Functional decentralization: the role of education infrastructures in dynamizing peripheral municipalities

Author: Lucas Lisbonis
Msc. Economics, University of Barcelona
E-mail: llisboli9@alumnes.ub.edu

Supervisor: Rosina Moreno
Department of Econometrics, Statistics and Applied Economics, University of Barcelona.
E-mail: rmoreno@ub.edu

Abstract
Many European countries have focused on redistributing wealth across their territory through the decentralization of public services. This so-called functional decentralization is often carried out through investments in various education infrastructures in order to attract migrants into peripheral regions. However, do education infrastructures actually influence gross migratory patterns? It is not clear whether universities or specialized schools only play a role in a student's short-term migration or if they also have an effect on the residential migration of households. Nonetheless, it is a key determinant of the potential success of a functional decentralization policy. This study will aim at defending the hypothesis that universities and specialized schools do have an important and positive impact on the gross migration inflows into a particular area. Studying such flows in the French province of Bouches-Du-Rhône, some evidence was found that specialized schools in agriculture have a decisive role in attracting migrants. This implies that the current policies focusing on the decentralization of the university system may not be as effective as the one focusing on specialized education infrastructures.

Keywords: Functional decentralization, education, migration flows, gravitational model

1. Introduction

Decentralization is a major challenge that Europe has been facing ever since its creation. However, each member of the union deals with decentralization in a different manner: some countries like the Netherlands have very few policies to promote decentralization, others like Germany have clear objectives to achieve it. Different countries have opted for different positions on the debate between centralization and decentralization. De Vries (2003) explains that supporters of centralization endorse a Leviathan state à la Hobbes as it can provide with a more efficient and stable control of the territory. He also explains that, on the opposite side of the debate, supporters of decentralization argue that local issues require local solutions and that a centralized power will never be able to fully understand the needs of its citizen at the local level.

France is an example of a European country that has been supporting decentralization for a long time. Since 1982, the decentralization policy has been initiated on 3 different grounds (Ohnet, 1996): institutional, territorial and functional. The *institutional* decentralization aims at avoiding the aggregation of powers in one city and so major banks and executive bodies were transferred to different cities across France. The *territorial* decentralization aims at allocating resources in a more equal way across different regions. Finally, and most interestingly for this research, the *functional* decentralization aims at spreading budgets and infrastructures: for example, through the creation of autonomous universities in peripheral cities. The establishment of universities in peripheral cities is thought to drive dynamism in these remote cities. The aim of this research is to show to what extent such infrastructures drive gross residential migrations and, consequently, the decentralization of population.

So far, migration studies have primarily focused on the workers' mobility, income differences, natural amenities or differences in quality of life to understand gross migratory patterns. A recent trend is for example the study of migratory patterns for highly skilled workers (Bahar et al., 2019, Caviggioli et al., 2020, Miguélez and Moreno, 2014). A diverging trend from the traditional application of gravitational models in migration studies is the study of higher education as a driver of students' migration. While previous studies focused on analysing how different characteristics of education infrastructures influenced the place of enrolment of a student (Bruno and Improta, 2006, Cullinan and Duggan, 2016, Faggian and Franklin, 2014, Ritsilä and Ovaskainen, 2001, Sa et al., 2004), the purpose of this study is to understand the general effect of education infrastructures on gross migration flows. The reasoning behind this choice is to provide evidence that education plays an important role in the migratory patterns of, not only students, but also households. In the previous literature of migration studies, the focus has been mainly on income, housing and amenities as primary drivers of the gross migration flow. Education infrastructure has been studied solely in a framework of student migration and therefore, no evidence has been provided on the role of education as a driver of gross residential migration. This study will aim at showing that education infrastructures are an important tool to achieve the goal revitalization of peripheral municipalities through the migration of households.

The idea that education infrastructure drives, not only students' migration, but also total residential migration, comes from a theory developed by Jacob Mincer (1978): migration is not only an individual decision; most of the times, the decision is taken at the level of the family. For example, a father might be moving to a new municipality where there is a high school so that his daughter can study, although this destination may not provide him with benefits in terms of income or amenities. This is the notion of "tied mover". Similarly, the "tied stayer" is the family member that will remain in the commune, even though it may be subject to a lot of push factors. As such, Jacob Mincer makes the distinction between "public" (familial) decisions and "private" (individual) decisions. This leads to the idea that people move as a family and not as individuals and therefore, the utility of an individual must also include the utility of other family members. As we have seen in the previous paragraph, gravity models mainly focused on the role of education as a main driver of individual decisions. Following Jacob Mincer's reasoning, education infrastructures should play an important role not only for the decision of students but also to any household susceptible to require this infrastructure.

Understanding how education infrastructures influence gross migratory patterns is key to understand the effect of education infrastructures on functional decentralization. Previous authors have already mentioned this; Florax (1992) explains how demand for education drives a considerable economic boost. He differentiates between national and regional demands and how they create dynamism around a particular university. He argues that the national demand can bring new talents who will migrate to a remote province to study. He shows how this attraction of talent can create poles of employment around the universities. He also argues that the regional demand for education can dynamize provinces where outflows of young people is important. He takes for example how the Dutch government limited the outflow of students from the remote region of Wageningen to the more attractive and dynamic Amsterdam: through ambitious policies of functional decentralization in the 1970s, the government managed to dynamize the region by investing considerably into the Wageningen university which was, before the 1970s, a rather small institution. Similarly, in France, the "law on Higher Education and Research" (2013) has granted much more freedom and budget to the development of regional universities. This law is in the context of the "Act II of decentralisation" (2003) whose focus is to bring the laws of 1982 mentioned earlier to an even higher level of decentralisation. More recently, the "NOTRe" (2015) law has focused on redistributing budgets and transferring more power to local education infrastructures. Clearly, in France, the development of universities in remote areas is seen as a key factor to attract talents into peripheral areas.

This paper will aim at answering the following question: are education infrastructures important in understanding gross migratory patterns? The main hypothesis is that universities and specialized schools are drivers of gross migration. In order to provide empirical evidence that support or reject the hypothesis, we analyse the different gross migratory patterns in the province of Bouche-Du-Rhône in Southern France. The main reason why Bouche-Du-Rhône was chosen is its wide heterogeneity of location which could bring external validity to our results: it contains Marseilles, one of the most populated cities of France, as well as a large amount of traditionally agricultural municipalities so that we can study the impact of specialized schools in these remote municipalities. Moreover, gravitational models are very demanding in terms of data processing when the number of potential flows is large. Since Bouche-Du-Rhône is a province with a small number of municipalities (compared to other provinces) this greatly limits the number of potential flows and, therefore, enables to perform statistical analyses with ease.

After this introduction, the section "literature review" will aim at explaining the reasoning behind this hypothesis. To verify this hypothesis, the methodology for this study will follow the framework of Beine, Bertoli and Fernandez-Huertas Moraga (2014) and thus, some problems may arise such as multilateral resistance to migration since the data cannot include all potential destinations. This problem will be thoroughly described in the "methodology" section. Another challenge of this study is to control for many traditional drivers of gross migration flows. Omitting for example, income or housing or natural amenities could lead to a problem of omitted variable bias. In the literature of gravity models with gross migration flows, three important notions arise: expulsing ability, accessibility and attractiveness (Beine, Bertoli and Fernández-Huertas Moraga, 2014) as we will see in the "data" section. Finally, the different results will be presented to support evidence that education infrastructures are indeed pull factors of gross migration flows, and then we will conclude.

