Università degli Studi di Messina

Department of Economics

Ph.D. Economics, Management and Statistics - XXXIV Cycle

Ph.D. Coordinator: Prof. *Edoardo Otranto*

Curriculum: *Economics* **SSD**: *SECS-P/02*

DOES CORRUPTION INFLUENCE YOUNG BRAIN DRAIN? EVIDENCE FROM ITALY

Ph.D. Candidate:

Dott.ssa *Alessandra Patti*

ID: 492348

Supervisor:

Prof. *Emanuele Millemaci*

Dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Department of Economics

Academic Year 2020-2021

Author Declaration

I Alessandra Patti declare that this thesis, titled “Does Corruption Influence Brain Drain? Evidence from Italy”, submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy at Department of Economics of University of Messina is my own work unless otherwise referenced or acknowledged. I certify that:

- This work was performed during my research activity as full time Ph.D. student
- Where any part of this thesis has previously been submitted to conferences, this has been clearly stated
- Where I have consulted the published works, this is always clearly attributed.
- Where I have quoted from the work of others, the sources are always given.

Messina 15/02/2022 *Alessandra Patti*

List of Contents

Acknowledgments

"When the winds of change blow,
some people build walls and others build windmills"
Chinese Proverb

First and foremost, I would express my gratitude to Professor Emanuele Millemaci who has guided and supported me from the first time I asked him to be my supervisor during my career. In fact, thanks to his wise advice, capacity to teach clearly, human touch and professionality, he has helped me to find the right direction to achieve my goals, satisfy my research interests and to learn enthusiastically. I would extend my gratitude to Professor Edoardo Otranto, who was my professor of Advance Methods for Statistics during my M.Sc. course and has inspired me to develop interests toward the research world. Also, I would thank Professors Monica Milasi and Maria Bernadette Donato who provide me insightful guidelines at the beginning of my Ph.D. course. I also thank my colleagues with whom, although for the first year of the Ph.D. course, I have shared lessons, launch breaks, travels by bus to attend lessons at the University of Catania, joyful moments but also worries related the Ph.D. concerns. A heartly thanks to my family who has constantly supported me with encouragements and love and have motivated me to never give-up in arduous moments. A special thanks to my boyfriend, Ugo, who is always by my side, support and endure my fortunes and misfortunes with reassuring confidence, lovely courtesy, and heart-warming smile. I would express my gratitude to Valeria who has spent tons of encouraging words and love to withstand my worries but also my happiest moments. Finally, I would also thank the first-year students of my first Principles of Economics course, who give me the opportunity to teach and learn the beauty of sharing ideas, knowledge, experiences.
I dedicate this Ph.D. to my courage, determination and passion for what I love most and to those who I encountered during this journey as well as my friends who have teach me to strive incessantly to achieve 'the best I can'

List of Tables

List of Figures

Abstract

In recent years, young Italian brain drain within provinces has increased at higher speed than ever. While is premature to assess whether this process is transitory or permanent, it is undoubted that it is relevant and needs to be analysed by researchers and monitored by policy makers constantly. Previous empirical studies have demonstrated that net skilled migration is influenced by economic factors, such as the search for higher income per capita and job opportunities, and, with a less extent, by the search of places endowed with more amenities. In the crossroad between these factors, this dissertation investigates the role of corruption, as proxy for meritocracy, as a key element of influence over young skilled mobility. To this end, a comprehensive framework of analysis, based on the comparison of results get by traditional and novel empirical methodologies, is used. Hence, the present dissertation develops its study in three chapters. Chapter 1 offers a detailed review of the literature that has analysed the general causes of skilled mobility and discusses the novel elements introduced by the current study. Then, Chapter 2 investigates the relationship between corruption and the Italian skilled mobility by exploiting traditional data and dynamic panel model. Evidence suggests that high corruption positively affects skilled flows from origin province, ceteris paribus. Although the adopted model is robust because it is widely exploited among researchers and controls for endogeneity, its main limit is represented by the fact that it does not fully exploit the potentialities offered by bilateral data of a tri-panel dataset, losing great part of information. Thus, Chapter 3 deals with the trade-off between robustness and completeness handled by the traditional method presented in Chapter 2 and adopts a gravity framework with a novel Pseudo Poisson. Results suggest the existence of push and pull mechanisms of corruption at play on young skilled mobility. Besides, evidence proves that sensitivity of the prospective tertiary students towards corruption varies according to their field of study of interest. Also, corruption widely affects long-distance skilled flow from the Centre-South to the North of Italy. Chapter 3 enriches its analysis with bilateral data on enrolments at university, additive research questions and results by adopting a remedy that does not give up completeness for gaining robustness in the empirical analysis. Results of the novel model of Chapter 3 present similarity with the results of the traditional model of Chapter 2, demonstrating that the coexistence of robustness and completeness features is possible if models are correctly implemented.

Keywords: Brain Drain, Corruption, Panel Data, system-GMM, Gravity, Pseudo Poisson Max Likelihood

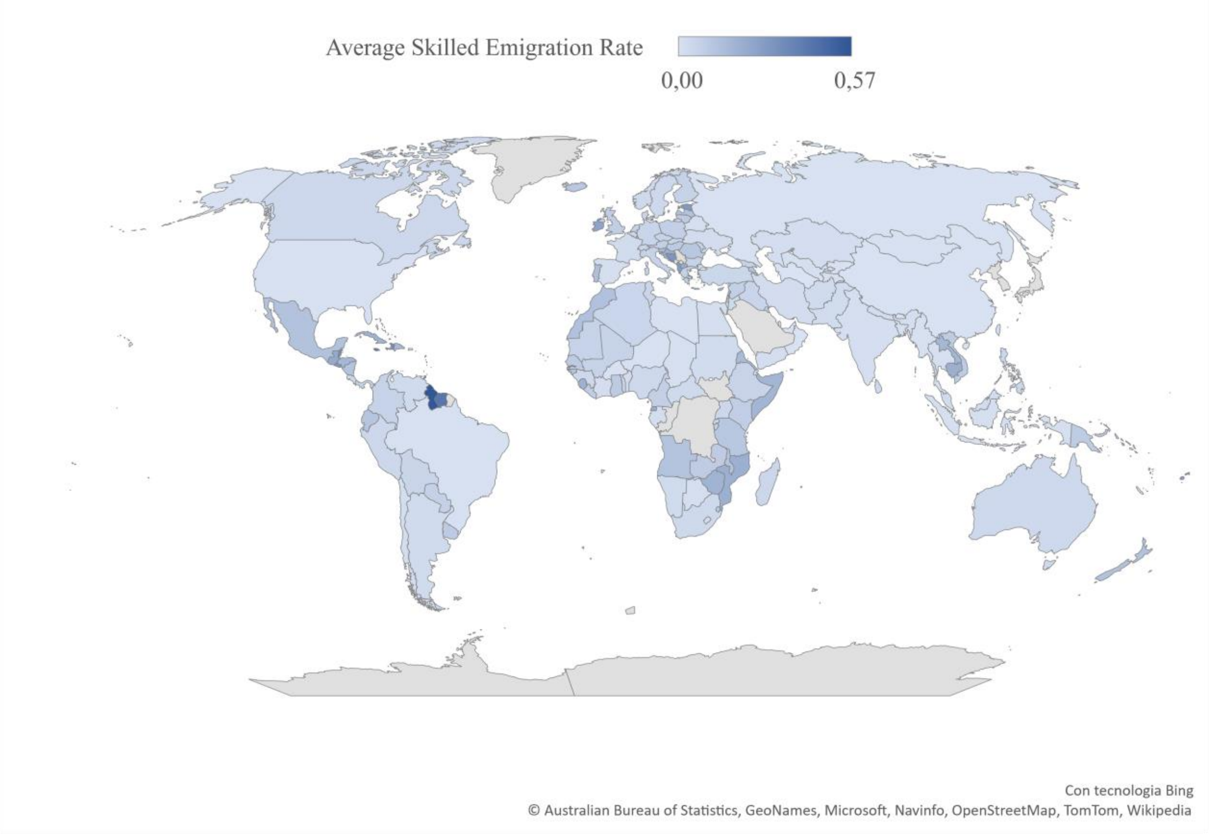
Introduction

Net skilled migration constitutes a bulk of interest for researchers of several fields of study including (endogenous) economic growth. The economic growth path of a country depends crucially by its human capital accumulation. According to Solow, 1956, technological change was the main determinant of economic growth. However, Solow did not consider technological change as an endogenous factor, as later Arrow, 1962, Romer, 1987 and Lucas, 1988, argued. In fact, they indicated technological change and human capital accumulation, through schooling and learning-by-doing, as main endogenous elements of growth. Nonetheless, in the quest for growth, increasing human capital has usually been considered an adequate policy (Beine et al., 2014). Usually, researchers have long been sustained the pessimistic view of brain drain. In fact, brain drain from sending places is conceived as human capital depletion with negative spill-over effects meanwhile brain migration to receiving places turns to be a gain because such places acquire and plausibly retain skilled individuals (Beine et al., 2014). In contrast, a new wave of research has emerged around the idea that net skilled migration also generates beneficial effects for sending places, by partly or totally compensating for the costs of losing talents (Beine et al., 2014). More precisely, skilled individuals who emigrate from native places leave out opportunities, in terms of employment or living standards, that become higher for those who remain. Besides, the costs of those who emigrate is attenuated if origin places receive larger remittances and other benefits from skilled returns. In this manner, inequalities between sending and receiving places can be mitigated. Thus, several studies have identified economic and socio-cultural factors, such as the search for job opportunities and higher income, better quality of life and quality of institutions as relevant causes for skilled mobility (Poprawe, 2015; Nifo and Vecchione, 2014; Charron et al., 2013; Dotti et al., 2013; Biagi et al., 2011). On the other hand, few part of it has already demonstrated the negative effects of corruption on growth (Simovic, 2021; Corrado and Rossetti, 2018; Lisciandra and Millemaci, 2016) and even fewer part of it has proved corruption as remarkable social factor that positively affects skilled flows from source contexts (Cooray and Schneider, 2016; Poprawe, 2015; Dimant et al., 2013). Thus, understanding how corruption influences skilled mobility is an issue of crucial importance for researchers and policymakers due to many socio-economic consequences, such as more unequeal re-distribution of human capital within territiories, that derive from it. Although it is simple to assume that high corruption positively affects skilled flows, there is not obvious evidence that skilled individuals are attracted to destinations where corruption is relatively low. This is partly explained by the fact that resident students from origin have perfect information of the unlawful cases of their context (such as achieving goals without merit but just for favouritism, nepotism and/or political connections) and decide to move to destinations where is expected that corruption is lower. To presume it, they collect information on their designed destinations from newspapers, web, media, parents and friends, who already live-in but such information is likely to be reported in misleading and incomplete way. Hence, the push effects of corruption are more defined than its pull effects. For that reason, the aim of this dissertation consists of analysing whether corruption acts not only as push factor, that determines skilled flows from origin, but also as pull factor that attracts skilled individuals to destinations, after they gather sufficient information on the quality of the context of their designed destinations. This happens because young skilled individuals, at the beginning of their university career, are more sensitive to legal concerns for living better-off in society. Besides, they know that the opportunities of their career depend on the competitiveness and meritocracy of the local labour market; in fact, they are looking forward to having chances to live in places where deserving rewards are likely to occur fairly. Thus, the decision of students to move from their native province is complex because there are innumerable motivations behind. Moving for "study reasons" may be related to move in search of "higher quality of life", where higher quality of life measures a mix of economic and social factors related to welfare, social mobility, efficiency and availability of services and infrastructures (Nifo and Vecchione, 2014).

Hence, the present section of this dissertation illustrates stylized facts on human mobility and corruption that help to give an idea of their spread and relevance from broad and narrow views (Italy).

Recent times have experienced a significant variation of the average migration across countries. In fact, two-thirds of the world's migrants move to North America, Western and Eastern Europe and to the oil-exporting Gulf Cooperation Council (GCC) countries in the Persian Gulf whereas one-third of the world's skilled migrants move to North America and to the northern part of West Europe, making Europe the elected destination for more than one over three skilled emigrants in the world (World Bank, 2018). Precisely, the average skilled emigration rate increases by roughly 5 percent (%) from 2010 to 2017 worldwide and skilled people tend to concentrate in a handful of destinations with ongoing labour shortages and high demand for specific sectors and occupations. In fact, as Figure 1 depicts, the average skilled emigration rate tends to be higher in the southern area (darker-blue area) rather than in the northern countries (lighter-blue area) of the world in 2017[1].

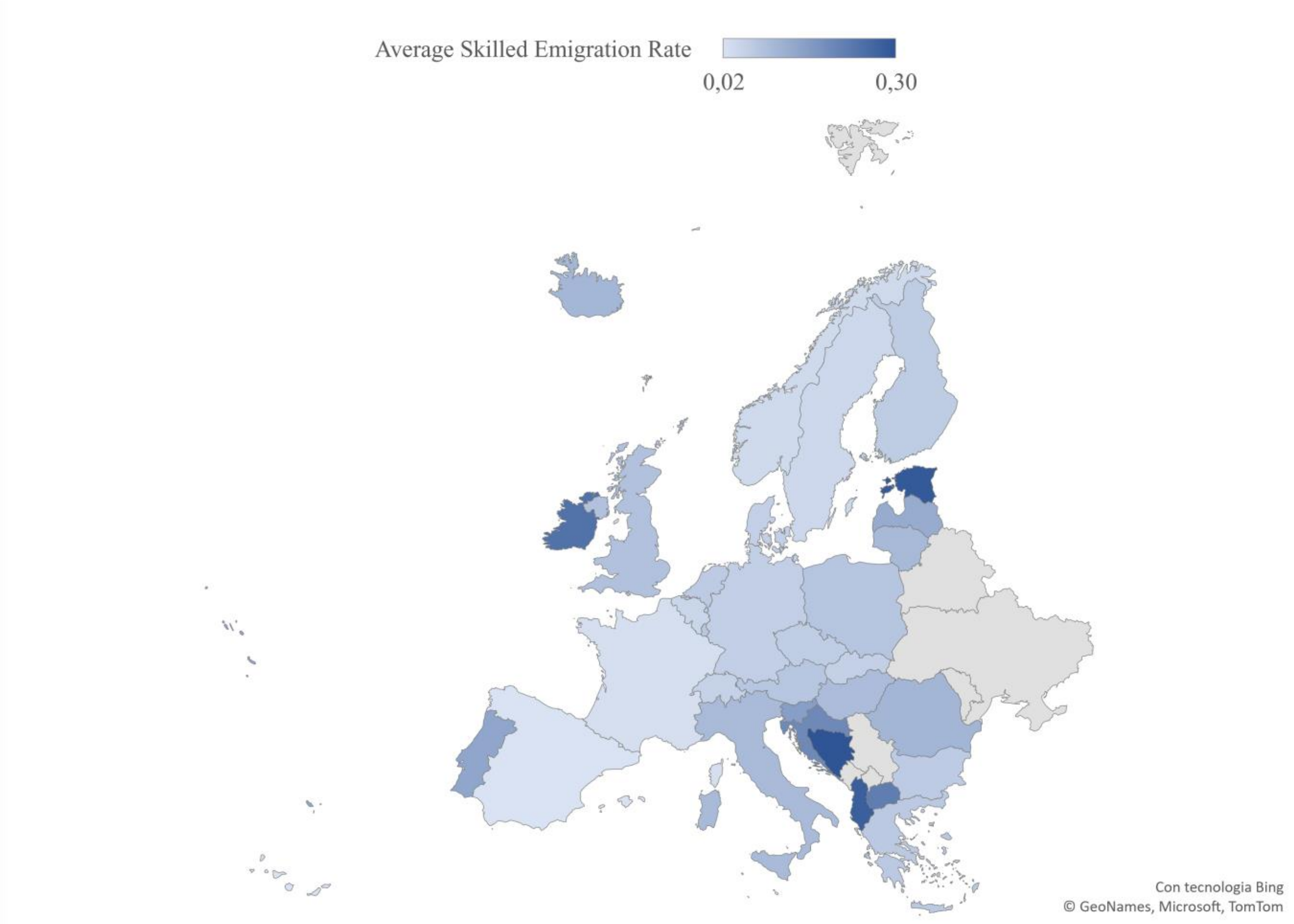
Figure 1 *Average Skilled Emigration Rate Worldwide 2017 (source: OECD)*



[1] Data are taken by OECD (2022) "*International Migration Statistics Database*" and re-elaborated by the author with Microsoft Excel with Bing Technology to realize geographical charts. Grey-shaded area indicates missing data for that country.

In addition, as the following Figure 2 demonstrates, the average skilled emigration rate reaches the 11 percent (%) between countries of the Eurozone in 2017. Precisely, Eastern nations (plus Ireland and Portugal) register roughly 30 percent (%) of the average skilled emigration rate against 2 percent (%) and 15 percent (%) reported by western area (for example, Italy reports almost 11 percent -%-)[2].

***Figure 2** Average Skilled Emigration Rate of Eurozone 2017 (source: OECD)*

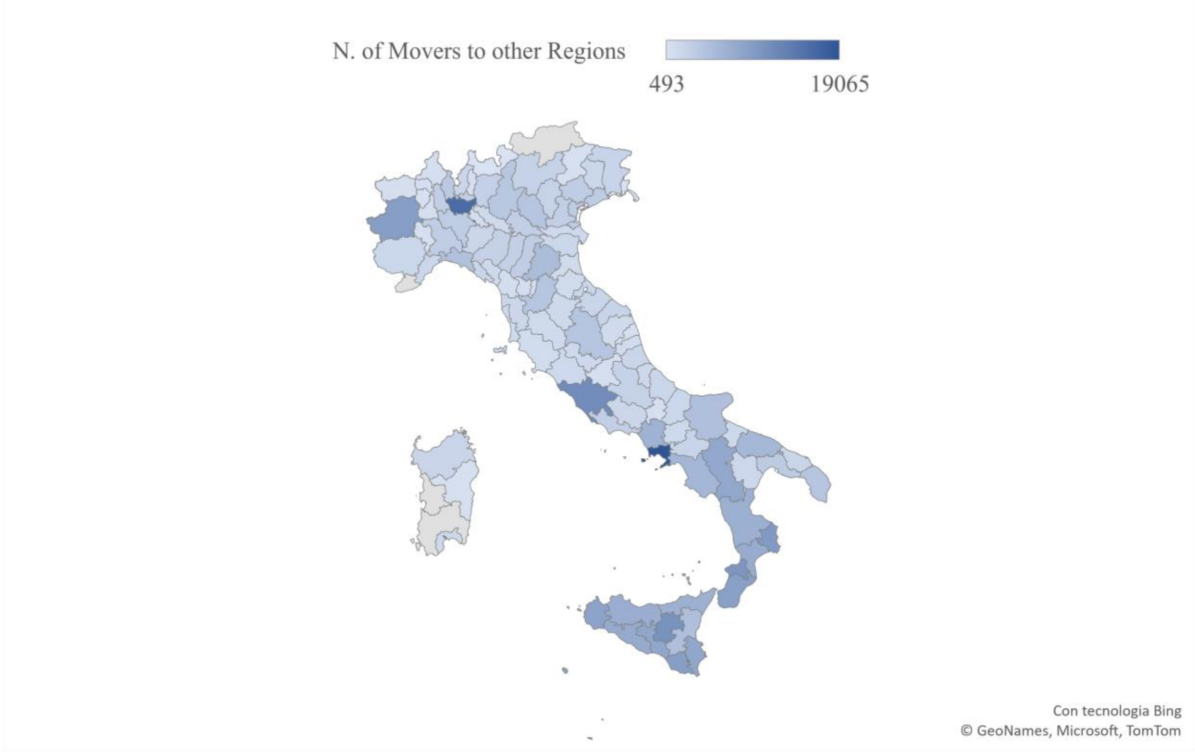


Thus, net skilled migration is widespread phenomenon that interests many countries all around the world. Skilled mobility fuels the socio-economic dualism status for developing-developed countries and for the North and South of developed countries. As matter of fact, an interesting case that is representative of above-mentioned North-South division is proven by Italy, which is one of the developed countries in Europe that has one of the highest levels of skilled emigration rate (11 percent

[2] Data are taken by OECD (2022) "*International Migration Statistics Database*" and re-elaborated by the author with Microsoft Excel with Bing Technology to realize geographical charts. Grey-shaded area indicates missing data for that country.

-%- against 3 percent -%- in France and 6 percent -%- in Germany in 2017[3]), characterized by ample movements of individuals within its main territories. Precisely, the Italian interprovincial mobility is a relevant phenomenon for the South. From 2002 to 2017, As reported by SVIMEX[4] (Italian Association for the Industrial Development of Mezzogiorno), the number of movers from Mezzogiorno is about 2 million, out of which approximately 150.000 in 2017 only. Figure 3 depicts these data, where the dark-blue shaded area indicate that the southern territories, with some exceptions, of Italy register higher number of movers to other regions in 2017.

Figure 3 *Number of Movers to other Italian Regions 2017 (source: ISTAT)*



Given that numbers, almost 50 percent (%) of those who move are students (66.557) while 33 percent (%) are already graduated (21.970). Besides, from 2010 to 2017, the net skilled migration rate from the South increase more than double than the net skilled migration rate from the North and it still follows an increasing path: as illustrated by Figure 4, southern skilled movements rise of 4 percent (%) against to 2 percent (%) of skilled movements occurring from the North[5].

[3] Data are provided by OECD (2022) "*International Migration Statistics Database*"
[4] Annual Report: "*Rapporto Svimex 2018 sull'Economia del Mezzogiorno*", Il Mulino, available at SVIMEZ
[5] Data are taken by M.I.U.R and elaborated by the author with STATA

Figure 4 *Migration Rate per Italian Area 2010-2017 (source: MIUR)*

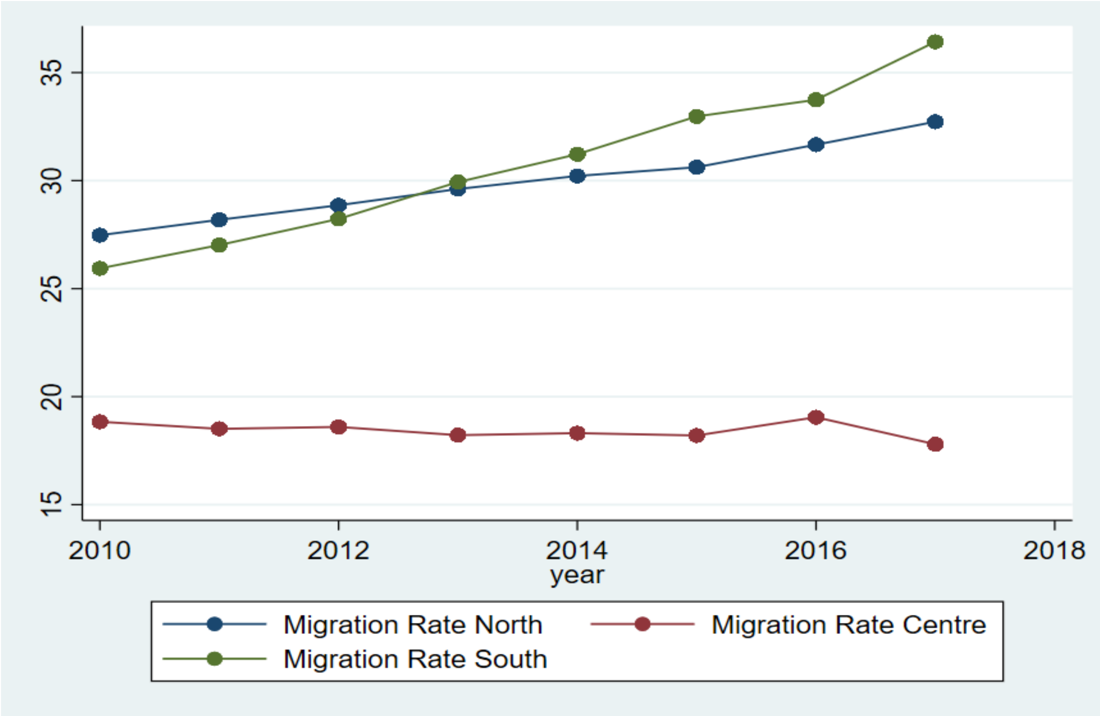


Figure 5 *Migration Rate within Italian Macroarea from 2010 to 2017 (source: MIUR)*

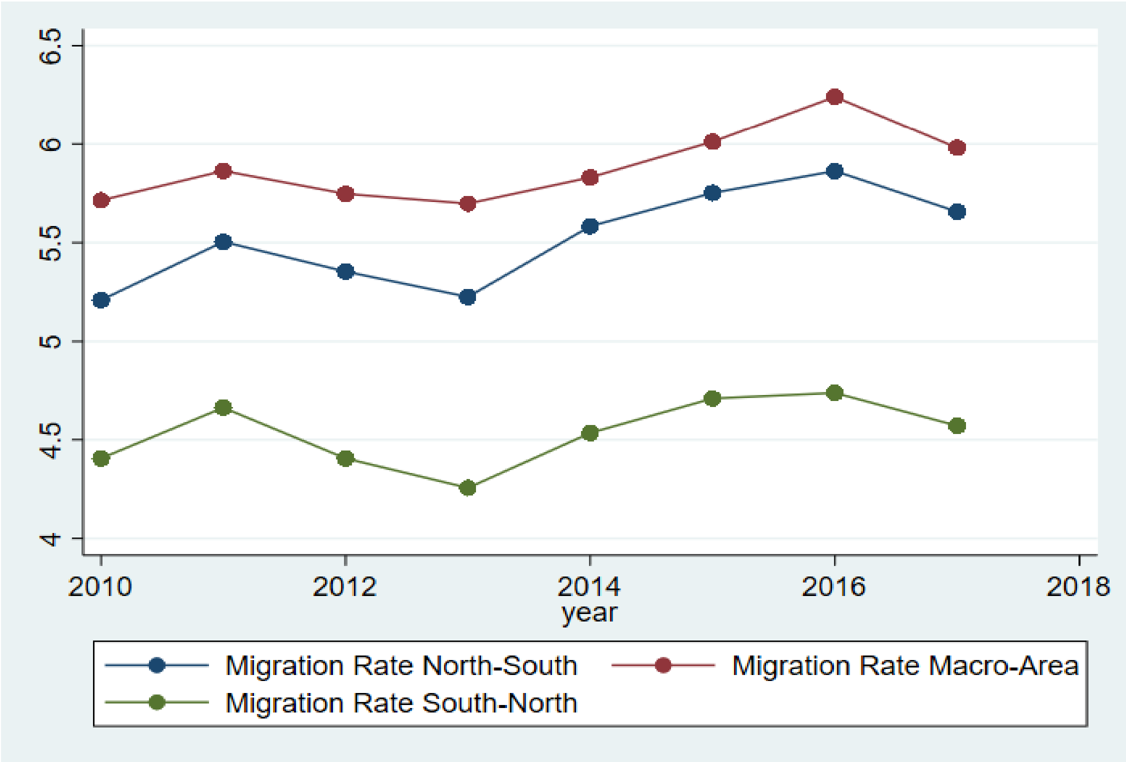
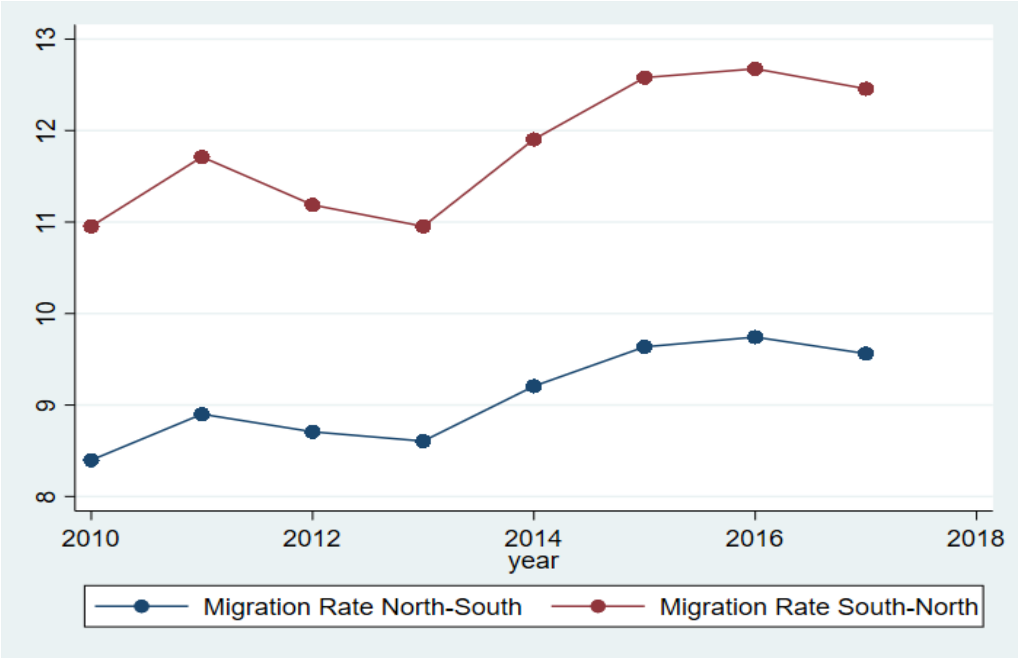


Figure 6 *Net Skilled Migration related Total Migration per Macroarea of reference (source: MIUR)*

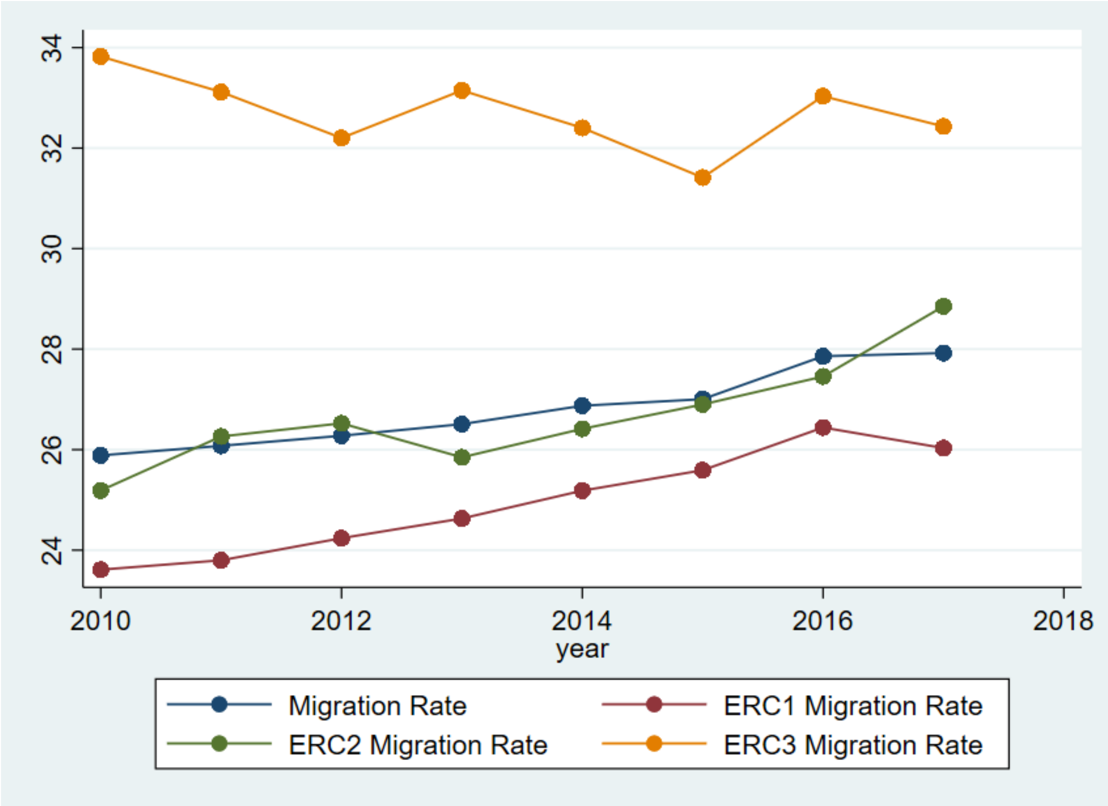


In addition, vast part of inter-macroarea skilled mobility occurs from the Centre to the North and from the Centre-South to the North of Italy. As Figure 5 demonstrates, almost 0.5 percent (%) of the total movements happen from the North to the Centre or to the South against the 4.5 percent (%) of total movements that occur from the South to the Centre-North of the country[6]. Besides, Figure 6 shows that between 11 percent (%) and 13 percent (%) are skilled movements occurring within provinces from the South while between 8 percent (%) and 10 percent (%) represent the provincial skilled movements from the Centre-South to the North[7]. This implies severe economic consequences on growth of each different macroarea: as reported by SVIMEX, from 2007 to 2017, the difference in economic growth between the Centre-North and the South was 9.6 percent (%) against 5.3 percent (%) if skilled mobility from the South never existed, that certifies the economic dualism of territories. In addition, as Figure 7 reports, Italian skilled movements can be categorized per students who enrol to courses of different fields of study: from 2010 to 2017, the number of prospective tertiary students of Social Science (ERC-1) who emigrate rise by 8 percent (%) as well as the number of prospective

[6] Data are taken by M.I.U.R and elaborated by the author with STATA
[7] Data are taken by M.I.U.R and elaborated by the author with STATA

tertiary students of Physical Science (ERC-2) rise by 16 percent (%) whereas the number of prospective tertiary students of Life Science (ERC-3) decreases by a mild 5 percent (%)[8].

