European Economic Review 137 (2021) 103781

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Immigration and nationalism: The importance of identity☆

Francesco Flaviano Russo

University of Naples Federico II and CSEF, Via Cintia, Monte Sant' Angelo, 80126, Napoli, Italy

ARTICLE INFO

JEL classification:
D72
J61

Keywords:
Elections
Culture
Language
Religion

ABSTRACT

I study how immigration affected electoral outcomes in Italy accounting for changes in the immigrants identity, measured as their cultural distance from the natives. I consider three metrics of cultural distance based on language, on religion and on genetic traits which are related to the vertical transmission of values and norms across generations. I find that the increased cultural distance between immigrants and natives determined more votes for nationalist, anti-immigration, parties, although this is not the result of increased religious distance. The immigrant share does not explain electoral outcomes once either of the three distance metrics is accounted for.

1. Introduction

Immigration to European countries has increased substantially in the past 15 years. In Italy, for instance, the share of foreign-born residents rose from 2.3% of the population in 2002 to 8.5% in 2018. During the same period there was a dramatic increase in the number of men and women crossing the Mediterranean on shabby boats, risking their lives to reach the European coast-mostly Italy, Spain and Greece. This increase was partly due to the Syrian War, that generated a massive displacement of individuals trying to reach Europe seeking asylum. An additional explanation is the collapse of the Libyan regime in 2011, that loosened the enforcement against human traffickers from countries such as Eritrea, Gambia, Somalia and Sudan. These events inevitably put immigration on top of the political agenda, especially in Italy. Nationalist political parties such as Lega Nord (henceforth Lega), headed by Mr. Matteo Salvini, tried to take political advantage, promising to stem what they conveniently defined as an "Invasion". Lega was extremely successful in the 2018 political elections, obtaining roughly 5.7 million votes, or 17% of the total, making it the second most voted party after the Movimento 5 Stelle, and allowing Mr. Salvini to become Minister of the Interior, with responsibility over immigration policy.

In a previous contribution, Barone et al. (2016) identify a causal relationship between increased immigration and increased votes for center-right parties in Italy, using panel data at the municipality level. Similar results appear in Mendez and Cutillas (2014) for Spain, Halla et al. (2017) for Austria, Gerdes and Wadensjo (2008), Harmon (2018) and Dustman et al. (2016) for Denmark, Otto and Steinhardt (2014) for the city of Hamburg, Edo et al. (2019) for France, Dinas et al. (2019) for Greece and Tabellini (2020) for the US at the beginning of the 20th century.

Increased immigration is often associated with changes in the migrant stock composition. Immigrants from different countries face different difficulties learning the host country language. They may also carry different social norms, have a different religion and so on. I expect such differences to produce different (possibly racist) responses from native Italians, since culturally distant immigrants could be perceived as a threat to the natives' identity (Sherif, 1953; Hainmueller and Hopkins, 2014). It is therefore fundamental to account for the immigrants' characteristics in order to better understand the electoral response to immigration.

☆ I would like to thank the journal editor, David Levine, an anonymous associate editor, Mario Carillo, Rosario Crinò, Marcello D'Amato, Ilyana Kuziemko, Giuseppe Russo, Alberto Zazzaro, two anonymous referees and the seminar participants at the University of Naples, the ESWM 18 meeting and the 2019 CSEF-IGIER symposium for the helpful comments. All remaining errors are mine.

E-mail address: francescoflaviano.russo@unina.it.

https://doi.org/10.1016/j.euroecorev.2021.103781
Received 30 March 2020; Received in revised form 31 March 2021; Accepted 21 May 2021
Available online 10 June 2021

In this paper I propose such an analysis. I account for immigrants' identity through their cultural distance from the native population, computed according to three metrics based on religious, linguistic and genetic similarity (Cavalli-Sforza et al., 1994). This last metric, in particular, is related to the vertical transmission of values and norms across generations (Spolaore and Wacziarg, 2015), and it was shown to be related to differences in opinions on, among others, religion, democracy, freedom of expression and trust (Desmet et al., 2011).

