



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M. FANNO"**

**CORSO DI LAUREA MAGISTRALE IN
ECONOMICS AND FINANCE**

TESI DI LAUREA

**THE SOCIAL AND ECONOMIC EFFECTS OF DISCRIMINATION
AGAINST IMMIGRANTS.
NEW EMPIRICAL EVIDENCE FROM ISTAT SURVEY DATA.**

RELATORE:

CH.MO PROF. Nunziata Luca

LAUREANDA: Palmarin Elisa

MATRICOLA N. 1131054

ANNO ACCADEMICO 2016 – 2017

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.

Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

ABSTRACT

The thesis uses data drawn from the ISTAT survey “Condizione e Integrazione sociale dei cittadini stranieri” (2016), carried out in 2011 and addressed to foreign citizens resident in Italy. This work has a double objective: to assess whether and which is the effect that perceived discrimination has on migrants’ health and how health, in turn, affects their employment status. In both cases a first attempt to implement instrumental variable strategy is made, so to tackle reverse causality and omitted variable bias. The obtained estimates indicate that perceived discrimination negatively affects health status, both mental and physical, and that it has a negative impact on migrants’ employment status, which is positively correlated with health. Furthermore, a third model concerning the relationship between perceived discrimination and socio-cultural integration is estimated, always adopting an IV approach. The estimates so obtained do not allow a clear interpretation, but there seems to be a positive correlation between discrimination and migrants’ behaviours signalling a higher level of social integration.

CONTENTS

1. Introduction	1
2. Literature review	7
3. Data description	13
4. Empirical strategy	
4.1 An instrumental variable approach.....	33
4.2 Controls.....	35
4.3 Weights.....	37
4.4 Generalised Method of Moments (GMM).....	38
5. Main findings	
5.1 The factors associated with life satisfaction.....	41
5.2 The association between perceived discrimination and health.....	44
5.3 The association between health and employment status.....	51
5.4 The association between perceived discrimination and social integration	55
6. Conclusion	61
References	63

1. Introduction

International migration is a global phenomenon and for some European Countries it can be said to be also a recent one. Italy, together with Spain and Ireland, is part of the so-called new migration Countries, where migration inflows have been consistently growing in the last decades (T.K. Bauer *et al.*, 2000). According to the Eurostat data the share of immigrants over native population was as low as 1.7% in 1998, while the same data for 2017 reports that 8.3% of the population resident in Italy is non-nationals. Italy experienced the first net migration inflow in 1972, after the Northern Countries stopped to encourage immigration following on from social tensions and the fear of recession after the first oil shock. This inflow was mainly due to return migration, but already in the early 1990s the type of migration had changed and most immigrants were non-nationals and, particularly, came from non-EU Countries.

This demographic change affecting some European Countries has important effects on native populations, which perceive both the economic and the socio-cultural consequences of immigration. Feelings and worries of the natives are reflected in the European political arena, where, in these last years, immigration policy has been a central topic in the electoral campaigns. In response to the recent increased migrants' flows, many parties with protectionist views on immigration were born in the European Union. Among them there are the Front National in France, the Dutch Freedom Party, the United Kingdom Independence Party and the Lega Nord (Northern League) in Italy. Marine Le Pen, leader of the Front National, which obtained the 21% of the votes in the French Presidential Election in 2017 and lost at the second ballot, had proposed in February 2014 a referendum for stricter immigration regulation (G. Barone *et al.*, 2014).

Immigration policy involves establishing rules to determine quotas, admission mechanisms, procedures for citizenship acquisition and rules to regulate the inflows of asylum seekers. There exists a summary indicator, developed by the Fondazione Rodolfo De Benedetti, measuring the degree of strictness of migration policy in a Country. This index takes into consideration seven different dimensions of migration legislation, which are likely to affect economic-driven migration from outside the European Union. Four of these seven aspects are: the number of entry requirements, the number of years of stay necessary to obtain a permanent residence permit, the existence of selection migration policies for high-skilled immigrants and, finally, the existence of quotas or point-based systems. This last aspect refers to different policies that can be implemented in order to put a cap to the number of migrants who are accepted by a Country every year. The system of quotas is simply based on the "first-come first-served" principle, while point-based mechanisms, which are adopted by a limited number

of Countries, award points on the basis of education, experience and language abilities. To construct the strictness migration policy index, the scores of each of the above categories are made comparable through a conversion in cardinal numbers, ranging from 1 to 6. The scores so obtained are then averaged out to obtain the final index, which can measure from a minimum of 1 to a maximum of 6, for stricter policies. Table 1 reports the 2010 index for eleven European Union Countries.

Table 1- Strictness of migration policies in major EU immigration destinations (2010)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Country	Quota/PSB system	N° entry requirements	First entrance	N° staying requirements	N° adminin. bodies involved	Years to obtain permanent residence	Selective policies for high-skilled	Overall index 2010
Austria	quota	6	before	4	1 or 2	5	no	4.3
Finland	no	3	before	2	1 or 2	4	limited	2.6
France	no	2	after	2	1 or 2	5	yes	0.7
Germany	no	2	after	4	1	5	yes	1.0
Greece	quota	2	after	5	1	10	limited	3.5
Ireland	quota	4	after	4	2	10	limited	3.8
Italy	quota	3	after	3	1	6	limited	2.5
NL	no	4	before	2	2	5	no	3.0
Portugal	quota	5	after	5	1 or 2	5	yes	3.7
Spain	quota	6	after	7	2	5	limited	3.9
UK	PBS+quota	4	before	3	1	5	limited	3.2

Source: IZA 2011, The Integration of Migrants and its Effects in the Labour Market

Notes:

(2) Number of certificates and procedures to be admitted as a foreigner

(3) Residence permit asked before or after entering the country for the first time

(4) Number of certificates and procedures to legally reside in the territory

(5) Number of administrative bodies involved to obtain a residence and a work permit

(7) Number of categories positively selected (existence of fast-track, simplified procedures or exclusions from quotas).

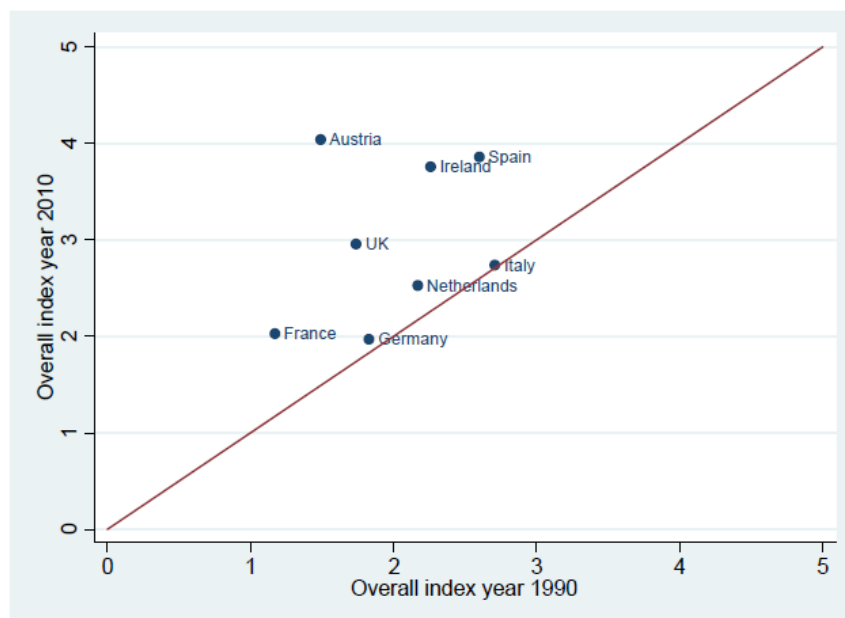
Italy 2010 index measured 2.5, which is one of the lowest values after France and Germany, indicating loose migration policies. What is also relevant to consider is the evolution of the strictness index over time. Figure 1 shows that Italy, as well as many other Countries, is above the 45 degree line, indicating tightened migration restrictions in 2010 with respect to 1990.

The first systematic Italian Immigration Act is known as the Turco-Napolitano Act, after the then Social Affairs and Interior Ministers, and was passed in 1998. It was based on four pillars: preventing and combating illegal entry; regulating new flows of foreign workers; promoting the integration of immigrants holding a valid residence permit; and granting basic individual rights to illegal immigrants. This Law was partially reformed in 2002 by the right-wing government. Law No.189/2002, known as the Bossi-Fini Act, which made migration

policy rules more restrictive (T. Caponio, 2005). For instance, it made it compulsory to have an employment contract before arriving in Italy, as a necessary condition to obtain a residence permit. Moreover, the number of years of stay in Italy necessary to require a permanent residence permit was increased to six, but it was reduced to five again in 2003, as a result of a European Directive on the status of non-EU nationals who are long-term residents (Directive 2003/109/EC).

The strictness of migration policies may represent the first obstacle to the integration of non-EU immigrants within society. Indeed, long and complex procedures to enter a foreign Country or to obtain residence permits hinder the access to the labour market and to the educational system. These obstacles, together with different cultural backgrounds, make it difficult for immigrants to integrate within society.

Figure 1-Evolution of the index of strictness of migration policies in the largest immigration Countries in the EU



Source: IZA 2011, The Integration of Migrants and its Effects in the Labour Market

The Migrant Integration Policy Index (MIPEX) is a tool which measures policies that promote migrants' integration in 38 different Countries: European Union Member States and Australia, Canada, Iceland, Japan, South Korea, New Zealand, Norway, Switzerland, Turkey and the US. The MIPEX is produced thanks to the project "Integration policies: Who benefits? The development and use of indicators in integration debates", which is led by the Barcelona Centre for International Affairs (CIDOB), the Migration Policy Group (MPG), and over 35 national partners with co-funding from the European Commission.

In order to compute MIPEX, eight different areas of policy intervention are identified: labour market mobility, education, political participation, access to nationality, family reunion, health,

permanent residence and anti-discrimination. MIPEX is based on 167 policy indicators, consisting of 167 questions related to the eight policy areas, which were designed to benchmark current laws and policies against the highest standards. Each question has three possible answers which can be awarded from one to three points for those policies meeting the highest standards. Questions belonging to a same area are grouped into four sub-areas, concerning similar aspects of the policy, whose score is computed as the average of the questions scores. Thereafter, the points are averaged out at a policy area level and then the average of the points of the eight areas is obtained. Finally, the initial 1, 2, 3 ranking is converted into a 0, 50, 100 scale: the closer a Country score is to 100 the closer the implemented policy is to the top standards. Table 2 shows both the overall and the specific scores for Italy in comparison with the EU-15 average.

Table 2-MIPEX 2014 in Italy and in the EU-15

	Italy	EU-15
Overall Score	58	61
Family Reunion	72	61
Labour Market Mobility	66	68
Permanent Residence	65	64
Health	65	52
Anti-discrimination	61	68
Political Participation	58	60
Access to nationality	50	59
Education	34	49

Source: mipex.eu

The reported MIPEX refers to 2014. Italy overall score is 58, slightly under the EU-15 average. Italy score is higher than the EU-15 in the family reunion and health areas. As to the latter, Italy is one of the fourteen MIPEX Countries that provides language support in the healthcare system. Moreover, in Italy all residents have the same healthcare coverage as nationals in law and in practice. On the other hand, the policy areas in which Italy is definitely below the EU-15 level are anti-discrimination, education and access to nationality. An education MIPEX lower than 50 signals that Italian schools are not helping pupils integrate, which may be due to the fact that teachers cannot recognise immigrants' needs and treat them as any other students, without any additional support. As far as access to nationality is concerned, except for marginal cases, in Italy non-EU immigrants can get Italian citizenship after ten years of residence. This data is far above the most frequent residence requirement in the MIPEX Countries, which is five years, and the average, which is seven years. As to anti-discrimination, a poor score may

indicate that potential victims are too poorly informed and supported to take even the first step in the long path to justice or that institutions have weak enforcement mechanisms.

This thesis focuses precisely on what may be the effects of discrimination on immigrants' life in Italy. What is original about this work is the fact that Italian data, provided by the ISTAT survey "Condizione e Integrazione Sociale dei Cittadini Stranieri" (2016), is used and that this data focuses on immigrants' own perceptions. This represents something new with respect to previous works, most of which are focused on explaining natives' attitudes towards immigrants and consider either the social or the economic effects of migration. The present analysis investigates both social and economic effects of discrimination against immigrants.

This thesis has a double objective and is thus organised in two interlinked parts. In the first part the objective is to isolate the impact that insults and threats against migrants have on their perception of discrimination to observe what is the effect that these exogenous variations in discrimination perception have on immigrants' health. On the other hand, the second part aims at capturing the effect that immigrants' health has on their employment status, always isolating changes in health status determined by offences and aggressions. Using more technical terms, each one of the two models just described is estimated through an instrumental variable strategy, in which the instruments used are always verbal insults and threats, assaults. A third model is also presented, which estimates the association between discrimination and social integration, always adopting an instrumental variable strategy.

The thesis continues with a section that presents the existing literature dealing with the topic of migration. Following are sections that describe the data used and detail the empirical methodology. These are then followed by the main results, which are divided into four parts. The first one analyses which are the determinants of migrants' life satisfaction, the second one presents the association between discrimination and health, the third one illustrates the impact that immigrants' health has on their employment status and the final section presents the results concerning the association between discrimination and integration. These three last sections report estimates using both weighted least-squares and Generalised Method of Moments.

2. Literature review

Literature dealing with the economic impact of migration usually focuses on identifying what are the determinants of the public opinions on immigrants. Scholars find three main channels which affect natives' perception of migrants: their impact on the labour market, on the fiscal system and, thirdly, on the cultural and social sphere.

First of all, as to migration effects on the labour market, two interlinked aspects are to be considered: the impact on wages and on unemployment. Starting from economic theory, the fact that in a competitive labour market the labour demand curve is downward sloping implies that immigration, that is to be viewed as an increase in the number of workers in the economy, negatively affects the wages of natives. Though, the amount of such a reduction depends on the elasticity of the labour supply curve. Borjas (2003) implements a simple theoretical model to describe what are the consequences of migration in the host Country economy, when the labour supply curve is perfectly inelastic. He assumes that labour is homogeneous, hence natives and migrants are perfect substitutes in production, that the production function depends on just two inputs, capital and labour, and that the supply of capital is inelastic. Moreover, the model works on the assumption of a competitive economy, in which the prices of inputs correspond to the value of their marginal product, starting from an equilibrium condition. In such a scenario, the increase of workforce due to migration flows causes the wages of native workers to fall, but the overall economy's output increases thanks to what Borjas call "immigration surplus", that is an increase in national income accruing to natives. The effects are the same if the assumption of a rigid supply curve is removed. The additional effect of migration flows in an economy in which the labour supply curve is elastic is that they increase the unemployment rate of natives, namely of those natives who are not willing to work at the lower, post-immigration wage. In both models the overall economic output increases and the gain is absorbed by employers, who, thus, have a conflict of interest with workers.

Peri G. (2010) empirically analyses the impact of migration on employment and income (GDP) in the United States. He considers the effect of net variations in migration on the change in various components of the output at the state level (GSP), among which natives' employment and total factor productivity. Peri's findings show that a 1% increase in net migration flows has a positive but not statistically significant effect on total employment in the short run, but it has a positive and significant impact in the long run (one decade). Moreover, the estimated coefficient for 1% variation in net migration is higher than 1, meaning that migration creates new jobs for natives (crowding in).