Figure 7 *Net Skilled Migration per field of study (source: MIUR)*



Hence, in the last ten years, Italian provinces have experienced an increasing trend of skilled movements that has achieved a peak to higher levels as never before. It is not possible to know if this phenomenon will stop or rise even further, although the advent of external shocks may cause a reverse trend of skilled mobility. An example is provided by the widespread of Coronavirus disease that has determined a likely return of skilled individuals to their homeplaces, as a recent survey of Il Sole 24 Ore reports: on 23 percent (%) of skilled emigrants interviewed, 1 student over 5 has decided to return in Italy and 1 student over 4 has decided to return to his/her origin place. Undoubtedly, the return of

[8] Data are taken by M.I.U.R and elaborated by the author with STATA

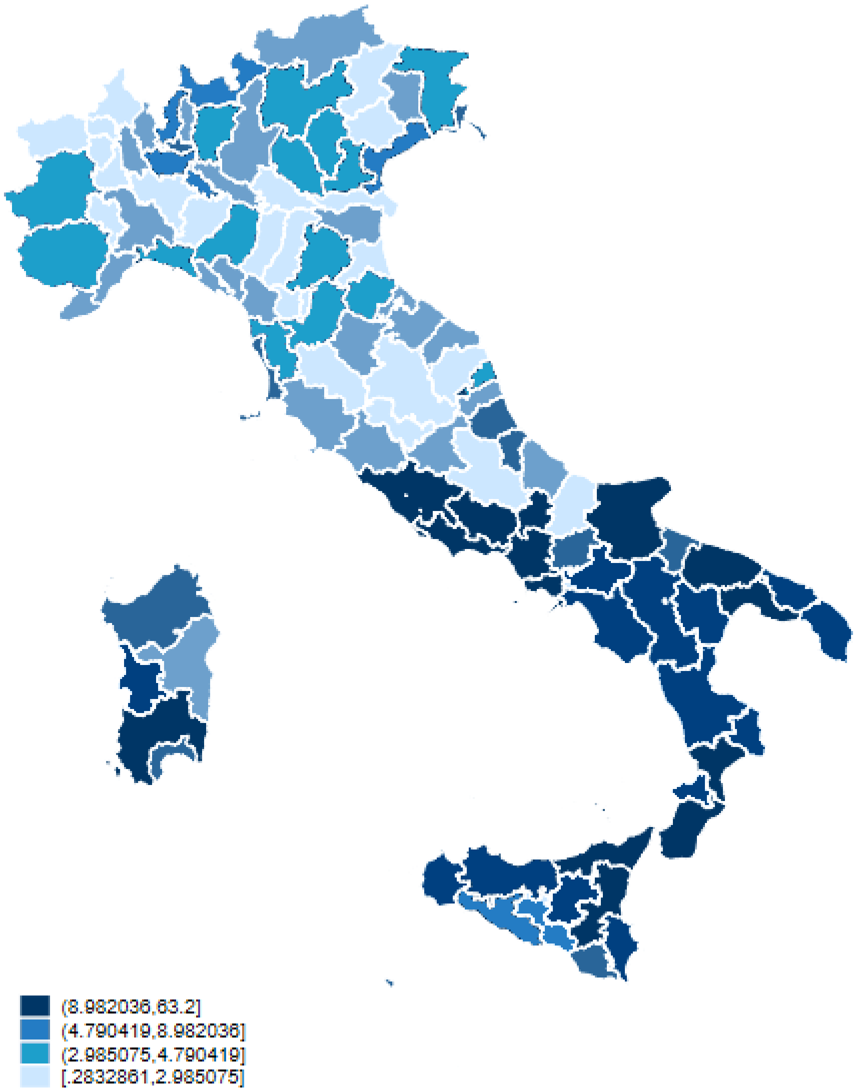
talents can reduce the harmful effects of losses of resources at origin but also remittances, sent by workers to support their families at origin, are effective remedies against the negative economic consequences derived by skilled escapes. As reported by Bank of Italy, remittances received in Italy rise and account from 0.3 percent (%) in 2010 to 0.5 percent (%) of GDP in 2017[9].

Similarly, corruption is a widespread phenomenon alongside brain drain in Italy. As reported by the Corruption Perception Index (CPI), published by Transparency International Italia (TI), Italy is ranked on 52nd place over 180 global countries, on 20th place over 27 European countries, for a considerable presence of corruption in 2020[10]. Besides, as evidenced by the Global Corruption Barometer (GCB), that is one of the most detailed survey of people's views on corruption and experiences of bribery in the 27 EU countries, almost 34 percent (%) of interviewed people believe that corruption increased in the previous 12 months in Italy (2021)[11] against the 26 percent (%) of people who thought that corruption increased in the previous 12 months in Denmark (2021)[12] and 16 percent (%) in Finland (2021)[13]. In addition, GCB reports that almost 3 percent (%) of users paid a bribe and used personal connections to access public services in the last 12 months in Italy (2021) against the 1 percent (%) of users of public services in Denmark and Finland respectively (2021)[14].

Furthermore, corruption within Italian provinces varies and is higher for provinces that are in the southern part of the country. As Figure 8 reports, from 2010 to 2017, the annual average corruption is between 9 percent (%) and 60 percent (%) for the centre-southern provinces whereas it is below 9 percent (%) for the northern provinces[15].

[9] Data are available at: Personal remittances, received (% of GDP) - Italy | Data (worldbank.org)
[10] The list of countries ranked for Corruption Perception Index (CPI) for year 2020 is available at: Italy Transparency.org
[11] Data of the Global Corruption Barometer for Italy are available at: Results European Union Transparency.org
[12] Data of the Global Corruption Barometer for Denmark are available at: Results European Union Transparency.org
[13] Data of the Global Corruption Barometer for Finland are available at: Results European Union Transparency.org
[14] These reported data can be viewed at: Results - European Union - GCB - Transparency.org
[15] The map is realized by the author with STATA with data provided by RE.GE ISTAT upon specific request

__Figure 8__ Annual Average Corruption for Italian provinces 2010-2017 (source: RE.GE ISTAT)



Hence, descriptive statistics reveal that both phenomena in Italy are widespread within provinces and need to be monitored to reduce their negative sides. Although part of the hypotheses presented here have already been tested in the scanty literature on corruption and migration within provinces, especially the push effects of corruption, the purpose of this study involves filling the gap on conceptual and methodological aspects with the existent studies on these two phenomena. Thus, this thesis offers a comprehensive framework of analysis where traditional and novel methods are used and discussed in the following three main Chapters.

Chapter 1 provides a broad view of the literature that has analysed the relevant causes of human mobility. Specifically, the first part presents the recent strand of literature that has investigated the general causes of the brain drain and Italian brain drain whereas the second and the third parts focus on studies that have examined the effects of corruption on skilled mobility by using dynamic panel models and gravity models, respectively. Each part highlights the sources of novelty introduced by the present work and compare them with the existent international and national literatures. Besides, each part introduces the main research hypotheses that are analysed throughout Chapter 2 and 3. Also, Chapter 1 offers a detailed description of the novel measure of corruption adopted and describes and compares, with the existent studies, the procedures used to deal with endogeneity in the present study. Then, Chapter 2 introduces the first research question by exploiting unilateral data and the dynamic model of system of Generalized Method of Moments (system-GMM). Evidence suggests that high corruption positively influences net skilled migration from origin province, ceteris paribus. Although the model is robust because it is widely practiced among researchers and it controls for serial autocorrelation and endogeneity issues, its main limit is represented by the fact that it does not fully exploit the potentialities offered by bilateral data of a tri-panel dataset, losing great part of information. Thus, Chapter 3 deals with the trade-off between robustness and completeness handled by the traditional method presented in Chapter 2 and adopts a gravity framework with Zero Inflated Poisson (ZIP) and a novel Pseudo Poisson Maximum Likelihood (PPML) model. This part introduces more than one research question for studying the relationship between corruption and brain drain within origin and destination provinces. Results suggest the existence of not only push but also pull mechanisms of corruption at play on young skilled mobility, filling the gap with the international literature that supports the existence of the push effect of corruption only. Besides, another interesting result achieved is that sensitivity of the prospective tertiary students towards corruption varies according to their fields of study of interest. Also, evidence proves that corruption significantly affects long-distance movements, from the Centre-South to the North mainly. Although Chapter 3 enriches the analysis with additive research questions, bilateral data on enrolments at university, empirical results, endogeneity is likely to occur. Hence, Chapter 3 circumvents such limit and adopts a remedy, that is relatively recent in the literature of gravity models, that does not give up completeness for gaining robustness in the empirical analysis. Results provided by the robust model of Chapter 2 against the ones reported by complete but pioneering model of Chapter 3 are quite similar. Although the present dissertation yields more information on the relationship between these two phenomena, it innovates on the methodological side because it proves that the coexistence of robustness and completeness is possible in the empirical analysis if models are correctly implemented.

Furthermore, this dissertation offers interesting cues for adopting public policies necessary to stem and overthrow inequalities within provinces affected by corruption and net skilled migration mostly. Policy makers should carry-out law enforcements in places where corruption is high and invest more on education in places endowed by local university characterized by conspicuous amount of skilled mobility (Ciriaci 2013). In addition, they can implement a series of strategies to retain skilled flows from origin such as i) encouraging competitiveness of wages and productivity in private and public sectors and improving meritocracy to reduce red tape and cronyism (Dotti et al., 2013), ii) using diaspora engagement programs to enhance connectivity with skilled native emigrants in ways that maximize externalities, such as the transfer of ideas, knowledge, technologies to the origin (Gould, 2018), iii) encouraging skilled returns with fiscal benefits and policies aimed to reduce gender discrimination in education and employment (Khan et al., 2017).

In sum, fighting against corruption and young brain drain is not only an important short and medium-term policy concern but has even more relevant consequences in the long run, as it might have long-lasting effects on human capital accumulation and distribution of wealth that generate inequalities within territories (in that case, within Italian provinces).

Chapter 1

Review of Literature on Brain Drain and Corruption

Brief Introduction[16]

Chapter 1 offers a comprehensive overview of the empirical literature of the last ten years that has dealt with net skilled migration, with a special focus on the Italian case. A conspicuous part of the literature has systematically indicated the economic factors, such as income per capita and unemployment, and the socio-cultural factors, such as quality of university and quality of life, as main determinants of young skilled mobility but fewer part of it has already demonstrated corruption as one of the key factors at play. To this end, many empirical studies that use dynamic panel models and gravity models are reported and discussed to emphasise the elements of novelty introduced by the present study, such as completeness related data, research questions and the methods adopted. The present Chapter is organized as follows: Section 1.1 presents an overview of the literature that has treated the topic on skilled mobility. Then, Section 1.2 reports studies that describe the relationship between corruption and skilled migration with dynamic panel models in contrast to Section 1.3 that describes studies that, instead, use gravity models. Section 1.4 is dedicated to the description of the variable of corruption adopted by the present study and compares it with measures implemented by other studies. Then, Section 1.5 presents an overview of literature of dynamic panel and gravity models that deal with endogeneity and compares the methods exploited by the present study to solve the reverse causality issue. A brief bibliographic analysis of the literature is also performed with the implementation of data mining technique and is reported in Appendix A of the present thesis.

[16] This *Brief Introduction* has the mere role of providing a general description of the contents presented by Chapter 1 while the aim and the purpose of the whole thesis are pinpointed into the previous *Introduction* section.

1.1 Overview of Literature on Skilled Mobility

Human mobility is an interesting but complex phenomenon due to not uniquely identifiable cause. The first wave of international economic studies on brain drain dates back to the late 1960s and refers to welfare analyses with theoretical frameworks: in fact, these contributions argue that the impact of skilled mobility on source places is neutral and emphasizes the benefits of free migration to the world economy due to the fact that skilled emigrants not only leave some of their assets in their origins but also send remittances to compensate skilled losses suffered by native places (Docquier and Rapoport, 2012; Grubel and Scott, 1966; Johnson, 1967; Berry and Soligo, 1969). Then, the second wave arrives less than a decade later with Bhagwati (1970), who explores the main causes and (negative) consequences of brain drain for those left behind, by analysing (negative) externalities (such as domestic market rigidities, informational imperfections etc.) that concur to increase inequality across countries, where rich nations becomes richer at the expense of poor ones (Docquier and Rapoport, 2012; McCulloch and Yellen, 1977; Bhagwati 1974).

On the other hand, the third wave starts in the late 1990s and offers a more evidence-based view of brain drain's causes and consequences. It is proved that human movements were mainly determined by economic reasons and skilled individuals move from their origin places in search of higher economic benefits rather than higher amenities and leisure activities (Docquier and Rapoport, 2012). Nowadays, human flows occur not only to improve one's economic position but also to improve social position as the most recent literature demonstrates (Faggian et al., 2017). Several studies have involved researchers to analyse how the re-distribution of human capital stocks affects the growth trajectories (Faggian et al., 2017; Faggian and McCann, 2009). Great part of the research on migration have opened a wide debate whether this phenomenon should be studied according to the disequilibrium (Greenwood, 1975; Greenwood and Hunts, 1984) or to the equilibrium (Graves, 1984) models jointly. The first approach interprets migrations as mainly determined by economic factors while the second approach considers migrations mainly influenced by search for higher quality of life. This has opened the "job versus people" debate, whether people's location decisions are mainly determined by the search for job opportunities or by the search for better quality of life. From this issue, a conspicuous part of the international literature has developed the analysis on the main determinants and patterns of students' mobility with across countries and regional studies[17]. Specifically, economists care about brain drain because it has been and continues to be an area of tremendous policy concerns in many countries: economic and socio-cultural differences due to the unequal distribution of skills across territories may create a dualism condition that trigger the economic welfare of a country (Gibson and McKenzie, 2011). Many scholars devoted time to study skilled flow across countries whereas scanty and scarce literature with within-provinces observations is devoted to this analysis. Hence, this thesis provides a newer method for this study, that has been treated in the international literature, and paves the way for adopting a detailed approach that may reveal valuable insights that have been left out so far. The country chosen is Italy by dint of available and easy access/consultation of Italian provincial data for performing the within-provinces analysis.

As the Italian skilled mobility rises in recent times, research identifies whether this phenomenon is caused by socio-cultural or economic factors mainly. Three main streams of research can be identified. The first group of studies considers the individuals' characteristics, such as personal preferences and attitudes, as key elements of brain mobility. An example is offered by Nifo et al., 2011, who demonstrate how Italian skilled mobility is characterized by positive self-selection, due to personality traits, attitudes and socio-cultural background of students' families, that enhances the propensity of these 'best and brightest" to move away. This condition triggers harmful effects to their origin places, creating a vitious circle of human capital depletion that, in turn, decreases growth. Similarly, Ciriaci, 2013, analyses the main determinants of Italian interregional skilled mobility for pre- and post-graduate students. The results of this study suggest that the decision of pre-graduate students to move from their native places is mainly determined by their family's background; in fact,

[17] **Appendix A** reports a brief bibliographic analysis conducted with data mining approach with data derived by Web of Science and analyzed via VOS-viewer. It provides evidence of the development of recent studies conducted on Italian brain drain mainly.

well-educated and high-income families can better afford costs of maintaining students who live far from their homeplaces.

However, a second branch of this literature, respect to the previous studies, adds socio-cultural variables, such as search for better quality of university and/or quality of life, for explaining the likely occurrence of the students' mobility. Ciriaci, 2013 supports the thesis that quality of university has a strong influence over students' migration due to its pull effects for pre-graduates (students may decide to look for a better university) and signal effects to firms for after-graduates (students may also decide to live in the same region of the university where they graduated). This result is also in line with the study conducted by Bacci and Bertaccini, 2020, who suggest that the tendency to emigrate varies among Italian students because undergraduate students enrol to bachelor courses in search of better contextual conditions, whereas students who enrol from bachelor to master's degrees are mainly influenced by the characteristics that are intrinsic of the universities (e.g., quality of structures and presence of competitors). According to Dotti et al., 2013, universities are potential catalysts for brain gain and regional development. In fact, the attractiveness of provinces with universities is linked to the prospects of job vacancies for graduates. Students tend to remain in the place where they get graduation, especially if there are conspicuous opportunities to get a job and higher standards of life. Furthermore, Nifo and Vecchione, 2014, discover that better quality of local institutions, which define the level and quality of essential services such as health, legality, transports and culture encourages students to emigrate to places where local governments are efficient and guarantee safe and security.

However, a third branch of this literature includes studies that evaluate skilled mobility caused by economic factors, such as the presence of favourable labour market outcomes. Mayda, 2009, finds that higher income opportunities at destinations work as magnet for young students, who move from their origin places. Besides, Etzo, 2011, demonstrates that per capita GDP turns out to be a strong pull factor while unemployment rate is an important push factor for skilled individuals in the sending regions. Similarly, Biondo and Monteleone, 2012, evidence that productive job environments and better wages decreases the propensity of graduates to return to their native places, confirming the study conducted by Dotti et al., 2013 also. Besides, Biondo and Monteleone suggest that the Italian brain drain is a singular phenomenon because it lacks brain exchange and does not facilitate the creation of value coming from the brain gain. This is more evident for the southern rather than centre-northern area, pointing out the well-known socio-economic dualism between the two parts of Italy. Also, Piras, 2016 and Basile et al., 2018, report that brain drain is mainly caused by high unemployment rate and low GDP at origin.

An alternative analysis of the influencing factors on Italian skilled mobility is offered by Bonasia and Napolitano, 2012, who prove that not only economic but also specific-regional features, such as CO^2 emissions, juvenile crimes and housing prices positively affects skilled flows from their origin places. Similar in principle to Biondo and Monteleone, 2012, Iammarino and Marinelli, 2014, demonstrate that dynamic labour markets, where there is the likely match of education and job at the early stage of the graduates' career, influences the decisions of the students to move from places where, instead, there is a conspicuous mismatch between education and job usually. In addition, Michaeli et al., 2021, prove that skilled migration can exercise a remarkable negative effect on social capital accumulation because, given the heterogeneous level of human civicness at origin, it can determine civicness drain from origin to destination places, where law enforcements preserve civicness from uncivil free-riding condition, leading to a poverty trap within different areas.

However, none of these studies places it-self in an intermediate position in the 'job versus people' debate by considering both economical and socio-cultural factors at play, none of these studies has directly investigated the effects of corruption, intended as proxy for meritocracy, over the Italian skilled mobility in recent years. Thus, this dissertation wants to expand knowledge upon the scanty investigation of corruption and Italian brain drain to state an intermediate position in the 'job versus people' debate, by analysing not only the effects of corruption but also the effects of socio-cultural (i.e., quality of life and quality of university) and economic (i.e., GDP per capita, employments) variables. In doing so, the present thesis uses upgraded data on Italian students' flows within provinces and hires traditional and novel methodologies to compare results achieved. Hence, each chapter contributes to study the causal relationship between corruption and net skilled migration in quite dissimilar way because both analyses are grounded on different assumptions, hypotheses, and empirical models: Chapter 2 adopts variables that are in line with the traditional literature and uses a dynamic panel framework to test whether high corruption acts as push factor over skilled mobility from origin provinces. Although it uses unilateral data, the model of system-GMM is robust enough because returns results that do not suffer from endogeneity and serial autocorrelation problems. However, the model does not exploit the information provided by bilateral data as, instead, done by Chapter 3. Hence, Chapter 3 tends to adopt more complex framework of analysis, that is the gravity set-up, to test whether high corruption acts as push while low corruption acts as pull factor for incentivizing/attracting students from/to origin/destination places. To this end, it adds completeness to the results previously achieved and it uses models that belong to the Pseudo-Poisson family. However, to maintain both completeness and robustness, Chapter 3 performs a relatively novel procedure that tends to control the likely occurrence of bias of estimates. To this end, it acts as pioneer, amidst the gravity models, to face endogeneity and to defeat the challenging trade-off between robustness and completeness. Results achieved can be compared with the ones obtained by the traditional model of Chapter 2 and they can be evaluated for their similarity. In this manner, this study demonstrates that such trade-off can be overcome and both robustness and completeness features can coexist in an empirical analysis. This outcome represents an added value to the existent traditional literature, that, instead, tends to adopt empirical models which favour robustness than completeness for studying corruption and brain drain jointly. Besides, this thesis permits to discover that traditional models, such as the sys-GMM, that are generally used in international cross-countries analyses, can be used to study how phenomena are distributed and studied from a narrow perspective also.

1.2 Review Literature of Brain Drain and Corruption – Dynamic Panel Framework

Chapter 2 grounds its empirical analysis based on the panel data frameworks mainly used by Dimant et al., 2013, Ketterer and Rodriguez-Pose, 2015 and Cooray and Schneider, 2016, in their studies[18].

Dimant et al., 2013, in cross-country dataset, demonstrate that corruption is the main push factor of skilled and not-skilled migration and tends to diminish the returns to education. In doing so, they handle Pooled OLS and Fixed Effect (FE) to estimate data for 111 global countries from 1985 to 2000. They use the dependent variable of net skilled migration rate taken by Defoort, 2008[19] and the International Country Risk Guide (ICRG) index as measure for corruption[20]. However, this study presents one limit due to the way in which corruption is measured. In fact, ICGR data purport to measure the risk of political instability caused by corruption rather than the perceived corruption, which may be somewhat different. Thus, it is hard to believe that changes in political institutions would affect corruption and, in turn, the perception of it. It is likely to occur that changes in institutions would cause experts to expect changes in corruption, although they do not observe them directly. Thus, this relationship reveals something about experts' theories on corruption rather than the observations on it, and, for that reason, this index should be treated with caution (Treisman, 2007). Then, other limits of this study rely on the usage of outdated data and quite simple methodologies.

Ketterer and Rodríguez-Pose, 2015, in cross country panel dataset, discover that corruption negatively affects the attractiveness of migrants in local places. They adopt FE and the system-GMM to estimate data for 254 global countries between 1995 and 2009. They use the dependent variable of net migration flows and the Worldwide Governance Indicator (WGI) index as measure of corruption.

[18]Each study is performed within international contexts. In fact, Dimant et al., 2013 evaluate 111 global countries in a cross-country panel dataset. Ketterer and Rodriguez-Pose, 2015, study 254 global countries in a cross-country panel dataset. Cooray and Schneider, 2016, use the international Brucker et al., 2013 panel dataset.

[19]Defoort (2008) provides estimates of the rates of skilled and average migration to six main receiving countries of Australia, Canada, France, Germany, United Kingdom and United States. Here the net skilled migration rate refers to the ratio of the number of skilled emigrants (with post-secondary education) to the total number of skilled natives aged 25 years or older, while the average migration rate is defined as the ratio of the total number of emigrants to the total number of natives aged 25 years or older.

[20]The authors have rescaled the ICGR corruption index, so that higher values correspond to higher corruption levels.

Comparable to Dimant et al., 2013, one limit of this study is represented by the usage of WGI index as measure for corruption. In fact, Arpaza (2009) conveys three main critiques related the usage of this index. In fact, the first critique argues that such index encompasses six dimensions (Voice and Accountability, Political Stability, Government Effectiveness, Regulatory Quality, Rule of Law and Control for Corruption) created with the aggregation methodology that does not allow to distinguish amidst the already-cited different dimensions. Hence, it is presumed that "Control for Corruption" may not be able to measure corruption over time and across countries properly. Then, other critiques refer to the unreliability of information provided by the different data sources of WGI as well as the difficulty to get access to them. In addition, other critiques highlight that each component may be biased toward the perspective of elites and the level of development of a country (i.e., higher income countries get better scores because they are richer or are growing faster than other countries). In this way, it should be possible to achieve a misleading conclusion that richer countries present lower levels of corruption, that it is not true. Besides, another limit of this study is represented by the usage of recent data since the current data cover the period from 1995 to 2009 only. However, this work enriches its analysis by conducting the system-GMM as robustness check of its main results. In this way, potential biases arising from reverse causality problems are controlled fully.

On the other hand, Cooray and Schneider, 2016, in cross-country dataset, find that as corruption increases, the emigration rate of high-skilled individuals increases as well while the emigration rate of medium and low-skilled migrants, increases at initial levels of corruption and then decreases beyond a certain threshold. In conducting such analysis, they handle FE to derive the main results and they add robustness check analysis by using the system-GMM model with data of 20 OECDs from 1980 to 2010. Specifically, they evaluate data for 1995-2010 period of 115 origin countries taken by the dataset of Brucker et al., 2013[21]. They use the emigration rate of men and women over

[21]This dataset covers 20 OECD member states on the immigrant population aged 25 years and older by gender, educational levels and country of birth from 1980 to 2010 (in 5 years intervals). Data are available on the stock of immigrants coming from 195 countries. See Brucker et al. (2013) for more details.

25 years old selected per educational levels (high, medium and low skilled individuals) as dependent variable and the Corruption Perception Index (CPI) as measure of corruption. Although CPI is one of most used indices in empirical works, it presents one limit related to the data sources used for constructing it. It is not based on observations that consider the occurrence of corruption but on inferences made by respondents of surveys based on their conventional understanding on corruption (i.e., CPI considers the degree to which public perceive corruption among politicians and officials) and does not return a reliable measure of such phenomenon (Treisman, 2007). As Ketterer and Rodriguez-Pose, 2015, the authors use FE, system-GMM and the instrumental variable (IV) estimation to check the consistency of their results. However, another limitation of this analysis is represented by the usage of the measure of latitude as instrument of corruption. In fact, as argued by Stephenson, 2015, in a critical article on a study conducted by Mauro, 1995, latitude is not a valid instrument because it has direct effects on a country's economic performance, which may in turn affects corruption[22].

Although Chapter 2 uses the dynamic panel model of system-GMM, as already used by the above-cited studies, it adopts an original panel dataset of 880 total observations from 2010 to 2017 and selects a sample of 375 Italian source provinces, that have one local university, and tests if:

H_1: High corruption acts as push factor over young skilled mobility from origin province

Hence, as Dimant et al., 2013, Cooray and Schneider, 2016, Chapter 2 analyses whether high corruption at origin province, endowed with one local university, positively influences prospective tertiary students to enrol to university located far from their origin provinces, ceteris paribus. The novel elements that enrich this analysis consist of i) adopting a dependent variable that indicates the incidence of students who enrol to university located at 2 hours and/or 200 km from their source provinces respect to the total number of enrolled students who move or not from their origin

[22]An interesting critical article written by Stephenson, 2015, against the research on the validity of instruments for corruption (Mauro, 1997), can be found here: [Invalid Instrumental Variables in Corruption Research: A Lament | GAB | The Global Anticorruption Blog]

provinces, ii) using, as main independent variable, a more objective-based measure of corruption that considers the total corruption cases[23], iii) including, as independent variables, economic factors, such as the average population, real GDP per capita and employment, and quality-related factors, such as quality of life, quality of university, size of university and law deterrence that capture performance and resources availability at origin. Besides, dummies that refers to the presence of transport facilities, such as airports, ports and high-speed train stations, that ease the movements within provinces, are included. Details on data sources, variables and the model adopted with dynamic panel framework are reported in Chapter 2.

1.3 Reviewed Literature on Brain Drain and Corruption – Gravity Framework

Chapter 3 grounds its empirical analysis based on the literature of gravity models adopted by Mayda, 2009, Dotti et. al., 2013, Van Bouwel and Veugelers, 2013, Beine et al., 2014 and Poprawe, 2015[24]. Mayda, 2009, although studies the effects of average income and income dispersion on international (not-skilled) migration at origins and destinations, performs a cross-country analysis based on OLS with year effects, country effects and standard errors clustered by country-pair, plus a Tobit regression with country-pair fixed effects of panel data that cover information of 14 OECDs from 1985 to 1990. One limitation of this study relies on the strategy adopted to deal with endogeneity: although Mayda supports the fact that reverse causality is not a serious issue and she addresses endogeneity by relating the dependent variable of emigration rate to the lagged values of (log) per worker GDP at home and abroad, it seems unrealistic to claim that wages at home and abroad are strictly exogenous. Furthermore, Van Bouwel and Veugelers find that quality of higher education has a positive and significant effect on the size and direction of students' flows between 31 European countries. However, they demonstrate that the driving force for student mobility appears to be the lack of

[23]The variable of corruption used in Chapter 2 is the same as the one used in Chapter 3 and indicates the total number of corruption cases according to the Italian penal law (art. 312-322). Besides, Chapter 2 conducts its main analysis by also using corruption variable that indicates cases made by known authors while Chapter 3 uses this specification in the robustness analysis and permits to compare results obtained. For more details, see **Section 1.4** of **Chapter 1**.

