Immigrant labour market outcomes: new insights from a lack of language proficiency in Italy

Pieroni, Luca and d'Agostino, Giorgio and Lanari, Donatella and Scarlato, Margherita

Department of Political Science, University of Perugia, Department of Economics, Roma Tre University, Department of Medicine, University of Perugia, Department of Economics, Roma Tre University

2022

Immigrant labour market outcomes: new insights from a lack of language proficiency in Italy

Pieroni L.*, d'Agostino G.**, Lanari D. ***, Scarlato M.**

* Department of Political Science, University of Perugia
** Department of Economics, Roma Tre University
*** Department of Medicine, University of Perugia

Abstract

In this paper, we examine how immigrants' proficiency in the Italian language affects their labour market performance using hitherto unexploited immigration survey published by the Italian Institute of Statistics. With respect to immigrants with good proficiency in the Italian language, our empirical findings suggest that weak language proficiency reduces employment by about 25-30 percentage points. Language proficiency in Italian also significantly affects the hourly wages of immigrants. The point estimates suggest an hourly wage gap of more than 30% between immigrants with good proficiency in the Italian language compared to those with weak proficiency, irrespective of gender. Robustness checks confirmed our estimates.

Keywords: Immigrants, Language skills, Employment, Hourly wages
JEL codes: J15, J20, J31

---

1. Introduction

Immigration has become an important socioeconomic and public policy issue in Southern Europe with the recent increase in migration flows. It is a fact that immigrants often possess a weak position within the labour market. The evidence shows that the employment rate of immigrants is typically lower than that of natives and that their job opportunities mainly consist of low-skilled (and often precarious) work (Adserà and Pytliková, 2016). Italy, in particular, has registered an exceptional increase in low-skilled immigrant in recent cohorts, with a proportion of low-skilled immigrants that is much higher than that found in other European countries(Barone and Mocetti, 2011; Mariani et al., 2021).

Language skills are certainly an important aspect of the individual human capital of immigrants and a determinant of employment success. Job selection on employment may be based on a lack of proficiency in the destination country's language, inducing the migrant to work in jobs that require a lower education level than the level achieved in the origin country. This can lead to lower job performance and, in turn, amplifies differences in the employment rate and wages.

In this paper, we examine how weak language proficiency of immigrants affects their labour market performance in terms of employment and wages. Our laboratory is Italy and, unlike English or Spanish, immigrants' knowledge of Italian is generally more limited.[1] The sparse estimates we found for the effects of immigrants' language problems on employment outcomes are also controversial. Dustmann and Fabbri (2003) found a decrease of 22% in employment probability in the UK for immigrants who lack English skills. Gonzales (2010) found significant negative effects of the lack of host country language skills on employment in Spain, while Yao and van Ours (2015) found that language problems affect wages but not the employment probability of immigrants in the Netherlands.[2] It is undeniable that although the language fluency of immigrants is just one component of a larger integration process for migrants, and a lack of language proficiency has spillover effects on job satisfaction. The main explanation of this relationship is that a lack of language skills may induce the selection of migrants into jobs of a lower professional level (Bloemen, 2013).

However, the consequences of language proficiency that have received the most attention in the labour market relate to the effect on wages. Chiswick and Miller (2015) motivated it for the role of wages in summarizing economic status, as well as the historical availability of data on wages and language proficiency of immigrants in several developed countries. Empirical evidence has unanimously shown a significant relationship between the language proficiency of immigrants and wages, varying from $10\% - 20\%$ percent for the US, the UK and Australia and slightly less for Germany and Spain, although these point estimates increased when unobserved heterogeneity and measurement errors were included (Yao and van Ours, 2015; Guven and Islam, 2015; Chiswick et al., 2005).[3]

Thus, we are contributing to this debate, due to the availability of Italian data on labour market outcomes that include immigrants' language proficiency, which have as to now gone unexploited. We base our analysis on data from the Italian survey on immigrants published by the Italian Insti-

---

[1]See Chiswick (2008) for a discussion on the importance of studying less common languages.

[2]Conversely, the change in language policy seen in Catalonia in the 1980s, which promoted the study and use of Catalan instead of Spanish, increased the employment probability by 4-5 percentage points (Rendon, 2007).

[3]See Table 5.5.of Chiswick and Miller (2015) for an extended analysis of the empirical estimates.

tute of Statistics (IIS): the Conditions and Social Integration of Foreign Citizens that was conducted between 2011 and 2012 - and published in 2014 (CSIFC 2011-2012). In addition, we exploit data from the Income and Living Conditions of Households with Foreigners that was conducted in 2009 and refers to the year 2008 for wages (ILCHF 2009). Both surveys include questions on the social and economic conditions of interviewees and measures of language proficiency of immigrants. The data collected from both surveys are partly overlapping and complementary. For instance, the CSIFC 2011-2012 survey includes relevant information about immigrants' household members in the country of origin and in Italy but limits the employment information to labour market outcomes. The ILCHF 2009 is the first nation-wide survey on the socioeconomic conditions of the foreign population living in Italy, which contains information about the wages of immigrants.[4] Both datasets include self-reported language proficiency, even though the information from interviewees does not overlap completely. In the CSIFC 2011-2012, self-reported language proficiency is measured using ordered modalities that record writing and reading problems, whereas the ILCHF 2009 classifies a general perception of knowledge of the Italian language. Despite this limitation, a plausible aggregation over modalities of language proficiency provides interesting insights into the relationship between language skills and labour market performance, comparing the evolution of these effects on employment probability between 2009 and 2012.

Our empirical strategy exploits the assumption that the language proficiency of immigrants is known to be age different. Adult immigrants typically make the decision to migrate after they have obtained their education in the country of origin and are mainly motivated by the absence of employment opportunities, while immigrants who arrived during their childhoods, following family immigration, typically learnt the language of the host country quickly. The key challenge for causal estimates is that age at arrival may conceivably correlate with language acquisition, and it explains why the literature generally uses age at arrival as an instrumental variable (IV) to investigate the relationship between the language proficiency of immigrants and labour market performance. The underlying assumption is that children who were exposed to the host country language will be more proficient as adult. However, as acknowledged by Bleakley and Chin (2004, 2010) age at arrival may have other implications for integration, not only host language proficiency. Thus, to exclude other influences that are not directly linked with the cost of linguistic acquisition, we propose interacting age at arrival with a linguistic country distance measure of the language speaking origin country in comparison to the Italian, using a complementary linguistic proximity indicator built by Adserà and Pytliková (2015). The idea is that immigrants coming from countries with more similar languages might need relatively less language training or less effort to learn Italian compared to those whose native language is more distant from Italian.

Based on these considerations, we first show that a lack of Italian language proficiency is associated with a significantly large decrease in employment probability. Our findings mirror those of existing studies, with the employment reduction ranging from $25 - 30$ percentage points for immigrants with weak Italian language proficiency. Second, we focus on the effects of language proficiency on hourly wages. We find that a good knowledge of the host country language significantly increases the expected hourly wages of immigrants (almost 35% in the IV baseline es-

---

[4]It is worth noting that the definitions of immigrant and foreigner do not completely overlap because foreigners include individuals who were born in Italy but do not have Italian citizenship, while some immigrants are individuals who are born abroad and have moved to Italy, which may also be the case for some Italian citizens. Although the survey refers to foreigners, in this paper we use the two terms, immigrants and foreigners, interchangeably.

timates.). However, some aspects of the (unadjusted)wage gap are persistent, even considering a scenario in which all immigrants have good Italian proficiency. This means there are other factors that explain the difference in wages between Italians and immigrants. We also extend our analysis to investigate factors influencing immigrants' knowledge of the host country language, such as gender. While there is a large body of literature showing the different effects of gender on the relationship between language skills and labour market performance (e.g., Yao and van Ours (2015)), we emphasise that gender selection into occupations helps to explain large differences in wages. Robustness analyses validated our main findings.

The remainder of this paper is organised as follows. Section 2 describes the background of immigration in Italy and the evolution of immigrants' employment and wages in Italy. Section 3 discusses the data and presents descriptive statistics of the surveys used in the empirical section, while baseline results and robustness are discussed in Section 4. Section 5 concludes.

2. Background

Currently, Italy is one of the main immigration countries in Europe, with more than five million foreign residents in 2017. The majority of foreign nationals residing in Italy are extra-European citizens, although immigrants arriving from Central and Eastern Europe (e.g. Romania) from the second half of the 1990s onwards represent the most significant group. The immigration flow was encouraged when Romania entered the European Union (January 2007), so that Romanian, along with Bulgarian, immigrants acquired the right to reside and work in Italy; thus, 'irregular' immigrants became 'regular' immigrants. Now, Romanians represent 23.3% of the total immigrants in Italy, while Albanians are the second most significant immigrant group, representing 10% of all immigrants. Emigration from Northern Africa, particularly Morocco, represents the third largest group at 8%. Moreover, data on migration flows in the last 20 years show that emigration from Asia (China, India, Philippines) and Latin America (Peru, Ecuador) also increased sharply (Pieroni et al., 2021).

