Emigration prospects and educational choices: evidence from the Lorraine-Luxembourg corridor.

Michel Beine[a], Vincent Fromentin[b]
Javier Sánchez Bachiller [c] University of Luxembourg, IZA, CES-Ifo.,
[b] University of Lorraine., [c] University of Luxembourg

March 12, 2024

Abstract

A large literature has documented the incentive effect of emigration prospects in terms of human capital accumulation in origin countries. Much less attention has been paid to the impact on specific educational choices. We provide some evidence from the behaviour of students of the University of Lorraine located in the North-East of France and close to Luxembourg, a booming economy with attractive work conditions. We find that students who paid attention to the foreign labour market at the time of enrolment tend to choose topics that lead to occupations that are highly valued in Luxembourg. These results hold when accounting for heterogeneous substitution patterns across study fields through the estimation of advanced discrete choice models. Incentive effects of emigration prospects are also found when accounting for the potential endogeneity of the interest for the foreign labour market using a control function approach based on the initial locations of these students at the time of enrolment. Consistently, students showing no attention to the foreign labour market are not subject to the incentive effect of emigration prospects.

JEL Classification: C25, F22, J61.

Keywords: Brain gain, Emigration prospects, Educational Choices, Discrete choice modelling, Labour Markets.

1 Introduction

Over the last decades, there has been a significant increase in the observed international mobility of skilled workers. While migration has long been the missing piece of globalization, the emigration rate of workers with a college degree has been multiplied by a factor of three over the last 30 years (Docquier and Rapoport (2012)). This increase first reflects a strong increase in the demand of skills in developed economies, spurred by many factors including the skill-biased technological progress. Firms located in industrialized countries are nowadays in high competition to attract the talented workers that they need to develop their projects and have increasingly searched beyond the domestic labour market. This increase in the labour demand has been matched by a higher propensity of skilled individuals to move abroad over time. Today, college graduates are better informed about foreign work opportunities, as technology makes it easier to obtain information about job offers abroad and tends to reduce physical and psychological moving costs.

The international migration of skilled workers, especially between developing and developed countries, has been coined the brain drain phenomenon. In the seventies, the traditional view was that the brain gain was detrimental to the migrants' origin countries, as it led to a depletion of their human capital. This traditional view inspired the proposed Baghwati tax through which destination countries would compensate the origin countries for the loss incurred by the brain drain (Bhagwati (1976)). Nevertheless, this view has been nuanced over time through the identification of several additional effects generated by the brain drain. An important effect is the so-called incentive effect of migration in terms of human capital investment.

The incentive effect of migration in terms of human capital level arises from the higher opportunities offered to individuals by the foreign labour market. The attractiveness of foreign opportunities are higher for educated individuals, basically for two complementary reasons. First, the wage premium between the domestic and foreign labour markets is clearly increasing with respect to the skill level. Second, immigration policies that act as powerful sorting devices in many destination countries are more favourable for skilled individuals. In turn, compared to an autarchic situation in which foreign options are unavailable, emigration prospects lead a larger number of individuals to invest in education, raising the global level of human capital in the source countries before emigration takes place. Whether the ex-post level of human capital increases or not, i.e. whether the brain drain results in a brain gain, depends on a set of country-specific factors, such as the level of the emigration rate, the quality of the higher education system and the quality of economic institutions (Beine et al. (2008)).

The precise nature of the incentive effect of the brain drain has been clarified at the end of nineties in a set of theoretical works (Stark et al. (1997), Mountford (1997), Vidal (1997), Beine et al. (2001)). Subsequently, these theoretical results have received some empirical support, initially based on macroeconomic data Beine et al. (2008). These empirical macroeconomic studies have also been complemented by analyses based on individual data, showing that the incentive effect is much more than an academic curiosity (Batista et al. (2012), Abarcar and Theoharides (2020) among others). However, this literature has focused on one specific type of incentive effect, namely the impact on the global human capital level, but has neglected to consider the impact in terms of the type of acquired skills. The incentive effect implies that individuals should not only increase their education level, but also invest more in the skills that are relatively more rewarded on the foreign labour market compared to the domestic market. In this paper, we bring evidence in favour of such an effect.

To this aim, we take advantage of an original survey conducted by DRAPEQ which covers students from the University of Lorraine after graduation. The University of Lorraine is one of the most important universities in France and provides a comprehensive offer in terms of subjects and degrees. The Lorraine region is located in the North-East of France and is contiguous to the Grand-Duchy of Luxembourg, the richest country in the world with a booming labour market based on the development of financial services and technological products. Due to its geographical and linguistic proximity, the absence of mobility restrictions for French citizens, and the existence of convenient bilateral agreements in the area of income taxation and social security, Luxembourg has been by far the preferred foreign option for fresh college graduates originating from this region. The survey data include precise information on the topic studied by the graduates of the University, as well as a rich set of individual characteristics. Combining this information with data capturing the relative attractiveness of corresponding professional occupations between the Luxembourgish and the French labour markets, we investigate whether the students internalized that information when choosing their subject at the start of their higher education studies. The attractiveness of each occupation is measured through average wages as well as through its employability rate in both countries' markets. An interesting aspect of the survey that we exploit is that it includes an explicit question about the attention the students paid to the foreign labour market in general, and to Luxembourg in particular, when eliciting their study subject.

Our findings can be summarized as follows. We provide evidence of an effect of emigration prospects on the investment on skills that are relatively more rewarded for abroad. Namely, students of the University of Lorraine tend to enroll more in degrees that lead to higher employability in Luxembourg. We also find an effect related to higher wages, although employability seems to be the more robust factor of attractiveness. The incentive effect of emigration is observed for students stating that they paid some attention to the foreign markets in general, and to Luxembourg in particular at the time of enrolment. Students who did not consider these options are not subject to an incentive effect of emigration. These results suggest that the acquisition of information about foreign options is key to generate an incentive effect of emigration prospects on human capital investment. Our results are robust to several phenomena. The findings hold when we capture in our estimations the fact that some set of topics are more similar than others, implying a higher degree of substitutability in the educational choices. They are also similar when we account for the possibility that the interest shown for Luxembourg might be related to factors driving the education choices.

Our paper is directly related to three separate existing strands of literature. First, we contribute to the empirical literature on the brain gain in general and on the incentive effect in terms of human capital accumulation in particular. Following early evidence based on macroeconomic data (Beine et al. (2008)), a set of contributions have assessed the existence of the incentive effect based on individual data.[1] Batista et al. (2012) bring the first causal evidence in the case of emigration from Cape Verde, showing that an increase in the individual probability of emigration tended to boost educational achievement at the secondary education level. Shrestha (2017) takes benefit of a change in the educational requirement of recruitment of the British Army of Nepali citizens and shows that this led to an increase in the proportion of men completing their secondary education. This resulted in a net increase in the ex-post human capital level, which in turn generated beneficial effects for the local economy. Chand and Clemens (2019) exploit a quasi-natural experiment in which a sudden surge in discrimination against islanders of Indian ethnic in Fidji led to a large emigration wave of this group of individuals and to an important investment in skills on their side, resulting ultimately in a brain grain.

There is also a limited number of contributions suggesting that emigration prospects induce investment in specific skills in origin countries. Abarcar and Theoharides (2020) exploit variations in visa restrictions from the US for nurses originating from the Philippines. They show that in regions traditionally prone to send nurses abroad, expansions (restrictions) in emigration prospects boosted (decreased) enrolment in nursing education programs and resulted in an increase (decrease) in the stock of graduates in this field. Using evidence from university students in seven different

[1]While most papers look at the impact of emigration prospects on contemporaneous levels of human levels, some studies focus also on the inter-generational effects of such emigration shocks. See, for instance, Theoharides (2018) or Dinkelman and Mariotti (2016).

countries, Kulka et al. (2023) find that there is a positive correlation between the level of international applicability of human capital and migration intentions. Based on a survey on secondary school students in Tonga, Gibson and McKenzie (2010) report that students considering to go abroad were more keen to study science subjects and to improve their English language skills. We contribute to this strand of the brain drain literature by providing an analysis involving a comprehensive set of study topics, which in turn allows to pin down the incentives of economic rewards of skills on the human capital investment. Indeed, our analysis is based on enrolment of students in a major university offering a comprehensive set of degrees, which we match with economic rewards of corresponding professional occupations in the domestic and foreign market. A second original point of our study, albeit less important, is that we provide evidence of an incentive effect of emigration prospects on human capital investment between developed countries. Almost all the contributions providing evidence of such an effect consider South-North emigration prospects. This is quite understandable since the incentive effect is driven by the magnitude of wage differentials. Nevertheless, we show that such an incentive effect might also occur within a context of neighbouring regions of developed economies when prospects of improvement of employment of migrants are non negligible.

Our paper is also related to a second important strand of literature focusing on the determinants of choice in terms of specific skill acquisition. Along the human capital theory (Becker (1962)), beyond preferences for specific topics as well as other factors such as their social background, students act as rational forward-looking agents and tend to choose the topics that are more rewarded on the labour market (Chapman (2012), Cameron and Heckman (1998b), Cameron and Heckman (1998a), Gibbons and Vignoles (2012)). Along the signalling theory, degrees also act as signalling devices of future productivity, allowing workers to grab higher wages in the labour market (Spence (1978)). Investments in specific skills by students are also explained in sociology by the rational choice theory, which involves long-term benefits such as income or job prestige (Breen and Goldthorpe (1997)). While there is a large body of evidence indicating that higher returns lead students to choose specific topics on the domestic labour market, to the best of our knowledge, no empirical contribution has looked at the specific role of the foreign labour market. Our contribution fills this gap by bringing some evidence that prospects of emigration play a role in this choice. In that sense, our paper brings a bridge between the literature on brain drain and the literature devoted to the choice of skill investment in higher education.

Finally, our paper is also connected to the brain waste literature. If domestic students tend to invest more in skills that are rewarded abroad and if industrial structures between the domestic and the foreign economies are different, the existence of an incentive effect can lead, at least in the short run, to suboptimal outcomes at origin, even with moderate brain drain rates.[2] In our context, this concern is relevant, since there are substantial differences in the industrial structures between the Lorraine region and Luxembourg. While Lorraine is characterized by a traditional industrial structure based on usual manufacturing sectors, Luxembourg is an economy dominated by a booming financial sector and related specialized services such as consulting, auditing, IT infrastructure and research (Statec (2023)). Most of the literature on the emigration of high skilled workers has addressed the question of the brain waste from the perspective of the receiving countries. The brain waste results in an under-utilization of human capital, at least assessed from the nominal education degrees. In some professional occupations such as medical ones, this phenomenon is related to the lack of recognition of credentials between countries. Nevertheless, the observed mismatch between jobs and education can also be explained by the difference in education quality between countries, especially in the case of South-North brain drain (Mattoo et al. (2008)). In contrast to this literature, the issue of the brain waste here applies directly to the sending country and results from a short-run mismatch of skills associated with the incentive effect of emigration.

The paper is organized as follows. Section 2 presents the model and its testable implications. Section 3 gives details about the context, presents our original data as well as the other data used in the econometric analysis. Section 4 presents the results while section 5 concludes.