[24] Chapter 3 poses quite dissimilar hypotheses derived from the study conducted in Chapter 2

educational opportunities in the home country. In doing this analysis, they use a cross-section of bilateral data of 31 countries for the year 2007 with a gravity model that accounts for country fixed effects. However, one drawback of this study is represented by the usage of a composite index for measuring the quality of countries' higher education. It uses indicators based on rankings of different universities that may be highly correlated with each other[25]. Besides, this type of variable returns a measure of research quality whereas students arguably care about teaching quality mainly.

Then, Beine et al., 2014, perform an original study that evaluates the choice of location of skilled and not-skilled individuals determined by two factors mainly: the costs of moving and the search of better quality of university. An interesting aspect related the costs of moving is the network effect from source to destination countries. In fact, the presence of compatriots at destination tends to act as a magnet for people. Interestingly, this effect is found to increase with the level of education of the network at destination. The higher the level of education of migrants in the host country, the higher the flows of students of the same nationality. Hence, people move to certain destinations because they can rely on people of their same nation too. Network effect is relevant because implies the reduction of the costs of moving between 40 percent (%) and 55 percent (%), as argued by Beine et al., 2014. In doing such analysis, they use a Poisson regression for panel data of 180 origin and 13 destination OECD countries from 2004 to 2007. Besides, to control for the correlated effect problem provided by the network, the authors combine Poisson with IV regression and hire "guest worker program" in the 60's and 70's as instrument for migration flows[26]. However, such instrument may be a poor predictor of the value of migrants' network after the advent of globalization and the evolved necessities of the subsequent modern migration flows.

[25] This study uses the *Shanghai Ranking* that considers the number of universities a country has in the top 200 of this ranking. Then, it uses the *Higher Education Supplemental Ranking* which considers quality of teaching based on peer review, recruiters' review, citations per academic staff, staff per students and proportion of international staff per students.

[26] Guest worker programs were implemented after the second world war in many industrialized countries to attract economic migrants for the explicit purposes of working in specific industries like coal mines or steel factories. They were mostly dropped at the beginning of the 70's. Those bilateral agreements led to the building of important diasporas in the destination countries before 2000 (Beine et al., 2014).

Dotti et al., 2013, demonstrate that Italian students move to study in different regions not only to attend higher-quality universities but also to enrol at universities located in provinces where they expect to find better job opportunities, once they graduate[27]. To this end, they use bilateral data of enrolments of Italian students at university, between 2003 and 2009, aggregated at Italian provincial levels. They use the Zero-Inflated Negative Binomial (ZINB) model, which is a modified version of Poisson's family models, that considers the highly skewed distribution of the dependent variable and the large number of zeros. Besides, they use one-year lagged dependent variable and one-year lagged explanatory variables to mitigate the risk of endogeneity. However, one limit is given by the variables adopted for indicating quality of university. Although they use alternative variables for quality of university, such as the presence of Rectorate, the average university fees and the average research quality at origin and destination[28], part of these variables should be refined because they are more quantitative than qualitative measures of performance: in fact, the presence of Rectorate, that denotes university with multiple campuses, is more suitable to indicate size than quality of university, while average fees do not capture students' performance but economic background of students' families (i.e., universities use the criterion of ISEE to evaluate the economic situation of students' families and to determine tuition fees for students)[29].

Another study worth to be mentioned is the one presented by Poprawe, 2015, who, in cross-section and cross-country analysis with bilateral data, finds that high level of corruption, at native places, encourages (not skilled) emigration to different destinations. The author uses, as dependent variable, a count value of migration flows taken by World Bank's Global Migration Database Ozden et al.,

[27] In doing this analysis, Dotti et al., (2013) have been able to study the actual attractiveness of Italian universities and relate it to provincial labour markets by constructing an attraction index for each Italian province based on university enrolment flows. For more details, see Dotti et al., 2013.

[28] These cited variables are taken by Il Sole 24 Ore and are re-elaborated by the authors.

[29] ISEE (Indicator of Equivalent Economic Situation) is a criterion to evaluate the economic situation of families. It gives the opportunity to take advantages of social benefits. It is used in largely in Italian universities for determining the students' fees.

2011[30] and CPI index as measure of corruption at origin and destination. In doing so, this study evaluates data of 230 world countries from 2000 to 2010 and adopts the Negative Binomial Model to derive its main results. Then, the author adds robustness checks made with the Extreme Bound Analysis (EBA) and Pseudo Poisson Maximum-Likelihood (PPML) with heteroscedasticity-robust standard errors à la Santos Silva and Tenreyro, 2010. However, Poprawe's work main limit is represented by the CPI index as measure of corruption, as already explained in Section 1.3. Hence, Chapter 3 begins its main analysis by formulating quite different hypotheses of Chapter 2, such as the following:

H_1: High corruption acts as push factor for young skilled mobility from origin provinces

H_2: Low corruption acts as pull factor for young skilled mobility to destination provinces

This work tests hypotheses that are mutually exclusive because the occurrence of the first event does not necessarily implies the occurrence of the second event. In fact, as already explained in the *Introduction,* the push condition of corruption is more verified than its pull side. In fact, the pull effect of corruption can occur if the decision of skilled individuals to move is based on perfect information provided by their family members and friends or collected by social media and newspapers, proving that the designed destination is characterized by lower corruption. Until now, the traditional literature on corruption and brain drain has evidenced the push effect of corruption only. Hence, this study tries to fill the gap with the literature to enlarge knowledge on the pull effect of corruption.

However, alternative hypotheses should be considered. On the one hand, students who do not migrate, from high corrupted origin, are those who are initiative-taking to pursue better quality of life and do not engage in corrupted activities, creating a virtuous circle that lowers the existent corruption. On the other hand, students who do not migrate, from high corrupted origin, are those who are probably less sensitive to corruption and engage in corrupted activities, creating a vicious circle that increases

[30] This dataset cover data for 230 world countries. For more details see: Ozden, C., Parsons, C. R., Schiff, M., Walmsley, T. L. (2011) "Where on earth is everybody? The evolution of global bilateral migration 1960–2000", *World Bank Economic Review*, 25(1), 12–56.

the existent corruption even further. However, even if the two hypotheses are not verified, these ones cannot be necessarily interpreted as evidence that corruption is not an important factor because the null outcome can be determined by the interplay of the two opposite effects derived by escaping from corruption at origin and fighting against corruption by remaining in the native place.

On the premises of these hypotheses, Chapter 3 investigates an aspect that has been poorly handled by scholars for the Italian case so far. In past times, prospective tertiary students of Medicine or Engineering, even due to the physical absence of certain courses in the university of origin province, tended to emigrate more than prospective tertiary students of Law or Economics. However, in last years the number of prospective tertiary students of Social and Physical Sciences who emigrate increases more than the number of students of Life Science, as described by Figure 4 in *Introduction*. Thus, Chapter 3 investigates whether corruption differently influences skilled mobility trends and tests the third original hypothesis as follows:

H_3. Sensibility to corruption varies among prospective tertiary Italian students who choose to enrol to courses belonging to different fields of study (Social Science, Physical Science and Life Science).

Thus, by comparing the magnitude of the estimated coefficients, we analyse whether prospective tertiary students of Law or Political Science are more sensitive to corruption and migrate more than prospective students of Biology or Medicine and/or differences of these types are not discernible. Results and plausibly explanations are reported in Section 3.3 of *Results* of Chapter 3.

Furthermore, poor part of the literature has examined the Italian skilled movements occurring between places located at different distances. For example, Michaeli et al. 2021, retain that long-distance skilled mobility from the South to the North of Italy is driven by the search of higher civicness due to stringent enforcements of civic behaviour in the North that makes migration more attractive for the southern skilled individuals. A tentative analysis is offered by Biagi et al., 2011, who examine the main determinants of long-and short-distances (not skilled) migration within Italy. Thus, Chapter 3 reproduces this framework and investigates if corruption maintains its effects over long-distance movements by testing the following fourth hypothesis:

H_4: Corruption maintains its push and pull effects when long-distance skilled mobility is considered

Biagi et al., 2011, find that economic factors have greater influence over long-distance flows, while quality of life related factors affect short-distance flows. However, the analysis is based on data of Italian migration for two years (2001-2002), that are outdated and may not still reflect today's causes of mobility, especially for net skilled migration. Besides, these data are not sufficient for concluding that short movements occur due to search of better quality of life instead of better economic conditions in the long run. Another limit of this study is represented by the fact that those who do not migrate are not considered. In addition, this study does not take explicitly account of regional or provincial fixed effects nor they cluster standards errors with Negative Binomial Regression model (NBREG).

Thus, Chapter 3 tries to overcome the limits of the empirical study conducted in Chapter 2, which uses a traditional model that is robust but does not exploit information of bilateral data fully, by introducing the novel Pseudo Poisson Maximum Likelihood with High Dimensional Fixed Effect (PPMLHDFE or PPML), with the gravity set-up, used by Santos Silva and Tenreyro, 2006.

PPML presents advantageous statistical properties that renders this model suitable for bilateral data because well-manages heterogeneity within observations. Besides, bilateral data permit to assess separately the sign and magnitude of each influencing variable, contrary to variables expressed in difference terms (Biagi et al., 2011). PPML well-handles the problem of excess of zeroes that determines some correlation between the covariates and the error terms, leading to the inconsistency of the estimates. In doing this, PPML minimizes the bias in Monte Carlo simulations under a set of alternative stochastic processes and uses adjusted standard errors with robust estimates. Besides, the estimations allow for the insertion of multiple fixed effects and the interaction term of fixed effects and time to control for the time trend.

Thus, Chapter 3 introduces elements of innovations that consist of i) the adoption of a count dependent variable that indicates the number of prospective tertiary students who decide to enrol to the university. This variable takes into account those students who enrol to their local university of their native province as well as those who, instead, decide to enrol to university located to different destinations, ii) same variable of corruption used in Chapter 3 that indicates the total corruption cases within Italian provinces[31], iii) the adoption of a newer procedure that deals with endogeneity in gravity model, by adding robustness to the results without giving-up completeness, iv) the usage of alternative variables for quality of university and quality of life respect to the ones used by the traditional literature. In contrast to Ciriaci, 2013 and Dotti et al., 2013, who use Sole 24 Ore and CENSIS as data sources, quality of life and university variables of Chapter 3 are created by using ALMALAUREA data and by taking the average of the standardized values that consider the factors that influence students' decisions to move. The detailed description of the data sources and methods used to derive these final variables are included in the *Additive Notes* of Appendix C.

1.4 Measure of Corruption[32]

Corruption is a complex phenomenon and its definition varies throughout the literature. Thus, measuring corruption on a single scale is difficult. Usually, scholars use different proxies depending on the dimension of interest (country, regional or provincial levels). Traditional literature adopts perception of corruption indices in the absence of more objective proxies, which is typically the case in cross-country surveys. The most used perception-based indices in international crime-based studies are the Corruption Perception Index (CPI), the Worldwide Governance Indicator (WGI) and the International Country Risk Guide (ICRG). Cooray and Schneider (2016), Poprawe (2015), in studying the effects of corruption on skilled flows, adopt the CPI to measure corruption in their studies. CPI is provided by Transparency International (TI) and measures how public users perceive the spread of corruption among politicians and officials. CPI ranks countries from 0 to 10, where 0 indicates that there is high level of perceived corruption while 10 indicates low levels of perceived corruption. To avoid the distorting effect on scoring that could be caused by external events, such as

---

[31] For more details, see **Section 1.5** of **Chapter 1**

[32] This section offers an overview of variables of corruption used by the current literature and compares them with the variable of corruption adopted by the present study. Hence, this part permits to better evaluate the pros and cons of such newer variable and its explanatory power in the empirical analyses conducted in Chapter 2 and 3 respectively.

the exposure to criminal scandals, the score combines the experts' assessments of the last two years only and does not return a more comprehensive measure of corruption in the medium and long-run.

On the other hand, Ketterer and Rodriguez-Pose (2015), in analysing the effect of quality of local government on migration, use WGI for measuring the level of efficiency of local government. WGI is provided by the World Bank and is an aggregated indicator which ranks countries from -2.5 - to indicate high level of perceived corruption - to 2.5 - to indicate low level of perceived corruption- and encompasses information on six dimensions of governance: Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law and Control Corruption. Its main limit consists of the fact that last dimension does not distinguish between "petty" and "grand" corruption and returns a too generic measure of it (Treisman, 2007).

Furthermore, Dimant et al. (2013), in studying the effect of corruption on skilled migration across countries, use the ICRG index as valid tool to measure corruption. ICRG is provided by the Political Risk Services Group (PRSG). It measures the extent to which public officials do illegal payments in return for import and export licenses, controls, tax assessment, policy protection or loans. It ranks countries from 0, indicating high presence of corruption, to 6, indicating low presence of corruption. However, such indicator returns an inaccurate measure of corruption because it tends to evaluate investment risk of corruption rather than corruption per se (Grundler and Potrafke, 2019). Another study conducted by Nifo and Vecchione (2014) uses an alternative index to measure corruption. Specifically, they take cue from the World Governance Indicator (WGI), proposed by Kaufmann et al. (2010), and create the Institutional Quality Index (IQI) that encompasses six elementary indices for Voice and Accountability, Government Effectiveness, Regulatory Quality, Rule of Law and Corruption. This index was created through three main phases of normalization, attribution of weights and aggregation practices[33]. This elementary index summarizes data on: crimes committed against the Italian public administration, the number of local administrations overruled by the federal

[33] For more details, see Nifo and Vecchione, 2014

authorities and the Golden–Picci Index, which measures the corruption level based on ‘the difference between the amounts of physically existing public infrastructure […] and the amount of money cumulatively allocated by government to create these public works’ (Golden and Picci, 2005; Nifo and Vecchione, 2014). The Golden and Picci Index is used to measure corruption in public sector and it is based on data taken by surveys of experts. However, this index does not capture observations of the frequency of corruption but inferences that heavily depend on the momentary public opinion and the media coverage of specific criminal cases. Hence, this IQI index may return distorted values on the evaluation of corruption per se.

Furthermore, Auer et al. (2020) implement a hybrid approach that combines country-level and individual perception measures of corruption to evaluate corruption as adverse effect that encourages people to emigrate. Thus, they adopt the Coppedge et al.’s (2018) V-Dem[34] eight (8) indices, which encompass measures for executive embezzlement, executive bribes, legislative and judicial corruption, with a range from 0, indicating low level of corruption, to 1, indicating high level of corruption, alongside the individual-based indices as proxy for indicating the direct exposure to corruption (having to bribe someone or witness incidents). Although country-level measures undeniably provide valuable and comparative information on the prevalence of corruption at the country level within the same country, there are differences between individuals with regard to the degree they encounter or perceive corruption (Auer et al., 2020; Treisman, 2007). Although this method permits to ensure that the perception of corruption is comparable across individuals that are in similar contexts, its main limitation can be represented by the fact that perception of corruption cannot be compared across individuals who are in different places.

However, this dissertation adopts a more objective value for corruption as is allowed by the usage of within-provinces analysis, contrary to the subjective-based measures that have not within-provinces observations, as the current literature proves.

[34] For more details, see Auer et al., 2020

This objective-based variable is used as proxy of meritocracy, that provides the idea of students' preferences to live in fair contexts where merit is fully awarded and the possibility to move upward in the society and achieve a respected position is likely to occur without the necessity to exploit any type of political connection, social grouping and/or cronyism. Specifically, the variable of corruption considers the number of crimes reported to prosecution departments resulting in criminal proceedings. It encompasses corruption crimes from 2010 to 2017 reported from the Annals of Criminal Statistics (Re.Ge) published by the National Institute of Statistics (ISTAT). This measure collects the number of crimes according to in the Italian penal code from art. 314 to art. 322bis, which cover bribery by and to public officials, judicial bribery, promised bribery and incitement to bribery as well as embezzlement, misuse of public funds, undue receipt of economic benefits and extortion by virtue of office. Part of these types of crimes are very often related to bribery. Besides, all reported crimes, grouped together in the official statistics, are those referred to a court. This avoids the distortions usually existing in the number of crimes reported simply to the police, which do not account of posterior dismissed charges[35]. However, this proxy presents some criticisms. First, it can be evaluated as a measure of crime detection due to the effort of prosecution departments to investigate and impose criminal charges and may determine the underestimation of such phenomenon even in the dynamic framework. Second, the number of detected crimes can be affected by the different quality of activities performed by different prosecution agencies (Treisman, 2007). However, the Italian prosecution system is homogenous in enforcing rules against corruption because it takes benefits from a centralized judiciary and the legal authorities tend to work independently from political groups. Then, as corruption increases, underreporting or reduced number of investigations can occur. This may be due to the lack of trust towards the judiciary or the time constraints facing the investigation that become more binding. However, this negative result can be mitigated by the spill-over effects of

[35] There is a conspicuous evidence of robustness of this proxy. For more details, see Lisciandra and Millemaci, 2017.

the investigative activity in crime reporting, maintaining this issue still debatable (Lisciandra and Millemaci, 2017).

Hitherto, a relative dearth of studies on crime have used more subjective than objective measures of corruption. Although the objective measure of corruption in this thesis presents more pros than cons in terms of reliability than the subjective indices do, an interesting study conducted by Lisciandra et al., 2021, proposes the usage of an original corruption risk indicator (CRI) at the procurement level[36] in Italy, obtained through a methodology that is replicable across time, space and proven robust to many validation tests[37]. In fact, CRI is used both as a risk-based assessment at the contract level and as reputational instrument to assess contracting authorities and track their performance, to conduct an evaluation of large funding programs and support the enactment of risk-based audits, by using scores that are obtained through a two-stage "residual approach" via non-parametric (Data Envelopment Analysis – DEA) and parametric analyses. CRI presents different advantages respect to alternative international measures of corruption based on crime statistics in public procurement predominantly[38]. In fact, CRI i) does not suffer neither from underreporting nor potential bias deriving from different investigation efforts or conviction rates across a given country, ii) overrides the subjectivity bias that is typical of perception measures and iii) its methodology is flexible and can be replicable in a broad range of institutional contexts. However, CRI cannot be considered a standard and direct measure of corruption risk rather "a measure of ignorance" being a residual with no guarantee that it captures corruption exclusively and its performance is highly dependent on the availability of data and the accurate identification in the second stage of all significant inefficiency-related factors.

Henceforth, the design of hybrid measures provided by advanced methodologies should be considered for building-up and obtaining a reliable measure of corruption in the future research agenda.

---

[36] Public procurements account for 12% of the GDP of OECDs which make governmental activity a fertile ground for corruption (Lisciandra et al., 2021).

[37] For more details, see Lisciandra et al., 2021.

[38] i) red flags (Fazekas and Kocsis, 2020)
ii) discrepancy of actual and estimated costs (Golden-Picci, 2005; Olken, 2009)

1.5 Review of methods to deal with Endogeneity[39]

This dissertation provides an outstanding contribution to deal with the endogeneity with the gravity set-up especially. In fact, while the literature of dynamic panel models offers solutions to deal with endogeneity frequently, this issue is seldom threated in gravity models and resolutive options requires quite demanding methodologies. Hence, this part offers a complete review of econometric solutions to deal with endogeneity and motivates the use of each methodology, that is reported in Chapter 2 and Chapter 3 to face this issue, respectively.

In the context of dynamic panel models, an example of endogeneity treatment is provided by the work of Dimant et al., 2013, who examine the likely existence of reverse causation between corruption and migration. They perform pooled and fixed-effects Instrumental Variable (IV) estimations with quality of judicial institutions and the degree of democratic participation as instruments of corruption. Results indicate that corruption is not endogenous to net skilled migration. Another example is offered by Fratesi and Percoco, 2014, who evaluate the link between regional disparities and migration flows, focusing on the skill content of migration. To detect the reverse causality between disparities and migration, they implement the IV-regression with instruments à la Card, 2001 and they find that interregional migrations are diverging forces in regional growth mainly. In addition, Ketterer and Rodriguez-Pose, 2015, use the system-GMM model to demonstrate that regional quality of government, such as the fight against corruption, is a potential source of attractiveness to skilled migrants. Hence, to address the potential endogeneity related institutional variables, they introduce explanatory variables with a 1-year lag and implemented the IV-regression. Also, Cooray and Schneider, 2016, adopt the system-GMM to study the effects of corruption on the emigration rate of low, medium, and high-skilled individuals. They find that corruption has a linear effect on the emigration rate of those with high skills and non-linear effect on those with medium and low skills.

[39] This sections present works that have dealt with endogeneity by using dynamic and gravity frameworks respectively. Hence, the aim of this section is to illustrate the general methods and to motivate the methodologies adopted by Chapters 2 and Chapter 3 to deal with endogeneity especially.

To ensure the causality direction from corruption to emigration, they use latitude and initial level of corruption as instruments of corruption in the IV-regression.

Furthermore, Li et al., 2017, examine how skilled labour emigration affects home country's institutional development and they adopt the system-GMM with two-years lags instruments to control endogenous variables. Besides, Nifo et al., 2020, study how heterogeneous agents, who face different migration costs, make different educational choices. Hence, in their dynamic panel framework, to ensure the causality direction from migration costs to educational choices, they use two-years lags of key regressors by employing a two Stage Least Square (2SLS) IV model.

Based on these works, Chapter 2 system-GMM exploits the lagged first differences in level equations as instruments to address the reverse causality between corruption and young net skilled migration.

On the other hand, in the context of gravity models, the most common approach used to mitigate the effect of endogeneity consists of introducing lagged variables in the regressions. In fact, Dotti et al., 2013, who studied how quality of universities and the local labour market conditions in the destination places affected students' mobility behaviour, use lagged explanatory variables to mitigate the risks of endogeneity. Same procedure was adopted by Pu et al., 2017, for studying the Chinese migration network system, where any change of regional characteristics influence outflows and inflows from and to the region itself and potentially its neighbouring regions. To avoid the problem of endogeneity, they puts five years lags on the dependent and explanatory variables within their four spatial autoregressive interaction models. Besides, Arpaia et al., 2018, in examining the bilateral trade in the European area affected by macro-economic variables such as population, real GDP per capita and unemployment, introduce the lagged exogeneous variables of population at origin, GDP per capita and unemployment for considering reverse causality. Furthermore, in the context of bilateral trade, Ghodsi, 2019, introduces lagged explanatory variables, one-year lag to reduce the endogeneity of trade. Besides, Anderson and Yotov, 2020, use two specifications: first, they create an instrument for lagged trade by using a restricted form of gravity model that only included the standard gravity variables. Then, they use the second to fifth lags of the newly constructed trade variable as instruments for the lagged dependent variable in the unrestricted gravity specification. Also, Gu et Shen, 2020, construct an eigenvector spatial filtering (ESF) hurdle gravity model (ESF-HGM) to examine the determinants of China's skilled internal migrations between 2010 and 2015. To redeem the possible endogeneity, they use not only the ESF-HGM model but also the lag of the independent variables.

However, another strategy frequently used to deal with endogeneity is the Instrumental Variable (IV) regression. An example is offered by Biagi et al., 2010, who analyse the long- and short-distance labour mobility in Italy. In doing so, they use a negative binomial model (NBREG), augmented by three instruments (football team of destination country, industry employment rate and presence of cash machines per 1,000 inhabitants) to control for potential endogeneity. In addition, Bergstrand et al., 2015, examine the effects of economic integration agreements, international borders and bilateral distance in international trade. To avoid biased results, they implemented panel pair fixed effects, that capture the cross-sectional negative impact of bilateral distance on trade flows. Besides, D'Ambrosio et al., 2018, study whether migrants promote co-inventorship between regions and foreign countries, and if the social capital of their respective communities favours such innovation networking. They addressed possible endogeneity by issuing more stringent fixed effects to their PPML. Finally, Zhang, 2020, investigates the role of migrants' taste in international trade and adopts two sets of estimations based on level and difference equations with two instruments. Specifically, he estimates the gravity equation in levels by two-stage least square method. Then, he further estimates the gravity equation by taking first differences between the two cross-sections to eliminate any potential time-invariant omitted variable. The results shows that the potential endogeneity leads to an overestimation of the taste bias but the effect is negligible because the differences remain within one standard deviation.

To sum up, Chapter 2 uses quite natural way to manage endogeneity because of the adoption of sys-GMM to deal with whereas Chapter 3 engages in a more demanding option that consists of exploiting a two-stage IV procedure à la Wooldridge, 2018 as already used by Drivas et al., 2020. For this newer procedure, a detailed analysis is offered in the subsection 3.3.1 *Robustness Check* of section 3.3 *Results* of Chapter 3.

Chapter 2

Unforgetful heads make a weary pair of heels– A Dynamic Panel Approach

Brief Introduction[40]

As young skilled mobility within Italian provinces has increased in last years, a conspicuous part of empirical studies has devoted its attention to the likely factors of influence. Many researchers have detected real GDP per capita and unemployment as key impacting variables, while others have identified quality of life and amenities-related factors as determinants of skilled mobility. To this end, Chapter 2 wants to investigate the role of corruption, used as proxy for meritocracy, as influencing factor over prospective tertiary students' decision to move or not from their origin province. In doing so, Chapter 2 uses a sample of resident enrolled students within 110 Italian provinces from 2010 to 2017 and adopts the dynamic panel model of one-step system of Generalized Method of Moments (system-GMM). Findings suggest that corruption acts as push variable that determines skilled mobility from origin.

Chapter 2 is structured as follows: Section 2.1 presents data used, Section 2.2 presents the model adopted, while Section 2.3 is divided in two Subsections, 2.3.1 and 2.3.2, that report the strategy used for the model choice and the empirical results get with the system-GMM respectively. Finally, Section 2.4 draws conclusions.

2.1 Data

The original panel dataset is fully balanced and collects data with origin specification (i) and consists of 880 total observations aggregated at provincial level, made of 110 Italian provinces[41] (N) for years 2010-2017 (T), for a total of 880 observations. Besides, this first empirical analysis is based on a

[40] This *Brief Introduction* has the mere role of providing a general description of the contents presented by Chapter 2 while the aim and the purpose of the whole thesis are pinpointed in the *Introduction* section.

[41] Since the analysis encompasses 54 universities for 8 years period, the observations are equal to 54 x 8 = 432 that, in turn, are 432-54= 378 final observations.

restricted sample of 378 observations because it takes into consideration young skilled students who decide to move from their origin province, with one local university[42], in year t. The initial number of origin provinces with one local university included in the same analysed is 51 plus 3 ancillary campuses of Como for Insubria University, of Ascoli Piceno for Politecnica delle Marche and of Forlì-Cesena for University of Bologna), for a total of 54 local universities at origin provinces.

The reason of the adoption of a restricted sample of provinces with local university relies on examining the voluntary skilled mobility not forced by the absence of local university but by other non-specified reasons. Hence, we do not take into consideration origin provinces that have no local university, otherwise skilled movements to different destinations become justifiable. Also, we exclude those students that move away from Italy to attend tertiary education abroad because it would be beyond the purpose of this study and requires different assumptions, variables and models.

The sources of each data collected in this panel dataset vary according to the variables of interest. The main source of the dependent variable of young net skilled migration is provided by the Italian Education, University and Research Department (MIUR), while the main data sources on corruption, quality of Italian universities and quality of life-related variables are the Italian Institute of Statistics with the Annals of Criminal Justice (RE.GE Istat), the Italian Centre for Investments and Social Studies (CENSIS) and Italian Institute of Statistics (Istat) respectively. Sources and descriptive statistics of each variable are reported in Table B.1 and Table B.2 of Appendix B respectively.

2.2 Empirical Model

The following part tests the main hypothesis already presented in Section 1.2 of Chapter 1. The econometric specification considered belongs to the family of panel data models and it can be expressed by Equation 1 as follows:

$$\boldsymbol{ln\, y_{i,t} = \alpha + \beta\, ln(X_{i,t}) + \delta_i + \eta_t + \varepsilon_{i,t}} \qquad (1)$$

[42] A detailed list of Italian provinces with university is reported by Table B4 of Appendix B

where ln y_{it} is the logarithmic ratio of those prospective tertiary resident students who move from origin province i respect to the total enrolled students at time t, ln $X_{i,t}$ is the logarithm vector that encompasses economic and quality-related variables that affect young skilled mobility, including corruption and the control variables, δ_i represents the time-invariant province-specific component, while η_t captures takes into account the relevant time effects and ε_{it} represents the *i.i.d.* error term. Besides, the adoption of the log-linear form permits an easier economic interpretation of the parameters estimated and allows to smooth-out the outliers and to obtain a bell-shaped distribution of regression residuals (Mayda, 2009).

The dependent variable of skilled mobility from origin, in the Equation (1), is represented by two different values which have not been used in the literature so far. These novel variables combine spatial values, such as distance and time, with the number of prospective tertiary students, who enrol at university, to model skilled flows from origin province. In doing so, the first dependent variable, that considers long-distance movements if two hours (120 minutes) of travel from origin are implied, is constructed by taking the logarithm of the ratio of the number of enrolled students, who travel two hours to arrive to the university located far from origin, over the total number of enrolled students, while the second dependent variable, that considers long-distance movements if two hundred kilometres (200 km) of travel from origin are considered, is created by taking the logarithm of the ratio of the number of enrolled students, who travel two hundred kilometres to university located far from origin, over the total number of enrolled students. The decision to use distance relies on the theory of recent migration studies that considers the physical distance between two locations as one main determinant of migration, which affects migration costs (C_i). In fact, if the distance between two places tends to increase, the related costs of transfers increase and may deter movements. (Mayda, 2009). However, far distances between two locations can be even travelled in few hours due to faster connections provided by the transportation facilities. Thus, time, which is used interchangeably, seems a preferable proxy of distance because does not under-evaluate the effects of transportation facilities that ease the burden of not being in proximity.

Then, the logarithm of $X_{i,t}$ vector includes economic, socio-cultural, and quality related variables of origin province. First, corruption, used as proxy for meritocracy, that refers to system or society in which people can reach positions of success, power and influence on the basis of their abilities and merits, refers to the number of crimes reported to prosecution departments[43]. This variable is novel in the current literature and differs from the perception of corruption indices already used in the empirical analyses conducted by Dimant et al., 2013, Ketterer and Rodriguez-Pose, 2015 and Cooray and Schneider, 2016[44].