The second determinant of natives' perception of migrants is their fiscal impact. If there existed no distortionary taxes and transfers, immigration would cause similar reactions in similar social groups, because it would have a similar impact on labour market and labour income. Though, government redistributive policy causes natives post-tax income to change in different ways in response to migration and thus different political coalitions may organise around migration. In a theoretical framework, migrants are likely both to pay taxes and to benefit from government transfers and public services, hence *a priori* they could have a negative as well as a positive impact on the fiscal balance. However, the sign of their actual contribution depends on their skill composition. If, like in the Italian case, migrants are mainly low-skilled, their access to the labour market will lower low-skilled wages and, what we are more interested in, they will decrease the post-tax income of the more skilled workers. Indeed, since high-skilled workers usually perceive higher income and progressive tax schemes are in place, they are going to bear migration fiscal costs by paying higher taxes to finance migrants' public assistance. Hanson *et al.* (2005) analyse how exposure to migrants' fiscal pressure affects natives' support to free migration in the United States. They measure fiscal exposure as a dichotomous variable accounting for two different aspects: fiscal pressure and migrants' share. This binary variable takes value one if welfare generosity of a certain State is above the national median welfare spending per native and, at the same time, the ratio of immigrants to natives is above a specific threshold. To measure the effect of this variable, interacted also with education levels, on natives' support for migration, Hanson *et al.* estimate a probit model. Their main finding is in line with theoretical explanations: college educated natives have a more restrictive policy view in those States in which the welfare system is very generous.

So far, only studies concerning migrants in the United States have been presented, though also the existing literature treating the migration phenomenon in the European Union is large. Indeed, due to the increasing flows of migrants, coexistence, discrimination and integration have become important issues in public policy and election campaigns of the European governments in the last years.

Economists traditionally focus on the immigration impact on the labour market, on wages and on fiscal transfers, but they do not fail to take into account also social and cultural aspects affecting natives' perceptions. For example, Card, Dustmann and Preston (2009) consider two dimensions, which they call the 'direct economic effect' and the 'compositional amenities effect'. They find that both economic concerns, related to job opportunities and wage reduction, and compositional effects, concerning the changing composition of local communities, play a role in shaping natives' perception of immigrants. Though, compositional amenities concerns contribute to a larger extent to explain the variation in opinions towards

immigration policy between different demographic groups, such as high-low educated, or younger-older people. The same two channels are considered by Senik *et al.* (2011), who use data from the first round of the European Social Survey to investigate whether the perceived share of immigrants in the population has a negative impact on the support for the welfare state. A probit model is constructed using as independent variables not only the perceived share of immigrants, but also some interaction terms between the perceived share and, on one hand the “taste channel” and on the other hand the “economic channel”.

Also D’Hombres and Nunziata (2015) exploit the European Social Survey dataset to analyse natives’ attitude toward migrants. In particular, they focus on studying the different approaches that different natives’ demographic groups have toward migrants. They exploit exogenous changes in compulsory-education legislation to identify the effect of education on attitudes towards immigrants. What emerges from this study is that more educated people appear to have more positive attitudes toward immigration, which can be explained both by economic and by non-economic factors. These findings suggest that policies boosting education could be a valid instrument to enhance integration in society characterised by large immigration flows.

The two different ways to interpret natives’ attitudes toward migrants, which, when negative, is discrimination, lead back to the interpretations of two Nobelists: Becker (1957), claiming that the only driver of discrimination is taste, and Arrow (1973). The latter, in his article “The theory of discrimination”, focuses on racial discrimination and devotes a section to analyse discrimination from employers. Arrow defines discrimination in the labour market as the evaluation process not only of the worker’s productivity, but also of some of his/her personal characteristics. Because of their prejudices, employers prefer to hire native workers rather than migrant ones and are willing to pay an opportunity cost, even though detrimental to their profits. Arrow explains the source of discrimination saying that discriminatory behaviours are led not only by employers’ distaste, but also by some rational beliefs based on economic theory. Indeed, in Arrow’s model, employers, who cannot know some unobservable workers’ characteristics, like ability and productivity, choose whom to hire on the basis of their preconceived ideas, such as that non-natives are less likely to be highly productive. Hence, there exists a rational justification to give them a lower wage. This behaviour may generate a vicious cycle, since foreign workers will be less willing to participate the labour market, or, discouraged, they will underinvest in their human capital. The kind of theory just described, which explains race inequality through stereotypes, is called statistical discrimination.

All the studies I have mentioned so far are focused on searching for the causes that drive natives’ behaviour towards immigrants, thus adopting host Country nationals’ point of view.

The two studies which follow, instead, are ones of the few articles which speak about the effects of discrimination on immigrant population, bringing attention to the opinions of immigrants themselves.

K. Vancluysen and M. Van Craen (2010) replicate a study conducted on US migrants to analyse the relationship between perceived discrimination and integration using data about Flemish Belgian migrants of Moroccan and Turkish descent. The relationship between these variables has been the focus of many studies because in the social science literature there exist two opposing theories explaining the relationship between integration and discrimination. On one hand, there is the assimilation theory (Gordon, 1964), stating that the more an individual is integrated in the host society and the less she/he is discriminated, precisely because he/she assimilated natives' attitudes and habits. On the other hand, the ethnic competition theory (Portes *et al.*, 1980) says that the more integrated a migrant is and the more she/he will perceive discrimination. In fact, a higher level of familiarity with the culture of natives means a greater consciousness of discrimination. Hence, according to the first theory, integration should have a positive effect on discrimination, whereas according to the latter one it should have a negative impact. K. Vancluysen and M. Van Craen regress measures of perceived personal discrimination and of ethnic group discrimination on indicators of both structural (occupational position, educational attainments) and socio-cultural integration (chatting with native neighbours, majority language proficiency). They find that migrants more socially integrated are less likely to perceive group discrimination, though no relation is found between social-cultural integration and personal experiences of discrimination. Hence, no clear support was found for either the assimilation theory or the ethnic competition theory.

The sociologist Mirna Safi investigates what are the determinants of migrants' life satisfaction and why it is lower than natives' satisfaction even after many years spent in the host Country (2009). She finds out that being an immigrant has a negative effect on happiness and life satisfaction and that, unlike it can be thought, this negative effect does not fade away with time. To observe an increase in migrants' life satisfaction it is necessary to wait for a 20-year migration duration, or to take into consideration those second generation migrants with at least one native parent (generation 2,5). The difference in life satisfaction between natives and immigrants can be partially explained by the impact of perceived discrimination and the same difference about second generation and generation 2,5 is fully explained by discrimination. The relation between life satisfaction and discrimination may not be causal because of omitted variables (personality traits, physical attractiveness, sexual orientation) bias. In order to overcome this problem, Safi instruments perceived discrimination with the affiliation with two religious communities (Judaism and Islam). Such an instrumental variable is informative

because defining themselves as members of a religious community, which may be perceived as discriminated against, strongly correlates with perceived discrimination. The same negative correlation between life satisfaction and discrimination is presented also in one of the next sections, though through a simple weighted least squares model exploiting data about Italian immigrants.

D. W. Johnston and G. Lordan, in their article in the *Journal of Health Economics* (2010), develop the topic of discrimination and analyse the impact of discrimination on various health outcomes of Pakistani and Bangladeshi Muslims resident in the United Kingdom. They use data referring to such a treatment group from 1999 to 2004, because they want to exploit the exogenous change in natives' attitude towards Muslims consequent to the 9/11 terrorist attack. Johnston and Lordan implement a difference-in-differences estimator considering the difference between two different time periods, before and after September 11th 2001, and between two demographic groups: Pakistani and Bangladeshi Muslims, the treatment group, and non-Muslim Indians, the control group. The findings of this study tell us that both objective indicators of health status, such as blood pressure, BMI (Body Mass Index), or cholesterol, and subjective ones, like perceived health status, worsened for Muslims relative to the control group after the terrorist attacks. This negative impact is explained through both direct and indirect effects of discrimination and health. Direct effects consist of the increased stress that people discriminated against must bear: higher levels of stress negatively affect health by, for instance, increasing blood pressure. On the other hand, indirect effects concern the fact that discrimination may push people to withdraw from the labour market or to be willing to work less hours than they could, for example preferring to work part-time rather than full-time; such worse economic outcomes may increase the level of stress and negatively impact on health. Also this thesis investigates the impact that discrimination has on health, though adopting an instrumental variable strategy.

A second paper by F. Roy and W. M. Rodgers (2011) analyses the impact of increased discrimination following the terrorist attacks of September 11th 2001 on economic outcomes of Muslim workers in the United States. Three different treatment groups were identified, characterised by different levels of risk of targeting: the group at highest risk is composed of Muslims coming from Middle-Eastern Countries, such as Syria, Iraq and Saudi Arabia, Iran and Afghanistan. The other two groups are formed by adding to immigrants of the first group immigrants coming from different Muslim-majority Countries, but less linked with the origins of the terrorists. The comparison group which is selected to implement the difference-in-differences estimator is made of first and second generation immigrants coming from non-Muslim-majority Countries. Roy and Rodgers find out that the employment gap among target

group men aged between 16 and 25 and the comparison group widened shortly after 9/11. This result, though, does not hold for older Muslim immigrants, for whom there is no change in the employment rate. Also the number of hours worked per week reduced after the terrorist attacks for the younger immigrants in the target group, while the same variable did not change for the older. The negative effects of 9/11 on both economic outcomes appear to dissipate with time: by the end of 2004 they had almost disappeared. However, it is difficult to determine whether this negative impact on employment rate and on the number of hours worked per week is due to a rise in discrimination or to the antiterrorism programs which were put in place. Roy and Rodgers identify two main pieces of evidence suggestive of a growth in discrimination against immigrant men: first, Muslims with age and nativity profiles closer to the terrorists' ones experienced larger declines in their employment, second, the decline in the employment and earnings of Muslims associated with 9-11 began to dissipate after 2002, just when some of the salient anti-terrorism programs and laws were initiated.

Relative to the existing literature, this paper presents an original and new analysis for two main reasons: it focuses on immigrants' point of view and opinions rather than on natives' ones and, secondly, it uses Italian data. Although papers describing immigrants' perceptions already exist, even if they are quite rare, none of them is based on data from the ISTAT survey "Condizione e Integrazione Sociale dei Cittadini Stranieri", which was published in 2016.

3. Data description

The ISTAT survey “Condizione e Integrazione Sociale dei Cittadini Stranieri” (“Condition and social integration of foreign citizens”) was carried out in 2011 and addresses households resident in Italy with at least one foreign citizen among their members.

For the survey purposes, a ‘household’ is defined as a set of cohabitant people, who may be married, relatives or tied by emotional bonds, and ‘foreign citizens’ are identified on the basis of their citizenship and not of their Country of birth.

Italian citizens from birth, who are members of the surveyed families, were not interviewed, but information about their sociodemographic features (gender, age, country of birth, etc.) were collected anyway.

The sample considered is composed of 9.533 households, a total of 25.326 individual observations, living in 833 different Italian municipalities. Foreign citizens are 20.379 of them, 696 are naturalized Italian and 4.251 are Italian citizens from birth, for whom all observations, except for the ones relative to sociodemographic variables, are missing.

For people aged less than 14, as well as for people temporary missing, answers were collected as proxies, that is a family member of age was interviewed at their place.

The survey is divided into seven main sections:

1. Family
2. Education
3. Migration history
4. Working life
5. Discrimination
6. Health
7. Integration

I am going to focus mainly on the last three sections, analysing also the background information, such as education and occupation.

The survey makes a distinction between “foreign immigrants” and “naturalized immigrants”. More precisely, observations are gathered in eight different target groups (Table 3): foreign immigrants, second generation non-naturalized immigrants, first generation naturalized immigrants, second generation naturalized immigrants, people born abroad who had Italian citizenship at birth, people born in Italy who had Italian citizenship at birth, Italian citizens born abroad and Italian citizens born in Italy.

Table 3-Sample individuals divided by target groups

Target groups	
Foreign first generation	17.544
Foreign second generation	2.821
Naturalized first generation	497
Naturalized second generation	199
Ex-Italian nationals born abroad	1
Ex-Italian nationals born in Italy	13
Italian nationals born abroad	154
Italian nationals born in Italy	4.097
Total	25.326

A further distinction that is made is about generations. Excluding Italian nationals responding to the survey, the variable ‘generation’ can take on five different values: 1, 1.25, 1.5, 1.75 and 2. By focusing on these five generations, I am excluding those target groups including ex-Italian nationals and Italian nationals. It is important to notice, in order to correctly understand the tables that follow, that target groups 1 and 3, referring to first generation immigrants, actually include observations belonging to generations 1, 1.25, 1.5 and 1.75. Belonging to a certain generation conveys information both about citizenship and about the age of the immigrant.

Analysing the data, I observe that people belonging to generations closer to the second one are younger (Table 4). For example, migrants belonging to the first generation are older than 18, while migrants of generation 1.75 can be newborn, by conventional definition. People belonging to generation 1.25 are older than 13 and those belonging to generation 1.75 can be older than 6, always by definition. Despite in the ISTAT dataset information about age is aggregated in groups and precise ages of the respondents are not available, their specific age is deductible by comparing the sociological definition of the different generations with the available data. Moreover, as to citizenship, the share of naturalized migrants, that is those who got Italian citizenship, increases as they belong to generations closer to 2. Indeed, as it is shown in Table 5, while the proportion of naturalized migrants vis à vis the share of foreign ones is about 2% for the first and the 1.25 generations, it becomes 4% for generation 1.5 and almost 6% for generation 1.75. Finally, the same share reaches about 7% for second generation migrants. In order to better understand the composition of naturalized migrants, I constructed a table crossing data about EU citizenship and data about naturalization (Table 6). Only non-naturalized people were supposed to specify whether they are EU citizens or not and what emerges is that 65% of non-naturalized migrants are not EU citizens.

The following chart gives information about Italian legislation for the acquisition of Italian citizenship, which is regulated by law n.91 of February 5th 1992.

Italian Citizenship

In Italy citizenship is acquired on the basis of the *ius sanguinis* principle, that is children of Italian nationals get Italian citizenship at birth (art.1 law 91/92).

In really specific cases also *ius soli* is applied: when a child is born by unknown or stateless parents or when he/she is abandoned, then he/she acquires Italian citizenship (art. 1, comma 1 letter b and comma 2 law n. 91/92).

People aged less than 18 can become Italian citizens also when adopted by Italian nationals, or when their parents are naturalized Italian. In this latter case, in order for Italian citizenship to be acquired, it is necessary that the children live together with their parents (art.14 law 91/92).

People of age can become Italian citizens by marrying an Italian national: after two years of marriage they can ask for Italian citizenship. In case children are born during the marriage, the length of stay to be entitled to get the citizenship reduces.

A second case giving the right to foreign nationals to get Italian citizenship is by length of residence in Italy. Non-EU citizens can acquire Italian citizenship after having being resident in Italy at least for ten years, while for EU citizens the term is just four years.