2 Underlying Model and testable implications

In order to understand the way the mechanism of the incentive effect of emigration works, we assume that, at the start of their tertiary education cycle, prospective students of the University of Lorraine tend to choose the educational program that is associated to the largest expected utility. In line with the Random Utility Maximisation (RUM) approach, each prospective student n maximizes her utility over all possible educational programs j ($j = 0, 1, ..., J$) offered at the University of Lorraine. Formally, the utility of individual n of choosing program j is expressed as U_{jn} and can be additively decomposed into a deterministic component V_{jn} and a stochastic component ϵ_{jn}:

[2]Such a negative effect is nevertheless less obvious in the long run due to the existence of the skill biased technological progress. In the long term, specific investment in skills that were initially in shortage might induce the creation of specific activities with beneficial consequences for the economic development of the origin country.

$$U_{jn} = V_{jn} + \epsilon_{jn}. \tag{1}$$

2.1 Expected returns to skill in the domestic and foreign labour markets

An important component of the deterministic component of utility V_{jn} is the expected return on the domestic and foreign labour markets. The foreign labour market of interest is entirely captured by the characteristics of the Luxembourgish market, given that wages and employment in Luxembourg are, by far, much more attractive than those of any alternative location in the neighbouring regions. The expected return of skill j on the French labour market is denoted by $\mathbb{E}(w_{jn})$ and is determined by the expected wage of skill j, understood as the expected wage in an occupation closely related to that skill (denoted by w_j), and the probability of employment associated with skill j (denoted by $\Pr(e_{nj} = 1)$). We can thus express it as:

$$\mathbb{E}(w_{jn}) = \Pr(e_{nj} = 1) \times w_j \tag{2}$$

Similarly, the expected return of skill j for graduated student n on the foreign market is given by the expected wage in Luxembourg for the occupation associated to this type of skill w_j^*, the associated probability of employment $\Pr(e_{nj}^* = 1)$ as well as the probability of migrating to Luxembourg $\Pr(mig_n = 1)$ for individual n:

$$\mathbb{E}(w_{jn}^*) = \Pr(mig_n = 1) \times \Pr(e_{nj}^* = 1) \times w_j^* \tag{3}$$

In Luxembourg, the expected wage for individual n graduating with skill j depends on $\Pr(mig_n = 1)$, the probability of being allowed to migrate and work in Luxembourg for each individual. This probability is equal to 1 for French and for other European Union nationals due to the free mobility agreements at the European level. For extra-EU students, international mobility is subject to the restrictions of the Luxembourgish immigration policy. This might result in a lower expected wage compared to a native student.[3] Immigration policy in Luxembourg belongs to the category of employer-driven systems. The possibility to get an immigration visa mainly depends on getting a firm job offer from a Luxembourgish employer, as well

[3]For instance, while foreign EU workers can become cross-border workers (i.e. work under a Luxembourgish labour contract whilst living outside the country) this possibility does not exist for non-EU workers. Given the relatively higher cost of living in Luxembourg, in particular the high housing costs, this generates a mitigation of the expected net gain of migration compared to EU foreign workers.

as on additional checks from the immigration authorities.[4] In short, the fact that the worker is from outside the European Union creates some additional uncertainty about the probability of crossing the border and exerts downward pressure on the expected wage abroad. This expected wage also depends on $\Pr(e^*_{nj})$, the probability of finding a job in an occupation related to skill j for worker n. We assume absence of discrimination. Since choices regarding the type of skill to acquire are made before university enrolment, we assume that individuals have homogeneous information about this probability. This probability therefore depends only on the magnitude of the labour demand for that skill j in Luxembourg. The other important component of the expected wage is the return for skill j in the foreign labour market, w^*_j.

Let us assume that individuals have precise information about the attractiveness of each skill on the Luxembourgish labour market. The presence of the foreign labour market attractiveness component in the underlying utility of skill j is directly related to the existence of the incentive effect of emigration prospects on human capital. This effect is likely to vary across individuals depending on whether they paid attention to the foreign alternatives of the labour market. Intended stayers, i.e. individuals with a strong preference to stay in France after graduation, would pay little attention to the variation of $\mathbb{E}(w^*_{jn})$, in contrast with individuals looking at work conditions in Luxembourg. We account for such a heterogeneity by interacting $\mathbb{E}(w^*_{jn})$ with variable I_n capturing whether individual n observed the foreign labour market in general, and the Luxembourgish one in particular, at the time of university enrolment.

For the sake of simplicity, we first assume that $\Pr(mig_n = 1) = 1$ for all individuals, i.e. everyone is able to get a work permit in Luxembourg without any restriction. The expression for the utility associated to skill j for individual n takes the following form:

$$V_{jn} = \sum_{i=1}^{2} \alpha_i (I_{in} \times \log[\mathbb{E}(w^*_{jn})])) + \beta \log[\mathbb{E}(w_{jn})] + \delta_j \tag{4}$$

where δ_j is a degree-specific constant capturing common factors across individuals that influence the level of attractiveness of skill j.

Substituting expressions (2) and (3) in equation (4), we get the following expression for V_{jn}:

[4] For instance, this entails the fact that the position cannot be filled by a native worker. Depending on the type of visa, it also requires that the wage offered is above a minimum level.

$$V_{jn} = \beta_1 \log[\Pr(e_j)] + \beta_2 \log(w_j) + \alpha_1(I_n \times \log[\Pr(e_j^*)]) + \alpha_2(I_n \times \log(w_j^*)) + \delta_j \quad (5)$$

In this specification, the incentive effect associated to the foreign location (i.e. the attractiveness exerted by the Luxembourgish labour market) is associated to parameters α_1 and α_2. In particular, the existence of an incentive effect on the choice of study topics should be reflected by $\alpha_1 > 0$ and/or $\alpha_2 > 0$. In other terms, higher attractiveness of skill j in the foreign market raises the probability of enroling in the study field associated with that skill in the origin country.

2.2 The structure of the error term

We can then derive the choice probabilities for each skill j by looking at the stochastic component of equation 1. In the RUM approach, each student n is supposed to choose the skill that gives the maximum level of her (expected) utility. P_{jn}, the probability that individual n chooses skill j is given by:

$$P_{jn} = \Pr(U_{jn} > U_{kn}, \forall j \neq k), \quad (6)$$

which can be expressed as :

$$P_{jn} = \Pr(\epsilon_{jn} - \epsilon_{kn} < V_{jn} - V_{kn}, \forall j \neq k), \quad (7)$$

Expression 7 makes clear that in order to solve the maximisation program, one has to assume a particular probability distribution $f(\epsilon_{jn})$ for the stochastic component of the utilities. If we assume that ϵ_{jn} follows an extreme value distribution of type-1 following McFadden (1973), the derived choice probability for alternative j takes the following form:

$$P_{jn} = \frac{e^{V_{jn}}}{\sum_{k=1}^{K} e^{V_{kn}}}. \quad (8)$$

Other specifications for $f(\epsilon_{jn})$ lead to more complex solutions for equation (7). In particular, the solution depends on the way we assume the stochastic component of various study topics to be correlated within a given subset. This will be empirically explored in section 4.2.

3 Context and Data

3.1 Lorraine and Luxembourg within the Great Region

In this study, we use survey data on students' enrolment from the University of Lorraine. The University of Lorraine is located in the new French region 'Grand Est' and in the departments of Moselle and Meurthe-et-Moselle, both part of the historical region of Lorraine up to 2014 territorial reform. The region and the departments are neighbours of the Grand Duchy of Luxembourg (hereafter Luxembourg) and belong to the so-called Great Region, encompassing regions of Belgium, France, Germany and Luxembourg. For a couple of decades, due to its attractiveness, Luxembourg has been a country of intense immigration, with a proportion of immigrants close to 50%. About 50000 French nationals are immigrants in Luxembourg. Furthermore, Luxembourg has been the most common destination for cross-border commuters in the EU (in relative terms), with 212,000 incoming cross-border commuters on a daily basis. France is the main provider of cross-border workers, with about 112,000 individuals crossing the border every day. All in all, French nationals represent about a quarter of the total labour force of the country. Most of them commute from neighbouring areas located in the Department of Lorraine across the Luxembourgish border. A significant share of these workers graduated from the University of Lorraine.

3.2 Enrolment and survey data

The key data that we use to explain the elicitation of educational choices by the students is based on an annual survey conducted by OVU (Observatoire de la Vie Universitaire), a central service of the University of Lorraine. According to the Shanghai Ranking of higher education institutions, the University of Lorraine is in the top 300 of universities worlwide. It is home of about 60,000 students each year. It is one the most important comprehensive institutions in France and by far the most important one in the north-east part of France (Région Grand-est).

3.2.1 Location and individual characteristics of graduates

The initial purpose of this survey is to get first-hand information on the success of integration in the labour market of students graduating from the University. For that purpose, OVU conducts a large survey on the students freshly graduated from the University who have decided to enter the labour market. Our population of interest therefore involves former graduates of the University who have completed their education process and have joined the labour market. The survey includes bachelor and master graduates. We use the 2019 wave. This means that the educational choices of these students were elicited between 3 and 6 years before the survey, i.e. between 2013 and 2016. The survey includes 3038 graduates.

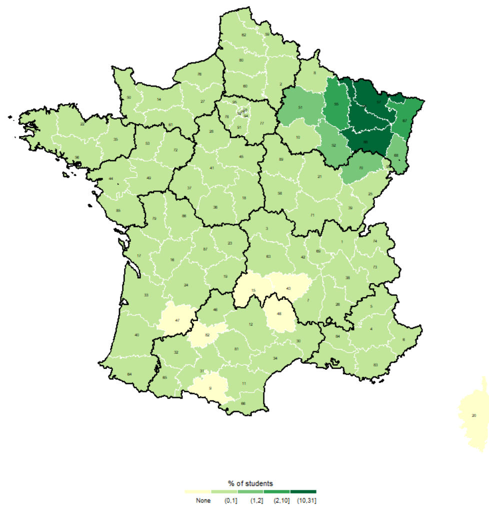


Figure 1: Origins of the graduates

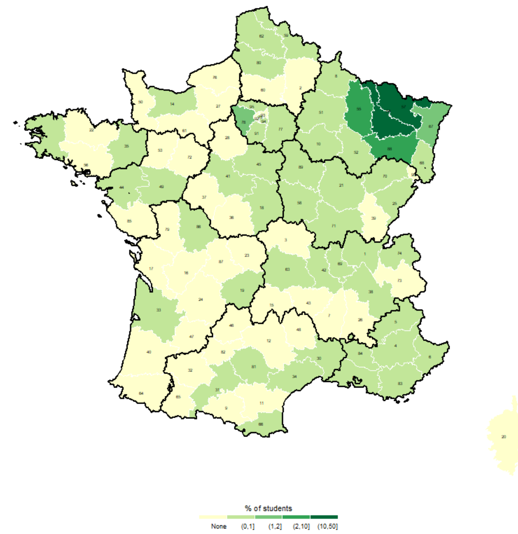


Figure 2: Share of interested in Luxembourg by origin of students

Figures 1 and 2 provide heat maps based on the initial origin of the native students.[5] Figure 1 gives the intensity of enrolment of native students with respect to their region of origin. While the overwhelming majority of the native students come from the Grand Est Region that includes the department of Lorraine, a substantial proportion of students come from regions outside the Grand Est. This is explains by the fact that University of Lorraine is a comprehensive university providing a very broad set of study topics. This feature is important for our empirical investigation since a sound discrete choice analysis of study topics requires a choice set as large as possible.

The survey data provides the details on the completed degree as well as information about the individual characteristics of the graduates. This includes individual characteristics such as gender, age, some information about their background such as the type of secondary degree or the postal code of parental address. This last piece of information turns out to be useful to capture student's location at the time of the elicitation of study topics. The data also include some information about their

[5]For the sake of exposition, foreign students and students from the overseas French territories (e.g. Guadeloupe, Martinique) are not represented here. Non-French students represent 14.3 % of the total enrolment (see Table 1).

current status like current location, the type of work, the job location, the type of contract and, if possible, their wage. Table 1 provides some summary statistics of the individual characteristics of the graduates included in the survey. We have a balanced sample in terms of gender and a proportion of foreign students (14.3%) in line with the share observed in the French system of higher education. About two thirds of the graduates originate from the Grand Est region, and about half come from the Lorraine department. About a fifth of the students had a strong or very strong interest for Luxembourg at the time of enrolment. 10% of graduates work in Luxembourg. We have also a balanced sample in terms of level of education, with about three fifths of the students graduating from a Master degree. The sample includes students from a broad set of disciplines. While the faculty of sciences hosts the highest number of students, there is a significant proportion studying social sciences and law.

To our request, the survey was supplemented with a couple of questions capturing the interest of the students for Luxembourg at the time of enrolment in the university. In particular, we capture their initial interest for the foreign labour market and the Luxembourguish one with the following questions. The first question takes the following form: 'At the start of your studies, did you consider a professional integration abroad?'. Then, for those answering positively, we ask the following daughter question: 'Was Luxembourg part of the countries of interest?'. The answer to that question considers four levels of intensity, from 'not at all' to 'yes, absolutely'. The variable based on that last question allows US to capture the variation of variable I_n in equation (4). Figure 2 provides the intensity of the interest for Luxembourg depending on the original location of native students, presenting the share of students for that region responding positively to the last question. The map makes clear that this interest is not random and is higher for native students having grown close to Luxembourg.[6] The endogeneity of this variable and its potential impact of our results will be addressed in section 4.3.