2. Literature review

As discussed earlier, when studying the gravity models of migration around education infrastructure, the previous literature focused on the distinct flows of students. Leppel (1993) has studied the enrolment decisions at Widener University in Pennsylvania. Through the use of a gravity model, she showed that the distance to the university was the most relevant factor in the enrolment decision of a student. The distance between universities and the place of residency becomes an important part of the utility of the migrant. She has set the grounds for numerous studies on accessibility and choice of enrolment (Alm and Winters, 2009, Kjellström and Regnér, 2006, Spiess and Wrohlich, 2010).

It is clear that students tend to study close to where they live. However, the interest for this study is to know whether students simply choose to study close to where they already live or if minimizing the distance to the university was an important objective of migration before the start of their studies. To answer this question, researchers have looked at the students' spatial mobility and have required the use of surveys to understand their intention of migrating prior to the decision of enrolment. Venhorst et al. (2010), studying spatial mobility of the Dutch students between the 4 NUTS I regions of the Netherlands, found out that college graduates are indeed less mobile and not willing to migrate regardless of ability. They argue that, indeed, the important familial ties tend to restrict the mobility of younger students. Kodrzycki (2001), through a thorough survey analysis, confirms these results and adds that the majority of young students study close to where they have been living with their family during high school. Some universities may be more effective than others at attracting students, but the decision to migrate was indeed made prior to the start of the studies. This result is very important for our study because it indicates that students may not migrate alone. As a summary of the previous two paragraphs:

- Students are more likely to enrol in a university close to where they were living at the moment of enrolment.
- Reducing the distance between the residency and the future place of study is a primary objective for migration.
- Students migrate prior to the start of the studies.

Since families, before the time of enrolment, aim at minimizing the distance between the residency and the place of study of the children, then the education infrastructures are a driver of the migration of families. These three above-mentioned facts lead to the hypothesis that not only students but also households are attracted to universities. This hypothesis can be linked to Jacob Mincer's theory of the "tied mover" mentioned earlier: households should also, when choosing where to migrate, minimize distances to universities. It is indeed observed that families choose the best location for their children to study. If both the results of the previous literature and Jacob Mincer's theory hold, it should be observed that education infrastructures have a high impact on gross migration flows. As a consequence of the previous argument, we can set the following hypothesis:

Hypothesis 1: Universities are important pull factors of gross immigration flows.

This hypothesis has already been contested by past research. For example, Sa et al. (2004) used a gravity model to study the students' enrolment decision in the Netherlands between 40 NUTS III regions and 13 universities. Although they acknowledge that distance is an important factor at the national level, they found that distance did not have a major role in the regional demand for universities which would contest our hypothesis that households try to minimize distances to universities when migrating. Therefore, in the Netherlands, they believe that policies should not focus on the geographical decentralization of the university system. Nevertheless, they seem to highlight the relevance of distance when accounting for specialized curricula: specialized universities (such as the Deflt technical university), and especially in remote areas (such as the agricultural university of Wagenignen) generate a regional demand highly elastic to distance. This finding is supported by Cullinan and Duggan (2016) when studying Higher Education Institutes (HEI) in Ireland. Acknowledging the importance of distance, they also confirm that students may be willing to live in a remote region to attend a specific course. Specifically, in remote areas where population growth is low, they suggest that the presence of specialized schools, for example in agriculture, would help dynamize migration in these remote areas. Once again linking this result to the theory of the "tied mover", the presence of an agricultural school should matter for the decision of families to move to a certain area. It should be observed that agricultural schools have a positive impact on gross immigration flows in remote areas. Thus, our second hypothesis stands as follows:

Hypothesis 2: Specialized schools are important pull factors of gross immigration flows

These two hypotheses will be assessed through a gravity model with both universities and agricultural schools as key explanatory variables. Each variable may help understand the current debate whether distance matters and whether education infrastructure are relevant in a context of functional decentralisation. All in all, this would provide a contribution to the literature on student migration and functional decentralization.

3. Methodology

3.1. The empirical framework

In order to understand the issue, one can study the flows through a gravitational model using the framework proposed by Beine, Bertoli and Fernández-Huertas Moraga (2014). The utility of the migrant is defined as:

$$U_{ijkt} = w_{jkt} - c_{jkt} + \epsilon_{ijkt} \quad (1)$$

with w_{jkt} being the benefits of moving from the origin j to the destination k at time t, and c_{jkt} the costs of moving from origin j to destination k at time t.

We can write m_{jkt}, the migration flow as:

$$m_{jkt} = p_{jkt} s_{jt} \quad (2)$$

with p_{jkt} being the share of individuals moving from origin j to destination k at time t, and s_{jt} the population at origin j at time t.

Assuming that $\in_{ijkt}$ follows an independent and identically distributed Extreme Value Type-1 distribution, the expected migration flow derived from the expected probability to migrate from j to k for the individual i at time t+1 is as follows (McFadden, 1974):

$$E(m_{jkt}) = \frac{e^{w_{jkt} - c_{jkt}}}{\sum_{l \epsilon D} e^{w_{jlt} - c_{jlt}}} s_{jt} \quad (3)$$

with l being any destination from the set D of all possible destinations.
s_{jt} the population at origin j defined as the expulsing ability,
$e^{w_{jkt}}$ the attractiveness of the destination k from origin j,
$e^{-c_{jkt}}$ the accessibility,
$\sum_{l \epsilon D} e^{w_{jlt} - c_{jlt}}$ the expected utility of potential migrants from the origin.

3.2. Methods of Estimation

The large presence of zero-flows puts into question the nature of the distribution of the residuals. Assuming normality and using OLS would lead to biased results. Assuming the extreme value distribution type 1 is common for gravity models (as we saw above in our framework) and the traditional estimator in the presence of a large number of 0 in the dataset is the Poisson estimator.

However, the assumption of equi-dispersion, the variance being equal to the mean, can cause problems with over-dispersed data (Cameron and Trivedi, 2001). The *negative binomial estimator* is a more suited estimator as it accounts for the heterogeneity inherent to gravitational models of migration.

Alternatively, the *zero-inflated estimator* may be better in case the data presents an abnormally large number of 0 flows because this model differentiates between two categories of 0 flows:

- municipalities that have never been and probably will not ever generate migration flows (we can think of two villages, one all the way to the West of the province, and another one all the way to the East).
- municipalities that have not been but that may start generating flows (we can think of two villages relatively close to each other but for some reason are not generating flows at a particular time).

However, there could be problems of heteroskedasticity as the above-mentioned estimators use the maximum likelihood estimation. Santos-Silva and Tenreyro (2010) argue that heteroskedasticity may cause the non-existence of the maximum likelihood estimates. Thus, in case of a large heteroskedasticity (which is often the case in migration studies) and of non-existence of the maximum likelihood estimates, the *Pseudo Poisson Maximum Likelihood (PPML)* technique can drop some of the regressors to solve this problem. More recently, an alternative estimator has been developed by Correia, Guimarães and Zylkin (2019): the PPML with multiple fixed effects and multiway clustering. This estimator can be very useful as it is robust to the convergence issues mentioned above.