Table 4-Sample individuals divided by generation and age

Age	Generation				
	2	1.75	1.5	1.25	1
0-4	1.404	147	-	-	-
5-9	937	414	54	-	-
10-14	452	377	417	8	-
15-19	147	200	469	261	18
20-24	51	82	264	462	591
25-29	9	23	76	294	1.549
30-34	5	8	30	129	2.346
35-39	3	7	11	53	2.535
40-44	1	3	5	16	2.286
45-49	2	5	2	11	1.793
50-54	1	3	2	4	1.343
>55	8	5	10	7	1.721
Total	3.020	1.274	1.340	1.245	14.182

Table 5-Sample division into target groups and generations

Target group	Generation				
	1.75	1.5	1.25	1	2
Foreign first generation	1.201	1.282	1.219	13.842	-
	94,30%	95,70%	97,90%	97,60%	-
Foreign second generation	-	-	-	-	2.821
	-	-	-	-	93,40%
Naturalized first generation	73	58	26	340	-
	5,70%	4,30%	2,10%	2,40%	-
Naturalized second generation	-	-	-	-	199
	-	-	-	-	6,60%

Table 6-EU citizenship on non-naturalized migrants

	EU citizenship		Total
	No	Yes	
Non- naturalized			
Frequency	9.041	4.801	13.842
%	65,32%	34,68%	100,00%

To make sure that all the observations for all the variables are available and to simplify the interpretation of the data, only observations referring to first generation migrants are taken into account. As already mentioned, first generation migrants are defined as people aged more than 18 who are either foreign citizens who migrated to Italy, or naturalized Italian citizens. On the whole, 14.182 observations out of 25.326 are kept (see Table 7). A disadvantage of keeping only this data is that we have a loss of information, for instance all the information about children are lost.

Table 7-Number of observations in the different subsamples

	N° observations
Whole sample	25.326
Foreign and naturalized migrants (whole sample excluding ex-Italian nationals and Italian nationals)	21.061
First generation migrants	14.182

Table 8-Missing observations

	Available obs.	Missing obs.
DISCRIMINATION		
Discrimination	14.144	38
Discrimination at workplace	11.704	2.478
Discrimination in the health care system	13.549	633
Discrimination in public offices	13.821	361
Discrimination by neighbours	13.833	349
Verbal insults	14.182	0
Threats and assaults	14.182	0
INTEGRATION		
Interest in Italian politics	14.182	0
Italian TV news	13.177	1.005
Italian newspapers	6.254	7.928
Italian television language	13.177	1.005
ITALIAN LANGUAGE		
Ability to read Italian	14.182	0
Ability to write Italian	14.182	0
Ability to speak Italian	14.182	0
Ability to understand Italian	14.182	0
Ability to understand Italian TV news	13.688	494
Language spoken with family	13.115	1.067
Language more frequently spoken with friends	13.842	340
WELLBEING		
Life satisfaction	14.182	0
Loneliness	14.182	0
Mental Health	14.182	0
Perceived Health Status	13.956	226
Absence of chronic health problems	13.878	304
CULTURE		
Religion	13.714	468
EMPLOYMENT		
Active in the labour market	10.895	3.285
Occupation	14.022	160
RESIDENCE		
Region	6.154	8.028
Province	6.154	8.028
Municipality	6.069	8.113

Despite first generation immigrants were supposed to answer to all the questions in the survey, many answers are missing: for some variables less than 14.182 observations are available,

which will affect the estimation models. The above table (Table 8) conveys synthetic information about the missing observations for questions related to different dimensions.

Some sociodemographic features follow, to better understand the composition of this subsample. First of all, 42% of the observations is men, while 58% is women.

The first three foreign Countries with the highest birth rate of migrants in our sample are Romania (23%), Albania (11%) and Morocco (9%). Table 9 reports these numbers and shows the share of migrants' citizenship, while Table 10 displays the distribution of the main sociodemographic characteristics in the sample.

Table 9-Citizenship distribution in the sample

Citizenship	N	%	Cum %
Romania	3334	23.5	23.5
Poland	517	3.6	27.2
Albania	1639	11.6	38.7
Ukraine	796	5.6	44.3
Moldova	331	2.3	46.7
Macedonia	198	1.4	48.1
Morocco	1260	8.9	56.9
Tunisia	409	2.9	59.8
Egypt	128	0.9	60.7
China	446	3.1	63.9
Philippines	378	2.7	66.5
India	250	1.8	68.3
Bangladesh	212	1.5	69.8
Ecuador	199	1.4	71.2
Peru	198	1.4	72.6
Oceania	4	0.0	72.6
Others	3883	27.4	100.0
Total	14.182	100.0	

What is important to notice in Table 10 is the uneven distribution in terms of macro-area of residence of surveyed migrants: 36.4% of respondents lives in Northern Italy, 46,3% in the South or in the Isles and only 17,4% comes from the Centre. This distribution in the sample is even more weird if we think that the population of Italian migrants is much more concentrated in the North, while in the South the density of immigrants' population is really low. The reasons

of such an unbalanced sample and how I coped with it will be explained in the “Empirical strategy” section.

Table 10-Distribution of the main sociodemographic variables in the sample

	N	%
Sex		
Male	6.027	42,5
Female	8.155	57,5
Age distribution		
Under 30	2.158	15,22
30-54	10.303	72,65
Over 54	1.721	12,14
Employment status		
Employed	9.696	68,37
Unemployed	1.199	8,45
Inactive	3.287	23,18
Education level		
No education	2.225	15,69
Middle level	10.196	71,89
Tertiary education	1.761	12,42
Religion		
Muslims	3.682	26,85
Others	10.032	73,15
Municipality size		
Main municipalities	3.583	25,26
>10.000	7.278	51,32
< 10.000	3.321	23,42
Macro-area of residence		
North	5.161	36,39
Centre	2.460	17,35
South and Isles	6.561	46,26

As to where migrants currently live in Italy, information about the region, the province and the municipality in which they live were suppressed. The only information available is the first and, possibly, the subsequent provinces and municipalities in which migrants lived once arrived in Italy. For those migrants who are still living in the same municipality since their arrival, no specific information is available. So, observations as to the first municipality is available only for 6.154 individuals, that is all those who have changed municipality of residence during their stay. Nevertheless, some of the municipalities contained in the dataset are not matched to the correct province, or they do not even exist. In this latter case the respondent indicated either his/her village instead of the municipality (e.g. Borgo San Bernardino, Roggiano Valtravaglia), or the name of the municipality was incorrectly specified. However, only 99 observations in

the whole sample (85 in the subsample retained) are excluded due to this kind of problems. Excluding these incorrect observations and observations for those individuals who are still living in the same municipality since their arrival, only 6.069 observations are left (see Table 8). Moreover, data about municipalities is not easily manageable because ISTAT codes, containing both the municipality number and the code of the province to which the municipality belongs, are not used: municipalities are indicated with a numeric code representing the progressive number of that municipality in its province. However, the variable ‘municipality’ could be constructed starting from the available data so that each municipality is identified by a unique value corresponding to the ISTAT code. Also, ISTAT dataset allows to distinguish among metropolitan municipalities, municipalities with more than 10.000 inhabitants and the smallest ones, counting less than 10.000 residents. On the basis of this distinction I constructed a table (Table 11), which reports some characteristics of the sample divided by municipality size. This highlights that a larger share of immigrants with a low education, as well as a higher proportion of unemployed, is to be found in the smaller municipalities. It is also interesting to notice that there is a higher concentration of Muslims in the lower categories and that, independent of the kind of municipality, women are more than men. The last characteristic considered in Table 11, Index of Mental Health, is almost the same for metropolitan municipalities and for the ones with more than 10.000 inhabitants, but it appears to be slightly above the sample average, thus indicating a better mental health status, in the smallest municipalities. This index goes from a minimum score of 7,6 to a maximum of 69,4 and the mean for the whole sample is 53,48. Even if not reported here, t-tests have been run and they indicate that the difference between the Index of Mental Health between Southern observations and observations from Northern or Central Italy are statistically significant.

Table 11-Characteristics of ISTAT Sample by Municipality Size

Municipality size	Mean Sex	Mean Age	Mean Occupation	Mean Education	Mean Muslim Religion	Mean Mental Health
Metropolitan municipalities (25%)	0,58	0,54	0,08	0,15	0,19	53,37
>10.000 (51%)	0,58	0,50	0,12	0,13	0,28	53,37
< 10.000 (23%)	0,56	0,48	0,12	0,09	0,33	53,83

Unlike municipality data, data about provinces is easily manageable because each of the provinces indicated corresponds to the correct ISTAT code. Excluding, also in this case, those provinces where migrants live since their arrival, which are unknown, 6.154 observations are

left. Thanks to the availability of this data I was able to group all the observations into the twenty Italian regions. Nevertheless, information about the current region of residence was suppressed from the data, so the variable I constructed indicates whether we have information about the penultimate province and hence region, in which the respondents lived. This is the most up-to-date information available about residence. In other words, ISTAT dataset contains residence information, even though not really specific and precise, about migrants who moved at least once from their initial province of residence. This kind of information is available for 6.154 individuals, while there is no information about the current residence of 8.028 respondents.

Exploiting this new information, I constructed two tables: Table 12 presents the same sociodemographic characteristics reported in the previous table, but distinguishing by region, whereas Table 13 illustrates the same characteristics but distinguishes between migrants for whom a recent residence information is available (who moved) and migrants for whom no up-to-date information is provided (who did not move). In Table 12 regions are ordered by GDP level as reported by ISTAT in “Conti economici territoriali” (December 12th 2016): the first regions are the ones with a higher GDP *per capita*. It can be noticed that in the bottom part of the table, in those regions with the lowest GDP *per capita*, the share of employed migrants is very high, always above 90%, while the proportion of migrants with tertiary education is quite low, being below 15%. In Table 13 it is interesting to notice that there are statistically significant differences for almost all the parameters considered between migrants who moved and those who did not change their residence since their first arrival.

The same migrants’ characteristics are presented also in Table 14, which divides migrants according to their macro-area of residence, North, Centre, South and Isles.

Table 15 presents results concerning education and employment distinguishing by sex. As far as the education level is concerned, 16% of the observations has no education, while 12% has a university degree or a PhD. Among women 15% of them has a tertiary education, while the corresponding percentage for men is 9%. As to first generation migrants’ current job, most of them (59%) is subject to an employer and inactive people are 23%. Inactive people are spread over the different age groups, only 2% of them is currently attending a school or a university and 85% of them is women. 83% of men is working, whereas only 58% of women is. The ISTAT survey includes also questions aimed to get specific information about the job of each respondent; one particular question is about the kind of job, such as employed, professional, self-employed, but only 295 people answered.

There is some inconsistency in the data, because if the occupational condition of respondents is crossed with the number of employees some state to have, it emerges that even people who

claim to be inactive or unemployed said they have some employees. What's more, only 1.033 observations are available for this question. Anyway, the dataset contains information about the specific occupation of every respondent. It is interesting to notice that 29% of women are employed in cleaning, personal and cultural services and another 25% in domestic activities, whereas 22% of men is workers specialised in the construction industry or in buildings' maintenance.

Table 12-Characteristics of ISTAT Sample by Region

Region	Female	Older than 40	Employed	Tertiary education	Muslim Religion	Index of Mental Health
Trentino-Alto Adige (1%)	56%	56%	84%	19%	40%	54,02
Lombardy (13%)	49%	52%	89%	16%	27%	53,19
Valle d'Aosta	58%	58%	90%	15%	34%	53,97
Emilia Romagna (7%)	53%	55%	87%	15%	33%	52,62
Lazio (11%)	52%	53%	90%	13%	15%	53,98
Veneto (9%)	51%	52%	88%	16%	24%	54,06
Liguria (1%)	47%	60%	92%	15%	37%	52,83
Toscana (5%)	57%	58%	89%	18%	25%	53,05
Friuli-Venezia Giulia (1%)	62%	47%	91%	18%	18%	55,06
Piemonte (4%)	53%	61%	89%	14%	25%	52,06
Marche (2%)	53%	48%	86%	9%	27%	52,26
Abruzzo (5%)	59%	53%	89%	16%	26%	53,17
Umbria (1%)	58%	57%	89%	18%	29%	51,34
Basilicata	62%	67%	95%	15%	23%	51,77
Sardinia (2%)	61%	54%	89%	18%	27%	52,51
Molise	64%	62%	89%	11%	20%	53,45
Campania (11%)	57%	58%	94%	13%	22%	53,01
Puglia (6%)	41%	56%	92%	9%	28%	52,52
Sicilia (7%)	44%	56%	92%	8%	40%	53,44
Calabria (3%)	50%	54%	91%	10%	33%	51,72

Table 13-Summary statistics in subsamples based on residence information

	All	Did not move	Moved	T-test (p-value)
Observations	14.182	8.028	6.154	(H ₀ : mean ₁ -mean ₂ =0)
Age (mean)	8,67 [2,00]	8,52 [2,09]	8,87 [1,87]	0.0000
% Female	57,50% [0,49]	61,48% [0,49]	52,31% [0,50]	0.0000
Education level (mean)	2,45 [1,09]	2,42 [1,08]	2,49 [1,09]	0.0001
% Employed	88,99% [0,31]	88,12% [0,32]	90,00% [0,30]	0.0017
Ability to read Italian (mean)	3,03 [0,96]	2,92 [0,99]	3,18 [0,90]	0.0000
Ability to write Italian (mean)	2,81 [0,99]	2,71 [1,00]	2,95 [0,95]	0.0000
Ability to speak Italian (mean)	3,36 [0,82]	3,27 [0,87]	3,48 [0,73]	0.0000
Ability to understand Italian (mean)	3,38 [0,81]	3,29 [0,85]	3,50 [0,72]	0.0000
% Italian television language	76,66% [0,42]	75,15% [0,43]	78,64% [0,41]	0.0000
% Italian TV news	84,09% [0,37]	80,65% [0,40]	88,58% [0,32]	0.0000
% Police control at least once a month	78,48% [0,41]	77,92% [0,41]	79,20% [0,41]	0.0842
% Police very or quite effective	80,35% [0,40]	80,16% [0,40]	80,60% [0,40]	0.5124
% Discriminated at work	17,89% [0,38]	14,23% [0,35]	21,90% [0,41]	0.0000
% Discriminated by neighbours	6,88% [0,25]	5,27% [0,22]	8,99% [0,29]	0.0000
% Discriminated in the Healthcare system	3,24% [0,18]	2,72% [0,16]	3,91% [0,19]	0.0001
% Discriminated in public offices	8,43% [0,28]	7,01% [0,26]	10,28% [0,30]	0.0000
% Discriminated by verbal insults	11,16% [0,31]	8,50% [0,28]	14,64% [0,35]	0.0000
Length of stay (mean)	15,87 [6,81]	14,54 [6,42]	17,61 [6,91]	0.0000

The variables whose values are not expressed as percentage are categorical ones: age is aggregated (4=15-19 yrs, 12=more than 55yrs), education level (0=no education, 5=tertiary education), Italian proficiency indices (1=not good at all, 4=very good). Length of stay is a discrete variable, going from a minimum of 5 years to a maximum of 74. Column 4 reports p-values for the test of the equality of means, under the null hypothesis that the means for the two subsamples (columns 2 and 3) are equal. Standard deviations in square brackets.

