (designated as **Org_Unit_Name**). Thus, our cross-sectional unit of observation is a **Course** offered at institution **HEI/Org** for a degree **DEG**. We have data for 12 years, which makes this a panel with a total of 51636 observations.

A database for the period 2005 to 2016 was created which compiles information gathered from several sources, including the Directorate General for Higher Education - Direção-Geral de Ensino Superior (DGES[3]), Directorate General of Education and Science Statistics of the Ministry of Education and Science - Direção-Geral de Estatísticas da Educação e Ciência (DGEEC[4]);

[3] http://www.dges.gov.pt
[4] http://www.dgeec.mec.pt

PORDATA[5], and the National Statistical Institute - Instituto Nacional de Estatística (INE[6]). The data collected from each source is detailed bellow.

- **DGEEC**

Data for the number of unemployed graduates registered in the Institute for Employment and Vocational Training (Instituto do Emprego e Formação Profissional, IEFP[7]) in June of 2016 (last year available) determined our course selection, since only courses that had data for unemployment were included in the database as well as their corresponding organic unity within each higher education institution. It is relevant to mention that, even though the registered unemployment, provided by the IEFP, does not represent the total level of national unemployment, which is published by the INE, the database from IEFP offered a greater level of detail related to unemployment of graduates, which was why we chose it. In bold we have the designations of the variables in the database.

- Total of unemployed graduates by year, course and by Higher Education Institution, from 2005 to 2016 (**Unemp**).

This database included data between 1984 and 2016. The period from 1984 to 2004 was not included in the database because it did not offer as much detail as for 2005-2016. We have computed the unemployment rate as explained below:

- Unemployment rate *per* course *per* year (**UR**) - we calculated the ratio of the total number of unemployed *per* course and Higher Education Institution over the total number of graduates *per* course and Higher Education Institution, in each academic year from 2005 to 2016[8] . The total number of graduates was also available at this source.

We have also obtained from this source:

[5] https://www.pordata.pt/
[6] https://www.ine.pt
[7] https://www.iefp.pt/
[8] The total number of graduates for each course for the year 2016 was only available in DGES, which allowed for the calculation of the unemployment rate for this year.

- **Vacs -** Total number of vacancies for each course for Private Higher Education Institutions, which complements the variable “Vacs” defined bellow, which had data for Public HEI.

- **DGES**

From this source we gathered information regarding:

- **Tot_Cand -** The total number of candidates for each course in all six options of application.
- **Cand_1stOp -** The percentage of candidates that put the course as their first option of application.
- The percentage of male **(Cand_Male)** and female **(Cand_Female)** candidates.
- The percentage of candidates **(Cand_Dist_Per)** from the district of origin (**Cand_District**) with the highest number of candidates for each course. From this variable, we built a dummy variable (**Dummy_dist**): 1 if the district of origin with the highest percentage is in the same NUTSIII as the higher institution for which the candidate applied; 0 otherwise

The previous first two variables are good proxies for course demand. Additionally, from this data source we also have:

- **Tot_Grad** - Number of graduates by gender **(Male_Grad, Fem_Grad)**
- **Vacs -** The total number of vacancies of Public Higher Education Institutions, *per* course, in the first phase of the National Contest for Admissions to Higher Education (Concurso Nacional de Acesso ao Ensino Superior, CNAES) from 2005 to 2016.

Our variable of interest is the strength index (**INDF**), a ratio between the number of 1^{st} option candidates and the number of vacancies, for a given course associated with a specific organic unit from a Higher Education Institution. This is the best variable to represent demand instead of **Tot_Cand** or **Cand_1stOp,** since it is ultimately the most inclusive amongst the three and the one calculated and used by the universities.

- **PORDATA**

We have used the Nomenclature of Territorial Units for Statistics (Nomenclatura das Unidades Territoriais para Fins Estatísticos, NUTS) of 2013, at the second level **(NUTSII)**, which includes 7 territorial unities, and also at the third level **(NUTSIII)**, which includes 25 territorial unities, to classify in terms of geographical/regional location the organic units corresponding to each course. We have taken the classification of NUTSII and III from PORDATA.

- **INE**

To assess the economic contribution of courses' supply and demand to a geographical region, we have taken from INE the following variables, at the NUTS II level, from 2005 to 2016:

- **GDP_pc -** Gross Domestic Product (GDP) *per capita*, expressed in PPS (purchasing power standards) and in percentage of UE28 (UE28=100).

In order to measure population (quantity) effects over the candidates' district of origin, we have also taken from INE our last independent variable, Population Density (Total Population/km^2), considered at the NUTS III level, from 2005 to 2016. This variable is designated by **Dens_Pop** in the database.

3.2 Methodological Concerns

Due to changes that were caused by the Bologna process, a total of 24.63% courses changed name, sometimes academic degree, or even organic unit, during our period of analysis. In order to simplify the presentation of the results and to provide a better analysis of the evolution of each course over time, all courses of the same academic degree, same organic unit and/or Higher Education Institution, and same name, were merged, leading to the prevalence of the most recent designation and academic degree over the others. In the particular case of first cycle courses with those characteristics, the information for each bachelor and undergraduate degree that existed before the Bologna Process was implemented were added to the data of the adapted undergraduate degrees. Even though the decision to apply this process was easy with most courses, there were other cases that required a deeper and time-consuming research of information, in order to provide any indication of a course that had its name or organic unit changed. Since there was no specific official source with this type of information compiled and organized, the confirmation of such merges could only be obtained through the consultation of DGES and/or the websites of the Higher Education Institution.

For the cases affected by the referred changes, when the most recent course designation or institution was a result of the Bologna Process: **Cand_Male** and **Cand_Female** have values obtained from the average of the absolute values of each course over the total number candidates for each extinguished and remaining course, *per* year. Also, the district of origin (**Cand_District**) chosen in such cases is the one with the highest absolute values amongst all of the “versions” of the prevailing course. For the variable **Cand_Dist_Per** though, the final values exhibited in the database are the result of the average of the absolute values of the district with the highest number of candidates over the total number of candidates of each course with the same correspondence.

The process of adding data from several sources into a single database was sometimes cumbersome due to several factors such as the existence of a great variation of courses and fields of studies but also the need to carefully confirm and review particular cases, when discrepancies of nomenclature were detected for the same course amongst different sources (e.g. different codes for a course with the same name and institution). It is also relevant to point out the difficulty in accessing data in a friendly format from some public entities which implied inserting, in the case of some variables, each entry manually, thus requiring also a careful review of possible errors on the introduction of those values. For other variables, we used text mining techniques, in particular the software Python.

Table 1 shows descriptive statistics of our database. Furthermore, Tables A1 to A11 can be found in Appendix with more detailed statistical information.

Table 1 - Descriptive Statistics

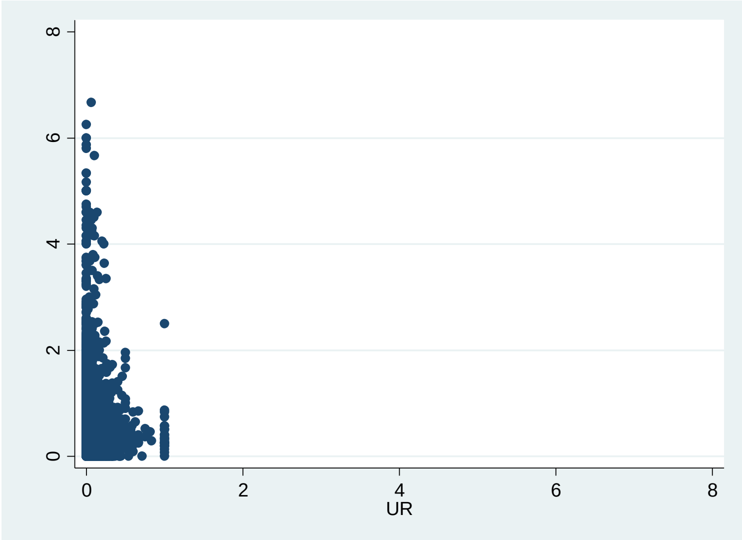
Variable	Observations	Mean	Standard	Minimum	Maximum

			Deviation		
Dens_Pop	51636	522.1679	400.5806	13.9	949.7
Vacs	13818	49.62028	41.66212	0	550
Unemp	51350	0.8444985	2.14437	0	49
UR	51350	0.0477862	0.1346974	0	8
Emp	51636	1256.175	426.7426	96.294	1732.233
GDP_pc	51636	81.15595	20.71101	61.8	118.5
Tot_Cand	11004	225.4008	259.6363	1	3426
Cand_1stOp	11004	17.95602	11.86981	0	100
Cand_Male	11004	43.09678	23.65747	0	100
Cand_Female	11004	56.9163	23.65836	0	100
Cand_District	11004	9.872319	4.838796	1	20
Cand_Dist_Per	11004	47.01645	19.64452	9	100
Dummy_dist	11004	0.8178844	0.3859573	0	1
Tot_Grad	29298	27.23292	37.5545	1	709
Male_Grad	29298	10.78664	19.22353	0	415
Female_Grad	29298	16.44628	24.92966	0	478
Indf	10947	0.4727495	0.4661556	0	6.666667

3.3 The Relationship between the Strength Index and the Unemployment Rate

In order to "anticipate" how the unemployment rate and the strength index are correlated, several scatter plots were obtained. The strength index (indf) is plotted on the *y*-axis of the scatter plot and the unemployment rate (UR) on the *x*-axis. On Figure 1 we see that a negative relationship between the unemployment rate and the strength index can be expected. The strength index variable reaches its' maximum value when the unemployment rate variable is at its' lowest.

Figure 1 - Unemployment Rate and Strength Index Relationship



Figures 2, 3, and 4 show the relationship between our main variables of interest by degree (degree 2, 3 and 5, respectively). After comparing these three figures, we may foresee that a higher unemployment rate will be associated to degree 5 courses (Integrated Masters) as opposite to degree 3 courses (Undergraduate Degrees after the Bologna Process) that show a tendency to lower unemployment rates and therefore higher demand (strength index). Regarding degree 2 (Undergraduate Degrees before the Bologna Process), the results are between degree 3 and 5.