The linguistic origin root of these group of immigrates is generally very different from the Italian language. We use a recent classification by Adserà and Pytliková (2015), who captures the linguistic proximity between two languages based on information from Ethnologue (Lewis, 2009), to measure the specific proximity of Italian. The index for the Italian language is a sum of weights capturing the maximum number of shared linguistic family tree branches and it is characterised by an increasing categorical level of interior linguistic proximity between 0 and 1 (0.1; 0.25; 0.45; and 0.7), where 1 is a common language.[5] For data limitation, we collapse the Italian language proximity measure equals to 0.25 and 0.45 in a single level of language proximity (level 2). Thus, in our specification, 0.7 is defined as a level 3 if both languages share the highest linguistic tree

---

[5]The linguistic proximity index ranges from 0 to 1 depending on how many levels of the linguistic family tree are shared by both the destination and origin country. To construct the index, we first define a set of increasing weights: the first equal to 0.1 if two languages are related at the most aggregated linguistic tree level, for example Indo-European versus Uralic (Finnish, Estonian, Hungarian); the second is equal to 0.15 if two languages belong to the same second linguistic tree level, for example Germanic versus Slavic languages; the third is equal to 0.20 if two languages belong to the same third linguistic tree level, for example Germanic West versus Germanic North languages; and the fourth is equal to 0.25 if both languages belong to the same fourth level of linguistic family tree, for example Scandinavian West (Icelandic) versus Scandinavian East (Danish, Norwegian and Swedish), German versus English, or ItaloWest (Italian, French, Spanish, Catalan and Portuguese) versus RomanceEast (Romanian).
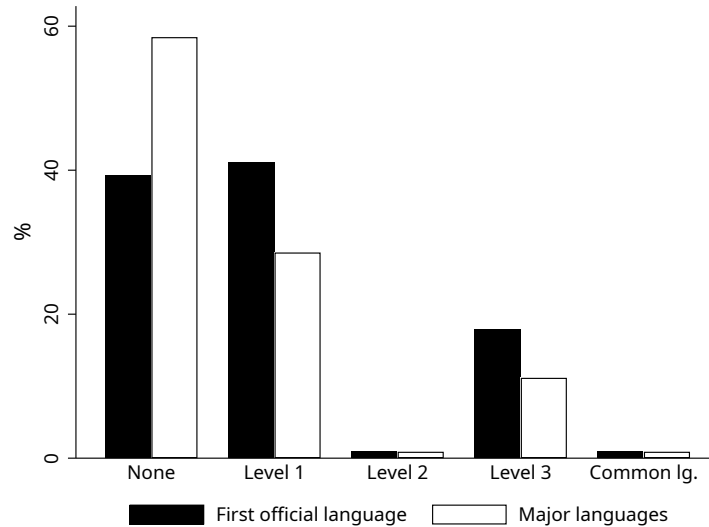
Figure 1: Distribution of country-pairs by Italian linguistic proximity Index

Note: The linguistic index equals 0 if two languages do not belong to any common language family, 0.1 if they are only related at level 1; 0.25 and 0.45 at level 2 and 0.7 at level 3. The index equals 1 if the two countries share a common language.

level.[6] We use two linguistic proximity indexes for immigrants to country-pairs: i.) the first compares the Italian language and the official language of the origin country of immigrants; ii.) the second compares the proximity of the Italian language with all the major local languages spoken in the origin country.

Figure 1 summarizes the distribution of the linguistic proximity between Italy and the origin countries. Around 40% (59% in the case of major languages) of the observations do not share any branch of the linguistic family tree and 41% are only related at the most aggregated level (Level 1). Only around 19% of the countries (12% when we consider all the major languages) show a linguistic proximity of a third level at least.

In according with the previous literature, the lack of fluency in the host country language, which is correlated with the distance from the linguistic root, represents a primary obstacle for immigrants in the economic assimilation process and in finding better job opportunities (Chiswick and Miller, 2003; Dustmann and Fabbri, 2003).[7] This central labour market determinant for the immigrant's success is reinforced in Italy by a large use of temporary resident permits that could have led skilled migrants to decide not to invest in a long search for higher-status jobs or in acquiring the language proficiency relevant to the host country, as they were unlikely to remain in Italy long enough to enjoy a return on these investments (Kalter and Kogan, 2006).[8]

---

[6]Level 1 is the linguistic proximity index which equals to 0.1 if two languages, which include the Italian language, are only related at the most aggregated level of the linguistic tree family.

[7]This disadvantage has been attributed to the difficulties experienced by immigrants in transferring formal schooling, experience and training obtained overseas upon their arrival in the host country (Chiswick and Miller, 2009; Clark and Drinkwater, 2008; Friedberg, 2000).

[8]The absence of any particular rules for regularization of foreign workers can be explained by the fact that, in its initial phase, immigration was considered to be a temporary phenomenon that would not involve large numbers of
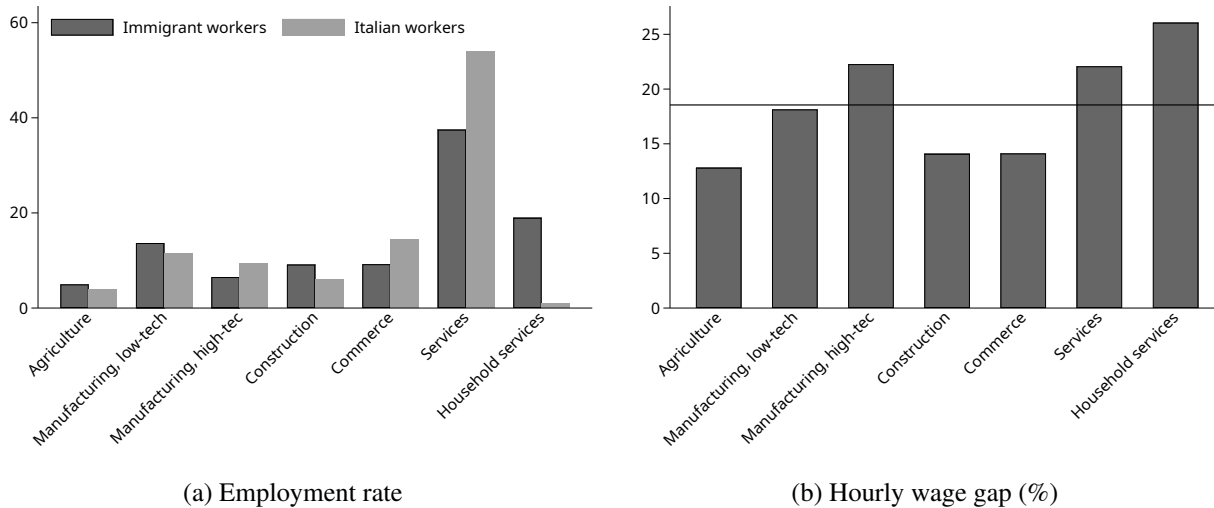
| (a) Employment rate | (b) Hourly wage gap (%) |

Figure 2: Labour market performance by sectoral employment (Mean of years 2015-2016)

2.1. Some stylized facts

Italy constitutes an interesting case study for our research question, which evaluates the effect of language proficiency on the employment and wages of immigrants although, as argued in Barone and Mocetti (2011), the lack of country-specific skills of immigrants and their limited knowledge about the functioning of the labour market could partly explain the differences in labour market performance.

Figure 2 shows some descriptive evidence for employment prevalence and wage gap between immigrant and Italian workers by sector and, within manufacturing, by high-tech/low-tech sectors using the 2015-2016 average of the Italian Labour Force Survey (LFS). Compared to Italian workers, there is a higher employment rate among immigrants in household services (e.g. domestic services and personal assistants) (18.96%), construction (9.13%) and low-skilled sectors (13.67%).[9] In many of these sectors, immigrants benefit from the seasonal nature of some activities, namely tourism and agriculture. Data analyses confirm that newly arrived migrants are mainly absorbed into specific segments of the labour market, those in which natives often refuse to work (Ponzo et al., 2015).

The data also show a stable differential hourly wage in favour of native workers, irrespective of the employment sector. On average, the hourly wage gap was close to 18.6%, estimated by the Italian Institute of Statistics for 2014 (ISTAT, 2016).[10] It is worth noting that in the high-tech manufacturing sector and in household services, we find a higher level of hourly wage gaps (22.08%

---

immigrants, as Italy was seen as merely a stage on the journey towards their final destination, i.e. traditional European immigration countries.

[9]The regularisation implemented by the Law 189/2002 accounted for a significant prevalence of female applicants from Romania, Ukraine, Moldavia, Poland and Ecuador working as domestic help and carers.

[10]The indicator is calculated as: (wages natives-wages immigrants)*100/wages immigrants.