It is important to notice that none of these 4 estimators mentioned above are absolutely superior to one another as argued by Brümmer, Kareem and Martinez-Zarzoso (2016): each estimator has its strengths and weaknesses.

3.3. Problem of multilateral resistance

This problem is common to the literature of gravitational models. Indeed, collecting data on all potential alternative destinations is impossible. For individual data, the problem is less important since all alternative destinations have zero values (because the individual moves to only one destination). With aggregated data, the problem is more important since some of the unobserved alternative flows have non-zero values. Irrespective of whether the flows are on an international level or at the municipal level, there will be non-zero flows to other destinations that are not observed in the data collected (Beine, Bertoli and Moraga, 2014).

In 2012, 36670 French towns represent 36,670*36,669 = 1,295,964,000 potential flows of differentiated direction. These flows represent 4,826,751 migrants; out of a total French population of 64,754,017 people (over 1 year old). Out of these 1,295,964,000 potential flows, only 514,933 flows are non-zero flows. So, 99.96% of the potential flows are zero flows. The database omits the outflows abroad (code 99), which represent around 2% of the total outflows.

In order to limit the scope of the study we will only consider intra-provincial flows since each province has on average around 150 communes, the potential flows being much smaller than at the national scale. The problem with this approach is that many non-zero outflows are omitted. On the national scale, the original database omits only 2% due to international outflows; on the provincial scale, the database for the study will omit all inter-provincial outflows which represent 33% of the outflows of a province on average. This number was approximated using an estimation of the aggregated outflows par commune and the sum of all intra-provincial outflows per commune of the original database.

In order to limit the scope of the study, only intra-provincial flows in the province of Bouches-Du-Rhône will be considered (see "data" section below). Its total outflows are represented at 65% by its intra-provincial flows (similar to the national average) and it has important and various cities such as Marseilles (one of the three biggest agglomeration of France). The data also divides Marseilles in 16 districts to make the analysis more precise. For cities generating a lot of intra-provincial outflows, this is not so much of a problem. But for big cities or cities with large inter-provincial flows, there is much less representation: for Marseilles, intra-provincial flows represent only 55% of the total outflows from Marseilles. As a big city, many of the migrants from Marseilles migrate to another big city in France. There is also a lot of omitted relevant outflows for cities close to the border like Arles, on the eastern bound of the province where the intra-provincial outflows represent only 46% of the total outflows. Therefore, there is a clear problem of omission of relevant flows.

3.4. Controlling for unobserved heterogeneity

The omission of relevant outflows can lead to biased results. This is known as the multilateral resistance to migration. This problem was first mentioned in trade theory: when looking at trade flows, the estimator should take into account the accessibility and attractiveness of a region compared to all potential alternatives. By accounting for these factors, the flows between the observed regions should increase. Bertoli et al. (2011) applied this concept to gravitational models and raised awareness to this problem. Royuela and Ordoñez-Cuenca (2016) have identified 4 different approaches to account for the multilateral resistance to migration:

- the traditional approach to solve this issue is to include dummy fixed effects for origin and destination to understand migration policies that are common to both origin and destination municipalities (Mayda, 2010).
- Another approach is to account for pairs of origin and destination. By including fixed effects per flow ID, one can account for historical migratory patterns, geographical characteristics, transportation… (Ortega and Peri, 2013).
- Beine and Parsons (2012) argue that dummies for pairs of destination and time together with dummies for origin, can control for the heterogeneity at the destination: this enables to better estimate push factors of migration.
- Ortega and Peri (2013) also argue that dummies for pairs of origin and time together with dummies for destination can control for the heterogeneity at the origin: this enables to better estimate pull factors of migration.

The last method from the list above is particularly interesting for this study. By controlling for the push factors and heterogeneity at the origin, the resulting estimation of the pull factors may be unbiased by the problem of multilateral resistance. Therefore, 2 models can be drawn: model 1 will be based on the traditional approach and model 2 will be based on the method of Ortega and Peri (2013) in order to estimate the attractiveness of education infrastructures at the destination.

Model 1: Panel model with origin, destination and time fixed effects:
$$ln\, m_{jkt} = D_j + D_k + D_t + \beta_1 X_{jkt} + e_{jkt}$$

Model 2: Panel model with origin-time and destination fixed effects:
$$ln\, m_{jkt} = \alpha + D_{jt} + D_k + \beta_1 X_{kt} + e_{jkt}$$

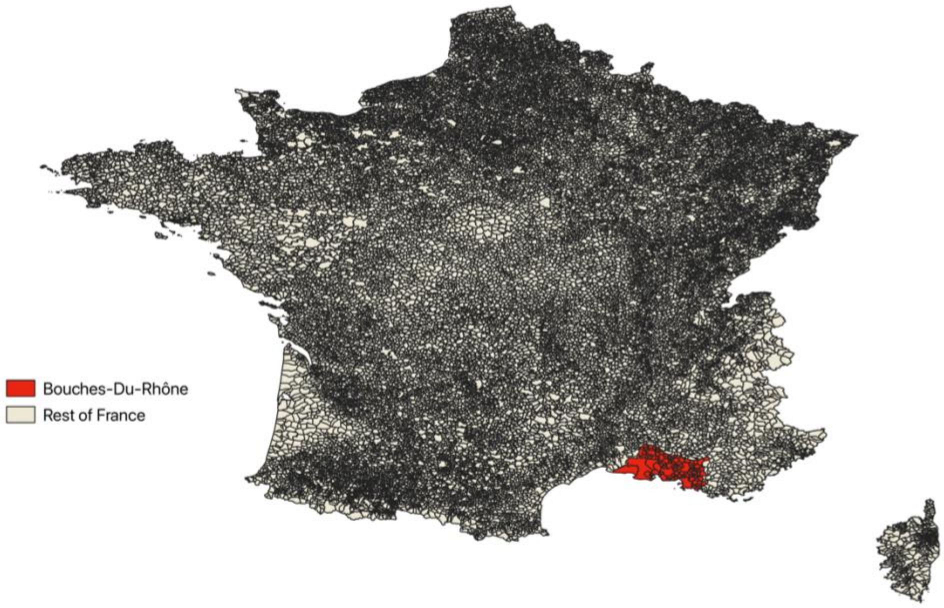
Several problems of endogeneity may appear in gravitational models. A common source of endogeneity is the problem of simultaneity: to avoid that a shock may affect both the flows and our explanatory variables at the same time, we will regress the flows between time t and t+1 on the explanatory variables at time t. That way, we use the census data at year t for the independent variables and at year t+1 for the flows. In the data section, we will see how the data collection method for the residential flows accounts for this problem of simultaneity.