Table 14-Summary statistics in subsamples based on residence information (macro-areas)

	North	Centre	South and Isles
Observations	5.161	2.460	6.561
Age (mean)	49,97% [0,50]	52,56% [0,50]	49,86% [0,50]
% Female	56,50% [0,50]	58,09% [0,49]	58,07% [0,49]
Education level (mean)	14,67% [0,35]	14,23% [0,35]	9,97% [0,30]
% Employed	85,58% [0,35]	89,20% [0,31]	91,63% [0,28]
Ability to read Italian (mean)	3,15 [0,97]	3,22 [0,89]	2,87 [0,95]
Ability to write Italian (mean)	2,92 [1,01]	2,99 [0,92]	2,66 [0,97]
Ability to speak Italian (mean)	3,3900 [0,87]	3,5200 [0,76]	3,2700 [0,79]
Ability to understand Italian (mean)	3,4100 [0,85]	3,5400 [0,75]	3,3000 [0,78]
% Italian television language	72,92% [0,44]	76,73% [0,42]	79,60% [0,40]
% Italian TV news	81,50% [0,39]	89,30% [0,31]	84,85% [0,36]
% Police control at least once a month	84,74% [0,36]	82,08% [0,38]	72,19% [0,45]
% Police very or quite effective	87,10% [0,34]	84,15% [0,37]	73,62% [0,44]
% Discriminated at work	17,64% [0,38]	15,04% [0,36]	19,17% [0,39]
% Discriminated by neighbours	6,83% [0,25]	5,65% [0,23]	7,39% [0,26]
% Discriminated in the Healthcare system	3,27% [0,18]	3,08% [0,17]	3,28% [0,18]
% Discriminated in public offices	8,70% [0,28]	8,57% [0,28]	8,16% [0,27]
% Discriminated by verbal insults	10,44% [0,31]	9,96% [0,30]	12,18% [0,33]
Length of stay (mean)	16,23 [6,58]	16,51 [6,92]	15,35 [6,89]

The variables whose values are not expressed as percentage are categorical ones: age is aggregated (4=15-19 yrs, 12=more than 55yrs), education level (0=no education, 5=tertiary education), Italian proficiency indices (1=not good at all, 4=very good). Length of stay is a discrete variable, going from a minimum of 5 yrs to a maximum of 74. Standard deviations in square brackets.

Table 15-Distribution of education and employment for men and women

	Sex		Total
	Male	Female	
Education level			
No education	1.102	1.123	2.225
	49,50%	50,50%	100,00%
	18,30%	13,80%	15,70%
Middle level education	4.381	5.815	10.196
	43,00%	57,00%	100,00%
	72,70%	71,30%	71,90%
Tertiary education	544	1.217	1.761
	30,90%	69,10%	100,00%
	9,00%	14,90%	12,40%
Total	6.027	8.155	14.182
	42,50%	57,50%	100,00%
	100,00%	100,00%	100,00%
Employment status			
Employed	5.003	4.693	9.696
	51,60%	48,40%	100,00%
	83,00%	57,50%	68,40%
Unemployed	526	673	1.199
	43,90%	56,10%	100,00%
	8,70%	8,30%	8,50%
Inactive	498	2.789	3.287
	15,20%	84,80%	100,00%
	8,30%	34,20%	23,20%
Total	6.027	8.155	14.182
	42,50%	57,50%	100,00%
	100,00%	100,00%	100,00%

The first survey section of great interest is the one about discrimination, which is divided into four different parts: discrimination at the workplace, discrimination in searching for a job, discrimination in everyday life and insults. Answers are based on the concept of discrimination defined as the way of treating some people in a less favourable way relative to the others because of their physical, mental, or personal characteristics.

Table 16 illustrates how answers to discrimination questions are distributed in the sample. The different kinds of discrimination considered in the survey are reported in the table in decreasing order: the first items are the ones for which migrants perceive to be more discriminated against.

Table 16- Discrimination perception (sample distribution)

	N	%
Perceived discrimination		
No	10.057	71,10
Yes	4.078	28,90
Total	14.144	100
Discrimination at work		
No	9.610	82,11
Yes	2.094	17,89
Total	11.704	100
Verbal insults		
No	12.599	88,84
Yes	1.583	11,16
Total	14.182	100
Discrimination in search of a house		
No	11.255	88,92
Yes	1.402	11,08
Total	12.657	100
Discrimination in search of a job		
No	11.740	90,59
Yes	1.220	9,41
Total	12.960	100
Discrimination in public offices		
No	12.656	91,57
Yes	1.165	8,43
Total	13.821	100
Discrimination in the residence neighbourhood		
No	12.881	93,12
Yes	952	6,88
Total	13.833	100
Threats and Insults		
No	13.590	95,83
Yes	592	4,17
Total	14.182	100
Discrimination in the Health Care system		
No	13.110	96,76
Yes	439	3,24
Total	13.549	100
Discrimination in asking for car insurance		
No	11.167	97,73
Yes	259	2,27
Total	11.426	100

Respondents to questions about discrimination at work are 11.704, because people who have never worked in Italy during their stay are excluded. 18% of people interviewed in this section has felt discriminated: 19% of the responding men and 16% of the responding women. People who are currently working are 1.711 and 48% of them felt discriminated in their current job, mainly by their employer. Respondents were asked which is the reason why they are discriminated against and they could choose among eleven alternative answers. The most frequent discrimination reason migrants indicated is being foreign. Another alternative reason why they are discriminated against is gender: 23% of women feels discriminated because they are women, while the corresponding percentage for men is just 3%.

Respondents to the subsection “discrimination in search of a job” are all the 14.182 individuals. 1.220 individuals, that is 9% of the subsample, felt discriminated in searching for a job. Also in this case, as in the previous subsection, 90% of people felt discriminated because they are foreign. Another important remark is that 26% of them felt discriminated because of the way they spoke Italian.

Table 17 illustrates how some main characteristics are distributed in two different subsamples: the one of migrants who have ever felt discriminated against and the one of those who have never experienced discrimination. It is interesting to notice that the mean education level and the employment condition are not statistically different in the two subsamples. Instead, most of the other variables considered appear to be different in a statistically significant way. This means that in the models that will be subsequently implemented, these factors need to be controlled for.

As to the third subsection, it investigates whether migrants felt discriminated in their everyday life: in public places, in asking for a loan or for buying a house, or by their neighbours. Observations for all the individuals are available. The frequency of people feeling discriminated against in these contexts is quite low, as reported in Table 16.

The next section to be analysed is health: starting from eating habits to chronic health problems. Observations are available for all the individuals. As to eating habits, 54% of respondents eat both Italian and their Country typical food. Moreover, 25% smokes almost every day. To the question “How healthy do you feel?” only 13.956 people answered, because 226 people refused to. 12% of all the observations is affected by chronic health problems (see Table 18). When asked whether they feel less productive at work possibly because of their depressive/anxious status, almost 4% of respondents refused to answer and only 6% of the responding individuals feel they are not as productive at work as they could be because of that. Only 0.21% of the sample went to psychiatric centres in the last three months.

Table 17-Summary statistics in subsamples based on perceived discrimination

	Never discriminated against	Discriminated against	T-test (p-value)
Observations	10.057	4.087	(H ₀ : mean ₁ -mean ₂ =0)
% Older than 40	0,5090 [0,4999]	0,4896 [0,4999]	0,0365
% Female	0,5966 [0,4906]	0,5195 [0,4996]	0,0000
Education level (mean)	2,4499 [1,0808]	2,4521 [1,0000]	0,9119
% Employed	0,8927 [0,3095]	0,8843 [0,3199]	0,1912
Ability to read Italian (mean)	3,0069 [0,9730]	3,0952 [0,9143]	0,0000
Ability to write in Italian (mean)	2,8118 [0,9941]	2,8268 [0,9640]	0,4554
Ability to speak Italian (mean)	3,3173 [0,8517]	3,4803 [0,7027]	0,0000
Ability to understand Italian (mean)	3,3374 [0,8406]	3,5011 [0,6879]	0,0000
% Italian television language	0,7680 [0,4221]	0,7657 [0,4236]	0,7808
% Italian TV news	0,8605 [0,3752]	0,8389 [0,3374]	0,0000
% Police control at least once a month	0,7905 [0,4069]	0,7730 [0,4189]	0,0294
% Police very or quite effective	0,8255 [0,3795]	0,7512 [0,4323]	0,0000
Length of stay (mean)	15,5657 [6,8981]	16,6701 [6,5179]	0,0000
Macro-area of residence (mean)	2,1086 [0,9014]	2,0726 [0,9088]	0,0317

The variables whose values are not expressed as percentage are categorical ones: education level (0=no education, 5=tertiary education), Italian proficiency indices (1=not good at all, 4=very good). Length of stay is a discrete variable, going from a minimum of 5 yrs to a maximum of 74. Standard deviations in square brackets.

Table 18 reports the mean values, as well as the minimum and the maximum, for three health measures that are going to be used in the following models. As it can be observed, migrants are on average quite healthy.

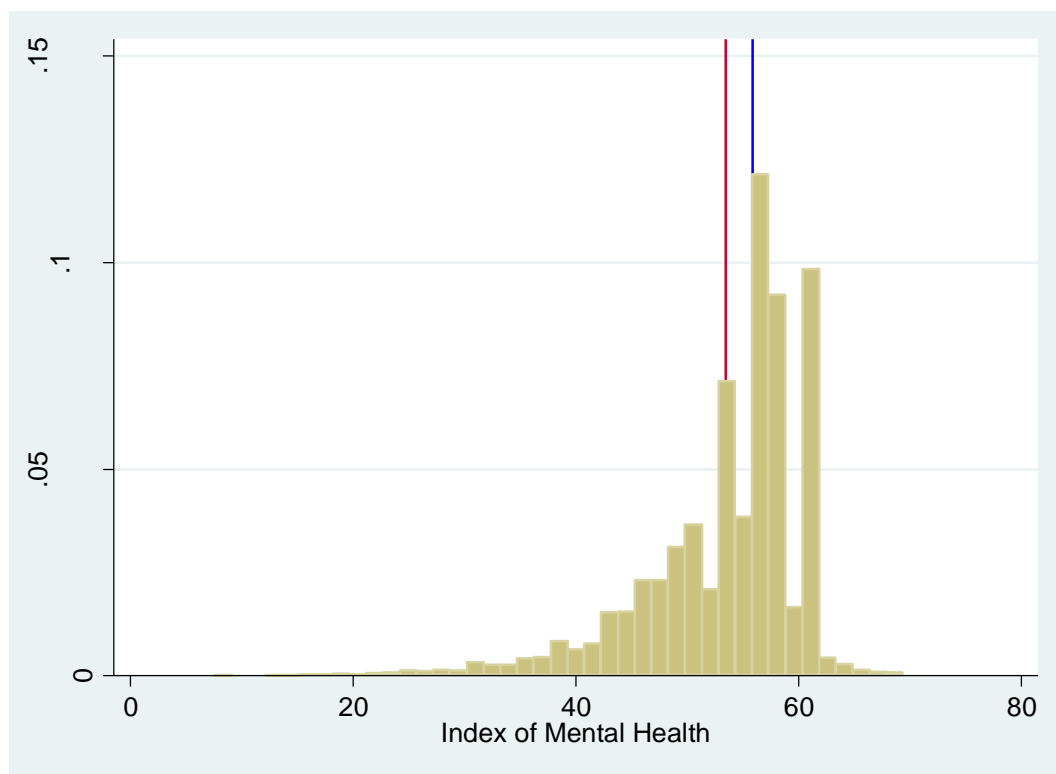
The ISTAT dataset includes also an index of mental health for every respondent, which is distributed as displayed in Figure 2. The higher the value of the index and the better the respondent's mental health. Figure 2 shows, through a vertical blue line, the median value of the mental health index, which is 55,9, and through a red line its mean value, which is 53,48. It

is to be noticed that the mean is lower than the median, hence the index of mental health distribution is negatively skewed.

Table 18- Summary statistics of the main health indicators

	Mean	Standard deviation	Minimum value (not healthy at all)	Maximum value (very healthy)
Index of Mental Health	53,48	7,1484	7,6	69,4
Perceived health status	4,11	0,7807	1	5
Absence of Chronic health problems	0,88	0,3273	0	1

Figure 2-Distribution of Index of Mental Health in the sample



Other interesting dimensions which are covered by the survey are migrants' life satisfaction and happiness. Together with the mental health index, these measures convey information about migrants' wellbeing, which could represent an important factor favouring integration. While all the observations are available for the variable indicating the level of life satisfaction, we can observe the happiness level of only 13.615: some observations are missing because respondents refused to answer.

The last section of interest is the one about integration. One way to measure integration is to observe migrants' ability to master Italian language. Different aspects of Italian language can be investigated, such as oral production, oral comprehension as well as written production and comprehension. In order to analyse the sample, I considered only oral ability. As results in Table 19 show, most migrants can both speak and understand Italian very well. As a further measure of integration, the share of migrants speaking Italian at home is added: this percentage tells that less than half of the population in the sample, about 36%, speaks Italian at home. For this last variable considered observations are missing for those respondents who have no relatives in Italy.

Another way to measure integration is to consider some migrants' private behaviours. One of these may be the habit of watching Italian television. 13.177 people out of 14.182 usually watch TV, 76% of them usually watch Italian channels and 84% of them watch Italian TV news. However, only 44% read newspapers to get informed. A second way to measure integration is to analyse the degree of participation in Italian facts. Because of this reason, Table 20 contains the variables "Interest in Italian politics", "Reading Italian newspapers", "Watching Italian TV news", "Television language" and "Going to the cinema". Only 37% of respondents are interested in Italian politics, but 94% of the migrants who usually read newspapers read Italian ones. For this last variable the 7.928 missing observations refer to those respondents who do not use to read any newspaper.

Table 19-Ability to master Italian language (sample distribution)

	N	%
Ability to speak Italian		
Not good at all	574	4,05
Not good	1.386	9,77
Quite good	4.569	32,22
Very good	7.653	53,96
Total	14.182	100
Ability to understand Italian		
Not good at all	533	3,76
Not good	1.334	9,41
Quite good	4.505	31,77
Very good	7.810	55,07
Total	14.182	100
Language spoken with family		
Other language	8.447	64,41
Italian	4.668	35,59
Total	13.115	100

Table 20-Indicators of social integration (sample distribution)

	N	%
Interest in Italian politics		
No	8.838	62,32
Yes	5.344	37,68
Total	14.182	100
Reading Italian newspapers		
No	367	5,87
Yes	5.887	94,13
Total	6.254	100
Watching Italian TV news		
No	2.097	15,91
Yes	11.080	84,09
Total	13.177	100
Television language		
Italian	3.075	23,34
Other	10.102	76,66
Total	13.177	100
Going to the cinema		
Never	12.982	91,54
At least once a year	1.200	8,46
Total	14.182	100

4. Empirical strategy

The final aim of this empirical analysis is to assess whether discrimination against immigrants has any effect on different dimensions of immigrants' life, namely with health, employment status and socio-cultural integration.

4.1 An instrumental variable approach

In the first model, I proceeded to analyse the impact that immigrants' perceived discrimination has on immigrants' health. Estimating this relationship through simple OLS would lead to biased and inconsistent coefficient because of reverse causality. In fact, on one hand it is intuitive to think that migrants who have ever felt discriminated against are more likely to suffer from health problems, particularly from mental health disturbs, but on the other hand also the opposite could be true, that is people affected by health problems may be more likely to be targeted. Hence, in order to tackle reverse causality, I chose to implement an instrumental variable strategy. To find suitable instruments for perceived discrimination was hard, but this work, which represents a first attempt to an IV estimation, suggests two instrumental variables. The instruments selected are 'verbal insults' and 'threats and assaults'. In other words, three different measures of health, mental health, perceived health status and absence of chronic health problems, are regressed on the exogenous variations in the discrimination perception due to insults and threats received.

'Verbal insults' and 'threats and assaults' are assumed to be strongly correlated with perceived discrimination: migrants who experienced this kind of offense are clearly more likely to feel discriminated against in the host society, it could be said that they are more sensitive to discrimination. This assumption is confirmed by the F-statistics presented at the bottom of each table, which are all above the conventional threshold of 10.

The regressions that are going to be presented are just a first step towards causality. In fact, the assumption of exogeneity of the instruments may be criticized: there may be some reverse causality between the instruments and the dependent variable, or some omitted variables may correlate with both the instruments and the dependent variable.