(a) Employment rate

(b) Hourly wage gap (%)
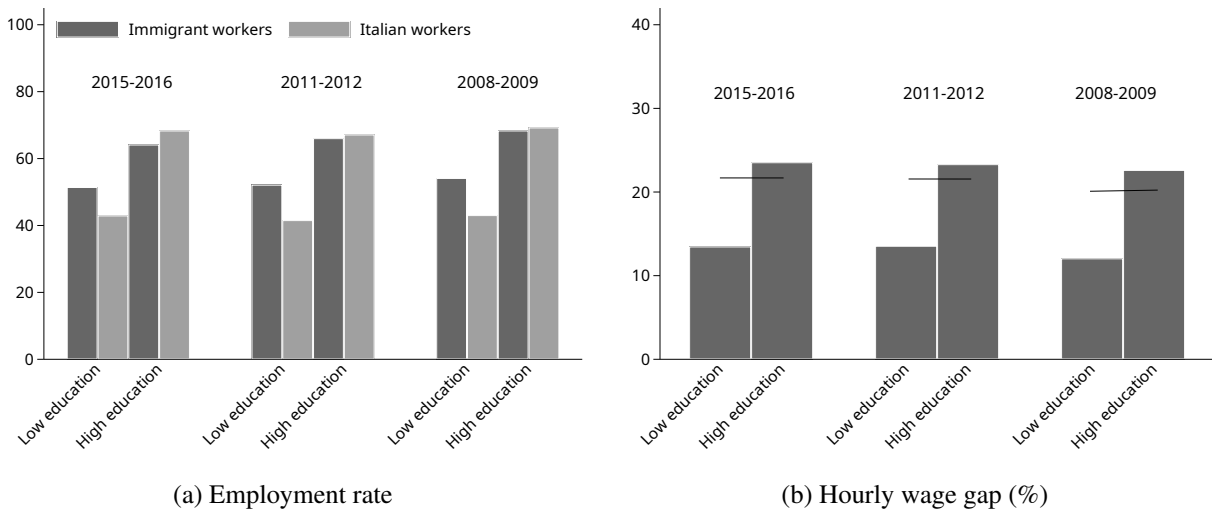
Figure 3: Labour market performance by educational level

(a) Employment rate

(b) Hourly wage gap (%)
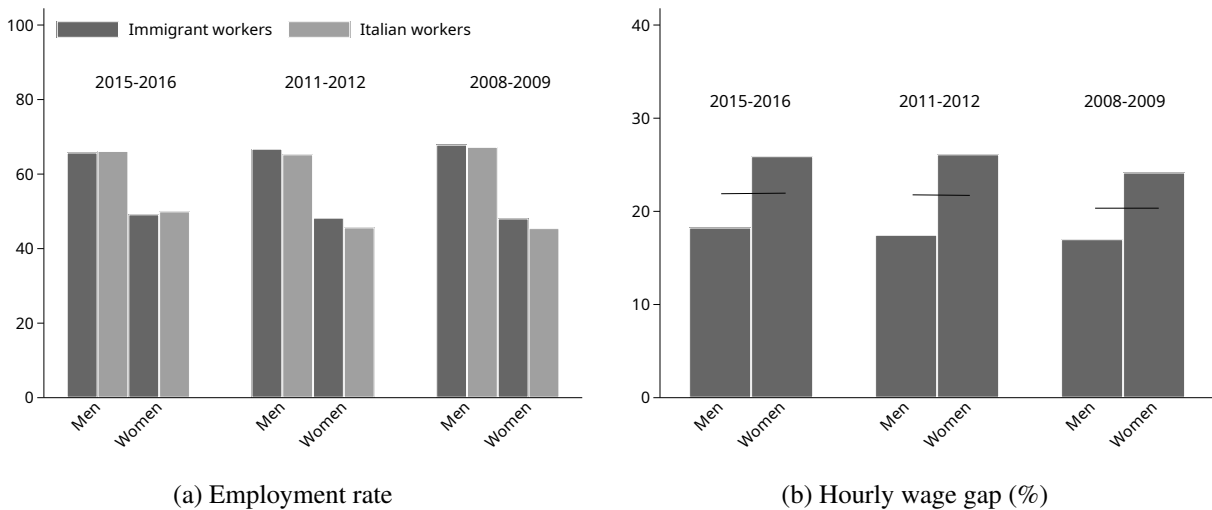
Figure 4: Labour market performance by gender

and 26.06%). In the case of the wage gap for the high-tech manufacturing sector, this may suggest that the tasks and responsibilities performed by immigrant workers, on average, are not completely equivalent to those performed by natives, although they may have equivalent educational backgrounds.

We show in Figure 3 and Figure 4 the differences in the employment rate and hourly wage gaps between immigrants and Italians by educational level and gender. The descriptive statistics concern the years 2008-2009, 2011-2012 and 2015-2016. The first two couples of years correspond to the years the surveys were conducted. The most recent wave shows that Italian workers with high education are more than 10 percentage points above the employment rate mean (68.4% vs. 57.9%). In relative terms, the employment rate for immigrants with high education is above the mean, even if the prevalence decreased 2.5 percentage points comparing the first (2009) and the last years (mean 2015-2016). As expected, Italians with low education had an employment rate of 43% in 2015-2016 compared to 51.5% for immigrants. In addition, the data suggests a mismatch between the potential association between the educational level of immigrants and their economic achievement. Higher educated immigrants have lower hourly wages than Italian low-educated workers, although they are largely higher than the wages of low-educated immigrants.

The observed employment and wage gaps between natives and immigrants suggest some specificity when investigated by gender (Figure 4). The employment rate by gender has remained stable over the years, although it is known that the unemployment rate increased with the downturn in the Italian economy.[11] Data show that the hourly wage gap mean with respect to natives has increased compared to previous years, reaching 16.5% among men and 22.5% among women.

3. Data and variables

3.1. Sources

Two national surveys were used to estimate the effect of Italian language skills on the labour market outcomes of immigrants. The first survey is the CSIFC 2011-2012, published by the IIS in 2014. The reference population of this survey were immigrants permanently or temporarily living in Italy between May 2011 and December 2012.[12] A sample of 12,000 households living in about 800 Italian cities was used for the interviews. The second is the ILCHF 2009, financed by the Italian Minister of Labour and Social Policies and conducted by the IIS in 2009. This survey made use of the methodological framework of the survey on Income and Living Conditions, carried out yearly in 27 EU countries (plus Norway and Iceland), and coordinated by Eurostat (e.g., EU-SILC). The questionnaire, data collection and correction procedures, as implemented and improved by EU-SILC, have been adapted to the specific needs implied by the foreigners' survey.

The final datasets were adjusted for some data issues. First, we only considered immigrants aged 15-64, excluding those who retired in 2011 and 2008, respectively, for the CSIFC and ILCHF. Second, although the survey included information on Italians living in households with immigrants, we excluded them from the successive analysis because they are not representative of the Italian population. Using this strategy, we obtained a CSIFC dataset of 17,298 immigrants, 14,544 of

---

[11]Immigrant men experienced a doubling of unemployment (e.g. from 7.8% in 2009 to 15.1% in 2015-2016).

[12]The interviews were extended until February 2013, increasing the sample for some big cities (e.g., Milan, Rome and Naples).

whom were first-generation immigrants, who we considered in our investigation.[13] The ILCHF dataset included 8,877 first-generation immigrants.

Table 1 lists the composition of samples by groups of immigrants' countries of birth compared with the Population Census (2011) and adjusted for changes in 2012 by official demographic statistics (e.g., the average of resident immigrants between 2011 and 2012). It appears that both surveys have similar group composition as of the 2011 Census. Only the American immigrant group appears to be slightly over-represented in the ILCHF survey (16.8% vs. the 7.9% of the Population Census).

Table 1: Composition of samples by country of birth

Immigrant groups	CSIFC survey (2011-2012)		ILCHF survey (2009)		Pop. Census (2011)	
	N.	%	N.	%	N.	%
F-Y*, Albany, Romania	5649	38.80	2930	33.00	1,613,015	37.40
Other-Europe	2961	20.30	1557	17.50	693,388	16.20
Asia	2118	14.50	1161	13.10	742,994	17.20
China & India	(774)	(5.30)	(490)	(5.50)	(323,221)	(7.50)
Africa	2840	19.50	1733	19.50	918515	21.30
Tunisia & Morocco	(1785)	(12.30)	(1030)	(11.60)	(525,189)	(12.20)
America	976	6.70	1496	16.8	342,718	7.90
Total	14,544		8,877		4,569,317	

Notes: The table compares the percentage of immigrants by country of origin according to the CSIFC survey of 2011-2012, the ILCHF survey of 2009 and the average of the immigrant population, as extracted by the population census of 2011 and population flows in the 2012. Subsamples of immigrants by continents are listed in parentheses. *F-Y: Former Yugoslavia

3.2. Variables

Both datasets provide information about the socio-demographic characteristics of immigrant household members' in the country of origin and in Italy and, of interest for the purposes of our study, Italian language proficiency. In addition, the surveys list interesting labour market outcomes. The CSIFC reports the outcomes of employment of immigrants. Aiming to investigate the effect of a lack of language proficiency on the occupation of immigrants, we focus on immigrants who engage actively in the labour market and consider as the outcome variable the share (in percentage) of the total labour force that is employed. The ILCHF, instead, reports individual monthly wages and the number of hours worked per month, and this information is used to obtain our outcome of interest, "hourly wages" (*in Log*).

Concerning language abilities, for the CSIFC survey, the household member answered the following questions: *"Is it difficult for you to read in Italian?"* and *"Is it difficult for you to speak*

---

[13]Here, second-generation immigrants are defined as those born in Italy to at least one foreign citizen or those born abroad but who completed a cycle of study in Italy.

in Italian?".[14] The answer choices were: often, sometimes, rarely, never. Following Yao and van Ours (2015), we defined a dummy variable that equals one if the individual answered that he/she often or sometimes had problems either in speaking or reading and zero otherwise (rarely, never). A four-modality framework for language proficiency is also used in the ILCHF survey, although it more generally asked respondents to evaluate whether their Italian is *Poor*, *Sufficient*, *Discrete* or *Good*. Also in this case, we obtained a binary outcome for language proficiency (LP) by collecting respondents who answered that they have poor or sufficient Italian language proficiency ($LP = 1$) and those who answered they have discrete or good language abilities ($LP = 0$).