Figure 2 - Unemployment Rate and Strength Index by Degree 2

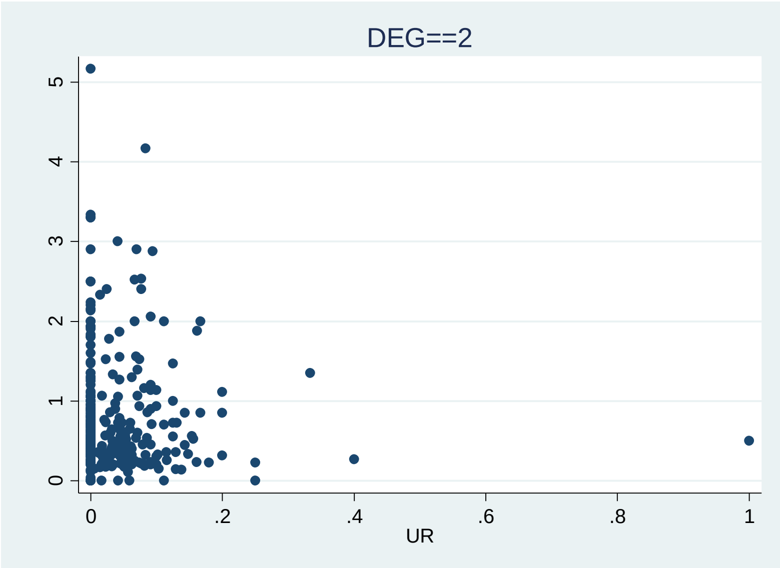


Figure 3 - Unemployment Rate and Strength Index by Degree 3

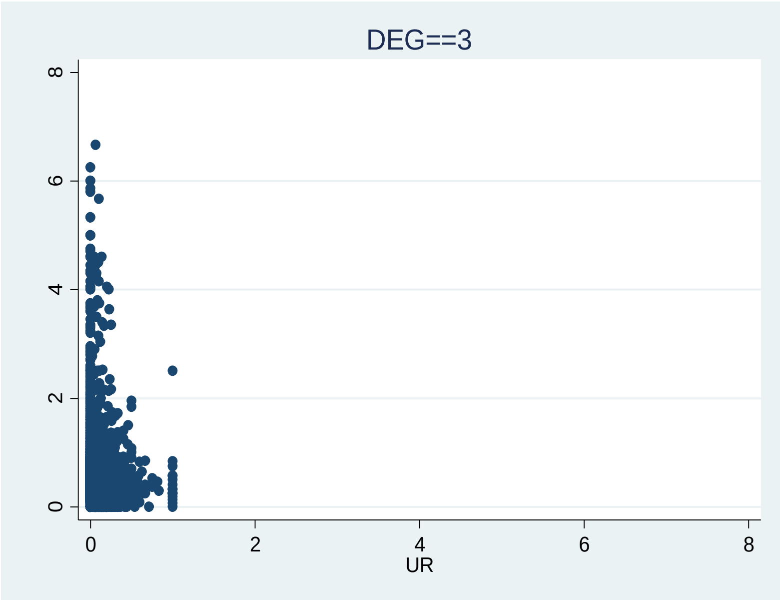
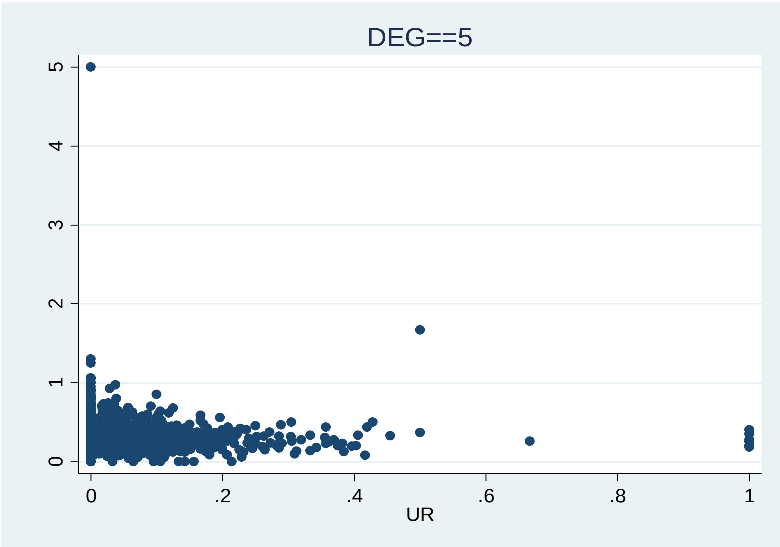


Figure 4 - Unemployment Rate and Strength Index by Degree 5



When it comes to the candidates' gender, even though there seems to be a higher number of courses chosen by the female candidates (Figure 5), the demand for courses with low unemployment rates is mostly prevalent amongst male candidates (Figure 6). This might mean that in general the majority of male candidates tend to choose to apply to courses with low unemployment rates.

Figure 5 - Unemployment Rate and Strength Index by Female Candidates

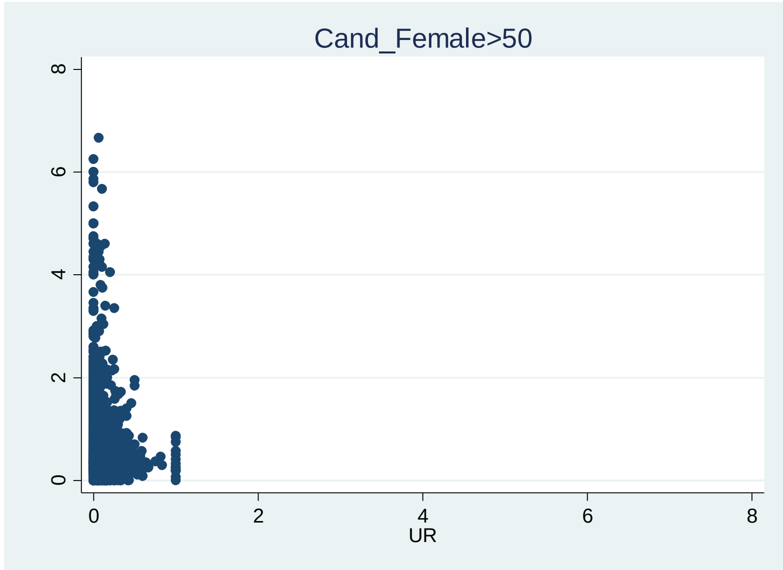
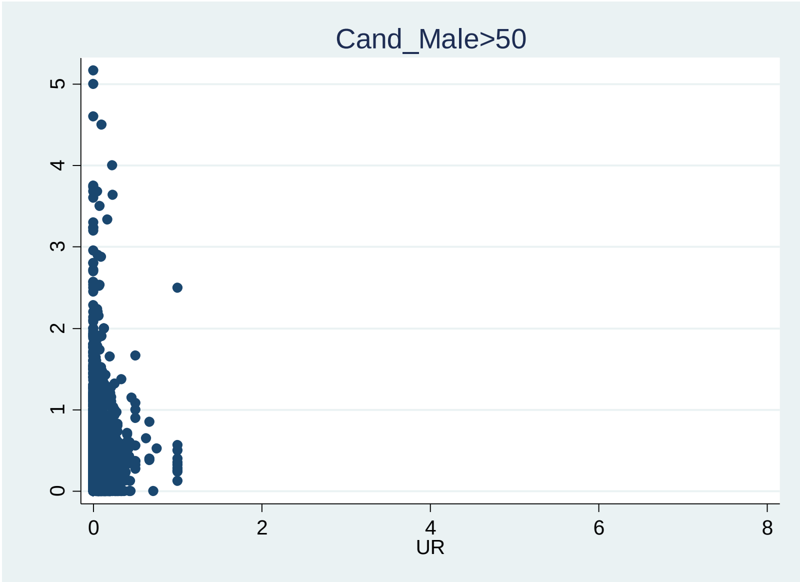


Figure 6 - Unemployment Rate and Strength Index by Male Candidates



A lower dispersion is associated to courses where the number of female graduates surpasses the number of male graduates (Figure 7), as the majority of graduates (mostly women), tend to graduate from courses with low unemployment rates. In comparison to this, there are more courses with a medium to high unemployment rate when the number of male graduates is higher (Figure 8).

Figure 7 - Unemployment Rate and Strength Index by Female Graduates

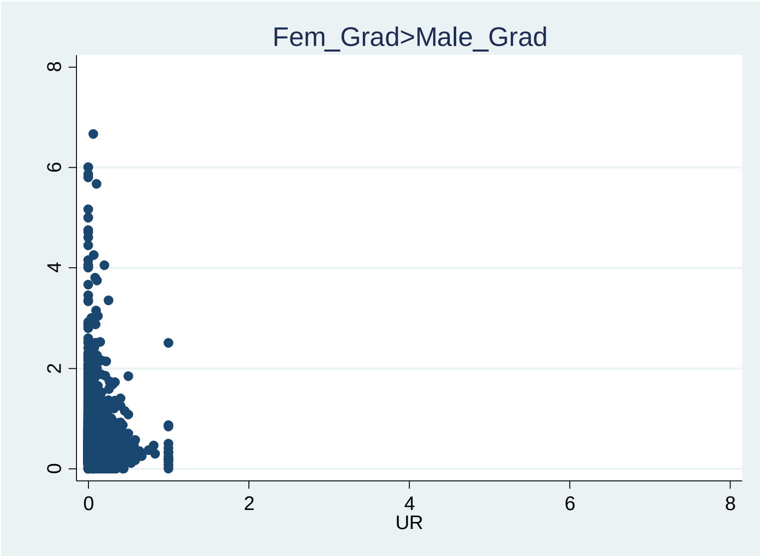
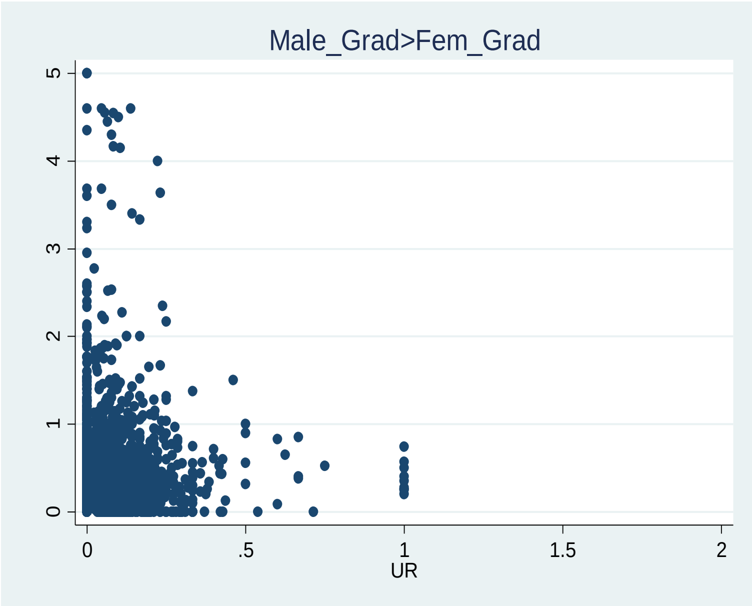


Figure 8 - Unemployment Rate and Strength Index by Male Graduates



Among fields of study, certain differences can be observed as well. The ones that seem to show a lower number of courses are CNAEF_1 (Figure 9) and CNAEF_6 (Figure 10). For the Education courses (CNAEF_1) there is a residual portion that reaches a peak in the strength index, but the majority of courses are situated at a low position for both variables. For the Agriculture courses (CNAEF_6), there are more courses with a medium level of unemployment rate with low strength index and residual courses with a peak of unemployment rate.