Besides, Chapter 2, as well as Chapter 3, uses the variable of corruption that considers not only total corruption cases but also corruption cases made by known authors (*authors*). Hence, both variables are used interchangeably, by proving their consistency in terms of signs and statistical significances. Then, economic variables, such as the average population, real GDP per capita and employment are inserted to control for scale and wealth effects at origin province. In particular, the average population is used to indicate the capacity of origin province to host facilities and services (Cooray and Schneider, 2016). Besides, the real GDP captures not only the wealth effects of the origin province but also controls for the capacity of high-income families to bear higher educational costs of those students who decide to move away (Ciriaci, 2013). Also, employment indicates the opportunities present in the local labour market (Ciriaci, 2013)

Furthermore, the model includes variables that account for quality of university. Size of university is a categorical variable with values of 1, 2 and 3 indicating small, medium and large size university. This variable is used as quantitative measure for proxy of quality of university because indicates the hosting capacity of universities in terms of availability of services (such as libraries, study-rooms, gym and other recreative corners) and didactic courses, programs and careers' opportunities (Ciriaci, 2013) offered to students. Besides, the additive value used for indicating the quality of university is

[43] For more details on this variable, see the dedicated Section **1.5** of **Chapter 1**

[44] Dimant et al., (2013) use the ICRG, while Ketterer and Rodriguez-Pose use WGI (2015) and Cooray and Schneider (2016) adopt CPI index. For more details, see Section 1.2 of Chapter 1

a traditional average index that accounts the standardized values for scholarships, infrastructures, internationalization and digitalization offered by each university provided by CENSIS. Specifically, scholarship indicates the number of grants-in-aid to the most deserving students whose families are not able to sustain the costs of their studies. Infrastructure indicates whether the designed university have more than one campus and offer accommodation services for those students who are not resident in the city where the university is settled. Then, internationalization refers to the capacity of university to offer international courses to promote global learning among students worldwide. Besides, digitalization refers to the usage of digital assets that enhance not only the students' learning experiences but also smooth the organizational and bureaucratic processes of the entire system. Furthermore, the model adds a proxy for quality of life, that is hospital migration. Hospital migration denotes the quality of essential public services, such as healthcare, in the origin province: poor performance of the healthcare system encourages skilled flows to place where this service has a better performance. This consideration is in line with the literature, that states that skilled individuals move, on average, to places where living conditions are better and more satisfactory (Nifo and Vecchione, 2014; Dotti et al., 2013; Biagi et al., 2011; Mayda, 2009).

In addition, the model includes also dummies that indicate the presence of transport facilities, such as airport, port and high speed-speed railways at origin. This study contemplates the insertion of these variables because ease mobility between provinces. In addition, these transport facilities control for aspects related urbanization. As urban area become bigger, urban sprawl is likely to occur. Urban sprawl implies that population becomes dispersed over an increasing geographical area and it generates significant distances that, in turn, determine the necessity of infrastructure facilities to connect each other area.

Finally, law enforcement variables are used as control from corruption crimes. In fact, an efficient judicial system prevents actions determined by illegal behaviours, by enforcing controls and convicting authors of crimes, guaranteeing safety and security of all the individuals in the society.

Also, the effectiveness of law enforcement corrects for the potential misreported and underestimated dark numbers of corruption cases. Time dummies for 2010-2017 period are inserted as well.

The suitable dynamic model for the balanced panel dataset used in this study belongs to the family of the Generalized Methods of Moments (GMM) estimators. Also, it seems plausible that all explanatory variables used in the above presented empirical model are not strictly exogenous and potential endogeneity can happen. Specifically, the problem of reverse causality between net skilled migration and corruption is likely to occur (Cooray and Schneider, 2016). For the aforementioned reasons, this work uses the Arellano-Bover (1995) and Blundell and Bond (1998) on-step system GMM (sys-GMM) estimator that well fits panel dataset and controls for the joint endogeneity of the explanatory variables through the usage of internal instruments. The system-GMM has become increasingly popular in the fields of empirical research of labour economics, development economics, macroeconomics, industrial organizations and international economics (Bun and Sarafidis, 2013). This estimator is designed for situations with i) small T and large N panels, meaning few time periods and many individuals, ii) a liner functional relationship, iii) one left-hand-side variable that is dynamic, depending on its own past occurrences (in the present study, it is the variable designed to study young skilled mobility and its lagged terms), iv) independent variables that are not strictly exogenous, meaning they are correlated with past and possible current realizations of the error, v) fixed individual effects, vi) heteroskedasticity and autocorrelation within individuals. While the Arellano-Bond estimator, difference GMM (difference GMM or diff-GMM) starts by transforming all regressors by differencing (Hansen, 1982), the system-GMM makes an additional assumption that first differences of instrument variables are uncorrelated with the fixed effects. In fact, system-GMM is an efficient and robust approach for panel data estimation because it involves the adoption of lagged first differences as instruments for the equation in levels and lagged levels as instruments for the equation in first differences to use fully all moment conditions. The required assumption for instruments in system GMM states that the validity of the additional instruments depends on the condition that changes in the instrumenting variables are uncorrelated with the fixed effects. In fact, it is required that throughout the study period, individuals sampled are not to far from stready states, that deviations from long-run means are not systematically related to fixed effects (Roodman, 2009). Besides, system GMM carries out the Hansen test for over-identifying restrictions (Null Hypothesis H_0: the instruments are not correlated with the residuals) and the Arellano-Bond test for second-order correlation in the first differenced residuals to prove further its robustness. The reasons to adopt the one-step system GMM rely on two main considerations: first, the system GMM performs better than the difference GMM estimator because the instruments in the level model remain good predictors for the endogenous variables even when the series are highly persistent (Blundell and Bond, 1998); second, the one-step GMM is favourite to the two-step GMM because simulation studies, such as the Monte Carlo simulation (e.g., Bond and Windmeijer, 2005), have found very modest efficiency gains with the two-step approach and bias increasing with the number of overidentifying restrictions in presence of small samples (Hwang & Sun, 2018). Finally, robust standard errors permit to control for heteroskedastic errors.

2.3 Results

2.3.1 *Model Choice*

First, this analysis performs estimations by using Pooled OLS and the static panel models of Random Effects (RE) and Fixed Effects (FE) to address the proper choice to adopt a more suitable model.

The estimated sample of observations is restricted and consists of 378 origin provinces endowed of one local university at least. Besides, many variables used in the model are in logarithmic form, such as the dependent variables that denotes young net skilled migration, corruption, the average population and the real GDP per capita, because it makes easier their economic interpretation when they are expressed in ratio or monetary terms. Exceptions are represented by the employment rate, size of university, which is a categorical value, quality of university, quality of life and dummies for transport facilities, which are already rescaled measures and time dummies.

Table 1 reports results obtained with Pooled Ordinary Least Squares (OLS), Fixed Effect (FE) and Random Effect (RE), by using the logarithm of skilled movements that considers time (in minutes) as benchmark for measuring the physical distance between two provinces, while Table 2 reports results obtained with the already cited models, by using the logarithm of skilled movements that considers distance (in kilometres) as benchmark for measuring the physical distance between two provinces. Besides, both Table 1 and Table 2 present, from column I to column III, the variable that account total corruption cases, while, from column IV to VI, corruption considers the number of corruption cases made by known authors (*authors*). On average, all models exhibit the expected positive sign for corruption while its statistical significance it is seldom reached. Also, average population preserve its expected negative sign and it is statistically significant barely. Besides, GDP per capita and employments maintain their positive and negative signs respectively but are statistically significant rarely. Even size of university, although it has the expected negative sign, is not statistically meaningful. Contrary to size of university, quality of university is statistically significant but does not display the expected sign and provides puzzling interpretation. Besides, the proxy for quality of life has unstable signs and shows statistical significance barely.

Finally, all models include the dummies of presence of transportation facilities of airport, TAV and ports. Also, law enforcement is added to control for the potential dark number of corruption. Even though the results below present similarities in signs and significances under each different specification, it is possible to select one of them by means of standard statistical tests for panel data. In fact, pooled OLS is different from the RE and FE because it imposes intercepts to be the same for all provinces. Consequently, the Pooled-OLS estimator is not consistent when there is fixed unobserved heterogeneity across observations. Besides, RE requires cross-sectional effects not to be correlated with the independent variables while FE does not.

Table 2.1 *Pooled OLS, RE and FE with log. brain drain (t=120 minutes)*

Y = log of brain drain (t=120)	**Pooled OLS** I	**RE** II	**FE** II	**Pooled OLS** IV	**RE** V	**FE** VI
log of corruption	.2482**	.09231	.06721			
	{0.103}	{0.077}	{0.077}			
log of corruption (authors)				.1683*	.06046	.04802
				{0.088}	{0.060}	{0.059}
log of population	-.1271	-.2912	.6184	-.1567	-.3099	.4894
	{0.318}	{0.315}	{1.994}	{0.312}	{0.318}	{1.968}
log of real GDP per capita	.2623	.6762	.7027	.2169	.6399	.5796
	{1.463}	{1.050}	{1.840}	{1.490}	{1.055}	{1.810}
employment	-.009416	-.03716	-.07855	-.007054	-.03648	-.07785
	{0.033}	{0.034}	{0.047}	{0.033}	{0.034}	{0.047}
unisize 1	-.85***	-.9528***		-.8325**	-.9481***	
	{0.313}	{0.272}		{0.314}	{0.272}	
unisize 2	-.5379**	-.5165**		-.5282**	-.5131**	
	{0.215}	{0.239}		{0.218}	{0.241}	
quality of university	.277***	.1097	.0201	.2733***	.1047	.01713
	{0.099}	{0.067}	{0.069}	{0.101}	{0.065}	{0.066}
hospital migration	.2492**	.1265	-.3682	.2465**	.1217	-.3832*
	{0.120}	{0.117}	{0.227}	{0.122}	{0.118}	{0.226}
dummies for airport, port and TAV	YES	YES	YES	YES	YES	YES
law enforcement	YES	YES	YES	YES	YES	YES
N	378	378	378	378	378	378

Notes: With all methods time dummies are included and standard errors reported in parenthesis are heteroskedasticity-robust. For the case of the pooled-OLS estimation, standard errors also account for potential clustering of provinces observations. ***, ** and * denote coefficients are significant at 1%, 5% and 10%, respectively. Results are in **b/se***

Table 2.2 *Pooled OLS, RE and Fe with log. brain drain (d=200)*

Y = log of brain drain (d=200)	**Pooled OLS** I	**RE** II	**FE** III	**Pooled OLS** IV	**RE** V	**FE** VI
log of corruption	.07864	-.03979	-.0782			
	{0.116}	{0.108}	{0.112}			
log of corruption (authors)				.07985	-.01307	-.03946
				{0.094}	{0.085}	{0.086}
log of population	-.6232*	-.5921*	-.6182	-.6319*	-.5859*	-.4161
	{0.358}	{0.343}	{3.401}	{0.355}	{0.346}	{3.395}
log of real GDP per capita	.6403	.2972	-.4722	.6115	.3088	-.3056
	{1.557}	{1.180}	{3.521}	{1.571}	{1.197}	{3.534}
employment	-.03545	-.0149	-.01161	-.03498	-.01575	-.01319
	{0.039}	{0.035}	{0.048}	{0.040}	{0.035}	{0.048}
unisize1	-.2088	-.2374		-.2021	-.2392	
	{0.415}	{0.400}		{0.414}	{0.401}	
unisize2	-.0346	-.06183		-.02815	-.06193	
	{0.334}	{0.364}		{0.334}	{0.364}	
quality of university	.4079***	.2272**	.07709	.4083***	.2314**	.08208
	{0.123}	{0.102}	{0.107}	{0.123}	{0.101}	{0.109}
hospital migration	.2476	.1829	-.2355	.2418	.1828	-.2204
	{0.149}	{0.156}	{0.364}	{0.148}	{0.156}	{0.368}
Dummies for airport, port and high-speed train station	YES	YES	YES	YES	YES	YES
law enforcement	YES	YES	YES	YES	YES	YES
N	378	378	378	378	378	378

Notes: With all methods time dummies are included and standard errors reported in parenthesis are heteroskedasticity-robust. For the case of the pooled-OLS estimation, standard errors also account for potential clustering of provinces observations. ***, ** and * denote coefficients are significant at 1%, 5% and 10%, respectively. Results are in **b/se***

Furthermore, the models adopted so far can be mis-specified because they assume a static setting, contrary to the dynamic set-up, which account serial correlation in the idiosyncratic error term of the linear panel data model. To evaluate whether the dynamic set-up can be more consistent than the static framework, Breusch-Pagan Lagrange multiplier test, Hausman test and Wooldridge test are performed. The first test provides guidelines to choose between pooled-OLS and RE, while the second and the third test permit to choose between RE and FE and between static or dynamic setting respectively. Table 3 reports the results obtained through test the following hypotheses:

Table 2.3 *Panel Data Choice Test*

	Stat	**P-Value**	**H_0**
Breusch-Pagan Test	407.52	0.000	**Rejected**
H_0: $\sigma_\delta=0$			
Hausman Test	19.89	0.069	**Rejected**
H_0: $Cov\ (\delta, X) =0$			
Wooldridge Test	2.927	0.089	**Rejected**

Breusch and Pagan test rejects the null hypothesis (H_0) of zero cross sectional variance, confirming that RE is more preferred than Pooled OLS. Then, the Hausman test rejects the null hypothesis of zero significance difference between RE and FE, thus FE is more preferred to RE. Finally, the Wooldridge test rejects the null hypothesis of absence of serial correlation in the idiosyncratic errors of the linear panel data model and indicates that the dynamic set-up is most suitable than static models. To this end, the following subsection 2.3.2 will present the results obtained with the adoption of the dynamic system-GMM model.

2.3.2 *Empirical Results with system GMM*

Results presented in Table 4 uses as main independent variable the logarithm of corruption while Table 5 uses the logarithm of corruption (authors). The dynamic set-up is confirmed to be appropriate because of significant autoregressive parameter and the rejection of the null hypothesis of absence of serial autocorrelation in the Arellano-Bond (1) test. Besides, the serial autocorrelation is of order 1 since the Arellano-Bond (2) test cannot reject the null of absence of serial correlation of order 2 or higher. The lagged dependent variable of brain drain with time=120 and/or distance=200 specifications are GMM-style variables that are instrumented with their lagged terms. However, adding higher moment restrictions and allowing for the lagged first differences to be used in level equations as instruments, system-GMM may have caused the overfitting condition of the endogenous variables due to the instrument proliferation. All estimates obtained meet the requirements that the overall number of instruments (# *instruments*) are not higher than the number of provinces (# *groups*). Moreover, Hansen test of the overidentifying restrictions indicates that instruments are valid since are uncorrelated with the error term and other instruments are omitted from estimations correctly.

Columns I-IV in Table 4 use the logarithm of brain drain with the threshold of travel time (t=120 in minutes) as dependent variable, meanwhile, columns V-VIII use the logarithm of brain drain with the threshold of travel distance (d=200 in km). The coefficient of the lagged dependent variable is statistically significant and the Arellano-Bond (1) test rejects the null of serial correlation. The significance of the lagged dependent variable for brain drain with threshold of time travel (t=120) proves the influence of network. In fact, the past occurrence of skilled flows is relevant for current skilled flows to tackle decisions to move from their origin to destination: in fact, network tends to diminish the non-monetary costs, such as being homesick, of moving from native places (Beine et al., 2014). Then, the logarithm of corruption exhibits the expected positive sign and it is statistically significance. Hence, high corruption acts as main push factor over skilled mobility from source provinces, ceteris paribus. This result is in line with past studies conducted by Cooray and Schneider, 2016, Poprawe, 2015 and Dimant et al., 2013, who demonstrate that corruption influence skilled and average migration from source places. As expected, prospective tertiary students, as movers from their origins, escape from places where merit is not fully recognized and the up-ward social mobility or the access to the job market (especially for high-profile and prestigious job positions) is likely to occur through family ties and/or political affiliations.

In addition, the logarithm of average population shows-off negative sign and it is not statistically significance. The logarithm of average population is used as proxy for indicating the capacity of cities to offer services: as city gets bigger, skilled mobility is reduced because larger cities become more attracting places for emigrants. Besides, real GDP per capita has the expected positive sign but is not statistically meaningful. This variable is used to control for wealth effects, and it denotes the capacity of higher income families to sustain the educational costs of students who decide to leave home (Ciriaci, 2013). Also, employment exhibits the expected negative sign but it is not statistically significant. Intuitively, higher opportunities to get a job in origin place refrain skilled individuals to move away (Piras, 2016; Fratesi and Percoco, 2014; Docquier et al., 2012; Mayda, 2009).

Furthermore, young students prefer to remain/move to origin/destination where the designed university provides better structures and higher quality of educational programs. For example, size of university has the expected negative sign but it is not statistically significant. This fact demonstrates that high availability of spaces such as classrooms, study-rooms, cafeteria or entertainment corners and libraries refrain young student to move away.

Moreover, quality of higher education variable is inserted. However, this coefficient presents positive sign and it is statistically significant. Although this result deserves deeper investigation, tentative interpretations are given. Positive sign may reflect the fact that traditional courses, with several entry barriers, such as admission tests or compulsory attendances with specific requirements (for example, highest grade of high school diploma, impressive knowledge of foreign languages, IT expertise, good mathematical and logical knowledge) force students to move to different universities with lower difficult entry requirements. In addition, skilled students prefer to attend course programs in university which have strong job-placements programs and give the opportunities to get job in the labour market where the university is located (Ciriaci, 2013; Dotti et al., 2013). Obviously, these conjectures should be analysed with more detailed data and therefore should be taken with caution.

Besides, hospital migration has the expected positive sign and it is statistically significant. Lower quality of the healthcare determines search for better quality of medical assistance elsewhere.

Furthermore, law enforcement is added on columns IV and VIII of Table 4. Law enforcement shows-off the expected negative sign but it is not statistically significant. The introduction of these variables is useful to control the underestimated number of crimes that are neither detected nor punished by the legal authorities. Besides, dummies of transportation facilities are added on columns III and VIII of Table 4. Hence, cities with airport, high-speed train stations and port prevent from isolation, and, on average, skilled individuals prefer to move to places where transport facilities permit faster connections and easier return to home, even for few periods.

Results presented on Table 5 consider the effect of corruption (authors) over skilled mobility. Also, the dependent variables used vary interchangeably: the logarithm of skilled movements with the threshold of travel time (t=120), from columns IX to XII, and the logarithm of skilled movements with the threshold of travel distance (d=200) from columns XIII to XVI are included. Again, the lagged dependent variables for both dependent variables are statistically significant and the Arellano-Bond (1) test rejects the null of serial correlation.

Corruption (authors) preserves its positive sign and its statistical significance for all specifications, confirming the result presented in Table 4 previously. Besides, all aforesaid independent variables maintain, on average, the same signs and statical significances, demonstrating the stability of results of above-mentioned model. Also, law enforcement variables are added on columns XII and XVI and dummies of infrastructure facilities are inserted on columns XI and XV of Table 5.

Table 2.4 *Main results of System-GMM with corruption*

	I	II	III	IV	V	VI	VII	VIII
lagged log of brain drain (t=120)	.2935**	.2645**	.2518**	.2392**				
	{0.113}	{0.117}	{0.116}	{0.117}				
lagged log of brain drain (d=200)					.2537	.2341	.2106	.2327
					{0.247}	{0.269}	{0.270}	{0.266}
log of corruption	.2314**	.2012*	.1967*	.2074**	.2362**	.204**	.1817	.2083**
	{0.102}	{0.102}	{0.103}	{0.103}	{0.100}	{0.100}	{0.109}	{0.101}
log of population	-.2054	-.06529	-.03042	-.1486	-.3	-.1631	-.2302	-.2493
	{0.203}	{0.161}	{0.204}	{0.172}	{0.326}	{0.272}	{0.301}	{0.255}
log of real GDP per capita	.4957	.01834	-.3266	.3169	.7159	.2625	-.06432	.4471
	{0.871}	{0.933}	{0.906}	{0.934}	{0.951}	{0.935}	{0.924}	{0.931}
employment	-.03653	-.02554	-.01458	-.03656	-.04809	-.03682	-.02706	-.04801
	{0.029}	{0.027}	{0.027}	{0.026}	{0.032}	{0.031}	{0.032}	{0.031}
unisize1	-.2065	-.00916	-.02914	-.0574	-.05354	.1527	.08447	.1612
	{0.257}	{0.254}	{0.257}	{0.258}	{0.368}	{0.345}	{0.336}	{0.347}
unisize2	-.1068	-.03105	-.1283	-.009419	.05997	.1414	.06375	.1997
	{0.188}	{0.186}	{0.208}	{0.200}	{0.195}	{0.191}	{0.221}	{0.199}
quality university	.1808**	.1775**	.1863**	.1767**	.2547**	.2512**	.266**	.2553**
	{0.085}	{0.087}	{0.089}	{0.088}	{0.118}	{0.120}	{0.122}	{0.120}
hospital migration		.2746**	.3293**	.2716**		.2731*	.3121*	.2638*
		{0.115}	{0.126}	{0.113}		{0.140}	{0.160}	{0.135}
dummies for airport, TAV & port	NO	NO	YES	NO	NO	NO	YES	NO
law enforcement	NO	NO	NO	YES	NO	NO	NO	YES
N	378	378	378	378	378	378	378	378
A-B test (1)	0.002	0.003	0.003	0.002	0.053	0.075	0.081	0.074
A-B test (2)	0.194	0.198	0.183	0.220	0.269	0.252	0.252	0.234
Hansen (p-value)	0.352	0.236	0.235	0.225	0.296	0.114	0.096	0.078
# instruments	33	33	37	39	28	28	31	31
# groups	54	54	54	54	54	54	54	54

Note**:** The lagged dependent variables, *log. brain drain d=200* and *log. brain drain d=200 (t-1)* are treated as predetermined variables (GMM-style option in xtabond2), while the other independent variables are treated as exogenous (IV-style option of command xtabond2). A maximum of six lags are used as instruments for the GMM-style endogenous variable. Time dummies are included in all specifications. Standard errors reported in parenthesis are heteroskedasticity-robust. ***, ** and * denote coefficients are significant at 1%, 5% and 10%. Thresholds of distance and time are expressed in km and minutes, respectively. Results are in **b/se***

Table 2.5 *Main results of System-GMM with corruption (authors)*

	IX	X	XI	XII	XIII	XIV	XV	XVI
lagged log of brain drain (t=120)	.3149***	.2667**	.2667**	.2576**				
	{0.117}	{0.120}	{0.120}	{0.121}				
lagged log of brain drain (d=200)					.3166***	.2999***	.2921***	.3008***
					{0.103}	{0.106}	{0.109}	{0.106}
log of corruption (authors)	.1546**	.1215*	.1215*	.137*	.1703**	.1461*	.1279	.1543*
	{0.071}	{0.071}	{0.071}	{0.074}	{0.082}	{0.079}	{0.078}	{0.081}
log of population	-.2262	-.06145	-.06145	-.1796	-.2848	-.1609	-.2113	-.2529
	{0.194}	{0.195}	{0.195}	{0.166}	{0.237}	{0.201}	{0.214}	{0.179}
log of real GDP per capita	.425	-.34	-.34	.2877	.5696	.2149	-.07023	.4032
	{0.840}	{0.894}	{0.894}	{0.916}	{0.846}	{0.851}	{0.845}	{0.853}
employment	-.03524	-.01407	-.01407	-.03612	-.04343*	-.03253	-.02329	-.04412*
	{0.028}	{0.026}	{0.026}	{0.026}	{0.025}	{0.025}	{0.025}	{0.024}
unisize1	-.1965	-.0258	-.0258	-.05533	-.05215	.1375	.07404	.1363
	{0.244}	{0.249}	{0.249}	{0.247}	{0.336}	{0.296}	{0.298}	{0.296}
unisize2	-.0948	-.1196	-.1196	-.003393	.05918	.1315	.05547	.1813
	{0.182}	{0.206}	{0.206}	{0.196}	{0.168}	{0.162}	{0.194}	{0.170}
quality university	.1658*	.1752*	.1752*	.1669*	.2251**	.2261**	.2354**	.2318***
	{0.084}	{0.089}	{0.089}	{0.087}	{0.089}	{0.088}	{0.088}	{0.084}
hospital migration		.3256**	.3256**	.266**		.252**	.2846**	.2405**
		{0.126}	{0.126}	{0.112}		{0.106}	{0.123}	{0.100}
dummies for airport, TAV & port	NO	YES	YES	NO	NO	NO	YES	NO
law enforcement	NO	NO	NO	YES	NO	NO	NO	YES
N	378	378	378	378	378	378	378	378
A-B test (1)	0.002	0.003	0.003	0.002	0.000	0.000	0.001	0.001
A-B test (2)	0.180	0.173	0.173	0.208	0.166	0.145	0.146	0.134
Hansen (p-value)	0.296	0.225	0.225	0.192	0.275	0.167	0.174	0.142
# instruments	33	37	37	39	30	30	33	33
# groups	54	54	54	54	54	54	54	54

Note**:** The lagged dependent variables, *log. brain drain d=200* and *log. brain drain d=200 (t-1)* are treated as predetermined variables (GMM-style option in xtabond2), while the other independent variables are treated as exogenous (IV-style option of command xtabond2). A maximum of six lags are used as instruments for the GMM-style endogenous variable. Time dummies are included in all specifications. Standard errors reported in parenthesis are heteroskedasticity-robust. ***, ** and * denote coefficients are significant at 1%, 5% and 10%. Thresholds of distance and time are expressed in km and minutes, respectively. Results are in **b/se***

2.3.3 *Additional Check*

The first robustness check is related the usage of simplified regression in which controls variables are excluded to manage the potential problem of endogeneity. Hence, the newer regressions avoid the insertions of the logarithm of GDP per capita, employment, the logarithm of population and university size variables. The lagged dependent variables use from 1 up to a maximum of 6 lags with the collapse option to avoid, in relatively small sample, the biases that arises as the number of instruments climbs toward the number of observations. Results are showed in Table B.5 (that uses as main independent variable the logarithm of corruption) and Table B.6 (that uses as main independent variable the logarithm of corruption cases for known authors) of *Appendix B.* Signs and statistical significances for the coefficients inserted are preserved. Specifically, the logarithm of corruption (and corruption made by known authors) exhibits the expected positive sign and its statistical significance for all regression reported from column I to VII of Table B.5 and from column IX to XVI of B.6, equals to the results demonstrated by Table 4 and 5 above. However, the p-value of Hansen test of over-identifying restrictions, with exceptions, results higher than 0.1, indicating valid instruments.

Then, a second analysis is performed with the backward orthogonal deviations in order to avoid any potential correlation between the instruments used and the error terms. Essentially, the instruments are replaced with their deviations from past means. Since the resulting instruments depend on all past values of the underlying variables, the regressors in the transformed equation should not be similarly transformed, otherwise the instruments may be correlated with the error. Results are exhibited in Tables B.7 and B.8 of *Appendix B.* The estimated coefficients present similar signs and statistical significances with the estimated coefficients of Tables 4 and 5 above. Also, the logarithm of corruption (and corruption made by known authors) has positive sign and it is strongly significant. Besides, the p-value of the Hansen Test of over-identifying restrictions reported on both Tables of *Appendix B* is higher than 0.1, demonstrating that the instruments are still valid.

2.4 Conclusion

Chapter 2 investigates why young Italian brain drain occurs frequently within Italian origin provinces in recent times. One spotted factor of influence is corruption, used as proxy for meritocracy that indicates the degree to which the advancement in the society is based on the individual's merits and capabilities rather than favouritism and other types of unfair activities in their source contexts. Thus, we examine the relationship between corruption and brain drain with the dynamic panel model set-up and we found evidence that, as corruption at origin increases, students 'mobility from origin increases as well. Hence, highly corrupted environment at origin is detrimental for its human capital accumulation because, if continuous, its effects lead to worse economic outcomes in the long run. This result is in line with the existent literature that uses dynamic panel framework with the system-GMM as empirical model (Dimant et al., 2013; Cooray and Schneider, 2016). The source of novelty of this study consists of adopting updated data on prospective tertiary students' enrolments at Italian university and performing a within-country analysis on Italian net skilled migration which is relative scanty in the literature. However, one limit of this work relies on the exploitation of unilateral data for origin provinces without consider bilateral movements. Although the empirical strategy adopted here is robust, it lacks completeness because does not exploit information fully as, instead, Chapter 3 does. Thus, Chapter 2 provides guidance to assess the validity of results achieved by Chapter 3, that, instead, exploits a novel model that offers completeness through its ability to manage bilateral data of a tri-panel dataset.

In order to reduce the negative consequences arising from high corruption and persistent brain drain from source provinces, policy remedies can be suggested: for example, education policies are needed to incentive higher education at origin whereas labour policies should encourage jobs creation to fill the mismatch between demand of highly skilled and labour supply at origin; besides, social policies are needed to enhance connectivity with native movers in order to promote transfer of idea, knowledge, technologies and trade. Besides, governmental policies are essential for enforcing laws and assuring order and security. Thus, policymakers should be acquainted with influencing factors that determine skilled flows and they should implement supportive policies for places that suffer from severe brain drain to limit the harmful effects due to waste of the "bests and brightest".

Chapter 3

Better the devil you do not know than the devil you know when you move – A Gravity Approach[45]

Brief Introduction[46]

Chapter 3 investigates whether corruption can play a dual role as push and pull factor over Italian skilled mobility. To this end, Chapter 3 introduces a gravity model with the Zero-Inflated Poisson and the novel Pseudo-Poisson Maximum Likelihood with High Dimensional Fixed Effects with bilateral data aggregated at Italian provincial level from 2010 to 2017. Findings suggest the existence of push and pull mechanisms at play, as high corruption at source province incentivizes net skilled migration to destinations that, instead, exhibit lower corruption. Moreover, sensitivity of the prospective tertiary students to corruption varies according to their fields of study of interest. Besides, corruption and brain drain plausibly explain the existence of dualism condition of the economically advanced North and the more lagging South of Italy. Chapter 3 presents the following structure: Section 3.1 presents data, Section 3.2 introduces the model, Section 3.3 reports the results and presents two Subsections that are, Subsection 3.3.1, which presents additive robustness check analyses and Subsection 3.3.2, which deals with the endogeneity issue. Finally, Section 3.4 provides conclusions.

3.1 Data

Italian brain drain was studied by constructing an original tri-panel dataset that is balanced and consists of 20.808 total observations, made by the combinations of 51 Italian origin provinces (*i*), where there is one university at least (we consider the larger one) and 51 destinations provinces (*j*),

---

[45] The present work has been already presented at 62nd edition of RSA Conference organized by the Italian Economics Society (S.I.E) on October 29th, 2021, in an online parallel session of "Italian Economy". Still, the same work is going to be presented at SIDE Italian Society of Law and Economics (ISLE) on December 16th, 2021, on online parallel session "Corruption" at University of Trento.

[46] This *Brief Introduction* has the mere role of providing a general description of the contents presented by Chapter 3 while the aim and the purpose of the whole thesis are pinpointed in the *Introduction*.

analysed from 2010 to 2017 (T)[47]. Our dependent variable, the number of prospective tertiary students, has a considerable number of zeros (0s), about 16.332 over 20.808, and it is not comparable with cross-section and cross-country studies.

The reason to adopt a restricted sample of provinces with one university relies on examining the voluntary skilled mobility not forced by the absence of local university but by other non-specified reasons[48]. For this reason, we do not take into consideration origin provinces that have no local university, otherwise skilled movements from these latter places become justifiable. Also, we exclude those students that move away from Italy to attend tertiary education abroad because it would be beyond the purpose of this study and requires different methodology to be implemented.

The data source used for brain flow is the Italian Education, University and Research Department (MIUR), while data on provincial corruption cases, data on quality of Italian universities and data on provincial economic features are taken from different sources such as the Italian Centre for Investments and Social Studies (CENSIS), the Italian Ministry of Education (MIUR), the Italian University Group of ALMALAUREA and National Institute of Statistics (ISTAT), respectively. Data sources and descriptive statistics of data are provided by Table 1 and Table 2 of Appendix C respectively.