Our interest variable should reflect the attention paid to the Luxembourguish market at the time of enrolment and the willingness to consider foreign options after graduation. Our survey includes also some information about the place of work after graduation. About 10% of our students in the sample work in Luxembourg. In order to assess the informational content of the interest variable, we can compute the probabilities of working in Luxembourg conditional on the interest paid at the time of enrolment. We find that the proportion of graduates with a very strong interest is 56.6%, with an interest 13.9%, with little interest 6.85% and with no interest at

[6]This variation is even more pronounced when including foreign students and French overseas students that are not accounted for in this map

Table 1: Students' data summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Age	3,038	24.947	3.356	20	58
Female	3,038	0.492	0.500	0	1
Foreigner	3,038	0.143	0.350	0	1
Parents: contiguity to LU	3,038	0.474	0.499	0	1
Parents: distance to LU	3,038	481.3	1,257.9	0.00004	12,220
Origin: GrandEst	3,038	0.683	0.466	0	1
Origin: Lorraine	3,038	0.474	0.499	0	1
Interest in Grand Est	3,038	0.672	0.470	0	1
Interest in FR	3,038	0.444	0.497	0	1
Interest abroad	3,038	0.307	0.461	0	1
Interest in LUX	3,038	0.204	0.403	0	1
LU as a deciding factor	3,038	0.055	0.229	0	1
Working in LU	2,759	0.104	0.305	0	1
Level: Master	3,038	0.586	0.493	0	1
Faculty: Arts	3,038	0.063	0.243	0	1
Faculty: Law, Econ., Mng.	3,038	0.314	0.464	0	1
Faculty: Social Sciences	3,038	0.195	0.396	0	1
Faculty: Sciences	3,038	0.411	0.492	0	1
Faculty: Physical	3,038	0.017	0.128	0	1

Summary stats from raw data. Number of observations reflect the number of students answering that question. The interest questions are nested: Students answering they were not interested in working in Grand Est are then asked whether they were rather interested in working in France or abroad. Those interested by working abroad are asked whether their interest was in Luxembourg. The proportion for students having an interest for Lux are those of the two highest modalities (Strong and Very Strong).

all 5.49%. This suggests that this variable captures the propensity to internalize information about future work opportunities.

3.2.2 Educational Topics

One important piece of information concerns the educational choices of the students. The survey collects information about the level of the final degree (bachelor or master) as well as the specific topic chosen. The data gives quite detailed information about the completed educational programs. In our dataset, this amounts to 178 different programs, including general categories or majors (e.g. law, management, chemistry, engineering) but also more precise subcategories or minors (e.g. real estate law, management in entrepreneurship, ...).

Given the high degree of dimensionality of the choices, we consolidated the degrees into 58 different categories that contain all degrees that closely share the main topic (major) and that belong to the same educational level (bachelor or master).[7] Our criterion of consolidation is based on the share of common topics of each original degree. For example, in the database there are several commerce-related degrees (sharing the 'commerce' major) which are differentiated only by their specialization (minor): 'commerce and distribution', 'commerce of alimentary goods' or 'commerce of goods and services'. These are grouped into a broader category labelled 'commerce'. Table 10 in appendix C gives the details of the consolidation process for each original degree and each consolidated one. The 58 consolidated degrees represent the alternatives of choice that we modelled using a discrete-choice econometric approach.

3.3 Degree-specific labour market indicators

To be able to identify the effect of both the local and foreign labour markets might have on the educational choice of students of the University of Lorraine, we need to compute wages associated to each degree. The same holds for their employability prospects, which are captured by the labour demand indicators. The association between degrees on the one hand and wages or employability on the other hand involves several steps.

The first step is to link degrees with skills and jobs. Skills are identified by ROME codes.[8] In order to associate each degree with its corresponding ROME codes we

[7]Beyond the computational constraints associated to choice sets with a large number of alternatives, the need to consolidate is due to the fact that two very close degrees will be hardly distinguishable by any determinant.

[8]ROME stands for Répertoire Opérationnel des Métiers et des Emplois.

use the *France Compétences* online tool from the French Education Authority (Autorité nationale de financement et de régulation de la formation professionnelle et de l'apprentissage), which contains information on the skills acquired in each degree and the accessible jobs after graduation (and their corresponding ROME codes).[9] As an illustration, the degree in Economics has two associated ROME codes: 'Banking/finance customer relations' and 'Socio-economic studies and forecasts'. This means that training in economics provides graduates with the skills and knowledge required to work in those jobs. One degree can be associated to one or several ROME codes.[10] From now on, we refer to this correspondence as the 'ROME-degrees' correspondence.

The second step is to link jobs identified by ROME code to 'professional categories' (PCS) as well as the broader 'professional families' (FAP). The identification of professional categories is needed since wage data are captured by professions. We use a correspondence table compiled by the French Ministry of Labour.[11] This correspondence identifies for each professional family (FAP), which occupations (PCS) are considered to be part of that family and which jobs (ROME) are related to those occupations. We provide an example of this 'FAP-PCS-ROME' correspondance table in Table 2. As an illustration, the professional family of secondary school teachers (FAP W0Z90) is composed by teachers (PCS 341a) and general teachers (PCS 422a), with their respective job categories of general secondary education (ROME K2107) and technical and vocational education (K2109).

Table 2: FAP – PCS – ROME correspondence

Professional family (FAP)	Occupation (PCS)	Job (ROME)
Secondary school teachers (W0Z90)	Secondary school teachers (341a)	General secondary education (K2107)
Secondary school teachers (W0Z90)	General secondary school teachers (422a)	Technical and vocational education (K2109)

Note: In parentheses, the corresponding FAP, PCS or ROME code to each of the categories.

Finally, we also need to establish a correspondence between ROME codes and

[9]The tool can be found under https://www.francecompetences.fr/recherche_certificationprofessionnelle/

[10]The number of ROME codes for a given degree ranges from 1 to 7.

[11]DARES – La nomenclature des familles professionnelles (Version 2009). Table de correspondance FAP/PCS/ROME.

ISCO codes, in order to harmonize data across countries. This table is provided by the French Employment Agency.[12] We refer to this as the 'ISCO-ROME' correspondence.

With these correspondences, we can relate jobs and occupations to each degree and, thus, build indicators of wage and employability that gather the information of these occupations for each degree and country.

3.3.1 Domestic and foreign skill prices

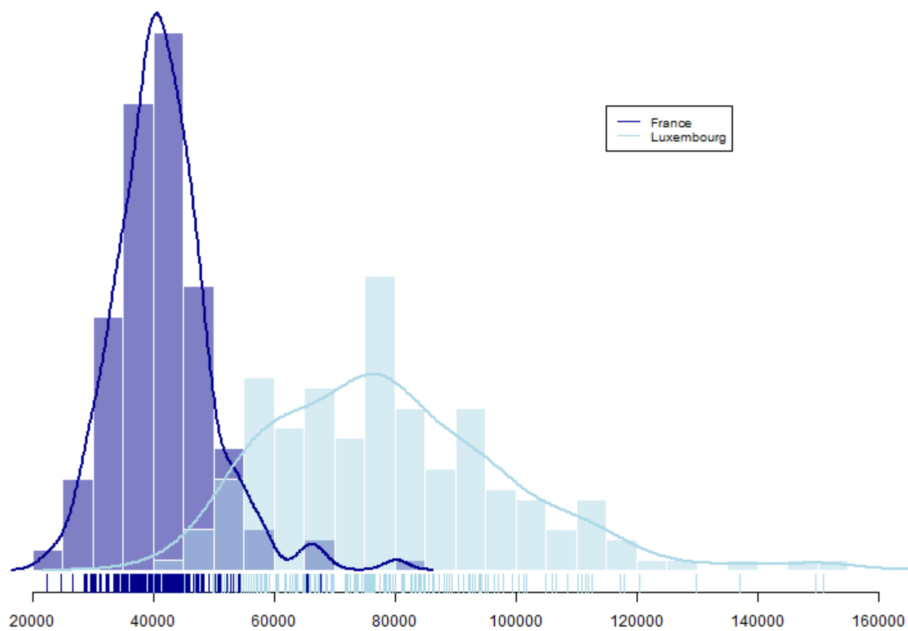
For France, wages are available for each occupation. For that purpose, we use the ECMOSS database 'Coût de la main d'œuvre et structure des salaires' provided by the National Statistical Agency (INSEE), which provides the salary by professional categories (PCS). This dataset is based on a survey gathering salary data for each occupation. We first calculate the average salary for each occupation. We then use the PCS–ROME correspondence explained above and calculate a ROME-specific salary as a weighted average of all the occupations related to it, weighted by the number of times we observe each PCS in the ECMOSS database. We then rely on the ROME-degrees correspondence to calculate a simple average wage per degree from the ROME-specific salaries we just calculated.

For Luxembourg, we use data from the 'Structure of Earnings Survey' carried out by National Staistical Agency (STATEC). This survey includes a sample of companies in Luxembourg, and covers all economic activities (with the exception of the agriculture sector). The survey includes individual salaries based on employee profiles, the characteristics of the occupations and the profiles of their employers. We compute the average salary by job (at ISCO-4 level), translate the data to ROME codes using the ISCO-ROME correspondence, which then allows us to finally calculate the average salary by degree.

[12] https://www.francetravail.org/files/live/sites/peorg/files/documents/Statistiques-et-analyses/Open-data/ROME/Correspondance_ROME_ISCO08.xlsx

Figure 3 reports the distribution of wages for both countries.[13] The comparison of both densities illustrates the wage premium of working in Luxembourg rather than in France. The average monthly premium after taxes amounts to about 38000€, i.e. a 47% top-up for French workers in Luxembourg. The distribution of (gross) wages in Luxembourg also exhibits a higher variance.

Figure 3: Distribution of wages in France and Luxembourg



Note: The average (gross) wage for France is of 41.554€/month, with a standard deviation of 7.991. For Luxembourg, (gross) wages have an average of 79.338€/month and a standard deviation of 19.613.

[13]The density smoothing is calculated using a Gaussian kernel with Silverman's rule-of-thumb. The final smoothing parameter is of 6000.

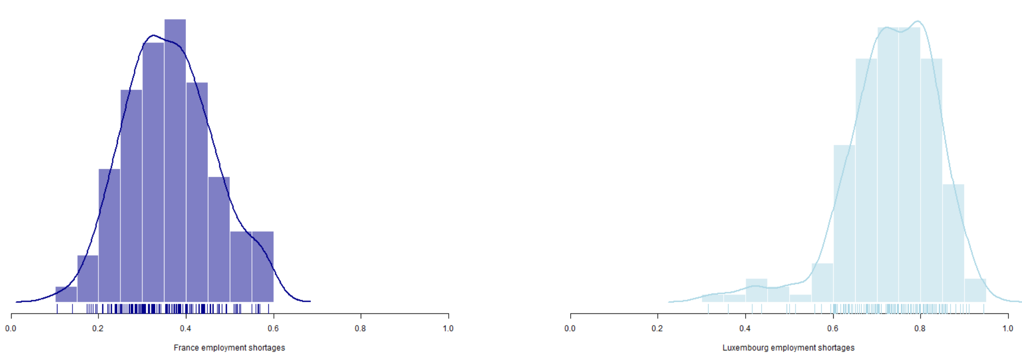
3.3.2 Labour demand indicators

Our measure of employability of graduates is based on indicators of labor market tightness on both countries. The indicator of tightness on the French labour market comes from the BMO survey (Besoins en Main-d'Œuvre). It measures hiring intentions, which reflect the labour needs of companies and opportunities for job seekers. They are also conditioned by recruitment difficulties, as assessed by employers. This database contains a percentage indicator of difficulties in recruiting workers for each occupation. Using the FAP-PCS-ROME correspondance, we assign a measure of market tightness to each job.[14] Finally, similar to what we did for wages, we take a simple average over all ROME codes using the ROME-degrees table to assign a final figure to each of the degrees.

The data for Luxembourg comes from the Agence pour le développement de l'emploi (ADEM), which is the public employment service in Luxembourg. This data relates to job offers without assignment, namely the percentage of vacancies that did not find suitable candidates. These unassigned offers are broken down by ROME code, which can therefore be associated with each degree in a similar fashion to what we did above using the ROME-degree table and taking averages.

Figure 4 reports the density plots for both countries.[15] Note that since the definition of both measures is not exactly the same, comparisons should be taken with caution. Nevertheless, it seems apparent that in Luxembourg there is a much higher need of workers across most occupations, making finding a job in that market easier than in France.

Figure 4: Distribution of labour shortages in France and Luxembourg



[14]In case of missing information in the correspondence for some ROME codes, we use the BMO average at the 3- or 2-digit ROME level

[15]The density smoothing is calculated using a Gaussian kernel with Silverman's rule-of-thumb. The final smoothing parameter is of 3000.

Note: The market shortage indicator in France has an average value of 36.2%, with a standard deviation of 9.9%. For Luxembourg, the average value is of 73.4% and the standard deviation of 10.4%.

4 Modeling educational choices and incentive effects

Our estimations rely on a discrete choice model based on the RUM model (equations 1 and 4) of educational choices. In the benchmark estimations, we will report results based on a multinomial logit (MNL) specification that assumes an extreme value distribution of type-1 for ϵ_{jn}. The flexibility of this model explains its overwhelming popularity in the education and migration literatures. Nevertheless, this flexibility comes at the cost of oversimplifying assumptions that can be questioned in our context. The remaining sections explore therefore the robustness of the benchmark results to other approaches lifting some of the underlying assumptions of the MNL model.