Figure 9 - Unemployment Rate and Strength Index by CNAEF_1

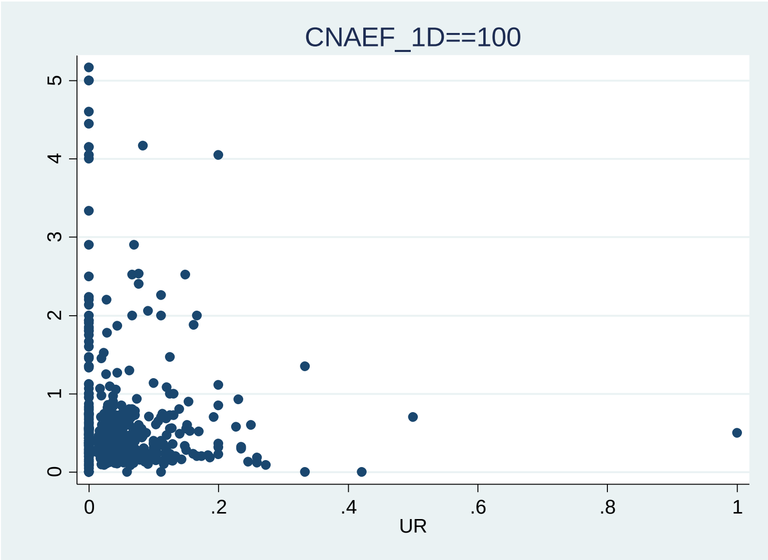
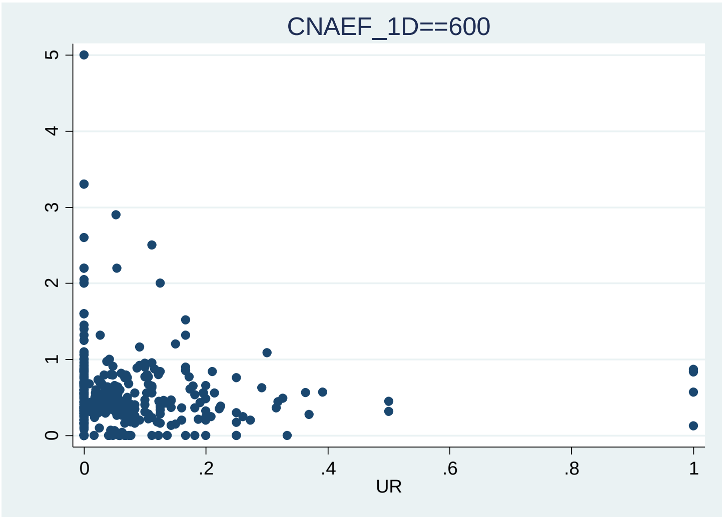


Figure 10 - Unemployment Rate and Strength Index by CNAEF_6



The diagrams that appear to show a higher concentration of courses correspond to the CNAEF_2 – Arts and Humanities (Figure 11), CNAEF_3 – Social Sciences. Commerce and Law (Figure 12) and CNAEF_5 - Engineering. Manufacturing and Construction (Figure 13).

For the first one, most courses are situated at a low strength index and low unemployment rate position, while for the second field of study courses are more present at a low/medium strength index and low unemployment rate position in the diagram. CNAEF_5 shows a higher dispersion of courses with medium/high unemployment rates and low strength index.

Figure 11 - Unemployment Rate and Strength Index by CNAEF_2

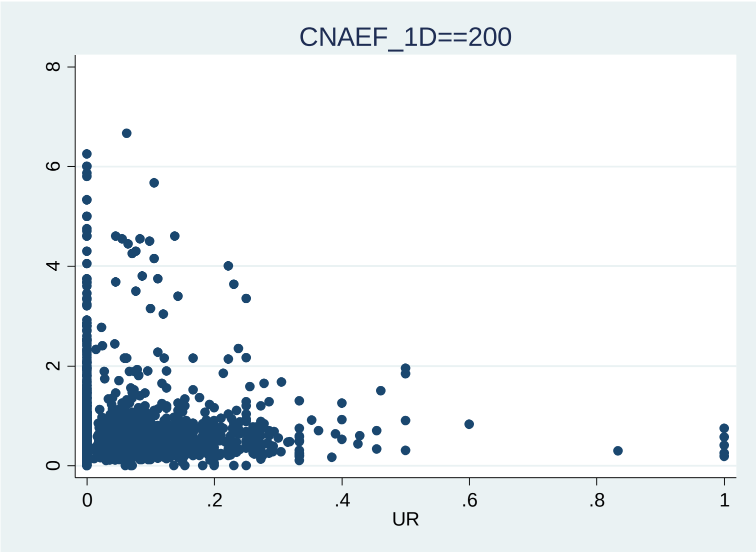


Figure 12 - Unemployment Rate and Strength Index by CNAEF_3

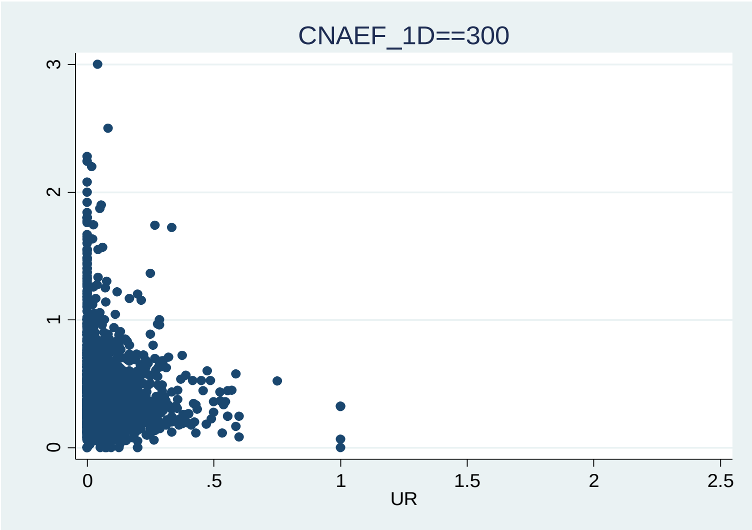
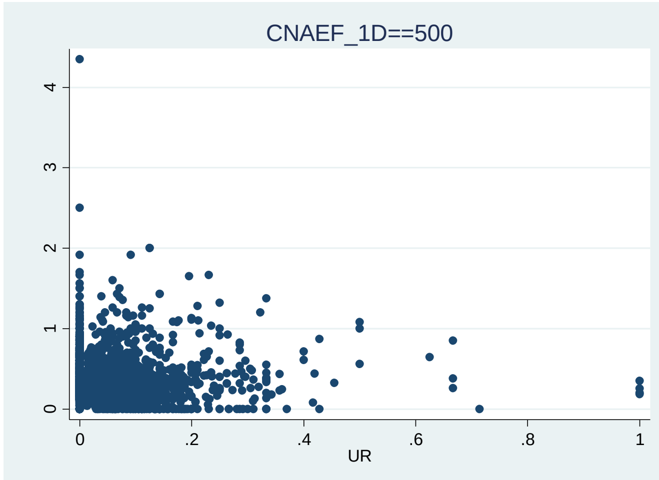
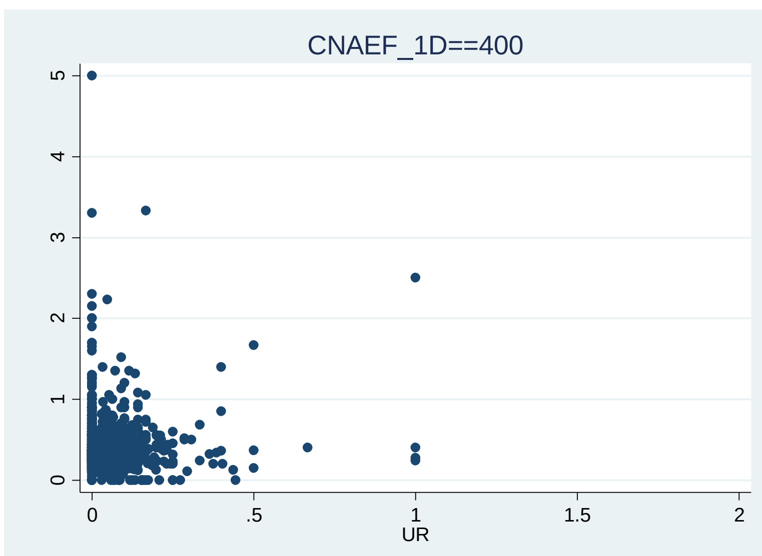


Figure 13 - Unemployment Rate and Strength Index by CNAEF_5



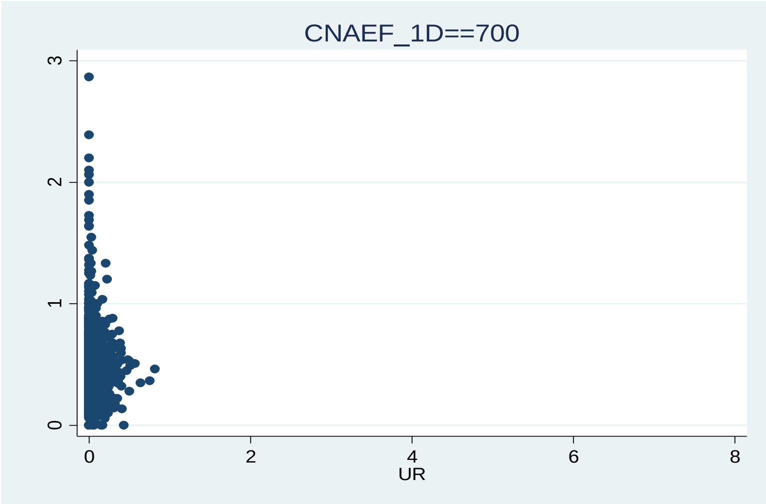
Most CNAEF_4 – Sciences, Mathematics and Informatics (Figure 14) follow the general trend of both low unemployment rates and strength index, with a great concentration in the left inferior position of the diagram.