4. Data

4.1. Database

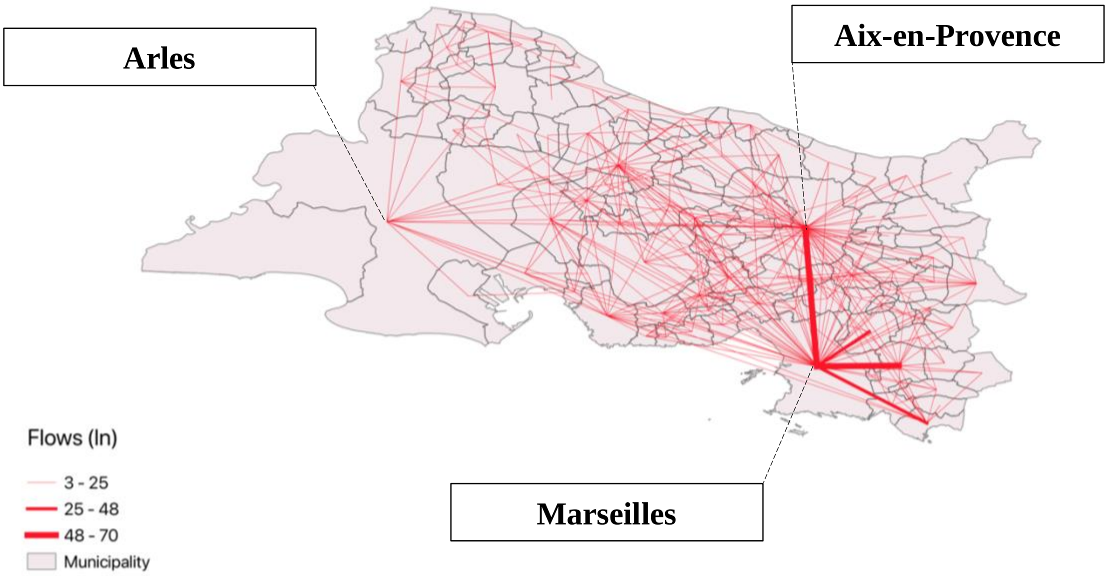
The province of interest is located in the South-Eastern part of France in a region called Provence-Alpes-Côte d'Azur (PACA) (since 2017, the region changed its name to Région Sud). In 2012, 5,000,000 people were living in the PACA region and 2,000,000 in the province Bouches-Du-Rhône (see Figure 1). The province of interest contains the capital of the whole region: Marseilles. It is also the third largest metropolitan area in France with a city containing 870,000 inhabitants in 2012. It is divided into 16 "arrondissements" which will be accounted for in the study. They will be referred to as districts and they are numbered from 1 to 16. The 1st district is located in the ancient centre of the city, on the old harbour. The counting increases as the distance from this old harbour increases. Therefore, the 16th district is the furthest away from the oldest part of the town.

Figure 1: Map of French municipalities



The data on flows are obtained from the Institut National de la Statistique et des Etudes Economiques (INSEE) through the national and yearly residential census. The data are available for the 36,658 municipalities of both metropolitan France and overseas. In 2016, a reform of the territory greatly decreased the number of municipalities to 35,885 which is the reason why the time scope of our study is until the 1st of January 2016. The residential flow between two municipalities was computed for each individual: where this individual is residing and where he/she was residing one year before. Thus, the census residential flows at year t are all flows that happened between January at year t-1 and January at year t. And therefore, any child less than 1 year old is excluded from the census. The limitation of this data collection method is that it does not account for people who do not register at the municipality they live in and it does not account for people that may change municipalities more than once in a year. The data are available every year since 2013 and directly published on the website of the INSEE. It is available for the entirety of the French territory but in this study the data was collected solely for the province of Bouches-Du-Rhône. It includes the 119 municipalities of the province knowing that one of the municipalities is Marseilles which will be subdivided into 16 districts as explained earlier. Each of the maps in this study will represent Marseilles as one unique municipality as you can see in Figure 2; but in the dataset, Marseilles will indeed be divided into 16 different districts. Therefore, there is a total of 134 units of observation. This means that there are 8911 intra-provincial routes and so 17,822 flows of differentiated direction. These intra-regional flows do not represent the total number of migrants in the province since some migrants decide to migrate to Paris or other extra-provincial cities. In 2013, the total number of outflows from Bouches-du-Rhône was 423,963 migrants. The intra-provincial flows in 2013 represent 254,978 of these migrants. This means that the intra-provincial flows represent 60% of the total outflows of migrants. Therefore, 40% of these outflows are not represented in the data as they are directed to any city outside of the Bouches-du-Rhône. This study will focus solely on intra-regional flows which will lead to a problem of omitted flows and so of multilateral resistance to migration. Therefore, the 5 models described in the previous section will play a major role in limiting this bias.

Figure 2: Map of flows between the municipalities of Bouches-Du-Rhône in 2012



The data collected will spread across 4 years: 2012, 2013, 2014, 2015. So, the total number of observations will amount to 71,288. Population, average income, unemployment rates, historical housing, number of universities and of agricultural schools are collected from the INSEE municipal database. The data on natural amenities are collected through the software QGIS and Google Maps distances are obtained through Google Sheets.

Bouches-du-Rhônes is a province containing a large city, Marseilles, divided in 16 districts in our dataset, and also a fair amount of smaller towns like Arles and Aix-en-Provence, and also a fair amount of villages. This province has a great diversity of cities and landscapes so that results can be extended to other provinces. The small scope of the study is a limit to the external validity of the results but the use of municipal data is essential to understand the role of education infrastructures in migration. Figure 2 clearly shows that the majority of large flows are outflows or inflows from or to Marseilles (the capital and largest city of the province). The following sections will aim at justifying why this is the case.

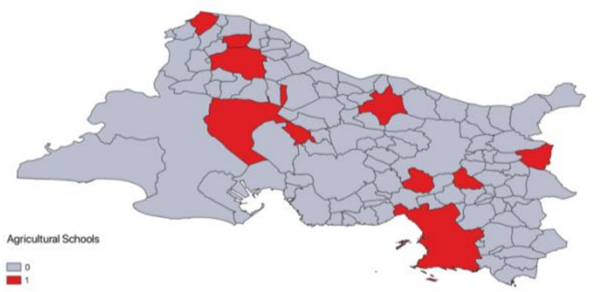
4.2. Key variables: Education infrastructures

There is often a problem of multicollinearity between different education infrastructures as pointed out by Tay (2014). This could lead to a problem of inefficiency as the variance inflation mechanism overestimates standard errors and biases the T-statistics. It is therefore important to select relevant education infrastructures. It is common that the specialized schools in agriculture mentioned above are located in different regions where agriculture is a prominent sector. Whereas general universities are rather located in cities. Thus, including both variables would lead to interesting results and the multicollinearity would not be a major issue. However, primary, secondary and high schools are strongly correlated with the presence of universities so it could lead a serious problem of multicollinearity. For the purpose of this study and to compare with results of the previous literature, universities are selected over the other types of infrastructure. As displayed in the appendix Table 2A, the correlation matrix reveals that only historical housing and population are strongly correlated. This result can be explained by the fact that most centres of activity in the region are also historical centres (ex. Arles, Marseilles, Aix-en-Provence…). However, it is still important to control for housing since the correlation is not equal to 1, there are many variations in migration flows captured by the proxy for housing.