4.1 Benchmark results

Table 3 provides the estimations of equation (1-5) using the multinomial logit model. This model is based on the choice of an Extreme Value Distribution of type-1 for ϵ_{jn}. This distribution assumes that the independence from irrelevant alternatives (IIA) holds, whose validity might be questioned in this context. However, due to their flexibility, the multinomial logit estimations provide a first assessment of the existence of the incentive effect of emigration prospects in terms of educational choices. In this specification, beyond the economic factors of attractiveness in terms of employability and wages on the domestic and foreign markets, we account for a set of dummies. We include a master dummy that captures the relative attractiveness of a master degree compared to a bachelor one. We also include faculty dummies that account for the relative attractiveness of broad type of topics such as sciences or arts. Note also that in discrete choice models the utilities V_{jn} are unitless. Therefore, for comparison purposes across models, it is interesting to report normalized coefficients, i.e. coefficients expressed as a ratio of another one. Thus, we also report the scaled coefficients of the incentive effects related to wage and employability in Luxembourg, as a ratio of the impact of employability in France.

Column (2) of Table 3 includes the estimation results of the full model. The comparison with the more parsimonious specifications (columns 1, 3 and 4 of Table 3) suggests that the inclusion of all economic factors and all types of dummies is relevant. We find clear support for an incentive effect of the prospects of working in Luxembourg, since topics associated with a higher level of employability in Luxembourg tend to be chosen more often by students beyond the level of attractiveness from the domestic market alone. While employability on the domestic market remains the most important factor, estimations of columns (1-3) suggest that employability in Luxembourg plays a significant role. Regarding the wage level in Luxembourg, there is less overwhelming evidence of its importance in students' decisions, although all estimated coefficients are positive with a subset of these being significant.

While these estimations are supportive of an incentive effect, they rely on a set of assumptions and need therefore to be checked for robustness. This is discussed in the following sections.

Table 3: Incentive effect of Luxembourg: benchmark results

	Dependent var: probability of enrolment in topics			
	(1)	(2)	(3)	(4)
Empl France	3.67***	4.83***	4.74***	–
	(0.192)	(0.27)	(0.273)	
IntLux*Empl Lux (α_1)	1.61***	2.09***	2.530***	–
	(0.466)	(0.51)	(0.478)	
Wage France	0.062	0.549***	–	0.187
	(0.138)	(0.139)		(0.145)
IntLux*Wage Lux (α_2)	0.330*	0.282	–	0.610***
	(0.191)	(0.207)		(0.195)
Master	0.264***	0.187***	0.258***	0.248***
	(0.045)	(0.044)	(0.039)	(0.044)
Arts	–	0.195	0.278*	0.188
		(0.160)	(0.158)	(0.160)
Law, Econ and Man.	–	0.231	0.408***	0.460***
		(0.151)	(0.145)	(0.151)
Human and Soc Sc.	–	1.010***	1.008***	0.888***
		(0.149)	(0.146)	(0.150)
Sciences	–	0.249	0.370**	0.845***
		(0.152)	(0.148)	(0.147)
scaled (α_1)	0.438***	0.432***	0.533***	–
scaled (α_2)	0.089*	0.058	–	–
Obs	3038	3038	3038	3038
Nber of topics	58	58	58	58
Log-Lik.	-12147.82	-12046.24	-12054.41	-12209.73
LRT (p-val)	0.0000	–	0.0003	0.0000

Notes: Multinomial Logit estimation. Dependent variable: probability of enrolment in topics. Master dummy captures topics leading to a master degree (reference level: bachelor). Arts, LEM, HSS and Sciences dummies capture topics belonging to faculties (reference level : faculty of physical education). IntLux is a dummy identifying students with a very strong or strong interest for Luxembourg at time of enrolment (reference level: weak or no interest). LRT prvides p-value of a Likelihood ratio test of model against model of column (2). Scaled coefficients α_1 and α_2 are normalized estimates of incentive effects as a ratio of the coefficient of employability in France.

4.2 Accounting for heterogenous substitution patterns

The multinomial logit model that yields the estimations in Table 3 rests on an important assumption, namely the hypothesis of Independence of Irrelevant Alternatives (IIA). In our context, the IIA hypothesis implies that the substitution rate between study fields is the same. As an illustration, under this hypothesis, an increase in the economic attractiveness of, say, mathematics, will have the same (negative) impact on the probability of enrolment in French Literature, Computer Science or Biomedicine. In the real world, we might be concerned that this assumption is violated for a set of different reasons. First, students have specific preferences for categories of topics. For instance, students might be interested by topics related to the understanding of societies such as sociology, economics or management. In this case, we might expect that substitutions between these topics might be higher compared to any of this topic with ones belonging to a different category. A second reason is the background of students. Some topics might require some specific background. This is, for instance, the case in quantitative fields, such as mathematics. It might be expected that substitution between topics belonging to these categories will be higher. Or, put differently, students with little background in mathematics will exhibit a low substitution rate from, say, French literature, to Physics even in the presence of an increase in the attractiveness of this latter topic.

We analyse the robustness of our results with respect to the incentive effect of prospective emigration by estimating alternative models that allow for a deviation from the IIA assumption. In the discrete choice literature, the way to deal with this is to specify a different distribution for ϵ_{jn} in equation (1). We consider two alternative models, the (multinomial) Nested Logit model (NL) and the Cross-Nested model (CNL). We expose here in a non-technical way the main features of the two alternative models as well as their specific contribution in terms of implied substitution patterns. Appendix A provides the technical details of the estimation of these models for the interested reader.

The NL model specifies the categories of alternatives that are expected to exhibit similarities in the stochastic component of utilities (ϵ_{jn}). Topics included among each category are supposed to be more similar compared to topics outside the category. Categories are reflected by nests in the model and are chosen ex-ante based on theoretical reasons. However, since the MNL model is nested in the NL model, likelihood ratio tests can be used to validate the choice of the nests. We use two alternative dimensions to define the nest structure. In the first NL model, we consider nests based on topics with and without a significant quantitative dimension. Topics such as chemistry, physics and economics belong to the quantitative nest, while law or literature belong to the non-quantitative nest. In the second NL model, we make a distinction between topics addressing societal issues and topics without this dimension. Topics like sociology or economics belong to the first category, while literature, mathematics and medicine belong to the second one. Appendix A provides a classification of each field along both dimensions.

The second model is an extension of the NL model. Instead of partitioning the set of topics using one dimension, the Cross-Nested Logit Model (CNL) combines the various dimensions to define overlapping nests. In our context, each topic belongs to one of the four possible nests that combine quantitative and societal criteria. For instance, economics belongs to a nest including quantitative and societal topics; mathematics belongs to a quantitative- non-societal nest; sociology to a non-quantitative-societal nest and so on. This approach allows for a more flexible way of capturing complex substitution patterns across topics. Once again, substitution is supposed to be higher between topics within the same nests than across the nests.

Table 4 provides the estimations for the various models. The specification follows the MNL model of column (2) from Table 3 that is best supported by the data. This specification includes both economic factors of attractiveness in both markets as well as the full set of degree and faculty dummies. Columns (2) and (3) of Table 4 provide the NL estimates with each partitioning criterion. Columns (4) and (5) provide the CNL estimations combining both criteria. Since these models are highly non-linear, it is desirable to constraint some coefficients such as the similarity parameters. This is done in the estimations of column 5 for the μ parameters of the non-quantitative nest. The bottom line is that the results support the relevance of each underlying criterion.[16] This suggests that the IIA hypothesis and the homogeneity of substitution patterns between topics are rejected by the data.

The estimation results of Table 4 support once again the existence of the incentive effect of emigration prospects. Most of the estimations support an incentive effect associated to employability in Luxembourg. Nevertheless, there is also moderate support for an incentive effect in terms of wage conditions, for instance from the estimations of the best CNL model (column 5). Overall, these results show that the evidence of an incentive effect drawn from the MNL estimations in Table 3 holds when we account for potential deviations from the IIA hypothesis.

[16]In statitical terms, all null hypotheses $H_0 : \mu = 1$ are rejected in favour of $H_A : \mu > 1$.

Table 4: Incentive effect of Luxembourg: heterogenous substitution patterns

	Dependent var: probability of enrolment in topics				
	(1)	(2)	(3)	(4)	(5)
Empl France	4.83***	1.030***	4.87***	1.37***	2.38***
	(0.27)	(0.145)	(0.225)	(0.149)	(0.152)
IntLux*Empl Lux	2.09***	0.222***	1.920***	0.263	0.41*
	(0.510)	(0.090)	(0.451)	(0.188)	(0.235)
Wage France	0.062	-0.013	-0.211*	0.134**	-0.129*
	(0.138)	(0.021)	(0.125)	(0.052)	(0.066)
IntLux*Wage Lux	0.282	0.061**	0.386**	0.095***	0.334***
	(0.207)	(0.028)	(0.170)	(0.024)	(0.068)
Master	0.187***	0.061***	0.261***	0.049***	0.154***
	(0.044)	(0.009)	(0.037)	(0.018)	(0.019)
Arts	0.195	0.027**	0.249**	0.034	0.109**
	(0.160)	(0.014)	(0.120)	(0.050)	(0.055)
Law, Econ and Man.	0.231	-0.080***	0.133	-0.118**	-0.179***
	(0.151)	(0.151)	(0.114)	(0.049)	(0.056)
Human and Soc Sc.	1.010***	0.083***	0.834***	0.082*	0.204***
	(0.149)	(0.149)	(0.110)	(0.046)	(0.048)
Sciences	0.249	-0.215***	-0.586***	-0.356***	-0.544***
	(0.152)	(0.152)	(0.120)	(0.057)	(0.068)
scaled (α_1)	0.432***	0.215***	0.394***	0.192	0.172*
scaled (α_2)	0.058	0.059**	0.079**	0.069***	0.140***
μ_{quant}	–	3.82***	–	3.21***	1.60***
		(0.355)		(0.530)	(0.085)
μ_{noquant}	–	13.40***	–	99.2***	20***
		(2.020)		(11.1)	(1.18)
μ_{soc}	–	–	1.35***	3.21***	2.36***
			(0.027)	(0.231)	(0.107)
μ_{nosoc}	–	–	1	2.36***	2.23***
				(0.157)	(0.146)
Obs	3038	3038	3038	3038	3038
Nber of topics	58	58	58	58	58
Log-Lik.	-12046.24	-11729.18	-11936.97	-11468.53	-11451.3
LRT (p-val)	–	0.00	0.00	0.00	0.00

Notes: col (1) Multinomial Logit estimation. cols (2) and (3) Nested Logit estimations. In col(2), nest dimensions: quantitative and non-quantitative topics. In col(3), nest dimensions: social and non-societal topics. Cols (4) and (5) : Cross-Nested Logit estimations. Participation parameters set to 0.5. In col(3), μ_{nosoc} constrained to 1. In col (4) unconstrained estimation. In col (5) constrained estimations with bound set to 20 for μ parameters. Tests based on Null hypothesis $\mu = 1$. Dependent variable: probability of enrolment in topics. Master dummy captures topics leading to a master degree (reference level: bachelor). Arts, LEM, HSS and Sciences dummies capture topics belonging to faculties (reference level : faculty of physical education). IntLux is a dummy identifying students with a very strong or strong interest for Luxembourg at time of enrolment (reference level: weak or no interest). LRT provides p-value of a Likelihood ratio test of model against MNL model of column (1). Scaled coefficients α_1 and α_2 are normalized estimates of incentive effects as a ratio of the coefficient of employability in France.

4.3 Placebos

We also run a set of placebo tests to test the existence of a potential incentive effect of the Luxembourgish labour market among students with no interest for Luxembourg. Since they claim that they did not have some interest for the Luxembourgish labour market, the intuition would suggest that the variations in attractiveness of various topics associated to this market should not affect their choice. In doing so, we estimate the following equation (5):

$$V_{jn}^{(pl)} = V_{jn} + \gamma_1[(1 - I_n) \times \Pr(e_j^*)] + \gamma_2[(1 - I_n) \times \log(w_j^*)] \tag{9}$$

and test whether the γ coefficients are significant and sensible. Table 5 report the results for these estimations. γ_1 is insignificant across all specifications, which suggests that variations in employability had no impact on the enrolment of students with no interest for Luxembourg. γ_2 coefficients exhibit a negative sign which is counterintuitive. This holds irrespectively of the inclusion or exclusion of the coefficient for students showing interest in Luxembourg at enrolment. All in all, these results support the fact that, in contrast to those paying attention to Luxembourg, other students not interested by moving abroad were not subject to the incentive effect of foreign opportunities.

