THE
ECONOMIC
JOURNAL

The Economic Journal, **125** (*August*), F82–F114. Doi: 10.1111/ecoj.12232 Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

IMMIGRATION, HOUSING DISCRIMINATION AND EMPLOYMENT*

Tito Boeri, Marta De Philippis, Eleonora Patacchini and Michele Pellizzari

We use a new data set on Italy and a novel identification strategy to analyse the relationship between migrants' employment status and the percentage of non-Italians living nearby. Our data contain information at the very local level and are representative of both legal and illegal migrants. Identification exploits the physical characteristics of local buildings as a source of exogenous variation in the incidence of migrants. We find that migrants residing in more immigrant-dense areas are less likely to be employed. This penalty is higher if the migrants leaving nearby are illegal and it is not mitigated if they are from own ethnic group or more proficient in Italian.

In the 10 years predating the Great Recession Europe received twice as many immigrants (relative to the resident population) than the US and 4–5 times as many as Japan (OECD, 2009). The main motivation of these flows was finding a job. Family reunification and asylum seeking were fairly marginal. The main destination of migrants was Southern Europe, where the stock of foreign born increased by some 5 million in Spain and 3 million in Italy within a decade, doubling the migrant population in several urban areas over a few years. Most of these migrants came in illegally hoping for a subsequent regularisation and there was no major social housing programme activated in the destination countries responding to these mass migration flows, just while house prices were booming. Newcomers did not locate uniformly across the board: they often found residence where other persons of the same nationality or ethnic group were already living, increasing the residential concentration of migrants in urban areas. How did these residential location patterns affect the economic integration of migrants, notably their probability of finding a job? Is there scope for policies reducing residential concentration of migrants in specific urban areas?

While in the US research related to the effects of residential concentration of migrants on their economic and social integration is long standing (Borjas, 1995; Cutler and Glaeser, 1997; Ross, 1998; Cutler *et al.*, 1999; Card and Rothstein, 2007), in Europe, notably in Southern Europe, there is a very thin scientific literature on this issue, despite the highly controversial public debate. This paucity of studies for Europe is mainly due to a lack of detailed data on ethnic minorities and migrants, especially at the city level (Bisin *et al.*, 2011). Moreover data collection is particularly problematic

* Corresponding author: Tito Boeri, Department of Economics, Bocconi University, via Roentgen 1, 20136 Milan, Italy. Email: tito.boeri@unibocconi.it.

We thank the European Bank for Reconstruction and Development (EBRD) for financing this project, Francesca Pissarides for her contribution to the survey design and Carlo Erminero & Co. for their professional services. We are grateful to Massimo Baldini and Marta Federici for having given us access to microdata of their experiment on discrimination in the Italian housing market. We benefited from the valuable comments of Francesco Fasani, Gianmarco Ottaviano, Giovanni Peri, Fabiano Schivardi, Stuart Rosenthal, John Yinger and seminar participants at Bocconi University. Gaetano Basso, Vittorio Bassi, Barbara Biasi, Andrea Catania, Hristo Dachev, Rachele Poggi and Alessandro Vecchiato provided skilful research assistance. The usual disclaimer applies. Michele Pellizzari acknowledges financial support from the Swiss National Science Foundation through the NCCR-LIVES programme.

given the relatively large share of illegal immigrants in Southern Europe, who are not represented by surveys drawing from population registers. In this context, a key issue is the relatively large share of illegal migrants in Southern Europe, who are not represented by surveys drawing from population registers.

In this article, we contribute to filling this gap by estimating the causal effect of the local concentration of immigrants on their employment prospects using a new and unique survey conducted in 2009 in eight cities located in the North of Italy. Italy is a particularly interesting case to study, as migrants appear to be very highly concentrated in their residential locations. Based on official data from the 2001 census, the coefficient of variation in the number of resident migrants across census tracts is twice as large as that of natives (1.793 against 0.966 for natives). At the same time, the fraction of illegal migrants is large: Italian migration policies require that migrants have a job before entering the country and getting a legal permit.[1] This requirement is never fulfilled as migrants come in illegally and then search for a job, hoping for a migration amnesty or a letter of an employer allowing them to leave the country and come back legally. The poor design and enforcement of Italian migration policies make the case even more interesting. Illegal status may indeed interact with residential location and affect the relationship between housing concentration of migrants and employment.

The focus on a limited group of eight cities in Northern Italy allowed us to design the sampling frame very carefully, also including illegal migrants, and to use a comprehensive questionnaire for the interviews.

Our study is novel along several dimensions. First, thanks to a particular sampling frame which randomly draws blocks from the continuum of map locations within cities (see subsection 2.1), our survey covers both legal and illegal migrants. Around 20% of migrants in our data are illegally resident in the country and they are far from being a random subgroup of the entire population. Compared to the legally resident, illegal migrants appear to be on average men, younger, slightly less educated, less proficient with the Italian language and more likely to rely on informal networks to look for employment.

Second, the data are available at a very detailed level of geographical disaggregation, namely, we can identify the exact address where each interviewed person resides. Hence, we can define residential concentration more accurately than in most previous studies, that is at the level of the individual block.

Third, by merging our survey with data from the national census, we are able to obtain information about various physical attributes of the buildings in each block, which are valid instruments for migrants' residential concentration in Italy. This is the main identification strategy used in this article to uncover the causal effect of migrant density on labour market outcomes. As part of our analysis, we also develop a methodological contribution that extends to non-linear models the procedure of Chernozhukov and Hansen (2008) to construct weak-instrument robust confidence intervals from reduced-form estimates. Moreover, the fine geographical disaggregation of our data allows identifying the same parameter using an alternative approach based on the comparison of blocks within narrowly defined groups, as in Bayer *et al.* (2008).

[1] The quota on residence permits are very tight.

Our main results show that migrants who reside in areas with a high concentration of non-Italians are less likely to be employed compared to similar migrants who reside in areas with a lower concentration of migrants. The magnitude of these effects is non-negligible: in our preferred specification, a 1 percentage point increase in the share of immigrants residing in the block reduces the probability of being employed by 2 percentage points or about 2.3% over the average.

While our data do not allow us to evaluate which particular mechanism is behind the negative effects of residential concentration on the employment prospects of migrants, we can nevertheless assess the nature of the externalities associated with large shares of migrants living nearby and rule out some explanations provided by the earlier literature. We find that residential concentration is not associated with a higher probability of finding jobs through friends. This suggests that network effects in job search may be fairly limited. At the same time we do not find support for the view that employers may redline areas with a high concentration of migrants as a form of statistical discrimination: native employment probabilities appear to be unaffected by the concentration of migrants in their block. We find that a large share of illegal, as opposed to legal, migrants in the block strengthens the negative effects on employment, which is only minimally reduced by a higher proficiency in Italian of neighbouring migrants. This may suggest that there are relevant congestion externalities in the informal labour market, the only market illegal migrants can have access to. This explanation, when confirmed by further evidence, would have relevant policy implications on the design of migration policies more than on social housing programmes. In particular, it suggests that migration policies, preventing the acquisition of legal status by those migrants who come without already having a job, are likely to increase the negative externalities associated with immigration.

The structure of the article is as follows. Section 1 discusses the theoretical foundation for a relationship between residential proximity of individuals from the same ethnic group and the probability of finding a job, and reviews the empirical evidence. Our data are described in Section 2. Section 3 outlines the identification strategy, whereas Section 4 is devoted to presenting our empirical results. Finally, Section 5 briefly characterises the normative implications of our results and concludes.

1. Theoretical Mechanisms and Existing Evidence

In this Section, we briefly survey both the theoretical and the empirical literature related to our work. The purpose of this review is twofold. First, we want to describe the theoretical mechanisms that may justify the existence of a causal link between the geographical incidence of migrants and their employment. Second, we want to document the variety of empirical strategies that have been used to address the difficult identification problems of this literature.

Partly because of the large differences in the econometric methodologies, the empirical findings are mixed and particularly scant for Europe, notably for Southern Europe, where illegal migration is pervasive.

1.1. *Theoretical Mechanisms*

In this subsection, we review the theoretical mechanisms that may explain why living in neighbourhoods with a high concentration of migrants may affect their labour market outcomes. Several theories have been proposed, pointing to either positive or negative effects. The latter are generally attributed to the fact that the neighbourhoods are seen as ghettos, spatially and socially separated from the majority society. Positive effects, on the other hand, are normally associated with the idea that migrant neighbourhoods can be launching pads, helping the newcomers to establish themselves in the majority society.

The ghettos hypothesis is long-standing in the literature. The underlying mechanism operates either on the labour supply side or via the demand of employers.

On the supply side, commuting and information frictions associated with the distance of ghettos from major centres of employment reduce the effectiveness (Ihlanfeldt, 1997; Wasmer and Zenou, 2002), the intensity (Smith and Zenou, 2003; Patacchini and Zenou, 2006) and the spatial horizon (Coulson *et al.*, 2001; Brueckner and Zenou, 2003; Gautier and Zenou, 2010) of job search. Social interactions may also be at work.[2] Ethnic minorities are over-represented among the unemployed, hence they have fewer connections to employed workers making it more difficult to access information about jobs (Hellerstein *et al.*, 2008). Also, infrequent interactions with natives reduce incentives to acquire host-country specific human capital (Chiswick, 1991; Chiswick and Miller, 1995; Lazear, 1999). Social distance and physical distance are self-reinforcing in this context, because migrants living far from business centres rely mainly on their strong ties, who are more likely to be unemployed, rather than on their weak ties, who are known to be the main source of connections to jobs (Zenou, 2013).[3] Limited access to local services, such as childcare facilities, also places individuals in migrant neighbourhoods at a disadvantage (Musterd and Andersson, 2005).

On the demand side, employers may discriminate against residentially segregated workers because of the stigma or prejudice associated with their residential location (Boccard and Zenou, 2000). This procedure, often labelled redlining, can encompass both prejudices against social or racial groups and statistical discrimination. Distant workers may also have relatively low productivity due to the long commuting, especially where the transport system is unreliable and particularly in jobs which involve long breaks during the day (such as waiter/waitress). Firms may then choose not to hire workers residing beyond a certain distance from their locations (Wilson, 1996; Zenou, 2002). Finally, employers may also discriminate against ghetto residents to satisfy the prejudices of their local customers (Borjas and Bronars, 1989).

A positive association between large shares of resident migrants and their employment is predicted by the literature on the cumulative causation of migration flows (Walker and Hannan, 1989; Massey, 1990; Massey and Espinosa, 1997; Massey *et al.*, 1998; Massey and Zenteno, 1999). This theory postulates that each act of migration creates social capital among those to whom the migrant is related, inducing new people to migrate and, thus, creating a network that can be useful in job search

[2] See the excellent literature reviews by Ioannides and Loury (2004) and Ioannides (2012).

[3] According to Granovetter (1973, 1974, 1983), weak ties are acquaintances who are not necessarily connected with one another by family or friendship links.

(Ioannides and Loury, 2004; Pellizzari, 2010). This mechanism is likely to be particularly strong within ethnic minorities, whose members often concentrate in specific jobs (Loury, 1977; Edin *et al.*, 2003; Damm, 2009; Patacchini and Zenou, 2012). In these theories, ethnic social networks mainly play the role of facilitating the transmission of information (Phelps, 1972) and, by doing so, they help newcomers to settle down in the receiving country (Bonacich and Light, 1988; Waldinger, 1996; Portes, 1998). Ethnic niches also often provide a refuge for immigrants who are discriminated against in the primary labour market (Li, 1998) and immigrant entrepreneurs may greatly benefit through reduced risk and costs of hiring members of their same groups (Bach and Portes, 1985; Bailey and Waldinger, 1991; Newman, 1999; Wang, 2004). Additionally, ethnic social networks may play a role in disseminating information about welfare eligibility, thus increasing take-up rates among migrants (Bertrand *et al.*, 2000; Pellizzari, 2013). Finally, ethnic networks shape the norms of individual co-ethnic members, potentially affecting their labour market outcomes through, for instance, peer group pressure (Granovetter, 1985), an effect which is likely to be more important when newcomers are more skilled and there is more human capital in the co-ethnic community (Borjas, 1995; Cutler and Glaeser, 1997; Edin *et al.*, 2003; Kahanec, 2006; Damm, 2009).