Firstly, as to reverse causality: migrants in bad health conditions may be more likely to be insulted or threatened. However, even if this is possible, it is unlikely. Moreover, migrants in the sample at hand are not so unhealthy to trigger insults and threats (see Table 17).

Secondly, as to omitted variables, it could be argued that there may be some omitted factors which correlate both with health and with the probability to be insulted or threatened

and aggressed. For instance, it could be argued that immigrants living in wealthy neighbourhoods, where living standards are high, are more likely to be healthy (e.g. because of lower levels of stress) and, at the same time, less likely to be offended, because in those areas people are more educated or police is more effective. In the ISTAT survey data, no information was provided about individual income, which would have been a good control for living standards, though I tackled the above problem by controlling for police controls' frequency and effectiveness. These two variables can be considered proxies of individual income levels and of the quality of the residence neighbourhood, since usually in wealthier residence areas police controls are more frequent and effective. Moreover, in the instrumental variable regression it is controlled for migrants' smoking and drinking habits, and for their BMI (Body Mass Index), which are factors affecting both health status and the probability to get insulted or aggressed. For instance, migrants who are used to drinking a lot are more likely to be targeted by natives. Hence, once it is controlled for these variables, being insulted or aggressed is assumed to be uncorrelated with the dependent variable health, it is assumed to be random.

The same rationales explained so far hold for the third model, estimating the association between perceived discrimination, always instrumented with insults and threats, and social integration. In this model too, once that the controls are included in the regression to tackle reverse causality between the instruments and the dependent variable, it can be assumed that threats and insults are associated with integration only through discrimination.

Also to estimate the second model, which focuses on the association between immigrants' health and their employment status, I opted for an instrumental variable strategy. In this second model the reverse causality between health and employment status is more straightforward than in the previous case. Indeed, healthier people are more likely to be employed, as well as employed individuals have more chances to be in a better health status. This last proposition is meaningful particularly if it is thought of psychological health: employed people may be more satisfied with their lives and experience lower levels of stress and frustration. Once again, this reverse causality bias is tackled by using verbal insults and threats received as instruments. It is to be noticed that the first stage equation of this second model is just the reduced form of the previous model, because the dependent variable of the first model, health, which is now the endogenous regressor, is directly regressed on the instruments, 'verbal insults' and 'threats and assaults'.

As to the validity of these instruments, it could be argued that they are correlated with the error term in the structural equation, because there may be some factors influencing both the probability to be offended or aggressed and the chances to be employed. For example, it could be argued that migrants living in wealthier neighbourhoods are less likely to be

unemployed as well as to be offended, because natives living in those areas are more educated or because migrants can develop better social networks which help them find a job. Like in the previous “discrimination-health” model, this omitted variable bias is tackled by controlling for police controls’ frequency and effectiveness. Other important characteristics which may correlate with both employment status and offences are controlled for: the ability to master Italian language, as well as unhealthy habits (e.g. consuming alcohol away from meals).

Furthermore, there may be reverse causality bias between offenses and employment status. In fact, people who have ever been insulted or aggressed may be discouraged, or humiliated, hence they may be less likely to be employed; at the same time, though, migrants’ employment status may affect their chances to be targeted. However, this last option is quite unlikely, because employment status is unobservable: from the outside it cannot be told whether a migrant is employed or unemployed. However, there may be some individual unobservable characteristics, such as personal appearance, which can affect both the probability of being employed and of being offended. In order to control for these unobservable features, I included in the regression a dummy variable indicating whether the respondent has ever worked in Italy. In this way, it can be distinguished between those migrants who have never worked in Italy, perhaps because of some personal characteristics which make them not suitable to work, and those who have the good characteristics to get a job, but are currently not working. Also in this second model it is assumed that, once all the controls mentioned are included in the regression, ‘verbal insults’ and ‘threats’ are random.

4.2 Controls

The controls included in the regressions are many, since the ISTAT survey provided a lot of information about respondents. All the controls are categorical variables, which are inserted in the regression as dummies for each single value. The various controls can be aggregated into five different sets: sociodemographic, cultural, lingual features, health determinants, information about residence location and employment information. In the first group of controls information about age, gender, marital status, education level, length of stay in Italy measured in years, and the BMI (Body Mass Index) is included. Cultural features concern respondents’ religion and their Continent of origin; more detailed information about origins could not be used because, due to privacy reasons, more specific information, like Country of birth, was provided only in aggregated form in the dataset. The third set of controls consists of indicators measuring immigrants’ ability to master Italian language: it includes categorical variables measuring Italian proficiency in writing, speaking, understanding and reading. It was important

to include this kind of controls because Italian language proficiency affects the likelihood to find an employment, as well as the probability to be discriminated against: people speaking a bad Italian are intuitively less likely to find a job and they could object of discrimination more easily. Health determinants include smoking and drinking habits of the respondents, which are well-known to be harmful for health. The first control is a categorical variable indicating whether the individual used to smoke in the past, still smokes or has never smoked. Variables concerning drinking habits instead consider how frequently the respondents are used to drinking beer, wine and fizzy drinks. Finally, the last group of controls deals with residence information. Also in this case, ISTAT dataset does not provide very specific information because of privacy reasons. That is why no information was available about the specific province, or region, where the immigrants are currently living. The most precise and recent indicator that I was able to create was about the penultimate province or region where the respondent lived, but such an indicator was available just for few observations. Hence, the variables that I included in the residence controls are the size of the residence municipality, distinguishing among metropolitan municipalities (Turin, Milan, Venice, Genoa, Bologna, Florence, Verona, Rome, Naples, Bari, Palermo, Catania and Cagliari) and those with more or less than 10.000 inhabitants, and macro-area of residence, indicating whether respondents live in Northern, Central or Southern (Isles included) Italy. As to the previous distinctions, Northern Italy includes Piemonte, Valle d'Aosta, Lombardy, Liguria, Trentino Alto Adige, Veneto, Friuli Venezia Giulia and Emilia Romagna, Centre includes Tuscany, Umbria, Marche and Lazio, Southern and Isles is composed of Campania, Basilicata, Calabria, Abruzzo, Molise, Puglia, Sicily and Sardinia. Moreover, trying to add as more information about residence location as possible, two further controls used are the frequency of police controls in the residence area and the effectiveness of such controls. The idea is that these indicators tell, to some extent, how safe a neighbourhood is: the more effective the police is and the lower the criminality rate is expected to be, as well as the more frequent the police controls are and the lower the criminality rate is likely to be. Both 'police controls' frequency' and 'police effectiveness' are categorical variables, representing the answers to the following questions, respectively: "In your opinion, how often does the police passes in the road where you live, both on foot and by car?" and "All things considered, do you think that the police manage to control criminality in the area where you live?".

A further control, indicating whether immigrants have ever worked in Italy, is included in the second model. The reason why this variable was added is to control for some migrants' characteristics which may prevent them from finding a job and at the same time affect discrimination likelihood.

4.3 Weights

One of the most important controls is macro-area of residence, because it conveys information not only about residence, but also about economic background, since Northern regions have higher GDP per capita than Southern ones (in 2012 ISTAT reported that in the South real GDP per capita was 43.2% lower than in the North-Centre Italy). Observations related to the South and Isles are oversampled with respect to the other macro-areas, despite the distribution of foreign residents in Italy is much more concentrated in the North and in the Centre. This resulting sample is explained by the sampling strategy that ISTAT used (“Condizione e Integrazione Sociale dei Cittadini Stranieri-Aspetti Metodologici dell’Indagine”, 2016). ISTAT chose to select data through stratified sampling, divided into two different selection stages: the first one concerning the sampling of municipalities with a minimum number of foreign citizens residents and the second one selecting from each municipalities’ stratum a certain number of households. The households sampled were allocated to six different geographical areas (North-East, North-West, Centre, South-East, South-West and Isles) through an allocation strategy which combined both proportional and uniform allocation. The latter, consisting of selecting a number of households from each stratum so to maintain the same proportion as in the population, was given a weight of 0.2 in the allocation strategy. On the other hand, the uniform allocation was attributed a greater weight, 0.8. Uniform allocation consists of selecting a certain number of households from each municipalities stratum regardless of those strata weight in the whole population. This method favours strata that have less weight in the population, in this case Southern Italy and Isles, by affording them the same level of importance as the more relevant strata. This reduces the global effectiveness of the sample, but it allows to make estimates at macro-area level with greater precision, which is exactly why uniform allocation was chosen.

Drawing estimates on the whole sample using such an unbalanced sample would lead to results whose external validity is poor, precisely because the sample does not reflect the distribution of the actual immigrants’ population in Italy. Therefore, in order to face the oversample problem, all the regressions that are estimated in the next sections are weighted. Weights used are already included in the dataset. These weights are constructed through a procedure based on the so-called calibration estimators. They are basically constructed starting from direct weights, that are the inverse of the probability of being included in the sample, then they are corrected by the number of missing answers and, at the end, final weights are determined through a problem of constrained optimization, where the constraints are equality conditions between some sample estimates and some known values, deduced from external

sources (ISTAT, “Condizione e Integrazione Sociale dei Cittadini Stranieri-Aspetti metodologici dell’indagine”, 2016).

4.4 Generalised Method of Moments (GMM)

As already explained, in estimating the different models, two variables are identified to instrument the endogenous regressors: verbal insults and as threats or assaults suffered. It is precisely because of overidentification and robust standard errors, which are used since in linear probabilities models the errors are heteroskedastic, that I chose to estimate the model through the Generalised Method of Moments (GMM) rather than through two-stage least squares (2SLS). Anyway, the 2SLS estimator is just a special case of the more general GMM estimator. The starting assumption to obtain both estimators is the exogeneity of the instruments: instrumental variables Z_i must be uncorrelated with the error terms u_i , that is $E(Z_i u_i) = \mathbf{0}$. The same relation can be rewritten as:

$$E\{Z_i(y_i - x_i\hat{\beta})\} = E\{g_i(\beta)\}, \quad (1)$$

where g_i is a set of moment conditions $L \times 1$, being L the number of instruments. The exogeneity of the instruments means that there are L moment conditions that will be satisfied at the true value of β :

$$E\{g_i(\beta)\} = 0. \quad (2)$$

The intuition behind GMM is to choose an estimator of β to solve $\bar{g}(\hat{\beta}) = 0$, where $\bar{g}(\hat{\beta})$ is the sample analog of $g_i(\beta)$.

If the number of instruments L is exactly the number of endogenous regressors K , then we have as many equations, the L moments conditions, as we do unknowns, the K coefficients in $\hat{\beta}$. In this case of exact identification, it is possible to find $\hat{\beta}$ that solves $\bar{g}(\hat{\beta}) = 0$ and the GMM estimator boils down to the 2SLS estimator (C.F. Baum *et al.*, 2003).

If, instead, the number of instruments is higher than the number of endogenous variables, $L > K$, we have an overidentification case and it is not possible to find a $\hat{\beta}$ that sets all the moment conditions to zero, because there are more equations than unknowns in the system. Hence, an $L \times L$ weighting matrix is used to construct a quadratic form in the moment condition. This way we obtain the GMM objective function $J(\hat{\beta})$:

$$J(\hat{\beta}) = n\bar{g}(\hat{\beta})'W\bar{g}(\hat{\beta}). \quad (3)$$

$\hat{\beta}$ is found by minimising the above function; the GMM estimator is obtained by putting to zero the first order derivative of the function $J(\hat{\beta})$, obtaining the following formula:

$$\beta_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'y . \quad (4)$$

The optimal weighting matrix to be chosen is the inverse of the covariance matrix of the moment conditions S, where S is:

$$S = \frac{1}{n}E(Z'uu'Z) = \frac{1}{n}E(Z'\Omega Z) \quad (5)$$

Hence:

$$\beta_{GMM} = (X'ZS^{-1}Z'X)^{-1}X'ZS^{-1}Z'y . \quad (6)$$

The 2SLS estimator is a special case of GMM estimator if the more restrictive assumption of homoskedasticity of the errors is made, that is if it is assumed that $\Omega = \sigma^2 I$.

The standard IV estimates are consistent even in presence of heteroskedasticity, but the conventional IV estimates of the standard errors are inconsistent, preventing valid inference. Therefore, albeit the conventional IV estimator is consistent even in presence of heteroskedastic errors, it is inefficient. The usual approach today when facing heteroskedasticity of unknown form is to use the generalized method of moments (GMM), introduced by Hansen in 1982 (C.F. Baum *et al.*, 2003).

In the models that are going to be presented in this thesis, GMM is used precisely because estimating weighted regressions implies assuming robust standard errors.

5. Main findings

5.1 The factors associated with life satisfaction

As already pointed out, one of the original aspects of this thesis is that it uses data describing migrants' own approach and perceptions of the host society in which they live. This enabled to conduct a study focused on analysing what is the impact that natives' attitude has on migrants' life, which is something new relative to most of the existing studies examining the determinants of natives' behaviours towards migrants.

This first paragraph presents what are the factors that positively or negatively affect migrants' life satisfaction. This first passage highlights some important correlations among the variables which will be used to implement the different models in the following sections.

Two dependent variables, life satisfaction and loneliness, are regressed on multiple covariates using a weighted least squares model. The weights used were already included in the ISTAT dataset and their aim is to balance our sample, in which migrants living in Southern Italy are overrepresented. Loneliness, beside life satisfaction, was chosen as a dependent variable because it gives a taste of how migrants perceive their lives and how they are happy and satisfied with living in Italy. This variable represents the answer to the following survey question: "Do you feel lonely here in Italy?" and takes the form of a binary variable, equal to one if the answers were "very lonely" or "quite lonely" and to zero if the respondent said "not very lonely" or "not lonely at all". As far as life satisfaction is concerned, it is measured through a dummy variable too. People could assess their own level of satisfaction through a scale from zero (not satisfied at all) to ten (very satisfied); those respondents whose score was above the sample average (7,56) were attributed the value one, those whose score was below the mean had the value zero. This is how the binary variable 'life satisfaction' was constructed.

Many controls are included in the regression: sociodemographic, cultural and lingual features, health status information, such as perceived health status, mental health and BMI (Body Mass Index), information about the length of stay in Italy, about the residence location and, finally, information about employment status and occupation. However, the covariates I want to bring attention to are all the variables measuring discrimination. Different kinds of discrimination are considered. First of all, discrimination the individual can experience in different places of everyday life, like at work, in the hospitals, in public offices, or in the residence neighbourhood by the neighbours themselves. Secondly a more 'random', we could say, type of discrimination is taken into consideration, that is discrimination that may take place in the roads, not connected with any specific activity or institution, but just related to racial

hatred and sourced from aesthetic and cultural differences. This kind of discrimination is measured through ‘verbal insults’ and ‘threats and assaults’. These variables correspond to the survey questions asking, respectively:” In Italy, have you ever been insulted, humiliated or offended in such a way that made you feel hurt?” and “In Italy have you ever been threatened or physically aggressed by known or unknown people, in a way that really scared you?”. These variables, as well as all the covariates concerning discrimination, are measured by dummies, taking value one if the migrant has ever been discriminated against and zero otherwise.