Table 5: Incentive effect of Luxembourg: placebos

	Dep. var: probability of enrolment in topics		
	(1)	(2)	(3)
	MNL	MNL	CNL
Empl France	4.86***	4.79***	2.49***
	(0.192)	(0.27)	(0.151)
IntLux*Empl Lux	1.83***	–	–
	(0.516)		
(1-IntLux)*Empl Lux	-0.127	-0.22	-0.376
	(0.307)	(0.30)	(0.113)
Wage France	0.997***	1.03***	-0.003
	(0.163)	(0.151)	(0.068)
IntLux*Wage Lux	-0.238	–	–
	(0.218)		
(1-IntLux)*Wage Lux	-0.841***	-0.852***	-0.132**
	(0.136)	(0.13)	(0.057)
Master	0.307***	0.321***	0.200***
	(0.051)	(0.048)	(0.023)
Arts	0.178	0.185	0.123**
	(0.160)	(0.160)	(0.057)
Law, Econ and Man.	0.372**	0.360**	-0.166***
	(0.154)	(0.153)	(0.066)
Human and Soc Sc.	1.100***	1.090***	0.230***
	(0.151)	(0.121)	(0.051)
Sciences	0.343**	0.363**	-0.531***
	(0.154)	(0.153)	(0.071)
μ_{quant}	–	–	1.56***
			(0.088)
μ_{noquant}	–	–	20***
			(1.090)
μ_{soc}	–	–	2.32***
			(0.106)
μ_{nosoc}	–	–	2.21***
			(0.135)
Obs	3038	3038	3038
Nber of topics	58	58	58
Log-Lik.	-12034.44	-12039.01	-11453.43

Notes: Cols (1) and (2): Multinomial Logit estimation. Col (3) CNL with 4 nests. μ_{noquant} constrained to 20. Dependent variable: probability of enrolment in topics. Master dummy captures topics leading to a master degree (reference level: bachelor). Arts, LEM, HSS and Sciences dummies capture topics belonging to faculties (reference level : faculty of physical education). IntLux is a dummy identifying students with a very strong or strong interest for Luxembourg at time of enrolment (reference level: weak or no interest).

4.4 Endogeneity of interest variable

A final concern about our benchmark results regarding the incentive effect is the potential endogeneity of our interest variable. This variable is interacted with the variables relative to the attractiveness of the Luxembourgish labour market, namely employment and the wage level. In fact, it could be argued that this variable is endogenous, as it could be correlated with unobserved factors that also affect the choice of study field. While this issue might concern only a small subset of individuals, the following example can be used to clarify its nature. Suppose that an individual has a strong preference for matters related to the ocean. This individual will at the same time express none or very little interest for Luxembourg since it is landlocked, but also a strong preference for topics such as maritime law or naval engineering. This joint influence could, in principle, bias the estimation of the parameters associated to the idea of an incentive effect.

Endogeneity issues in discrete choice models such as ours have been addressed in the literature. See Guevara and Ben-Akiva (2010) for a review of the methods dealing with endogeneity in discrete choice models. The typical approach relies on a control function (CF) approach in the estimation of model (5). We provide the details of this approach in Appendix B.

The CF approach is the equivalent of an instrumental variable (IV) estimation for non-linear models such as the discrete choice models (see Wooldridge (2015) for a general description of the CF approach). In a nutshell, it requires in a first stage the use of an instrument that is used to predict the endogenous variable. The residuals of this first-stage regression are then included in the estimation of the choice model. The inclusion of this additional term allows for correction of the endogeneity bias. Furthermore, the coefficient of this residual variable is indicative of the size and magnitude of this bias.

There are nevertheless two complications to the usual CF approach in our context. The first one is that our endogenous variable is interacted rather than included autonomously in equation (5). The solution to this is to use as the instrument the product of the instrument and the variable. More specifically, if Z_n is the instrument of the variable capturing the degree of interest for Luxembourg expressed by individual n, we use $Z_n * EmplLux$ as the instrument for $IntLux_n * EmplLux$. The same applies to the wage in Luxembourg. The second complication is that, due to the specification of model (5), we end up with two endogenous variables instead of one. The estimation of multiple endogenous variables in empirical work is often not advised. Therefore, we proceed to successive CF estimations, considering either $interest_n * EmplLux$ or $interest_n * wageLux$ the endogenous variable alone.

The implementation of the CF requires the choice of an instrument. This instrument

Table 6: Distance, contiguity and interest for Luxembourg

	Dependent Var: Interest for Luxembourg					
	All students			with Interest for abroad		
	(1)	(2)	(3)	(4)	(5)	(6)
Contiguity	0.185***	0.349***	–	0.258***	0.603***	–
	(0.047)	(0.038)		(0.098)	(0.075)	
Log distance	-0.111***	–	-0.145***	-0.194***	–	-0.239***
	(0.022)		(0.019)	(0.034)		(0.03)
Female	-0.183***	-0.168***	-0.190***	-0.227***	-0.199***	-0.231***
	(0.037)	(0.037)	(0.037)	(0.068)	(0.070)	(0.068)
Foreign	0.508***	0.295***	0.510***	0.587***	0.242**	0.573***
	(0.075)	(0.058)	(0.080)	(0.193)	(0.010)	(0.134)
Constant	2.015***	1.439***	2.265***	3.589***	2.266***	3.936***
	(0.119)	(0.033)	(0.095)	(0.193)	(0.069)	(0.133)
Nber obs.	3036	3036	3036	931	931	931
R^2	0.050	0.034	0.044	0.119	0.072	0.111

Notes: Dependent variable: interest for Luxembourg expressed at the time of enrolment. Scale: 1-4, with 1 if no interest and 4 if strong interest.
Distance is minimal distance from home at time of enrolment to closest point on the Luxembourguish border. Contiguity : 1 if lived in a department contiguous to Luxembourg.

should predict the interest for Luxembourg while not being correlated with some preferences with respect to the field of study. In our approach, we rely on the location of the parents of the student at the time of enrolment. The idea is that the location of the parents and, therefore, the initial living place of the student is the result of the location choice of the parents which can be considered exogenous to any preference regarding study fields. We then use distance between this location and Luxembourg as a predictor of the interest in luxembourg variable. As a preliminary piece of evidence, Table 6 provides some evidence that distance-related variables predict the probability as well the magnitude of the interest expressed with respect to Luxembourg. The results suggest that students having lived in a contiguous French department to Luxembourg tend to express a higher interest for the country. Also, the higher the distance to Luxembourg, the lower this interest becomes. These preliminary findings suggest that contiguity or distance can be used to generate instruments in the control function approach.

Table 7 presents the CF estimates of equation (5). We use the multinomial logit specification. We provide five different estimations depending on the instrumented variable(s) and the choice of the instrument (contiguity and/or distance to Luxembourg). The first stage estimates corresponding to the CF estimations are provided in Table 14 in Appendix B.
By tackling only one of the endogenous variables at a time, as done in columns (1)–(4), results are mainly unaffected, suggesting a very low impact of the endogeneity problem. We got the same type of result in column (5) when we include simultaneously the interaction of interest in Luxembourg and wages and employability instrumented by measures of distance and contiguity.

4.5 Additional checks and extensions

- Excluding foreigners
- Excluding Non EU graduates (to do)
- Higher level of the interest variable
- Using variable Lux as a deciding factor
- Using variable Lux as a deciding factor for native students
- Using wage data for under 30.

Table with 6 columns with estimates from MNL.
In this section, we look at various variants of our benchmark set-up to check the robustness of our estimates of the incentive effect. We also run sample specific regressions to check that the incentive effect varies with the expected change from the full sample, which brings further support for the validity of our estimates.

4.5.1 Native students and EU students only

The incentive effect might vary across origins of the students. In particular, for specific reasons, it can be different between native students on the one hand and foreign students on the other one. It can also be different between EU and non EU students for other reasons. Given the nature of the incentive effect, we can expect that native students will be subject more to the incentive effect than foreign students. One of the reasons is that foreign students also contemplate an additional location alternative, i.e. their origin country. It is well documented that return rates of foreign

Table 7: Incentive effect of Luxembourg: endogeneity of interest

	Dependent var: probability of enrolment in topics				
	(1)	(2)	(3)	(4)	(5)
Empl France	4.83***	4.83***	4.83***	4.83***	4.83***
	(0.27)	(0.27)	(0.27)	(0.27)	(0.27)
Int.*Empl Lux	2.09***	2.09***	2.09***	2.09***	2.09***
	(0.510)	(0.510)	(0.510)	(0.510)	(0.510)
Wage France	0.549***	0.549***	0.549***	0.549***	0.549***
	(0.139)	(0.139)	(0.139)	(0.139)	(0.139)
Int*Wage Lux	0.282	0.282	0.282	0.282	0.282
	(0.207)	(0.207)	(0.207)	(0.207)	(0.207)
Master	0.187***	0.187***	0.187***	0.187***	0.187***
	(0.044)	(0.043)	(0.043)	(0.043)	(0.043)
Arts	0.195	0.195	0.195	0.195	0.195
	(0.160)	(0.160)	(0.160)	(0.160)	(0.160)
Law, Econ and Man.	0.231	0.231	0.231	0.231	0.231
	(0.151)	(0.151)	(0.151)	(0.151)	(0.151)
Human and Soc Sc.	1.010***	1.010***	1.010***	1.010***	1.010***
	(0.149)	(0.149)	(0.149)	(0.149)	(0.149)
Sciences	0.249	0.249	0.249	0.249	0.249
	(0.152)	(0.152)	(0.152)	(0.152)	(0.152)
$\hat{\nu_{jn}}$	0.000*	0.000	-0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Endog. var. 1	Int*Empl	Int*Empl	Int*Wage	Int*Wage	Int*Empl
Endog. var. 2	–	–	–	–	Int*Wage
Instrument 1	Contig.	Dist	Contig.	Dist	Contig.
Instrument 2	–	–	–	–	Dist
Nber Obs	3038	3038	3038	3038	3038
Nber of topics	58	58	58	58	58
Log-Lik.	-12046.24	-12046.24	-12046.24	-12046.24	-12046.24

Notes: col (1) Multinomial Logit estimation. $\hat{\nu_{jn}}$ is the residual of a first stage estimation regressing the endogeneous variable(s) on the instrument(s) indicated on the two last lines. First stage estimations are reported in Table 9 in the Appendix C. Instruments are contiguity and/or distance between initial location of the students at the time of enrolment and Luxembourg. Scaled coefficients α_1 and α_2 are normalized estimates of incentive effects as a ratio of the coefficient of employability in France.

students are quite high, even in attractive destinations, which has been flagged as an issue for the hosting country that supports a substantial part of the cost of education (reference Chaloff). The reason is that beyond economic incentives, individuals have strong preferences for living in their country. Therefore, the incentive effect exerted by Luxembourg might be mitigated by the existence of this important alternative location. We explore this by running the MNL model of equations (1-4) excluding foreign students. Foreign students represent about 14% of students in our sample. Column (1) of Table XXX provides the new estimates. We find that the incentive effect associated to employability in Luxembourg is stronger in the sample of native students compared to the full sample.
We also consider a sample of EU students only since we might expect that the incentive effect will be higher compared to the non EU students. The reason is that for non EU students, due to visa restrictions and other regulations governing for instance the residence of cross-border workers in Luxembourg, $P(mig = 1)$, the probability of working in Luxembourg will be lower than 1. This in turn should lower the expected foreign wage and the foreign wage premium, making the incentive effect less important.Column (2) of Table XXX provides the new estimates obtained on a sample of EU students only. We find that ....

4.5.2 Alternative measures of the interest for Luxembourg

In the benchmark estimations, we have used an interest variable for Luxembourg based on the two higher levels of this variable (strong and very strong interest). This variable is important as it captures the possible efforts of collecting infrmation about the Luxembourguish labour market at the time of enrolment. In a variant to the benchmark results, we use only the highest modality of that variable, looking specifically at the students who stated a very strong interest for Luxembourg. We might expect the incentive effect to be higher for these ones. Column (3) of Table XXX provides the estimates and confirms this expectation.
In our survey, we have also another variable that we did not use so far. We ask more directly whether Luxembourg was a deciding factor for educational choices at the time of enrolment. Note that only 5.5% of the students replied positively, which means that we consider here a very specific section of the population of interest. Once again, we might expect the incentive effect to be much higher for these ones. Column (4) of Table XXX provides the estimates and confirms this expectation.
*** combination of the 2 ***

4.5.3 Using wage data for young workers

5 Conclusion

In this paper, we assess a new kind of incentive effect of emigration prospects in terms of human capital accumulation. The existing literature has mostly looked at whether attractive emigration prospects induced individuals to invest more in education in their origin country. Evidence of such an incentive effect has been provided in terms of the general level of human capital level, but much less in terms of the specific type of human capital. Furthermore, the incentive effect has been explored mostly in a context of South-North migration prospects, i.e. emigration from developing to developed countries. No evidence has been provided in the context of human mobility between two developed countries.