3.2 Empirical Model

The following part test the main hypotheses already presented in Section 1.3 of Chapter 1. Our study wants to reproduce the bilateral net skilled migration within provinces with the gravity set-up[49]. In analogy with the Newton's law of gravity resident student flows can be predicted according to the following formula[50]:

[47] A detailed list of Italian provinces for origin and destination is illustrated by **Table C.3** of **Appendix C**
[48] A detailed list of Italian university for origin and destination is illustrated by **Table C.4** of **Appendix C**
[49] A theoretical framework that encompasses the characteristics of gravity models is illustrated in *Additive Notes* of **Table 12** of **Appendix C**
[50] Also, Dotti et al., 2013, report the same gravity model with similar specification. For more details, check Dotti et al., 2013

$$I_{ij} = K\ \frac{M_i^{\beta_1} M_j^{\beta_2}}{d_{ij}^{\beta_3}}$$

[1]

Where I_{ij} represents the interaction term of the number of students of origin province *i* who enrol to university located in province *i* or *j*, K is a proportionality constant, M_i is the mass of origin province, M_j is the mass of destination province, d_{ij} indicates the physical distance between the two provinces. In addition, β1 is the coefficient that estimates flows from mass origin *i*, β2 flows that are attracted to mass destination *j* and β3 is an impedance factor reflecting distance decay between masses *i* and *j* (Dotti et al., 2013). Hence, rewriting the above conditions according to the gravity model specification, we estimate the econometric model with the designed variables in the following form:

$$Enrolled_{ij} = f\,(\text{time}\ pop_{ij}\ corruption_{ij}\ rgdppc_{ij}\ employment_{ij}\ unisize_{ij}\ quniversity_{ij}\ qlife_{ij}\ infrastructures_{ij}\ law_{ij})$$

[2]

Equation (2) states that young students' decisions to move will be made upon evaluation of economic and quality-based factors for origin and destination provinces. The dependent variable is a non-negative count variable and represents the number of prospective tertiary students who enrol at university. Besides, Chapter 3 inserts additional count dependent variables, denoted with ERC-1, ERC-2 and ERC-3, for indicating prospective tertiary students who enrol to courses of Social Sciences (ERC-1), Physical Sciences (ERC-2) and Life Sciences (ERC-3)[51]. All independent variables are two-fold featured for origin and destination provinces. The main independent variable is corruption, used as proxy of meritocracy and is the same that is used in Chapter 2, refers to the total number of corruption cases against Italian Public Administration (PA) and punished according to rules dictated by the Italian Penal Law[52]. In doing robustness analysis, Chapter 3 uses the variable of corruption cases committed by known authors (authors). In this way, it is possible to draw considerations on the validity of the sign and statistical significance of corruption when robust and complete models are used in Chapter 2 and Chapter 3 respectively.

[51] A detailed list that describes each graduate course per study field is reported by **Table C.5** of **Appendix C**

[52] For more detailed information on corruption, see **Section 1.5** of **Chapter 1**

Differently from Chapter 2, Chapter 3 adopts the gravity framework and inserts its designed gravity variables. In fact, Chapter 3 adds time, expressed in minutes, as proxy of Euclidean distance (km) that indicates the time necessary to travel from source to destination provinces by car. Besides, time captures general migration costs that a skilled person should sustain when he/she moves to places far from their origin (Dotti et al., 2013). Besides, gravity model uses the average population, as mass variable, which describes the capacity of origin and destinations to draw flows. High average population indicates the presence of noticeable urban agglomerates that attract several people because urban cities offer more services, facilities and leisure activities than rural cities do (Beine et al., 2014). Then, Chapter 3 adds economic values of real GDP per capita and employment per origin and destinations. Specifically, employment controls for safety and welfare status while real GDP per capita does not only capture wealth effects but also the capacity of high-income families to bear students' costs of studying far from their home (Ciriaci, 2013). Besides, as done by Chapter 2, Chapter 3 uses size of the university as quantitative measure of quality of university. Larger-size universities are preferable because they offer more students' services (libraries, study-rooms cafeteria, gym) and didactic, extra-didactic and international programs (inclusive of careers' opportunities network) than smaller-sized (Ciriaci, 2013). In addition, quality of university is added and is quite dissimilar to the one used by Chapter 2. This newer variable evaluates the practical aspects of quality of university and is the average of the standardized age, grade, expected income per capita, expected time to find a job when students get graduation from their three-years course program at least[53]. All these factors adapt to the students' expectations because, in evaluating their university choice, students arguably care on the 'working prospects' formation of skills required to find a valuable job tomorrow.

As done by Chapter 2, Chapter 3 inserts quality of life variable. It is given by the average of the standardized values of mortality rate, working formation, gender difference in employment, the presence of green urban areas, childcare and eldercare[54]. All indicators are related personal health

[53] For details related this variable, see **Table C.13** of **Appendix C**

[54] Data source is ISTAT

and social well-being, that, in turn, affect the standards of living of individuals[55]. Also, the present work uses dummy variables for the presence of airport, port, and high-speed railway stations at origin and destination. These infrastructure facilities are crucial to favour mobility within places. Finally, variables of law enforcement are used as a control for the underestimated number of crimes unpunished. In fact, law enforcements indicates the efficiency of judicial system to detect and punish illegal behaviours, by convicting criminals to guarantee safety and security of citizens in the society. This study empirically represents Equation (2) by adopting the Zero Inflated Poisson (ZIP) and the Pseudo-Poisson Maximum Likelihood with multiple High Dimensional Fixed Effects (PPPMLHDFE or PPPML) proposed by Long (1997) and Santos Silva and Tenreyro (2010), respectively. These models belong to the General Poisson family models. General Poisson belongs to the Generalized Linear Model (GLM) class of popular non-linear regression models based on the exponential family of distributions. Hardin and Hilbe (2018), present the exponential family as given by:

$$f_y(y;\theta,\varphi\,) \;=\; exp \quad \left\{\frac{y\theta-b(\theta)}{a(\varphi)}\;c(y,\varphi)\right\} \qquad [3]$$

Where a (·), b (·), and c (·) are specific functions and $\theta \; and \; \varphi$ are parameters. For these models,

$$E(y) \; = \mu \; = b'(\theta) \qquad [4]$$

and

$$V(y) = b''(\theta)a(\varphi) \qquad [5]$$

Given a set of n independent observations, each indexed by i, we can relate the expected value to a set of covariates (x_i) by means of a link function g (·). More specifically, it is assumed that

$$E(y_i) = \mu_i = g^{-1}(x_i\beta) \qquad [6]$$

and in the case of Poisson regression, Equation [6] is re-stated as

$$E(y_i) = \mu_i = exp(x_i\beta) \qquad [7]$$

and the likelihood for the GLM is written as

[55] For details related this variable, see **Table C.13** of **Appendix C**

$$L(\theta, \varphi; y_1, y_2 ..., y_n) = \prod_{i=1}^{n} \quad exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)} + c(y_i, \varphi)\right\} \qquad [8]$$

Estimates for β are obtained by solving the first-order conditions for maximization of the likelihood,

$$\beta^{(r)} = (X'W^{(r-1)}X)^{-1} X'W^{(r-1)}z^{(r-1)} \qquad [9]$$

where **X** is the matrix of explanatory variables, $W^{(r-1)}$ is a weighting matrix, $z^{(r-1)}$ is a transformation of the dependent variable while r is an index for iteration (Hardin and Hilbe, 2018).

Chapter 3 faces the challenging trade-off between robustness and completeness already performed by the empirical analysis of Chapter 2. Although the dynamic system-GMM provides robust results, it suffers from completeness because it does not exploit bilateral data. Thus, a study which uses a panel dataset with three dimensions (origin, destination and time) aggregated at provincial level, may result penalised if robust but simplex models are adopted. For that reason, Chapter 3 adopts these two models, one of which, PPML, is a cutting-edge model for the current gravity literature known so far. ZIP allows us to take into consideration the highly skewed distribution of our dependent variable and allows for overdispersion assuming that there are two different types of observations in the data: i) those who have a zero count with a probability of 1 (0 group), and ii) those who have counts predicted by the standard Poisson (not 0 group). Observed zero could be from either group and if the zero is from the 0 group, it indicates that the observation is free from the probability of having a positive outcome (Long, 1997).

In addition, we handle PPML and then, we compare the results obtained with ZIP. Presenting both these models appears useful as robustness check of their validity. Besides, there are several advantages related the adoption of PPML as: i) relaxing assumption of knowledge of distribution of the non-negative dependent variable, ii) providing more natural way to deal with great number of zeros of the dependent variable as ZIP, iii) dealing better with sources of heterogeneity within larger panel-type dataset instead to resort to log-linear regressions (Correia et al., 2019), iv) allowing flexibility with multiple fixed effects and interactions (Fally, 2015), v) "`ppmlhdfe`" STATA command provided by Correia et al. (2020), allows less-time consuming estimation of parameters of interest even in presence of multiple fixed effects, by dropping problematic observations to avoid multicollinearity (Santos Silva and Tenreyro, 2010). In fact, recent article of Correia, Guimaraes, and Zylkin (2020) discusses the necessary and sufficient conditions for the existence of estimates in a wide class of GLM models and show that, in the case of Poisson regression, it is always possible to find MLE estimates if some observations are dropped from the sample. Hence, they promote the PPML as valid tool that can easily detect and discard separated observations that do not convey relevant information for the estimation process. Thus, PPML is a promising procedure because proceeds fast in nonlinear estimations with high-dimensional covariates (Correia et al., 2019). Hence, the following section reports results obtained from the estimation of the econometric specification reported in Equation (2) with ZIP and PPML. The estimations are carried-out with STATA software, by implementing the command ZIP (Greene, 2012; Long, 1997) and the user-written command PPMLHDFE (Correia et al., 2020).

3.3 Results

The main results of ZIP and PPML are reported in the following Table 3. In accordance with the gravity framework, such models use time, in place of distance, as more accurate measure of distance within provinces, and the average population as mass variables of source and destination provinces. Besides, *Notes* report the additional variables included in the regression but not displayed in Table 3. It includes law deterrence, used as a control variable for the stability of judicial enforcement at origin *i* and destination j. Besides, both models contain the fixed effects of the Centre and the South and years. The decision to insert these fixed effects grouped for macro-area permits to detect common and relevant fixed effects that cannot be revealed if, instead, provinces are used. To address the likely correlation between the error term over time for a given province, cluster-robust standard errors are used to check the statistical significance of the parameters.

Column (I) uses as dependent variable, the number of prospective tertiary students, column (II) uses instead the number of prospective tertiary students who enrol to courses of Social Science (ERC-1), column (III) of Physical Science (ERC-2) and column (IV) of Life Science (ERC-3). Even columns V to VIII display estimates performed via PPML with the same count dependent variables. Both models present similar results for signs especially. However, PPML present estimates that have more stability because present higher statistical significances.

Variables of time and the average population of origin and destinations are statistically significant and have the expected signs: time exhibits negative sign, meanwhile, the average population of source and destinations has positive sign. Signs and magnitude of these variables are consistent with the predictions of gravity model. In fact, greater mass means that the city is large enough to offer services and entertainment activities that attract skilled people to join rather than smaller cities, while higher time necessary to travel between far-distanced provinces, does not incentivize emigration due monetary and non-monetary costs of leaving home, contrary to Biagi et al. (2011) who, instead, use the variable of distance for indicating both the physical distance and general costs to move away.

Then, corruption exhibits the expected positive sign for the origin and negative sign for the destination provinces. Besides, it is strongly significant for both specifications, meaning that higher corruption at origin province positively influences, on average, young skilled mobility from origin, while lower corruption at destination, attract skilled flows to destinations, ceteris paribus.

Although the result of the push effect of corruption is in line with the result achieved in Chapter 2 and with the studies of Dimant et al. (2013), Dotti et al., (2013), Nifo and Vecchione, (2014), Cooray and Schneider (2016), that demonstrate that an increase in corruption at origin tends to increase skilled flows from origin, the result of the pull effect of corruption is rarely considered by the recent literature. Hence, this work wants to contribute to fill the gap with this lacking part that has not focused on the pull side of attraction of corruption over skilled flows mainly.

Furthermore, this study attempts to understand if the sensitivity to corruption is homogeneous or not among prospective tertiary students: results suggest that students, who decide to enrol to courses of Social Sciences (ERC-1) and Physical Sciences (ERC-2) are more sensitive to corruption at origin (which is statistical significant at 5% and 10% confidence level on columns II and III, significant at 5% and 2% on columns VI and VII) than students who enrol to courses of Life Science (ERC-3) and for whom corruption is not a significant influencing factor (columns IV and VIII) for deciding to rest or move away. To understand the intensity of the phenomenon under examination, we interpret the magnitude of the coefficients with their standard deviations. A unit increase of standard deviation in the corruption levels of origin results in an average increase of skilled mobility between 7.8 and 10 percentage points, between 5 and 6.8 percentage points for students of Social Science, between 8 and 16 percentage points for students of Physical Science and between 2 and 4.9 percentage points for students of Life Science. This novel result can be explained in two ways: first, students who enrol to courses of Social Science are, on average, more lawful oriented than those who enrol to courses of Life Sciences. Hence, students of Social Sciences are more sensitive to meritocracy concerns and they are more prone to condemn the misuse of legal powers. Then, a second explanation relies on the fact that sensitivity to corruption reflects the existence of heterogeneous conditions of the job market for these three categories. In fact, in Italy, the employment rate for those who get a graduation in the fields of Social Science and Physical Science is lower than the employment rate for those who get graduation in the field of Life Science.[56] Hence, in the first scenario, where the job mismatch between the labour demand and supply is very pronounced, the competition among participants, after they graduated, to get a job arises and, in turn, events of corruption, such as bribery and cronyism, are likely to occur. In such scenario, skilled individuals tend to be more sensible to corruption, rather than skilled persons who are going to work in the second scenario. This motivation seems to be more important and prevail over the consideration that, in Italy, those who get graduation in Medicine are more likely to work in public sector, where events of corruption occur usually, than those who get graduation in Economics, Engineering and Law, who are likely to work in the private sector[57].

Then, real GDP per capita has the expected positive sign for origin and negative for destination. In fact, positive sign represents high-income families who can afford high educational costs for their young students once they decide to enrol to universities placed far-away from their homeplace. This result is in line to Ciriaci and Palma, 2008 and Ciriaci, 2013.

However, for ZIP, real GDP per capita is not statistically significant neither for origin nor for destination. Also, system-GMM results present alternate positive sign for real GDP per capita but it is not statistically significant. For PPML it is statistically meaningful at origin (5% and 10% confidence levels on columns V VI and VII) and at destination (10% confidence on column VII).

Besides, for ZIP and PPML, employment rate reveals the expected negative sign at origin and positive sign for destination. It is statistically meaningful for origin, while is statistically significant for destination for PPML only (on columns V, VI and VII). Also, results get by system-GMM confirm the negative sign and significance of employment at origin. Intuitively, students make their migration

[56] According to Almalurea, in 2017, the average employment rate for those who get graduation after 1 year in studies of Social Sciences (ERC-1) is 49.7% while it is 57.2% in studies of Physical Science (ERC-2) against 69.5% in studies of Life Science (ERC-3). Computations are made by own elaboration based on data available at: [Condizione occupazionale dei laureati (almalaurea.it)](#).

[57] According to Almalurea, in 2017, the average of graduates in Medicine (ERC-3) who work in the public sector is 25.6% against 14.0% of those who are graduated in Law (ERC-1), 5.7% in Economics (ERC-1) and 5.1% in Engineering (ERC-2). Data are available at: [Condizione occupazionale dei laureati (almalaurea.it)](#).

choices by comparing job opportunities offered by native and destination labour markets and choose the one that offers higher and better chances of employment (Dotti et al., 2013).

Then, size of university has the expected negative sign for origin and positive sign for destination. Size of university is statistically significant for origin and destination for PPML (1% and 10% confidence level on columns V, VI, VII and VIII). Also, results get by system-GMM confirm the negative sign and significance of size of university at origin. Thus, if the size of university of origin is large, the number of resident students who move decreases because they prefer to attend universities that offer many didactic and job-partnered courses near their home. Besides, the increase of enrolments permits to large universities to become even larger because they acquire additional resources, depleting them from small universities, that, in turn, become even smaller (Ciriaci, 2013). Then, we added another variable for education that is the average of the standardized values for quality of university. As expected, the sign of quality of university is negative for origin and positive for destination provinces. Quality of university is statistically significant for origin (1% confidence level on columns V, VI and VIII) and seldom significant for destination (5% and 10% on columns VII and VIII). This suggests that high quality of university at origin tends to incentivize skilled individuals to remain to their local university if the educational quality of its didactic and extra-didactic courses is satisfactory. This result is in line with the main findings achieved by Ciriaci (2013), who recognizes the role of university as a key driver of economic development, through its role in knowledge, production and human capital accumulation, and an attraction pole for talents. Then, we add the variable of quality of life. The sign of this variable is negative for origin and positive for destination, with small exceptions. Although it is seldom significant (on columns III and VII) for origin and destination, higher quality of life at origin is associated, on average, with lower skilled mobility to different places. In fact, students prefer to live in communities with efficient services for their safety and security (Beine et al, 2014; Nifo and Vecchione, 2014; Dotti et al., 2013). Both quality of university and quality of life variables are newer respect to the ones used in Chapter 2 because the traditional variables present missing data related provinces of destinations. For that reason, signs and significances achieved here cannot be compared with quality of university and quality of life variables used in Chapter 2.

Furthermore, dummies for airport, port and high-speed railway stations are added to control whether their presence ease the transfers within provinces. Airport displays the expected positive sign for origin. It is seldom significant, but its presence allows us to control for the possibility to travel in fewer hours than needed using cars. High-speed railway and ports present similar patterns. Port is an essential logistic infrastructure needed to connect islands to mainland, preventing from isolation. In addition, it facilitates transport by car and by train and does not discourage the discontinuous mobility. In sum, ZIP and PPML maintain, for all variables inserted, same signs but sometimes different statistical significances of the estimates. Estimates performed with PPML are, on average, more statistically meaningful that the ones executed with ZIP. Besides, PPML and ZIP present estimates that have, on average, same signs and statistical significance as the ones provided, with some exceptions, by the dynamic model of system-GMM in Chapter 2.

For sake of completion, Chapter 3 performs several additive checks to assess the robustness of the methodologies adopted so far, without giving-up completeness.

__Table 3.3__ Main Results with Zero Inflated Poisson and Pseudo-Poisson Maximum Likelihood

	ZIP	ZIP	ZIP	ZIP	PPML	PPML	PPML	PPML
	I	II	III	IV	V	VI	VII	VIII
	Enrolled	ERC-1	ERC-2	ERC-3	Enrolled	ERC-1	ERC-2	ERC-3
time	-.02225***	-.02284***	-.01729***	-.0151***	-.02673***	-.02846***	-.02593***	-.02423***
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
population	2.58e-07**	3.46e-07***	3.46e-07***	7.54e-08	3.03e-07***	3.77e-07***	2.65e-07**	1.79e-07
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
population_j	3.64e-07***	2.52e-07***	2.95e-07**	5.01e-07***	2.83e-07***	1.88e-07**	3.47e-07***	4.10e-07***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
corruption	.001469**	.0009438*	.001603**	.0005472	.001951***	.001288**	.003353***	.00101
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
corruption_j	-.00311***	-.002182***	-.003786***	-.002017***	-.003373***	-.00231***	-.0055***	-.002443***
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
real GDP per capita	.0000138	.0000127	-7.66e-06	4.90e-06	.0000214**	.0000196*	.0000225*	.0000205
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
real GDP per capita_j	-8.19e-06	3.11e-06	1.10e-06	-.0000114	-.0000104	3.82e-06	-.0000232*	-.0000254
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
employment	-.0469***	-.03719***	-.03909***	-.03518**	-.05837***	-.05432***	-.05377***	-.07205***
	{0.013}	{0.013}	{0.014}	{0.014}	{0.013}	{0.013}	{0.015}	{0.014}
employment_j	.02083	.006165	.009782	.01051	.03291**	.02658*	.02968	.0511***
	{0.014}	{0.015}	{0.016}	{0.015}	{0.014}	{0.015}	{0.018}	{0.016}
size university	-.1436*	-.1179	-.05357	.002883	-.2933***	-.2512***	-.3805***	-.2128*
	{0.086}	{0.084}	{0.105}	{0.108}	{0.081}	{0.081}	{0.099}	{0.115}
size university_j	.3751***	.2002**	.5945***	.2171*	.5308***	.3506***	.8872***	.458***
	{0.095}	{0.091}	{0.124}	{0.121}	{0.091}	{0.088}	{0.112}	{0.137}
std. quality university	-.3936**	-.3704**	-.263	-.5743***	-.4515***	-.5002***	-.256	-.6431***
	{0.164}	{0.164}	{0.214}	{0.192}	{0.157}	{0.168}	{0.201}	{0.234}
std. quality university_j	-.04433	-.01318	-.4454*	.267	.09124	.2046	-.4224*	.5262**
	{0.173}	{0.172}	{0.256}	{0.217}	{0.154}	{0.168}	{0.216}	{0.219}
std. quality life	-.421	-.329	-.6991**	-.2478	-.3295	-.1113	-.7565**	-.2235
	{0.279}	{0.267}	{0.305}	{0.357}	{0.277}	{0.262}	{0.310}	{0.416}
std. quality life_j	.2607	-.01667	.5656**	.1987	.2077	-.241	.8157***	.4393
	{0.240}	{0.262}	{0.272}	{0.311}	{0.235}	{0.258}	{0.265}	{0.341}
Dairport	.045	.03806	.04933	.1911	.04778	.05861	.04842	.03189
	{0.116}	{0.107}	{0.136}	{0.153}	{0.113}	{0.114}	{0.135}	{0.158}
Dairport_j	.1044	.1167	-.008307	.3003*	.08623	.05885	-.05277	.3503*
	{0.131}	{0.119}	{0.166}	{0.154}	{0.130}	{0.126}	{0.157}	{0.182}
DTAV	-.0616	-.1039	.04328	.1975	-.2519	-.2952*	-.3534	-.03559
	{0.192}	{0.185}	{0.232}	{0.200}	{0.172}	{0.162}	{0.220}	{0.198}
DTAV_j	.1386	.2066	-.01834	-.2797	.4581***	.5152***	.5659***	.1758
	{0.180}	{0.174}	{0.232}	{0.232}	{0.151}	{0.146}	{0.187}	{0.195}
Dport	.283**	.2893**	.2047	.1712	.3582***	.403***	.2298	.4739***
	{0.124}	{0.140}	{0.154}	{0.160}	{0.121}	{0.138}	{0.143}	{0.161}
Dport_j	-.04648	-.09002	.1137	-.118	-.08843	-.173	.2069	-.325
	{0.147}	{0.161}	{0.194}	{0.193}	{0.151}	{0.166}	{0.190}	{0.206}
N.	20808	20808	20808	20808	20808	20808	20808	20808
Wald Chi Square	545.044	597.845	367.115	202.422	364.188	375.918	242.792	236.564
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo-R2					0.8725	0.8883	0.8305	0.8067

Notes: For both model's specifications, we inserted law deterrence variables for origin and destination provinces as control values. In addition, fixed effects of dummies for macro-areas of North, Centre and South of origin and destination are included with interaction with year (2010-2017 period). Standard errors account for clustering of provinces. ***, ** and * denote coefficients that are statistically significant at 1%, 5% and 10%, respectively. Results are in **b/se***

Results are in **b/se**

3.3.1 *Robustness Check*

In order to compare the consistency of the results provided by ZIP and PPML, this sections present three robustness checks that consists of i) the usage of the alternative variable of corruption cases made by known authors (*authors*) as in Chapter 2, ii) the addition of the interaction term of distance and transport infrastructures to consider the effective distance to be travelled from origin to destinations, iii) the introduction of the constraint for long-distances skilled movements from the Centre-South to the North of Italy. Results are on Tables 6, 7 and 8 in Appendix C respectively.

Thus, Table 6 presents the same model used so far but the usual variable of corruption is replaced by corruption (authors), that denotes the number of convicted authors for illegal activities related corruption. In this way, it is possible to compare the results get by Chapter 2 with Chapter 3 when the alternative variable for corruption (*authors*) is used. Hence, this robustness check is added to maintain coherence between the analyses conducted. Although this alternative variable of corruption (*authors*) becomes unstable because loses statistical significance in certain circumstances (with exception for destination provinces), its signs are usually preserved and confirm the results previously obtained: an increase in number of persons involved in corruption offences has a positive influence over young skilled movements toward destination provinces that have smaller number of convicted authors of crimes. Then, the remaining independent variables preserve signs as expected but lose some statistical significance (for example, the real GDP per capita loses significance in all cases per origin and destination provinces, size of university in four cases for origin provinces, the standardized quality of university in six cases for destination provinces and the standardized quality of life in six cases for origin and destination provinces).

Then, the following analyses are performed to enrich information related the effects of corruption over young Italian brain drain also.

Table 7 reports the interaction of transport infrastructures with distance, expressed in kilometres (km). Interestingly, values are, on average, statistically significant and indicate that their presence reduce the travelling distance (km) between provinces, as its negative sign mainly suggests. In fact, provinces with airports render long-distances achievable in less time. Same consideration is valid for high-speed train stations. On the other hand, the interaction term of distance and ports presents positive sign for origin and negative sign for destinations. In fact, connections by ports are not fast as those provided by airports and trains and time necessary to achieve long-distances is double than time spent by airplanes and/or trains. Hence, cities with ports exercise low attractiveness to fast mobility. Besides, as Table 7 illustrates, the sign of corruption is preserved although loses, in certain cases, statistical significance. At the same time, the sensibility to corruption of students enrolled to courses of different study-fields remain unchanged and in line with the main results previously achieved. For the remaining variables, even though part of these of become unstable (for example, real GDP per capita loses significance in all cases for origin and destination provinces, size of university in three cases for origin provinces and standardized quality of university in sixth cases for origin and destination provinces), the sign of all the estimated coefficients is preserved mainly.

Table 8 contains the specification of movements to non-adjacent macro-area. To evaluate the long-distance movements, we group the Italian provinces into three macro-area, namely the North, the Centre, and the South. This categorical variable, designed for origin and destination j, assumes the values of 1 for identifying the North, of 2 for the Centre and of 3 for the South. For all regressions (columns I to VIII), we insert the condition of skilled movements by using as benchmark the value of the Centre (if macroarea >2 and macroarea_j<2) for indicating skilled flows from the Centre-South to the North. For ZIP and PPMLHDFE, time and population variables preserve their expected signs and are statistically meaningful. Corruption maintains both the positive sign and statistical significance for origin as well as the negative sign and statistical significance for destinations. Besides, the sensibility to corruption for students enrolled in different studies fields remain unchanged even when long-distance moments are considered, further confirming its validity.

For the remaining variables, even if part of these one become unstable with ZIP, the results presented by PPMLHDFE are more stable because they not only preserve the signs of the estimated coefficients but also exhibit statistical significance at 10% and 5%. In sum, we assess that high corruption in centre-southern provinces encourage skilled mobility from the Centre-South to the North of Italy, because northern destinations exhibit, on average, lower level of corruption, ceteris paribus.

3.3.2 Endogeneity

A critical concern in empirical analysis is endogeneity which causes biased and inconsistent results. In the context of dynamic panel models, examples on the methodology to adopt are provided by Dimant et al., 2013, who use the fixed effects IV estimation and instrument corruption by the quality of judicial institutions and the degree of democratic participation. In addition, Ketterer and Rodrigues-Pose, 2015, perform a two-stage IV procedure and instrument local government efficiency and voice and accountability indicators with their past values. Also, Cooray and Schneider, 2016, carry-out system GMM and IV and instrument corruption with the variable of latitude (that is criticized).

In the context of gravity models, the common approach used to mitigate the effect of endogeneity relies on the introduction of lagged variables of main variables as instruments and/or the usage of the Instrumental Variable (IV) regression with alterative instruments. Mayda, 2009, use the lagged value of per worker GDP at home and abroad as instrument on emigration rate. Dotti et al., 2013, opt to instrument the dependent variable of enrolled students with its past values. Besides, Biagi et al., 2011, use the two-stage GMM and instrument GDP and unemployment rate with three alternative instruments that are the performance of the football teams in the destination province, the industry mix employment rate and the number of ATM machines per 10,000 inhabitants. Also, Beine et al., 2014, use Poisson with IV regression and instrument network with guest worker program in 60 and 70 years as bilateral agreements on net skilled migration.

On the premises of the already cited works, this study controls for the potential reverse causality between corruption and skilled brain migration by mixing traditional strategies, such as the use of lagged values of the main variables as instruments, with novel procedures. In addition, it controls for possible bias determined by the network effect. The methodology adopted is similar to the one proposed by Drivas et al., 2020[58]. The instruments used for corruption and network are their respective one-year lagged values.

For the first causality case, we apply a two-stage residual estimation that is the equivalent of two-stage least square (2SLS) for count data. In the first stage, we regress our endogenous variable of corruption (with dual specification *i* and *j*), with its one-year lag instrument, upon the exogenous variables of the model. Once we recover the predicted residuals (with dual specification *i* and *j*) of this estimation, we plug them into the original models (first, ZIP then PPML) in the second stage. The inference is based on bootstrapping over all the two-step procedure with 200 replications. Results are shown in Table 9, with ZIP in the second stage, and Table 10, with PPML in the second stage, of Appendix C. Both tables present similar results: the sign of the coefficient of corruption at origin is positive while for corruption at destination is negative, as expected. However, the coefficient for corruption at origin is not statistically significant while it is statistically meaningful for corruption at destination. However, the bootstrap standard errors are relatively low and the estimated residuals of coefficients for corruption at origin and destination are not statistically significant at 5% significance level. Thus, there is weak evidence of endogeneity of corruption not as push but pull factor influencing brain drain once the model corrects for the eventual endogeneity, even if endogeneity is not a severe problem in this study.

Then, we consider the effects of network over skilled mobility. The variable of network is itself endogenous because past decisions tend to influence the current decisions to move. Hence, one-year lag of the dependent variable of enrolled students is used as instrument for network because it avoids the omitted variable bias and it provides the idea of continuity trend in mobility from origins to destinations. The model used is the same two-stage procedure already presented above. The results are presented in Table 11, with PPML in the second stage[59], of Appendix C. All the estimated

[58] For more details, see Drivas et al. 2020

[59] We have also used ZIP in the second stage but problem of convergencies arise. Thus, PPML results to be an ideal model to account for endogeneity bias with the two-stage procedure suggested by Drivas et al, (2020) for gravity models

coefficients have the sign as expected. Besides, the coefficient for corruption at origin and network are not statistically significant while the coefficient for corruption at destination is statistically meaningful. However, the bootstrap standard errors are relatively low and the estimated residuals for network, corruption at origin and destination are not statistically significant at the 5% significance level. Thus, although there is weak evidence of endogeneity for corruption not as push but pull factor[60] of influence over skilled mobility, network is not a determinant factor over the decisions of skilled individuals to move to places where native communities are deep-seated, contrary to the conclusions achieved by Beine et al., 2014[61]. Again, this result proves that endogeneity is not a severe problem to these gravity models and the results obtained are robust[62]. This fact proves that trade-off between robustness and completeness in empirical analysis can be overcome if models are correctly implemented.