Figure 14 - Unemployment Rate and Strength Index by CNAEF_4



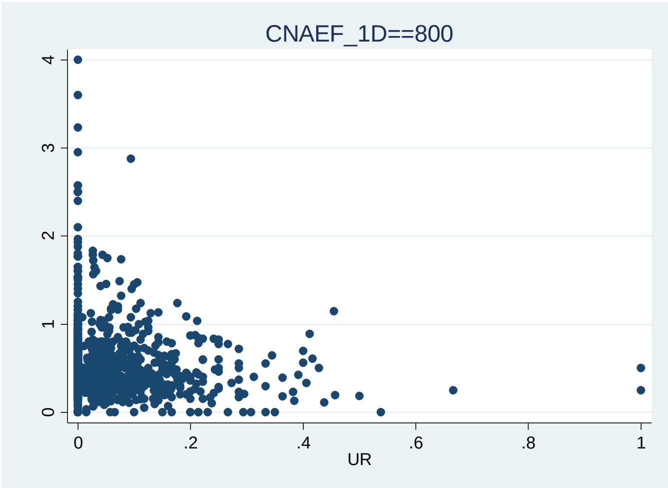
An interesting disposition of dots is seen for CNAEF_7 - Health and Social Protection courses in Figure15, as this is the field of study that shows the lowest levels of unemployment rates associated with high levels of the strength index.

Figure 15 - Unemployment Rate and Strength Index by CNAEF_7



Lastly a pattern of disposition very similar to the one for the field of study of Engineering, Manufacturing and Construction (CNAEF_5) is seen for CNAEF_8 - Services (Figure 16), with perhaps a lower number of courses.

Figure 16 - Unemployment Rate and Strength Index by CNAEF_7



4. Econometric Methodology

In order to analyze the relationship between the strength index (**INDF**), the ratio between the number of 1st option candidates and the number of vacancies, and all its above mentioned determinants, in particular the unemployment rate (**UR)**, we use standard panel linear regression models.[9] We add to the list of covariates not only its contemporary value but also one and two lagged terms. The idea is to capture effects from a covariate to the strength index that may persist for more one or two academic years. As explained in the Data section, our object of interest is a given course "*i*" associated with a specific organic unit from a Higher Education Institution observed in year "*t*". Here "*t*" is read as academic year *t/t+1*. The general model that combines both course cross-sectional and time-series dimensions can be written as follows:

$$\mathbf{INDF}_{it} = \beta_0 + \beta_1 \mathbf{UR}_{\boldsymbol{it}} + \beta_2 \mathbf{UR}_{\boldsymbol{i,t-1}} + \beta_3 \mathbf{UR}_{\boldsymbol{i,t-2}} + \beta_4 \boldsymbol{X}_{\boldsymbol{it}} + \beta_5 \boldsymbol{X}_{\boldsymbol{i,t-1}} + \beta_6 \boldsymbol{X}_{\boldsymbol{i,t-2}} + u_{it}$$

For cross-sections $i = 1, \dots, n$ and periods $t = 3, \dots, T$

$\boldsymbol{X}$ is a K_X-dimensional vector that represents all the explanatory variables, except **UR.**

$\boldsymbol{\beta_0}$ is the model's intercept. The $\boldsymbol{\beta_j}$**'s** are the partial slopes associated to the *j*th regressor, after controlling for all other terms. That is, ceteris paribus, a unit change in the *j*th regressor implies a change of $\boldsymbol{\beta_j}$ units of the strength index (points of ratio between the number of 1st option candidates and the number of vacancies).

u_{it} is the model's error term and includes all unobserved components that also affect $\mathbf{INDF}_{it}$.

In linear regression models using panel data is important to determine the statistical properties of the potential unobservable course-specific effects. These individual effects α_i are assumed to be time-invariant and fixed or random, such that it is typically decomposed as:

$$u_{it} = \alpha_i + \varepsilon_{it}.$$

The appropriate statistical tests were conducted to infer whether or not there exists course-specific effects and in the case of existing if are of type fixed or random. In the case of fixed effects, these are constants that are estimated; in the case of random effects, it is assumed that the unobserved individual effects are distributed as independent $N(0, \sigma_\alpha^2)$ and the idiosyncratic

[9] The estimations of the models were obtained using the software Stata.

disturbances are independent $N(0, \sigma_{\varepsilon}^{2})$. For further details about the estimation and inference of panel data models see, for example, Wooldridge (2006) and Arellano (2003).

In the next sections, we present the results obtained for the estimation of the linear regression panel models for the variable of interest strength index. We only report the estimated coefficients that were found to be statistically significant according to the usual t-tests (z-statistic). The covariates that help to explain the strength index are the unemployment rate, the Gross Domestic Product *per capita*, the population density, and the dummy variable that reports the cases where the district of origin with the highest percentage is in the same NUTSIII as the higher institution for which the candidate applied. In Table 2, we have the estimated pairwise correlation coefficients for all the variables of the panel data models. We conclude that there is a negative correlation between the strength index and the unemployment rate. This is not a causality statement but is interesting in itself to what comes next in the following sections.

Table 2 - Correlation Matrix for the Dependent and Independent Variables

	indf	UR	GDP_pc	Dens_Pop	Dummy_dist
indf	1.0000				
UR	-0.0232	1.0000			
GDP_pc	-0.0203	-0.1003	1.0000		
Dens_Pop	-0.0029	-0.0832	0.6850	1.0000	
Dummy_dist	0.0434	-0.0805	0.2684	0.3846	1.0000

5. Results

In this section, we analyze the results of our benchmark model, as well as the results of the robustness exercises.

5.1. Benchmark Estimation

In order to assess how the strength index is influenced by the unemployment rate, controlling for other variables that are also determinants of the demand for Higher Education courses, we estimated two equations, one using random effects, other with fixed effects, to see which panel data method we were going to use. We specify models that include individual effects (random versus fixed) because we have a strong evidence of its existence. For the random effects model, the LM test has a pvalue of zero meaning that the variance of the individual is not zero. For the fixed effects model, the F test also has a pvalue of zero, which indicates that the individual effects are not all the same and equal to zero. Table 3 gives the results of the random effects estimation, while in Table 4 we have the fixed effects estimation. According to the Hausman test (Prob>chi2=0.0191) we reject the null hypothesis at 5%, hence we proceed with the fixed effects estimator. The fixed effects estimator is also going to be used in the robustness exercises, in the next section.

Table 3 - Estimation Results – Random Effects

indf	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
UR.	-0.1004267	0.0254582	-3.94	0	-0.1503239 -0.0505295
UR_L2.	0.0302431	0.0297479	1.02	0.309	-.0280617 .0885479
GDP_pc	0.001545	0.0005633	2.74	0.006	.000441 .002649
Dens_Pop	-0.0000613	0.0000369	-1.66	0.097	-.0001337 .0000111
Dummy_dist	0.0531792	0.0117614	4.52	0	.0301272 .0762311
_cons	0.3279108	0.0412179	7.96	0	.2471252 .4086964
$Sigma_u$	0.39311345				
$Sigma_e$	0.20737888				Number of obs = 9053
rho	0.78229681	(fraction of variance due to u_i)			

Table 4 – Estimation Results – Fixed Effects

indf	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
UR	-0.0807643	0.0257877	-3.13	0.002	-0.1313148 -0.0302137
UR_L2.	0.0613828	0.0303816	2.02	0.043	.0018269 .1209386
GDP_pc	0.0017486	0.0007264	2.41	0.016	.0003246 .0031726
Dens_Pop	0.0024209	0.0006045	4	0	.0012359 .0036059
Dummy_dist	0.0519754	0.0123183	4.22	0	.0278282 .0761225
_cons	-0.6514853	0.2193438	-2.97	0.003	-1.08145 -0.221514
*Sigma*_u	1.047666				
*Sigma*_e	0.20737888				Number of obs = 9053
rho	0.96229561	(fraction of variance due to $u_{\_i}$)			

In Table 4 we can see that all independent variables are statistically significant in explaining the strength index (indf) since we have a p-value smaller than 0.05 for all coefficients. Specifically, for our variable of interest – the unemployment rate - results also show that at the same period if the unemployment rate (UR) increases by one unit the strength index decreases by 0.08 units. This means that candidates take in consideration the information about each course unemployment rate. The effect to next year does not apparently occur. On the other hand, regarding the unemployment rate with a two-year rate lag, the relationship with the strength index is positive. If in period *t* the UR changes one unit then two years later the INDF goes up by 0.06. This seems to be counterintuitive but since we have two effects (contemporaneous and lagged) we need to estimate the overall effect. This overall effect is the sum of both coefficients which equals -0.02. Hence, the effect of UR on INDF is the expected one. After an increase of the UR, students adjust their preferences and the end effect on the INDF is negative. The interesting piece is that at the beginning they do penalize those courses but later on (two years after) then seem to "forget" and the rejection of those courses becomes less strong.

It is also interesting to see in Table A2 (in the Appendix) that from 2005 to 2016 the total number of unemployed graduates (Unemp) tends to increase over the years, while the average strength index decreases. This relationship becomes more evident when in 2009, 2015, and 2016, the total number of unemployed graduates drops and the strength index increases. Nevertheless, the average unemployment rate increases through the years, with exception of 2009, 2014 and 2016.

GDP *per capita* and population density have the expected (positive) signs. The higher the income level of a region, relative to other regions, the higher the number of candidates in 1st option. The higher the population density the higher the number of candidates, in this case specifically, the higher the number of 1st option candidates (since the number of potential candidates is higher).