1.2. *Empirical Studies*

The early empirical literature on the effects of residential concentration on the employment prospects of migrants treats residential location as exogenous and documents a strong negative effect of residential segregation on labour market outcomes (Borjas, 1987; Chiswick and Miller, 2005; Kahanec, 2006). In particular, the correlation between the employment of ethnic minorities and their physical distance from major business centres, according to the spatial mismatch hypothesis has been extensively investigated (Kain, 1968; Ihlanfeldt, 1991; Ihlanfeldt and Sjoquist, 1998; Zenou, 2008).

However, residential location is obviously endogenous and any causal inference made in this literature is questionable. Self-selection and unobserved heterogeneity, rather than distance to jobs, may explain the association of lower employment and higher residential concentration among migrants and ethnic minorities more generally. Causality might actually run from employment to job access, as better labour market outcomes of workers in some neighbourhoods may attract firms into the area (Ihlanfeldt, 1991). As noted by Ihlanfeldt (1992), if the simultaneity between employment and residential location is ignored, the estimated effect of job access on employment will likely be biased towards zero.

Two main strategies have been pursued to deal with these endogeneity problems, in particular to those related to endogenous sorting into neighbourhoods, one based on observational studies and one using experimental (or quasi experimental) variation.

1.2.1. *Observational studies*

A relatively large set of observational studies address the problem of sorting by exploiting cross-metropolitan variation in the incidence of migrants and assuming that sorting across metropolitan areas is orthogonal to the outcome under consideration (Evans *et al.*, 1992; Cutler and Glaeser, 1997; Ross, 1998; Gabriel and Rosenthal, 1999; Bertrand *et al.*, 2000;

Weinberg, 2000, 2004; Card and Rothstein, 2007; Ross and Zenou, 2008). The common finding of these studies is a negative employment effect of residential concentration.

Another approach consists in analysing young workers residing with their parents, who are assumed to have chosen their place of residence for their children (Borjas, 1995; Raphael, 1998). These studies also find a negative link between the incidence of migrants in one's location of residence and their employment. However, if parents and children share similar unobservable traits and/or parents decide where to reside considering the employment prospects of their children, the youth approach becomes invalid.

A different identification strategy is based on instrumental variables and uses lagged immigrant density to instrument its current level. The key identification assumption in this approach is the orthogonality between the factors that influenced immigrants' settlements in the past and in the present, apart from their effect through the current presence of immigrants. Such a strategy has been extensively used in the US literature (Walker and Hannan, 1989; Altonji and Card, 1991; Conley and Topa, 1999; Massey and Zenteno, 1999; Falcon and Melendez, 2001; Mouw, 2002; Munshi, 2003; Falcon, 2007). Patacchini and Zenou (2012) used this identification strategy on data for the UK they show a positive employment effect of ethnic population density. This is one of the very few studies in Europe. Some evidence on the role of ethnic networks in finding a job can be found in Frijters *et al.* (2005) and Battu *et al.* (2011), although these studies are not explicitly focused on residential concentration.

Bayer *et al.* (2008) use an alternative approach. They draw on data from the US Census, disaggregated at the level of the city block and city blocks are grouped into small sets of adjacent areas. This enables them to condition on block-group fixed effects in their regression analysis to isolate block-level variation in neighbour attributes. Their identifying (untestable) assumption is the absence of correlation in unobservables across blocks within block groups. They find evidence of significant social interactions operating at the block level: residing in the same *versus* nearby blocks increases the probability of working for the same employer by over 33%. Their results also indicate that this referral effect is stronger when individuals are similar in socio-demographic characteristics and when at least one individual is well attached to the labour market.

1.2.2. *Experimental (and quasi-experimental) studies*

Edin *et al.* (2003) and Damm (2009) make use of natural experiments in Sweden and Denmark, respectively. In both countries the residential choices of migrants were limited by governmental policies that either explicitly randomised (conditional on a set of observables) or were arguably exogenous to labour market conditions.

Edin *et al.* (2003) document immigrant earning gains of about 13% following a standard deviation increase in local ethnic group size. Damm (2009) finds a similar effect for earnings (about 18%), while the effect on employment is negative for high-educated individuals and virtually zero for the low educated.[4]

While the use of random assignment across municipalities is attractive, these studies are not without their shortcomings. These include small sample sizes, the large margin

[4] Recently, Beckers and Borghans (2011) exploit a similar natural experiment in the Netherlands and find similar, although stronger, results.

of error in the definition of the treated population and the use of municipalities as the geographical unit of interest.[5]

In the US, some papers were inspired by the two major programmes of residential mobility: the Gautreaux programme, implemented in Chicago (1976–90), and the Moving to Opportunity programme (MTO), implemented in five major cities (Baltimore, Boston, Chicago, Los Angeles and New York) between 1994 and 1999.[6] Assessing the employment effects of the Gautreaux programme, Harris and Rosenbaum (2001) find higher employment but no difference in wages or hours worked for those who moved to the suburbs compared with those who moved to the central city. Mendenhall *et al.* (2006) study the effect of the programme on low-income black females and find no difference between movers to suburbs and movers to the central city. Katz *et al.* (2001) find no effect of MTO on either employment or earnings.

Finally, Holzer *et al.* (2003) uses exogenous variation in job access generated by the unanticipated opening of a new transit line to control for sorting across neighbourhoods and he finds that employment effects are positive and greatest for those residing nearest to the origin of the new transit road.

2. Data and Descriptive Evidence

2.1. *Data*

Our analysis is based on data from a new survey of immigrants, which was carried out between October and November 2009 in eight cities in Northern Italy: Alessandria, Brescia, Bologna, Lucca, Milano, Prato, Rimini and Verona. The cities were chosen non-randomly to represent agglomerations of different sizes (large, medium-sized and small) while at the same time guaranteeing a good degree of representativeness of the entire population of the North of Italy, where more than 60% of the non-Italian residents are located.

Table A1 in the Appendix reports some key characteristics of these cities, comparing them to the averages in the country and showing that they offer a good representation of the population of the North of Italy.

The sampling procedure of our survey was designed with the intention of reaching particularly hard-to-trace segments of the population, namely immigrants, both legal and illegal. Migrants are grouped into three macro regions of origin and the survey guarantees representative results only within these three subpopulations: European new member states (NMS),[7] Western Balkan countries (WBC)[8] and all other countries of origin.[9]

The sampling strategy consists of three main steps: in the first stage, we sample neighbourhoods separately in each of the eight cities and then, in the second stage, we

[5] Conley and Topa (1999) and Bayer *et al.* (2008) show that the relevant neighbours are those in the close vicinity.

[6] The Gautreaux programme targeted black families residing in poor neighbourhoods and handed them rental vouchers to move to predominantly white or racially mixed areas. The MTO programme was inspired by the Gautreaux programme but the target was inner-city low income families with children living in public housing.

[7] Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia.

[8] Albania, Bosnia, Croatia, Macedonia, Kosovo, Montenegro and Serbia.

[9] The focus on EU new member states and the Western Balkan countries was imposed by the European Bank for Reconstruction and Development (EBRD), the sponsor of the study.

select one block of buildings in each of the sampled neighbourhoods where, in the final stage, the individuals to be interviewed are randomly chosen.

The neighbourhoods are selected with sampling probabilities that are proportional to the share of legal migrants resident in the area, as measured by the official population registers. Then, a purposely designed algorithm randomly selects one point on the official map of each sampled neighbourhood and the blocks that are closest to such points are included in the survey.[10] In order to increase sample size additional blocks are selected based on a proximity criterion. Namely, we also include in the survey blocks that are adjacent to one (or more) of the randomly selected blocks where the share of dwellings occupied by immigrant households is higher than a fixed threshold.[11]

In each selected block, a census of residential units is carried out on the basis of a combination of conversations with the buildings' janitors and short door-to-door visits. The census provides a list of apartments for each of the four groups (NMS, WBS, other non-Italians and Italians). It is used to randomly select four households for each of the above population groups. One adult (older than 18 years old) in each household is randomly chosen for the interview. Hence, a maximum of 16 persons are eventually interviewed in each block. However, in most blocks there are fewer than 16 interviews because there were fewer than four persons in some of the population groups.[12]

Table 1 summarises the sampling procedure. Each city is divided into three districts: central, mid-central and peripheral. The first three columns of the Table indicate for each city and district the number of sampled neighbourhoods, which, ignoring the blocks selected with the proximity criterion, is equivalent to the number of blocks. The fourth column simply sums over the first three and reports the total number of sampled neighbourhoods. The average number of interviews/observations per neighbourhood is shown in column 5. In columns 4 and 5, we also show in parentheses the total number of neighbourhoods in the city (column 4) and the average population in the neighbourhoods (column 5), so as to give an indication of the coverage of our sample.[13]

The census of the residential units in each block is a particularly precious source of information. Official population registers from the city councils only consider legal immigrants, whereas our census includes both legal and illegal residents, living, either permanently and temporarily, in the considered blocks.

Although the survey includes both migrants and natives, for this study we only consider the subsample of migrants.[14] Interviewees are asked questions on individual and family characteristics, reasons behind migration, living and work conditions,

[10] The website http://v.controul.com/app/ shows exactly which blocks were chosen in each neighbourhood. Blocks are defined as portions of urban surface that are built-up and continuous, that is not interrupted by areas for traffic circulation or allocated for public use (e.g. parks).

[11] Since the randomly selected blocks that satisfy the threshold criterion are usually adjacent to several other blocks, only the one adjacent block with the highest incidence of immigrants is selected.

[12] Non-response bias is very low: interviewers are asked to visit the selected households several times and at different times of the day. In case the selected individual refuses to answer or is unreachable, a replacement unit is drawn from a reserve list.

[13] Note that sampled blocks are much smaller than sampled neighbourhoods and have, on average, a population of 208 residential units.

[14] Some information on natives will be used in Table 10.

Table 1
Sampled Neighbourhoods by City and District

	Central (1)	Mid-central (2)	Peripheral (3)	Total* (4)	Observations (mean)† (5)
Alessandria	2	3	1	6 (23)	5.1 (4,073)
Bologna	2	5	7	14 (90)	6.5 (4,135)
Brescia	2	3	0	5 (30)	4.8 (6,482)
Lucca	2	2	6	10 (79)	4.3 (1,093)
Milano	4	8	19	31 (87)	6.3 (14,879)
Prato	0	2	4	6 (35)	2.8 (5,334)
Rimini	2	3	1	6 (57)	6.7 (2,455)
Verona	0	4	5	9 (23)	3.8 (11,528)
Total	14	30	43	87 (424)	5.5 (6,247)

Notes. *Total number of neighbourhoods in the city in parentheses. †Average number of residents per neighbourhood (from city registers) in parentheses.

cultural integration and compliance with immigration laws. Especially for the questions about legal status, the interviewers were very carefully instructed to insist on the fact that the survey was carried out exclusively for research purposes, that the data would remain fully anonymous and that none of the institutions involved in the organisation of the survey were in any way connected with the immigration authorities, the fiscal administration, the police or the Ministry of Internal Affairs (which is the institution that issues work and residence permits).

We code as illegal migrants those who declare they do not have a permit of stay or refuse to answer the question on legal status and those who declare they do not have access to the Italian health system or they do not have the required documents to go back to their home country (definition 1). In all cities, undocumented migrants represent a sizeable proportion of total migration: from 12% in Bologna to over 29% in Brescia. Since around 6% of such individuals are from EU countries of recent access (e.g. Bulgaria and Romania) and can get the Italian permit of stay with fewer restrictions, we also consider a more restrictive definition (definition 2) that replicates the first one but excludes all immigrants from new member states from the pool of illegal.[15]

2.2. *Descriptive Statistics*

Given the peculiar sampling structure of our study, in Table 2 we compare our data with other surveys that might be used to conduct studies of migration, namely the official labour force surveys (LFS) and a survey of migrants carried out by the institute Iniziative e Studi sulla Multietnicità (ISMU), which is relatively popular in Italy (Dustmann *et al.*, 2010).