The main problem with this analysis is that the immigrants' settlement choices can be endogenous to political outcomes, for instance if they avoid cities with a high concentration of nationalists so as to avoid discrimination. The solution adopted in many previous contributions entails using an exogenous instrument built around the *Enclave* theory (Card, 2001) or *Chain Migration* hypothesis. According to this theory, immigrants tend to settle where other immigrants of the same nationality already live, for family reunions and because expat communities can provide help with housing and employment. Since past settlement choices are exogenous to current political developments, it is therefore possible to use the immigrant share in a given city in a reference year in the past to artificially redistribute incoming immigrants at the national level in subsequent years, obtaining an artificial, exogenous, immigration flow. This is the so-called "Shift-share" or Bartik instrument.

This instrument is problematic if a city had no immigrants of a given nationality in the reference year, because it predicts that no immigrant of that nationality will ever be allocated there. This is particularly troublesome when recent immigration flows are coupled with a significant diffusion over the territory, like in Italy. For instance, in 2004 there were only 51 cities with immigrants from Afghanistan, but 1019 in 2017. The shift-share instrument, in this case, would be equal to zero in 968 cities with a positive, and perhaps big, share of immigrants from Afghanistan, with very little predictive power.

To overcome this issue, I propose an alternative instrument based on the distance between each city and the enclaves for immigrants of a given nationality in a reference year. The simple logic is that if, say, Milan is the major enclave for Egyptian immigrants, then there will be more new immigrants from Egypt that decide to settle in the outskirts of Milan rather than in Naples. I call this modification the *Spatial Migration Chain* hypothesis. This new instrument is based on two main ideas: first, the reasons why a given city is an enclave for immigrants of a given nationality are deeply rooted and, as such, not a function of current electoral outcomes. Second, the distances between each city and those enclaves are exogenous. On top of that, since the total number of incoming immigrants might itself depend on the overall political situation at the national level, I use an alternative, *artificial shift* component equal to the number of emigrants from each country multiplied by the fraction of those immigrants that, in the past, chose Italy as destination. Once I have an instrument for the shares of immigrants, I can construct exogenous weighted average linguistic, religious and genetic distances between immigrants and natives using these artificial shares as weights.

I estimate two empirical models. The first entails a regression, at the city level, of the vote shares for nationalist parties/coalitions on the immigrant share and on the cultural distances between immigrants and natives, for the immigrants that live in the given city. The second entails a regression of the same vote shares on the average immigrant share and on the average cultural distances, computed over all cities within a small distance. This last empirical specification is useful to address two fundamental issues: first, voters might not only be influenced by what happens in their city, but also by what happens nearby, especially in case of small towns, whose citizens typically commute to work or shop. Second, there is a potential selection problem, because some immigrants might choose to settle in a given city because of their personal, unobserved, characteristics, which are potentially different from the average characteristics of an individual from their origin country.

I find that, in both empirical models, genetic and linguistic distance between immigrants and natives are positively and significantly associated with the vote share for nationalist parties, and that the immigrant share does not explain electoral outcomes once linguistic and genetic distances are included in the regression. Conversely, religious distance does not robustly explain nationalist votes. These results are indeed new to the literature, which currently identifies the share of immigrants as a determinant of the nationalist vote.

A small note before proceeding. I use the term nationalist as synonym with anti-immigration, although it must be clear that nationalist parties do not push for stricter immigration policies only, but typically also for trade restrictions and, more generally, for identitarian policies.

The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 describes the dataset. Section 4 describes the empirical models, the strategy to deal with the endogeneity of the main regressors (Section 4.1), and the algorithm to construct an instrument for cultural distance (Section 4.2). Section 5 summarizes the empirical results. Section 6 analyzes the robustness of the results and discusses several extensions. Section 7 concludes. In a companion appendix, I discuss the identification of the empirical model and I summarize several additional empirical results.

2. Related literature

This paper is related to the literature on the effect of immigration on voting, more closely to the few works that explicitly account for the identity of the migrants. To my knowledge, only three papers do so. The first is Barone et al. (2016), who find a positive impact of immigration on votes for center-right parties in Italy, which is stronger in case of bigger religious differences. The second is Brunner and Kuhn (2018), who find a positive relationship between the share of culturally distant immigrants (measured following Inglehart and Baker, 2000) and anti-immigration votes in Switzerland. The third is Tabellini (2020), who finds that the inflow of immigrants determined the election of more conservative politicians and the implementation of anti-immigration policies in the US between 1910 and 1930, highlighting increased linguistic and religious distances as the most likely determinants.