Table A2 (in Appendix) shows that both GDP *per capita* and population density averages tend to increase from 2005 to 2016. The exceptions for these patterns are specific time periods like 2008, 2009-2012 and 2013-2016 for GDP *per capita* and 2009 and 2012-2015 for population density. Also, in Table A3 (also in the Appendix) we note that the year with the highest percentage of 1st option candidates is 2014 (21.94%) while the year with the lowest value is 2008 (20.23%).

The district dummy also presents a positive sign, i.e., when we see that the percentage of students from the district of origin with the highest percentage is in the same NUTSIII as the higher institution for which the candidate applied, the strength index increases. This difference in INDF of being in the same NUTSIII is measured to be 5.2%.

In Table 5 we have the top-5 and bottom-5 estimated course-specific effects α_i from the fixed effects model. These are the largest (in absolute value) differences across courses' INDF due to their intrinsic unobserved characteristics because in a fixed effects model, the constant term is in fact $\beta_0 + \alpha_i$. At most, this difference in INDF is equal to 5.8 (4-(-1.8)) points, which can be considered a significant quantity. The maximum value is for the undergraduate course of Music at the Universidade de Aveiro and the lowest for the Masters course of Economics at Universidade do Algarve. Interesting in itself, there is something in Music offered at the Universidade de Aveiro that makes students value a lot. On the opposite, students seem to “dislike” the Undergraduate degree in Economics course at the Universidade do Algarve. A deeper analysis on why this is so is worth considering.

Table 5 – Estimated Course-Specific Effects

Highest		Lowest	
Undergraduate of Music at Universidade de Aveiro	4.013137	Undergraduate of Economics at Universidade do Algarve	-1.802077
Integrated Masters of Engineering of Biosystems at Universidade de Évora	3.825616	Undergraduate of Management at Universidade dos Açores	-1.79664
Undergraduate of Dance at Universidade de Lisboa	3.275786	Undergraduate of Law at Universidade de Coimbra	-1.785825
Undergraduate of Translation and Interpretation (Chinese/Portuguese) at Instituto Politécnico de Leiria	2.779801	Undergraduate of Portuguese and Lusophone Studies at Universidade de Coimbra	-1.777329
Undergraduate of Audiovisual Communication Technology at Instituto Politécnico do Porto	2.655707	Undergraduate of Electric Engineering at Universidade Lusófona de Humanidades e Tecnologias	-1.772844

5.2. Robustness Exercises

In this section, we perform robustness exercises to our benchmark estimations. We perform six different exercises – by type of higher education degree, NUTSII, by the gender of the candidates, gender of the graduates, fields of study and formation, and specific field of health related courses.

5.2.1. Type of Degree

Table 6 shows the results of our estimations by type of degree - undergraduate degree (before the Bologna process), undergraduate degree, after the Bologna process, and integrated Master. Regarding the estimation for the pre-Bologna undergraduate degrees, all independent variables have shown to be nonsignificant. The most obvious reason is the sample size being so small.

A possible explanation is the lack of data, since 2005, the year that our data sample begins, was still a phase of transition from the Bologna process, but phasing out. From our database, pre-Bologna undergraduate degrees represent only 15.3% of total courses, while post-Bologna undergraduate degrees represent 38.9 % in 2016 (see Table A6 in Appendix). In opposition, the estimation for the post-Bologna degrees shows significance for all variables and with the signs and estimated values very close to the benchmark model. 41.9% of courses in our sample are post-Bologna degrees (see Table A6 in Appendix), which can explain the significance of the results.

Note that, even though Bachelor, Master and PhD degrees are presented at table A6 (in Appendix), their respective courses were not considered for neither benchmark estimation or robustness exercises mainly due to their lack of information within our candidates related variables. The estimation for the integrated Masters degrees reveals that only the unemployment rate and GDP *per capita* are significant, with the same signs as in the benchmark estimation as well. In particular, we observe the contemporaneous negative relationship between INDF and UR.

Table 6 – Estimation Results by Degree – Fixed Effects

indf	**Undergraduate Degree (pre-Bologna)**	**Undergraduate Degree (post-Bologna)**	**Integrated Master**
Number of obs	21	8079	946
UR	3.290522	-0.0744521***	-0.147509***
	(0.109)	(0.009)	(0.001)
UR_L2	0.6653958	0.0756605**	-0.0293978
	(0.682)	(0.026)	(0.542)
GDP_pc	0.0297511	0.0017051**	0.00214*
	(0.505)	(0.037)	(0.060)
Dens_Pop	0.0039597	0.0028409***	0.0000643
	(0.619)	(0.000)	(0.942)
Dummy_dist	0	0.0540761***	-0.0029082
		(0.000)	(0.940)
_cons	-4.409432	-0.7272649***	0.108417
	(0.522)	(0.002)	(0.822)

Notes: ***, **, * denotes significance at the 1%, 5%, and 10% levels, respectively.

5.2.2. Regions NUTSII

In this section, we analyse the estimations obtained for each NUTS II, which are presented in Table 7. Only for NUTSII Açores we find non-significant coefficients for all variables. Açores is the least populated NUTII region; hence, the reduced amount of data is not enough to draw significant results. Also, in our database Açores represents only 0.8 % of the total in 2016 (see Table A7 in Appendix).

In the other NUTSII regions where variables have significant coefficients, the sign is the same as in the benchmark estimation. For Alentejo the significant variables were the unemployment rate and the district dummy. For Algarve, it is the unemployment rate with the two-year lag. For Lisboa, it is the GDP *per capita* only. For the Centro region, it is population density and the district dummy. For the Norte region, it is the population density only, and for Madeira it is the unemployment rate and GDP *per capita.* Again, for the cases where UR is statistically significant, a negative sign is obtained. Notice that for 2016, Algarve and Alentejo represent 3.7% and 6.6 %, respectively, of the total NUTSII in our database (see Table A7 in Appendix). On the other hand, Lisboa, Norte and Centro correspond to 32%, 33.1%, and 23.3 %. From 2005 to 2016 the majority of candidates for each course comes from Lisboa (21.6% to 26.6 %), followed by Porto (12.8% to 19.4%), and Coimbra (8.7% to 9.7%). Faro, Leiria, and Évora are candidates' districts of origin with some of the lowest percentages of the total, with ranges of 4.5% to 5.4%, 3.3% to 4.8% and 2.3% to 2.9%, respectively (see Table A4 in Appendix). In table A5 (in the Appendix) we notice that 32% of courses correspond to courses of NUTS III Área Metropolitana de Lisboa and 18.5% to Área Metropolitana do Porto (see Table A5 in Appendix), while only 4.3%, 3.7%, 2.9% and 2% represent Região de Aveiro, Algarve, Alentejo Central, and Região de Leiria, respectively.

Table 7 - Estimation Results by NUTSII – Fixed Effects

Indf	Alentejo	Algarve	Lisboa	Centro	Norte	Madeira	Açores
Number of obs	841	445	2085	2851	2622	68	141
UR	-0.1789568**	-0.0298803	-0.0398491	-0.0899478	-0.0425863	-4.040571*	2.413798

	(0.029)	(0.861)	(0.344)	(0.102)	(0.302)	(0.084)	(0.203)
UR_L2	0.1569734	0.5868642***	0.042045	0.0523134	0.006207	0.6736905	2.498707
	(0.103)	(0.006)	(0.364)	(0.424)	(0.905)	(0.777)	(0.140)
GDP_pc	0.0007517	0.0048513	0.0024678***	-0.0015039	0.0035552	0.0461214***	-0.007717
	(0.858)	(0.125)	(0.002)	(0.591)	(0.217)	(0.007)	(0.621)
Dens_Pop	0.0076322	0.0042976	0.0002824	0.0075675**	0.0032254***	0.0004949	0.0047692
	(0.677)	(0.764)	(0.735)	(0.017)	(0.000)	(0.964)	(0.969)
Dummy_dist	0.0717231**	0.0122771	0.0779891	0.0722669***	0.0244662	0	0
	(0.022)	(0.895)	(0.286)	(0.001)	(0.163)		
_cons	0.171599	-0.2142408	-0.20411	-0.3163732	-1.194292 ***	-2.261462	0.9518235
	(0.736)	(0.878)	(0.782)	(0.289)	(0.002)	(0.443)	(0.942))

Notes: ***, **, * denotes significance at the 1%, 5%, and 10% levels, respectively.

5.2.3. Candidates by Gender

Table 8 - Estimation Results by Gender – Candidates (Fixed Effects)

indf	Male	Female
Number of obs =	2790	6112
UR	-0.0511472	-0.060732**
	(0.329)	(0.030)
UR_L2	-0.1238798*	0.1262638***
	(0.072)	(0.000)
GDP_pc	0.0013853	0.0018595**
	(0.329)	(0.022)
Dens_Pop	0.0033318***	0.0023047***
	(0.005)	(0.001)
Dummy_dist	0.0302076	0.0557952***
	(0.214)	(0.000)
_cons	-1.064598**	-0.5858673**
	(0.021)	(0.013)

Notes: ***, **, * denotes significance at the 1%, 5%, and 10% levels, respectively.

As we can see in Table 8, only for female candidates we had all coefficients with a significant sign (and equal to the benchmark estimation).

In the case of male candidates, we have only population density and the unemployment rate with a two-year lag with significant coefficients. Population density still presents a positive relationship with the strength index, while the unemployment rate with a two-year lag has a negative sign, contrary to the benchmark estimation, but coherent in terms of the overall effect.

In table A3 (in the Appendix) we notice that from 2005 to 2016, the percentage of female candidates (varies from 57.02% to 62.77%) is always higher than the male candidates' percentage (from 37.23% to 42.98%). Also, through this period we see a trend in the increase of vacancies to follow the increase in the number of total candidates (see Table A3 in Appendix). Nevertheless, the number of total candidates always surpasses the number of vacancies for each year, even in the years of 2011 and 2012, where the number of vacancies increased quite significantly from 47460 to 77436, when compared to the average of previous years (approx. 46110).

5.2.4. Graduates by Gender

Table 9 presents the results for estimation regarding graduate students by gender. The results show that for female graduates, all independent variables are statistically significant, and with the same sign as in the benchmark estimation, except for the unemployment rate. Regarding male graduates, both unemployment rates' variables show a negative (significant) sign, which in the case of the unemployment rate with a two-year lag is contrary to the benchmark estimation. This also happened in the previous sub-section, in the case of male candidates. In fact, when we look at the values of both male and female candidates and male and female graduates we see that they have almost symmetric values.