As argued in the introduction, part of the benefits to moving to a particular location is the access to certain education infrastructure. As a proxy for the education infrastructure, one could include various infrastructures such as: nursery school, primary school, secondary school, high school and universities. Additionally, one could include specialized schools like preparatory schools or high schools for agricultural studies: this comes from previous results by Cullinan and Duggan (2016) supporting the idea that students tend to migrate towards specialized education infrastructure. Figure 3 shows the spatial distribution of universities across the province. The province has 11 universities with 6 universities in Marseilles (2 in the 1^{st} district, 1 in the 3^{rd}, 1 in the 4^{th}, 1 in the 9^{th} and 1 in the 13^{th}), 2 in Aix-en-Provence, 1 in Arles, 1 in Salon-de-Provence and 1 in La Ciotat. There is a high concentration of universities in Marseilles, however, they are spread relatively evenly within the districts of this large agglomeration. Concerning the agricultural schools, there are 10 schools with only 1 for each of the following municipalities: the 8^{th} district of Marseilles, Saint-Martin-de-Crau, Grans, Lambesc, Puyloubier, Barbentane, Eyragues, Saint-Rémy-De-Provence, Gardanne and Cabriés. Therefore, they are spread evenly across the province and generally located in rural and remote areas with little economic dynamism, as discussed later in the control variables section.

Figure 3: Universities

Figure 4: Agricultural Schools



4.3. Controls of the gravitational model

This section presents the main control variables to give some insights on their geographical distribution. Table 1 provides main statistics of the variables as well as their definition. As mentioned earlier data was collected through the INSEE database and the shapefiles for the municipalities on data.gov.fr. Figures 3 to 6 offer the QGIS maps for all these variables, with the values for year 2012 only, in order to observe their spatial distribution.

The gross migration flow depends on the number of inhabitants at the origin. The model predicts that the more inhabitants at the origin, the bigger the gross outflow out of this origin. Therefore, controlling for the number of inhabitants at the origin is essential. As observed in Figure 5, Marseilles and its surrounding municipalities are the most populated so the model would predict that outflows from Marseilles are large. As shown from the map below, the coastal areas and the South of the province in general are the most populated. We can relate it to the flow map above, where the highest flow density is also located in the South of the province. This notion is defined as expulsing ability in the literature of gravity models.

Accessibility is important in gravity models. This presumes that the more accessible the destination is to the origin, the bigger the migration flow. However, in other fields, this notion has been discussed: Blijie (2005) argues that for highly connected regions, accessibility cannot predict residential migration as people would rather commute everyday than migrate. The data in this study will use the google time travel distances in minutes by car. As shown in Table 1, the average google time distance between two municipalities is 41 minutes, with a minimum of 5 minutes (between Cadolive and Saint-Savournin, two neighbouring municipalities) and maximum of 1 hour 48 minutes (between Ceyreste, South-East bound, and Boulbon, North-West bound). The North-West to South-East trajectory is the one benefitting from fewer transportation possibilities as those are the least dynamic parts of the province as we will in the paragraphs below.

In the framework developed by Beine, Bertoli and Fernández-Huertas Moraga (2014), the household maximizes its utility weighting the benefits of moving to a particular destination over the costs of moving to that destination. Therefore, defining the utility of the household is essential. As argued in the introduction, education infrastructure plays an essential role in the attractiveness of a particular location. However, the heterogeneity of locations must be taken into account in order to fully understand the role of education infrastructure. Omitting certain relevant variables would lead to a problem of omitted variable bias. Throughout the literature on gravitational models, researchers have identified 4 essential variables to control for the heterogeneity of location: income, employment, housing and natural amenities (a summary of all variables is available in appendix Table 1A).

In migration studies, income is often referred as the main proxy for the pull factor of attractiveness of a region. To represent this, Mayda (2010) used the average salary; similarly, here we will use the average salary in the labour force per municipality. Income is very unheavenly distributed in the province, the Eastern part having higher average salaries than the Western part. This is linked to the structure of employment as the Western part of the province is predominantly of agricultural tradition whereas the Southern and Eastern parts gain much more wealth from industries and maritime trade. This large heterogeneity in salaries is an important variable to control for the heterogeneity within the province. As shown in figure 6, Marseilles as a whole, has a very low average income but, across districts, average income varies a lot: the 8th district being the richest with 16.2€/h (close to the maximum in the province of 19.1€/h in Ventabren) and the 15th district being the poorest with 10.8€/h (close to the minimum of 10.0 in Saintes-Marie-de-la-Mer). The richest municipalities in terms of average income are located on the Eastern part and the poorest on the Western part. This can be explained by the fact that the western part of the province is of agricultural tradition and suffers from high levels of unemployment as discussed in the next paragraph.

Human capital endowment is a key characteristic to understand workers' migration flow. Controlling for the sector of activities of a certain municipality is essential to understand inflows or outflows of migrants. The rural exodus is a clear example that unemployment in the agricultural sector is a push factors for emigration (Piras, 2015); some sectors of employment like agriculture can be push factors while some other sectors like services and administration can be pull factors (Royuela and Ordoñez, 2016). It is important to control for the structure of employment in the analysis. However, due to problems of multicollinearity with the education infrastructures, the study will only focus on unemployment rates per commune. Although the unemployment rate has not been a traditional control in gravity models, Beine et al. (2013) recently underlined the importance of this factor in migration studies. They especially focus on the role of unemployment in the context of the Schengen agreements and the introduction of the European monetary union: as workers' mobility increased, the short-term migratory patterns are much more reactive to short-term economic downturns. Therefore, they underlined the importance of accounting for the unemployment rate to understand short-term workers' mobility. As shown in Figure 7, the Western bound of the province suffers from low employment mainly due to the large rural exodus of the firms towards the richer regions of the East. Aix-en-Provence and Marseilles, the two largest cities, have very high levels of unemployment, almost 40%. This is an increasing problem in France as the two cities have reached abnormally high levels of unemployment compared to the national average of 10.6% in 2012. Traditionally, Marseilles has always been a relatively poor city with low-income residents living in the historical centre (1st, 2nd or 3rd districts) and high-income residents living in the outskirts (7th or 8th districts). This is an unusual distribution for a European city as poor social classes live in the historical centre (Bouzouina, 2007).

Rents and historical centres have often been a focus in migration studies. With a monocentric model, Alonso (1964) argues that minimizing rent is one of the primary objectives of the migrant. This is one of the reasons for the rising discontentment of the local population: since the 1960s, the entire region and especially the capital, Marseilles, is experiencing gentrification in its historical centres. It has caused the emigration of the disadvantaged social categories to out of historical centres (Jourdan, 2006). Thus, the number of houses built before 1919 will be used as a proxy for the cost of renting. Moreover, since Bouches-du-Rhône is a province with a lot of historical heritage, the number of houses before 1919 will also be used as a proxy for the presence of historical centres as they are considered an important driver for urban development (Koster and Rowendal, 2017). As shown in Figure 8, historical houses are vastly spread throughout the province with historical towns such as Salon-de-Provence, Baux-de-Provence, Fontvieille and of course Arles and Aix-en-Provence for which the foundation dates back from the 2nd century BC and Marseilles, previously known as Massilia which dates back from the Greek colonization of the region in the 7th century BC. These historical amenities have a considerable impact on attractiveness through prices and is believed to have a negative impact on migration. Especially since the province's lower income residents traditionally live in historical centres as discussed above and gentrification has pushed many to migrate outside of these historical centres as argued above.