Clearly, the groups of immigrants with less proficiency in Italian differ in terms of observable characteristics from those without language problems. These individual and household differences concern the immigrants' actual and past conditions, if they never studied in the country of origin and the reason they immigrated to Italy. In addition, other control variables may differ. We recorded the most common control variables, such as age, marital status, education level, macroregion and area of residence.

A more complete set of background variables, including family networks and migration plans, are included in the empirical analysis. As the literature on immigration suggests, immigrant networks decrease immigration costs (i.e. information, assimilation or visa costs through reunification policies) so networks may positively influence the likelihood of new migrants finding a job and being more satisfied with their actual job (Massey et al., 1993; Munshi, 2003; Piras, 2019). In this respect, the family network, or the number of close relatives who already live in Italy, may play an important role in explaining the labour market performance of the immigrant population. Much along the same lines, migration plans may also be important for inclusion in the analysis. Indeed, if migrants plan to return home, they will invest less in language, but they may also put less effort into finding a good, well-paid job. Finally, in order to account for individual (unobserved) abilities obtained in the country of origin, which may affect the likelihood of a finding a good job and getting a good wage in the destination country, we include a proxy for immigrants who have never worked in the country of origin.

Our dataset is completed by interacting the variables 'age at arrival' and 'Italian language distance', which will be used as an instrumental variable in the IV estimations. Age at arrival is calculated as the difference between the year in which the person immigrated to Italy and their year of birth. Instead, the Italian language distance exploits information about the levels of the linguistic family tree that the language of the source country share with the Italian language, taking the first official language in each country for the baseline estimations, and the major local language spoken in each country as a robustness check. Thus, we define a linguistic distance index using the complementary classification of the distribution of country-pairs by the Italian linguistic proximity index summarised in a binary variable: 1 when a linguistic distance is at least of second level (level 2) of the linguistic family tree compared to the Italian language, which identifies a greater linguistic distance, and 0 when there is non-significant linguistic distance.

---

[14]The data of the CSIFC survey has been accessed and the analyses including language proficiency have been carried out by the authors at the Istat's Laboratorio per l'Analisi dei Dati ELEmentari (ADELE) in Rome.

3.3. Descriptive statistics and main bivariate relationships

Table 2 shows the descriptive statistics for the outcomes of interest included in our analysis by gender and Italian language proficiency. Immigrant women have a lower probability of being employed than men. When performance is measured by the share of employed in labour forces, the probability of women with a higher linguistic proficiency in employment compared to men in the same linguistic category is 5 percentage points lower (86.5% for men and 81.3% for women). Within gender groups, employment differences can be also investigated for women with more or less linguistic proficiency. The probability of a woman with weak linguistic proficiency being employed is reduced by a further 4 percentage points (from 81.3% to 76.9%).

ILCHF statistics for the hourly wages (in log) are also reported inTable 2. Our sample restriction to first-generation immigrants may underestimate mean wages with respect to labour force statistics. However, if we calculate the gender pay gap of immigrants by language proficiency using the ILCHF, we find that those with weak proficiency in the Italian language have a gender pay gap of 18.2%, while this is reduced to 15.7% for the sub-sample of immigrants with good proficiency in the Italian language.[15] Appendix A lists descriptive statistics for Italian language proficiency and for all covariates by gender and language proficiency.

Table 2: Descriptive statistics of labour market outcomes

	CSIFC survey (2011-2012)				ILCHF survey (2009)			
	Men		Women		Men		Women	
	Language proficiency							
	Good	Weak	Good	Weak	Good	Weak	Good	Weak
Share of employed in labour forces (%)	86.5	83.1	81.3	76.9				
Hourly wages (in euro)					8.67	7.61	7.10	6.61
Log hourly wages					2.16	2.03	1.96	1.86

Notes: In this table, we compare the labour market outcomes of the CSIFC 2011-2012 and the ILCHF 2009. In the CSIFC survey (2011-2012), Italian language proficiency (*LP*) is defined as a dummy variable, which equals one if the individual answered that he/she often or sometimes had difficulty either speaking or reading Italian (Weak) and zero for rarely and never modalities(Good). In the ILCHF survey, Italian language proficiency (*LP*) is defined as a dummy variable, which equals one if the individual answered that he/she had poor or sufficient language abilities (Weak) and zero for those with discrete or good language abilities (Good).

Now, we delve deeper into the descriptive statistics of the variable that will be used below as an instrument in the IV approach. The age at arrival of the first-generation immigrants in both surveys appears to be different by gender, with a greater probability density for younger men, which is overcome by women after age 35 (Appendix B, Figure B1). This evidence confirms the argument that women generally arrive later in host countries mainly for family reunification, while men usually emigrate earlier for work opportunities. The increasing demand for immigrant women to work in the household service sector as caregivers, and often employed without permits to stay

---

[15]The weighted estimate is about 17%.

in Italy, generated in illegal immigrants the expectation for successive regularisations (Salmasi and Pieroni, 2015).

Figure 5 shows the probability of having weak language proficiency based on age classes at arrival in both surveys. We distinguish between immigrants with a greater linguistic distance and immigrants with a lesser linguistic distance using the first official language index as a comparison. This figure illustrates that increasing age at arrival trends are steeper for the group of immigrants coming from countries characterised by a greater language distance. It suggests that the interaction variable may be a good instrument for estimating the causal relationship between language proficiency and labour market outcomes for the Italian case, thereby reducing the bias due to the differences in effort required to acquire commands in Italian.

Clearly, the accuracy of the instrument 'age at arrival' interacted with 'more linguistic distance' may be affected by the heterogeneous composition of immigrants on the probability of having weak proficiency in the Italian language. As argued by Hermansen (2017), different groups of immigrants may also follow different patterns, which may be hidden if we use the aggregate variable of age at arrival of all immigrants as an instrument to obtain estimates for these groups. Figure B2, in Appendix B, reports the probability of having language problems for immigrants with a greater linguistic distance. The Figure suggests that the different immigrant groups mimic the patterns shown in Figure 5 of a positive relationship between age at arrival interacted with 'more linguistic distance' and Italian language proficiency.

(a) CSIFC survey (2011-2012)   (b) ILCHF survey (2009)
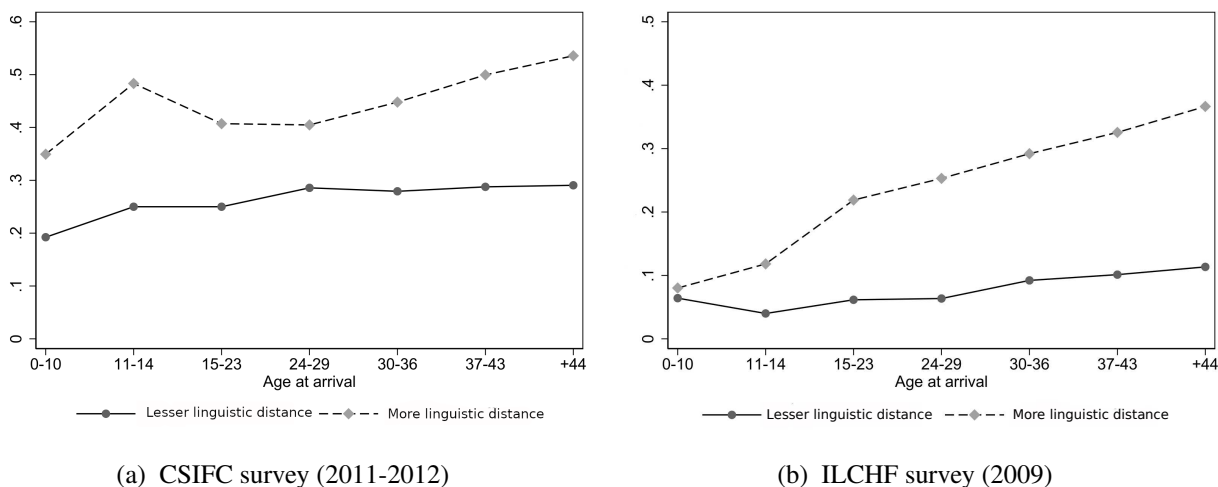
Figure 5: Probability of having weak language proficiency and age at arrival for immigrants with lesser or more linguistic distance

Notes: Estimation is obtained after conditioning for age dummies.

4. Estimating the effect of language proficiency on labour market outcomes

4.1. OLS estimates

The main premise of our analysis is that the labour market is less accessible to immigrants and less profitable based on low proficiency in the Italian language. This section provides evidence

to support this premise. The simplest way to formalise the relationship between weak language proficiency and the labour market performance of immigrants is through a regression model:

$$y_i = \alpha_1 + \alpha_2 LP_i + X'\alpha_3 + \varepsilon_i \qquad (1)$$

where i denotes each immigrant in the dataset, y_i denotes an indicator of employment, job satisfaction or hourly wage performance, LP_i is a dummy variable that is equal to one when individual i has weak proficiency in the Italian language and X is the vector that includes all individual and household characteristics listed in Appendix B. The parameter of interest α_2 should measure how much less or more likely labour market outcomes are for immigrants with Italian language proficiency with respect to immigrants with good proficiency in the Italian language.