Table 21-Weighted Least Squares Regression of the Factors Associated with Life Satisfaction and Loneliness

Dependent variables	(1) Life satisfaction	(2) Loneliness
Discrimination at work	-0.053*** (0.017)	0.051*** (0.016)
Discrimination in the residence neighbourhood	-0.023 (0.024)	0.030 (0.025)
Discrimination in the Health Care system	-0.014 (0.032)	0.064* (0.034)
Discrimination in public offices	-0.042* (0.022)	0.0038 (0.022)
Threats/assaults	-0.061* (0.032)	-0.032 (0.029)
Verbal insults	0.0022 (0.019)	0.062*** (0.020)
Gender (1=female)	0.037*** (0.014)	0.0071 (0.014)
Married	0.061*** (0.011)	-0.097*** (0.012)
Employment status (base level: dependent employee)		
Collaboration	-0.16*** (0.049)	-0.029 (0.042)
Self-employed	-0.040* (0.022)	-0.036** (0.017)
Unemployed	-0.13*** (0.044)	0.090*** (0.034)
Inactive	-0.028 (0.042)	0.055* (0.033)
Index of Mental Health	0.13*** (0.012)	-0.11*** (0.011)
Perceived Health Status	0.11*** (0.017)	-0.078*** (0.017)
Observations	9,436	9,436
R-squared	0.122	0.121

*The model is estimated through weighted least squares. The dependent variables are equal to one if, respectively, respondents’ valuation of their life satisfaction, from 1 to 10, was above the sample average and if they defined themselves “very lonely” or “quite lonely”. Also the covariates of interest are binary: they take value one if the respondent has ever felt discriminated against in the various contexts (hospitals, public offices, workplace). All the controls are dummies or factor variables. Only part of the controls is displayed in the table. All the controls include: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin), lingual (ability to speak, understand, read and write in Italian) features, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness), employment status information and dummies concerning the specific occupation, information about smoking and drinking habits. Robust standard errors are displayed in brackets (*** p<0.01, ** p<0.05, * p<0.1).*

The observations used in this regression are less than 14.182 because some of the variables included have some missing values, as displayed in Table 8 in the third Section. In particular, ‘discrimination at work’ is available only for 11.704 observations and religion information, included in the cultural controls, is missing for 468 individuals. These values are missing simply because the respondents refused to answer to these questions.

It is important to make clear that the results presented in Table 21 represent simple correlations and not causal relationships between the variables. In fact, the exogeneity of the explanatory variables of interest, discrimination, cannot be guaranteed by implementing a weighted least squares model. The estimates generated by this model, indeed, may be subject to an omitted variable bias: despite as many controls as possible were included in the regression, some controls are unobservable, like being optimistic, or simply they were not provided in the dataset, such as household income, which certainly affects the level of individual life satisfaction.

Results concerning the estimated coefficients relative to the controls are in line with the findings of the previous studies about life satisfaction. In fact, it results that women are more satisfied with their lives than men, that married people are happier and less lonely and, finally, unemployment negatively affects life satisfaction and favours loneliness. Furthermore, as it was predictable, a higher index of mental health, indicating a better mental health status and a better perceived health status positively contribute to life satisfaction and make it less likely to feel lonely. Some other controls were included in the regression, but they were not reported in the table because they were not statistically significant, or because they were not of interest. For example, a higher BMI (Body Mass Index), corresponding to overweight or obese people, make people feel lonelier and its estimated coefficient is highly statistically significant. Also, even though contrary to what could be expected, education level is one of the controls, but it emerged that higher education level has no statistically significant impact nor on life satisfaction neither on loneliness.

As to discrimination covariates, even though not all the different discrimination measures have a significant impact on the two dependent variables, all the estimated coefficients have the expected sign. In fact, discrimination is negatively associated with life satisfaction and positively contributes to loneliness. It is interesting to notice that discrimination at work is the only covariate that is significant at a 1% significance level in both regressions: being discriminated against at work implies a higher level of dissatisfaction and loneliness. Loneliness is positively correlated with discrimination at work, as already said, and with discrimination in the health care system and verbal insults. On the other hand, life satisfaction has a negative correlation with discrimination at work, in the public offices and threats and

assaults. Particularly, migrants who have ever been threatened or aggressed are almost 8% less likely to define themselves satisfied with their lives and those who have been insulted or offended have 8% probability more to feel lonely with respect to those who have never been discriminated against.

These findings are in line with those obtained in a study conducted in 2011 by the sociologist M. Safi, who found out that discrimination is a major determinant in explaining migrants' lower life satisfaction with respect to natives. As it is reported in this study, it is not just the migration phenomenon itself that explains immigrants' subjective despair, but rather the hostility they may experience in everyday life in the host Country.

These preliminary results already give an idea of the real impact that discrimination has on immigrants. Discrimination is not just something "culturally wrong or unfair", but it has statistically significant consequences on migrants' life and behaviours. In the following sections, the new findings will point out that the impact of discrimination establishes concrete barrier in the process of integration and accounts for a psychological, but also physical, cost in immigrants' well-being.

5.2 The association between perceived discrimination and health

The aim of the present paragraph is to show what is the association between discrimination and health, exploiting the variations in discrimination perception deriving from being offended or harassed.

The literature shows that health gaps between migrants and natives exist in many developed Countries. The most common explanations to this phenomenon concern the lower socioeconomic status of immigrants, hence different health inputs and lower quality hospitals, or their lower ability to purchase medications and medical procedures. Yet, another potentially important determinant of racial and ethnic health gaps is discrimination (D.W. Johnston, G. Lordan, 2011).

On the basis of a concept developed by the WHO (World Health Organisation) in the Ottawa Charter (1986) and of a definition given by the scientific magazine "The Lancet", health can be defined as the capability to manage conditions of illness or of well-being by adapting to the external environment and using one's own resources, so to satisfy individual and social needs. Using economic terms, health has value, because, even though it has no exchange value, it positively influences individual utility, interpreted as the satisfaction of one's needs. According to the Grossman model (1972), health can be seen both as a consumption and as an investment good. It can be defined a consumption good, because, as already mentioned, it

positively affects individual utility, exactly like any other consumption good. However, it can be defined an investment good because it preserves or increases the level of individual human capital, thus making individuals productive and able to get some income, which is one of the factors that “produce” health in turn. Health is an outcome determined by various inputs, which are more or less manageable by the policymaker. These inputs include: the health care system, individual behaviours, socio-economic factors and environmental and genetic conditions. The latter aspects are indirectly influenced by education and income level.

What this paragraph wants to show is that, when referring to an immigrants’ population, a further indirect factor affecting the environmental conditions in which migrants live exists: discrimination. The analyses that follow will show that perceived discrimination is associated not only with lower values of the mental health index, or of perceived health status, but also with the incidence of chronic health problems. The association between perceived discrimination and health is estimated through two different models: a weighted least squares model, to point out interesting correlations among the variables, and a generalized method of moments (GMM) model, which gets closer to a causal interpretation of the coefficients, since instruments are used to tackle reverse causality bias.

The first results are presented in Table 22, which reports the estimates obtained by running a weighted least squares model of three different health indicators on ‘perceived discrimination’.

The dependent variables chosen are: index of mental health, perceived health status and chronic health problems. All of them are binary variables. Mental health is measured through the 12-item version of the General Health Questionnaire (GHQ) contained in the ISTAT questionnaire. The GHQ is a commonly used self-reported measure of mental health and consists of questions regarding the respondent’s emotional and behavioural health over the past few weeks. The twelve items in the GHQ are: ability to concentrate, sleep loss due to worry, perception of role, capability in decision making, whether constantly under strain, problems in overcoming difficulties, enjoyment of day-to-day activities, ability to face problems, whether unhappy or depressed, loss of confidence, self-worth, and general happiness (D. W. Johnston, G. Lordan, 2011). The higher the score obtained in the GHQ and the better the mental health status of the respondent. In the following model, the index of mental health was turned into a binary indicator by attributing the value one to all the individuals whose score was above the sample average of 53,8 (see Figure 2) and the value zero to all those below this average.

As far as ‘perceived health status’ is concerned, it is the answer to the survey question: “How healthy do you feel in general?”. Respondents could answer using a scale from one (very healthy) to five (not healthy at all), but I turned also this categorical variable into a dummy by

assigning the value one to all the respondents who declared a health status above the sample average, corresponding to a score of 1,89, and zero otherwise.

Finally, the third binary variable is ‘absence of chronic health problems’. It takes value one if the respondent is not affected by chronic problems and zero if she/he is.

The covariate of interest is perceived discrimination, measured as a dummy, taking value one if the individual has ever felt discriminated against and zero otherwise. It was constructed exploiting the information about discrimination in different contexts provided by the ISTAT survey. The variable ‘perceived discrimination’, thus, takes value one if the respondent has ever been discriminated against, since her/his arrival in Italy, in any of these contexts: at the workplace, while searching for a job, in public offices, in search of a house, in buying a car insurance, in the hospitals, in the residence neighbourhood. It is important to highlight that the variable at hand indicates perceived discrimination, not discrimination *tout court*, because measuring discrimination represents an empirical difficulty. In fact, perceived discrimination may not correspond to true discrimination, since it is a subjective perception, thus it can be measured with error.

The controls included in the model below are: sociodemographic, cultural, lingual features and residence information. In addition, controls concerning smoking and drinking habits, already described in Section 4, were added, since they are obviously correlated with health outcomes. Not all these controls, but only the ones whose estimated coefficients were considered interesting are included in Table 21. No employment information is included in the controls’ set because it would represent a bad control, that is it may be considered as an outcome variable itself, thus it could bias the estimates. Moreover, as Table 17 points out, employment status does not statistically differ between migrants who have perceived discrimination and those who have not, so it is not needed to include it among the controls.

In this first weighted least squares model, due to potential reverse causality, the estimated coefficients, presented in Table 22, do not point out any causal relationship between the variables, but just simple correlations.

Starting from observing the correlations among the controls and health, it can be noticed that, as it was expected, overweight and obese people are more likely to report a lower perceived health level and a higher number of chronic problems than normal weight people. It is also interesting to observe that married people report better mental and perceived health. A further interesting aspect is that people who used to smoke in the past are more likely to be affected by chronic health problems at present, while people who currently smoke have a lower index of mental health, on average, and report worse perceived health status. Moreover, people who define police more effective in the area where they live have higher mental health indices. This

could mean that police high effectiveness makes people feel safer and less stressed, though this cannot be affirmed because this coefficient may be biased: migrants living in a wealthier neighbourhood are more likely to be in a better health status and also to benefit from higher police effectiveness.

Table 22- Weighted Least Squares Regression of the Association between Discrimination and Health

Dependent variables	(1) Index of Mental Health	(2) Perceived Health Status	(3) Absence of chronic problems
Perceived discrimination	-0.14*** (0.014)	-0.087*** (0.010)	-0.074*** (0.0098)
Gender (1=female)	-0.040*** (0.014)	-0.038*** (0.010)	-0.033*** (0.010)
Married	0.047*** (0.013)	0.016* (0.0091)	-0.0044 (0.0087)
BMI (base level: normal weight)			
Underweight	0.043 (0.035)	0.023 (0.021)	0.028 (0.019)
Overweight	-0.0034 (0.014)	-0.016 (0.010)	-0.026*** (0.0096)
Obese	-0.0013 (0.022)	-0.086*** (0.019)	-0.072*** (0.018)
Smoking habits (base level: never smoked)			
Smoked in the past	-0.0032 (0.018)	-0.0086 (0.014)	-0.041*** (0.014)
Smoking at present	-0.044*** (0.016)	-0.015 (0.011)	-0.0054 (0.011)
Alcohol consumption away from meals (base level: never)			
Everyday	-0.24*** (0.086)	-0.082 (0.078)	-0.15* (0.085)
Sometimes in a week	-0.089** (0.037)	0.014 (0.019)	-0.0068 (0.019)
More rarely	-0.048*** (0.018)	0.0099 (0.012)	-0.0017 (0.012)
Police effectiveness (base level: not effective at all)			
Not very effective	0.044 (0.037)	0.031 (0.030)	0.018 (0.026)
Quite effective	0.066* (0.034)	0.037 (0.028)	0.039 (0.023)
Very effective	0.13*** (0.035)	0.053* (0.029)	0.028 (0.024)
Observations	11,960	11,865	11,840
R-squared	0.075	0.117	0.107

The model is estimated through weighted least squares. The dependent variables are equal to one if, respectively, the index of mental health is higher than the sample average, if the perceived health is very good or good and if the individual has no chronic health problems. Also the covariates of interest are binary: they take value one if the respondent has ever felt discriminated against in the various contexts (hospitals, public offices, workplace). All the controls are dummies or factor variables. Only part of the controls is displayed in the table. All the controls include: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin), lingual (ability to speak, understand, read and write in Italian) features, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness), information about smoking and drinking habits. Robust standard errors are displayed in brackets (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

As far as the covariate of interest is concerned, discrimination, it results that it is negatively correlated with all the three outcomes and in all the cases it is statistically significant: at a 1% significance level. It is interesting that discrimination correlates not only with mental health, which was predictable, but also with the incidence of chronic health problems: migrants being discriminated against are also those people who report chronic health problems.

However, these results represent simple correlations, as already said, because reverse causality and omitted variable bias makes these weighted least-squares estimates biased and inconsistent. Hence, to get closer to causality, the same association between the variables is estimated using the Generalised Method of Moments (GMM) and selecting ‘verbal insults’ and ‘threats and assaults’ as instruments. They are assumed to be valid and informative for the reasons explained in Section 4. Being the number of instruments higher than the number of endogenous variables, this is an overidentified model, which allows to conduct a Hansen’s test to verify the validity of the selected instruments. The p-value of this test, whose null hypothesis assumes all instruments are valid, is shown in the bottom lines of the following table.

The controls included in this model, even if not displayed in the above table, are the same used in the weighted least squares model.

It is to be observed that the number of observations is lower than 14.182, the whole sample, in all the three regressions. This is due to the fact that some observations are missing for some of the variables included in the regressions, simply because the respondents refused to answer certain survey questions. Particularly, for the dependent variables ‘perceived health status’ and ‘absence of chronic problems’ respectively 13.956 and 13.878 observations are available. The number of the missing values for the other variables used in the regressions are displayed in Table 8 in Section 3.

Table 23 illustrates the estimated coefficients of the first and second stage equations, distinguishing between the three different variables used to measure health: index of mental health, perceived health status and chronic health problems. Another distinction that is made in the table is the one between columns 1 and 2 and columns 3 and 4. In fact, the first two columns report the estimates using a weighted model, while the last columns show the estimates of a model without weights. As it can be observed, the estimates are qualitatively the same, but they are slightly different in quantitative terms. As to the index of mental health, for example, the second stage coefficient in the model without weights is slightly lower in absolute value. This means that for the observations of the Southern subsample, the correlation between perceived discrimination and health is weaker: once weights are added to the model in order to balance the sample, to attribute lower weight to the oversampled Southern observations, the estimated coefficient increase in absolute value.