Several other works also find a positive relationship between immigration and nationalist votes, but without accounting for the immigrants' characteristics: Halla et al. (2017) in Austria, Otto and Steinhardt (2014) in the city of Hamburg, Edo et al. (2019) in

Table 1
Immigration and elections: Literature summary.

Paper	Country and period	Result	Coeff	Role of culture
Gerdes and Wadensjo (2008)	Denmark 1989–2001	Higher immigrant share ⇒ More votes for nationalist parties	0.04	None
Mendez and Cutillas (2014)	Spain 1998–2008	No effect of immigrant share on votes for right-wing parties	0	None (but continent of origin matters)
Barone et al. (2016)	Italy 2001–2008	Higher immigrant share ⇒ More vote for center-right parties	1.26	Religion amplifies the effect
Dustman et al. (2016)	Denmark 1986–1998	Higher immigrant share ⇒ More votes for nationalist parties	1.23	None
Mayda et al. (2016)	US 1994–2012	More immigrants ⇒ Less votes for republicans	1.8	None
Steinmayr (2016)	Austria 2015	Hosting refugees ⇒ Less votes for nationalist parties	4.4	None
Bratti et al. (2017)	Italy 2016	Closer to refugees shelters ⇒ More votes for nationalist parties	0.13	None
Halla et al. (2017)	Austria 1981–2011	Higher immigrant share ⇒ More votes for nationalist parties	0.16	None
Otto and Steinhardt (2014)	Hamburg 1987–2000	Higher immigrant share ⇒ More votes for nationalist parties	0.34	None
Brunner and Kuhn (2018)	Switzerland 1970–2010	Higher immigrant share ⇒ More anti-immigrant vote	0.17	Culturally distant immigrants explain the effect
Harmon (2018)	Denmark 1989–2001	Higher immigrant share ⇒ More seats for nationalist parties	2.8	None
Vertier and Viskanic (2018)	France 2017	Hosting refugees ⇒ Less votes for nationalist parties	15.7*	None
Dinas et al. (2019)	Greece 2015	Exposure to refugees ⇒ More votes for extreme-right parties	2*	None
Edo et al. (2019)	France 1988–2017	Higher immigrant share ⇒ More votes for extreme-right	0.4	None
Tabellini (2020)	US 1910–1930	Higher immigrant share ⇒ Higher prob to elect a conservative	2.5	Religious and linguistic distance explain the effect

Notes: Coeff is the percentage point variation in votes following 1% increase in the right-hand-side variable except for the entries followed by a star (*) that refer to dummy right-hand-side variables. The papers are listed in chronological order.

France, Gerdes and Wadensjo (2008), Harmon (2018) and Dustman et al. (2016) in Denmark, Dinas et al. (2019) in Greece, Mendez and Cutillas (2014) in Spain (but only for immigrants from Africa), Bratti et al. (2017) in Italy. Two recent works actually find a negative relationship between the inflow of refugees and nationalist votes: Steinmayr (2016) in Austria and Vertier and Viskanic (2018) in France. Mayda et al. (2016) highlight contrasting effects of immigration on electoral outcomes in the US: naturalized immigrants typically vote democrat, yet increases in non-naturalized immigrants lead to more republican votes. Table 1 summarizes this literature.

I contribute to this literature by using cultural characteristics to disentangle the effect of the changing composition of the stock of immigrants from the effect of its size. The main result is that the immigrant shares do not affect electoral outcomes once their cultural distance from the natives is properly accounted for. Moreover, I show that the immigrants' characteristics in neighboring towns influence electoral outcomes, addressing the potential selection of immigrants into municipalities based on unobservables.

Some related, survey-based studies also find a significant empirical relationship between immigration increases and anti-immigration sentiment, although not looking at electoral results. Examples include Dustmann and Preston (2001) for the UK, Card et al. (2012) for the US, Slotwinski and Stutzer (2019) for Switzerland and Mayda (2006) and Facchini and Mayda (2009) for a panel of countries. The analysis also relates to some recent works explaining the raise of populist parties in Europe, such as Guiso et al. (2019).