Table 3 illustrates the ordinary least squares (OLS) estimated coefficients for the two surveys and for men and women, separately. The coefficients of the share of employment in the CSIFC 2011-2012 survey are of the expected sign, although they are, in the large part, close to zero. Even when α_2 is significant at the usual 5%level, the magnitude of the differences in the employment due to language proficiency is up to 4 percentage points. More importantly, the inclusion of the conditional variables corrects for the unobserved heterogeneity. For example, in the sample of women with language problems, a reduction in the upward bias is evident when passing from the unconditional differences to the point estimated reported in Table 2 (i.e. the difference is about 3 percentage points). In addition, the conditional OLS point estimates for hourly wages suggest that immigrants' weak proficiency in the Italian language causes a 2.8% reduction, a result that appears to be driven by male immigrants. Overall, these correlations suggest that language abilities in destination country may play a role in determining the labour market outcomes of immigrants.

Table 3: Estimation results, conditional OLS

	Total sample	Women	Men
CSIFC survey (2011-2012)			
Share of employed in labour forces	-0.030***	-0.040***	-0.019*
	(0.008)	(0.013)	(0.010)
Observations	*12,114*	*5,927*	*6,184*
ILCHF survey (2009)			
Log hourly wages	-0.028**	-0.018	-0.030**
	(0.011)	(0.017)	(0.014)
Observations	*4,228*	*1,988*	*2,240*

Notes: Robust clustered standard errors are in parentheses. The asterisks stand for the *p-value* significance levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

4.2. IV estimates

Estimates from equation 1 may not identify the causal effect of weak language proficiency on labour performance due to other and the most important sources of endogeneity. For example,

self-reported language skill measures are subject to substantial measurement error. Interviewed immigrants generally tend to over-estimate their language abilities either due to their incomplete knowledge concerning the language proficiency of the destination country or due to a positive propensity to show that they are integrated within the immigration country through their language skills. This propensity leads to downward bias, which is alleviated when the language evaluation is carried out by the interviewer directly (Dustmann and van Soest, 2001).

To deal with this main source of bias, we use an IV estimator. As discussed above, a typical IV proposed in the literature is age at arrival in the host country (Bleakley and Chin, 2004, 2010; Guven and Islam, 2015) or age at arrival interacted with no linguistic command in childhood for immigrants in host countries (Yao and van Ours, 2015). We follow the latter strategy and adapt the instrument to less common languages than Italian to satisfy the exclusion restrictions of the model (1). In particular, we propose interacting age at arrival with a linguistic distance index using language speaking origin country to Italian language, as argued in Section 2, since immigrants coming from countries with languages that are more similar to Italian may need relatively less language training or less effort to learn Italian, compared to those whose native language is more distant from Italian. The main identifying assumption in using this interacted variable as an instrument should be that labour market outcomes are affected only through the costs of language acquisition of Italian but have no further direct or indirect effects on the labour market. In the baseline estimations, we use the immigrants' first official language in the country-pairs classification of the linguistic distance from Italian, whereas robustness checks are obtained by comparing the main language and testing the violation of the assumption when two types of immigrants experience different assimilation trajectories.

The results of the IV estimates for employment in (Table 4) are larger and statistically significant with respect to the OLS estimates presented in Table 3. We find that immigrants with weak language proficiency had a significantly decreased employment rate, with a magnitude of the point estimates of around 28 percentage points. This suggests that measurement errors accounting for Italian language abilities in employment are very important.

Furthermore, given the representativeness of the sample, we apply the model to the subsamples of men and women and ascertain whether weak immigrant language proficiency affected the employment rate differently. Columns 2-3 of Table 4 show that the results are very close when the two samples are analysed. The table also reports the F-test for the relevance of the IV, age at arrival interacted with linguistic distance. For all estimates the F-statistics are larger than the rule of thumb (e.g., F-statistics exceeding 10), indicating that our estimates do not suffer from weak instruments.[16]

By extracting the share of employed in labour forces from the ILCHF survey (2009) as an outcome, we perform an additional estimation using the correspondent Italian language proficiency described in Section 3. The results of the point estimates in reducing the employment rate for immigrants with weak language proficiency in Italian [(β=-0.205; s.e.=0.097)] by 21 percentage points are slightly smaller in magnitude, although within the confidence intervals of those obtained from the CSIFC survey (2011-2012).

---

[16]We also perform a Durbin-Wu-Hausman test for endogeneity of language problems, where significant F-statistics suggest that the language indicator is endogenous. Based on the data, we do not reject the null hypothesis of exogeneity for all IV estimates in this paper, although to save space we do not report it extensively in the tables. All estimates and tests are available from the corresponding author upon request.

Table 4: Estimation results by IV

	Total sample	Women	Men
CSIFC survey (2011-2012)			
Employment rate	-0.276***	-0.287***	-0.267***
	(0.049)	(0.078)	(0.058)
Weak instrument test	324.676⋆	147.878⋆	211.286⋆
Observations	*12,114*	*5,927*	*6,184*
ILCHF survey (2009)			
Log hourly wages	-0.345***	-0.311**	-0.350**
	(0.109)	(0.145)	(0.150)
Weak instrument test	62.831⋆	37.555⋆	30.622⋆
Observations	*4,203*	*1,975*	*2,228*

Notes: Language problems are defined as having either speaking or reading problems. Robust clustered standard errors are in parentheses. The asterisks stand for the *p-value* significance levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$). We also report the relevance of the IV by F-statistics under the testing hypothesis of a weak instrument. The star indicates the rejection of the weak instrument associated with the F-statistic that exceeds 10 (Angrist and Pischke, 2009).

In the second block of Table 4, we employ the same model to evaluate weak proficiency in the Italian language on hourly wages. Thus, we are effectively asking how much hourly wages decrease when an immigrant has difficulty speaking or reading Italian. The results show that the magnitude of these effects in decreasing hourly wages is of about 34.5% and that, consistent with our expectations, there is a large downward bias induced by measurement errors associated with Italian language abilities. Regarding gender differences, the point estimates show that the wages of immigrants with language problems are reduced by 35% for men and by 31% for women.

4.2.1. Robustness

In this section, we present several robustness checks for our findings. Panel a of Table 5 shows the point estimates of weak proficiency in the Italian language as an exogenous variable (i.e. not as an instrument) in the OLS labour market outcome equations. This specification controls for the direct effects of age at arrival for more distant immigrant language on the relevant labour market variables, which are independent from the Italian language proficiency of immigrants (i.e., it is assumed that the direct non-language effects are the same between immigrants who grew up in countries with distant or not linguistic roots from Italian language). Although the estimated parameters of the model (1) may be affected by the inclusion of the age assimilation of the host language as a right-hand side variable, we do not find any statistically significant differences with the point estimates presented in Table 3.

A second robustness check is performed by introducing age at arrival as a supplementary instrument in the labour market equations. As shown Figure 5, we cannot rule out that early arrival in Italy may contribute to language proficiency for immigrants with language proximity, but we

Table 5: Robustness checks

	Employment rate			Log hourly wages		
	Total sample	Women	Men	Total sample	Women	Men
a) Age at arrival * Imm. ling.distance						
as a right-hand side variable	-0.022***	-0.037***	-0.010	-0.021*	-0.012	-0.025*
	(0.008)	(0.013)	(0.010)	(0.011)	(0.018)	(0.014)
Observations	*12,114*	*5,927*	*6,187*	*4,203*	*1,975*	*2,228*
b) IV estimates using a larger						
index of language proximity	-0.241***	-0.243***	-0.219***	-0.357***	-0.246**	-0.390**
	(0.057)	(0.087)	(0.071)	(0.121)	(0.129)	(0.197)
Weak instrument test	221.220★	120.298★	133.163★	44.916★	43.485★	13.425★
Observations	*12,114*	*5,927*	*6,184*	*4,203*	*1,975*	*2,228*
c) IV estimates using						
an additional instrument	-0.256***	-0.262***	-0.242***	-0.321***	-0.320*	-0.329**
	(0.044)	(0.072)	(0.052)	(0.073)	(0.106)	(0.097)
Cragg-Donald F *statistic*	229.208	98.092	132.805	54.798	31.882	25.940
Kleibergen-Paap $F-statistic$	206.344	94.692	127.030	49.112	27.958	27.053
Kleibergen-Paap $LM statistic$	388.467	194.718	234.331	94.832	55.408	53.015
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hansen J statistic	3.797	1.652	2.259	3.862	1.625	2.264
	(0.062)	(0.199)	(0.112)	(0.058)	(0.202)	(0.132)
Observations	*11,997*	*5,887*	*6,110*	*4,203*	*1,975*	*2,228*
d) V estimates restricted						
to reading problems	-0.284***	-0.263***	-0.269***			
	(0.048)	(0.076)	(0.058)			
Weak instrument test	380.238★	169.413★	216.915★			
Observations	*12,114*	*5,927*	*6,187*			

Notes: Language problems are defined as having either speaking or reading problems. Robust clustered standard errors are in parentheses. The asterisks stand for the *p-value* significance levels (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$). We also report the relevance of the IV by $F-statistics$ under the testing hypothesis of a weak instrument. The star indicates the rejection of the weak instrument associated with the $F-statistic$ that exceeds 10 (Angrist and Pischke, 2009).

remove the hypothesis that there may be an age at arrival effect on language proficiency for immigrants who grew up speaking a distant language from Italian. Panel b of Table 5 shows the point estimates of the effect of weak proficiency in the Italian language on labour market outcomes by using two instruments. There is a no significant difference with respect to the baseline estimates. Weak proficiency in the Italian language reduces employment probability by about 25 percentage points for both female and male immigrants and hourly wages by 33%. Adding an additional instrument yields robust and consistent coefficients, within the confidence interval incluses in the benchmark estimates.