While the LFS data only capture legal migrants, being sampled from the population registers, the ISMU survey also includes illegal migrants but its sampling frame is radically different from ours (Cesareo and Blangiardo, 2009). In particular, the ISMU survey was carried out between October 2008 and February 2009 in 32 cities all over

[15] We also consider two alternative definitions that only use information on permits of stay and results are almost identical.

Table 2

Comparison with Other Data Sources

Variable	fRDB-EBRD* (1)	LFS† (2)	ISMU‡ (3)
Share of migrants	0.75	0.07	1.00
Share of migrants from new member states	0.25	0.17	0.13
Share of migrants from Western Balkans	0.25	0.19	0.17
1 = illegal migrant (def. 1)	0.20	0.00	0.11
1 = female migrants	0.47	0.51	0.51
1 = primary education	0.38	0.46	0.30
1 = secondary education	0.48	0.39	0.45
1 = tertiary education	0.10	0.10	0.21
1 = employed	0.85	0.47	0.68

Notes. *These statistics refer to the whole sample (1,137 observations), not just to the sample used for the empirical results. †The LFS data, being sampled from the population registers, only capture legal migrants. Moreover, it is not representative at the level of the single municipality. For these calculations, the sample has been restricted to the regions of the North of Italy. ‡The ISMU data include both regular and irregular immigrants. It is based on 12,000 interviews conducted between October 2008 and February 2009 at popular social venues for migrants, such as language schools, assistance centres, etc.

Italy. Immigrants were interviewed in places where they usually meet or go to seek assistance, such as language schools, immigrant assistance centres and trade unions. The advantage of this sampling method is that it makes it much easier to reach illegal immigrants, thus allowing for larger sample sizes.[16] However, such an advantage comes at the cost of representativeness, as migrants who are likely to be found in the places covered by the ISMU survey might be very different from the rest.

By construction, migrants are over-represented in our data compared to the LFS, both overall and for each of the subgroups that we consider (NMS, WBS and others), which are equally represented (by construction).[17] Also, we find slightly more illegal migrants compared to ISMU, although the difference is minor. Female migrants are under-represented in our data compared to both the LFS and ISMU, while the education distribution is remarkably similar. Our interviewees are also more likely to be in employment, a result that is due for the most part to the presence of illegal residents, who are necessarily employed in the shadow sector.

We now focus on our data and present in Table 3 a description of the main variables used in our empirical exercise of Section 4.

On average, migrants are younger than natives, with an average age of about 37 years old, which compares to about 43 for Italians. Moreover, the incidence of females is much lower than among natives (46% against 52%). Immigrants into Italy do not appear to be a particularly low-skilled group; more than half of them have at least a degree of secondary education. About 20% of our surveyed immigrants are illegal, according to our preferred definition (definition 1). In terms of labour market performance, roughly 88% are employed, which compares to a much lower employment rate for natives (about 50% in Northern Italy. See Table A1). Almost 60% of migrants obtained their jobs through friends.

[16] The ISMU survey consists of 12,000 interviews to both legal and illegal immigrants.
[17] The ISMU survey covers only immigrants.

Table 3
Descriptive Statistics

Variable	Mean (1)	SD (2)	Observations (3)
Socio-demographic characteristics			
Age	37.40	8.47	478
1 = female	0.46	–	478
years living in Italy	9.87	7.30	478
1 = at least secondary education	0.57	–	478
1 = illegal immigrant	0.19	–	478
Labour market outcomes			
1 = employed	0.88	–	478
1 = found work through friends	0.58	–	406
Residential population (at the block level)			
% of non-Italians	16.65	10.20	478
% of immigrants from same origin	5.94	5.58	478
% of illegal immigrants*	3.17	4.11	478
% of legal immigrants†	13.49	8.97	478
Block characteristics			
Time to city centre (min.)‡	28.21	13.02	467
House price§	2,396	658	468
% commercial buildings¶	0.06	0.10	478

Notes. *Computed as the share of immigrants in the block times the share of illegal immigrants interviewed in the block. †Computed as the share of immigrants in the block times the share of legal immigrants interviewed in the block. ‡*Source.* local transport authorities. §Euro per square metre. *Source.* Agenzia del Territorio. ¶Number of commercial buildings over total number of buildings in the census tract.

We measure migrant population density by the percentage of non-Italian households living in the considered blocks. On average there are about 17% of non-Italian households in the surveyed blocks, with a standard deviation of more than 10 percentage points. We can also define immigrant density more restrictively as the percentage of households from one's same area of origin in the block. The mean of this variable in our sample is just below 6%, with a standard deviation of 6 percentage points.

Moreover, Table 3 reports summary statistics for the estimated share of legal and illegal immigrants in the block. We construct these proxies by multiplying the share on non-Italians in the block, from our census, by the share of illegal and legal immigrants actually interviewed in the block.[18]

Finally, the bottom panel of the Table reports some summary statistics at the block level. We obtain house prices per square metre at the neighbourhood level from the Agenzia del Territorio, a government agency that records housing transactions and complements them with surveys of real estate agents to construct indexes of housing costs. Time-to-travel to the city centre is computed by combining information on the household address and the centre of the city, which given the strong historical heritage of all the eight cities in our survey (as most cities in Italy) is very easy to identify.[19] We

[18] Definition 1 of illegal migrants is used for this calculation.
[19] The historical centres of the ancient Roman or medieval cities still remain today the most important commercial areas in the majority of Italian cities and certainly in those that are covered in our survey.

then use the online websites of the local transport authorities to compute the time (in minutes) necessary to travel to the centre by public transport.

Finally, we include some variables obtained by merging our survey with the auxiliary database of census tracts from the 2001 official census of the Italian population. Beside aggregate population variables, the database also contains a large set of descriptive characteristics of the buildings in the tract and it is the source used to construct our instrument in Section 3.[20] In Table 3, we show the share of commercial buildings in the block as a proxy for the presence of jobs in the neighbourhood.[21]

In order to get a first glance at the pattern of immigrant density in our data, Table 4 reports a selected set of statistics separately for immigrants living in areas characterised by high and low-densities of migrants, defined as blocks where the percentage of non-Italians lies in the top and bottom 25% of the observed distribution.[22] Columns 1 and 2 report the mean and the standard deviation of some immigrant characteristics in high and low-density neighbourhoods respectively. Columns 3 and 4 show the difference between the first two columns, unconditional and conditional on city and district dummies respectively.[23] Interestingly, the differences are minimal. The only few statistically significant conditional differences show that immigrants residing in areas with higher migrant population density arrived in Italy more recently and are slightly older. We do not find evidence that more educated immigrants sort into less segregated areas.

This pattern is not related to the availability of subsidised rents, as migrants typically have access to social housing after 15–20 years on a waiting list. Moreover, all major programmes of social housing were discontinued in Italy in 1993, just before the mass immigration waves. The compulsory employer contribution to the social housing fund (Gescal) was indeed redirected to the financing of pensions. In 1998, competencies over social housing were transferred to regional authorities that were encouraged to sell off their housing stock. Most of the sales (totalling some 150,000 housing units) were carried out while the stock of migrants was increasing at an average yearly rate of some 25%. The total stock of social housing available for subsidised rent is currently of about 350,000 units in the North of Italy, where most of the migrants are located. The ratio of the rental stock of social housing to the population in Italy is about 1.2% compared with 7% in France. The total stock of subsidised housing (including also private, rent-regulated, housing) in Italy is less than 5% of the housing stock compared with 20% in France and the UK. Nowadays inflows are very limited as local authorities are privatising the stock as soon as there are free units, that can be sold at a higher price. Thus, the waiting time to get social housing is between 10 and 15 years for natives. Moreover the municipalities with a stronger presence of migrants require a relatively long minimum legal residence period before applications to social housing can be made. Although migrant families have income and asset conditions qualifying

[20] The link with our survey is based on the actual addresses of the residential units occupied by the individuals in our sample

[21] Unfortunately, the data do not include commercial square metres in the block.

[22] According to the distribution of immigrants in the considered blocks, the threshold level for the high-density neighbourhoods (top 25% of the distribution) is 25.5% of foreign households and that for low-density (bottom 25%) is 7.5% of foreign households.

[23] Estimates are produced by OLS.

Table 4

Immigrants by Immigrant Density in the Block

Variable	High immigrant density† (1)	Low immigrant density† (2)	Difference: Unconditional (3)	Difference: Conditional (4)
1 = female	0.474	0.474	0.000	0.109
	(0.046)	(0.046)	(0.063)	(0.094)
Age	39.198	36.707	2.491*	−0.214
	(0.820)	(0.798)	(1.265)	(1.684)
Years in Italy	8.310	11.069	−2.759***	−3.338**
	(0.468)	(0.727)	(1.047)	(1.509)
1 = at least secondary education	0.647	0.569	0.078	0.122
	(0.044)	(0.046)	(0.072)	(0.084)
1 = illegal migrant	0.198	0.233	−0.034	−0.236***
	(0.037)	(0.039)	(0.071)	(0.055)
1 = employed	0.914	0.810	0.103**	0.081
	(0.026)	(0.036)	(0.050)	(0.060)
1 = found work through friends	0.606	0.570	0.036	0.132
	(0.048)	(0.051)	(0.083)	(0.096)

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Estimates are produced by OLS (robust standard errors, clustered by census tract, in parentheses). †High and low-immigrant density blocks are those where the percentage of non-Italians lies in the top and bottom 25% of the observed distribution respectively. The first two columns report the means (SEs in parentheses) of the indicated variable in the two samples. The last two columns report the unconditional or conditional (on city and district dummies) difference.

them for priority in the queue, their waiting period is therefore longer than that of natives as they have to wait first for the regularisation and then for the completion of the minimum legal residence period.

One of the most interesting features of our data is the possibility of identifying illegal immigrants. Throughout our analysis we use many alternative measures of immigrant density, including the distinction between the share of illegal or legal immigrants in the block (see subsection 4.3). Table 5 provides some precious information on how legal and illegal immigrants differ. This will be useful in interpreting our results. Each cell reports the unconditional or conditional (on city and district dummies) difference between the means of the variable indicated in the first column of the Table across the samples of legal and illegal immigrants. All estimates are produced by OLS.

Compared to the legally resident, illegal migrants appear to be, on average, men, younger, slightly less educated. Moreover they are less likely to be employed and more likely to rely on informal networks to find a job. Especially when we restrict attention to the first definition, illegal immigrants also appear to be more recent migrants. Finally, they are less proficient in the Italian language. While subjective assessment of language proficiency is usually biased, our data contain an objective measure of the linguistic abilities of migrants, as a formal test of the knowledge of the Italian language was administered at the end of the personal interviews.[24]

[24] The test was optional and approximately 14% of the individuals in the sample refused to take it. A small amount of 5 euro was given to individuals taking the test. The test included question on language comprehension, of growing complexity. Final scores are standardised to have average of 500 and standard deviation of 100.

Table 5

Legal Versus Illegal Differences in Observable Characteristics

	Definition 1†		Definition 2‡	
Variable	Unconditional (1)	Conditional (2)	Unconditional (3)	Conditional (4)
1 = female	−0.139**	−0.118*	−0.256***	−0.236***
	(0.058)	(0.060)	(0.054)	(0.055)
Age	−2.705**	−2.735**	−2.894**	−2.757**
	(1.098)	(1.060)	(1.319)	(1.291)
Years in Italy	−1.194	−0.792	−0.687	−0.116
	(0.894)	(0.896)	(1.072)	(1.081)
1 = at least secondary education	−0.099*	−0.093*	−0.217***	−0.198***
	(0.060)	(0.054)	(0.068)	(0.065)
Knowledge of Italian§	−52.898***	−47.578***	−65.856***	−56.381***
	(12.810)	(11.270)	(14.408)	(13.252)
1 = found work through friends	0.177***	0.166***	0.225***	0.214***
	(0.060)	(0.062)	(0.067)	(0.069)
1 = employed	−0.125***	−0.118**	−0.164***	−0.161***
	(0.045)	(0.049)	(0.051)	(0.054)
Percentage of non-Italians in their block	−0.241	−1.074	0.325	−0.106
	(1.523)	(1.130)	(1.822)	(1.362)

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Robust standard errors, clustered by census tract, in parentheses. †Respondents are coded as illegal if either
(*i*) they do not have a permit of stay or do not answer the question, or
(*ii*) they declare they do not have access to the Italian health system or
(*iii*) they declare they do not have the documents required to go back to their country more often.
‡The same as definition 1 but excluding all EU-27 citizens. §Score in the Italian test, standardised to have mean 500 and SD 100. Each cell reports the unconditional or conditional (on city and district dummies) difference between the means of the variable indicated in the first column across the samples of illegal and legal immigrants.