3. Data

The empirical analysis focuses on Italy. The dataset covers the universe of municipalities. To account for the slow formation of beliefs, and thus political preferences, I merge the electoral outcomes with the average value of all other variables computed between the electoral year and the year before the elections. For simplicity, I will discuss the electoral results for only one of the two chambers that compose the Italian Parliament, the Camera dei Deputati. Despite the slightly different electoral rules, the results for the other chamber, the Senato della Repubblica, are almost identical.

3.1. National elections in Italy

I use data on electoral results from the ELIGENDO database, managed by the Italian Ministry of the Interior. I focus on the last four national political elections, held in 2006, 2008, 2013 and 2018. I consider two alternative measures of the nationalist vote. The first is the vote share for Lega, the most prominent anti-immigration party in Italy. Lega formed in 1989 from the merger of six small, northern-Italian, independentist, parties under the charismatic leadership of Mr. Umberto Bossi, who was head of the party for two decades. Its earlier political agenda was focused on increased political and economic autonomy for northern regions, on Euro-skepticism and, in general, on souverainism. Recently, under the leadership of Mr. Matteo Salvini, there has been a shift toward more extreme, right-wing positions, especially toward immigration.

The second measure is the vote share of an artificial coalition composed by Lega and by few smaller, right-wing parties, whose rhetoric and political platform are also heavily biased against immigrants. I will refer to this artificial coalition as XR (acronym for Extreme-Right). Not all parties competed in all elections, mostly because they did not survive for such a long period of time (the full list is available in appendix). In 2018, XR included Lega and four other parties: Fratelli D'Italia, historically bonded to the post-fascist party Movimento Sociale Italiano, as well as Casa Pound, Forza Nuova and Italia Agli Italiani, which are all right-populist, openly neo-fascist political movements.

Looking at the last election held in 2018, Lega alone was the second most voted party, with 17% of the votes. In some northern cities, it collected more than 50%. Interestingly, the vote shares for Lega increased also in most southern Italian cities, where it was almost non-existent few years before. Just to give a couple of examples, in Naples, the biggest city in southern Italy, almost 2.6% of the voters chose Lega in 2018, compared to 0.14% in 2013. In Bari, the second-largest southern Italian city, Lega got an astonishing 5% in 2018, up from 0.05% in 2013.