To shed some light of such an incentive effect, we take advantage of a survey conducted on graduates of the University of Lorraine, located in the North-East of France. The region of Lorraine is located near the country of Luxembourg, which enjoys a booming economy based on the development of financial activities and high-tech services to firms. The Luxembourgish labour market offers very attractive opportunities for workers of the Lorraine region, with minimal costs in terms of mobility, cultural and linguistic adjustment as well as administrative procedures. We leverage data on individual enrolment and graduation in a large set of study subjects and tests the existence of the incentive effect of migration prospects. We find evidence that students tend to invest more in human capital associated to occupations that offer high attractive returns in Luxembourg. The attractiveness of the Luxembourgish labour market is captured by two dimensions: employability (i.e. probability of employment) and wage conditions. We find more evidence in favour of the first dimension, even though some results support some evidence of an effect associated to wage conditions.

Our results are specific to students who stated that they paid attention to the foreign labour market at the time of enrolment, providing some evidence that the incentive effect depends on the acquisition of some information about foreign opportunities. Students that did not consider the foreign labour market in general, or Luxembourg in particular, do not seem to be affected by the attractiveness of the foreign labour market when making educational choices. The results are robust to a set of considerations that could affect the validity of the results. First, the initial interest for Luxembourg might be endogenous, which could bias the estimation of the incentive effect. We tackle this by taking advantage of the initial location of the students before enrolment and show that students living close to Luxembourg are more likely to pay attention to the foreign labour market. The incentive effect is still found when this source of endogeneity is taken into account in the estimations. Second,

we account for the heterogeneous substitution patterns between study topics by estimating a more advanced discrete choice model. Partitioning the choice set of topics along two dimensions (societal and quantitative topics), we find robust evidence of the incentive effect of emigration prospects.
The existence of such an incentive effect is much more than an intellectual curiosity and entails important potential implications for the economic development of regions and countries. To the extent that there are differences in the industrial structures, the existence of such an incentive effect might lead to an underinvestment in skills that are needed in the region of origin of the students. Therefore, at least in the short-run, this incentive effect might worsen the issue of skill mismatch and skill shortages observed in many regions of developed countries. Interestingly, our context provides a good example of such a case. While Luxembourg is an economy dominated by the development of financial and high-tech services, the Lorraine region is characterized by a more traditional industrial structure based on manufacturing activities. Like many regions in Western Europe, the Lorraine economy is shaped by skill shortages in many important sectors. Pole Emploi, the public organization in charge of the monitoring of the labour market in France, has often claimed that these skill shortages are amplified by the brain drain to Luxembourg. Nevertheless, the brain drain is only one face of the coin. Brain drain implies a depletion of human capital at origin in favour of foreign regions or countries after the acquisition of skills. What our incentive effect suggests is that, regardless of the intensity of the brain drain, there is an effect of emigration prospects in terms of the *composition* of skills that can also, at least in the short-run, be detrimental to the region of origin.
However, these pernicious consequences might be offset in the long run. Acquisition of new skills by individuals might induce a skill-biased technological change that might be beneficial over time for the region of origin. This will be the case provided that the magnitude of the brain drain is moderate. In that sense, the incentive effect of emigration prospects in terms of specific skills might lead to the same phenomenon of a long-run beneficial brain-drain identified in the previous literature (Beine et al. (2008), Mountford (1997); Docquier and Rapoport (2012)). Therefore, the implications of the incentive effect in terms of skills might be very different depending on the time horizons. We leave this investigation for further research.

References

Abarcar, P. and Theoharides, C. (2020). Medical worker migration and origin-country human capital: Evidence from u.s. visa policy. Working paper, Amberst College.

Batista, C., Lacuesta, A., and Vincente, P. (2012). Testing the 'brain gain' hypothesis: Micro evidence from cape verde. *Journal of Development Economics*, 97(1):32–45.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2):9–49.

Beine, M., Docquier, F., and Rapoport, H. (2001). Brain drain and economic growth: Theory and evidence. *Journal of Development Economics*, 64(1):275–289.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain drain and human capital formation in developing countries: Winners and losers. *The Economic Journal*, 118(528):631–652.

Bhagwati, J. N. (1976). Taxing the brain drain. *Challenge*, 19(3):34–38.

Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, 144(1):287–300.

Breen, R. and Goldthorpe, J. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society*, 9(5, Part 2):275–305.

Cameron, S. V. and Heckman, J. J. (1998a). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political economy*, 109(3):455–499.

Cameron, S. V. and Heckman, J. J. (1998b). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political economy*, 106(2):262–333.

Chand, S. and Clemens, M. (2019). Human capital investment under exit options: Evidence from a natural quasiexperiment. Discussion Paper 12173, IZA Institute of Labor economics.

Chapman, D. (2012). A model of student college choice. *Journal of Higher Education*, 52:490–505.

Dinkelman, T. and Mariotti, M. (2016). The long run effect of labor migration on human capital formation in communities of origin. *American Economic Journal: Applied Economics*, 8(4):1–35.

Docquier, F. and Rapoport, H. (2012). Globalization, Brain Drain, and Development. *Journal of Economic Literature*, 50(3):681–730.

Gibbons, S. and Vignoles, A. (2012). Geography, choice and participation in higher education in england. *Regional Science and Urban Economics*, 42:98–113.

Gibson, J. and McKenzie, D. (2010). How does high-skilled emigration affect small countries? microeconomic evidence from tonga. *Annals of Economics and Statistics*, (97/98):167–185.

Guevara, C. and Ben-Akiva, M. (2010). Addressing endogeneity in discrete choice models: Assessing control-function and latent-variable methods.

Kulka, A., Nikolka, T., Poutvaara, P., and Uebelmesser, S. (2023). International applicability of education and migration aspirations. CESifo Working Paper 10395, Munich.

Mattoo, A., Neagu, I. C., and Çağlar Özden (2008). Brain waste? educated immigrants in the us labor market. *Journal of Development Economics*, 87(2):255–269.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. New York: Academic Press.

Mountford, A. (1997). Can a brain drain be good for growth in the source economy?". *Journal of Development Economics*, 53(2):287–303.

Shrestha, S. (2017). No manleft behind: Effects of emigration prospects on educational and labour outcomes of non migrants. *Economic Journal*, 127(600):495–521.

Spence, M. (1978). Job market signaling. In *Uncertainty in economics*, pages 281–306. New York: Academic Press.

Stark, O., Helmenstein, C., and Prskawetz, A. (1997). A brain gain with a brain drain?". *Economics Letters*, 55(3):227–234.

Statec (2023). Luxembourg in figures. Technical report.

Theoharides, C. (2018). Manila to malaysia, quezon to qatar: International migration andthe effects on origin-country human capital. *Journal of Human Resources*, 53(4):1022–1049.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Vidal, J.-P. (1997). The effect of emigration on human capital formation. *Journal of Population Economics*, 11(4):589–600.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

Appendix A Accounting for unobserved heterogeneity in substitution

This section details how we account for the potential heterogenous substitutions across studied topics. The literature has extended the logit model and generated more complex models that take into account the fact that substitution across a subset of alternatives (here topics) can be higher or lower than with the rest of these alternatives. This issue is related to the well-known violation of the Independence of Irrelevant Alternatives (IIA) assumption that underlies the use of the logit (MNL) specification in equation 1:

$$U_{jn} = V_{jn} + \epsilon_{jn}. \tag{10}$$

The relevance of the MNL specification relies on the validity of the IIA hypothesis. In our context, IIA implies that any pair of topics exhibit the same substitution among the whole choice set of study fields. Statistically speaking, the validity of the IIA hypothesis implies that ϵ_{jn} follows a extreme value distribution of type 1, which in turn implies that there is no correlation of ϵ_{jn} across any pair of j alternatives. The logit model implies very restrictive substitution patterns that can be visualized by computing, the cross-elasticity, i.e. the change in the probability of choosing a particular topic linked to a change in the value of an attribute z_{jn} (e.g. wage or employability) specific to another topic (Train (2009)):

$$\frac{\partial P_n(j|C)}{\partial z_{kn}} = -\gamma_z P_n(j|C) P_n(k|C). \tag{11}$$

The corresponding elasticity is given by:

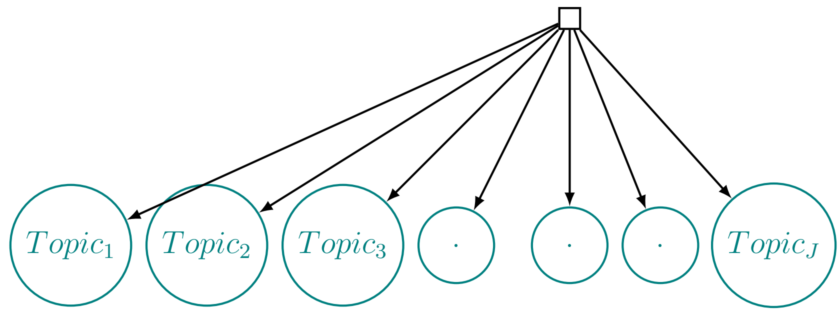
$$E_{j,z_{kn}} = -\gamma_z z_{kn} P_n(k|C), \tag{12}$$

where γ_z is the estimated effect of topic z. The cross-elasticity for destination j implied by the logit model is the same across all other topics (i.e., it does not depend on the specificity of topic j).

Figure 5 provides a graphical representation of the partitioning of the choice set of topics at work in the logit specification.

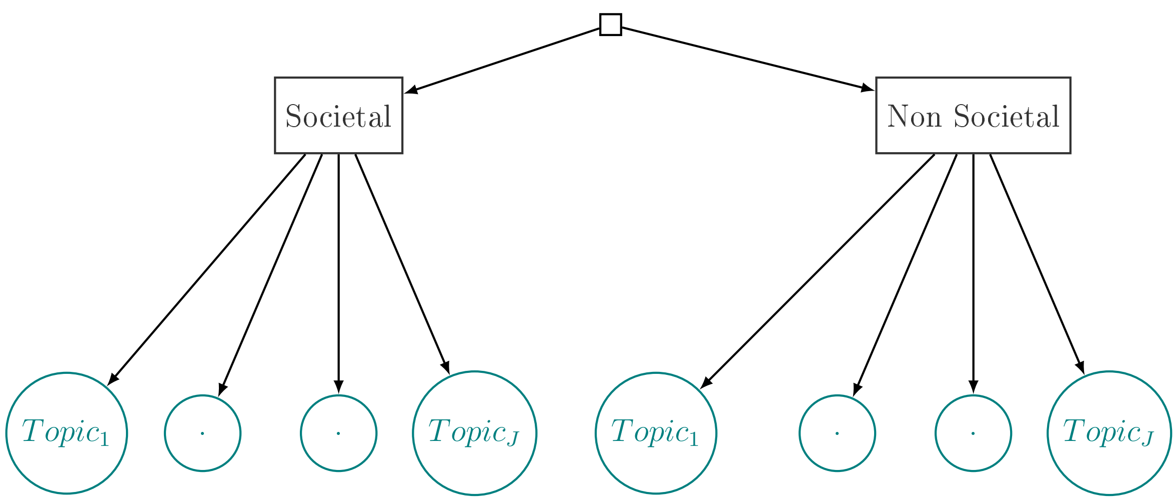
The Nested Logit model (NL) breaks down the hypothesis of uncorrelated ϵ_{jn} by creating nests of topics within which the substitution is supposed to be higher than with other topics outside the nest. This is done by assuming a new distribution for ϵ_{jn}, i.e. a specific version of the multivaratiate extreme value distribution (for more details, see Bierlaire (2006)). Under this distribution, each topic is assigned to a category of topics in which the unobserved similarity is supposed to be higher, i.e.

Figure 5: Graphical representation of the Logit model for Study fields.



the substitution is higher than with topics outside the category. In our estimation, we consider two types of categories. In the first approach, we suppose that students make a distinction between quantitative and non quantitative topics. In the second approach, we suppose that students make a distinction between topics related to the analysis of the society (societal) and other topics (non societal). Figure 6 graphically represents how the nested logit model partitions the choice set of topics.