Table 23-Instrumental Variable Regression of the Association between Discrimination against Immigrants and Immigrants' Health

	WEIGHTED		NOT WEIGHTED	
	(1)	(2)	(3)	(4)
	First Stage	Second Stage	First Stage	Second Stage
Dependent variable (1)	Perceived discrimination	Index of Mental Health	Perceived discrimination	Index of Mental Health
Verbal insults	0.46*** (0.018)		0.48*** (0.013)	
Threats and assaults	0.18*** (0.027)		0.18*** (0.021)	
Perceived discrimination		-0.25*** (0.056)		-0.19*** (0.042)
Observations	11,960	11,960	11,960	11,96
R-squared	0.204		0.212	
IV F-stat		525,104		975,099
Hansen's J test (p-value)				
Dependent variable (2)	Perceived discrimination	Perceived Health Status	Perceived discrimination	Perceived Health Status
Verbal insults	0.46*** (0.018)		0.48*** (0.013)	
Threats and assaults	0.18*** (0.027)		0.18*** (0.021)	
Perceived discrimination		-0.15*** (0.045)		-0.14*** (0.034)
Observations	11,865	11,865	11,865	11,865
R-squared	0.204		0.213	
IV F-stat		519,903		975,49
Hansen's J test (p-value)				
Dependent variable (3)	Perceived discrimination	Absence of chronic problems	Perceived discrimination	Absence of chronic problems
Verbal insults	0.46*** (0.018)		0.48*** (0.013)	
Threats and assaults	0.18*** (0.027)		0.18*** (0.021)	
Perceived discrimination		-0.16*** (0.042)		-0.12*** (0.031)
Observations	11,840	11,840	11,840	11,840
R-squared	0.204		0.213	0.080
IV F-stat		518,444		972,57
Hansen's J test (p-value)				

*The dependent variables are equal to one if, respectively, the index of mental health is higher than the sample average, if the perceived health is very good or good and if the individual has no chronic health problems. Also the endogenous variable is binary, it takes value one if the respondent has ever felt discriminated against in various contexts (hospitals, public offices, workplace). The instruments are dummies too, taking value one if the individual has ever been insulted or threatened. The controls included in the regression but not displayed in the table are: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin), lingual (ability to speak, understand, read and write in Italian) information, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness, information about smoking and drinking habits. Robust standard errors are displayed in brackets (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).*

Table 23 reports positive and statistically significant coefficients for the first stage equations. This means that migrants who have ever been offended or harassed are those migrants who perceive stronger discrimination. Both instruments have estimated coefficients which are highly statistically significant, all at a 1% significance level. Moreover, they appear to be strongly correlated with perceived discrimination, since the F-statistics reported at the bottom of the table are far above the conventional threshold of 10. In addition, as far as instruments validity is concerned, the p-values of the Hansen's test lead to accept the null hypothesis that instruments are valid.

As to the second stage estimates, they have the same signs as in the weighted least squares model in Table 22. Even isolating discrimination variations due to insults and threats there appears to be a statistically significant correlation between perceived discrimination and health. Moreover, this association is strong, especially when considering mental health, since the model estimates that migrants who perceive they have been discriminated against are 25% less likely to obtain a mental health index higher than the average of the sample. The same share is 15% when referred to perceived health status and 16% referred to the absence of chronic health problems.

Despite the corresponding results are not reported in the thesis, the same model was implemented using the dependent variable 'index of mental health' as a continuous variable. It was rescaled so that its minimum value was 0,76 and its maximum 6,94. The mental health coefficient so estimated is always negative and it measures -0,46. Also in this case, the coefficient is very high, since it indicates that migrants who were discriminated against report, on average, 0,46 points less in their mental health index, out of a maximum score of 6,94, with respect to not-discriminated-against ones.

It is interesting to notice that, while health conditions reported by the respondents refer to the present, questions about discrimination ask whether the migrants "has ever been discriminated against since her/his arrival in Italy". Thus, discrimination episodes may have occurred even long time ago, but its consequences on health may still be there. Hence, it can be concluded that discrimination can have long term effects on migrants' health status.

The channels through which discrimination is associated with health are different. Firstly, feeling discriminated against induces negative psychological consequences, which turns to be negative for physical health too. Secondly, due to 'taste' discrimination by the hospital personnel, immigrants may be provided with lower healthcare and assistance. Moreover, discrimination can imply a lower socioeconomic position, thus limited access to higher quality hospitals and healthcare.

5.3 The association between health and employment status

The aim of this paragraph is to analyse the association between health and labour market outcomes, in order to estimate the effects that health indirectly has on a relevant economic indicator, namely employment status.

Firstly, the association between these two variables is estimated through a weighted least squares model. Though, the estimates obtained using this estimation strategy may be biased and inconsistent because of potential reverse causality and omitted variable bias, but it is interesting to start from this simple model to observe what are the associations between some controls of interest and the dependent variable, employment status.

The controls included in the following regression are sociodemographic, cultural, lingual features, information about smoking and drinking habits, residence information and a dummy indicating whether the respondent has ever worked in Italy.

The sample size is reduced because the dependent binary variable, employment status, only accounts for employed and unemployed people, thus excluding inactive respondents from the original sample. People active in the labour market are 10.895. 'Employment status' takes value one if the individual is employed and zero if she/he is unemployed and refers to the current employment status of migrants.

What is to be primarily noticed observing Table 24 is the fact that all the three covariates measuring health are positively correlated with employment status. This is nothing surprising, because it is intuitive to think that healthier people are more likely to be employed. However, this estimate may be biased, because there may be some unobservable variables, that are not controlled for, which affect both health and employment status. For example, migrants living in wealthier areas can benefit from a more dynamic labour market, wider social networks and may also have easier access to better hospitals and healthcare services. Moreover, another source of bias may be reverse causality: employment status can affect health in the same way health affects employment status. In fact, employed people may report better mental health indices and better general health conditions due to lower levels of stress and frustration.

Other interesting estimates concern gender and marital status: being a woman is negatively correlated with being employed, as well as being married makes it less likely to have an employment. It is known that for a native Italian woman employment chances are lower than for men, then it is not surprising that the same chances may be even lower for migrant women. As expected, migrants who master a fluent Italian are more likely to be employed than those who are not able to speak Italian. Contrary to expectations, instead, it emerges that education

level has no statistically significant correlation with employment status: having a higher education level does not affect the chances of being employed in a significant way.

Table 24-Weighted Least Squares Regression of the Association between Health and Employment Status

Dependent variable	(1) Employment status	(2) Employment status	(3) Employment status
Index of Mental Health	0.076*** (0.010)		
Perceived health Status		0.044*** (0.015)	
Absence of chronic health problems			0.030* (0.015)
Gender (1=female)	-0.023** (0.011)	-0.024** (0.011)	-0.026** (0.011)
Married	-0.051*** (0.010)	-0.048*** (0.010)	-0.047*** (0.010)
Education level (base level: no education)			
Primary education	0.0048 (0.028)	0.0041 (0.028)	0.0066 (0.028)
Secondary education	0.00090 (0.021)	-0.00094 (0.022)	-0.00022 (0.022)
High school education	0.0019 (0.021)	0.00048 (0.021)	0.0020 (0.021)
Tertiary education	0.013 (0.023)	0.0096 (0.023)	0.013 (0.023)
Ability to speak Italian (base level: not good at all)			
Not good	0.18*** (0.067)	0.19*** (0.067)	0.19*** (0.067)
Quite good	0.18*** (0.069)	0.19*** (0.069)	0.19*** (0.069)
Very Good	0.15** (0.070)	0.17** (0.071)	0.17** (0.071)
Continent of origin (base level: Europe)			
Africa	-0.070*** (0.018)	-0.072*** (0.018)	-0.072*** (0.018)
Asia	0.0037 (0.017)	0.0061 (0.017)	0.0064 (0.017)
North America	0.11*** (0.030)	0.11*** (0.028)	0.11*** (0.028)
Central and South America	-0.032* (0.019)	-0.032* (0.019)	-0.030 (0.019)
Oceania	0.40*** (0.056)	0.35*** (0.055)	0.34*** (0.055)
Observations	9,240	9,174	9,153
R-squared	0.133	0.123	0.122

*The model is estimated through weighted least squares. The dependent variable is a dummy equal to one if the respondent is employed, to zero if she/he is unemployed. Inactive people are excluded from the sample. The covariates of interest are dummies too: they are equal to one if, respectively, the index of mental health is higher than the sample average, if the perceived health is very good or good and if the individual has no chronic health problems. Only part of the controls is displayed in the table. All the controls include: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin), lingual (ability to speak, understand, read and write in Italian) features, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness) and dummies concerning the specific occupation, information about smoking and drinking habits. Robust standard errors are displayed in brackets (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).*

The second model which is implemented to analyse the association between health and employment status is a GMM model, which uses two instrumental variables as well: ‘verbal insults’ and ‘threats and assaults’. The aim of this model is to tackle reverse causality and omitted variable bias and to get consistent estimates, as explained in the “Empirical strategy” section. The estimates obtained and presented in Table 25 constitute just a first step towards causality.

The controls included in this model are the same used in the weighted least squares regression and the great number of missing values is justified by the same reasons.

Also this table reports the estimates both for a weighted model and for a model without weights. In this case, both first and second stage estimated coefficients are generally lower, in absolute value, for the no weights model, indicating that the correlation between discrimination and health, as well as between health and employment status, is lower for the Southern observations, which are oversampled.

The first stage equation of this model is precisely the reduced form of the previous GMM model estimating the relationship between perceived discrimination and health. Indeed, this first stage equation estimates the direct relation between health and insults and assaults without “filtering” it through the level of perceived discrimination. As it emerges from Table 21, the direct effect that discrimination has on health is always negative, though its modulus and its statistical significance are different for the three measures of health. Particularly, as it is expected, the impact of perceived discrimination is stronger when considering psychological health: an immigrant who has ever been threatened or aggressed is 11% less likely to have a good mental health index and the same percentage increases to 13% for migrants who have experienced verbal insults. Moreover, these estimates are statistically significant at a 1% significance level and the F-statistics measures 31, far above the threshold level of 10. Instead, the correlation between discrimination and health is understandably weaker when considering the absence of chronic health problems as a measure of health. In fact, the estimated coefficient for the instrument ‘threats and assaults’ is not even significant and instruments are quite weak, since the F-statistics has value 12,5, just above the conventional threshold. On the whole, it can be said that being aggressed or offended is harmful for migrants’ health.

As to the second stage equation, which estimates the effect that exogenous variations of health due to insults and harassment have on employment, all the estimated coefficients indicate a positive contribution of health to employment. The coefficients that are obtained are even quite high: people with a better mental health index are 28% more likely to be employed, as well as people perceiving a better health are even 37% more likely to have a job.

Table 25-Instrumental Variable Regression of the Association between Health and Employment

	WEIGHTED		NOT WEIGHTED	
	(1) First stage	(2) Second stage	(3) First stage	(4) Second stage
Dependent variables (1)	Index of Mental Health	Employment status	Index of Mental Health	Employment status
Verbal insults	-0.13*** (0.023)		-0.10*** (0.017)	
Threats/assaults	-0.11*** (0.034)		-0.11*** (0.025)	
Index of Mental Health		0.28** (0.13)		0.19** (0.089)
Observations	9,240	9,240	9,240	9,240
R-squared	0.078		0.064	
IV F-stat		31,00		46,67
Hansen's test		1,00		1,00
Dependent variables (2)	Perceived Health Status	Employment status	Perceived Health Status	Employment status
Verbal insults	-0.079*** (0.018)		-0.071*** (0.014)	
Threats/assaults	-0.097*** (0.029)		-0.078*** (0.022)	
Perceived Health Status		0.37** (0.18)		0.27* (0.14)
Observations	9,174	9,174	9,174	9,174
R-squared	0.095		0.085	
IV F-stat		19,04		27,95
Hansen's test		1,00		0,896
Dependent variables (3)	Absence of Chronic Health Problems	Employment status	Absence of Chronic Health Problems	Employment status
Verbal insults	0.075*** (0.018)		0.053*** (0.012)	
Threats/assaults	0.033 (0.026)		0.035* (0.019)	
Absence of Chronic Health Problems		0.69* (0.39)		0.50* (0.28)
Observations	9,153	9,153	9,153	9,153
R-squared	0.084		0.068	
IV F-stat		12,05		15,3
Hansen's test		1,00		1,00

The model is estimated through the Generalised Method of Moments (GMM). The dependent variable is a dummy equal to one if the respondent is employed, to zero if she/he is unemployed. Inactive people are excluded from the sample. Also the endogenous variables are equal to one if, respectively, the index of mental health is higher than the sample average, if the perceived health is very good or good and if the individual has no chronic health problems. The model is overidentified, since two instruments are used: they are two dummies, taking value one if the individual has ever been insulted or threatened. The controls included in the regression but not displayed in the table are: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin) and lingual (ability to speak, understand, read and write in Italian) information, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness, information about smoking and drinking habits. Robust standard errors are displayed in brackets (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

The magnitude of these coefficients seems to suggest that some controls are probably missing in the regression. Hence, what can be affirmed with certainty is that health is positively associated with employment status in a statistically significant way, but as to the quantitative aspect, future analyses should make it clearer.

In other words, as it is intuitive to think, health is one of the inputs of human capital: bad health conditions, both psychological and physical, hinders the development of human capital and thus makes it more difficult to be employed. The indirect association between insults, harassment and employment status can be explained as follows: offences and threats, undermining psychological and physical health, lowers the human capital of those migrants who perceive they are discriminated against and thus reduces their chances of being employed.

5.4 The association between perceived discrimination and social integration

The last model that is going to be presented in this thesis aims at analysing what kind of correlation exists between discrimination and integration. It is important to highlight that the kind of integration that is considered is the so-called ‘social integration’, to be distinguished from structural integration. The latter refers, for example, to the educational attainments and to the immigrants’ position in the labour market. Instead, socio-cultural integration is “the social contacts that ethnic minorities have with individuals and agencies of the majority group” (K. Vanclyusen and M. Van Craen, 2010). Some indicators of this kind of integration are immigrants’ values, norms, opinions and proficiency of the majority language. Social integration can be said to be a much more subjective integration measure and, once more, this is in line with the spirit of this work, which uses data describing migrants’ subjective perceptions.

It is interesting to investigate the correlation between discrimination and integration because in the literature different studies have pointed out different correlations. According to the assimilation theory (Gordon, 1964) the more an individual is integrated in the host society and the less she/he is discriminated, whereas according to the ethnic competition theory (Portes *et al.*, 1980) the more integrated a migrant is and the more she/he will perceive discrimination. These theories basically state that discrimination and integration are to be considered jointly, because they are mutually determined. To say it in econometric terms, there exists reverse causality: discrimination contributes to determine integration as well as the integration process affects the probability of being discriminated against.

The association between discrimination and integration is first investigated through a weighted least squares model, whose estimates are illustrated in Table 26.

Table 26-Weighted Least Squares of the Association between Discrimination and Integration

Dependent variables	(1) Interest in Italian politics	(2) Watching Italian TV news	(3) Television language	(4) Cinema at least once a year
Perceived discrimination	0.075*** (0.013)	0.0065 (0.0092)	-0.022* (0.012)	-0.0057 (0.013)
Gender (1=female)	-0.13*** (0.013)	-0.0060 (0.0099)	0.065*** (0.012)	0.017 (0.012)
Married	0.013 (0.012)	-0.0077 (0.0089)	-0.085*** (0.011)	-0.057*** (0.012)
Education level (base level: no education)				
Primary education	-0.0087 (0.034)	-0.034 (0.028)	-0.024 (0.032)	-0.073*** (0.024)
Secondary education	0.026 (0.025)	0.0049 (0.020)	0.033 (0.024)	-0.033 (0.020)
High school education	0.078*** (0.025)	0.0051 (0.020)	0.019 (0.024)	0.028 (0.021)
Tertiary education	0.17*** (0.028)	0.022 (0.021)	-0.032 (0.027)	0.13*** (0.025)
Ability to speak Italian (base level: not good at all)				
Not good	-0.061 (0.065)	0.076 (0.076)	-0.0066 (0.073)	0.081* (0.047)
Quite good	0.013 (0.070)	0.23*** (0.075)	0.097 (0.074)	0.13** (0.054)
Very Good	0.0044 (0.073)	0.24*** (0.075)	0.092 (0.076)	0.096 (0.059)
Ability to understand Italian (base level: not good at all)				
Not good	0.098 (0.065)	0.023 (0.076)	0.040 (0.073)	-0.060 (0.050)
Quite good	0.081 (0.070)	0.0053 (0.076)	0.055 (0.075)	-0.11* (0.057)
Very Good	0.14* (0.073)	0.041 (0.076)	0.10 (0.076)	-0.036 (0.061)
Observations	12,106	11,368	11,368	12,106
R-squared	0.155	0.171	0.160	0.170

The model is estimated through weighted least squares. The dependent variables are binary: they are equal to one if, respectively, migrants are interested in Italian politics, are used to watching Italian TV news, to watch Italian TV channels and, finally, if they go to the theatre at least once a year. Also the covariate of interest is binary: it takes value one if the respondent has ever felt discriminated against in various contexts (hospitals, public offices, workplace). Only part of the controls is displayed in the table. All the controls used are: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin), lingual (ability to speak, understand, read and write in Italian) features, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness). Robust standard errors are displayed in brackets (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Despite the coefficients so estimated may be biased and inconsistent because of the reverse causality bias explained above, this first model gives a hint of what are the associations between the variables. In both these models the number of observations does not coincide with the overall sample, because some people refused to answer to questions related to integration and discrimination. More numeric details are provided in Table 8, in Section 3.