3. Empirical Model and Estimation Strategy

Our empirical analysis is primarily aimed at estimating the causal effect of the percentage of migrants in one's residential block on the employment status of migrants.[25]

Our empirical model is based on the following main equation:

$$y_{icdb} = \alpha_1 m_{cdb} + \boldsymbol{\alpha}_2 \boldsymbol{X}_{icdb} + \epsilon_{icdb}, \tag{1}$$

where y_{icdb} is an indicator of employment for migrant i in city c residing in district d and block b; m_{cdb} is the percentage of all non-Italians residing in block b of district d and city c; $\boldsymbol{X}_{icdb}$ is a set of observable individual characteristics, including district (central, mid-central, peripheral) and city fixed effects, and ϵ_{icdb} is the error term. The sample is restricted to migrants only.

The parameter of main interest in (1) is α_1, whose identification is possibly impeded by the presence of unobservable factors that influence both the location decisions of migrants and their labour market outcomes. For example, one might be worried that

[25] Unfortunately, the poor information on wages contained in our data prevents us from analysing the effect of wages. Indeed, the number of missing values is very high and for the valid observations wages are recorded in relatively wide intervals.

residentially segregated migrants are negatively selected, as only the very high ability can afford to live in native-dominated neighbourhoods and high-ability workers also experience better labour market outcomes, regardless of where they live. Such a mechanism would bias α_1 downwards in standard OLS. Additionally, there might also be unobservable factors at the block level that affect both the migrant's probability of locating in the block as well as labour market success, such as the availability of some public services (employment services, public transport). Finally, our regressor of interest, being based on conversations with buildings' janitors and door-to-door conversations, is likely to be affected by measurement error. Although it is difficult to assess the exact extent of mismeasurement, the assumption of classical measurement error seems quite plausible in our setting, so that the resulting bias should draw the estimated parameter closer to zero.

Overall, it is hard to establish whether the total bias in simple OLS (or probit) estimates of (1) would be positive or negative.

We address the two issues of measurement error and omitted variable bias differently. For measurement error, we collected additional auxiliary information about the implementation of the survey, namely individual characteristics of the interviewers and their evaluations of the overall quality of each single interview. Assuming that the measurement error is a linear function of such variables, it is possible to rewrite an augmented version of (1) which includes interviewer' characteristics as additional explanatory variables to eliminate the bias due to measurement error in m_{cdb}.[26]

The bias from omitted variables is the key identification issue in this literature and it has been addressed in many different ways by previous studies, as we discussed in subsection 1.2. Our identification strategy rests on the use of an instrumental variable that has never been previously proposed. Moreover, in subsection 4.2, we replicate our results using an alternative approach that mimics closely the prominent study by Bayer *et al.* (2008), which compares adjacent blocks within small groups of buildings. Given the particular sampling structure of our data, only a small subsample of our survey can be used for this purpose, so that the first approach, the instrumental variable strategy, is more powerful in our setting (see Section 1 for details).

Specifically, we use the building structure of the block 10 years before the survey to instrument the percentage of migrants currently residing in the area. Using the actual addresses of the residential units of the individuals in our sample, we have linked our data to an ancillary database of the 2001 Italian population census. Such an ancillary database contains a large set of descriptive characteristics of each single city block in Italy, including the total number of buildings and the total amount of square metres (i.e. the sum of the square metres of each floor in each building) in the block, broken down by residential and commercial space. We use these data to calculate the ratio of

[26] In order to clarify our approach to measurement error, assume m_{cdb} is the true variable and $m^*_{cdb} = m_{cdb} + u_{cdb}$ is its mismeasured analogue with u_{cdb} being the error of measurement. Further assume that u_{cdb} is the sum of a linear function of observables (in our specification, the individual characteristics of the interviewers and their evaluations of the overall quality of each single interview) and a purely random term: $u_{cdb} = \gamma Z_{cdb} + v_{cdb}$. Then, replacing and rearranging terms in (1) yields the following augmented version of the model $y_{icdb} = \alpha_1 m_{cdb} + \boldsymbol{\alpha}_2 \boldsymbol{X}_{icdb} - (\alpha_1 \gamma) Z_{cdb} - \alpha_1 v_{cdb} + \epsilon_{icdb}$ which allows reducing the bias from measurement error.

residential square metres per residential building in the block, a variable that takes high values in areas that are dominated by large residential buildings (lots of residential square metres for few buildings) and low values in areas of detached or semi-detached houses.[27]

The idea of this instrument builds on the literature on housing discrimination, which documents how migrants and other minorities find accommodation more difficultly than natives, both on the renting and the property markets (Yinger, 1986; Page, 1995; Ondrich *et al.*, 1999; Ahmed and Hammarstedt, 2008; Bosch *et al.*, 2010; Baldini and Federici, 2011; Hanson and Hawley, 2011). The relevance of our instrument rests on the presumption that such type of discrimination is taste-based and that natives, who predominantly populate the supply side of the housing market, dislike close interactions with migrants. As a consequence, they are less willing to rent or sell their properties to migrants, especially so where the urban structure is conductive of close interactions among residents, such as in neighbourhoods where residential space is concentrated in a small number of buildings.

The literature on housing discrimination provides numerous pieces of evidence in support of our instrumental variable strategy. Firstly, all papers find a sizeable degree of discrimination against migrants, both in the US and in Europe (Page, 1995; Ahmed and Hammarstedt, 2008; Bosch *et al.*, 2010; Baldini and Federici, 2011; Hanson and Hawley, 2011). Second, discrimination persists even when additional information about the potential renter/buyer is available (Ondrich *et al.*, 1999; Ahmed *et al.*, 2010; Bosch *et al.*, 2010). This is an important finding because it allows documenting the extent to which discrimination is driven by each of the two most commonly cited sources, namely statistical and taste-based discrimination. In our specific application, statistical discrimination arises if landlords, who are primarily natives, prefer not to rent their apartments to a migrant because they know that on average they are poorer than natives and thus less likely to pay their rents regularly. Without specific information about the individual applicant, landlords base their decisions on the average characteristics of migrants. Hence, statistical discrimination should disappear or at least be limited when information about the individual candidate is provided. Ahmed *et al.* (2010); Bosch *et al.* (2010) show that providing such information does not eliminate discrimination. As a consequence, it must be that it arises, at least partly, by the fact that natives simply dislike interacting with migrants, a mechanism that is commonly labelled as taste-based discrimination.

Discrimination in the Italian housing market has been recently documented by Baldini and Federici (2011), who selected a large sample of renting advertisements for housing units throughout Italy that were posted on the Internet and sent fictitious email requests to visit such units. The only distinctive feature of the email messages was the name of the perspective tenant, which could be either typical Italian or typical of Arab or Eastern European origin. Emails were sent to advertisers according to a random algorithm, so as to guarantee orthogonality of the characteristics of the fictitious perspective tenant and those of the apartments, a strategy that is common to

[27] As far as we know, Bauer *et al.* (2011) is the only other paper that instruments migration at the neighbourhood level with some physical characteristics of the local buildings, although their specific instrument is different from ours and the context is also different.

other studies in this field (Carpusor and Loges, 2006; Ahmed *et al.*, 2010). The study then records responses to the email contacts and investigates whether the probability of a positive feedback varies with the ethnicity of the fictitious names.[28] The results show clear evidence of housing discrimination in the Italian market, especially in Northern Italy, which is where our sampled cities are located.

We have been kindly given access to the data of Baldini and Federici (2011) and have merged them with our instrumental variable at the level of the city to produce supporting evidence for our identification strategy.[29] We then run an OLS regression with the ratio of the average rate of positive response for migrants and natives as a dependent variable and the total amount of residential square metres over the total number of residential buildings in the city as a main regressor of interest. Additionally, we include the city average of all the controls used by Baldini and Federici (2011) as control variables, namely dummies for the week and the weekday when the email was sent, the log of the property size in square metres, the monthly rent per square metre, dummies for whether the advertisement was posted by an agency, whether the advertisement included pictures, whether the email included additional information about the perspective tenant (family composition, occupation) and whether the email included orthographic or grammar errors. There are 41 cities in the database and the regression weights them by the number of observations in the original microdata.

Figure 1 shows the partitioned regression equivalent of the above model, namely the variables on the axes are the residuals of regressions of the dependent variable (on the vertical axis) and the main regressor of interest (on the horizontal axis) on the control set. The graph indicates the existence of housing discrimination (as measured by a lower recall rate for possible renters with non-Italian names) in urban structures dominated by large residential buildings (as opposed to those populated by detached or semi-detached houses). Indeed, the results show that residential building structure is strongly and significantly correlated with discrimination against migrants in the housing market and it explains a sizeable 20% of the variation in relative response rates across cities, thus providing strong support to the logic behind our instrumental variable strategy.[30] Additional evidence of the relevance of the instrument is in the first-stage results that will be reported later in Table 7.

Another factor which plays in favour of the relevance of our instrument is connected to the characteristics of immigration to Italy. As stressed above, most immigrants come in illegally, and residence in buildings with a relatively low density may reduce the probability that their illegal status is detected. Although not statistically significant, illegal status is negatively correlated with the ratio of total residential square metres to the number of residential buildings in the block (see column 7 of Table 6).

Contrary to the relevance of the instrument, its exogeneity cannot be tested. However, a direct effect of the structure of the buildings in the neighbourhood on

[28] Contacts leading to an immediate appointment for a visit or ask additional information are classified as positive responses.

[29] Unfortunately, Baldini and Federici (2011) did not record the address or the neighbourhood of the advertised apartments and the city is the only geographical identifier that can be used for our purposes.

[30] We checked that our results are not driven by few outliers by using robust regression methods. When estimating the slope parameter in Figure 1 using OLS weighted by the inverse of the Cook's distance, our coefficient is still negative and significant ($b = -0.179$, SE = 0.060).

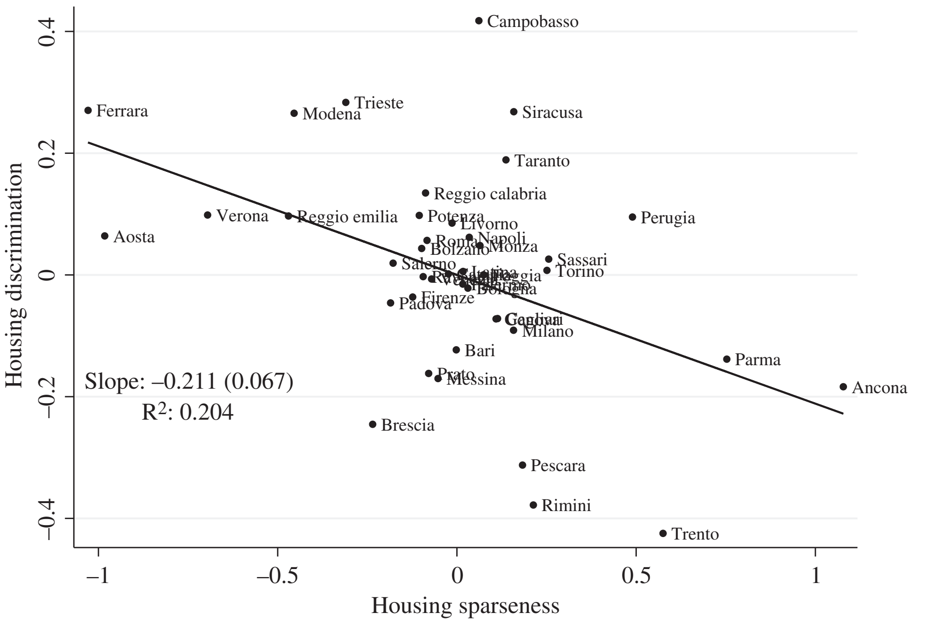

Fig. 1. *Housing Discrimination and Residential Building Structure*

employment is hard to imagine. Alternatively, the exogeneity of our instrument might be questioned on the basis of an indirect link with some neighbourhood characteristics that are omitted from (1).