[60] However, one limit can be represented by the number of replications for the bootstrap procedure that is not sufficient (200). It is possible that increasing the number of replications (500 or 1000) may rule out the statistical significance of the residuals for corruption at destination and strongly prove that our estimates do not suffer of endogeneity

[61] Disclaim for the usage of the instrument for network: one-year lag of the dependent variable of enrolled student may be replaced by other valid instruments that could confirm or not bias of the network effects over young skilled mobility

3.4 Conclusion

This dissertation is aimed at expanding the knowledge of the size and magnitude of corruption on young brain drain from origin to destination Italian provinces. The results suggest that corruption has noticeable influence over skilled mobility. This evidence is robust and persistent throughout the usage of ZIP and the ground-breaking PPML and return similar and significant results to the ones provided by the traditional system-GMM. Additive estimates, made as robustness checks, confirm that results do not suffer from severe biases and inconsistency. Thus, this study overrides the trade-off between robustness and completeness and demonstrate that novel models, if correctly implemented, shall endeavour to ensure the coexistence of both features in the empirical analysis. Evidence proves that high corruption at origins acts as push factor that incentivizes prospective tertiary students to move away (Chapter 2), while low corruption at destinations acts as pull factor that attracts prospective tertiary students to destinations, ceteris paribus (Chapter 3). Besides, models of Chapter 3, since exploit more information provided by bilateral data, add more findings to the ones previously achieved. First, Chapter 3 proves that sensitivity to corruption of prospective students vary according to the study fields they choose for their studies: results suggests that, on average, skilled students enrolled to courses of Social and Physical Sciences tend to be more susceptible to corruption than those who are enrolled to courses of Life Sciences. Also, corruption and brain drain partly explains one amongst all possible causes of the dualism condition of North and South of Italy: high corruption in the southern provinces tend to increase the number of skilled individuals who move from the Centre-South to the North of the country. Thus, human resources are not equally distributed and such condition determines negative consequences on the growth rate of Italy, that runs fast and slow in the North and the South, respectively. Thus, few considerations on the policy remedies to adopt to reduce both phenomena can be drawn. First, policymakers should invest more resources on higher education to promote meritocracy and to build-up lawful-oriented minds of prospective tertiary students who are going to be the tomorrow class of workers. To this end, they should promote the increase of investments addressed to universities that are considered as “gateway” of skilled mobility. Universities can attract but should be capable to retain talents also. To this end, policies should enact remedies to smooth the structural rigidity of public competitions or limited availability of scholarships at certain university, promoting the right to study and to work for all students. Thus, universities should readdress the (job) insecurity suffered by students and alumni by enacting more traineeships, fellowships and scholarship. In doing so, talents are attracted and encouraged to remain to their local university. Secondly, policymakers should regulate skilled mobility by monitoring skilled inflows and outflows and intervene when the outflows are greater than inflows. To this end, they should enforce rules that promote job opportunities and eliminate excess of bureaucratic norms. Besides, they should contemplate norms that regulate agreements between local SMEs and universities to promote mutual exchange of skills and job. Also, local active policies should be reinforced to eliminate disparities among underprivileged skilled minorities (women). Thus, corrections should induce a virtuous circle leading to the decrease of intranational skilled mobility.

Future research may develop an empirical analysis that investigates the effects of corruption on skilled individuals who emigrate abroad. It would be interesting to perform a comparative analysis that study whether corruption differently affects pre-graduate and post-graduated individuals to move abroad. In doing so, results can prove whether university or workplaces in Italy are identified as places where merit is recognized/rewarded or not. Another interesting research would be to evaluate the effects of remittances from native residents abroad on level of corruption at origin[63]. Finally, many studies of economics of crime evaluate corruption with perception indices, which return data on the conventional understanding of corruption causes. Hence, the challenging future task of this research is to refine and gather more objective-based measures of corruption and to examine the effective pattern they reveal, by constructing a Meritocracy Index that may encompass variables at micro, meso and macro-levels to study the phenomenon carefully.

---

[63] It is presumed that places that receive remittances from abroad are affected by economic and social issues. For more details, see Rapoport and Docquier, 2006

References

- Anderson, Tim; Hsiao, William; Estimation of dynamic models with error components, *Journal of the American Statistical Association,* 16, 598-606, 1981
- Anderson, Tim; Hsiao, William; Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47-82, 1982
- Angrist, Joshua; Pischke, Jorn; Mostly Harmless Econometrics: An Empiricist's Companion, *Princeton University Press*, 2008, Princeton
- Apaza, Carmen; Measuring Governance and Corruption through the Worldwide Governance Indicators: Critiques, Responses, and Ongoing Scholarly Discussion, *Political Science and Politics*, 42, 1, 139-43, 2009, Cambridge University Press
- Arellano, Manuel; Panel Data Econometrics, *Oxford University Press,* 2003 Oxford
- Arellano, Manuel; Bond, Stephen; Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations" *Review of Economic Studies*, 58,277-297, 1991
- Arellano Manuel; Bover, Olympia; Another look at the instrumental-variable estimation of error components models, *Journal of Econometrics*, 68:29-52,1995
- Auer, Daniel; Romer, Friederike; Tjaden Jasper; Corruption and desire to leave quasi-experimental evidence on corruption as a driver of emigration intentions, *IZA Journal of Development and Migration*, 11,7, 2020
- Basile, Roberto; De Benedictis, Luca; Durban, María; Faggian, Alessandra; Mínguez, Román; The impact of immigration on the internal mobility of natives and foreign-born residents: evidence from Italy, *Spatial Economic Analysis*, 16, 1, 9-26, 2021, Taylor & Francis
- Basile, Roberto; Girardi, Alessandro; Mantuano, Marianna; Russo, Giuseppe; Interregional migration of human capital and unemployment dynamics: evidence from Italian provinces, *German Economic Review*, 20, 4, 385-414, 2019, Wiley Online Library

- Beine, Michel; Noël, Romain; Ragot, Lionel; Determinants of the international mobility of students, *Economics of Education Review*, 41, 40-54, 2014, Elsevier
- Michel Beine, Frédéric Docquier, Cecily Oden-Defoort, A Panel Data Analysis of the Brain Gain, *World Development*, 39, 4, 523-532, 2011
- Biagi, Bianca; Faggian, Alessandra; McCann, Philip; Long and short distance migration in Italy: the role of economic, social and environmental characteristics", *Spatial Economic Analysis*, 6, 1, 111-131, 2011, Taylor & Francis
- Biondo, Alessio, Emanuele; Monteleone, Simona; Skonieczny, Giorgio; Torrisi, Benedetto; The propensity to return: Theory and evidence for the Italian brain drain, *Economics Letters*, 115, 3, 359-362, 2012, Elsevier
- Blundell, Richard; Bond, Stephen; "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87, 115-143, 1998
- Bonasia, Mariangela; Napolitano, Oreste; Determinants of interregional migration flows: the role of environmental factors in the Italian case, *The Manchester School*, 80, 4, 525-544, 2012, Wiley Online Library
- Bond, Stephen; Dynamic panel data models: a guide to micro data methods and practice, Cemmap WP CWP09, 2002
- Bond, Stephen; Windmeijer, Frank; Reliable Inference for GMM estimators? Finite Sample Properties of Alternative Test procedures in Linear Panel Data Models, *Econometric Reviews*, 24, 1, 1-37, 2005
- Bun, Manuel; On the diminishing returns of higher-order terms in asymptotic expansions of bias, *Economic Letters*, 79, 2003
- Cameron, Colin; Trivedi, Pravin; Regression analysis of count data. 821 *Cambridge University Press*, New York NY, 822, 2005

- Charron, Nicholas; Lapuente, Victor; Why do some regions in Europe have a higher quality of government? *The Journal of Politics*, 75, 3, 567-582, 2013, Cambridge University Press, New York
- Ciriaci, Daria; Does university quality influence the interregional mobility of students and graduates? The case of Italy, *Regional Studies*, 48, 10, 1592-1608, 2013, Taylor & Francis
- Ciriaci, Daria; Muscio, Alessandro; "University choice, research quality and graduates' employability: Evidence from Italian national survey data, *European Educational Research Journal*, 13, 2, 199-219, 2014, SAGE Publications Sage UK: London, England
- Ciriaci, Daria; Palma, Daniela; The role of knowledge-based supply specialisation for competitiveness: A spatial econometric approach, *Papers in Regional Science*, 87, 3, 453-475, 2008, Wiley Online Library
- Cooray, Arusha; Schneider, Friedrich; Does corruption promote emigration? An empirical examination, *Journal of Population Economics*, 29, 1, 293-310, 2016, Springer
- Corrado, Germana; Rossetti, Fiammetta; Public corruption: A study across regions in Italy, *Journal of Policy Modeling*, 40, 6, 1126-1139, 2018, Elsevier
- Correia, Sergio; Guimarães, Paulo; Zylkin, Tom; Fast Poisson estimation with high-dimensional fixed effects, *The Stata Journal*, 20, 1, 95-115, 2020, SAGE Publications Sage CA: Los Angeles
- Dimant, Eugen; Krieger, Tim; Meierrieks, Daniel; The effect of corruption on migration, 1985–2000, *Applied Economics Letters* 20.13:1270-1274, 2013
- Docquier, Frédéric; Rapoport, Hillel; Globalization, brain drain, and development, *Journal of Economic Literature*, 50, 3, 681-730, 2012
- Dotti, Nicola, Francesco; Fratesi, Ugo; Lenzi, Camilla; Percoco, Marco; Local labour markets and the interregional mobility of Italian university students, *Spatial Economic Analysis*, 8, 4, 443-468, 2013, Taylor & Francis

- Dotti, Nicola Francesco; Fratesi, Ugo; Lenzi, Camilla; Percoco, Marco; Marchés locaux du travail et mobilité interrégionale des étudiants universitaires italiens, *Regional Studies*, 2013
- Drivas, Kyriacos; Economidou, Claire; Karamanis, Dimitrios; Sanders, Mark; Mobility of highly skilled individuals and local innovation activity. *Technological Forecasting and Social Change*, *158*, 120-144, 2020
- Etzo, Ivan; The determinant of the recent interregional migration flows in Italy: A panel data analysis, *Journal of Regional Science*, 51, 5, 948-966, 2011, Wiley Online Library
- Faggian, Alessandra; Rajbhandari, Isha; Dotzel, Kathryn; The interregional migration of human capital and its regional consequences: a review, *Regional Studies*, 51, 1, 128-143, 2017, Taylor & Francis
- Graves, Philip; Sexton, Robert; Knapp, Thomas; A multi-disciplinary interpretation of migration: Amenity capitalization in both land and labour markets, *The Annals of Regional science*, 18, 2, 35-44, 1984
- Greene, William; Econometric analysis, *Boston Pearson*, 2008 Boston
- Gibson, John; McKenzie, David; Eight Questions about Brain Drain, *Journal of Economic Perspectives*, 25, 3, 107-128, 2011
- Hsiao, William; Analysis of Panel Data, *Cambridge University Press*, 2003
- Gould, Laurence; The systems psychodynamics of organizations: Integrating the group relations approach, psychoanalytic, and open systems perspectives, Routledge, 2018
- Greenwood, Michael; Gary, Hunt; Jobs versus amenities in the analysis of metropolitan migration, *Journal of Urban Economics*, 25, 1, 1-16, 1989
- Gründler, Klaus; Potrafke, Niklas; Corruption and economic growth: new empirical evidence, *European Journal of Political Economy*, Vol. 60,101-810, ISSN 0176-2680, 2019
- Hardin, James; Hilbe, Joseph Michael; Hilbe, James; Generalized linear models and extensions. *Stata press*, 2007

- Hwang, Jungbun.; Sun, Yixiao; Should We Go One Step Further? An Accurate Comparison of One-Step and Two-Step Procedures in a Generalized Method of Moments Framework, *Journal of Econometrics*, 207, 381-405, 2018
- Hwang, Jungbin; Sun, Yixiao; Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework," Journal of Econometrics, Elsevier, 207,2, 381-405, 2018
- Iammarino, Simona; Marinelli, Elisabetta; Education–job (mis) match and interregional migration: Italian university graduates' transition to work, *Regional Studies*, 49, 5, 866-882, 2015, Taylor & Francis
- Ketterer, Tobias; Rodríguez-Pose, Andrés; Local quality of government and voting with one's feet, *The Annals of Regional Science*, 55, 2, 501-532, 2015, Springer
- Khan, Haroon Ur Rashid; Khan, Anwar; Zaman, Khalid; Nabi, Agha Amad; Hishan, Sanil, Islam, Talat; Gender discrimination in education, health, and labour market: a voice for equality, *Quality & Quantity*, 51, 5, 2245-2266, 2017
- Lisciandra, Maurizio; Milani, Riccardo; Millemaci Emanuele; A corruption risk indicator for public procurement, *European Journal of Political Economy*, ISSN 0176-2680, https://doi.org/10.1016/j.ejpoleco.2021.102141, 2021
- Lisciandra, Maurizio; Millemaci, Emanuele; The Economic Effect of Corruption in Italy: A Regional Panel Analysis, *Regional Studies*, 51, 9, 1387-1398, 2017
- Long, John Scott; Freese, Jeremy; Regression models for categorical dependent variables using Stata7, 2006, Stata Press
- Long, John Scott; Regression models for categorical and limited dependent variables, 1997, Sage
- Mauro, Paolo; Corruption and Growth, *The Quarterly Journal of Economics*,110, 3,681-712, 1995, MIT Press

- Mayda, Anna Maria; International migration: A panel data analysis of the determinants of bilateral flows, *Journal of Population Economics*, 23, 4, 1249-1274, 2010, Springer
- Michaeli, Moti; Casari, Marco; Ichino, Andrea; De Paola, Maria; Marandola, Ginevra; Scoppa, Vincenzo; Civcness Drain, IZA Discussion Paper No. 11955, 2021
- Mocetti, Sauro; Porello, Carmine; How does immigration affect native internal mobility? New evidence from Italy, *Regional Science and Urban Economics*, 40,6, 427-439, 2010, Elsevier
- Nifo, Annamaria; Pagnotta, Stefano; Scalera, Domenico; The best and brightest. Positive selection and brain drain in Italian internal migrations, 2011
- Nifo, Annamaria; Vecchione, Gaetano; Do institutions play a role in net skilled migration? The case of Italy, *Regional Studies*, 48, 10, 1628-1649, 2014, Taylor & Francis
- Piras, Romano; 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, 2017, Wiley Online Library
- Piras, Romano; Internal migration in Italy: The role of migration networks, *Italian Economic Journal*, 6, 1, 157-195, 2020, Springer
- Poprawe, Marie; On the relationship between corruption and migration: empirical evidence from a gravity model of migration, *Public Choice*, 163, 3, 337-354, 2015, Springer
- Rodríguez-Pose, Andrés; Ketterer, Tobias D; Do local amenities affect the appeal of regions in Europe for migrants? *Journal of Regional Science*, 52,4,535-561,2012, Wiley Online Library
- Roodman, David; How to do xtabond2: An introduction to difference and system GMM in Stata, *The Stata Journal*, 12,4,766-767, 2012
- Silva, João, Santos; Tenreyro, Silvana; Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator, *Economics Letters*, 112, 2, 220-222, 2011, Elsevier

- Silva, João, Santos; Tenreyro, Silvana; On the existence of the maximum likelihood estimates in Poisson regression, *Economics Letters*, 107,2,310-312,2010, Elsevier
- Silva, João, Santos; Tenreyro, Silvana; The log of gravity, *The Review of Economics and Statistics*, 88,4,641-658, 2006, The MIT Press
- Silva, João, Santos; Tenreyro, Silvana; Windmeijer, Frank; Testing competing models for non-negative data with many zeros, *Journal of Econometric Methods*, 4,1,29-46, 2015, de Gruyter
- Torrisi, Benedetto; Pernagallo, Giuseppe; Investigating the relationship between job satisfaction and academic brain drain: the Italian case, *Scientometrics*, 124, 925-952, 2020, Springer
- Treisman, Daniel; What Have We Learned about the Causes of Corruption from Ten Years of Cross-National Empirical Research, *Annual Review of Political Science*, 10, 2007, Available at SSRN: https://ssrn.com/abstract=1077293
- Van Bouwel, Linda; Veugelers, Reinhilde; The determinants of student mobility in Europe: the quality dimension, *European Journal of Higher Education*, 3, 2, 172-190, 2013, Taylor & Francis
- Waltman, Ludo; Van Eck, Nees Jan; Noyons, Edward; A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4,4, 629-635, 2010
- Van Eck, Nees Jan; Waltman, Ludo; Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84,2,523-538, 2010
- William, Richard; Models for Count Outcomes, University of Notre Dame, 2021 Available at: https://www3.nd.edu/~rwilliam/
- Windmeijer, Frank; Silva, João, Santos; Endogeneity in count data models: an application to demand for health care, *Journal of Applied Econometrics*, 12, 3, 281-294, 1997, Wiley Online Library

Chapter 1 - Appendix A

Bibliometric Analysis

In last ten years, the interest toward brain drain has increased among researchers. In fact, from 2010 the number of publications on this topic has more than doubled until 2021. Evidence is provided by data reported on Fig. 1 and Fig. 2 which are derived throughout the insertion of keywords of “brain drain”, “net skilled migration”, “corruption”, “2010-2021” and refined search for “Italy” on Web of Science, a bibliographic database produced by Thomson Reuters. The research returns 119 publications divided into 100 articles, 1 book chapter, 13 proceeding papers and 5 other documents (Fig.1). The number of publications is highest in 2014, 2017 and 2019 while is lower in 2010, 2011 and 2013. Besides, the number of cited documents is increased from 2010 to 2021 and the total citations are 1344 (H index =19) (Fig.2)

Figure A.1 *Total number of publications per year (source: Web of Science)*

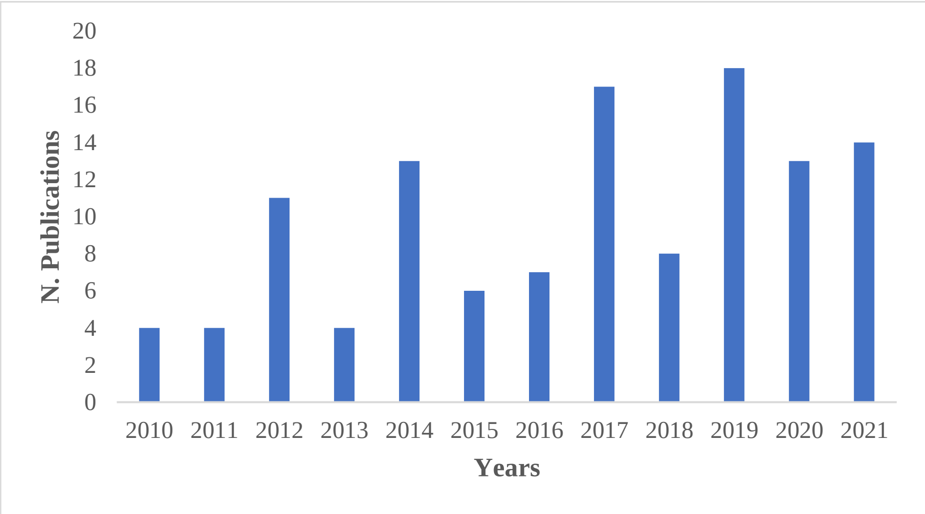
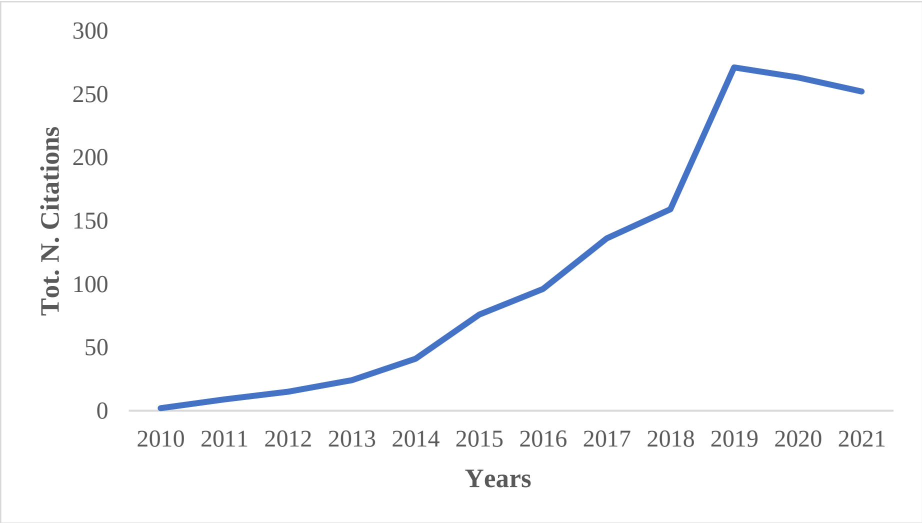


Figure A.2 *Total number of citations per year (source: Web of Science)*



In addition, the analysis includes the number of publications divided for the first 10 area of research and for the first 10 journals on which are published. As figure 3 demonstrates, the subject area that deeply focus on the topic is "Economics", followed by "Environmental Studies", "Geography", "Demography" and "Regional Urban Planning". In addition, figure 4 reports the journals on which these documents are published: International Migration, Regional Studies and Research Policy emerge with the highest numbers of contributions respect to the other journals.

Figure A.3 *Publications per research area (2010-2021) (source: Web of Science)*

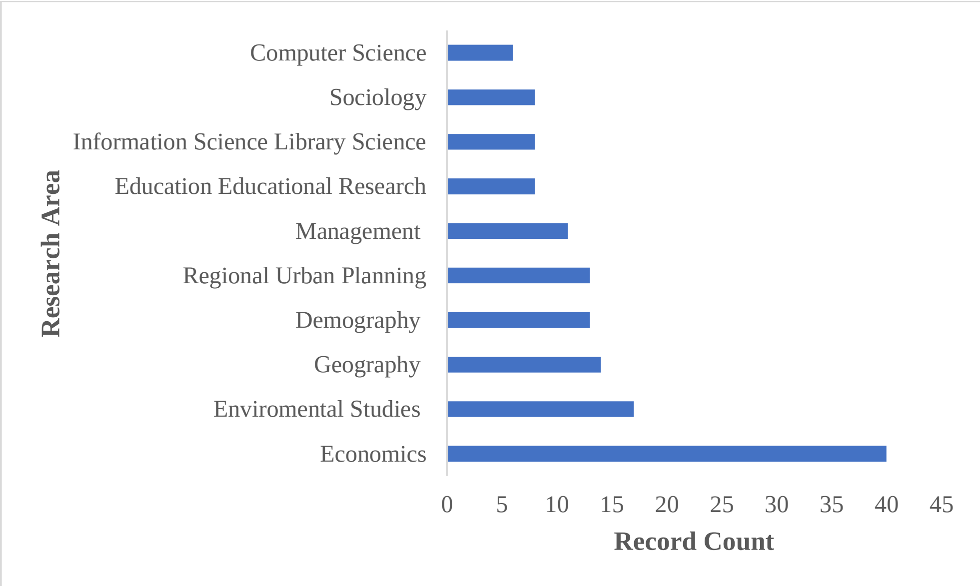
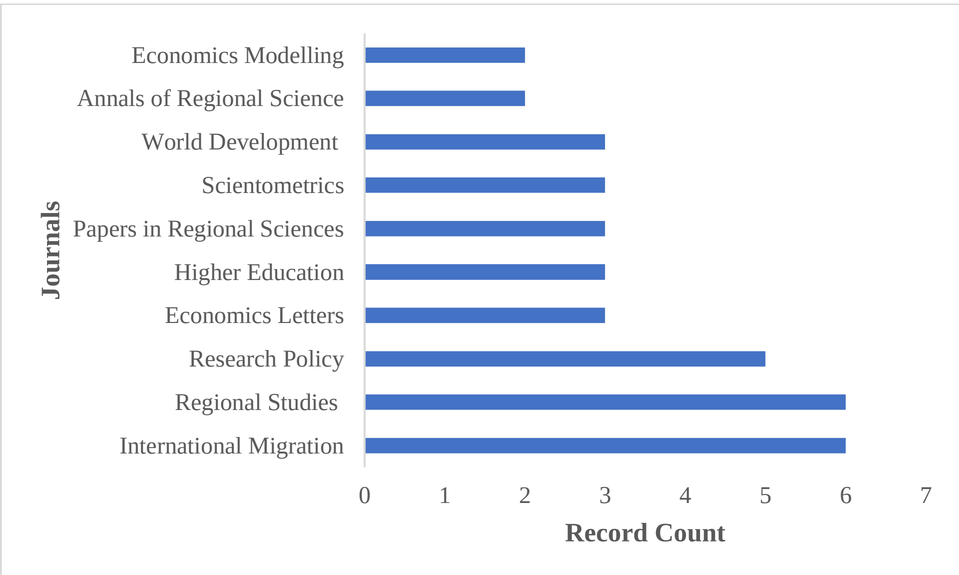


Figure A.4 *Publications per journals (2010-2021) (source: Web of Science)*



Then, additive analyses of keyworks are performed through the selection of words that frequently occur in the title and abstracts of the publications. These methods are i) network analysis, which is a data mining procedure that permits to visualize items that are grouped into clusters with labels and circles that indicate the weight of each item, and ii) the overlays analysis, which is a data mining procedure that permits to know whether the topics treated in the research are old or newer to the literature. All keywords, taken from Web of Science (as already presented above), have been processed into VOS-viewer software, to create the conceptual mappings in two-dimensional spaces, as suggested by Van Eck and Waltman, 2010.

Network analysis (Fig. 5) is made of keywords, that occur a minimum of 3 times in the above-mentioned sample of 119 publications, that are positioned in nodes connected each other by links. The greater the mass of each node, the higher the occurrence of the keyword of that node. Besides, the lower the distance between nodes, the higher is their relatedness. The higher the number of links, the higher is the connection between the two and/or more items. Each item belongs to a cluster. Fig. 5 shows four clusters that are differently coloured because indicates diverse fields of study: in fact, “red” refers to the “Brain Drain of Globalization” (related to Higher Education or Labour Markets), “green” indicates the “International Net skilled migration”, “yellow” refers to the “Interregional Net skilled migration” and blue indicates the “Brain Drain of Scientific Innovation”.

The overlay visualization (Fig. 6) permits to understand the chronological evolution of the theme studied in years. At the bottom of the figure, the Legenda shows if the topics published are old (blue) topics or latest (yellow) topics. The oldest topics in 2014 are related “International Net skilled migration” and “Skilled Remittances”. Then, the study interest moves toward the “Brain Drain of Globalization” – higher education, job opportunities, quality of life – and the “Brain Drain of Innovation” between 2016 and 2017. Recently, the interest of scholars on studying brain drain moved from broader context to a narrower context of “Interregional Net skilled migration” – with a special focus on the main determinants of skilled mobility that, in turn, could affect regional growth rate (in line with the research topic of this thesis).