In many gravitational models estimated in the previous literature, the presence of the coast line is considered. Lee and Lin (2018) argue that in a closed city, the coast line is an amenity that heavily influences income distribution. A common source of endogeneity in gravitational models is due to the diverse origins of the agglomeration forces: cities exist not only because of economic factors but also because of historical path dependence, geographical amenities… Omitting these natural features may lead to an omitted variable bias since these amenities can influence migration through our explanatory variables (Koster and Rouwendal, 2017). The coast is a traditional amenity in gravitational models (Lee and Lin, 2018). The southern bound of the province is called the Côte d'Azur and is a famous area for mass tourism. It is believed to be an attractive factor for residential migration. On the next page are four maps summarizing the spatial distribution of our main control variables as well as a table of descriptive statistics to have an overview of the data collected.

Figure 5: Population

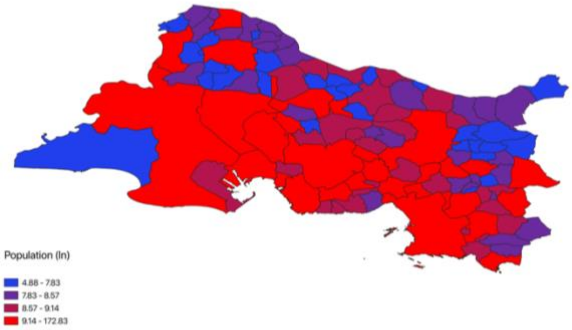


Figure 6: Income

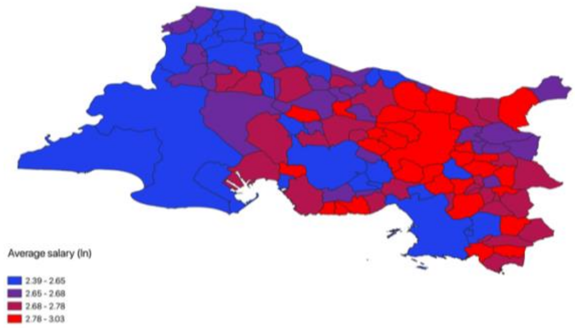


Figure 7: Unemployment

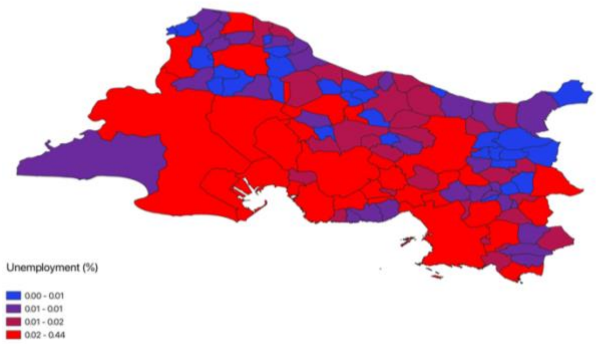


Figure 8: Historical housing

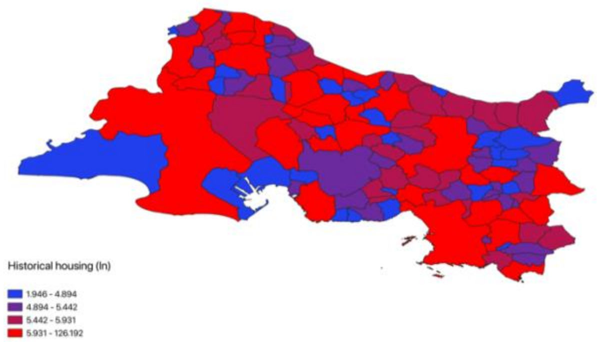


Table 1: Variable definition and descriptive statistics

	Mean	**Std. D.**	**Min**	**Max**	**Measurement**
flows	0.46	1.05	0	6.37	Log of the number of migrants between two regions in a year
pop	9.9	15.1	4.8	172.8	Log of the number of inhabitants
dist	3.73	0.47	1.5	4.7	Log of the distance in minutes by car
salary	2.7	.12	2.4	3.0	Log of the average salary
housing	6.4	11.1	1.9	126.1	Log of the number of houses built before 1919
unemp	3%	5%	0%	43%	Percentage of job seekers out of the labour force
uni	.06	.22	0	3.9	Log number of universities
a-school	0.052	0.18	0	0.69	Log number of agricultural schools

5. Results

Table 2: Pooled OLS estimation

	Poisson	Nbreg	Zinb	Ppml
dist	-1.470***	-1.761***	-1.019***	-1.470***
	(0.0103)	(0.0156)	(0.128)	(0.0165)
pop_o	0.706***	0.766***	0.421***	0.706***
	(0.0105)	(0.0138)	(0.131)	(0.0125)
salary_o	0.310***	0.233***	0.249***	0.310***
	(0.0499)	(0.0664)	(0.054)	(0.0632)
housing_o	-0.0597***	-0.0325***	-0.049***	-0.0597***
	(0.00698)	(0.00951)	(0.007)	(0.00803)
unemp_o	-11.49***	-12.69***	-6.101***	-11.49***
	(0.792)	(1.12)	(0.865)	(0.922)
coast _o	-0.0892***	-0.105***	-0.039**	-0.0892***
	(0.0155)	(0.0211)	(0.016)	(0.0195)
pop_d	0.648***	0.703***	0.120***	0.648***
	(0.0104)	(0.0139)	(0.012)	(0.0127)
salary_d	0.241***	0.0771	0.515***	0.241***
	(0.052)	(0.0694)	(0.057)	(0.0646)
housing_d	0.0761***	0.114***	0.063***	0.0761***
	(0.00766)	(0.0102)	(0.008)	(0.00912)
unemp_d	-1.627***	-1.907***	0.324**	-1.627***
	(0.127)	(0.186)	(0.138)	(0.155)
coast_d	-0.119***	-0.142***	-0.115***	-0.119***
	(0.0157)	(0.0216)	(0.016)	(0.0196)
a-school_d	-0.0775**	-0.0900**	-0.019	-0.0775**
	(0.0305)	(0.0394)	(0.033)	(0.0369)
uni_d	0.00844	0.0930**	-0.005	0.00844
	(0.0287)	(0.0425)	(0.030)	(0.0364)
Constant	-7.222***	-7.116***	-1.749	-7.222***
	(0.214)	(0.283)	(0.233)	(0.295)

**p < .10, **p < .05, ***p < .01*

Table 2 reveals how the pooled model performed on the different estimators mentioned in the section "methodology". Having a quick look at the table, one can observe that most of the signs for the traditional controls of the gravity are as expected. The population at the origin indeed shares a positive coefficient with the flow of migrants and same goes for accessibility. As expected, unemployment is a significantly negative pull factor meaning that migrants are not attracted to the municipalities with a high level of unemployment. It is also a negative push factor, perhaps, because there is a higher financial difficulty to move out of a municipality when unemployed. The coast line seems to have a negative impact on both outflows and inflows. It implies that people who live in a municipality by the sea are more likely to remain in the municipality but it also means that immigrants are not attracted by the sea. This could be linked to the idea of the sociologist Alain Corbin (1993) who argued that the coast line has not always been a factor of attraction for migration and that the coast can also be regarded as an inferior amenity, depending on where and when the study takes place. However, this result goes against most of the previous literature which considers the coast line a superior amenity. This result should be investigated in further researches specifically on the topic.