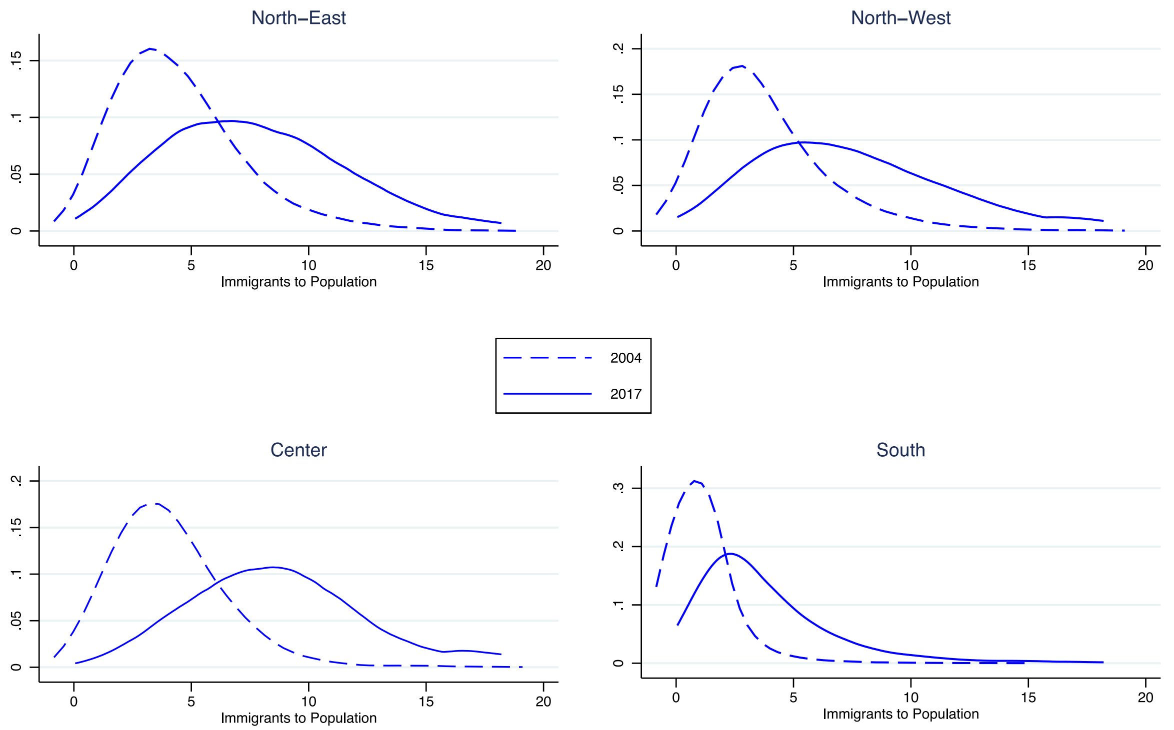

Fig. 1. Ratio of immigrants to population. **Notes**: Empirical distribution of the ratio of immigrants to population over municipalities in 2004 and 2017 by macroarea.
Source: Own elaboration based on data by ISTAT.

3.2. Immigration to Italy

The immigration data are from the Italian Statistical Institute (ISTAT). They refer to regular, registered immigrants, while there is no information on irregular immigration, which might be a potential determinant of electoral outcomes.[1]

The immigrant share increased from 3% in 2004 to 8% in 2017. Looking at the distribution over cities, the median share increased from 2.4% in 2004 to 5.8% in 2017. This increase was sharper in cities with smaller shares of immigrants. For instance, in Naples, the increase was from 1% in 2004 to 5.7% in 2017. To give a graphical idea of the inflow of immigrants, Fig. 1 plots the empirical distribution (over municipalities) of the ratio of immigrants to population, separately by macro-areas: North-East, North-West, Center and South.

Another important empirical regularity is the significant change in the composition of immigrants by nationality. For instance, back in 2004, 13.6% of all immigrants to Italy where from Albania, while in 2017 only 8.9%. Similarly, immigrants from Morocco were 12.7% of the total in 2004, while 8.3% in 2017. Conversely, Romanians increased from 8.9% in 2004 to 23.2% in 2017. Chinese immigrants also increased, from 4.4% in 2004 to 5.6% in 2017. This evidence is actually what motivates my empirical analysis. Fig. 2 provides a visual overview of this differential increase in immigration by nationality, mapping the changes in shares for the 4 most important nationalities of immigration in the Lazio region,[2] which includes the capital city Rome (home to 7.5% of all immigrants to Italy).

Together with the increased shares and with the changed identity, there has also been an increase in the number of nationalities in most Italian cities, especially in smaller ones. In Anzio, a small city close to Rome, the number of nationalities increased from 84 in 2004 to 107 in 2017. In Cirò marina, a beach destination in Calabria, from 27 to 44. Taking the national average, the number of nationalities increased from 17.8 to 25.7. The increase was much milder in large cities: for instance, in Rome the number of nationalities went from 166 to 173 and from 143 to 156 in Milan. Turin and Florence similarly experienced modest increases. This implies that some countries of origin were represented in increasingly many cities. For instance, Gambians were present in 149 cities in 2004, and in 1500 in 2017. This increase in the number of cities where immigrants settled, accounting for a dispersion across the national territory, is the reason why I needed to construct a new instrument.

[1] If the immigrants settled close to other immigrants of the same nationality, as predicted by the *Chain Migration* hypothesis, then the number of irregulars would be correlated with the number of regulars. However enforcement against illegal immigration might be asymmetric and potentially dependent on electoral results.

[2] Further pictures for additional regions can be found in appendix.