Once again, the percentage of female graduates, which changed from 65.22% in 2005 to 58.57% in 2016, surpasses the percentage of male graduates - 34.78% in 2005 to 41.43% in 2016-, which makes sense for most years, given that there is a correspondence between the initial candidates that eventually graduate (see Table A8 in Appendix).

Table 9 – Estimation Results by Gender – Graduates (Fixed Effects)

Indf	Male	Female
Number of obs =	2537	5590
UR	-0.1157296*	-0.0380561
	(0.057)	(0.189)
UR_L2	-0.1402547**	0.1487012***
	(0.043)	(0.000)
GDP_pc	0.0019342	0.0028111***
	(0.217)	(0.000)
Dens_Pop	0.0019797	0.0021702***
	(0.140)	(0.001)
Dummy_dist	0.0356609	0.0643606***
	(0.178)	(0.000)
_cons	-0.5211896	-0.6711977***
	(0.297)	(0.004)

Notes: ***, **, * denotes significance at the 1%, 5%, and 10% levels, respectively.

5.2.5. Fields of Study and Formation

In Table 10 we present the estimations by fields of study and formation. Only for CNAEF 4 - Sciences, Mathematics and IT and 6 – Agriculture we do not have any significant variable. For CNAEF 1 (Education) the unemployment rate at time *t* and the unemployment rate with a two-year lag are both significant, as well as population density. Population density now has a negative sign, contrary to the benchmark estimation. One possible explanation is the lack of demand for this type of courses by candidates when compared to other fields of studies.

Regarding CNAEF 2 - Arts and Humanities, both GDP *per capita* and population density are significant, with a positive sign. Regarding the field Social Sciences, Commerce and Law (CNAEF 3) the two unemployment rates are both significant, as well as the district dummy. For Engineering, Industry, and Construction (CNAEF 5) the geographical variables – population density and the district dummy are significant. In the case of CNAEF 7 – Health and Social Protection, only the unemployment rate with a two-year lag is non-significant. For CNAEF 8 – Services, only the district dummy is statistically significant.

For 2016, CNAEF 3 - Social Sciences, Commerce and Law represent the largest percentage of fields of study (25.5%), while CNAEF 6 – Agriculture represents the lowest, 2.4% (see Table A9 in Appendix).

Table 10 – Estimation Results by Fields of Study and Formation – Fixed Effects

	CNAEF_1D							
indf	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Number of obs	292	1408	2073	995	1997	322	1143	820
UR	-0.5683***	-0.0546	-0.1151***	-0.1229	-0.0748	-0.0146	-0.1010*	0.0295
	(0.001)	(0.424)	(0.004)	(0.170)	(0.241)	(0.919)	(0.080)	(0.764)
UR_L2	0.5858**	0.0742	0.1794***	-0.0919	0.0869	-0.1421	0.02045	0.1162
	(0.016)	(0.337)	(0.000)	(0.342)	(0.235)	(0.397)	(0.785)	(0.347)
GDP_pc	-0.0064	0.0068***	0.0000	-0.0034	0.0021	0.0066	0.00463***	0.0006
	(0.168)	(0.002)	(0.985)	(0.135)	(0.187)	(0.342)	(0.001)	(0.842)
Dens_Pop	-0.0100***	0.0064***	0.0008	0.0010	0.0036***	-0.0013	0.0030***	0.0022
	(0.005)	(0.000)	(0.435)	(0.570)	(0.006)	(0.850)	(0.009)	(0.417)
Dummy_dist	0.0486	0.0191	0.0465*	-0.0083	0.0713***	0.01928	0.0941***	0.0793*
	(0.410)	(0.635)	(0.065)	(0.833)	(0.006)	(0.737)	(0.000)	(0.070)
_cons	3.6317***	-2.4557***	0.0340	0.2252	-1.3626***	0.2770	-1.1550***	-0.2450
	(0.000)	(0.000)	(0.927)	(0.753)	(0.008)	(0.813)	(0.006)	(0.749)

Notes: ***, **, * denotes significance at the 1%, 5%, and 10% levels, respectively.

5.2.6. Specific Field – Health Related Courses

In this section, we analyze a specific field – Health related courses. Notice that for 2016, Health represents 6.8% (see Table A10 in Appendix) of total fields of study, while Medicine corresponds to 0.5% and Nursing to 1.6 % in our database (see Table A11 in Appendix).

Table 11 presents the results for the estimations regarding Health (general), Medicine, and Nursing courses. All signs are according to the benchmark estimations, except the coefficient of the unemployment rate with a two-year lag, which for the Health courses (in general) and also for Nursing courses, present a negative (significant) sign. Still, the most important fact is that the overall effect is negative.

Table 11 – Estimation Results by Specific Field – Health Related Courses (Fixed Effects)

	CNAEF_3D		
Indf	**Health**	**Medicine**	**Nursing**
Number of obs =	903	40	217
UR	-0.3028144***	-1.941724	-0.4768699**
	(0.000)	(0.462)	(0.023)
UR_L2	-0.2745772**	0.2747917	-1.025443***
	(0.033)	(0.888)	(0.005)
GDP_pc	0.0057713***	0.0010827	0.002866
	(0.000)	(0.329)	(0.419)
Dens_Pop	0.0018954	-0.0001297	0.0018116
	(0.115)	(0.862)	(0.607)
Dummy_dist	0.0915406***	0.0132736	0.1355601***
	(0.000)	(0.542)	(0.000)
_cons	-0.906204*	0.1538504	-0.2645767
	(0.066)	(0.812)	(0.695)

Notes: ***, **, * denotes significance at the 1%, 5%, and 10% levels, respectively.

6. Conclusions

In this work we inferred, using a sample of Portuguese pairs of courses/higher education institutions for the period 2005-2016, if information regarding the unemployment rate by course was a possible determinant of the choice of a higher education degree. Using control variables such as GDP *per capita* and population density of the region where the higher education institution is located, as well as the district of the majority of candidates by course, which are other possible determinants of the demand for higher education courses, we find that the relationship between the unemployment rate and the strength index (the variable we use to be our dependent variable) is negative, i.e., the higher the unemployment rate of that course, the lower the strength index. The strength index is a ratio between the number of 1st option candidates and the number of vacancies. If it has a negative relationship with the unemployment rate, this means that courses that exhibit a higher unemployment rate will have less demand. Other variables such as GDP per capita, population density, and the district dummy have the expected positive relationships with the strength index.

Additionally, we have performed several robustness exercises - by course degree, field of study (including specific fields like Health, Medicine and Education), by NUTSII regions, by male and female candidates and graduates. We have to underline the consistency of almost all results, regarding the relationships (signs) between the explanatory variables and the strength index. Specifically, for our variable of interest, the unemployment rate, in all cases the increase of the unemployment rate has an inverse relationship with the strength index. This means that for the candidates, their first option course selection is indeed influenced by the unemployment rate, and that its increase might be a discouraging factor on including a certain course as their first choice. Some of the most interesting findings were on the estimation results by gender. In both male candidates and graduates, an increase of the unemployment rate with a two-year lag has the opposite effect on the strength index variation. This could be explained by a progressive decrease of the number of male candidates and therefore graduates. The same logic might be applied to Health and Nursing courses. The offer of such courses may have been decreasing over the years as well as its' demand by potential candidates.

Our work was the first to test the impact of the policy measure of disclosing information about the unemployment rate by courses. Regarding future research, we would like to deepen the analysis by fields of study and degrees, and also by NUTS II regions, which can provide insightful policy information. Additionally, we can add more years to our database and try alternative model specifications.

Bibliography

1. Alves, N., Centeno, M., & Novo, A. A. (2010). Investment in Education in Portugal: Returns and Heterogeneity. *Banco de Portugal Economic Bulletin,* Spring, Economics and Research Department.
2. Amaral, A., & Teixeira, P. (2000). The rise and fall of the private sector in Portuguese higher education. *Higher Education Policy*, 13, 245-266.
3. Arellano, M. (2003). *Panel Data Econometrics.* Oxford University Press
4. Campos, M. M., & Reis, H. (2017). Revisiting the returns to schooling in the Portuguese economy. *Banco de Portugal Economic Bulletin*.
5. Cardoso, A. R., Portela, M., Sá, C., & Fernando, A. (2007). Demand for higher education programs: the impact of the Bologna process. *CESifo working paper No.2081*.
6. Correia, F., Amaral, A., & Magalhães, A. (2002). Public and Private Higher Education in Portugal: unintended effects of deregulation. *European Journal of Education*, 37(4), 457-471.
7. Gajderowicz, T., Grotkowska, G., Mycielski, J., Wincenciak, L. (2014). Social and economic determinants of higher education choices in Poland. *Ekonomia Journal*, 38, 9-29, Faculty of Economic Sciences, University of Warsaw.
8. Goulas, S., Megalokonomou, R. (2019). Which degrees do student prefer during recessions? *Empirical Economics*, 56, 2093-2125.
9. Hartog, J., Pereira, P., & Vieira, J. (2001). Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression estimators. *Applied Economics*, 33(8), 1021-1037.
10. Machado, M., Brites, R., Magalhães, A., & Sá, M. (2011). Satisfaction with higher education: Critical data for student development. *European Journal of Education*, 46(3), 415-432.
11. Portugal, P. (2004). Myths and Facts Regarding the Portuguese Labour Market - the Tragic Fate of College Graduates. *Banco de Portugal Economic Bulletin*.
12. Simões, C., & Soares, A. M. (2010). Applying to higher education: informations sources and choice factors. *Studies in Higher Education*, 35(4), 371-389. doi:10.1080/03075070903096490

13. Sousa, S., Portela, M., & Sá, C. (2015). Charactherization of returns to education in Portugal: 1986-2009. *Mimeo*.
14. Tavares, D., Tavares, O., Justino, E., & Amaral, A. (2008). Students' preferences and needs in Portuguese higher education. *European Journal of Education*, 43(1), 107-122.
15. Varga, J. (2006). The role of labour market expectations and admission probabilities in students' application decisions on higher education: The case of Hungary. Education Economics, 14(3), 309-327.
16. Vieira, J. (1999). Returns to education in Portugal. *Labour Economics*, 6(4), 535-541.
17. Wooldridge, J. M. (2006). *Introductory Econometrics: A Modern Approach.* Mason OH: Thomson/South-Western.