A third robustness check is performed by extending the index of language distance in the instrumental variable, which considers the major spoken languages in the origin countries of immigrants. However, the variability in some countries where there is more than one important language spoken may affect the magnitude of the point estimates of the model (1). The estimated coefficients confirm the magnitude of the effects of weak proficiency in the Italian language on employment and hourly wages, even if the point estimates of job satisfaction outcomes appear to be reduced, especially for men, with respect the baseline estimations.

In addition, we use the CSIFC survey (2011-2012) to investigate whether restricting immigrants' problems with the Italian language to reading led to different point estimates on the em-

ployment. We conducted this robustness check because the descriptive statistics showed that the percentage of immigrants with reading problems was significantly higher than the percentage of immigrants with speaking problems. We still show that the point estimates of the effect of weak proficiency in reading Italian on labour market performance do not evidence significant differences from the baseline estimates.

4.3. A focus on wage disparities of immigrants

In this section, we use the baseline IV coefficients to measure the contribution of immigrants' language problems in explaining the hourly wage variability of people residing in Italy. While the ILCHF 2009 survey may be merged with the IT-SILC to extract the wages of the Italian people conditional to the presence or absence of language problems, we have a lack of immigrant representativeness in the IT-SILC for the successive years. This leads to us developing a different estimation strategy, integrating the data and point estimates of the ILCHF 2009 survey with data from the Italian labour force surveys. Sections a) and b) of Table 6 summarise this information. We assume that the proportions of immigrants with poor (η_1) or good language (η_2) proficiency are approximately constant over time. We then calculate the different wages of immigrants using the estimated parameter α_2 in Table 4 (absolute term) and exploiting the mean hourly wage information (HW_{mean}) of immigrants (and Italians) from the LFS.[17]

We report the results related to immigrant wage estimates for the two-language proficiency groups in sub-section c) of Table 6. In particular, column 2 of this table shows that immigrants with weak proficiency in the Italian language have a mean wage of around 6.14 euros per hour of work. This increases for immigrants with good proficiency in the Italian language to 8.25 euros per hour of work, suggesting that immigrants with poor Italian language abilities are severely affected in terms of economic opportunities. The immigrants' wage disparities are emphasised by gender (columns 3 and 4). Estimated gains for immigrant women are reduced to 5.39 euros per hour of work against 6.55 euros per hour of work for men. In the successive columns, we also report the scenarios for the years 2011-2012 and 2015-2016. We found that the patterns are fully consistent with the scenario described in the 2009 sample, which confirms that Italian language proficiency represents one of the main sources of hourly wage inequalities between immigrants.

In the last two lines of Table 6, we list the contribution of a lack of language proficiency in immigrants to explain the hourly wage variability of people residing in Italy. We estimate it by separating the "within" contribution to the hourly wage disparity of immigrants determined by poor language proficiency with respect to other factors affecting hourly wage differences between immigrants and natives (i.e. the "between" contribution to the hourly wage variability of residents in Italy). The results indicate that in 2009 about 70% of the hourly wage disparity could be ascribed to poor proficiency in the Italian language, while other unobserved channels could explain the remaining 30% of the hourly wage differences between immigrants with good proficiency in Italian and Italian workers. The contribution of Italian language proficiency to the hourly wage disparity of immigrants appears to be mostly driven by men (78%). The findings are not surprising since, for example, in the German context using practical reading and writing tests, Himmler and Jäckle

---

[17]Formally, we obtain the wage estimates using the following equation: $W_{mean} = \eta_1 * W_{LP_0} + \eta_2 * W_{LP_1}$. We substitute the parameters from column 2 of Table 6, such that $0.36 * W_{LP_1} + 0.64 * (1.23 * W_{LP_1}) = 1019$. This also implies that $W_{LP_1} * (0.36 + 0.64 * 1.345) = 7.49$ and $W_{LP_1} = 6.14$. The wage for immigrants with good language proficiency is $W_{LP_0} = 1.345 * W_{LP_1} = 8.25$.

(2018) show that the migrant-native employment and (hourly) wage gaps disappear largely when language proficiency is taken into account.

A potential mechanism for explaining the differences in language skill effects can be attributed to the selection of immigrants into different occupations. For example, Berman et al. (2003) show that Hebrew language proficiency reduces the wage differentials (compared to native speakers) of programmers and technicians in Israel, while this does not hold for construction workers and gas station attendants. These results are also confirmed in Italy when the effects of language proficiency on wages are compared within immigrants' occupation by gender. As shown in Appendix C, immigrant women in Italy are mainly employed in the domestic sector as household workers (i.e. unskilled occupation), and receive the lowest wages, while manufacturing and construction mainly employ immigrant men (56%). This gender structure of employment for women smoothes the positive effect of a good command of Italian on wages and hourly wages and, in turn, limits the potential reduction of wage differentials compared to native speakers.

These differences in the impact of Italian language proficiency seem to decrease in the successive years under the financial crisis shock that, in turn, affected the real economy and made the expectation of employment and the careers of immigrants less important. Given the evidence that the composition by country of origin and gender of immigrants did not change recently (Appendix D), the mechanism based on the selection into different occupations may at least partly explain why a poor command of Italian impacts hourly wages less. Data from the LFS (various years) show that the long- term effects of the 2008 financial crisis reduced the proportion of immigrant men employed in the manufacturing sectors, where a good command of the host country language is required, particularly in the high-tech sector. This increased the mean wage differences between immigrants and natives, as a consequence of a complementary lack between language proficiency and further human capital dimensions, with a reduction in the wage convergence (Lehmer and Ludsteck, 2015). On the other hand, immigrant women experienced more stable household service employment, which supports the prediction that having a good command of Italian may not be a sufficient condition for immigrant women to reduce disparities with respect to natives.[18]

---

[18]Chiswick and Miller (2010) found equivalent results for workers with poor English skills employed in jobs that have very low English-language requirements in England.

Table 6: Conditional wage of immigrants by language proficiency and wage disparity: estimates

		Estimates 2009			Scenario 2011-2012			Scenario 2015-2016		
		All	Women	Men	All	Women	Men	All	Women	Men
a) ILCHF survey 2009										
Language proficiency (α_2)		(0.345)	(0.311)	(0.351)	(0.345)	(0.311)	(0.351)	(0.345)	(0.311)	(0.351)
Proportion of immigrants with weak language proficiency(η_1)		0.36	0.362	0.358	0.36	0.362	0.358	0.36	0.362	0.358
Proportion of immigrants with good language proficiency (η_2)		0.64	0.638	0.642	0.64	0.638	0.642	0.64	0.638	0.642
b) Labour Force surveys: Years 2009, 2011-2012 and 2015-2016										
Hourly wages of immigrant workers		7.49	6.46	8.03	7.53	6.66	8.17	7.56	6.80	8.24
Hourly wages of Italian workers		9.18	8.53	9.57	9.50	8.97	9.81	9.71	9.17	10.07
c) Conditional estimates for immigrant wage										
Hourly wages of immigrants with weak language proficiency	*	6.14	5.39	6.55	6.17	5.56	6.67	6.19	5.68	6.73
Hourly wages of immigrants with good language proficiency	*	8.25	7.07	8.85	8.30	7.29	9.01	8.33	7.44	9.09
Contribution of weak language proficiency to hourly wage disparity (%)	**	0.70	0.53	0.76	0.64	0.51	0.74	0.61	0.50	0.70
Other factors contributing to hourly wage disparity (%)		0.30	0.47	0.24	0.36	0.49	0.26	0.39	0.50	0.30

Notes: Hourly wages of immigrants with weak or good language proficiency are obtained as described in footnote 16. ** The contribution of Italian language problems to hourly wage disparities is obtained as a share of within contribution due to the weak Italian language proficiency effects to disparity of immigrants' hourly wages with respect to the whole disparity of hourly wages.

5. Conclusion

Fluency in the language of the destination country can facilitate the transfer of migrants' skills to the new labour market. We first argued that migrants with a low proficiency in Italian had particularly reduced opportunities in the labour market and documented that they were significantly affected by decreasing employment and increasing workplace unsatisfaction. We then showed that immigrants with Italian linguistic problems reduced their expected hourly wages relative to immigrants with good Italian linguistic proficiency, suggesting that the tightening of the linguistic abilities of immigrants in host countries may significantly constrain the adverse impact on the individuals that were affected. In particular, we found that Italian language problems explained more than half of the hourly wage inequalities between natives and immigrants, particularly for men. An important implication of this analysis for immigration policy is that immigrants proficient in Italian improve the opportunities in the economic, social and civic life of their new country.

The study is not without drawbacks. One such limitation is that we were unable to investigate the influence of different types of residence permits held by immigrant respondents given the absence of information about immigrants possessing a permanent residence permit, or lack thereof. In addition, we do not have data on undocumented immigrants, who represent about 10% of the immigrant population. This implies that for undocumented immigrants facing increasing difficulty in realising their expectations in the host country, the lack of language skills exacerbates the difficulty of obtaining (good) employment. Therefore, our results most likely underestimate the potentially larger differences in labour market outcomes between immigrants and natives in the general population. Taking these limitations into account, our results are certainly a conservative test of the impact of language proficiency on employment and wage disadvantages of immigrants in Italy.