In order to pursue with the fixed effect models, we first need to evaluate which estimator might be the most appropriated. As seen in the section "data", the large heterogeneity of the province might lead the Poisson estimator to be biased. Moreover, when predicting zero-flows with the negative binomial regressor, it seems that the number of predicted 0-flows (59,549) is higher than the number of actual 0-flows in the dataset (57,613). Burger et al. (2009) argue that the zero-inflated estimator is superior to the Negative Binomial estimator if the number of observed 0-flows is higher than the predicted 0-flows. However, we can see that this is not the case and so there is no need to inflate any regressor. Nonetheless, the validity of the negative binomial regressor can be discussed. As seen earlier, there is a large spatial heteroskedasticity in the dataset and although the negative binomial regressor is robust to heteroskedasticity, the pseudo-Poisson maximum likelihood proves to be more effective in reaching convergence in dataset with large heteroskedasticity. As a matter of facts, in this study, convergence was not reached with the negative binomial regressor when including fixed effects. So, it seems that the pseudo-Poisson maximum likelihood is the most relevant in this study.

As mentioned in the "methodology" section, results of the pooled model are believed to be biased. In order to discuss the results for the variables of interest on education infrastructure, we will have to take a look at the fixed effect models. Table 2 summarizes the results displaying only for the variables of interest (full table in appendix Table 2A).

Table 3: Results for the fixed-effects models

	Pooled	**Model 1**	**Model 2**
a-school_d	-0.0775**	0.382	1.326**
	(0.0305)	(0.899)	(0.608)
uni_d	0.00844	0.157	1.079
	(0.0287)	(0.737)	(0.847)
Dummies for years	**YES**	**YES**	NO
Dummies for origin	NO	**YES**	**YES**
Dummies for destination	NO	**YES**	NO
Dummies for time-origin	NO	NO	**YES**

The model 1 from the traditional approach is not providing with any evidence of the presence of pull factors for education infrastructures. Moreover, as seen earlier, this model may not control for the heterogeneity of the push factors at the origin and therefore, these results may be biased. However, model 2 from the approach of Ortega and Peri (2013) including dummies for time-origin and for destination provides us with reliable evidence that the presence of an agricultural school in the municipality drives a higher inflow of migrants into that municipality. Therefore, the agricultural schools are indeed a pull factor for gross migration as supposed in the hypothesis 2. In the pooled model, this coefficient might have been negative because this model was not controlling for the heterogeneity of location. Since agricultural schools are often located in remote and unattractive areas, the pooled model captured some of the negative effect of the location of the institution. Thanks to the fixed effects model, we managed to isolate the positive effect of such specialized institutions on the decision to migrate.

However, it seems that the presence of a university is not driving any variation in the inflow of migrants as its coefficient is not significantly different from 0. So, there is no evidence that a university is a pull factor of gross migration. This goes against the hypothesis 1 drawn in the literature review. A reason for this result may be due to the one of the limitations discussed earlier: since the census data relies on townhall registrations, students may keep their original address and may not register to the townhall close to their place of study. This means that the universities may not capture a lot of relevant flows. This would explain why there are some difficulties to capture the effect on the gross migration flows. On the contrary, with agricultural schools, the age of enrolment (on average 16) is usually lower than universities (on average 18) and therefore, the registered address may be completed by the family and may correspond better to the actual place of residency. Thus, the limitation of this study causes less problems to verify the hypothesis for specialized schools.

Moreover, another limitation to these results is a problem that Bruno and Improta (2006) already addressed in the previous literature: spatial dependence. The results of this study may suggest that there is indeed a large spatial dependence between the different municipalities and that the presence of a university does not only impact the immigration flows towards the university's municipality: it will also impact the immigration towards the neighbouring provinces and students may choose to commute every day from the residency to the place of study (as we saw earlier, some municipalities are only 5 minutes away by car). For further research, the introduction of students' commuting flows together with spatial dependencies might be relevant to understand better how specialized schools affect migration.

6. Conclusions

The case of the Bouches-du-Rhône has provided us with some evidence that specialized schools and especially agricultural schools are important drivers of migration. As it seems, the role of education infrastructures in migration is more complex than a mere attractive force. The results of our study tend to support that universities only have an impact on student migration and that gross migration flows are not affected by the presence of a university. Confirming the results of Cullinan and Dogan (2016), the specialized institutions seem to have much more weight than universities in the migration decision of households. Especially considering the fact that agricultural institutions are often located in remote and rural areas which are not attractive for migration. Controlling for the heterogeneity of location, we managed to capture the positive effect of a specialized institution on the attractiveness of its municipality.

Concerning the different policies that the governments could follow to achieve a successful functional decentralization, specialized schools have a key role in achieving this objective. Instead of focusing on the decentralization of the university system as suggested by most of the previous literature, governments should invest on the specialized schools of peripheral areas as they seem to be a more relevant driver of gross migration. This implies that the NOTRe French reforms of 2015, focusing on the decentralization of the general university system may not be the optimal method to dynamize peripheral areas. Contrarily, in Holland, the importance of specialized schools in gross migration has explained the success of the 1970s investment of the Dutch government in specialized schools of Wageningen. It is clear that policies aiming at developing specialized education infrastructures in peripheral municipalities will have a stronger impact on household migration compared to policies focusing on the decentralization of the general university system. Therefore, education infrastructures do have an important role in the general objective of decentralization that many European countries are pursuing and investing in the appropriate curriculum is key to the success of a functional decentralization policy.

References

1. Alm, J. Winters, J. V. (2009) “Distance and intrastate college student migration”, Economics of Education Review, 28, 728-738.

2. Alonso, W (1964) *Location and land use,* Harvard University Press, Cambridge, United States.

3. Bahar, D. Choudhury, P. Rapoport, H. (2019) “Migrant Inventors and the Technological Advantage of Nations”, Working paper.

4. Bouzouina, L (2007) “Concentrations spatiales des populations à faible revenu, entre polarisation et mixite”,Pensée Plurielle, 16, 59-72.

5. Bruno, G. Improta, G. (2006) “Using gravity models for the evaluation of new university site locations: A case study”, Computers & Operations Research, 35, 436-444.

6. Beine, M. Bertoli, S. Moraga, J. F. H. (2015) “A practitioners' guide to gravity models of international migration”, The World Economy, 39(4), 496-512.

7. Beine, M. Bourgeon, P. Bricongne J.-C. (2013) “Aggregate Fluctuations and International Migration,” CESifo Working Paper No. 4379, Munich.

8. Blijie, B. (2005) “The impact of accessibility on residential choice - empirical results of a discrete choice model”, ERSA conference papers, European Regional Science Association.

9. Brümmer, B. Kareem, F. O. Martinez-Zarzoso, I. (2016) “Fitting the Gravity Model when Zero Trade Flows are Frequent: a Comparison of Estimation Techniques using Africa's Trade Data”, Working Paper.