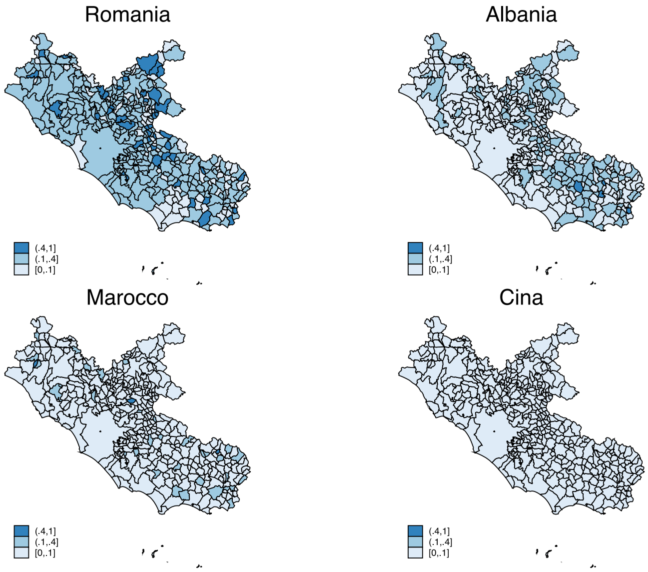

Fig. 2. Change in the shares of immigrants by nationality. **Notes**: Change in the shares of immigrants by nationality between 2004 and 2017 in the Lazio region. *Source:* Own elaboration based on data by ISTAT.

3.3. Distances between immigrants and natives

To measure the distance between immigrants and natives, I follow Spolaore and Wacziarg (2015), focusing on three dimensions: language, religion and genes.

The measure of linguistic distance is from Dyen et al. (1992). It starts from the identification of words with the same meaning and root between languages, called cognate words. For instance, "Acqua" in Italian and "Agua" in Spanish are cognate because they both mean water and they both come from the same Latin word "Aqua". The measure of linguistic distance between two given languages that I use is equal to one minus the percentage of cognate words for 200 common meanings (water, food, table etc.). The distance can be computed for each language pair, but not at the country level, since there could be more than one language spoken within a country. Spolaore and Wacziarg (2009, 2015) aggregated the language-pairs measures at the country level by computing weighted averages using the ethnic composition data by Alesina et al. (2003) as weights. The final linguistic distance measure should be interpreted as the percentage of different words for two randomly selected individuals from two given countries.

The measure of religious distance is from Mecham et al. (2006). They propose a classification of religions in broad categories, which they use to construct a religion tree. For instance, the Roman Catholic and the Greek Orthodox churches are both Christian churches; broader than this, Islam and Christianity are monotheistic religions, etc. Starting from the tree, they then compute the measure of religious distance for each religion pair as the difference between the maximum number of common classifications/nodes on the tree minus the actual number of common classifications, (standardized to be between 0 and 1). Similarly to linguistic distance, this measure can only be computed at the religion level and Spolaore and Wacziarg (2009) aggregate it at the country level using the ethnic composition in Alesina et al. (2003). This measure should be interpreted as the probability that two randomly selected individuals from two given countries share the same religion.

The last distance metric hinges on genetic similarity and it is based on the works by Cavalli-Sforza et al. (1994) and Spolaore and Wacziarg (2009). Genes change over time because of random variation, and some of those changes persist over time because of natural selection. When two populations separate, the random drift determines different evolutionary paths. Genetic distance is a measure of the resulting difference between the distributions of genes. It is therefore proportional to the time since two population separated or had a common ancestor. It ranges from 0, in the case of countries with a very recent common history, such as Algeria and Tunisia, to 3375, in the case of countries whose populations evolved independently since early human settlements, such as Fiji and Kenya.

I use genetic distance because it is correlated to the vertical transmission of values and norms across generations: Desmet et al. (2011) and Spolaore and Wacziarg (2015) show that it is related to how people respond to the World Values Survey questions, with smaller genetic distances associated with similar opinions over a wide range of subjects spanning from religion and democracy to the role of women. Thus genetic distance is a proxy for cultural distance. Once again, these genetic distance measures, available for population pairs, are aggregated by Spolaore and Wacziarg (2009) at the country level following the same logic described for linguistic distance.