Appendix A   Descriptive statistics: main predictors and covariates

| | | CSIFC survey (2014) | | | | ILCHF survey (2009) | | | |
| | | Women | | Men | | Women | | Men | |
Variables	Any language problems	No	Yes	No	Yes	No	Yes	No	Yes
Italian Language Proficiency									
Difficulty in speaking Italian (often or sometimes)			24.46		27.57				
Difficulty in reading Italian (often or sometimes)			37.19		33.02				
Language proficiency (Poor)							8.02		8.20
Language proficiency (Sufficient)							27.87		28.05
Language proficiency (LP) (Modality 1=weak)			24.46		27.57		35.80		36.20
Covariates									
Age		38.669	38.650	37.569	39.403	36.481	36.679	35.895	36.279
Marital status	Single	0.274	0.191	0.335	0.344	0.288	0.266	0.334	0.393
	Married	0.482	0.617	0.554	0.418	0.551	0.603	0.623	0.557
	Divorced	0.197	0.144	0.107	0.192	0.104	0.070	0.033	0.039
	Widowed	0.047	0.049	0.004	0.006	0.056	0.062	0.009	0.010
Type of household	Living alone	0.256	0.223	0.245	0.406				
	Couples with children	0.464	0.519	0.558	0.394				
	Couples without children	0.162	0.175	0.143	0.146				
	Father with children	0.002	0.003	0.022	0.037				
	Mother with children	0.115	0.080	0.032	0.017				
Family size	One					0.230	0.245	0.184	0.279
	Two					0.218	0.177	0.167	0.173
	Three					0.222	0.194	0.232	0.188
	Four					0.196	0.219	0.229	0.186
	>Four					0.133	0.165	0.188	0.174
Household relation	Head					0.615	0.483	0.744	0.734
	Spouse					0.287	0.407	0.074	0.072
	Head parents					0.017	0.023	0.008	0.010
	Head sons					0.048	0.043	0.094	0.075
	Head brother					0.025	0.031	0.056	0.074
	Other relative					0.008	0.013	0.024	0.034
Education	No education	0.038	0.187	0.062	0.213	0.029	0.107	0.048	0.154
	Primary	0.038	0.103	0.058	0.107	0.070	0.149	0.088	0.177
	Lower secondary	0.245	0.348	0.333	0.354	0.259	0.350	0.331	0.381
	Upper secondary	0.531	0.316	0.462	0.302	0.503	0.330	0.439	0.258
	Higher education	0.148	0.046	0.085	0.023	0.139	0.063	0.094	0.030
Number of children		1.229	1.552	1.168	1.1200				
Foreign degrees						0.940	0.975	0.922	0.970
Ever employed in the country of birth		0.483	0.714	0.438	0.402				
First time job seeker		0.024	0.030	0.024	0.018				
Childcare *		0.135	0.225	0.001	0.003	0.188	0.260	0.010	0.011
Domestic worker		0.217	0.168	0.027	0.055	0.228	0.260	0.027	0.055
Network		0.133	0.114	0.140	0.120	0.135	0.131	0.142	0.134
Family network		0.507	0.502	0.631	0.575				
Attending It course		0.019	0.133	0.030	0.039	0.080	0.030	0.075	0.027
Area of residence	Metropolitan areas	0.269	0.214	0.253	0.223				
	Municipalities (more 10.000)	0.215	0.280	0.246	0.240				
	Municipalities (less 10.000)	0.516	0.506	0.514	0.501				
Area of residence	Big					0.424	0.411	0.401	0.415
	Intermediate					0.417	0.433	0.453	0.438
	Small					0.159	0.157	0.146	0.147
Macro-region	North	0.356	0.338	0.501	0.537	0.374	0.321	0.430	0.328
	Centre	0.191	0.117	0.188	0.122	0.306	0.238	0.305	0.250
	South	0.453	0.545	0.403	0.616	0.319	0.441	0.265	0.422
Reason to immigrate	Absence of work	0.314	0.284	0.445	0.501				
	A higher income	0.164	0.128	0.202	0.256				
	To improve quality of life	0.122	0.147	0.121	0.105				
	Family reunion	0.292	0.385	0.113	0.081				
	War	0.009	0.008	0.029	0.016				
	Other	0.099	0.048	0.090	0.040				
Reason to immigrate	Work					0.551	0.539	0.790	0.892
	Family					0.392	0.432	0.151	0.082
	To study					0.018	0.004	0.018	0.005
	Other					0.039	0.025	0.041	0.022

Table continued

Variables	Any language problems	CSIFC survey (2014)				ILCHF survey (2009)			
		Women		Men		Women		Men	
		No	Yes	No	Yes	No	Yes	No	Yes
Region of birth	F-Y*, Albany, Romania	0.414	0.310	0.463	0.285	0.425	0.317	0.453	0.305
	Other Europe	0.292	0.146	0.114	0.086	0.291	0.198	0.110	0.066
	Asia	0.080	0.263	0.128	0.303	0.074	0.202	0.105	0.257
	Africa	0.115	0.247	0.240	0.305	0.111	0.221	0.268	0.337
	America	0.100	0.032	0.056	0.020	0.098	0.063	0.064	0.035
Migration plan	Remain in Italy	0.352	0.453	0.349	0.374				
	Origin country	0.377	0.291	0.341	0.346				
	Other country	0.271	0.256	0.310	0.280				
Remain to live in Italy						0.760	0.747	0.716	0.716
Sector	Agriculture					0.034	0.060	0.074	0.153
	Manufacture					0.115	0.072	0.228	0.214
	General services					0.003	0.002	0.008	0.010
	Construction					0.015	0.012	0.272	0.217
	Services					0.073	0.076	0.085	0.136
	Public services					0.185	0.143	0.164	0.150
	Others					0.574	0.635	0.109	0.121
Part time						0.146	0.131	0.045	0.075
Firm size	< 10					0.715	0.822	0.540	0.617
	11 − 15					0.053	0.046	0.120	0.108
	10 − 20					0.025	0.022	0.049	0.033
	21 − 50					0.062	0.029	0.101	0.073
	> 50					0.114	0.038	0.145	0.103
Contract	Permanent					0.692	0.582	0.737	0.600
	Temporary					0.208	0.218	0.205	0.270
	No contract					0.099	0.200	0.058	0.130

Notes: The Table compares the the Conditions and Social Integration of Foreign Citizens survey (CSIFC survey 2011-2012) and immigrants from the Income and Living Conditions of Households (ILCHF survey 2009). *F-Y: Former Yugoslavia

Appendix B   Instrumental variable, age at arrival
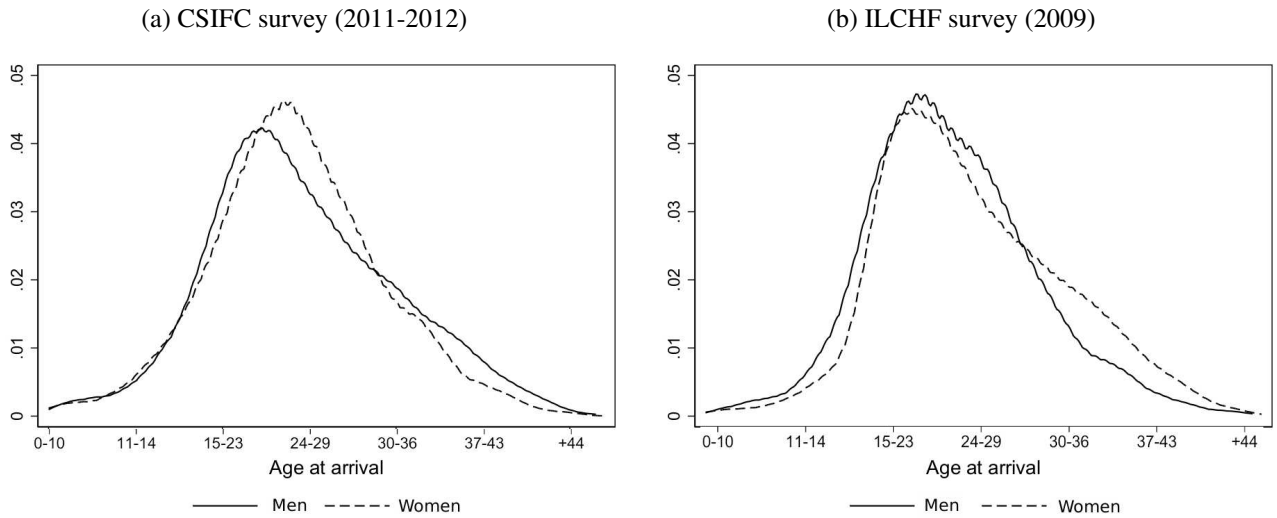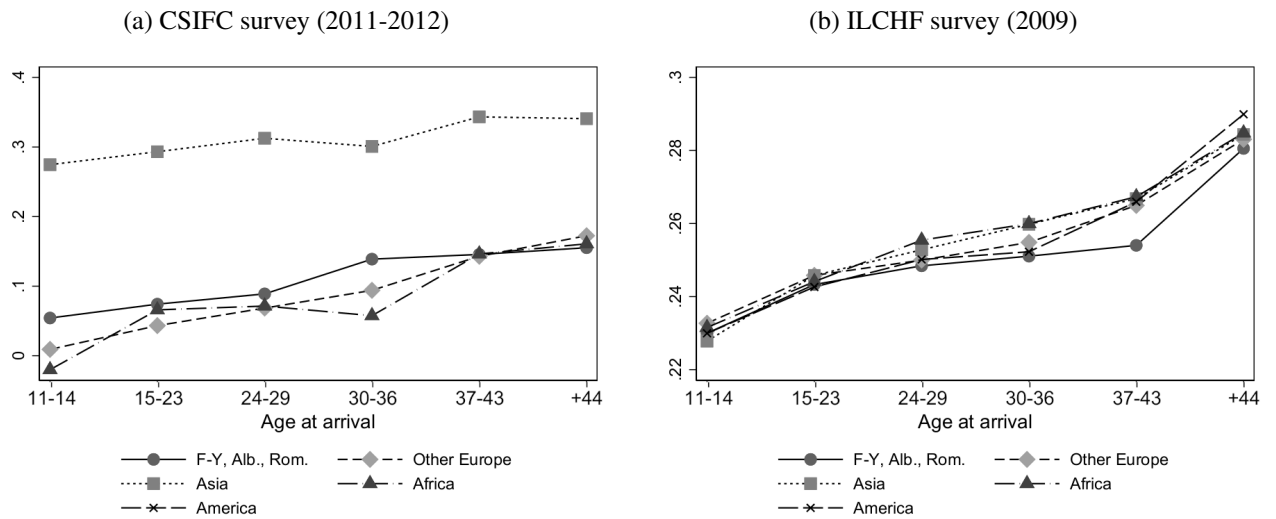
Figure B1: Kernel density plots of age at arrival