Of course, we cannot exclude this possibility *a priori* but we can provide evidence that the instrument is orthogonal to some of the most obvious suspects, such as house prices or distance from the city centre. Columns 1–6 of Table 6 report the results of a battery of OLS regressions at the level of the census tracts with our instrument as the dependent variable and some block characteristics that could influence the employment opportunities of local residents as explanatory variables. Specifically, we consider average house prices in the block, time to travel to the city centre by public transport, the share of commercial buildings in the block and the share of high-skilled population in the block.[31]

The estimates in Table 6 confirm the intuition that our building structure indicator does not correlate significantly with these observable neighbourhood characteristics that might affect employment. This result holds across a number of specifications, either conditioning or non-conditioning on city and district dummies (columns 1 and 2 respectively) and including the explanatory variables all together (columns 1 and 2) or one by one (columns 3, 4, 5 and 6).[32]

Column 7 of Table 6 investigates whether the instrument is related to the characteristics of the immigrants living in the block. It reports results of a regression of

[31] Skilled people are defined as those having completed at least secondary schooling.
[32] We have also investigated the correlation of our instrument with the percentage of caretakers in the block (i.e. the share of people answering in our survey that they are occupied as: caretakers, domestic workers, housekeeper, babysitters or cleaners) finding no significant association.

Table 6

Residential Building Structure and Other Block Characteristics

	Dep. variable: building structure†						
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
House prices‡	−0.000	−0.000	−0.000				
	(0.000)	(0.000)	(0.000)				
Travel time to city centre§	−0.000	0.001		−0.002			
	(0.006)	(0.009)		(0.008)			
Percentage of commercial buildings¶	−0.040	−0.034			−0.027		
	(0.071)	(0.072)			(0.072)		
Percentage skilled people††	0.410	0.776				0.519	
	(0.620)	(0.751)				(0.616)	
Characteristics of the immigrants (from frdb survey)							
Age							0.007
							(0.032)
1 = female imm							−0.008
							(0.203)
1 = skilled imm							0.038
							(0.091)
1 = illegal							−0.141
							(0.111)
District fixed effect	no	yes	yes	yes	yes	yes	yes
Observations	168	168	168	168	168	168	478
F stat. joint sign	0.21	0.35	–	–	–	–	0.86‡‡

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Robust standard errors in parentheses. †Ratio of total residential square metres to number of residential buildings in the block, normalised at the city level. ‡*Source.* Agenzia del Territorio. §Time to travel is measured in minutes by public transport and it is computed from the websites of the local transportation authorities. ¶ Ratio of commercial buildings over the total number of buildings in the block, normalised at the city level. ††Share of population in the block (both immigrants and natives) with at least secondary education. ‡‡This refers to all controls (including years since migration, country of origin (grouped) and age squared). All controls of (1) in column (7).

the instrument on all exogenous (observable) characteristics, controlling for district and city fixed effects. None of the correlations is significant. It is particularly interesting to note that the percentage of skilled immigrants in the block appears to be unrelated to the instrument.

Obviously, the evidence in Table 6 is by no means a formal test of exogeneity. Nevertheless, the lack of correlation of the instrument with some relevant observable neighbourhood and individual characteristics is suggestive that it is likely orthogonal to other unobservables of the same nature.

In principle, we could have included the neighbourhood characteristics considered in Table 6 in the control set of our main model but we prefer to exclude them as some of them may induce further endogeneity.[33] Note also that exogeneity is further guaranteed by the lagged nature of the instrument, that is measured 10 years prior to the survey generating our main data.

Our identification strategy departs significantly from the popular approach of using lagged values of the immigration-related variable in the different areas to instrument

[33] Including them, however, does not change our main findings.

its current values (Altonji and Card, 1991). The validity of such an approach rests on very specific assumptions about the relative degree of serial correlation in the error term of the main model and in the process generating the endogenous variable (Angrist and Krueger, 2001). These assumptions are very rarely spelt out and discussed and we believe that they would be hard to justify in our setting.

4. Empirical Results

Our main results are reported in Table 7, which shows probit estimates of model (1) where the dependent variable is a dummy indicator of employment and the percentage of immigrants in the block is the main regressor of interest. The basic set of controls includes a linear function of age, a gender dummy, a dummy for education at or above secondary level, a dummy for illegal status, dummies for origin (new member states, Western Balkans, other origins), dummies for years since migration in Italy (less than 5 years, 5–10 years, 15–20 years and more than 20 years) and city and district dummies.

Consistently with the dichotomous nature of the outcome variable, we adopt a probit model, although our discussion in Section 3 was framed in a linear setting in order to emphasise the fact that we do not exploit the non-linearity of the probit model for identification purposes.[34]

The first column of Table 7 reports the estimates of a simple specification of our model that only includes the basic controls and does not take into account neither the potential endogeneity of the main regressor of interest nor the bias due to measurement error. The estimated coefficient is negative, but very small and imprecise.

In column 2, the control set is augmented with several variables to control, at least partly, for mismeasurement, namely dummies for interviewers of Italian and of Albanian nationality (the two most common groups), a dummy for graduate interviewers and one for professionals,[35] a dummy for whether the interviewer and the interviewee are of the same gender and the self-reported evaluation of the level of understanding of the questions by the interviewee, ranging 0–10. The estimated coefficient is now substantially larger (-0.005 as opposed to -0.001) but still far from conventional levels of statistical significance.

Column 3 reports our preferred specification and, in addition to the interviewer's characteristics, it also instruments the percentage of non-Italians in the block with our indicator of residential building structure. The model is estimated by full-information maximum likelihood, thus producing jointly the estimates of the first stage and the main equation. The standard errors are clustered at the level of the census tract, which is the exact level of variation of the instrument and the same clustering is applied to all the estimates in Table 7. The first stage linear regression and the reduced form probit,

[34] The estimation results when using a linear probability model remain largely unchanged. The results are available upon request from the authors.

[35] About one-quarter of the interviewers are regular dependent employees of the survey company, while the others were hired (and trained) for this specific project, although they might have worked for the same or similar companies in the past.

Table 7

The Effect of the Local Share of Resident Migrants on Migrants' Employment

	Dependent variable: 1 = employed				
Variables	Probit (1)	Probit (2)	IV-Probit (3)	First stage (4)	RF-Probit (5)
% of non-Italians	−0.001	−0.005	−0.095***	–	–
	(0.010)	(0.011)	(0.026)		
Building structure†	–	–	–	−1.377**	0.196**
				(0.687)	(0.082)
Characteristics of the interviewee					
Age	0.263***	0.279***	0.165*	−0.229	0.280***
	(0.054)	(0.057)	(0.087)	(0.302)	(0.071)
1 = female	−0.446**	−0.681	−1.097***	−6.832***	−0.677
	(0.175)	(0.471)	(0.389)	(1.903)	(0.430)
1 = at least sec. edu	−0.068	−0.193	−0.109	0.164	−0.188
	(0.199)	(0.224)	(0.165)	(0.792)	(0.185)
1 = illegal immigrant	−0.605***	−0.522**	−0.489**	−1.711*	−0.492*
	(0.223)	(0.221)	(0.211)	(1.031)	(0.264)
Characteristics of the interviewer					
1 = Italian	–	−0.199	−0.594	−5.314	−0.136
		(0.445)	(0.585)	(4.500)	(1.367)
1 = Albanian	–	−0.572	−1.136	−8.873*	−0.445
		(0.573)	(0.700)	(5.345)	(1.374)
1 = graduate	–	0.229	−0.011	−2.105	0.283
		(0.390)	(0.428)	(3.161)	(0.359)
1 = professional‡	–	0.354	0.256	0.093	0.372
		(0.285)	(0.298)	(1.921)	(0.378)
1 = interviewer-interviewee same gender	–	0.238	0.874**	7.751***	0.211
		(0.456)	(0.410)	(2.063)	(0.383)
Quality of interview§	–	0.160***	0.109*	0.021	0.161**
		(0.062)	(0.062)	(0.258)	(0.063)
Observations	478	478	478	478	478

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Robust standard errors in parentheses, clustered by census tract. †Ratio of total residential square metres to number of residential buildings in the block, normalised at the city level. *Source.* 2001 Census. ‡Interviewer is a dependent employee of the survey company. §Interviewers self-reported evaluation of the level of understanding of the questions by the interviewee (0–10). Additional controls: age squared, city and district dummies, dummies for years since migration in Italy (less than 5 years, 5–10 years, 15–20 years and more than 20 years), dummies for origin (New member countries, Western Balkans, other origins). RF = Reduced Form.

obtained by replacing the first stage linear specification of the endogenous variable into the main model, are shown in columns 4 and 5 respectively.[36]

The main result of our study is the negative and significant effect of the percentage of migrants in the block on the probability of employment (column 3). In terms of size, the point estimate of the coefficient is −0.095[37] and it implies an average marginal

[36] In order to make the results of the reduced-form model comparable to those in column 1, we include the residuals of the first-stage regression among the regressors, otherwise the common normalisation to unity of the variance of the error term in the probit model would be inconsistent with the same assumption imposed in the IV-probit model. The standard errors of the estimates in column 5 are bootstrapped (stratifying by city) to account for this generated regressor.

[37] Results are very similar if we do not control for years since migration, which is potentially endogenous. The IV coefficient becomes −0.080 with a standard error of 0.039.

effect of about 2 percentage points, over an average employment rate of 88%, for each percentage point change in the share of immigrants residing in the block.[38]

This magnitude, however, needs to be taken with care. Our approach is likely to identify a Local Average Treatment Effect (LATE; Angrist and Imbens, 1994), as our estimates are identified by the subgroup of the immigrants residing in areas where the presence of immigrants is related to discrimination in the housing market or to the illegal status of immigrants (through the residential building structure). Note that this may actually be the LATE of policy interest for populations with fragile or uncertain attachment to the native markets but it still cannot be interpreted in a global sense. A local treatment effect of illegal migrants would also explain why the IV estimates are more negative than the OLS estimates. Illegal migrants can only compete for jobs in the informal sector and, due to the thin size of the illegal jobs market, congestion externalities are likely to be stronger in this context. In subsection 4.2, we use an alternative estimation strategy and we always find a negative, albeit smaller, effect of immigrant density on employment.

4.1. *Dealing with Weak Instruments in Non-linear Models*

In assessing the robustness of our main finding, it is important to notice the F-test of the excluded instrument in the first stage is just above 4, which, to some readers, may indicate a problem of weak instruments.

To tackle this issue we extend to non-linear models the reduced-form approach suggested by Angrist and Krueger (2001) and further developed by Chernozhukov and Hansen (2008) for linear models. To describe our procedure, consider our IV system of two equations:

$$y^*_{icdb} = \alpha_1 m_{cdb} + \boldsymbol{\alpha}_2 \boldsymbol{X}_{icdb} + \epsilon_{icdb}, \tag{2}$$

$$m_{cdb} = \beta_1 z_{cdb} + \boldsymbol{\beta}_2 \boldsymbol{X}_{icdb} + v_{icdb}, \tag{3}$$

where (2) is equivalent to (1), with the only difference that we now explicitly consider the dependent variable as a latent outcome and we indicate it with a star, following the common convention. Equation (3) is the first stage linear regression with z_{cdb} as the instrument.

The reduced-form model is obtained by replacing (3) into (2):

$$y^*_{icdb} = (\beta_1 \alpha_1) z_{cdb} + (\boldsymbol{\beta}_2 \alpha_1 + \boldsymbol{\alpha}_2) \boldsymbol{X}_{icdb} + \alpha_1 v_{icdb} + \epsilon_{icdb}, \tag{4}$$

which can be simply estimated as a probit under the usual distributional assumption $\epsilon_{icdb} \sim \text{i.i.d.N}(0,1)$. The only minor complication is the presence of the unobservable first-stage error v_{icdb} among the explanatory variables and it can be addressed as in Rivers and Vuong (1988) by replacing it with the estimated OLS residuals and appropriately adjusting the standard errors to account for the generated regressor. This is the exact procedure used to produce the estimates reported in column 5 of Table 7.[39]

[38] We compute this average partial effect by first calculating the marginal effect for every observation and then averaging over the entire sample.
[39] The standard errors are bootstrapped (stratifying by city).