Figure 6: Graphical representation of Nested Logit for study fields



A similar figure can be drawn for the second approach distinguishing quantitative topics from non-quantitative ones.
The Cross-Nested Model (CNL) also breaks down the hypothesis of uncorrelated ϵ_{jn} but combines the above chosen categories of topics by creating overlapping nests. In

our context, each topic might belong to four nests: societal-quantitative, non societal-quantitative, societal-non quantitative, non societal-non quantitative. Statistically, the CNL relies on the Generalised Multivariate Extreme Value Distribution with the following generating function G:

$$G(e^{\epsilon_{0n}}, ..., e^{\epsilon_{Jn}}) = \sum_{m=1}^{M} \left(\sum_{j=0}^{J} (\alpha_{jm}^{\frac{1}{\mu}} e^{\epsilon_{jn}})^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \tag{13}$$

with $\alpha_{jm} \geq 0, \frac{\mu}{\mu_m} \leq 1$ and $\forall j$, $\exists m$ such that $\alpha_{jm} \geq 0$.
In this model, the parameters μ_ms capture the similarity between the ϵ_{jn}s within nest m. The α_{jm} parameters are participation parameters, capturing the extent to which topic j belongs to nest m. In the CNL, μ_m and α_{jm} jointly capture the correlation between the topics.[17] This specification generalizes the NL model, in which each topic is assigned to a single nest (i.e., $\alpha_{jm} = 1$ for one m, and 0 for the others). In the CNL specification, this restriction is relaxed. The CNL imposes the normalisation constraint that $\sum_{m=1}^{M} \alpha_{jm} = 1 \ \forall j$. Therefore, the NL model might be seen as a linear restriction of the CNL model. In turn, the logit model can be obtained as a particular case of the NL with $\frac{\mu}{\mu_m} = 1$ for each m.
The partition of the choice set by the CNL can be represented by figure 7.
We set $\alpha_{j,NonSoc} = \alpha_{j,Soc} = \alpha_{j,NonQuantit} = \alpha_{j,Quant} = 0.5$ in order to comply with the normalisation constraint $\sum_{m=1}^{M} \alpha_{jm} = 1. \ \forall j$.
Table 8 lists all the topics with their respective assignment to each nest.

[17]See Bierlaire (2006) for a discussion of the conditions to define a GEV function and its properties. In particular this G has properties of non negativity and homogeneity, and complies with some limit properties and the sign of its derivatives. The CDF of the MEV distribution and the expected maximum utility can be directly derived from G.

Figure 7: Graphical representation of the CNL for study fields .

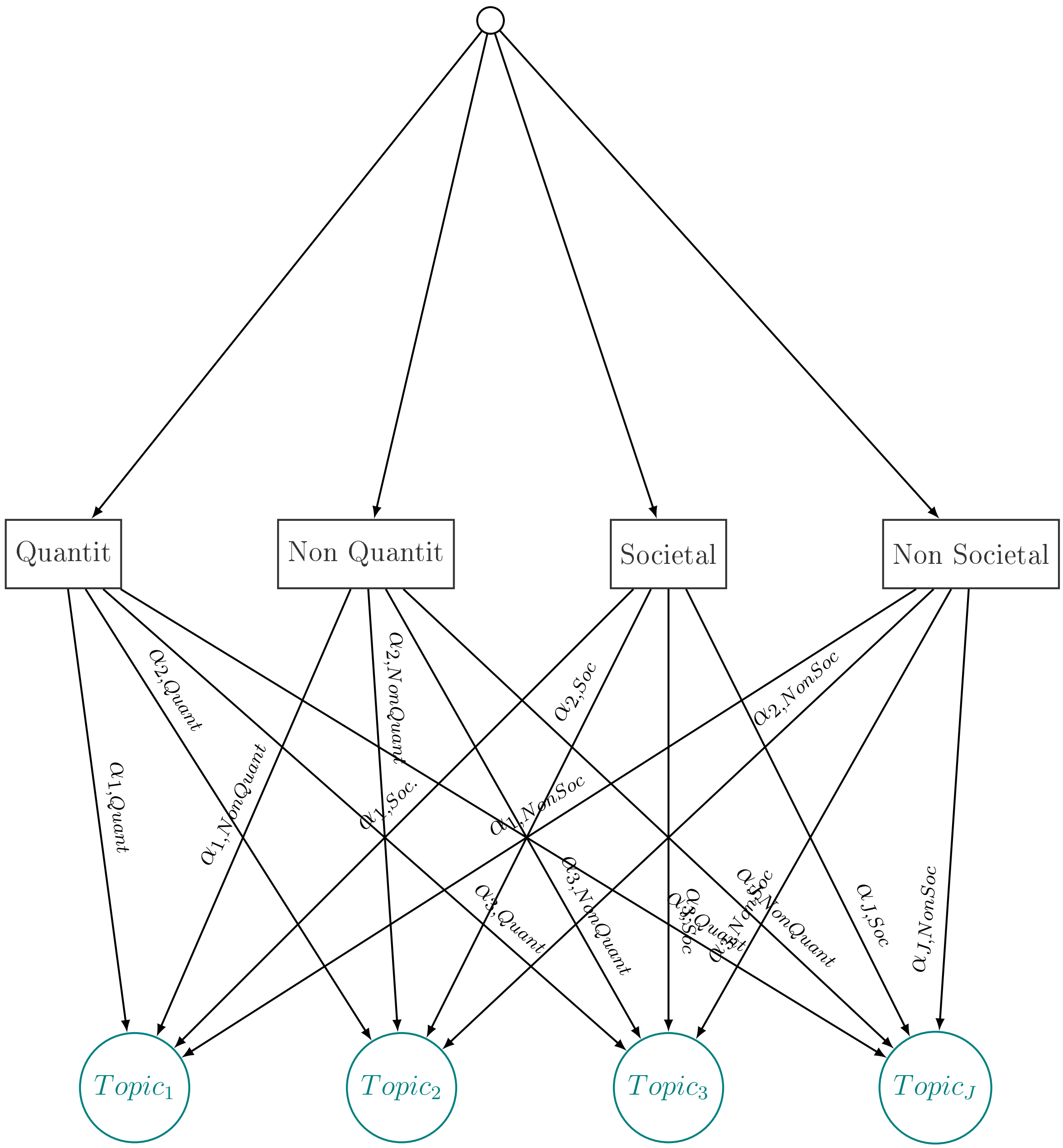


Table 8: Assignment of each study field to nests in the CNL model

Topic	$\alpha_{j,Quant}$	$\alpha_{j,NonQuantit}$	$\alpha_{j,Societal}$	$\alpha_{j,NonSocietal}$
Agronomy bachelor	0.5	0	0	0.5
Agronomy master	0.5	0	0	0.5
Applied economics master	0.5	0	0.5	0
Arts bachelor	0	0.5	0.5	0
Arts master	0	0.5	0.5	0
Biology bachelor	0.5	0	0	0.5
Biology master	0.5	0	0	0.5
Business law master	0	0.5	0.5	0
Chemistry bachelor	0.5	0	0	0.5
Chemistry master	0.5	0	0	0.5
Civil engineering bachelor	0.5	0	0	0.5
Civil law master	0	0.5	0.5	0
Communication bachelor	0	0.5	0.5	0
Communication master	0	0.5	0.5	0
Economics bachelor	0.5	0	0.5	0
Education master	0	0.5	0.5	0
Engineering bachelor	0.5	0	0	0.5
Engineering master	0.5	0	0	0.5
Fashion bachelor	0	0.5	0.5	0
Finance master	0.5	0	0	0.5
Foreing languages bachelor	0	0.5	0.5	0
Foreing languages master	0	0.5	0.5	0
Geography bachelor	0.5	0	0	0.5
Geography master	0.5	0	0	0.5
Health master	0.5	0	0.5	0
HR bachelor	0	0.5	0.5	0
HR master	0	0.5	0.5	0
History bachelor	0	0.5	0.5	0
History master	0	0.5	0.5	0
Industrial organization bachelor	0.5	0	0	0.5
Insurance bachelor	0.5	0	0	0.5
International bachelor	0	0.5	0.5	0
IT bachelor	0.5	0	0	0.5
IT master	0.5	0	0	0.5

Topic	$\alpha_{j,Quant}$	$\alpha_{j,NonQuantit}$	$\alpha_{j,Societal}$	$\alpha_{j,NonSocietal}$
Languages bachelor	0	0.5	0.5	0
Languages master	0	0.5	0.5	0
Law bachelor	0	0.5	0.5	0
Logistics bachelor	0.5	0	0	0.5
Logistics master	0.5	0	0	0.5
Management bachelor	0.5	0	0.5	0
Management bachelor	0.5	0	0.5	0
Management master	0.5	0	0.5	0
Marketing master	0	0.5	0.5	0
Maths master	0.5	0	0	0.5
MBA	0	0.5	0.5	0
Paramedical master	0	0.5	0.5	0
Physics bachelor	0.5	0	0	0.5
Psychology bachelor	0	0.5	0.5	0
Psychology master	0	0.5	0.5	0
Public law master	0	0.5	0.5	0
Public management master	0	0.5	0.5	0
Social Science master	0	0.5	0.5	0
Sociology bachelor	0	0.5	0.5	0
Sociology master	0	0.5	0.5	0
Sport sciences bachelor	0	0.5	0.5	0
Sport sciences master	0	0.5	0.5	0
Statistics bachelor	0.5	0	0	0.5
Trade bachelor	0.5	0	0.5	0

Notes: $\alpha_{j,m}$ is the participation of topic j to nest m. $\sum_{m=1}^{M} \alpha_s = 1$ where M is the number of nests. In the NL estimation, assignment of a destination to a nest is exclusive, which means , which means $\alpha_{j,m}$ equal to 0 or 1, and $M = 2$. In the NL, $\alpha_{j,m} = 1$ means that topic j is assigned to that nest m.

Appendix B Endogeneity and the control function approach

Several methods can be used for the treatment of endogeneity is discrete choice models. See Guevara and Ben-Akiva (2010) for a review. The control function (CF) approach is one of the main approach to estimate discrete choice models in which a variable of interest is endogenous. Control function approaches are typically used in non-linear models, as reviewed by Wooldridge (2015). They can be seen as the counterpart of the instrumental variable approach for non-linear models. Control function estimation and instrumental variable estimation converge in linear models. In order to estimate equation based on (1) and (**??**), we first need to instrument the interaction terms involving the endogenous variable, namely the interest variable I_n. We proceed to several CF estimations depending on which variable(s) is considered endogenous. For the sake of illustration, if we consider $I_n \times \log[\Pr(e_j^*)]$ as endogenous and if we use contiguity as the instrument of $I_(n)$, we proceed to two successive estimations. First, we estimate the following equation:

$$\begin{aligned} I_n \times \log[\Pr(e_j^*)] = \gamma_1 \times \log[\Pr(e_j^*)] + \gamma_2 \times \log(w_j) + \gamma_3(I_n \times \log(w_j^*)) \\ + \gamma_4(contig_n \times \log[\Pr(e_j^*)]) + \delta_{sj} + \nu_{jn} \end{aligned} \quad (14)$$

In this regression, the estimate of γ_4 is informative of the strength of the instrument used in the CF estimation. In particular, failure to reject $H_0 : \gamma_4 = 0$ reflects a weak instrument problem.
Then we estimate a second equation based on the logit model:

$$P_{kn} = \frac{e^{V_{kn} + \lambda \hat{\nu_{jn}}}}{\sum_{j=1}^{J} e^{V_{kn} + \lambda \hat{\nu_{jn}}}}. \quad (15)$$

where $\hat{\nu_{jn}}$ is the residual of the first stage equation 14. One appealing feature of the CF approach is that the estimate of λ in equation 15 is informative of the strength of the endogeneity problem as well as the direction of the bias associated to this endogeneity issue. In fact, testing for the hypothesis $H_0 : \lambda = 0$ provides a counterpart of the Hansen endogeneity test for non linear models (see once again Wooldridge (2015) on this).
Table 9 provides the first stage estimations (equation (14)) for the 6 different cases that we consider. These are respectively:

- (1)$I_n \times \log[\Pr(e_j^*)]$ instrumented by ($contig_n \times \log[\Pr(e_j^*])$,

- (2) $I_n \times \log[\Pr(e_j^*)$ instrumented by $(dist_n \times \log[\Pr(e_j^*])$,
- (3) $(I_n \times \log(w_j^*))$ instrumented by $(contig_n \times \log(w_j^*)$;
- (4) $(I_n \times \log(w_j^*))$ instrumented by $(contig_n \times \log(w_j^*))$;
- (5) both $I_n \times \log[\Pr(e_j^*)]$ and $(I_n \times \log(w_j^*))$ instrumented by $(contig_n \times \log(w_j^*)$ and $(dist_n \times \log[\Pr(e_j^*])]$;
- (6) both $I_n \times \log[\Pr(e_j^*)]$ and $(I_n \times \log(w_j^*))$ instrumented by $(dist_n \times \log(w_j^*)$ and $(contig_n \times \log[\Pr(e_j^*])$.

Table 9 provides the estimation results of the first stage regressions used in the CF approach.