For all distance metrics, I computed weighted averages for all municipalities and election years, using the immigrant shares by nationality as weights. More specifically, the weighted average cultural distance between immigrants and natives is computed as

$$D_{jt} = \sum_{k=1}^{K} S_{jt}^{k} D^{k} \tag{1}$$

where D^k is the distance (genetic, linguistic or religious) between natives and immigrants of nationality k for all $K = 180$ classified nationalities of immigration to Italy, and S_{jt}^{k} is the share of immigrants of nationality k in municipality j and election year t. Importantly, the three main cultural distance metrics that I use are not correlated: the correlation between genetic distance and linguistic distance is 0.278, between genetic distance and religious distance 0.251, and between linguistic distance and religious distance 0.086. This evidence implies that there is indeed something specific to each one of them that deserves a separate analysis.

Three caveats before proceeding. First, and foremost, the distances D^k in Eq. (1) are between an average Italian native and an average individual in the given foreign country k, which entails assuming no selection into migration. This is potentially problematic because migrants can differ from their fellow nationals, differences that may relate to why they emigrated or why they chose Italy. This is not a huge issue if migration chains are important, which implies that it is not so much the immigrants' individual characteristics that matter, such as speaking or not the language of the destination country or not, but rather that a family member or friend has settled into a given city in the past. Still, it is not possible to completely rule out the possibility of selection. In the appendix, I discuss all the available data on the immigrants' characteristics, concluding that selection does not seem to be an issue for my empirical results. That said, when interpreting the results, it is important to keep in mind that there might still be a small measurement error in cultural distance.

Second, there might be selection also at the municipality level: Immigrants from the same country might choose a specific city, rather than another, based on their personal characteristics. To attenuate this selection problem, I include a separate empirical analysis where electoral outcomes in a given city are explained by the average characteristics of the immigrants within a given small distance from that city.

Third, when computing distances, I do not account for the time that the immigrants spent in the country. This is potentially problematic because earlier, more integrated, immigrants might not induce anti-immigration feelings. However, the sharp immigration increase of the past 10 years suggests that many immigrants came to Italy recently; for several cities immigration is indeed a very recent phenomenon.

The distribution of genetic distance shifted to the right from 2004 to 2017, especially in southern regions, meaning there were many more municipalities with culturally distant immigrants in 2017. For 10% of the municipalities (around 800), mostly small towns with small immigrant shares at the beginning of the sample,[3] the genetic distance increase was bigger than 112%. Religious distance follows a similar pattern to genetic distance, albeit with smaller increases. Linguistic distance decreased instead in most cities, mainly as a consequence of the inflow of Romanian immigrants, although it increased in many small towns.

4. Estimation

Section 4.1 describes the empirical model and the strategy to deal with the endogeneity of immigration. Section 4.2 describes the algorithm used to construct the instrument for cultural distance. A full discussion of the data supporting the validity of the identification assumptions can be found in the appendix.

4.1. The empirical models

The first empirical model is the following:

$$V_{jt}^{l} = \theta_j + \lambda_{rt} + \beta_1^l I_{jt} + \beta_2^l D_{jt} + \beta_3^l G_{jt} + X_{jt}' \Gamma^l + \varepsilon_{jt} \tag{2}$$

where V_{jt}^{l} is the vote share for nationalist party or artificial coalition l in municipality j and election t, I_{jt} is the ratio of immigrants to population, D_{jt} is the weighted average cultural distance between immigrants and natives, either religious, linguistic or genetic, G_{jt} is the weighted average GDP per capita of the immigrants' origin countries, X_{jt} are control variables (GDP, firm dynamics, education, demographics,[4] social capital, religiosity and crime), θ_j is a municipality fixed effect and λ_{rt} an election-by-region fixed effect (each municipality j belongs to a region r), which controls both for elections specific factors and for the differential diffusion of political parties across regions and elections.

The inclusion of G_{jt} is important both to disentangle the effect of cultural distance from the effect of the immigrants' economic background, which might affect the compositional amenities (Card et al., 2012; Ottaviano and Peri, 2006) and labor market skills

[3] The inflow of a relatively small number of culturally distant immigrants can significantly increase cultural distance, especially in smaller towns. For instance, suppose that a small a family of two Nigerians (genetic distance from Italians 1443) decides to settle in a small municipality with 18 Albanian immigrants (genetic distance from Italians 85). Weighted average genetic distance will shift from 85 to $0.9 \cdot 85 + 0.1 \cdot 1443 = 200.8$ even if the new immigrants account for just 10% of all immigrants.