Equation (4) shows that the standard test statistics for the null hypothesis $\beta_1\alpha_1 = 0$ can be used to make inference about the statistical significance of the main parameter of interest α_1, extending the results Chernozhukov and Hansen (2008) to non-linear models.[40] In other words, one can interpret the usual z-statistics of the coefficient on the instrument in the first-stage model as a test of the statistical significance of α_1 that is robust to weak instruments, as no information about the strength of the correlation between the endogenous regressor and the instrument is used to derive it.

In our specific setting, the z-statistics of the main effect derived from the joint maximum likelihood estimation of (2) and (3) is 3.58 (column 3 of Table 7), whereas the same statistics in the reduced form model declines to 2.39 (column 5 of Table 7), which is approximately one third lower but still allows rejection of the null.

We also extend the procedure of Chernozhukov and Hansen (2008) to derive a weak-instrument robust confidence interval for α_1. Define a wide enough range of potential values for α_1, A, and for each $a \in A$ rewrite (2) as follows:

$$y^*_{icdb} = (\alpha_1 - a)m_{cdb} + am_{cdb} + \boldsymbol{\alpha}_2\boldsymbol{X}_{icdb} + \epsilon_{icdb}. \tag{5}$$

Then, replace the first instance of m_{cdb} with the first-stage (3):

$$y^*_{icdb} = [\beta_1(\alpha_1 - a)]z_{cdb} + a\ m_{cdb} + (\boldsymbol{\beta}_2\alpha_1 + \boldsymbol{\alpha}_2)\boldsymbol{X}_{icdb} + (\alpha_1 - a)v_{icdb} + \epsilon_{icdb}. \tag{6}$$

In the simple linear context, Chernozhukov and Hansen (2008) propose estimating (6) by moving the term am_{cdb} to the left-hand side, effectively transforming the dependent variable. By the same argument made above, the usual test statistics for the significance of the coefficient on the instrument in such a modified reduced-form equation tests the null $\alpha_1 = a$ and iterating over several values of a allows constructing a confidence interval for α_1 that does not use information about the strength of the correlation between the instrument and the endogenous variable.

In our setting y^*_{icdb} is not observable and it is not possible to transform the dependent variable as in Chernozhukov and Hansen (2008). However, we can leave am_{cdb} on the right-hand side of (6) and estimate it as a constrained probit, forcing the coefficient of the endogenous variable m_{cdb} to equal a. By doing so, the endogeneity of m_{cdb} becomes irrelevant for the consistent estimation of $[\beta_1(\alpha_1 - a)]$.

In practice, we proceed as follows:

(*i*) set A as the set of real numbers between −0.3 and 0.15, spaced 0.001;
(*ii*) estimate (6) for each $a \in A$ and retain the z-statistics for $[\beta_1(\alpha_1 - a)]$;[41]
(*iii*) construct the $1 - p$ confidence interval as the set of as such that the z-statistics is smaller than $c(1 - p)$ where $c(1 - p)$ is the $(1 - p)$th percentile of a χ^2_1 distribution.

Applying this procedure to our setting yields a 95% confidence interval for α_1 of [−0.300,−0.018], which compares with the narrower interval derived from the usual maximum likelihood asymptotics of [−0.147,−0.043]. What is important for our

[40] Given the validity of the instrument, namely its exogeneity and relevance, the null hypothesis implies $\alpha_1 = 0$.
[41] Notice that, under the null the term $(\alpha_1 - a)v_{icdb}$ disappears from (5), thus simplifying its estimation.

purposes is that, in both cases, the entire interval lies on the negative side of the real line and excludes the zero, thus reassuring about the robustness of our finding.

4.2. *Alternative Identification Strategy*

In this subsection, we compare our identification strategy with that of Bayer *et al.* (2008), which rests on the comparison of blocks within narrowly defined groups using a fixed-effect model.

In practice we estimate models similar to (1) including a set of fixed effects for narrowly defined groups of blocks and excluding observations in isolated blocks. In terms of econometric identification, the fixed effects are meant to control for local unobservables and, thus, play exactly the same role of our instruments for identification purposes.

The groups of blocks are defined on the basis of a geographical criterion within a circle of radius 1.5 kilometres. Besides being simple, such a criterion allows us to use for this analysis all the blocks selected through the proximity criterion (see subsection 2.1), as well as others that were randomly selected but happen to be very close to each other.[42]

Once we include group fixed effects, our regressor of interest – immigration density – only varies within non-isolated blocks, that are blocks with a non-empty set of adjacent sampled blocks within a circle of radius 1.5 kilometres. Therefore, this fixed-effect approach comes at the cost of reducing the size of the sample to 254 individuals and 78 census tracts. The non-isolated census tracts are coded into 28 groups, with on average nine observations per group.

In Table 8 we report results obtained using this alternative fixed-effect strategy.[43] Given the smaller sample size, the control set needs to be modified slightly to make the model more parsimonious otherwise the outcome would be perfectly predicted for too many individuals. In the footnote to the Table, we describe the new set of controls and, for brevity, in the Table we only report the coefficients of interest. For comparison, columns 1 and 2 of Table 8 replicate our main results in columns 2 and 3 of Table 7 conditioning on the more parsimonious set of controls. Columns 3 and 4 show results from the alternative identification strategy. In column 3, we do not include block-groups fixed effects, while in column 4 we do include them. In this sense, the results in column 4 should be compared with those in column 2.

We find that the estimated effects of residential segregation are still negative. Moreover, as in Table 7, the bias in α_1 seems to be positive. The magnitude of the estimated effect, however, is smaller. Indeed, the average marginal effect is of about 0.9 percentage points, about half the size of our IV estimates.

[42] Groups are defined exploiting information on the exact location of the census tracts matched to each sampled individual. Two individuals belong to the same group if the centroids of the census tracts that contain their exact address (or the centroids of their closest sampled census tract) are included in a circle of radius 1.5 kilometres.

[43] We report results from probit models. Results obtained using a logistic distribution, which is robust to the incidental parameter problem (Neyman and Scott, 1948), remain qualitatively unchanged.

Table 8

Estimates with Neighbourhood Fixed Effects

	Dependent variable: 1 = employed			
	All sample		Adjacent blocks†	
Variables	Probit (1)	IV-Probit (2)	Probit (3)	Probit (4)
Percentage of non-Italians	−0.001	−0.091***	−0.031**	−0.052**
	(0.010)	(0.028)	(0.013)	(0.024)
Block-pair fixed effects	no	no	no	yes
Observations	478	478	254	254

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Robust standard errors, clustered by census tract, in parentheses. †The sample is limited to individuals residing in census tracts where the closest (sampled) census track is within a circle of ray 1.5 kilometres. Additional controls: age, gender, dummies for education, legal status, year of arrival in Italy (categorical), whether from NMS or from Balkans, district and city fixed effects, interviewer evaluation of interviewed, whether interviewer is Italian or Albanian.

4.3. *Additional Evidence and Discussion*

Table 7 shows that the IV coefficient on our regressor of interest is, in absolute value, significantly larger that its non-IV counterpart, hence the overall endogeneity bias seems to be positive. This is the combined outcome of the many potential sources of endogeneity in our model, such as individual sorting, unobserved neighbourhood shocks or measurement error (see Section 3). Once these factors have been taken into consideration, we uncover a negative relationship between employment and immigrant residential density. As discussed in subsection 1.1, this finding can be rationalised by several alternative mechanisms and it is extremely difficult to discriminate between the various explanations.

Even though we cannot test here any particular mechanism, we present some additional results in Table 9 that provide further insights. Table 9 shows results obtained using alternative measures of immigrant residential density. For comparison, our baseline specification is reported in column 1 (see Table 7). In column 2, we replace the share of non-Italians residing in the block with the share of households belonging to the same ethnic group, which is the common proxy for ethnic networks used in the literature (see subsection 1.1). If ethnic groups find employment in particular jobs and industries in which own-ethnics are over-represented, one should expect a positive and significant effect from such a modified specification. Table 9 shows, instead, a negative effect that is even larger than our benchmark.[44] This finding is consistent with the descriptive evidence in Table 4, which suggests that informal hiring networks do not seem to play a major role in our setting, as those living in areas with higher shares of migrants are not (significantly) more likely to find jobs through friends.

[44] All IV estimates in Table 9 are statistically significant and all the corresponding 95% weak instrument robust confidence intervals lay entirely on the negative side of the real line.

Table 9

Extensions with Different Measures of Immigrant Density

	Dependent variable: 1 = employed				
	Benchmark (1)	Same origin† (2)	Illegal‡ (3)	Legal§ (4)	Language adjusted¶ (5)
Immigrant density	−0.095***	−0.160***	−0.299***	−0.107***	−0.080***
	(0.026)	(0.041)	(0.056)	(0.026)	(0.024)
Observations	478	478	478	478	478

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Robust standard errors in parentheses, clustered by census tract. †Share of immigrants in the block belonging to one's same origin (European new member states, Western Balkans and Other countries). ‡Computed as the share of immigrants in the block times the share of illegal immigrants interviewed in the block. §Computed as the share of immigrants in the block times the share of legal immigrants interviewed in the block. ¶See text (subsection 4.3). Additional controls: age, age squared, gender, a dummy equal 1 if the immigrant has at least a secondary education diploma, city and district dummies, dummies for years since migration in Italy (less than 5 years, 5–10 years, 15–20 years and more than 20 years), dummies for origin (new member states, Western Balkans, other origins).

At the same time, though, Table 5 shows that illegal immigrants are more likely to rely on informal networks to look for employment and in the last two columns of Table 9 we further investigate this issue. In particular, we estimate separately the employment effect of the share of legal and illegal migrants residing in the block.[45] The results show that the density of illegal immigrants living nearby exercises a more negative effect on employment (of all migrants) compared to the density of legal migrants. One possible explanation for this finding, which would also be supported by the evidence in Table 4, is that the ethnic network used by illegal immigrants is mostly composed by legal immigrants residing nearby, consistently with the design of migration policies in Italy and a simple integration process by which the more recent and younger immigrants are more likely to be illegal (see Table 5) and to rely on the existing (established) networks of social contacts, most of whom are legal.

The last column on the right-hand side reports the effect on employment of the share of non-Italians residing in the block, adjusted by their relative language skills. Specifically, we divide the number of immigrants in the block (from our census) by the average score obtained by actually interviewed migrants who took the language test in the block. Similarly, we divide the number of Italians in the block by the maximum test score. Hence, if all immigrants in the block were perfectly proficient in Italian, the adjusted and unadjusted immigrant shares would be equal. The lower the average immigrants' language proficiency, the larger the adjusted share of immigrants in the block.

When using this indicator we still find a negative effect on the employment prospect of migrants. However, the estimated average treatment effect is smaller (in absolute value) than in our baseline model. A possible interpretation of these findings is related to the idea that more language proficient migrants are better integrated into the labour market and constitute a better network for the transmission of job-related information.

[45] We have experimented with several alternative definitions of illegal immigrants and results change only marginally.

Table 10

The Effect of the Local Share of Resident Migrants on Natives' Employment

	Dependent variable: 1 = employed				
Variables	Probit (1)	Probit (2)	IV-Probit (3)	First stage (4)	RF-Probit (5)
% of non-Italians	0.005	0.006	−0.022	–	–
	(0.012)	(0.014)	(0.066)		
Building structure†	–	–	–	−1.573*	0.036
				(0.868)	(0.201)
Characteristics of the interviewee					
Age	0.375***	0.399***	0.385***	−0.190	0.401**
	(0.081)	(0.083)	(0.104)	(0.378)	(0.164)
1 = female	−0.899***	−0.989*	−1.303	−12.403***	−1.062
	(0.314)	(0.564)	(0.888)	(4.488)	(1.627)
1 = at least secondary education	−0.231	−0.367	−0.316	1.343	−0.356
	(0.283)	(0.315)	(0.348)	(1.584)	(0.485)
Characteristics of the interviewer					
1 = Albanian	–	−0.486	−0.976	−16.758**	−0.626
		(0.848)	(1.402)	(7.796)	(2.730)
1 = Italian	–	0.204	−0.213	−12.645*	0.066
		(0.725)	(1.227)	(6.939)	(2.798)
1 = graduate	–	−1.051*	−1.015*	0.378	−1.054
		(0.541)	(0.584)	(5.103)	(1.963)
1 = professional‡	–	0.307	0.402	3.380	0.337
		(0.354)	(0.380)	(2.447)	(0.481)
1 = interviewer-interviewee same gender	–	−0.005	0.403	14.870***	0.079
		(0.525)	(1.129)	(4.626)	(1.716)
Quality of interview§	–	0.289**	0.233	−1.573*	0.275*
		(0.114)	(0.165)	(0.875)	(0.153)
Observations	179	179	179	179	179

Notes. *p < 0.10, **p < 0.05, ***p < 0.01. Robust standard errors in parentheses, clustered by census tract. †Ratio of total residential square metres to number of residential buildings in the block, normalised at the city level. *Source.* Census 2001. ‡Interviewer is a dependent employee of the survey company. §Interviewers self-reported evaluation of the level of understanding of the questions by the interviewee (0–10). Additional controls: age squared, city and district dummies. On the subsample of Italians only.