Appendix

Table A1 - Classification of the Study Areas Explored in this Study divided by Three Groups

Major Groups CNAEF-1D	Fields of Study CNAEF-2D	Areas of Education and Formation CNAEF-3D
1 - Education	14 - Teacher/Trainer Training and Education Sciences	142 - Education Sciences 143 - Training of Early Childhood Educators 144 - Training of Teachers of Basic Education (1st and 2nd Cycles) 145 - Teacher Training in Specific Disciplinary Areas 146 - Training of Teachers and Trainers of Technological Areas
2 - Arts and Humanities	21 - Arts	210 – Arts 211 - Fine Arts 212 - Performing Arts 213 - Audio-visual and Media Production 214 - Design 215 - Craft
	22 - Humanities	221 - Religion and Theology 222 - Foreign Languages and Literatures 223 - Mother Language and Literature 225 - History and archeology 226 - Philosophy and Ethics 229 - Humanities – programs not classified in another training area
3 - Social Sciences, Commerce and Law	31 - Social and Behavioral Sciences	310 - Social and Behavioral Sciences 311 – Psychology 312 - Sociology and Other Studies 313 - Political Science and Citizenship 314 - Economy
	32 - Information and Journalism	321 - Journalism and Reportage 322 - Librarianship, Archival science and Documentation (LAD) 329 - Information and Journalism - programs not classified in another training area
	34 - Business Sciences	340 - Business Sciences 341 – Trade 342 - Marketing and Advertising 343 - Finance, Banking and Insurance 344 - Accounting and Taxation 345 - Management and Administration 346 - Secretariat and Administrative Work 347 - Framework on Organization

		349 - Business Sciences - programs not classified in another training area
	38 - Law	380 - Law
4 - Sciences, Mathematics and Informatics	42 - Life Sciences	421 - Biology and Biochemistry 422 - Environmental Sciences 429 - Life Sciences - programs not classified in another training area
	44 - Physical Sciences	441 - Physics 442 - Chemistry 443 - Earth Sciences
	46 - Mathematics and Statistics	461 - Mathematics 462 - Statistics
	48 - Informatics	481 - Computer Science 489 - Informatics - programs not classified in another training area
5 - Engineering, Manufacturing and Construction	52 - Engineering and Related Techniques	520 - Engineering and Related Techniques 521 - Metallurgy and Metalworking 522 - Electricity and Energy 523 - Electronics and Automation 524 - Chemical Process Technology 525 - Construction and Repair of Vehicles 529 - Engineering and Related Techniques - programs not classified in another training area
	54 - Manufacturing Industries	541 - Food Industries 542 - Textile, Clothing, Footwear and Leather Industries 543 - Materials (Wood Industries, Cork, Paper, Plastic, Glass and others) 544 - Extractive Industries 549 - Manufacturing Industries - programs not classified in another training area
	58 - Architecture and Construction	581 - Architecture and Urbanism 582 - Civil construction and Civil Engineering
6 - Agriculture	62 - Agriculture, Forestry, and Fisheries	620 - Agriculture, Forestry, and Fisheries 621 - Agricultural and animal production 622 - Floriculture and Gardening 623 - Forestry and Hunting 624 - Fisheries
	64 - Veterinary Science	640 - Veterinary Science
7 - Health and Social Protection	72 - Health	721 – Medicine 723 - Nursing 724 - Dental Sciences 725 - Diagnostic and Therapeutic Technologies 726 - Therapy and Rehabilitation 727 - Pharmaceutical Sciences 729 - Health - programs not classified in another training area
	76 - Social Services	761 - Support Services for Children and Young People 762 - Social Work and Orientation

8 - Services	81 - Personal Services	811 - Hotel and Restaurant 812 - Tourism and Leisure 813 - Sport
	84 - Transport Services	840 - Transport Services
	85 - Environment Protection	851 - Environmental Protection Technology 852 - Natural Environments and Wildlife 853 - Public Health Services
	86 - Security Services	861 - Protection of Persons and Property 862 - Safety and Hygiene at Work 863 - Military Security
9 - Unknown or not specified	99 - Unknown or not specified	999 - Unknown or not specified

Table A2 – Descriptive Statistics – 2005-2016

	UNEMP (Total)	INDF (Average)	UR (Average)	GDP_PC (Average)	DENS_POP (Average)
2005	2076	0.540	0.0232139215	85.01	522.255
2006	2290	0.492	0.0235539125	85.10	524.3125
2007	3021	0.475	0.0348754343	85.66	526.4288
2008	3218	0.463	0.0394179587	82.81	527.8010
2009	2893	0.469	0.0450009115	84.28	523.2639
2010	2918	0.462	0.0465616937	84.08	524.3451
2011	3133	0.301	0.0480159581	79.05	524.3890
2012	4038	0.303	0.0572947556	76.79	522.3766
2013	5814	0.295	0.0704299897	78.21	518.6983
2014	5254	0.292	0.0693482584	78.17	517.6656
2015	7190	0.306	0.0902300487	78.15	517.1264
2016	1520	0.306	0.0251330751	78.55	517.3528

Table A3 – Descriptive Statistics by Total of Candidates, Candidates by Gender, 1st Option Candidates and Vacancies – 2005 to 2016

	Tot_Cand	Cand_Male	Cand_Female	Cand_1stOp	Vacs
2005	173076	64439 (37.23%)	108641 (62.77%)	36684 (21.2%)	46767
2006	181234	68463 (37.77%)	112783 (62.23%)	38313 (21.14%)	44809
2007	228658	90926 (39.76%)	137743 (60.24%)	48235 (21.09%)	45221
2008	246042	105765 (42.98%)	140299 (57.02%)	49785 (20.23%)	45868
2009	237673	99937 (42.05%)	137745 (57.95%)	48970 (20.60%)	46538
2010	232562	97009 (41.71%)	135570 (58.29%)	48500 (20.85%)	47460
2011	205931	85263 (41.40%)	120687 (58.60%)	43702 (21.22%)	77436
2012	200907	82318 (40.97%)	118596 (59.03%)	42731 (21.27%)	73283
2013	176897	72515 (40.99%)	104405 (59.01%)	38475 (21.75%)	68170
2014	183772	72843 (39.64%)	110938 (60.36%)	40323 (21.94%)	63646
2015	207029	87564 (42.30%)	119467 (57.70%)	44757 (21.62%)	63602
2016	206529	88265 (42.74%)	118273 (57.26%)	45053 (21.81%)	62853

Table A4 –District of Origin (Cand_District) with the Highest Number of Candidates for each Course – 2005 to 2016

Cand_Dist/Ano	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Aveiro	4.6%	4.2%	4.9%	4.2%	4.6%	3.9%	4.3%	4.4%	4.6%	4.7%	4.3%	4.2%
Beja	2.0%	1.9%	1.7%	1.6%	1.7%	1.1%	17%	1.6%	1.1%	1.1%	1.2%	1.2%
Braga	8.1%	8.6%	8.6%	11.1%	8.7%	9.5%	9.5%	8.3%	8.3%	8.5%	8.8%	8.9%
Bragança	2.8%	3.3%	3.1%	1.5%	1.3%	0.9%	0.9%	1.3%	0.5%	1.3%	0.6%	0.5%
Castelo.B	6.1%	5.2%	5.0%	5.2%	5.2%	4.7%	4.1%	4.6%	4.1%	4.4%	3.9%	3.7%
Coimbra	9%	9.3%	9.4%	9.4%	9.4%	9.1%	9.2%	9.0%	9.1%	8.7%	9.7%	9.3%
Évora	2.7%	2.3%	2.9%	2.8%	2.8%	2.6%	2.7%	2.9%	2.5%	2.6%	2.5%	2.6%
Faro	5.1%	5.0%	5.4%	5.0%	5.2%	5.4%	5.0%	5.3%	5.2%	4.8%	4.7%	4.5%
Guarda	1.5%	1.8%	2.0%	1.4%	1.4%	1.6%	1.1%	1.0%	1.1%	1.1%	1.1%	1.3%
Leiria	3.6%	3.3%	3.9%	3.5%	3.6%	4.4%	4.5%	4.4%	4.7%	4.8%	4.1%	4.4%
Lisboa	21.6%	21.7%	21.6%	23.4%	22.9%	22.8%	22.4%	23.2%	24.9%	25.2%	26.1%	26.6%
Portalegre	1.5%	1.8%	1.9%	2.0%	1.6%	2.0%	1.6%	1.1%	1.3%	1.2%	0.9%	1.1%
Porto	12.8%	13.5%	13.2%	13.9%	16.9%	17.7%	17.6%	17.6%	19.1%	18.5%	19.4%	19.0%
R.A.Açores	1.1%	1.3%	1.6%	1.5%	1.7%	1.6%	1.4%	1.5%	1.7%	1.8%	1.8%	1.7%
R.A.Madeira	1.2%	0.9%	0.8%	0.8%	0.7%	0.7%	0.7%	0.7%	0.9%	1.0%	0.9%	1.1%
Santarém	4.6%	4.1%	3.5%	2.8%	3.3%	3.0%	3.3%	3.0%	2.4%	2.0%	2.6%	1.9%
Setúbal	2.9%	3.1%	2.9%	2.8%	2.5%	2.3%	2.5%	2.2%	2.3%	2.0%	1.8%	2.1%
VianaCastelo	1.2%	1.1%	0.8%	0.5%	1.1%	0.9%	1.0%	0.9%	1.1%	1.0%	1.2%	0.8%
Vila Real	3.9%	3.7%	3.2%	2.7%	2.1%	2.2%	2.8%	3.2%	2.0%	1.9%	1.8%	1.9%
Viseu	3.8%	4.0%	3.8%	3.8%	3.6%	3.5%	3.6%	3.8%	3.3%	3.2%	3.4%	3.4%