(a) CSIFC survey (2011-2012)   (b) ILCHF survey (2009)

Figure B2: Probability of having language problems based on age at arrival for immigrants with the most linguistic distance, heterogeneity

(a) CSIFC survey (2011-2012)   (b) ILCHF survey (2009)

Appendix C Employment composition (in percentage) by sector and gender

	Men		Women		Total	
	Italian	Immigrant	Italian	Immigrant	Italian	Immigrant
Agriculture	2.3	7.1	1.7	2.9	2.0	5.2
Manufacturing	33.2	33.5	18.6	11.5	26.8	23.8
Low-tech	*23.7*	*31.5*	*13.2*	*9.9*	*19.1*	*21.8*
High-tech	*9.5*	*2.0*	*5.4*	*1.6*	*7.7*	*2.0*
Construction	9.5	24.6	1.8	1.3	6.1	14.4
Commerce	20.8	18.7	19.2	16.3	20.1	17.6
Services	32.1	7.3	56.1	15.6	42.7	11.0
Household services	2.0	9.0	2.7	52.4	2.3	28.1

Source: Italian Labour Force Survey (LFS), year 2009.

Appendix D Composition by country of origin and gender of immigrants in Italy, 2009-2015

	2009		2010		2011		2012		2013		2014		2015	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
F-Y*, Albany, Romania	34.80	33.43	34.65	33.72	34.03	33.44	33.65	33.46	33.28	33.91	32.98	34.22	33.03	34.69
Other-Europe	14.01	24.73	13.99	24.56	14.01	24.98	13.98	24.4	13.74	23.97	13.43	23.75	13.06	23.45
Asia	17.50	14.22	18.11	14.45	18.91	14.79	20.31	15.52	20.99	15.96	21.71	16.42	22.29	16.68
China & India	(7.48)	(6.02)	(7.78)	(6.17)	(8.26)	(6.29)	(9.01)	(6.72)	(9.25)	(6.95)	(9.28)	(7.06)	(9.45)	(7.17)
Africa	27.39	17.55	26.89	17.37	26.61	16.92	25.59	16.97	25.67	16.76	25.53	16.33	25.36	16.11
Tunisia & Morocco	(15.57)	(10.41)	(15.09)	(10.31)	(14.65)	(9.98)	(14.34)	(10.20)	(13.84)	(9.89)	(13.22)	(9.42)	(12.75)	(9.18)
America	6.29	10.06	6.35	9.89	6.42	9.86	6.46	9.64	6.33	9.39	6.34	9.27	6.26	9.06
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Source: Italian Institute of Statistics, http://istat.it/. The values are expressed as percentages. Subsamples of immigrants by continents are listed in parentheses.

References

Adserà, A., Pytliková, M., 2015. The role of language in shaping international migration. The Economic Journal 125, F49–F81.

Adserà, A., Pytliková, M., 2016. Language and Migration. The Palgrave Handbook of Economics and Language, 342-372., London, UK.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, New Jersey.

Barone, G., Mocetti, S., 2011. With a little help from abroad: The effect of low-skilled immigration on the female labour supply. Labour Economics 18, 664–675.

Berman, E., Lang, K., Siniver, E., 2003. Language-skill complementarity: returns to immigrant language acquisition. Labour Economics 10, 265–290.

Bleakley, H., Chin, A., 2004. Language Skills and Earnings: Evidence from Childhood Immigrants. The Review of Economics and Statistics 86, 481–496.

Bleakley, H., Chin, A., 2010. Age at arrival, English proficiency, and social assimilation among U.S. immigrants. American Economic Review, Applied Economics 2, 165–192.

Bloemen, H., 2013. Language proficiency of migrants: The relation with job satisfaction and matching. IZA Discussion Paper 7366, Institute of Labor Economics (IZA) .

Chiswick, B.R., 2008. The Economics of Language: An Introduction and Overview. IZA Discussion Papers 3568. Institute for the Study of Labor (IZA).

Chiswick, B.R., Lee, Y.L., Miller, P.W., 2005. Immigrant earnings: A longitudinal analysis. Review of Income and Wealth 51, 485–503.

Chiswick, B.R., Miller, P.W., 2003. The complementarity of language and other human capital: immigrant earnings in Canada. Economics of Education Review 22, 469–480.

Chiswick, B.R., Miller, P.W., 2009. Earnings and occupational attainment among immigrants. Industrial Relations: A Journal of Economy and Society 48, 454–465.

Chiswick, B.R., Miller, P.W., 2010. Occupational language requirements and the value of English in the US labor market. Journal of Population Economics 23, 353–372.

Chiswick, B.R., Miller, P.W., 2015. International Migration and the Economics of Language, In: Chiswick, B.R., Miller, P.W. (Eds.). Handbook of the Economics of International Migration: The Immigrants, Chapter 5, 240-269. Elsevier., Amsterdam, Netherlands.

Clark, K., Drinkwater, S., 2008. The labour-market performance of recent migrants. Oxford Review of Economic Policy 24, 495–516.

Dustmann, C., Fabbri, F., 2003. Language proficiency and labour market performance of immigrants in the UK. The Economic Journal 113, 695–717.

Dustmann, C., van Soest, A., 2001. Language fluency and earnings: Estimation with misclassified language indicators. The Review of Economics and Statistics 83, 663–674.

Friedberg, R.M., 2000. You can't take it with you? Immigrant assimilation and the portability of human capital. Journal of Labor Economics 18, 221–251.

Gonzales, P.S., 2010. The impact of education and host language skills on the labor market outcomes of immigrants in Spain, in: Investigaciones de Economà de la Educaciòn 5. Asociaciòn de Economía de la Educaciòn, pp. 798–824.

Guven, C., Islam, A., 2015. Age at migration, language proficiency, and socioeconomic outcomes: Evidence from Australia. Demography 52, 513–542.

Hermansen, A.S., 2017. Age at arrival and life chances among childhood immigrants. Demography 54, 201–229.

Himmler, O., Jäckle, R., 2018. Literacy and the migrant's native wage gap. Review of Income and Wealth 64, 592–625.

Kalter, F., Kogan, I., 2006. Ethnic inequalities at the transition from school to work in Belgium and Spain: Discrimination or self-exclusion? Research in Social Stratification and Mobility 24, 259 – 274.

Lehmer, F., Ludsteck, J., 2015. Wage assimilation of foreigners: Which factors close the gap? evidence from Germany. Review of Income and Wealth 61, 677–701.

Mariani, R.D., Pasquini, A., Rosati, F.C., 2021. The immigration puzzle in italy: A survey of evidence and facts. Italian Economic Journal .

Massey, D.S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, J.E., 1993. Theories of international migration: A review and appraisal. Population and Development Review 19, 431–466.

Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the U. S. labor market. The Quarterly Journal of Economics 118, 549–599.

Pieroni, L., Pompei, F., 2008. Evaluating innovation and labour market relationships: the case of Italy. Cambridge Journal of Economics 32, 325–347.

Piras, R., 2019. Internal migration in italy: The role of migration networks. Italian Economic Journal 6, 157–195.

Ponzo, I., Finotelli, C., Malheiros, J., Fonseca, L., Salis, E., 2015. Is the economic crisis in southern Europe turning into a migrant integration crisis? Politiche Sociali 1, 59–88.

Rendon, S., 2007. The Catalan premium: language and employment in Catalonia. Journal of Population Economics 20, 669–686.

Salmasi, L., Pieroni, L., 2015. Immigration policy and birth weight: Positive externalities in Italian law. Journal of Health Economics 43, 128 – 139.

Yao, Y., van Ours, J.C., 2015. Language skills and labor market performance of immigrants in the Netherlands. Labour Economics 34, 76– 85.