Finally, in Table 10 we also document that the share of immigrants living nearby does not matter for the employment of natives, a result that is consistent with many papers (Card, 1990, 2005; Friedberg and Hunt, 1995; Angrist and Kugler, 2003; Bodvarsson *et al.*, 2008; Ottaviano and Peri, 2011).

The above analysis does not provide a direct support for any of the theoretical mechanisms outlined by the previous literature and summarised in subsection 1.1. In particular, the fact that the employment prospects of Italians are not affected by the concentration of migrants in their blocks is not consistent with the redlining of some areas by the employers. Moreover, the fact that concentrations of illegal migrants have a stronger negative effect on employment than concentrations of legal migrants does not lend support to the view that there are important congestion externalities in the search of legal jobs associated with residential concentration. Indeed, illegal migrants cannot compete with other migrants in the legal labour market. However, the informal labour market is significantly smaller than the legal market and, in this context, congestion externalities may be stronger. Newly arrived illegal migrants, who need a job to apply for legal status and who have a very low reservation wage before family unification, may crowd out from employment in the informal sector other migrants. This would also be consistent with a stronger negative effect of residential concentration in our IV estimates. Illegal migrants prefer to locate in blocks with a relatively low residential density in order to reduce the risk of detection. Thus, the type of externalities which are consistent with our findings call into play the relatively large share of illegal migrants in countries like Italy.

5. Conclusions

Europe, notably Southern Europe, experienced very sizeable inflows of immigrants in the decade before the Great Recession. Policies to promote the economic and social integration of newcomers must be supported by an analysis of the relationship between residential patterns and employment prospects of migrants. Due to a lack of data, this relationship has barely been investigated by the literature for Southern Europe. Moreover, data on residential concentration of migrants typically do not capture illegal migration, which is quite sizeable in Southern Europe, and do not generally permit identification of causal effects of residential concentration on labour market outcomes.

In this article, we take advantage of the information gathered by a survey covering both legal and illegal migrants in eight cities in the North of Italy, a region experiencing a doubling of its foreign population just while disinvesting in social housing. The survey provides detailed information on residential patterns of migrants, that can be matched with the official census data to obtain instruments allowing us to identify the causal effect of the percentage of migrants living nearby on the employment status of migrants.

Our analysis uncovers a negative externality, which is higher if the immigrants living nearby are illegal. The effect is sizeable. Our results suggest that if the incidence of migrants in the block increased from its median value (approx. 15%) to the 75th percentile of its distribution (approx. 25%), the employment rate of migrants in such a median block would drop by between 10 and 20 percentage points (from 88% to 78–68%), depending to the model specification and estimation strategy.

The relationship between residential proximity of individuals from the same ethnic group and the probability of finding a job is extremely complex and our findings can be rationalised by several alternative mechanisms. While our data do not allow us to evaluate with sufficient precision which particular mechanism is behind our main results, we can nevertheless rule out some of the explanations provided by the literature for the externalities associated with a large share of migrants in the block, and provide alternative tentative explanations.

For example, redlining by employers of areas with a large concentration of migrants is not supported by our analysis. We likewise do not find support to the view that residential concentration is a source of congestion externalities in job search in the legal market. However, our findings are consistent with a mechanism in which thin labour market externalities arise in the informal labour market, the only market illegal migrants have access to.

Analysing the interactions between illegal status, illegal work and residential concentration of migrants remains a key area of future research, particularly relevant for the new destination countries of immigration in Europe. This analysis has important policy implications in terms not only of scale and design of social housing programmes but also of policies deciding upon residence and work permits for migrants. Migration policies conditioning legal entry to the fact of having already a job in the destination country prior to the arrival may increase the negative externalities of residential concentration on employment of migrants that we document in this article. These policies induce new migrants who come illegally to desperately search for jobs to qualify for legal status and they may crowd-out other migrants. More realistic migration policies, allowing for legal status while searching for (legal) jobs, may reduce these negative spillovers of residential concentration.

Appendix A. Additional Results.

Table A1

Characteristics of the Sampled Cities

City	Size* (1)	Income per capita† (2)	Average age‡ (3)	Unemployment rate§ (4)	Employment rate§ (5)	Share of immigrants¶ (6)
Alessandria	93,676	13,648	46	0.065	0.45	0.11
Bologna	374,944	18,771	47	0.044	0.48	0.09
Brescia	190,844	15,812	45	0.048	0.48	0.16
Lucca	89,640	14,920	45	0.065	0.46	0.08
Milano	1,295,705	21,358	45	0.044	0.49	0.14
Prato	185,091	12,446	43**	0.057	0.51	0.14
Rimini	140,137	12,059	45††	0.070	0.46	0.09
Verona	265,368	15,220	44	0.049	0.48	0.13
Italy	60,045,068	12,953	43	0.112	0.43	0.06
Northern Italy‡‡	27,390,496	15,529	44	0.049	0.49	0.09

Notes. *Number of residents. *Source.* ISTAT, 2009. †Annual gross taxable income. *Source.* Tax declarations, 2007. ‡*Source.* ISTAT, 2007. §*Source.* ISTAT, 2001 Population Census. ¶*Source.* Population registers of city councils, ISTAT, 2009. Legal immigrants only. ***Source.* City Population Register, 2005. ††*Source.* City Population Register, 2009. ‡‡Norther Italy includes the following regions: Piemonte, Valle D'Aosta, Lombardia, Trentino Alto Adige, Veneto, Friuli Venezia Giulia, Liguria, Emilia Romagna.

IGIER-Bocconi University, CEPR, fRDB and IZA
LSE, CEP and fRDB
Cornell University, CEPR, EIEF, fRDB and IZA
University of Geneva, CEPR, fRDB and IZA

Additional Supporting Information may be found in the online version of this article:

Data S1.

References

Ahmed, A.M., Andersson, L. and Hammarstedt, M. (2010). 'Can discrimination in the housing market be reduced by increasing the information about the applicants?', *Land Economics*, vol. 86(1), pp. 79–90.
Ahmed, A.M. and Hammarstedt, M. (2008). 'Discrimination in the rental housing market: a field experiment on the internet', *Journal of Urban Economics*, vol. 64(2), pp. 362–72.

Altonji, J.G. and Card, D. (1991). 'The effects of immigration on the labor market outcomes of less-skilled natives', in (J.M. Abowd and R. B. Freeman, eds.), *Immigration, Trade, and the Labor Market*, pp. 201–34, chap. 7, Chicago: The University of Chicago Press.

Angrist, J.D. and Imbens, G.W. (1994). 'Identification and estimation of local average treatment effects', *Econometrica*, vol. 62(2), pp. 467–2.

Angrist, J.D. and Krueger, A.B. (2001). 'Instrumental variables and the search for identification: from supply and demand to natural experiments', *Journal of Economic Perspectives*, vol. 15(4), pp. 69–85.

Angrist, J.D. and Kugler, A.D. (2003). 'Protective or counter-productive? Labour market institutions and the effect of immigration on EU natives', ECONOMIC JOURNAL, vol. 113(488), pp. F302–31.

Bach, R.L. and Portes, A. (1985). *Latin Journey: Cuban and Mexican Immigrants in the United States*, Berkeley: University of California Press.

Bailey, T. and Waldinger, R. (1991). 'Primary, secondary, and enclave labor markets: a training systems approach', *American Sociological Review*, vol. 56(4), pp. 432–45.

Baldini, M. and Federici, M. (2011). 'Ethnic discrimination in the Italian rental housing market', *Journal of Housing Economics*, vol. 20(1), pp. 1–14.

Battu, H., Seaman, P. and Zenou, Y. (2011). 'Job contact networks and the ethnic minorities', *Labour Economics*, vol. 18(1), pp. 48–56.

Bauer, T.K., Flake, R. and Sinning, M. (2011). 'Labor market effects of immigration: evidence from neighborhood data', Institute for the Study of Labor (IZA).

Bayer, P., Ross, S.L. and Topa, G. (2008). 'Place of work and place of residence: informal hiring networks and labor market outcomes', *Journal of Political Economy*, vol. 116(6), pp. 1150–96.

Beckers, P. and Borghans, L. (2011). 'Segregation in neighbourhoods and labour market outcomes of immigrants: evidence from random assignment in the Netherlands', United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology.

Bertrand, M., Luttmer, E.F.P. and Mullainathan, S. (2000). 'Network effects and welfare cultures', *Quarterly Journal of Economics*, vol. 115(3), pp. 1019–55.

Bisin, A., Patacchini, E., Verdier, T. and Zenou, Y. (2011). 'Ethnic identity and labour market outcomes of immigrants in Europe', *Economic Policy*, vol. 26(65), pp. 57–92.

Boccard, N. and Zenou, Y. (2000). 'Racial discrimination and redlining in cities', *Journal of Urban Economics*, vol. 48(2), pp. 260–85.

Bodvarsson, O.B., Van den Berg, H.F. and Lewer, J.J. (2008). 'Measuring immigration's effects on labor demand: a reexamination of the Mariel boatlift', *Labour Economics*, vol. 15(4), pp. 560–74.

Bonacich, E. and Light, I. (1988). *Immigrant Entrepreneurs: Koreans in Los Angeles*, Berkeley, CA: University of California Press.

Borjas, G.J. (1987). 'Immigrants, minorities and labour market competition', *Industrial and Labour Relation Review*, vol. 40(3), pp. 382–92.

Borjas, G.J. (1995). 'Ethnicity, neighborhoods, and human-capital externalities', *American Economic Review*, vol. 85(3), pp. 365–90.

Borjas, G.J. and Bronars, S.G. (1989). 'Consumer discrimination and self-employment', *Journal of Political Economy*, vol. 97(3), pp. 581–605.

Bosch, M., Carnero, M.A. and Farré, L. (2010). 'Information and discrimination in the rental housing market: evidence from a field experiment', *Regional Science and Urban Economics*, vol. 40(1), pp. 11–9.

Brueckner, J.K. and Zenou, Y. (2003). 'Space and unemployment: the labor-market effects of spatial mismatch', *Journal of Labor Economics*, vol. 21(1), pp. 242–62.

Card, D. (1990). 'The impact of the Mariel boatlift on the Miami labor market', *Industrial and Labor Relations Review*, vol. 43(2), pp. 245–57.

Card, D. (2005). 'Is the new immigration really so bad?', ECONOMIC JOURNAL, vol. 115(507), pp. F300–23.

Card, D. and Rothstein, J. (2007). 'Racial segregation and the black-white test score gap', *Journal of Public Economics*, vol. 91(11–12), pp. 2158–84.

Carpusor, A.G. and Loges, W.E. (2006). 'Rental discrimination and ethnicity in names1', *Journal of Applied Social Psychology*, vol. 36(4), pp. 934–52.

Cesareo, V. and Blangiardo, G.C. (2009). *Indici di integrazione: un indagine empirica sulla realta migratoria italiana*, Milano: FrancoAngeli.

Chernozhukov, V. and Hansen, C. (2008). 'The reduced form: a simple approach to inference with weak instruments', *Economics Letters*, vol. 100(1), pp. 68–71.

Chiswick, B.R. (1991). 'Speaking, reading, and earnings among low-skilled immigrants', *Journal of Labor Economics*, vol. 9(2), pp. 149–70.