Figure A.5 *Network Analysis (source: Web of Science plus VOSViewer)*

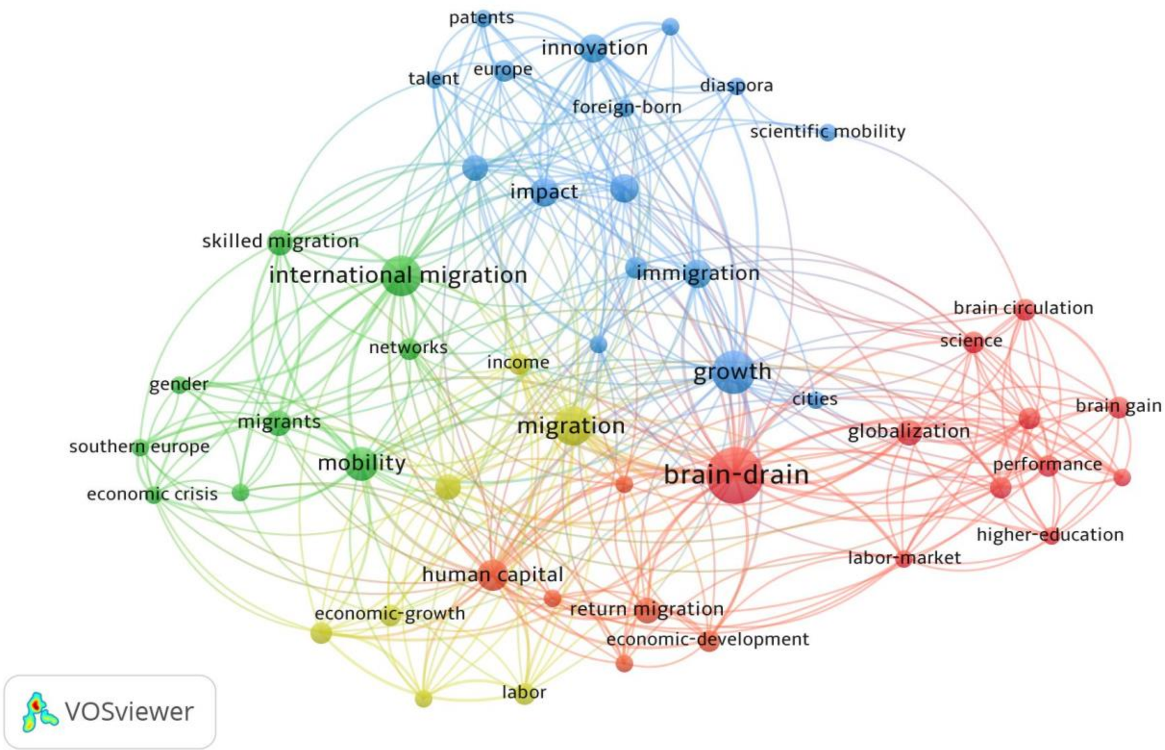
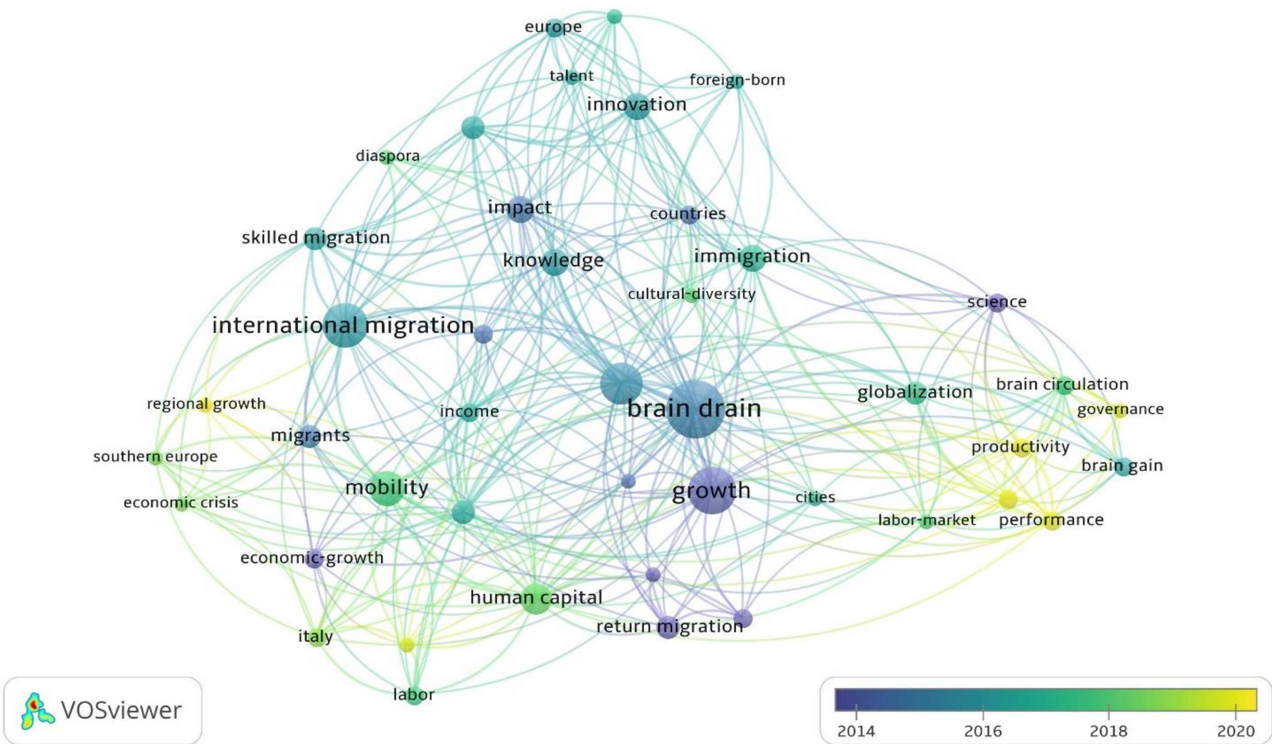


Figure A.6 *Overlay visualization of keywords (source: Web of Science plus VOSViewer)*



Chapter 2 - Appendix B

Table B.1 *Data Sources*

Variables	***Description***	***Source***
lbrain_t120	number of enrolled who travel 2 hours / tot. enrolled students	MIUR
lbrain200	number of enrolled who travel 200 km / tot. enrolled students	MIUR
lcorr_iap	corruption against P.A- artt. 314 to 322 of Italian tort law	RE.GE ISTAT
lcorr_authors	corruption by authors against P.A- artt. 314 to 322 of Italian tort law	RE.GE ISTAT
lpop	logarithm of average annual population	ISTAT
lrgdppc	logarithm of GDP per capita	ISTAT
employment	employment rate	ISTAT
uni_size	size of university with values =1 large, =2 medium, =3 small	CENSIS
zquality_univ	quality of university for services, scholarships, infrastructures, web, internationalization	CENSIS
zhospital_mig	hospital migration	ISTAT
Dairport	dummy for D=1 airport presence, D=0 otherwise	ISTAT
DTAV	dummy for D=1 HST service presence, D=0 otherwise	ISTAT
Dport	dummy for D=1 port presence, D=0 otherwise	ISTAT
disccrim	variable indicator of law enforcement	ISTAT
probofarr	variable indicator of law enforcement	ISTAT
probofconv	variable indicator of law enforcement	ISTAT

Table B.2 *Descriptive Statistics*

Variable	***Obs***	***Mean***	***Std. Dev.***	***Min***	***Max***
lbrain_t120	880	-2,548116	1,6221	-6,678028	0
lbrain200	880	-2,704234	1,757402	-7,371175	0
lcorr_iap	880	-10,05495	1,836392	-13,59804	0
lcorr_authors	880	-9,581675	1,869625	-13,59804	0
lpop	880	12,89305	0,74905	10,95172	15,287
lrgdppc	880	10,08998	0,2753795	9,481724	10,88089
employment	880	56,61151	10,58151	0	72,9
uni_size	880	1,872727	0,7645036	1	3
zquality_university	880	1,01E-09	1	-12,8147	2,580524
zhospital_migration	880	1,06E-09	1	-1,519476	3,684861
Dairport	880	0,3511364	0,4775969	0	1
DTAV	880	0,0727273	0,259836	0	1
Dport	880	0,2181818	0,4132464	0	1
disccrim	880	22,39125	5,762358	9	42,4
probofarr	880	0,4196229	0,1368231	0,1325305	1,057299
probofconv	880	0,2655381	0,1079817	0,0210307	0,8046912
DNorth	880	0,4272727	0,4949638	0	1
DCentre	880	0,2	0,4002275	0	1
DSouth	880	0,3727273	0,4838054	0	1

Table B.3 *Italian Provinces, Regions and Macroarea*[64]

Provinces	***Region***	***Macroarea***
Torino	Piedmont	North
Vercelli	Piedmont	North
Novara	Piedmont	North
Cuneo	Piedmont	North
Asti	Piedmont	North
Alessandria	Piedmont	North
Biella	Piedmont	North
Verbania	Piedmont	North
Aosta	Aosta Valley	North
Imperia	Ligury	North
Savona	Ligury	North
Genova	Ligury	North
La Spezia	Ligury	North
Varese	Lombardy	North
Como	Lombardy	North
Sondrio	Lombardy	North
Milano	Lombardy	North
Bergamo	Lombardy	North
Brescia	Lombardy	North
Pavia	Lombardy	North
Cremona	Lombardy	North
Mantova	Lombardy	North
Lecco	Lombardy	North
Lodi	Lombardy	North
Monza	Lombardy	North
Bolzano	Trentino-Alto Adige	North
Trento	Trentino-Alto Adige	North
Verona	Veneto	North
Vicenza	Veneto	North
Belluno	Veneto	North
Treviso	Veneto	North
Venezia	Veneto	North
Padova	Veneto	North
Rovigo	Veneto	North
Udine	Friuli-Venezia Giulia	North
Gorizia	Friuli-Venezia Giulia	North
Trieste	Friuli-Venezia Giulia	North
Pordenone	Friuli-Venezia Giulia	North
Piacenza	Emilia-Romagna	North

[64] Data source is provided by ISTAT

Parma	Emilia-Romagna	North
Reggio nell'Emilia	Emilia-Romagna	North
Modena	Emilia-Romagna	North
Bologna	Emilia-Romagna	North
Ferrara	Emilia-Romagna	North
Ravenna	Emilia-Romagna	North
Forlì-Cesena	Emilia-Romagna	North
Rimini	Emilia-Romagna	North
Pesaro	Marche	Centre
Ancona	Marche	Centre
Macerata	Marche	Centre
Fermo	Marche	Centre
Ascoli Piceno	Marche	Centre
Massa Carrara	Tuscany	Centre
Lucca	Tuscany	Centre
Pistoia	Tuscany	Centre
Firenze	Tuscany	Centre
Livorno	Tuscany	Centre
Pisa	Tuscany	Centre
Arezzo	Tuscany	Centre
Siena	Tuscany	Centre
Grosseto	Tuscany	Centre
Prato	Tuscany	Centre
Perugia	Umbria	Centre
Terni	Umbria	Centre
Viterbo	Lazio	Centre
Rieti	Lazio	Centre
Roma	Lazio	Centre
Latina	Lazio	Centre
Frosinone	Lazio	Centre
Caserta	Campania	South
Benevento	Campania	South
Napoli	Campania	South
Avellino	Campania	South
Salerno	Campania	South
L'Aquila	Abruzzo	South
Teramo	Abruzzo	South
Pescara	Abruzzo	South
Chieti	Abruzzo	South
Campobasso	Molise	South
Isernia	Molise	South
Foggia	Puglia	South
Bari	Puglia	South
Taranto	Puglia	South

Brindisi	Puglia	South
Lecce	Puglia	South
Trani	Puglia	South
Potenza	Basilicata	South
Matera	Basilicata	South
Cosenza	Calabria	South
Crotone	Calabria	South
Vibo Valentia	Calabria	South
Catanzaro	Calabria	South
Reggio di Calabria	Calabria	South
Trapani	Sicily	South
Palermo	Sicily	South
Messina	Sicily	South
Agrigento	Sicily	South
Caltanissetta	Sicily	South
Enna	Sicily	South
Catania	Sicily	South
Ragusa	Sicily	South
Siracusa	Sicily	South
Sassari	Sardinia	South
Nuoro	Sardinia	South
Cagliari	Sardinia	South
Oristano	Sardinia	South
Olbia-Tempio	Sardinia	South
Ogliastra	Sardinia	South
Medio Campidano	Sardinia	South
Carbonia Iglesias	Sardinia	South

Notes to Table B.3

1. The nomenclature used for identifying the Italian Provinces follows the one provided by ISTAT. Specifically, this work uses the nomenclature of the edition 2016, where, the new province of Sud Sardinia, ante 2016, results to be divided into four provinces of Olbia-Tempio, Ogliastra, Medio Campidano and Carbonia Iglesias.
2. For year 2017, we continue to use the divided provinces of Sud Sardinia by diving the number of enrolled students of Sud Sardinia by four and giving higher weights (in terms of number of students) to provinces with higher population density rate: Carbonia-Iglesias presented the highest rate while Ogliastra had the lowest one.

Table B.4 *Italian Provinces with University*[65]

Province	***University***	***Type of University***
Torino	Università degli Studi di Torino	Public
Torino	Politecnico di Torino	Public
Torino	Bra Scienze Gastronomiche	Public
Torino	Università degli Studi del Piemonte Orientale	Public
Aosta	Università degli Studi di Aosta	Public
Genova	Università degli Studi di Genova	Public
Milano	Castellanza LIUC	Private
Milano	Università degli Studi di Milano	Public
Milano	Politecnico di Milano	Public
Milano	Università Bocconi	Private
Milano	Università Cattolica	Private
Milano	IULM	Private
Milano	Università degli Studi di Milano Bicocca	Public
Milano	Università Humanitas Rozzano	Private
Como	Campus for Università dell’Insubria	Public
Brescia	Università degli Studi di Brescia	Public
Bergamo	Università degli Studi di Bergamo	Public
Pavia	Università degli Studi di Pavia	Public
Trento	Università degli Studi di Trento	Public
Verona	Università degli Studi di Verona	Public
Venezia	Cà Foscari	Public
Venezia	Iuav- Tolentini	Public
Padova	Università degli Studi di Padova	Public
Udine	Università degli Studi di Udine	Public
Trieste	Università degli Studi di Trieste	Public
Parma	Università degli Studi di Parma	Public
Modena	Università degli Studi di Modena e Reggio Emilia	Public
Reggio nell'Emilia	Università degli Studi di Modena e Reggio Emilia	Public
Bologna	Alma Mater Studiorum -Università di Bologna	Public
Forlì-Cesena	Campus for Università of Bologna	Public
Ferrara	Università degli Studi di Ferrara	Public
Pesaro e Urbino	Università degli Studi di Urbino	Public
Ancona	Università degli Studi delle Marche	Public
Macerata	Università degli Studi di Macerata	Public
Ascoli Piceno	Campus of Politecnica delle Marche	Public
Firenze	Università degli Studi di Firenze	Public
Pisa	Università degli Studi di Pisa	Public
Siena	Università degli Studi di Siena	Public
Siena	Università per Stranieri di Siena	Public

[65] Data source is provided by M.I.U.R

Perugia	Università degli Studi di Perugia	Public
Perugia	Università per Stranieri di Perugia	Public
Viterbo	Università degli Studi della Tuscia	Public
Roma	Università Roma "La Sapienza"	Public
Roma	Università Roma "Tor Vergata"	Public
Roma	Libera Università SS. Maria Assunta - LUMSA	Private
Roma	Libera Università degli Studi Sociali - LUISS Guido Carli	Private
Roma	Università degli Studi di Roma Foro Italico	Public
Roma	Università degli Studi "Roma Tre"	Public
Roma	Università Campus Bio Medico di Roma	Private
Roma	Università degli Studi Internazionali di Roma - UNINT	Private
Roma	UER - Università Europea di Roma	Private
Frosinone	Università degli Studi di Cassino	Public
Benevento	Università degli Studi del Sannio	Public
Napoli	Università "Federico II" di Napoli	Public
Napoli	Università Parthenope di Napoli	Public
Napoli	Università degli Studi di Napoli "L'Orientale"	Public
Napoli	Università Suor Orsola Benincasa	Private
Napoli	Università degli Studi della Campania "L. Vanvitelli"	Public
Salerno	Università degli Studi di Salerno	Public
L'Aquila	Università degli Studi dell'Aquila	Public
Teramo	Università degli Studi di Teramo	Public
Chieti	Università degli Studi di Chieti e Pescara	Public
Pescara	Università degli Studi di Chieti e Pescara	Public
Molise	Università degli Studi del Molise	Public
Foggia	Università degli Studi di Foggia	Public
Bari	Università degli Studi di Bari	Public
Bari	Politecnico di Bari	Public
Bari	Università LUM Jean Monnet	Private
Lecce	Università del Salento	Public
Potenza	Università degli Studi della Basilicata	Public
Cosenza	Università della Calabria	Public
Catanzaro	Università degli Studi di Catanzaro "Magna Graecia"	Public
Reggio di Calabria	Università degli Studi Mediterranea di Reggio Calabria	Public
Reggio di Calabria	Università per Stranieri "Dante Alighieri"	Private
Palermo	Università degli Studi di Palermo	Public
Messina	Università degli Studi di Messina	Public
Enna	Università KORE di Enna	Public
Catania	Università degli Studi di Catania	Public
Sassari	Università degli Studi di Sassari	Public
Cagliari	Università degli Studi di Cagliari	Public

Notes to Table B.4

1. The provinces reported with university are 51 + 3 (that are ancillary campuses and are: Como for University of Insubria, Ascoli Piceno for Politecnica delle Marche, Forlì-Cesena for University of Bologna) for origins.
2. Adjustments for “Modena and Reggio nell ‘Emilia” and “Chieti and Pescara” have been made. The campus of such universities is placed in both cities, and we divide them respectively. Hence, the study divides the number of enrolled students by half, giving more weight (in terms of number of students) to the province that presents the highest population density rate (for example, higher number of enrolled students is attributed to Modena because its population density rate is higher than the one present by Reggio nell ’Emilia. Also, higher number of enrolled students is given to Pescara because its population density rate is higher than the one presented by Chieti)
3. The analysis does not consider Telematic Universities, Schools of Superior Specialization and/or Schools of Excellence
4. The Academic Years evaluated starts from 2010-2011 to 2017-2018
5. This study reports a single-year format for the Academic Year, beginning with 2010 for the A.Y. 2010-2011 and ends with 2017 for the A.Y. 2017-2018
6. Telematic Universities are not considered because data are not available for the period of study of interest fully.

Table B.5 *Simplified sys-GMM regression (control variables avoided)*

	I	II	III	IV	V	VI	VII	VIII
lagged log of brain drain (t=120)	.3277***	.328**	.2491**	.2364*				
	{0.116}	{0.148}	{0.122}	{0.120}				
lagged log of brain drain (d=200)					.316	.2218	.186	.2036
					{0.226}	{0.270}	{0.284}	{0.284}
log of corruption	.329***	.25***	.2349**	.2513**	.355***	.299***	.2614***	.2797***
	{0.104}	{0.089}	{0.102}	{0.101}	{0.101}	{0.090}	{0.097}	{0.092}
quality university	.1846**	.06688	.08308	.07929	.2594**	.1573	.1881*	.1599
	{0.072}	{0.072}	{0.086}	{0.085}	{0.115}	{0.100}	{0.102}	{0.104}
hospital migration		.3378***	.4116***	.3589***		.3932**	.4521**	.3726**
		{0.116}	{0.117}	{0.108}		{0.176}	{0.191}	{0.166}
dummies for airport								
TAV & port	NO	NO	YES	NO	NO	NO	YES	NO
law enforcement	NO	NO	NO	YES	NO	NO	NO	YES
N	378	378	378	378	378	378	378	378
A-B test(1)	0.002	0.003	0.003	0.002	0.020	0.072	0.100	0.091
A-B test(2)	0.191	0.185	0.194	0.220	0.314	0.244	0.244	0.243
Hansen (p-value)	0.062	0.031	0.129	0.206	0.086	0.039	0.043	0.048
# instruments	24	24	28	30	21	21	24	24
# groups	54	54	54	54	54	54	54	54

Table B.6 *Simplified sys-GMM regression (control variables avoided)*

	IX	X	XI	XII	XIII	XIV	XV	XVI
lagged log of brain drain (t=120)	.3667***	.2742**	.2742**	.2595**				
	{0.120}	{0.127}	{0.127}	{0.124}				
lagged log of brain drain (d=200)					.3387***	.2986***	.2878**	.2929**
					{0.107}	{0.111}	{0.113}	{0.112}
log of corruption (authors)	.2209***	.1581**	.1581**	.1672**	.2558***	.2113**	.1786**	.195**
	{0.076}	{0.071}	{0.071}	{0.073}	{0.087}	{0.079}	{0.079}	{0.081}
quality university	.1702**	.07695	.07695	.07514	.2495***	.1455*	.1646**	.1445*
	{0.070}	{0.082}	{0.082}	{0.081}	{0.083}	{0.076}	{0.073}	{0.078}
hospital migration		.4045***	.4045***	.361***		.3623***	.3996***	.3431***
		{0.119}	{0.119}	{0.109}		{0.090}	{0.103}	{0.093}
dummies for airport								
TAV & port	NO	YES	YES	NO	NO	NO	YES	NO
law enforcement	NO	NO	NO	YES	NO	NO	NO	YES
N	378	378	378	378	378	378	378	378
A-B test(1)	0.002	0.003	0.003	0.002	0.000	0.001	0.001	0.001
A-B test(2)	0.168	0.180	0.180	0.204	0.216	0.142	0.139	0.134
Hansen (p-value)	0.083	0.098	0.098	0.167	0.055	0.053	0.084	0.076
# instruments	24	28	28	30	23	23	26	26
# groups	54	54	54	54	54	54	54	54

Table B.7 *Results with corruption (backward orthogonal deviation)*

	I	II	III	IV	V	VI	VII
lagged log of brain drain (t=120)	.3667**	.3063**	.2938*	.2845*			
	{0.153}	{0.149}	{0.150}	{0.147}			
lagged log of brain drain (d=200)					.4597	.3544	.3044
					{0.275}	{0.298}	{0.308}
log of corruption	.22**	.1966*	.193*	.2024**	.212**	.1945**	.1783*
	{0.095}	{0.098}	{0.098}	{0.099}	{0.088}	{0.093}	{0.103}
log of population	-.1819	-.05857	-.02337	-.141	-.1788	-.1166	-.1809
	{0.194}	{0.155}	{0.196}	{0.165}	{0.310}	{0.259}	{0.298}
log of real GDP per capita	.5002	.03448	-.2946	.3294	.5342	.2607	-.0353
	{0.795}	{0.881}	{0.859}	{0.882}	{0.784}	{0.776}	{0.803}
employment	-.03295	-.02431	-.01394	-.03515	-.03646	-.03116	-.02372
	{0.027}	{0.026}	{0.025}	{0.025}	{0.027}	{0.027}	{0.028}
unisize1	-.1868	-.009032	-.02733	-.05761	-.07519	.1133	.06502
	{0.233}	{0.243}	{0.247}	{0.245}	{0.299}	{0.319}	{0.316}
unisize2	-.1013	-.03186	-.1242	-.01181	.01893	.1063	.04411
	{0.168}	{0.175}	{0.197}	{0.188}	{0.147}	{0.172}	{0.200}
quality university	.1623**	.1666**	.175**	.1653**	.1827	.2138*	.2359*
	{0.077}	{0.083}	{0.085}	{0.082}	{0.115}	{0.121}	{0.128}
hospital migration		.26**	.3124**	.2559**		.2362*	.2815*
		{0.118}	{0.131}	{0.114}		{0.136}	{0.158}
dummies for airport, TAV & port	NO	NO	YES	NO	NO	NO	YES
law enforcement	NO	NO	NO	YES	NO	NO	NO
N	378	378	378	378	378	378	378
A-B test(1)	0.005	0.005	0.006	0.003	0.031	0.064	0.083
A-B test(2)	0.173	0.188	0.174	0.211	0.316	0.300	0.298
Hansen (p-value)	0.194	0.116	0.162	0.118	0.242	0.176	0.155
# instruments	33	33	37	39	28	28	31
# groups	54	54	54	54	54	54	54

Note**:** The lagged dependent variables, *log. brain drain d=200* and *log. brain drain d=200 (t-1)* are treated as predetermined variables (GMM-style option in xtabond2), while the other independent variables are treated as exogenous (IV-style option of command xtabond2). A maximum of six lags are used as instruments for the GMM-style endogenous variable. Time dummies are included in all specifications. Standard errors reported in parenthesis are heteroskedasticity-robust. ***, ** and * denote coefficients are significant at 1%, 5% and 10%. Thresholds of distance and time are expressed in km and minutes, respectively. Results are in **b/se***. STATA "ort" option of `xtabond2` is used.

Table B.8 *Results with corruption (authors) (backward orthogonal deviation)*

	IX	X	XI	XII	XIII	XIV	XV	XVI
lagged log of brain drain (t=120)	.3894**	.3122**	.3122**	.3071**				
	{0.150}	{0.151}	{0.151}	{0.146}				
lagged log of brain drain (d=200)					.3607***	.3266***	.3122***	.3291***
					{0.109}	{0.106}	{0.110}	{0.106}
log of corruption (authors)	.1451**	.1183*	.1183*	.1327*	.1655**	.1439*	.1269	.1519*
	{0.067}	{0.068}	{0.068}	{0.071}	{0.079}	{0.078}	{0.077}	{0.080}
log of population	-.2015	-.05337	-.05337	-.1704	-.2587	-.1504	-.2008	-.2415
	{0.183}	{0.186}	{0.186}	{0.157}	{0.232}	{0.196}	{0.211}	{0.176}
log of real GDP per capita	.4352	-.3045	-.3045	.3028	.5339	.2155	-.06363	.4059
	{0.762}	{0.841}	{0.841}	{0.858}	{0.808}	{0.815}	{0.818}	{0.818}
employment	-.03169	-.0134	-.0134	-.0346	-.04102*	-.03132	-.02259	-.04277*
	{0.026}	{0.025}	{0.025}	{0.025}	{0.023}	{0.024}	{0.025}	{0.023}
unisize1	-.177	-.02405	-.02405	-.05565	-.0572	.1286	.06972	.1247
	{0.222}	{0.238}	{0.238}	{0.233}	{0.321}	{0.289}	{0.292}	{0.288}
unisize2	-.08988	-.1153	-.1153	-.006186	.05004	.1236	.05115	.1708
	{0.162}	{0.194}	{0.194}	{0.182}	{0.158}	{0.156}	{0.189}	{0.163}
quality university	.1472*	.163*	.163*	.1544*	.2098**	.2178**	.2289**	.2228***
	{0.076}	{0.085}	{0.085}	{0.082}	{0.084}	{0.085}	{0.086}	{0.081}
hospital migration		.3073**	.3073**	.249**		.2439**	.2781**	.2319**
		{0.129}	{0.129}	{0.111}		{0.103}	{0.121}	{0.096}
dummies for airport, TAV & port	NO	YES	YES	NO	NO	NO	YES	NO
law enforcement	NO	NO	NO	YES	NO	NO	NO	YES
N	378	378	378	378	378	378	378	378
A-B test(1)	0.005	0.007	0.007	0.003	0.000	0.000	0.000	0.000
A-B test(2)	0.161	0.164	0.164	0.197	0.188	0.160	0.158	0.149
Hansen (p-value)	0.181	0.134	0.134	0.109	0.175	0.137	0.148	0.121
# instruments	33	37	37	39	30	30	33	33
# groups	54	54	54	54	54	54	54	54

Note**:** The lagged dependent variables, *log. brain drain d=200* and *log. brain drain d=200 (t-1)* are treated as predetermined variables (GMM-style option in xtabond2), while the other independent variables are treated as exogenous (IV-style option of command xtabond2). A maximum of six lags are used as instruments for the GMM-style endogenous variable. Time dummies are included in all specifications. Standard errors reported in parenthesis are heteroskedasticity-robust. ***, ** and * denote coefficients are significant at 1%, 5% and 10%. Thresholds of distance and time are expressed in km and minutes, respectively. Results are in **b/se***. STATA “ort” option of `xtabond2` is used.

Chapter 3 - Appendix C

Table C.1 *Data Sources*

Variables	***Notes***	***Source***
enrolled	number of resident students who enrol from origin province, with one local university, to university of destination provinces	MIUR
enrol erc1	number of resident students who enrol to courses of Social Science	MIUR
enrol erc2	number of resident students who enrol to courses of Physical Science	MIUR
enrol erc3	number of resident students who enrol to courses of Life Science	MIUR
time	time expressed in minutes to travel by car from origin to destination	ISTAT
pop/pop_j	average annual population of origin/destination	ISTAT
corruption/corruption_j	corruption of origin/destination vs PA (art 314-322 Italian penal law)	RE.GE ISTAT
rgdppc/rgdppc_j	real GDP per capita origin/destination, base GDP year 2010	ISTAT
employment/employment_j	employment rate of origin/destination	ISTAT
uni size/uni_size_j	number of enrolled =1 small, =2 medium, =3 large origin/destination	CENSIS
zquniv/zquniv_j	standardized value of quality of university of origin/destination	AlmaLaurea
zqlife/zqlife_j	standardized value of quality of life for origin/destination	ISTAT
Dairport/Dairport_j	dummy for airport, D=1 airport presence, D=0 otherwise	ISTAT
DTAV/DTAV_j	dummy for high-speed train, D=1 HST presence, D=0 otherwise	Ferrovie Stato
Dport/Dport_j	dummy for port, D=1 port presence, D=0 otherwise	ISTAT
Dnorth/Dnorth_j	dummy for North macro area origin/destination	ISTAT
Dcentre/Dcentre_j	dummy for Centre macro area origin/destination	ISTAT
Dsouth/Dsouth_j	dummy for South macro area origin/destination	ISTAT
disccrim/disccrim_j	variable indicator of law enforcement of origin/destination	ISTAT
probofconv/probofconv_j	variable indicator of law enforcement of origin/destination	ISTAT

Table C.2 Descriptive Statistics

Variable	***Obs.***	***Mean***	***Std. Dev.***	***Min***	***Max***
enrolledp1	20,808	67.76812	629.636	0	21007
enrol_erc1	20,808	35.98678	349.8259	0	11929
enrol_erc2	20,808	20.06695	191.2988	0	6649
enrol_erc3	20,808	11.71631	94.28154	0	2725
time	20,808	390.6581	244.3646	0	1026.9
pop	20,808	806643.1	783995.2	126202	4355725
pop_j	20,808	806643.1	783995.2	126202	4355725
corr_tot	20,808	41.3848	53.14326	1	413
corr_tot_j	20,808	41.3848	53.14326	1	413
rgdppc	20,808	26699.84	7906.145	12831.49	54500.27
rgdppc_j	20,808	26699.84	7906.145	12831.48	54500.27
employment	20,808	56.67339	10.4165	3.623.432	72.9
employment_j	20,808	56.67339	10.4165	3.623.432	72.9
uni_size	20,808	2.098039	.7209889	1	3
uni_size_j	20,808	2.098039	.7209889	1	3
zquniv	20,808	-1.04e-09	.3973304	-.7653301	183.594
zquniv_j	20,808	2.85e-10	.3568086	-.7678611	1.829
zqlife	20,808	.0026567	.2734514	-.7520893	1.398.549
zqlife_j	20,808	-1.71e-09	.3245846	-.843572	.9806968
disccrim	20,808	21.42328	5.567764	10.2	39.2
disccrim_j	20,808	21.42328	5.567764	10.2	39.2
probofconv	20,808	.2744877	.0990663	.042898	.6575092
probofconv_j	20,808	.2744877	.0990663	.042898	.6575091
Dairport	20,808	.5196078	.4996274	0	1
Dairport_j	20,808	.5196078	.4996274	0	1
DTAV	20,808	.1372549	.3441245	0	1
DTAV_j	20,808	.1372549	.3441245	0	1
Dport	20,808	.254902	.4358166	0	1
Dport_j	20,808	.254902	.4358166	0	1
Dnorth	20,808	.3921569	.4882431	0	1
Dnorth_j	20,808	.3921569	.4882431	0	1
Dcentre	20,808	.1960784	.3970381	0	1
Dcentre_j	20,808	.1960784	.3970381	0	1
Dsouth	20,808	.4117647	.4921648	0	1
Dsouth_j	20,808	.4117647	.4921648	0	1
macroarea	20,808	2.019608	.8964239	1	3
macroarea_j	20,808	2.019608	.8964239	1	3

Table C.3 *Italian Provinces, Regions and Macroarea of origin and destination*[66]

Province/Province_j	***Region/Region_j***	***Macroarea/Macroarea_j***
Torino	Piedmont	North
Vercelli	Piedmont	North
Novara	Piedmont	North
Cuneo	Piedmont	North
Asti	Piedmont	North
Alessandria	Piedmont	North
Biella	Piedmont	North
Verbania	Piedmont	North
Aosta	Aosta Valley	North
Imperia	Ligury	North
Savona	Ligury	North
Genova	Ligury	North
La Spezia	Ligury	North
Varese	Lombardy	North
Como	Lombardy	North
Sondrio	Lombardy	North
Milano	Lombardy	North
Bergamo	Lombardy	North
Brescia	Lombardy	North
Pavia	Lombardy	North
Cremona	Lombardy	North
Mantova	Lombardy	North
Lecco	Lombardy	North
Lodi	Lombardy	North
Monza	Lombardy	North
Bolzano	Trentino-Alto Adige	North
Trento	Trentino-Alto Adige	North
Verona	Veneto	North
Vicenza	Veneto	North
Belluno	Veneto	North
Treviso	Veneto	North
Venezia	Veneto	North
Padova	Veneto	North
Rovigo	Veneto	North
Udine	Friuli-Venezia Giulia	North
Gorizia	Friuli-Venezia Giulia	North
Trieste	Friuli-Venezia Giulia	North
Pordenone	Friuli-Venezia Giulia	North
Piacenza	Emilia-Romagna	North

[66] Data source is provided by ISTAT

Parma	Emilia-Romagna	North
Reggio nell'Emilia	Emilia-Romagna	North
Modena	Emilia-Romagna	North
Bologna	Emilia-Romagna	North
Ferrara	Emilia-Romagna	North
Ravenna	Emilia-Romagna	North
Forlì-cesena	Emilia-Romagna	North
Rimini	Emilia-Romagna	North
Pesaro	Marche	Centre
Ancona	Marche	Centre
Macerata	Marche	Centre
Fermo	Marche	Centre
Ascoli Piceno	Marche	Centre
Massa Carrara	Tuscany	Centre
Lucca	Tuscany	Centre
Pistoia	Tuscany	Centre
Firenze	Tuscany	Centre
Livorno	Tuscany	Centre
Pisa	Tuscany	Centre
Arezzo	Tuscany	Centre
Siena	Tuscany	Centre
Grosseto	Tuscany	Centre
Prato	Tuscany	Centre
Perugia	Umbria	Centre
Terni	Umbria	Centre
Viterbo	Lazio	Centre
Rieti	Lazio	Centre
Roma	Lazio	Centre
Latina	Lazio	Centre
Frosinone	Lazio	Centre
Caserta	Campania	South
Benevento	Campania	South
Napoli	Campania	South
Avellino	Campania	South
Salerno	Campania	South
L'Aquila	Abruzzo	South
Teramo	Abruzzo	South
Pescara	Abruzzo	South
Chieti	Abruzzo	South
Campobasso	Molise	South
Isernia	Molise	South
Foggia	Puglia	South
Bari	Puglia	South
Taranto	Puglia	South

Brindisi	Puglia	South
Lecce	Puglia	South
Trani	Puglia	South
Potenza	Basilicata	South
Matera	Basilicata	South
Cosenza	Calabria	South
Crotone	Calabria	South
Vibo Valentia	Calabria	South
Catanzaro	Calabria	South
Reggio di Calabria	Calabria	South
Trapani	Sicily	South
Palermo	Sicily	South
Messina	Sicily	South
Agrigento	Sicily	South
Caltanissetta	Sicily	South
Enna	Sicily	South
Catania	Sicily	South
Ragusa	Sicily	South
Siracusa	Sicily	South
Sassari	Sardinia	South
Nuoro	Sardinia	South
Cagliari	Sardinia	South
Oristano	Sardinia	South
Olbia-Tempio	Sardinia	South
Ogliastra	Sardinia	South
Medio Campidano	Sardinia	South
Carbonia Iglesias	Sardinia	South

Notes to Table C.3

- The nomenclature used for identifying the Italian Provinces follows the one provided by ISTAT. Specifically, this work uses the nomenclature of the edition 2016, where, the new province of Sud Sardinia, ante 2016, results to be divided into four provinces of Olbia-Tempio, Ogliastra, Medio Campidano and Carbonia Iglesias.
- For year 2017, we continue to use the divided provinces of Sud Sardinia by diving the number of enrolled students of Sud Sardinia by four and giving higher weights (in terms of number of students) to provinces with higher population density rate: Carbonia-Iglesias presented the highest rate while Ogliastra had the lowest one.