Table 9: First stage results

	Dependent var:			
	Int.*Emp Lux		Int.*Wage France	
	(1)	(2)	(3)	(4)
Contig.*Emp Lux	0.116***	–	–	–
	(0.002)			
Dist.*Emp Lux	–	-0.016***	–	–
		(0.001)		
Emp France	0.096***	0.191***	–	–
	(0.008)	(0.008)		
Contig.*Wage Lux	–	–	0.107***	–
			(0.002)	
Dist.*Wage Lux	–	–	–	-0.031***
				(0.001)
Wage France	–	–	0.149***	0.352***
			(0.006)	(0.006)
Master	0.021***	0.025***	0.239***	0.312***
	(0.001)	(0.001)	(0.022)	(0.022)
Arts	0.070***	0.136***	-0.014	-0.033
	(0.003)	(0.003)	(0.067)	(0.068)
Law, Econ and Man.	0.067***	0.130***	0.008	0.014
	(0.003)	(0.003)	(0.062)	(0.062)
Human and Soc Sc.	0.067***	0.130***	0.012	0.026
	(0.003)	(0.003)	(0.065)	(0.065)
Sciences	0.062***	0.120***	0.009	0.019
	(0.003)	(0.004)	(0.062)	(0.062)
Observations	176,204	176,204	176,204	176,204
R^2	0.215	0.202	0.218	0.216
Adjusted R^2	0.215	0.202	0.218	0.216

Notes: OLS estimation. The dependent variable is the endogenous variable in our main model. *Contig.* is a dummy variable indicating whether the original region of the student shares a border with Luxembourg. *Dist.* is the log distance from parents' residence to Luxembourg. Master dummy captures topics leading to a master degree (reference level: bachelor). Arts, LEM, HSS and Sciences dummies capture topics belonging to faculties (reference level : faculty of physical education). *Int* is a dummy identifying students with a very strong or strong interest for Luxembourg at time of enrolment (reference level: weak or no interest). Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01

Appendix C Data

Table 10: Original and Consolidated degrees at the University of Lorraine

Original degree	Level	Consolidated degree
Accounting - Control - Audit	Master	Management master
Agri-Food Industries: Management, Production, and Valorization	Bachelor	Industrial organization bachelor
Agronomy	Bachelor	Agronomy bachelor
Agrosciences, Environment, Territories, Landscape, Forest	Master	Agronomy master
Animal Productions	Bachelor	Agronomy bachelor
Animation, Management, and Organization of Physical and Sports Activities	Bachelor	Sport sciences bachelor
Applied Economics	Master	Applied economics master
Applied Foreign Languages	Master	Foreing languages master
Applied Foreign Languages	Bachelor	Foreing languages bachelor
Art History and Archaeology	Bachelor	History bachelor
Arts	Master	Arts master
Audiovisual, Interactive Digital Media, Games	Master	Communication master
Automated Systems, Networks, and Industrial Computing	Bachelor	IT bachelor
Bioindustries and Biotechnologies	Bachelor	Biology bachelor
Biological Engineering	Bachelor	Biology bachelor
Business and Market Economics	Master	Applied economics master
Business and Public Administration Management	Bachelor	Management bachelor
Business Law	Master	Business law master
Cartography, Topography, and Geographic Information Systems	Bachelor	Geography bachelor
Chemical Engineering - Process Engineering	Bachelor	Chemistry bachelor
Chemistry	Bachelor	Chemistry bachelor
Chemistry	Master	Chemistry master
Cinema and Audiovisual	Master	Arts master
Civil Engineering	Bachelor	Engineering bachelor
Civil Engineering	Master	Engineering master
Civil Engineering and Construction Professions	Bachelor	Civil engineering bachelor
Civil Law	Master	Civil law master
Civilizations, Cultures, and Societies	Master	Arts master

Original degree	Level	Consolidated degree
Clinical Psychology, Psychopathology, and Health Psychology	Master	Psychology master
Cognitive Sciences	Master	Education master
Communication and Valorization of Artistic Creation	Bachelor	Arts bachelor
Communication Professions: Advertising	Bachelor	Communication bachelor
Communication Professions: Communication Officer	Bachelor	Communication bachelor
Complex Systems Engineering	Master	Engineering master
Computer Methods Applied to Business Management (MIAGE)	Master	IT master
Computer Science	Bachelor	IT bachelor
Computer Science	Master	IT master
Construction and Building Trades Professions	Bachelor	Civil engineering bachelor
Criminal Law and Criminology	Master	Civil law master
Cultural Studies	Bachelor	Arts bachelor
Cultural Studies	Master	Arts master
Decision Support and Statistics Professions	Bachelor	Statistics bachelor
Design	Master	Engineering master
Design and Control of Processes	Bachelor	Industrial organization bachelor
Digital Professions: Web Design, Writing, and Realization	Bachelor	IT bachelor
Earth and Planetary Sciences, Environment	Master	Geography master
Earth Sciences	Bachelor	Geography bachelor
E-Commerce and Digital Marketing	Bachelor	Trade bachelor
Economics	Bachelor	Economics bachelor
Education Sciences	Master	Education master
Education Sciences	Bachelor	Psychology bachelor
Electrical and Energy Professions	Bachelor	Industrial organization bachelor
Electrical Engineering and Industrial Computing	Bachelor	Engineering bachelor
Electronics, Electrical Energy, Automation	Master	Engineering master
Energetics, Environmental, and Engineering Professions	Bachelor	Industrial organization bachelor
Energy	Master	Engineering master
Energy and Environmental Performance of Buildings Professions	Bachelor	Civil engineering bachelor

Original degree	Level	Consolidated degree
Energy Control, Electricity, Sustainable Development	Bachelor	Industrial organization bachelor
Engineering Sciences	Bachelor	Industrial organization bachelor
Entrepreneurship and Project Management	Master	MBA
Entrepreneurship Professions	Bachelor	Management bachelor
Environmental Management	Master	Geography master
Ergonomics	Master	Paramedical master
European and International Studies	Master	Social Science master
Fashion Professions	Bachelor	Fashion bachelor
Finance	Master	Finance master
Foreign and Regional Languages, Literatures, and Civilizations	Bachelor	Languages bachelor
French as a Foreign Language	Master	Languages master
Geography, Planning, Environment, and Development	Master	Geography master
Health	Master	Health master
Health Engineering	Master	Engineering master
Health Professions: Technologies	Bachelor	Industrial organization bachelor
History	Bachelor	History bachelor
History, Civilizations, Heritage	Master	History master
HR Professions: Assistant	Bachelor	HR bachelor
HR Professions: Training, Skills, and Employment	Bachelor	HR bachelor
Human Resource Management	Master	HR master
Humanities	Bachelor	History bachelor
Industrial and Technological Risks Management	Bachelor	Industrial organization bachelor
Industry Professions: Design and Improvement of Processes and Procedures	Bachelor	Industrial organization bachelor
Industry Professions: Design and Process of Material Forming	Bachelor	Industrial organization bachelor
Industry Professions: Industrial Logistics	Bachelor	Logistics bachelor
Industry Professions: Industrial Product Design	Bachelor	Industrial organization bachelor
Industry Professions: Industrial Production Management	Bachelor	Industrial organization bachelor

Original degree	Level	Consolidated degree
Industry Professions: Mechatronics, Robotics	Bachelor	Industrial organization bachelor
Industry Professions: Metallurgy, Material Forming, and Welding	Bachelor	Industrial organization bachelor
Information and Communication	Bachelor	Communication bachelor
Information and Communication	Master	Communication master
Innovation Management	Master	MBA
Instrumentation, Measurement, and Quality Control Professions	Bachelor	Physics bachelor
Insurance, Banking, Finance: Customer Relations Manager	Bachelor	Insurance bachelor
Insurance, Banking, Finance: Operational Supports	Bachelor	Insurance bachelor
Integrated Franco-German Master's in Management	Master	Management master
International Cooperation and Development	Bachelor	International bachelor
International Logistics and Transportation	Bachelor	Logistics bachelor
International Trade Professions	Bachelor	International bachelor
IT Professions: Design, Development, and Testing of Software	Bachelor	IT bachelor
IT Professions: Systems and Network Administration and Security	Bachelor	IT bachelor
IT Professions: Web Applications	Bachelor	IT bachelor
Journalism	Master	Communication master
Landscape Design: Conceptualization, Management, Maintenance	Bachelor	Agronomy bachelor
Languages and Societies	Master	Languages master
Law	Bachelor	Law bachelor
Legal Activities: Labor Law Professions	Bachelor	Law bachelor
Legal Activities: Real Estate Law Professions	Bachelor	Law bachelor
Life Sciences	Bachelor	Biology bachelor
Linguistics	Master	Education master
Linguistics	Bachelor	Psychology bachelor
Literature	Bachelor	Languages bachelor
Literature	Master	Languages master
Living Sciences	Master	Biology master
Logistics and Flow Management	Bachelor	Logistics bachelor
Logistics and Transport Management	Bachelor	Logistics bachelor

Original degree	Level	Consolidated degree
Maintenance and Technology: Industrial Control	Bachelor	Industrial organization bachelor
Maintenance of Industrial Systems, Production, and Energy	Bachelor	Industrial organization bachelor
Management	Bachelor	Management bachelor
Management and Accounting Professions: Accounting and Financial Management	Bachelor	Management bachelor
Management and Accounting Professions: Accounting and Payroll	Bachelor	Management bachelor
Management and Administration of Businesses	Master	Management master
Management and Development of Organizations, Sports Services, and Law	Bachelor	Sport sciences bachelor
Management and Organization Management	Bachelor	Management bachelor
Management Control and Organizational Audit	Master	Management master
Management of Business Activities	Bachelor	Trade bachelor
Management of Projects and Artistic and Cultural Structures	Bachelor	Arts bachelor
Marketing of Food Products	Bachelor	Trade bachelor
Marketing of Products and Services	Bachelor	Trade bachelor
Marketing Techniques	Bachelor	Trade bachelor
Marketing, Sales	Master	Marketing master
Mathematics and Applications	Master	Maths master
Mechanical Engineering and Production	Bachelor	Engineering bachelor
Mechanics	Master	Engineering master
Microbiology	Master	Agronomy master
Multimedia and Internet Professions	Bachelor	IT bachelor
Musicology	Bachelor	Arts bachelor
Natural Language Processing	Master	Engineering master
Network and Telecommunications Professions	Bachelor	IT bachelor
Networks and Telecommunications	Bachelor	IT bachelor
Notarial Law	Master	Business law master
Nutrition and Food Sciences	Master	Agronomy master
Operational Marketing Professions	Bachelor	Trade bachelor
Performing Arts	Bachelor	Arts bachelor
Philosophy	Master	Languages master
Physical Measurements	Bachelor	Physics bachelor
Physics	Master	Engineering master
Plastic Arts	Bachelor	Arts bachelor
Political Science	Master	Social Science master

Original degree	Level	Consolidated degree
Primary Education Teaching	Master	Education master
Process and Bio-Process Engineering	Master	Engineering master
Process Engineering for the Environment	Bachelor	Industrial organization bachelor
Production Management, Logistics, Purchasing	Master	Logistics master
Professional Optics	Bachelor	Physics bachelor
Psychology	Bachelor	Psychology bachelor
Public Health	Master	Health master
Public Law	Master	Public law master
Public Management	Master	Public management master
Public Works Professions	Bachelor	Civil engineering bachelor
Quality, Hygiene, Safety, Health, Environment	Bachelor	Biology bachelor
Quality, Industrial Logistics, and Organization	Bachelor	Logistics bachelor
Real Estate Professions: Management and Development of Real Estate Heritage	Bachelor	Management bachelor
Refrigeration and Air Conditioning Installations	Bachelor	Industrial organization bachelor
Science and Engineering of Materials	Master	Engineering master
Science and Engineering of Materials	Bachelor	Physics bachelor
Sciences and Techniques of Physical and Sports Activities	Bachelor	Sport sciences bachelor
Sciences for Health	Bachelor	Biology bachelor
Sectorial Management	Master	Management master
Economic and social administration	Bachelor	Management bachelor
Economic and social administration	Master	Public management master
Social Law	Master	Public law master
Social Sciences	Master	Social Science master
Social, Work, and Organizational Psychology	Master	Psychology master
Sociology	Bachelor	Sociology bachelor
Sociology	Master	Sociology master
Sound and Image Techniques	Bachelor	Industrial organization bachelor
STAPS: Adapted Physical Activity and Health	Master	Sport sciences master
Tax Law	Master	Business law master
Teaching, Education, and Training Professions, 2nd Degree	Master	Education master
Teaching, Education, and Training Professions, Practical	Master	Education master

Original degree	Level	Consolidated degree
Teaching, Education, and Training Professions, Supervision	Master	Education master
Technical Sales	Bachelor	Trade bachelor
Territorial Planning and Urban Planning Professions	Bachelor	Geography bachelor
Tourism and Leisure Professions	Bachelor	International bachelor
Trade and Distribution	Bachelor	Trade bachelor
Urban Planning and Development	Master	Geography master
Wood and Furniture	Bachelor	Agronomy bachelor

Notes: Our criterion of consolidation is based on the share of common topics of each original degree. This means sharing a common major and potentially differentiated only by their specialization (minor).