10. Burger, M. van Ort, F. Linders, G. (2009) “On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation”, Spatial Economic Analysis, 4(2), 167-190.

11. Cameron, C. Trivedi, P. K. (2001) "Essentials of Count Data Regression", A Companion to Theoretical Econometrics, 331-348, Blackwell, Oxford.

12. Carree, M. A. Kronenberg, K. (2014) “Locational choices and the costs of distance: Empirical evidence for Dutch graduates”, Spatial Economic Analysis, 9(4), 420-435.

13. Caviggioli, F. Jensen, P. Scellato, G. (2020) "Highly skilled migrants and technological diversification in the US and Europe", Technological Forecasting and Social Change, 154.

14. Corbin, A. (1993) "Le territoire du vide : l'Occident ou le désir du rivage", Revue d'Histoire Moderne et Contemporaine, 40(4), 691-695

15. Correia, S. Guimarães, P. Zylkin, T. (2020) "Fast Poisson estimation with high-dimensional fixed effects", The Stata Journal, 20(1).

16. Cullinan, J. Duggan, J. (2016) "A School-Level Gravity Model of Student Migration Flows to Higher Education Institutions", Spatial Economic Analysis, 11(3), 1-21.

17. De Vries, M. S. (2003) "The rise and fall of decentralization: A comparative analysis of arguments and practices in European countries", European Journal of Political Research, 38(2), 193-224.

18. Dong, X. Zheng, S. Kahn, M. E. (2020) "The role of transportation speed in facilitating high skilled teamwork across cities", Journal of Urban Economics, 115.

19. Eaton, J. Kortum, S. S., Sotelo S. (2012) "International Trade: Linking Micro and Macro", NBER Working Paper No. 17864.

20. Faggian, A. Franklin, R. (2014) "Human Capital Redistribution in the USA: The Migration of the College-bound", Spatial Economic Analysis, 9(4), 376-395.

21. Florax, R. J. G. M. (1992) *The University: A Regional Booster – Economic Impacts of Academic Knowledge Infrastructure*, Ashgate, Aldershot, United Kingdom.

22. Head, K. Mayer, T. (2014) "Gravity Equations: Workhorse, Toolkit, and Cookbook", Handbook of International Economics, 4, 131-195

23. Jourdan, S. (2006) "Marseille and the difficulties of gentrification", Journal of Mediterranean Geography, 111, 85-90.

24. Kjellström, C. Regnér H. (2006) "The Effects of Geographical Distance on the Decision to Enrol in University Education", Scandinavian Journal of Educational Research, 43(4), 335-348.

25. Kodrzycki, Y. (2001) "Migration of recent college graduates: evidence from the National Longitudinal Survey of Youth", New England Economic Review, 1, 13-34.

26. Koster, H. R. A., Rowendal J. (2017) "Historic Amentities and Housing Externalities: Evidence from the Netherlands", The Economic Journal, 127(605), 396-420.

27. Lee, S. Lin, J. (2018) "Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income", The Review of Economic Studies, 85(1), 663–694.

28. Leppel, K. (1993) "Logit Estimation of a Gravity Model of the College Enrolment decision", Research in Higher Education, 34(3), 387-398.

29. LeSage, J. P. Pace, K. R. (2016) “Spatial Econometric Modelling of Origin-Destination Flows”, Journal of Regional Science, 48(5), 941–967.

30. Mayda, A. M. (2010) “International migration: A panel data analysis of the determinants of bilateral flows”, Journal of Population Economics, 23(4), 1249-1274.

31. McFadden, D. (1974) “The measurement of urban travel demand”, Journal of Public Economics, 3(4), 303-328.

32. Miguélez, E. Moreno, R. (2014) “What attracts knowledge workers? The role of space and social networks”, Journal of Regional Science, 54(1), 33-60.

33. Mincer, J. (1978) "Family Migration Decisions", Journal of Political Economy, 86(5), 749-773.

34. Ohnet, J.-M. (1996) *Histoire de la décentralisation française,* libraire générale française, Paris, France

35. Piras, R. (2016) “A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches”, Papers in Regional Science, 96(3), 571-602.

36. Ritsilä, J. Ovaskainen, M. (2001) “Migration and regional centralization of human capital”, Applied Economics, 33(3), 317-325.

37. Royuela, V. Ordoñez-Cuenca, J. (2016) “Internal migration in a developing country: a panel data analysis of Ecuador (1982-2010)”, Papers in Regional Science, 97(2), 345-367.

38. Sa, C. Florax, R. J. G. M. Rietveld, P. (2004) “Determinants of the Regional Demand for Higher Education in The Netherlands: A Gravity Model Approach”, Regional Studies, 38(4), 375-392.

39. Santos-Silva, J. M. C., Tenreyro S. (2010) “On the Existence of the Maximum Likelihood Estimates for Poisson Regression”, Economics Letters, 107(2), 310-312.

40. Spiess, C. K. Wrohlich, K. (2010) “Does Distance Determine Who Attends a University in Germany?”, Economics of Education Review, 29(3), 470-479.

41. Tay, C. (2014) “An econometric model on bilateral trade in education using an augmented gravity model”, Journal of Industrial Engineering and Management, 7(2), 401-412.

42. Venhorst, V. Van Dijk, J. Van Wissen, L. (2010) “Do the Best Graduates Leave the Peripheral Areas of the Nehterlands?”, Journal of Economic and Social Geography, 101(5), 521-537.

Appendix

Table A1: Main variables, sources and expected sign

Variables	proxy	Expected sign	Source
Population	Population (ln)	+	INSEE
Accessibility	Google time distance (ln)	-	Google
Salary	Average salary (ln)	+	INSEE
Housing	Historical housing (ln)	-	INSEE
Unemployment	Unemployment rate (%)	-	INSEE
Amenities	Dummy for coast (ln)	+	Google
Education	Number of Universities (ln), Agricultural schools (ln)	+ +	INSEE INSEE

Table A2. Correlation matrix

	dist	pop	salary	housing	unemp	a-school	uni
dist	1						
pop	-0.077	1					
salary	-0.0859	-0.061	1				
housing	-0.0392	0.8169	-0.154	1			
unemp	-0.0516	0.7615	-0.288	0.708	1		
a-schools	0.0215	0.0965	-0.012	0.1074	-0.0109	1	
uni	-0.0009	0.429	-0.12	0.4913	0.6612	-0.0715	1

Table A4: fixed-effect models

	Model 1	**Model 2**
dist	-1.635***	-
	(0.0150)	
pop_o	0.547	-
	(0.447)	
salary_o	-0.0132	-
	(0.757)	
housing_o	-0.00251	-
	(0.0906)	
unemp_o	2.431	-
	(10.17)	
coast _o	-0.535	-
	(0.405)	
pop_d	-0.0918	0.0950
	(0.442)	(0.496)
salary_d	-0.577	-0.775
	(0.770)	(0.787)
housing_d	0.0693	0.0673
	(0.0932)	(0.0959)
unemp_d	2.797*	2.042
	(1.453)	(1.558)
coast_d	0.696	-0.279
	(0.581)	(0.424)
a -school_d	0.382	1.326**
	(0.899)	(0.608)
uni_d	0.157	1.079
	(0.737)	(0.847)