Chiswick, B.R. and Miller, P.W. (1995). 'The endogeneity between language and earnings: international analyses', *Journal of Labor Economics*, vol. 13(2), pp. 246–88.

Chiswick, B.R. and Miller, P.W. (2005). 'Do enclaves matter in immigrant adjustment?', *City and Community*, vol. 4(1), pp. 5–35.

Conley, T. and Topa, G. (1999). 'Socio-economic distance and spatial patterns in unemployment', *Journal of Applied Econometrics*, vol. 17(4), pp. 303–27.

Coulson, N.E., Laing, D. and Wang, P. (2001). 'Spatial mismatch in search equilibrium', *Journal of Labor Economics*, vol. 19(4), pp. 949–72.

Cutler, D.M. and Glaeser, E.L. (1997). 'Are ghettos good or bad?', *Quarterly Journal of Economics*, vol. 112(3), pp. 827–72.

Cutler, D.M., Glaeser, E.L. and Vigdor, J.L. (1999). 'The rise and decline of the American ghetto', *Journal of Political Economy*, vol. 107(3), pp. 455–506.

Damm, A.P. (2009). 'Ethnic enclaves and immigrant labor market outcomes: quasi-experimental evidence', *Journal of Labor Economics*, vol. 27(2), pp. 281–314.

Dustmann, C., Fasani, F. and Speciale, B. (2010). 'Legal status and income allocation of immigrant households', mimeo, UCL.

Edin, P.A., Fredriksson, P. and Aaslund, O. (2003). 'Ethnic enclaves and the economic success of immigrants: evidence from a natural experiment', *Quarterly Journal of Economics*, vol. 118(1), pp. 329–57.

Evans, W., Oates, W. and Shwab, R. (1992). 'Measuring peer group effects: a study of teenage behavior', *Journal of Political Economy*, vol. 100(5), pp. 966–91.

Falcon, L.M. (2007). 'Social networks and Latino immigrants in the labor market: a review of the literature and evidence', in (M. Montero–Sieburth and E. Melendez, eds.), *Latinos in a Changing Society*, pp. 83–101, Westport, CT: Praeger.

Falcon, L.M. and Melendez, E. (2001). 'Racial and ethnic differences in job searching in urban centers', in (A. O'Connor, C. Tilly and L.D. Bobo, eds.), *Urban Inequality: Evidence from Four Cities*, pp. 341–71, New York: Russell Sage Foundation.

Friedberg, R.M. and Hunt, J. (1995). 'The impact of immigrants on host country wages, employment and growth', *Journal of Economic Perspectives*, vol. 9(2), pp. 23–44.

Frijters, P., Shields, M.A. and Price, S.W. (2005). 'Job search methods and their success: a comparison of immigrants and natives in the UK', Economic Journal, vol. 115(507), pp. 359–76.

Gabriel, S.A. and Rosenthal, S.S. (1999). 'Location and the effect of demographic traits on earnings', *Regional Science and Urban Economics*, vol. 29(4), pp. 445–61.

Gautier, P.A. and Zenou, Y. (2010). 'Car ownership and the labor market of ethnic minorities', *Journal of Urban Economics*, vol. 67(3), pp. 392–403.

Granovetter, M. (1973). 'The strength of weak ties', *American Journal of Sociology*, vol. 78(6), pp. 1360–80.

Granovetter, M. (1974). *Getting a Job: A Study of Contacts and Careers*, Cambridge, MA: Harvard University Press.

Granovetter, M. (1983). 'The strength of weak ties: a network theory revisited', *Sociological Theory*, vol. 1, pp. 201–23.

Granovetter, M. (1985). 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, vol. 91(3), pp. 481–510.

Hanson, A. and Hawley, Z. (2011). 'Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in us cities', *Journal of Urban Economics*, vol. 70(2), pp. 99–114.

Harris, L. and Rosenbaum, E. (2001). 'Residential mobility and opportunities: early impacts of the moving to opportunity demonstration program in Chicago', *Housing Policy Debate*, vol. 12(2), pp. 321–46.

Hellerstein, J.K., Neumark, D. and McInerney, M. (2008). 'Spatial mismatch or racial mismatch?', *Journal of Urban Economics*, vol. 64(2), pp. 464–79.

Holzer, H.J., Quigley, J.M. and Raphael, S. (2003). 'Public transit and the spatial distribution of minority employment: evidence from a natural experiment', *Journal of Policy Analysis and Management*, vol. 22(3), pp. 415–42.

Ihlanfeldt, K.R. (1991). 'A primer on spatial mismatch within urban labor markets'. in (R. Arnott and D. McMillen, eds.), *A Companion to Urban Economics*, pp. 404–17, Boston, MA: Blackwell Publishing.

Ihlanfeldt, K.R. (1992). *Job Accessibility and the Employment and School Enrollment of Teenagers*, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Ihlanfeldt, K.R. (1997). 'Information on the spatial distribution of job opportunities within metropolitan areas', *Journal of Urban Economics*, vol. 41(2), pp. 218–42.

Ihlanfeldt, K.R. and Sjoquist, D. (1998). 'The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform', *Housing Policy Debate*, vol. 9(4), pp. 849–92.

Ioannides, Y.M. (2012). *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton, NJ: Princeton University Press.

Ioannides, Y.M. and Loury, L.D. (2004). 'Job information networks, neighborhood effects, and inequality', *Journal of Economic Literature*, vol. 42(4), pp. 1056–93.

Kahanec, M. (2006). 'Ethnic specialization and earnings inequality: why being a minority hurts, but being a big minority hurts more', Institute for the Study of Labor (IZA).

Kain, J.F. (1968). 'The spatial mismatch hypotheis: three decades later', *Housing Policy Debate*, vol. 3(2), pp. 371–92.

Katz, L.F., Kling, J.R. and Liebman, J.B. (2001). 'Moving to opportunity in Boston: early results of a randomized mobility experiment', *Quarterly Journal of Economics*, vol. 116(2), pp. 607–54.

Lazear, E.P. (1999). 'Culture and language', *Journal of Political Economy*, vol. 107(S6), pp. S95–126.

Li, W. (1998). 'Anatomy of a new ethnic settlement: the Chinese ethnoburb in Los Angeles', *Urban studies*, vol. 35(3), pp. 479–501.
Loury, G. (1977). *A Dynamic Theory of Racial Differences. Women, Minorities, and Employment Discrimination*, Lexington, MA: Lexington Books.
Massey, D. (1990). 'Social structure, household strategies, and the cumulative causation of migration', *Population Index*, vol. 56(1), pp. 3–26.
Massey, D., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, E. (1998). *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Oxford: Oxford University Press.
Massey, D. and Espinosa, K.E. (1997). 'What's driving Mexico-US migration? A theoretical, empirical, and policy analysis', *American Journal of Sociology*, vol. 102(4), pp. 939–99.
Massey, D. and Zenteno, R. (1999). 'The dynamics of mass migration', *Proceedings of the National Academy of Sciences*, vol. 10(12), pp. 5328–35.
Mendenhall, R., DeLuca, S. and Duncan, G. (2006). 'Neighborhood resources, racial segregation, and economic mobility: results from the Gautreaux program', *Social Science Research*, vol. 35(4), pp. 892–923.
Mouw, T. (2002). 'Racial differences in the effects of job contacts: conflicting evidence from cross-sectional and longitudinal data', *Social Science Quarterly*, vol. 31(4), pp. 511–38.
Munshi, K. (2003). 'Networks in the modern economy: Mexican migrants in the US labor market', *Quarterly Journal of Economics*, vol. 118(2), pp. 549–99.
Musterd, S. and Andersson, R. (2005). 'Housing mix, social mix, and social opportunities', *Urban Affairs Review*, vol. 40(6), pp. 761–90.
Newman, K.S. (1999). *No Shame in My Game: The Working Poor in the Inner City*, New York: Knopf and Russell Sage Foundation.
Neyman, J. and Scott, E.L. (1948). 'Consistent estimates based on partially consistent observations', *Econometrica*, vol. 16(1), pp. 1–32.
OECD (2009). *International Migration Outlook*, Paris: OECD Publishing.
Ondrich, J., Stricker, A. and Yinger, J. (1999). 'Do landlords discriminate? The incidence and causes of racial discrimination in rental housing markets', *Journal of Housing Economics*, vol. 8(3), pp. 185–204.
Ottaviano, G.I. and Peri, G. (2011). 'Rethinking the effects of immigration on wages'. *Journal of the European Economic Association*, vol. 10(1), pp. 152–97.
Page, M. (1995). 'Racial and ethnic discrimination in urban housing markets: evidence from a recent audit study', *Journal of Urban Economics*, vol. 38(2), pp. 183–206.
Patacchini, E. and Zenou, Y. (2006). 'Search activities, cost of living and local labor markets', *Regional Science and Urban Economics*, vol. 36(2), pp. 227–48.
Patacchini, E. and Zenou, Y. (2012). 'Ethnic networks and employment outcomes', *Regional Science and Urban Economics*, vol. 42(6), pp. 938–49.
Pellizzari, M. (2010). 'Do friends and relatives really help in getting a good job?', *Industrial and Labor Relations Review*, vol. 63(3), pp. 494–510.
Pellizzari, M. (2013). 'The use of welfare by migrants in italy', *International Journal of Manpower*, vol. 34(2), pp. 155–66.
Phelps, E.S. (1972). 'The statistical theory of racism and sexism', *American Economic Review*, vol. 62(4), pp. 659–61.
Portes, A. (1998). 'Social capital: its origin and applications in modern sociology', *American Review of Sociology*, vol. 24, pp. 1–24.
Raphael, S. (1998). 'The spatial mismatch hypothesis and black youth joblessness: evidence from the San Francisco Bay area', *Journal of Urban Economics*, vol. 43(1), pp. 79–111.
Rivers, D. and Vuong, Q.H. (1988). 'Limited information estimators and exogeneity tests for simultaneous probit models', *Journal of Econometrics*, vol. 39(3), pp. 347–66.
Ross, S.L. (1998). 'Racial differences in residential and job mobility: evidence concerning the spatial mismatch hypothesis', *Journal of Urban Economics*, vol. 43(1), pp. 112–35.
Ross, S.L. and Zenou, Y. (2008). 'Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economic theory', *Regional Science and Urban Economics*, vol. 38(5), pp. 498–517.
Smith, T.E. and Zenou, Y. (2003). 'Spatial mismatch, search effort, and urban spatial structure', *Journal of Urban Economics*, vol. 54(1), pp. 129–56.
Waldinger, R. (1996). *Ethnic Los Angeles*, New York: Russell Sage Foundation Press.
Walker, R. and Hannan, M. (1989). 'Dynamic settlement processes: the case of US immigration', *Professional Geographer*, vol. 41(2), pp. 172–83.
Wang, Q. (2004). 'Labor market concentration of asian ethnic groups in us metropolitan areas: A disaggregated study', *Population, Space and Place*, vol. 109(6), pp. 479–94.
Wasmer, E. and Zenou, Y. (2002). 'Does city structure affect job search and welfare?', *Journal of Urban Economics*, vol. 51(3), pp. 515–41.
Weinberg, B. (2000). 'Black residential centralization and the spatial mismatch hypothesis', *Journal of Urban Economics*, vol. 48(1), pp. 110–34.

Weinberg, B. (2004). 'Testing the spatial mismatch hypothesis using inter-city variations in industrial composition', *Regional Science and Urban Economics*, vol. 34(5), pp. 505–32.
Wilson, W.J. (1996). *When Work Disappears: The World of the New Urban Poor*, New York: Knopf.
Yinger, J. (1986). 'Measuring racial discrimination with fair housing audits: caught in the act', *American Economic Review*, vol. 76(5), pp. 881–93.
Zenou, Y. (2002). 'How do firms redline workers?', *Journal of Urban Economics*, vol. 52(3), pp. 391–408.
Zenou, Y. (2008). 'The spatial mismatch hypothesis', in (L. Blume and S. Durlauf, eds.), *A Dictionary of Economics*, 2nd edn., pp. 83–101, London: MacMillan Press.
Zenou, Y. (2013). 'Spatial versus social mismatch', *Journal of Urban Economics*, vol. 74(C), pp. 113–32.