Table C.4 *Italian Provinces for origin and destination with university*[67]

Province/Province_j	***University***	***Type of University***
Torino	Università degli Studi di Torino	Public
Torino	Politecnico di Torino	Public
Torino	Bra Scienze Gastronomiche	Public
Torino	Università degli Studi del Piemonte Orientale	Public
Aosta	Università degli Studi di Aosta	Public
Genova	Università degli Studi di Genova	Public
Milano	Castellanza LIUC	Private
Milano	Università degli Studi di Milano	Public
Milano	Politecnico di Milano	Public
Milano	Università Bocconi	Private
Milano	Università Cattolica	Private
Milano	IULM	Private
Milano	Università degli Studi di Milano Bicocca	Public
Milano	Università Humanitas Rozzano	Private
Brescia	Università degli Studi di Brescia	Public
Bergamo	Università degli Studi di Bergamo	Public
Pavia	Università degli Studi di Pavia	Public
Trento	Università degli Studi di Trento	Public
Verona	Università degli Studi di Verona	Public
Venezia	Cà Foscari	Public
Venezia	Iuav- Tolentini	Public
Padova	Università degli Studi di Padova	Public
Udine	Università degli Studi di Udine	Public
Trieste	Università degli Studi di Trieste	Public
Parma	Università degli Studi di Parma	Public
Modena	Università degli Studi di Modena e Reggio Emilia	Public
Reggio nell'Emilia	Università degli Studi di Modena e Reggio Emilia	Public
Bologna	Alma Mater Studiorum -Università di Bologna	Public
Ferrara	Università degli Studi di Ferrara	Public
Pesaro	Università degli Studi di Urbino	Public
Urbino	Università degli Studi di Urbino	Public
Ancona	Università degli Studi delle Marche	Public
Macerata	Università degli Studi di Macerata	Public
Ascoli Piceno	Università di Camerino	Public
Firenze	Università degli Studi di Firenze	Public
Pisa	Università degli Studi di Pisa	Public
Siena	Università degli Studi di Siena	Public
Siena	Università per Stranieri di Siena	Public
Perugia	Università degli Studi di Perugia	Public

[67] Data source is provided by M.I.U.R

Perugia	Università per Stranieri di Perugia	Public
Viterbo	Università degli Studi della Tuscia	Public
Roma	Università Roma "La Sapienza"	Public
Roma	Università Roma "Tor Vergata"	Public
Roma	Libera Università SS. Maria Assunta - LUMSA	Private
Roma	Libera Università degli Studi Sociali - LUISS Guido Carli	Private
Roma	Università degli Studi di Roma Foro Italico	Public
Roma	Università degli Studi "Roma Tre"	Public
Roma	Università Campus Bio Medico di Roma	Private
Roma	Università degli Studi Internazionali di Roma - UNINT	Private
Roma	UER - Università Europea di Roma	Private
Frosinone	Università degli Studi di Cassino	Public
Benevento	Università degli Studi del Sannio	Public
Napoli	Università "Federico II" di Napoli	Public
Napoli	Università Parthenope di Napoli	Public
Napoli	Università degli Studi di Napoli "L'Orientale"	Public
Napoli	Università Suor Orsola Benincasa	Private
Napoli	Università degli Studi della Campania "L. Vanvitelli"	Public
Salerno	Università degli Studi di Salerno	Public
L'Aquila	Università degli Studi dell'Aquila	Public
Teramo	Università degli Studi di Teramo	Public
Chieti	Università degli Studi di Chieti e Pescara	Public
Pescara	Università degli Studi di Chieti e Pescara	Public
Molise	Università degli Studi del Molise	Public
Foggia	Università degli Studi di Foggia	Public
Bari	Università degli Studi di Bari	Public
Bari	Politecnico di Bari	Public
Bari	Università LUM Jean Monnet	Private
Lecce	Università del Salento	Public
Potenza	Università degli Studi della Basilicata	Public
Cosenza	Università della Calabria	Public
Catanzaro	Università degli Studi di Catanzaro "Magna Graecia"	Public
Reggio di Calabria	Università degli Studi Mediterranea di Reggio Calabria	Public
Reggio di Calabria	Università per Stranieri "Dante Alighieri"	Private
Palermo	Università degli Studi di Palermo	Public
Messina	Università degli Studi di Messina	Public
Enna	Università KORE di Enna	Public
Catania	Università degli Studi di Catania	Public
Sassari	Università degli Studi di Sassari	Public
Cagliari	Università degli Studi di Cagliari	Public

Notes to Table C.4

1. The provinces reported with university are 51 for origin
2. Adjustments for “Modena and Reggio nell’Emilia” and “Chieti and Pescara” have been made. The campus of such universities is placed in both cities, and we divide them respectively. Hence, the study divides the number of enrolled students by half, giving more weight (in terms of number of students) to the province that presents the highest population density rate (for example, higher number of enrolled students is attributed to Modena because its population density rate is higher than the one present by Reggio nell’Emilia. Also, higher number of enrolled students is given to Pescara because its population density rate is higher than the one presented by Chieti)
3. The analysis does not consider Telematic Universities, Schools of Superior Specialization and/or Schools of Excellence
4. The Academic Years evaluated starts from 2010-2011 to 2017-2018
5. This study reports a single-year format for the Academic Year, beginning with 2010 for the A.Y. 2010-2011 and ends with 2017 for the A.Y. 2017-2018

19Table C.5 *Specifications for ERC-Study Field*[68]

ERC	***Denomination***	***Disciplines***
1	*Social Sciences* (**SH**)	Economics, Finance, Management, Sociology, Social Anthropology, Political Science, Law, Communication, Psychology and Human Behaviour
2	*Physical Sciences* (**PE**)	Mathematics, Physics, Chemistry, Computer Sciences and Informatics, Systems and Communication Engineering, Product and Processes Engineering, Universe Sciences and Astrophysics, Climatology, Ecology, Biogeochemistry
3	*Life Sciences* (**LS**)	Molecular and Structural Biology and Biochemistry, Genetics and Genomics, Cellular and Developmental Biology, Physiology, Pathophysiology and Endocrinology, Neurosciences and Neural Disorders, Immunity and Infection, Aetiology, Tropical Medicine, Public Health, Epidemiology, Pharmacology, Toxicology, Regenerative Medicine, Medical Ethics, Biodiversity, Biogeography, Marine Biology, Eco-toxicology, Microbial ecology, Population Biology, Biotechnology, Genetic Engineering, Synthetic and Chemical Biology, Industrial Biosciences, Environmental Biotechnology

[68] Data source is provided by M.I.U.R

Table C.6 *Robustness Check with corruption (authors)*

	ZIP	**ZIP**	**ZIP**	**ZIP**	**PPML**	**PPML**	**PPML**	**PPML**
	I	II	III	IV	V	VI	VII	VIII
	Enrolled	**ERC-1**	**ERC-2**	**ERC-3**	**Enrolled**	**ERC-1**	**ERC-2**	**ERC-3**
time	-.02212***	-.02274***	-.01709***	-.01503***	-.02659***	-.02837***	-.02565***	-.02415***
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
population	3.32e-07***	4.10e-07***	3.76e-07***	1.30e-07	4.08e-07***	4.58e-07***	4.07e-07***	2.75e-07**
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
population_j	2.78e-07***	1.88e-07**	2.30e-07*	4.35e-07***	1.84e-07**	1.17e-07	2.03e-07*	3.21e-07***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
corruption (authors)	.0002052	-4.98e-06	.0006322*	-.0000861	.0002298	.000044	.0007199**	-.0001829
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
corruption_j (authors)	-.0009411***	-.0006089***	-.001466***	-.0005728*	-.0009606***	-.0006096***	-.001756***	-.0005812*
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
real GDP per capita	9.72e-06	9.32e-06	-.0000111	2.23e-06	.0000164	.0000157	.0000154	.0000166
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
real GDP per capita_j	-4.78e-06	5.76e-06	4.15e-06	-9.58e-06	-6.10e-06	7.04e-06	-.0000167	-.0000223
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
employment	-.0394***	-.03128**	-.03423**	-.03053**	-.04914***	-.04736***	-.04024***	-.06512***
	{0.014}	{0.014}	{0.014}	{0.014}	{0.013}	{0.014}	{0.016}	{0.015}
employment_j	.01188	-.0004424	.001175	.005166	.02295	.0194	.01396	.04405***
	{0.015}	{0.015}	{0.017}	{0.015}	{0.015}	{0.015}	{0.019}	{0.017}
size university	-.1422	-.1146	-.04921	-.001058	-.2908***	-.2476***	-.38***	-.2125*
	{0.087}	{0.084}	{0.107}	{0.108}	{0.081}	{0.081}	{0.101}	{0.116}
size university_j	.3714***	.1945**	.5907***	.2199*	.5266***	.3456***	.8855***	.4571***
	{0.096}	{0.091}	{0.127}	{0.122}	{0.092}	{0.088}	{0.114}	{0.138}
tandardized quality universit	-.4055**	-.3814**	-.2796	-.5808***	-.4597***	-.5083***	-.268	-.6423***
	{0.169}	{0.166}	{0.217}	{0.193}	{0.161}	{0.169}	{0.209}	{0.237}
andardized quality university	-.03145	-.003302	-.4128	.2691	.09776	.21	-.4032*	.5176**
	{0.178}	{0.174}	{0.258}	{0.217}	{0.157}	{0.168}	{0.220}	{0.221}
standardized quality life	-.4098	-.3219	-.6647**	-.2481	-.3238	-.1047	-.7461**	-.2273
	{0.278}	{0.265}	{0.304}	{0.357}	{0.276}	{0.260}	{0.309}	{0.416}
standardized quality life_j	.2914	.008505	.596**	.2286	.2421	-.22	.8782***	.4736
	{0.243}	{0.264}	{0.275}	{0.314}	{0.238}	{0.259}	{0.269}	{0.346}
Dairport	.04438	.0374	.0537	.1916	.04682	.05849	.04478	.03084
	{0.116}	{0.106}	{0.138}	{0.154}	{0.114}	{0.113}	{0.138}	{0.159}
Dairport_j	.1213	.1279	.02262	.3144**	.09975	.06621	-.02341	.3614**
	{0.131}	{0.118}	{0.168}	{0.153}	{0.131}	{0.125}	{0.160}	{0.183}
DTAV	-.06265	-.1052	.07402	.1906	-.2642	-.3053*	-.3629	-.05547
	{0.190}	{0.183}	{0.228}	{0.195}	{0.172}	{0.162}	{0.221}	{0.195}
DTAV_j	.1609	.2234	-.01213	-.2522	.4802***	.5309***	.5941***	.2025
	{0.180}	{0.174}	{0.231}	{0.229}	{0.154}	{0.147}	{0.192}	{0.194}
Dport	.3089**	.3129**	.2095	.18	.3936***	.4324***	.2719*	.5009***
	{0.123}	{0.137}	{0.155}	{0.161}	{0.119}	{0.134}	{0.144}	{0.162}
Dport_j	-.1028	-.1321	.05628	-.1505	-.1461	-.2159	.1259	-.3699*
	{0.148}	{0.160}	{0.196}	{0.193}	{0.152}	{0.165}	{0.193}	{0.208}
N	20808	20808	20808	20808	20808	20808	20808	20808
Wald Chi Square Test	5557.91	6050.69	3547.11	2160.71	3753.81	3856.26	2376.61	2399.67
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo-R2					0.8717	0.8880	0.8282	0.8063

Notes: For both model's specifications, we inserted law deterrence variables for origin and destination provinces as control values. In addition, fixed effects of dummies for macro-areas of North, Centre and South of origin and destination are included with interaction with year (2010-2017 period). Standard errors account for clustering of provinces. ***, ** and * denote coefficients that are statistically significant at 1%, 5% and 10%, respectively. Results are in **b/se***

Table C.7 *Robustness Check with the interaction of distance and infrastructures*

	ZIP I Enrolled	ZIP II ERC-1	ZIP III ERC-2	ZIP IV ERC-3	PPML V Enrolled	PPML VI ERC-1	PPML VII ERC-2	PPML VIII ERC-3
time	-.04807***	-.04631***	-.04019***	-.04538***	-.04739***	-.04522***	-.04314***	-.0523***
	{0.006}	{0.006}	{0.010}	{0.006}	{0.006}	{0.006}	{0.011}	{0.007}
population	2.96e-07***	4.19e-07***	3.18e-07***	1.52e-07	3.09e-07***	4.55e-07***	1.58e-07	1.78e-07
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
population_j	2.30e-07***	1.01e-07	2.00e-07**	3.74e-07***	2.36e-07***	8.35e-08	3.76e-07***	3.79e-07***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
corruption	.0004787	.0000645	.0006148	-.0001855	.001133*	.0004493	.002511***	.0005969
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
corruption_j	-.001822***	-.001111**	-.002444***	-.001049	-.002345***	-.001291**	-.004413***	-.001719**
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
real GDP per capita	-.0000117	-.0000119	-.0000219*	-.0000191	-.0000117	-.0000182	-3.94e-06	-.0000122
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
real GDP per capita_j	.0000148	.0000255**	.0000161	5.90e-06	.0000175	.0000367***	-1.04e-06	1.69e-07
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
employment	-.02789**	-.0207	-.0169	-.01051	-.03515***	-.02854**	-.03443**	-.04897***
	{0.013}	{0.013}	{0.014}	{0.015}	{0.013}	{0.013}	{0.014}	{0.017}
employment_j	.006193	-.00674	-.006875	-.005997	.0136	.003827	.01241	.03414*
	{0.013}	{0.013}	{0.016}	{0.016}	{0.013}	{0.014}	{0.017}	{0.018}
size university	-.1986**	-.1655*	-.2351***	-.03884	-.3281***	-.2756***	-.4881***	-.1888
	{0.090}	{0.091}	{0.090}	{0.111}	{0.085}	{0.088}	{0.090}	{0.133}
size university_j	.3901***	.2182**	.6791***	.1894	.5242***	.3428***	.9566***	.3812**
	{0.100}	{0.096}	{0.112}	{0.125}	{0.099}	{0.097}	{0.116}	{0.158}
standardized quality university	-.5734**	-.5638**	-.3478	-.6193**	-.6981***	-.7616***	-.4729*	-.8689***
	{0.230}	{0.222}	{0.282}	{0.242}	{0.211}	{0.210}	{0.263}	{0.291}
standardized quality university_j	.06558	.1266	-.343	.2868	.2472	.386*	-.269	.6246**
	{0.225}	{0.216}	{0.301}	{0.258}	{0.194}	{0.201}	{0.256}	{0.260}
standardized quality life	-.4633*	-.3041	-.7904***	-.2572	-.4385	-.1699	-.9315***	-.3655
	{0.279}	{0.282}	{0.296}	{0.342}	{0.276}	{0.278}	{0.310}	{0.410}
standardized quality life_j	.3069	-.05663	.6693**	.2851	.3226	-.1733	.9733***	.6201*
	{0.253}	{0.285}	{0.268}	{0.315}	{0.250}	{0.282}	{0.272}	{0.368}
1.Dairport#c.dist	-.002864***	-.002638***	-.002057***	-.002624***	-.003034***	-.002775***	-.002864***	-.003528***
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
1.Dairport_j#c.dist	.002299*	.001325	.003318***	-.0001714	.002382*	.001418	.005683***	.0001322
	{0.001}	{0.002}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.002}
1.DTAV#c.dist	-.007416***	-.007973***	-.00646***	-.004531**	-.008064***	-.008791***	-.008332***	-.006684***
	{0.002}	{0.002}	{0.001}	{0.002}	{0.002}	{0.002}	{0.002}	{0.002}
1.DTAV_j#c.dist	.001957*	.003265**	.0006851	.0008236	.00516***	.007292***	.003464***	.004213***
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
1.Dport#c.dist	.004399***	.004536***	.003453***	.002825***	.005443***	.005883***	.005302***	.004756***
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
1.Dport_j#c.dist	-.005315***	-.004289**	-.008261***	-.002481*	-.004747***	-.003487**	-.00959***	-.002258
	{0.002}	{0.002}	{0.002}	{0.001}	{0.002}	{0.002}	{0.002}	{0.001}
N.obs	20808	20808	20808	20808	20808	20808	20808	20808
Wald chi2 test	1110.97	10866.70	8692.09	6128.40	6224.91	6895.78	3952.55	4292.53
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo-R2					0.9102	0.9207	0.8833	0.8396

Notes: For both model's specifications, we inserted law deterrence variables for origin and destination provinces as control values. In addition, fixed effects of dummies for macro-areas of North, Centre and South of origin and destination are included with interaction with year (2010-2017 period). Standard errors account for clustering of provinces.***, ** and * denote coefficients that are statistically significant at 1%, 5% and 10%, respectively. Results are in **b/se***

Table C.8 *Robustness Check with long-distance skilled movements from Centre-South to North*

	ZIP	**ZIP**	**ZIP**	**ZIP**	**PPML**	**PPML**	**PPML**	**PPML**
	I	II	III	IV	V	VI	VII	VIII
	Enrolled	**ERC-1**	**ERC-2**	**ERC-3**	**Enrolled**	**ERC-1**	**ERC-2**	**ERC-3**
time	-.01109***	-.01196***	-.007201***	-.006969***	-.01519***	-.01787***	-.0134***	-.01401***
	{0.002}	{0.002}	{0.002}	{0.002}	{0.002}	{0.002}	{0.002}	{0.002}
population	4.45e-08	-2.23e-09	1.78e-07	6.15e-08	2.17e-08	-1.80e-08	7.86e-08	1.61e-07
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
population_j	6.59e-07**	5.59e-07**	5.21e-07**	7.53e-07***	8.63e-07***	7.15e-07***	9.84e-07***	8.70e-07***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
corruption	.002295**	.002142*	.002016*	.001052	.002933***	.002663**	.003606***	.000986
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
corruption_j	-.002598**	-.0004577	-.00346***	-.002794**	-.004281***	-.001389*	-.007296***	-.002392**
	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}	{0.001}
real GDP per capita	.0000512	.000068	2.40e-06	-.0000169	.0000644	.0000705	.0000877	.0000196
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
real GDP per capita_j	-.0000366	-2.95e-06	-.0000483*	-.0000325	-.000018	.0000468*	-.0001072***	.0000505
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
employment	-.05295**	-.06446**	-.03964	.005224	-.06342**	-.07905***	-.05653*	-.05437*
	{0.026}	{0.026}	{0.027}	{0.022}	{0.028}	{0.028}	{0.031}	{0.029}
employment_j	.06794	.0618	.02415	.05548	.06271	.0729*	.04156	.07148*
	{0.046}	{0.042}	{0.035}	{0.040}	{0.040}	{0.041}	{0.044}	{0.038}
size university	-.1471	-.09926	.1235	.0003323	-.4015**	-.3617*	-.5229**	-.2718
	{0.198}	{0.211}	{0.189}	{0.169}	{0.189}	{0.201}	{0.204}	{0.174}
size university_j	.1024	-.102	.2694	-.1101	.3091	.008161	.7892***	-.004971
	{0.239}	{0.231}	{0.262}	{0.251}	{0.226}	{0.220}	{0.235}	{0.230}
standardized quality university	-.4872	-.2808	-.2667	-1.114***	-.5163	-.5007	-.2516	-1.128**
	{0.404}	{0.445}	{0.368}	{0.346}	{0.416}	{0.450}	{0.433}	{0.454}
standardized quality university_j	-.03791	-.09394	-.9129*	-.6405	.6538*	.9994**	.3474	.4144
	{0.492}	{0.492}	{0.491}	{0.461}	{0.361}	{0.420}	{0.345}	{0.354}
standardized quality life	-.7992	-.4636	-.8644	-.4922	-.8757	-.4774	-1.412**	-.3702
	{0.614}	{0.584}	{0.608}	{0.551}	{0.535}	{0.518}	{0.579}	{0.561}
standardized quality life_j	.7041	.0855	.707	1.506***	.3916	-.1663	.8206**	1.527***
	{0.430}	{0.444}	{0.456}	{0.422}	{0.339}	{0.380}	{0.371}	{0.342}
Dairport	.04641	.01021	.2715	.5028	.05664	.05363	.05942	-.01651
	{0.314}	{0.310}	{0.277}	{0.311}	{0.314}	{0.310}	{0.336}	{0.324}
Dairport_j	.4811	.1882	.6754*	-.05106	.907***	.4923	1.574***	.7017**
	{0.354}	{0.377}	{0.357}	{0.343}	{0.330}	{0.344}	{0.344}	{0.321}
DTAV	-.4774	-.5029	-.3614	.1536	-.6425	-.6914	-.7786	-.4617
	{0.570}	{0.510}	{0.588}	{0.456}	{0.591}	{0.509}	{0.733}	{0.636}
DTAV_j	.4008	.3239	.5708	.06311	.4189	.3958	.6328	-.1617
	{0.576}	{0.543}	{0.629}	{0.541}	{0.470}	{0.467}	{0.496}	{0.582}
Dport	.3218	.5675*	-.005406	-.3724	.7208**	1.014***	.4679	.5285
	{0.331}	{0.337}	{0.327}	{0.322}	{0.342}	{0.350}	{0.418}	{0.385}
Dport_j	-.4123	-.9637**	.08335	.2207	-1.153***	-1.688***	-.7224*	-1.097**
	{0.406}	{0.377}	{0.393}	{0.425}	{0.418}	{0.421}	{0.385}	{0.441}
N	7440	7440	7440	7440	7440	7440	7440	7440
Wald Chi Square Test	314.527	395.558	245.512	210.900	853.81	743.15	781.21	693.74
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo-R2					0.7947	0.8252	0.7397	0.7408

Notes: For both model's specifications, we inserted law deterrence variables for origin and destination provinces as control values. In addition, fixed effects of dummies for macro-areas of North, Centre and South of origin and destination are included with interaction with year (2010-2017 period). Standard errors account for clustering of provinces. The constraint South-Centre to the North is added as (macroarea≥ 2 & macroarea_j ≤2) ***, ** and * denote coefficients that are statistically significant at 1%, 5% and 10%, respectively Results are in **b/se***

Table C.9 *Bootstrap Results with ZIP in the 2nd stage*

	Obs Coef.	**Bootstrap Std. Err.**	**Z**	**P>|z|**	**Normal Based [95% Conf. Interval]**	
r(b_corruption)	0.0015866	0.0019282	0.80	0.411	-0.0021925	0.0053658
r(b_corruption_j)	-0.0047955	0.0015685	-3.06	0.002	-0.0078698	-0.0017213
r(b_corr_res)	-0.00004	0.0015124	-0.03	0.979	-0.0030042	0.0029242
r(b_corr_res_j)	0.002326	0.0012418	1.87	0.061	-0.0001078	0.0047598

N. observations: 18.207; N. of Replications: 200 based on 2.601 cluster in panelid

Table C.10 *Bootstrap Results with PPML in the 2nd stage*

	Obs Coef.	**Bootstrap Std. Err.**	**z**	**P>|z|**	**Normal Based [95% Conf. Interval]**	
r(b_corruption)	0.0016962	0.0018408	0.92	0.357	-0.0019118	0.0053042
r(b_corruption_j)	0.0053151	0.0016113	-3.30	0.001	-0.0084731	-0.002157
r(b_corr_res)	0.0005872	0.0014817	0.40	0.692	-0.0023169	0.0034913
r(b_corr_res_j)	0.0024111	0.0013222	1.82	0.068	-0.0001803	0.0050025

N. observations: 18.207; N. of Replications: 200 based on 2.601 cluster in panelid

Table C.11 *Bootstrap Results with PPML in the 2nd stage*

	Obs Coef.	**Bootstrap Std. Err.**	**z**	**P>|z|**	**Normal Based [95% Conf. Interval]**	
r(b_corruption)	0.0009282	0.0020431	0.45	0.650	-0.0030762	0.0049326
r(b_corruption_j)	-0.0053502	0.0016848	-3.18	0.001	-0.0086523	-0.002048
r(b_network)	0.000737	0.000147	0.50	0.616	-0.0002144	0.0003617
r(b_corr_res)	0.0007244	0.0016077	0.45	0.652	-0.0024267	0.0038755
r(b_corr_res_j)	0.0027396	0.0014912	1.84	0.066	-0.0001831	0.0056623
r(b_network_res)	0.0001788	0.0002561	0.70	0.485	-0.0003231	0.0006807

N. observations: 15.606; N. of Replications: 200 based on 2.601 cluster in panelid

Table C.12 *Additive Notes*

Theoretical Framework

Bilateral net skilled migration can be theorized with the utility maximization framework combined with the gravity set-up. Assumed that skilled students are rational, they are free to enrol to universities that belong to different provinces and their decision to move will be based on the comparison between the expected utilities of origin (*i*) and destination (*j*). Besides, individual utility is a function that encompasses socio-economic and quality of life variables plus the costs of moving, which are represented by distance between origin (*i*) and destination (*j*). Thus, the utility function for the skilled individual (*s*) at origin (*i*) province can be expressed as:

$$U_i^s = u(E_i, L_i) + \varepsilon_i^s \quad [1]$$

While the utility function for the skilled individual (*s*) at destination (*j*) is expressed as:

$$U_j^s = u(E_j, L_j) + \varepsilon_j^s \quad [2]$$

where the total utility U includes a deterministic part *u* and a stochastic part ε_i^s and ε_j^s The deterministic part *u* is a function of a vector of a wide range of economic (E) and quality of life (L) variables. Students will decide to move from location *i* to location *j* if the expected utility of the destination is greater than the expected utility for the origin plus the costs of relocating (which are expressed as function of distance):

$$E[U_j^s] \geqq E[U_i^s] + C(d_{ij}) \quad [3]$$

Hence, rewriting the above conditions according to the gravity model specification, we get:

$$Enrolled_{ij} = f(E_{ij}L_{ij}D_{ij}) \quad [4]$$

Where i= 1, 2, ..., 51 represents origin provinces with one local university, j= 1, 2, …, 51 represents destination provinces with one local university (with i ≠ j), E is a vector of socio-economic characteristics for origin and destination, L is a vector of quality-of-life characteristics for origin and destination and D_{ij} represents the distance between origin *i* and destination *j*.

Table C.13 *Additive Notes*

Methods used to build-up Quality of University and Quality of Life

The method used to derive quality of university and quality of life variables is a two-steps procedure that consists of taking the average of the standardized values related to the quality of university and quality of life. All variables are aggregated at Italian provincial level from 2010 to 2017.

The values used for creating quality of university are provided by ALMALAUREA and are age, grade, expected income per capita, expected time to find a job. In particular, age indicates how much old are the students when they take the bachelor's degree (3-years course program), grades indicates the final grade that the students achieve once they get graduated, expected income per capita returns an estimate of the income they would earn with their bachelor degree and the expected time to find a job indicates how much time is needed to find a job once the students get their bachelor degree. All these variables reflect the quality of university from students' perspective: in fact, according to students, the quality of university is the highest if they learn skills that are required to graduate early, to find a job easily and to earn a pleasantly income.

In addition, the values used for creating quality of life variable are taken by ISTAT and are mortality rate, working formation, gender difference in employment, the presence of green urban areas, childcare and elderly-care. Mortality rate indicates the incidence of deaths over the Italian population. It is used as proxy for quality of life because it identifies the health status of the designed population. Besides, working formation indicates the incidence of those who participate to working formation programs over the Italian population. It is used as proxy for quality because it represents the alphabetization and technological progresses needed to work in firms. In addition, gender difference in employment indicates how many women works respect to men. This value is used as proxy for indicating the existence of the equality condition within the Italian labour context. Then, green urban area indicates the presence of parks in the urbanized area. This value is exploited as proxy for quality of life because the presence of green area permits to follow good health-habits (breathing unpolluted air, walking, jogging, running, playing etc.) that positively affects individuals' lives. Finally, childcare and elderly care are primary services that cannot be overlooked in civilized societies. These are used as proxies for quality of life and states that high assistance offered to the public is associated with high quality of life where these services operate.

Once these variables are selected and collected, we standardized each of them to make an easier comparison among scores measured on different scales.

The standardization process consists of rescaling variables using the z-score, as expressed in the following formula:

$$z = \frac{X - \mu}{\sigma} \quad [1]$$

The z-score is obtained by i) subtracting the mean, , from the value to be converted, X and ii) dividing the numerator by the standard deviation, , of the denominator. Hence, the standardized values obtained have the mean of 0 and the standard deviation of 1.

Then, we take the arithmetic average of all the z-scores get from Equation (1) for creating the variables of quality of university and quality of life, by using the following formula:

$$Average\ for\ Quality\ of\ University = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{z_1 + z_2 + \ldots + z_n}{n}$$

$$Average\ for\ Quality\ of\ Life = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{z_1 + z_2 + \ldots + z_n}{n}$$

where z_1, z_2 ... z_n represent the standardized values selected and used for creating the variables of quality of university and quality of life, while *n* indicates the number of values inserted for creating the already cited two variables.