Table A5 – Descriptive Statistics (NUTS III) – year 2016

	Frequency	Percentage	Valid Percentage	Cumulative Percentage
Alentejo Central	126	2.9	2.9	2.9
Alentejo Litoral	6	.1	.1	3.1
Algarve	161	3.7	3.7	6.8
Alto Alentejo	48	1.1	1.1	7.9
Alto Minho	73	1.7	1.7	9.6
Alto Tâmega	7	.2	.2	9.8
Área Metropolitana de Lisboa	1379	32.0	32.0	41.8
Área Metropolitana do Porto	796	18.5	18.5	60.3
Ave	59	1.4	1.4	61.7
Baixo Alentejo	51	1.2	1.2	62.9
Beira Baixa	72	1.7	1.7	64.6
Beiras e Serra da Estrela	143	3.3	3.3	67.9
Cávado	251	5.8	5.8	73.7
Douro	112	2.6	2.6	76.3
Lezíria do Tejo	54	1.3	1.3	77.6
Médio Tejo	40	.9	.9	78.5
Oeste	41	1.0	1.0	79.5
Região Autónoma da Madeira	19	.4	.4	79.9
Região Autónoma dos Açores	33	.8	.8	80.7
Região de Aveiro	183	4.3	4.3	84.9
Região de Coimbra	324	7.5	7.5	92.4
Região de Leiria	87	2.0	2.0	94.5
Tâmega e Sousa	14	.3	.3	94.8
Terras de Trás-os-Montes	111	2.6	2.6	97.4
Viseu Dão Lafões	113	2.6	2.6	100.0
Total	4303	100.0	100.0	

Table A6 – Descriptive Statistics by Degree in 2016

	Frequency	Percentage	Valid Percentage	Cumulative Percentage
B	32	.7	.7	.7
L	657	15.3	15.3	16.0
L1	1675	38.9	38.9	54.9
M	1602	37.2	37.2	92.2
MI	127	3.0	3.0	95.1
D	210	4.9	4.9	100.0
Total	4303	100.0	100.0	

Table A7 – Descriptive Statistics by NUTS II in 2016

	Frequency	Percentage	Valid Percentage	Cumulative Percentage
Alentejo	285	6.6	6.6	6.6
Algarve	161	3.7	3.7	10.4
Área Metropolitana de Lisboa	1379	32.0	32.0	42.4
Centro	1003	23.3	23.3	65.7
Norte	1423	33.1	33.1	98.8
Região Autónoma da Madeira	19	.4	.4	99.2
Região Autónoma dos Açores	33	.8	.8	100.0
Total	4303	100.0	100.0	

Table A8 – Descriptive Statistics by Total of Graduates and Graduates by Gender – 2005 to 2016

	Tot_Grad	Male_Grad	Fem_Grad
2005	61328	21331 (34.78%)	39997 (65.22%)
2006	62825	21702 (34.54%)	41123 (65.46%)
2007	74472	28838 (38.72%)	45634 (61.28)
2008	74654	30178 (40.42%)	44476 (59.58%)
2009	67613	27762 (41.06%)	39851 (58.94%)
2010	67532	27253 (40.36%)	40279 (59.64%)
2011	66789	27238 (40.78%)	39551 (59.22%)
2012	69872	28145 (40.28%)	41727 (59.72%)
2013	68779	28054 (40.79%)	40725 (59.21%)
2014	62476	25583 (40.95%)	36893 (59.05%)
2015	62110	25326 (40.78%)	36784 (59.22%)
2016	59420	24617 (41.43%)	34803 (58.57%)

Table A9 – Descriptive Statistics by CNAEF 1 Digit in 2016

	Frequency	Percentage	Valid Percentage	Cumulative Percentage
Education	565	13.1	13.1	13.1
Arts and Humanities	661	15.4	15.4	28.5
Social Sciences. Commerce and Law	1099	25.5	25.5	54.0
Sciences. Mathematics and Informatics	517	12.0	12.0	66.0
Engineering. Manufacturing and Construction	651	15.1	15.1	81.2
Agriculture	103	2.4	2.4	83.6
Health and Social Protection	388	9.0	9.0	92.6
Services	318	7.4	7.4	100.0
Unknown or Unspecified	1	.0	.0	100.0
Total	4303	100.0	100.0	

Table A10 – Descriptive Statistics by CNAEF 2 Digits in 2016

	Frequency	Percentage	Valid Percentage	Cumulative Percentage
Teacher/Trainer Training and Education Sciences	565	13.1	13.1	13.1
Arts	321	7.5	7.5	20.6
Humanities	340	7.9	7.9	28.5
Social and Behavioral Sciences	394	9.2	9.2	37.6
Information and Journalism	87	2.0	2.0	39.7
Business Sciences	542	12.6	12.6	52.3
Law	76	1.8	1.8	54.0
Life Sciences	188	4.4	4.4	58.4
Physical Sciences	166	3.9	3.9	62.3
Mathematics and Statistics	53	1.2	1.2	63.5
Informatics	110	2.6	2.6	66.0
Engineering and Related Techniques	395	9.2	9.2	75.2
Manufacturing Industries	89	2.1	2.1	77.3
Architecture and Construction	167	3.9	3.9	81.2
Agriculture. Forestry. and Fisheries	85	2.0	2.0	83.2
Veterinary Science	18	.4	.4	83.6
Health	293	6.8	6.8	90.4
Social Services	95	2.2	2.2	92.6
Personal Services	177	4.1	4.1	96.7
Transport Services	7	.2	.2	96.9
Environment Protection	88	2.0	2.0	98.9
Security Services	46	1.1	1.1	100.0
Unknown or not specified	1	.0	.0	100.0
Total	4303	100.0	100.0	

Table A11 – Descriptive Statistics by CNAEF 3 Digits in 2016

	Frequency	Percentage	Valid Percentage	Cumulative Percentage
Education Sciences	113	2.6	2.6	2.6
Training of Early Childhood Educators	37	.9	.9	3.5
Training of Teachers of Basic Education (1st and 2nd Cycles)	254	5.9	5.9	9.4
Teacher Training in Specific	115	2.7	2.7	12.1
Training of Teachers and Trainers of Technological Areas	46	1.1	1.1	13.1
Arts	3	.1	.1	13.2
Fine Arts	66	1.5	1.5	14.7
Performing Arts	75	1.7	1.7	16.5
Audio-visual and Media Production	100	2.3	2.3	18.8
Design	69	1.6	1.6	20.4
Craft	8	.2	.2	20.6
Religion and Theology	8	.2	.2	20.8
Foreign Languages and Literatures	144	3.3	3.3	24.1
Language and Modern Literature	51	1.2	1.2	25.3
History and archeology	111	2.6	2.6	27.9
Philosophy and Ethics	25	.6	.6	28.5
Humanities – programs not classified in another training area	1	.0	.0	28.5
Social and Behavioral Sciences	1	.0	.0	28.5
Psychology	142	3.3	3.3	31.8
Sociology and Other Studies	124	2.9	2.9	34.7
Political Science and Citizenship	55	1.3	1.3	36.0
Economy	72	1.7	1.7	37.6
Journalism and Reportage	63	1.5	1.5	39.1
Librarianship. Archival science and Documentation (LAD)	23	.5	.5	39.6
Information and Journalism - programs not classified in another training area	1	.0	.0	39.7

Business Sciences	1	.0	.0	39.7
Trade	17	.4	.4	40.1
Marketing and Advertising	101	2.3	2.3	42.4
Finance. Banking and Insurance	27	.6	.6	43.1
Accounting and Taxation	77	1.8	1.8	44.9
Management and Administration	284	6.6	6.6	51.5
Secretariat and Administrative Work	23	.5	.5	52.0
Framework on Organization	8	.2	.2	52.2
Business Sciences - programs not classified in another training area	4	.1	.1	52.3
Law	76	1.8	1.8	54.0
Biology and Biochemistry	159	3.7	3.7	57.7
Environmental sciences	25	.6	.6	58.3
Life Sciences - programs not classified in another training area	4	.1	.1	58.4
Physics	36	.8	.8	59.2
Chemistry	41	1.0	1.0	60.2
Earth Sciences	89	2.1	2.1	62.3
Mathematics	43	1.0	1.0	63.3
Statistics	10	.2	.2	63.5
Computer Science	102	2.4	2.4	65.9
Informatics - programs not classified in another training area	8	.2	.2	66.0
Engineering and Related Techniques	2	.0	.0	66.1
Metallurgy and Metalworking	54	1.3	1.3	67.3
Electricity and Energy	53	1.2	1.2	68.6
Electronics and Automation	138	3.2	3.2	71.8
Chemical Process Technology	97	2.3	2.3	74.0
Construction and Repair of Vehicles	14	.3	.3	74.4
Engineering and Related Techniques - programs not classified in another training area	37	.9	.9	75.2
Food Industries	48	1.1	1.1	76.3
Textile. Clothing. Footwear and Leather Industries	9	.2	.2	76.6
Materials (Wood Industries. Cork. Paper. Plastic. Glass and others)	18	.4	.4	77.0
Extractive Industries	13	.3	.3	77.3

Manufacturing Industries - programs not classified in another training area	1	.0	.0	77.3
Architecture and Urbanism	94	2.2	2.2	79.5
Civil construction and Civil Engineering	73	1.7	1.7	81.2
Agriculture. Forestry. and Fisheries	1	.0	.0	81.2
Agricultural and animal production	68	1.6	1.6	82.8
Floriculture and Gardening	3	.1	.1	82.8
Forestry and Hunting	11	.3	.3	83.1
Fisheries	2	.0	.0	83.2
Veterinary Science	18	.4	.4	83.6
Medicine	23	.5	.5	84.1
Nursing	67	1.6	1.6	85.7
Dental Sciences	18	.4	.4	86.1
Diagnostic and Therapeutic Technologies	69	1.6	1.6	87.7
Therapy and Rehabilitation	70	1.6	1.6	89.3
Pharmaceutical Sciences	39	.9	.9	90.2
Health - programs not classified in another training	7	.2	.2	90.4
Support Services for Children and Young People	1	.0	.0	90.4
Social Work and Orientation	94	2.2	2.2	92.6
Hotel and Restaurant	14	.3	.3	92.9
Tourism and Leisure	86	2.0	2.0	94.9
Sport	77	1.8	1.8	96.7
Transport Services	7	.2	.2	96.9
Environmental Protection Technology	62	1.4	1.4	98.3
Natural Environments and Wildlife	15	.3	.3	98.7
Public Health Services	11	.3	.3	98.9
Protection of Persons and Property	8	.2	.2	99.1
Safety and hygiene at Work	30	.7	.7	99.8
Military Security	8	.2	.2	100.0
Unknown or not specified	1	.0	.0	100.0
Total	4303	100.0	100.0	