As far as the dependent variables are concerned, four different variables are used as integration measures: migrants' interest in Italian politics, their habit to watch Italian TV news, to watch Italian TV channels and, finally, whether they go to the theatre at least once a year. All these variables are dummies. The first one takes value one if migrants are interested in Italian politics and zero otherwise, as well as 'Italian TV news' and 'television language' have value one if the respondent does watch it and zero if she/he does not. Also 'cinema at least once a year' is a dummy equal to one if the respondent usually goes to the cinema at least once a year and to zero otherwise.

It is important to notice that the indicators of integration chosen concern what I would define "unobservable integration". In fact, the dependent variables just described identify migrants' private behaviours, that is those attitudes that migrants usually take on in their everyday life, which indicate what is their real level of cultural assimilation and not the effort they make to be socially accepted.

The explanatory variable of interest, discrimination, is a binary variable constructed as in the previous models.

The controls included in the following regression are sociodemographic, cultural, lingual features and residence information. Employment information is not included among the controls because it would be a bad control, since migrants' employment status can have been determined after they were discriminated against, hence this variable itself may be considered an outcome variable in the regression at hand. If 'employment status' was used as a control, the estimated coefficients would be biased, because migrants discriminated against may have lower probability to be integrated even if employed and in the absence of discrimination, due to, for example, some personal characteristics that are unobservable.

Table 25 illustrates the coefficients estimated through the weighted least squares model. What emerges looking at the correlations between the controls and the dependent variables is that the signs of the controls' coefficients are very different in the four regressions. It is interesting to notice that people with higher education levels are more likely to be interested in Italian politics or to go to the cinema once a year, but there is no significant correlation with watching Italian television for the more educated migrants with respect to the not educated ones. It is surprising to observe that understanding Italian language makes no difference in terms of socio-cultural integration relative to not understanding Italian at all, except for interest in Italian politics, whose coefficient is weakly significant. As far as the explanatory variable of interest is concerned, discrimination, results are mixed. In fact, its estimated parameter is negative and statistically significant in column 3, positive and significant in column 1, not significant in column 2 and 4. Hence, looking at these results, it is unclear what is the association between

discrimination and integration. These mixed findings may be due to reverse causality between the two variables, so the following model estimates the same association through an instrumental variable regression.

The instrumental variable estimator which is used is the GMM. The instruments selected are the same used in the previous model. Estimates are reported in Table 27.

The first stage equation of this model is exactly the same as in the first model, the one estimating the association between discrimination and health, thus coefficients are positive, highly statistically significant and the value of the F-statistics is above the conventional threshold of 10.

As to the second stage weighted estimates, they are very different from one another. Particularly, none of them is significant except for the one in the first panel regression, whose dependent variable is ‘interest in Italian politics’. The coefficient relative to this first regression is positive, thus indicating that those migrants who have ever been discriminated against are also the ones who are more interested in Italian politics: they are even 27% more likely, with respect to not-discriminated-against migrants, to care about Italian politics. This association could be interpreted as follows: migrants who perceive they are discriminated against make more effort in assimilating the Italian culture, thus they show a higher interest in Italian politics with the aim of being socially accepted.

As to the estimates in the subsequent panels, they are not significant and they have opposite signs. Also the estimates obtained through the model without weights are not always significant, but they all share the same positive sign. This signals that migrants discriminated against who live in the South are those migrants who have a higher level of social integration.

All things considered, in estimating these models ISTAT data does not point out any significant association that can be clearly interpreted. Such mixed results do not allow to draw any conclusion, nor in favour of the assimilation theory, neither of the ethnic competition theory.

Table 27-Instrumental Variable Regression of the Association between Discrimination and Social Integration

	WEIGHTED		NOT WEIGHTED	
	(1) First stage	(2) Second stage	(3) First stage	(4) Second stage
Dependent variable (1)	Perceived discrimination	Interest in Italian politics	Perceived discrimination	Interest in Italian politics
Verbal insults	0.47*** (0.018)		0.49*** (0.013)	
Threats/assaults	0.18*** (0.027)		0.18*** (0.021)	
Perceived discrimination		0.27*** (0.052)		0.27*** (0.039)
Observations	12,106	12,106	12,106	12,106
R-squared	0.197		0.204	
IV F-stat		541,396		1039,69
Hansen's J test (p-value)		0,9845		1,00
Dependent variable (2)	Perceived discrimination	Italian TV news	Perceived discrimination	Italian TV news
Verbal insults	0.47*** (0.019)		0.49*** (0.013)	
Threats/assaults	0.17*** (0.029)		0.17*** (0.022)	
Perceived discrimination		0.038 (0.037)		0.060** (0.024)
Observations	11,368	11,368	11,368	11,368
R-squared	0.194	0.168	0.202	0.050
IV F-stat		489,433		956,005
Hansen's J test (p-value)		1,00		0,853
Dependent variable (3)	Perceived discrimination	Italian TV channels	Perceived discrimination	Italian TV channels
Verbal insults	0.47*** (0.019)		0.49*** (0.013)	
Threats/assaults	0.17*** (0.029)		0.17*** (0.022)	
Perceived discrimination		-0.052 (0.047)		0.0055 (0.034)
Observations	11,368	11,368	11,368	11,368
R-squared	0.194		0.202	0.081
IV F-stat		489,433		956,005
Hansen's J test (p-value)		0,9879		1,00

Dependent variable (4)	Perceived discrimination	Cinema at least once a year	Perceived discrimination	Cinema at least once a year
Verbal insults	0.47*** (0.018)		0.49*** (0.013)	
Threats/assaults	0.18*** (0.027)		0.18*** (0.021)	
Perceived discrimination		-0.012 (0.055)		0.016 (0.036)
Observations	12,106	12,106	12,106	12,106
R-squared	0.197	0.166	0.204	0.133
IV F-stat		541,396		1039,69
Hansen's J test (p-value)		1,00		1,00

The model is estimated through the Generalised Method of Moments (GMM). The dependent variables are binary: they are equal to one if, respectively, migrants are interested in Italian politics, are used to watching Italian TV news, to watch Italian TV channels and, finally, if they go to the theatre at least once a year. Also the endogenous variable is binary, it takes value one if the respondent has ever felt discriminated against in various contexts (hospitals, public offices, workplace). The instruments are dummies too, taking value one if the individual has ever been insulted or threatened. The controls included in the regression but not displayed in the table are: sociodemographic (gender, age, education level, BMI, length of stay in Italy), cultural (religion, Continent of origin), lingual (ability to speak, understand, read and write in Italian) information, information about residence (size of the municipality, macro-area (North, Centre, South and Isles), frequency of police controls and degree of police effectiveness), information about smoking and drinking habits. Robust standard errors are displayed in brackets (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6. Conclusion

In the previous paragraphs it was investigated whether discrimination against immigrants has any correlation with some social and economic outcomes. The dimensions which are put in relation with discrimination are socio-cultural integration, health and employment status.

As to the first aspect, integration, modelling ISTAT data does not allow to draw any specific conclusion about the correlation between integration and discrimination, because the obtained estimates are not statistically significant and their signs are different. Hence, it cannot be said whether discrimination against immigrants slows down their integration process, or, as the ethnic competition theory would suggest, it works as a sort of stimulus, pushing migrants to assimilate Italian culture. This finding is not surprising if compared with previous studies: many of them reach opposite conclusions and others (see Vancluysen, K., Van Craen, M., 2010) came to mixed results, not univocally interpretable.

As far as the second dimension considered is concerned, health outcomes, it emerged that discrimination is negatively associated not only with mental health, as it was predictable, but also, even if in a weaker fashion, with physical health. This correlation is interesting, because it provides an alternative explanation to the fact that migrants' health status is usually worse than natives' one, which is usually ascribed to lower socioeconomic status. The present findings hint at the possibility that discrimination can be an alternative factor to explain migrants' worse health conditions. The channels through which discrimination can affect migrants' health are mainly two of them. First, the perception of being discriminated against certainly generates high levels of stress and frustration which could result in a poorer psychological and physical health. Secondly, if health professionals have a 'taste' for discrimination, then migrants are likely to be provided with lower health care standards. These findings are consistent with the ones by Johnston D.W. and Lordan G. (2011), who suggest which are the channels linking discrimination to worse health status.

The third aspect analysed in this thesis is the association between health and immigrants' employment status. Findings tell that people in a better health status are much more likely to be employed than sick ones. This is nothing new nor surprising, though if this result is put in relation with the previous findings, it can be seen that, since discrimination is negatively correlated with health, which is positively associated with employment status, then discrimination hinders migrants' chances to get employed. Also in this case there can be a double explanation to this correlation. First, discrimination, undermining psychological and physical health, lowers the human capital of those migrants who perceive they are discriminated

against and thus reduces their chances of being employed. Secondly, another channel through which discrimination can affect employment opportunities is employers' 'taste' for discrimination: employers may carry out discriminatory hiring practice, thus preferring to hire only native workers (Becker, 1957).

All the results that have just been summarised represent a first exploration of causal relationships between the variables. Indeed, despite every association is estimated using an instrumental variable strategy to overcome reverse causality and omitted variable bias, the assumption that the selected instruments are perfectly exogenous and uncorrelated with the structural equation error term may be criticized. It is hoped that this work can be considered a point of departure for deeper analyses aiming at identifying causal relationships.

Nevertheless, this thesis is characterised by some original elements. First of all, it is one of the few studies which conduct econometric analyses exploiting data about Italian immigrants. Secondly, and more important, the data used is drawn from the survey "Condizione e Integrazione Sociale dei Cittadini Stranieri" (2011), which collects very recent data. In fact, the survey was carried out in 2011 and the dataset was published only on October 2016. Therefore, this is one of the first studies, if not the very first one, manipulating this data.

In conclusion, it is hoped that this work will inspire future studies about Italian immigrants, able to point out causal relationships of social and economic interest.

References

- ANGRIST, J.D. and PISCHKE, J. (2008). Mostly harmless econometrics. *Princeton University Press*.
- ARROW, K., (1971). The theory of discrimination. *Industrial Relations Sections working paper no. 30A*.
- BARONE, G., D'IGNAZIO, A., DE BLASIO, G. and NATICCHIONI, P. (2014). Mr. Rossi, Mr. Hu and politics: the role of immigration in shaping natives' political preferences. *IZA Discussion Paper no. 8228, Institute for the Study of Labour*.
- BAUER, T. K., LOFSTROM, M. and ZIMMERMANN, K. F. (2000). Immigration policy, assimilation of immigrants and natives' sentiments towards immigrants: evidence from 12 OCED-Countries. *IZA Discussion Paper no. 187, Institute for the Study of Labour*.
- BAUM, C.F., SCHAFFER, M. and STILLMAN, S. (2003). Instrumental variables and GMM: estimation and testing. *The Stata Journal*.
- BORJAS, G.J. (2003). The economic analysis of migration (The handbook of Labour Economics, Volume 3, Part A, Chapter 28). *Elsevier Science B.V.*
- CAMERON, C. and TRIVEDI, P.K. (2010). Microeconometrics using Stata-revised edition. *STATA Press*.
- CARD, D., DUSTMANN, P. and PRESTON, I. (2009). Immigration, wages and compositional amenities. *NBER Working Paper no. 15521, National Bureau of Economic Research*.
- D'HOMBRES, B. and NUNZIATA, L. (2015). Wish you were here? Quasi-experimental evidence on the effect of education on attitude toward immigrants. *IZA Discussion Paper no. 8880, Institute for the Study of Labour*.
- DANCYGIER, R.M. and LAITIN, D. (2014). Immigration into Europe: economic discrimination, violence and public policy. *Annual Review of Political Science*.
- EICHHORST, W., FRATTINI, T., GIULIETTI, C., GUZI, M., NOWOTNY, K., KENDZIA, M.J., HUBER, P., MONTI, P. and VANDEWEGHE, B. (2011). The integration of migrants and its effects on the labour market. *IZA Research report no.40, Institute for the Study of Labour*.

EUROSTAT (2016). Non-national population by group of citizenship, 1 January 2016. Available online at [http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Non-national_population_by_group_of_citizenship_1_January_2016_\(%C2%B9\).png](http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Non-national_population_by_group_of_citizenship_1_January_2016_(%C2%B9).png)

HANSON, G.H., SCHEVE, K.F. and SLAUGHTER, M.J. (2005). Public finance and individual preferences over globalization strategies. *NBER Working Paper no. 11028, National Bureau of Economic Research*.

ISTAT (2016). Conti economici territoriali-anno 2015 [online]. Available at https://www.istat.it/it/files/2016/12/Conti-regionali_2015.pdf?title=Conti+economici

JOHNSTON, D.W. and LORDAN, G. (2011). Discrimination makes me sick! An examination of the discrimination-health relationship. *The Journal of Health Economics*.

MIGRANT INTEGRATION POLICY INDEX (2015). MIPEX in Italia [online]. Available at <http://www.mipex.eu/italy>.

MINISTERO DEGLI AFFARI ESTERI E DELLA COOPERAZIONE INTERNAZIONALE (2017). Cittadinanza. Available at http://www.esteri.it/mae/it/italiani_nel_mondo/serviziconsolari/cittadinanza.html.

PERI, G. (2010). The impact of immigrants in recession and economic expansion. *Migration Policy Institute*.

RABBY, F. and RODGERS, W.M. (2011). Post 9-11 U.S. Muslim labour market outcomes. *International Atlantic Economic Society*.

SAFI, M. (2010). Immigrants' life satisfaction in Europe: between discrimination and assimilation. *European Sociological Review-Volume 6, number 2, 2010*.

SENIK, C., STICHNOTH, H. and VAN DER STRAETEN, K. (2008). Immigration and natives' attitudes towards the welfare state: evidence from the European social survey. *Springer Science +Business Media B.V.*

THE LANCET (2009). What is health? The ability to adapt. *The Lancet-Editorial, Volume 373, March 7, 2009*

VANCLUYSEN, K. and VAN CRAEN, M. (2010). Integration and perceived discrimination: two competing hypotheses testes among persons of Moroccan and Turkish descent in Belgium. *INED, Institut national des Études Démographiques*.

ZINCONE, G. and CAPONIO, T. (2005). Immigrant and Immigration policy-making: the case of Italy. *IMISCOE Working Paper: Country report*.

Legislation

Council Directive 2003/109/EC of 25 November 2003 concerning the status of third-country nationals who are long-term residents.

L. 5 Febbraio 1992, n.91.