[4] See Card (2001) and Cattaneo et al. (2015) on the possibility that an immigrants' inflow might induce an outflow of natives.

Table 2
Summary statistics.

	City					5 km				
	Mean	Med	Std	25	75	Mean	Med	Std	25	75
Lega	11.48	7.61	12.11	0.49	18.75					
XR	17.59	15.88	12.55	6.91	25.74					
Gen Dist	321	287	168	210	398	321	298	129	229	390
Lang Dist	0.78	0.79	0.09	0.72	0.84	0.77	0.79	0.07	0.72	0.83
Relig Dist	0.72	0.75	0.13	0.67	0.79	0.71	0.74	0.09	0.68	0.78
GDPPC	7.38	6.31	3.69	5.25	8.12	7.37	6.61	2.88	5.56	8.31
Imm/pop	5.25	4.41	4.02	2.07	7.51	5.26	4.75	3.51	2.41	7.39

Notes: Lega is the vote share to Lega. XR is the vote share for the Extreme-right artificial coalition (see text). Gen Dist is the weighted average genetic distance between immigrants and natives (see text). Lang Dist is the weighted average linguistic distance between immigrants and natives. Relig Dist is the weighted average religious distance between immigrants and natives (see text). GDPPC is the weighted average GDP per capita at purchasing Power Parity of the Immigrants' origin country. Imm/pop is the ratio of immigrants to total population (pct terms). In the column city, the above variables are relative to each municipality, while in column 5 km they are instead averaged over all cities within 5 km from each municipality.

(Borjas, 2003; Mayda, 2006), and because genetic distance is also correlated with income differences[5] (Spolaore and Wacziarg, 2009). The second empirical model is instead:

$$V^l_{jt} = \theta_j + \lambda_{rt} + \beta^l_1 I^d_{jt} + \beta^l_2 D^d_{jt} + \beta^l_3 G^d_{jt} + X^{d\,\prime}_{jt} \Gamma^l + \varepsilon_{jt} \tag{3}$$

where the superscript d denotes the average value of the variable between all municipalities within d kilometers from municipality j, with the convention that, when $d = 0$, $I^0_{jt} = I_{jt}$ and $D^0_{jt} = D_{jrt}$. As a benchmark, I consider $d = 5$, since smaller distances translate into a very small number of neighboring municipalities, while bigger distances in a very big number (see Section 6 for details). Table 2 reports the summary statistics for the main variables included in these regressions. The main coefficient of interest is β^l_2, measuring the effect of changes in the cultural distance between immigrants and natives on electoral outcomes, controlling for changes in the stock of immigrants.

The problem with both empirical models is that the spatial distribution of immigrants is endogenous to electoral outcomes: immigrants might avoid settling in a city with a lot of nationalists simply because they are afraid of discrimination, the more so the more culturally distant they are from the natives (Dustmann and Preston, 2001). Thus both I^d_{jt} and D^d_{jt} are endogenous.

To deal with the endogeneity of the immigrant share I^d_{jt}, I simply use a Bartik instrument, following previous works such as Barone et al. (2016). According to the *Enclave* theory, or *Chain Migration* hypothesis, immigrants tend to settle in cities where other immigrants of the same nationality already live. One possible explanation is family reunions, with extended family members and friends joining earlier immigrants. Moreover, expat communities typically provide help with housing and job placement. The consequence is that the number of immigrants in the past, being exogenous to current developments, can be used as an instrument for the current number of immigrants. The idea behind the Bartik, or shift-share, instrument, is exactly to use the immigrant shares in the past (the *share* component) to artificially redistribute incoming immigrants (the *shift* component) in subsequent years, thereby producing an exogenous immigrant stock. The first stage regression results are summarized in Table 3.

Dealing with the endogeneity of cultural distance D^d_{jt} is more complicated. In principle, I could follow the same logic of the shift-share instrument and compute weighted average distances between immigrants and natives using, as shares, the exogenous number of immigrants by nationality constructed with the Bartik instrument divided by the total artificial number of immigrants in that municipality. But doing so requires having immigrants' shares by nationality for each municipality in the base year, and this is not possible because immigration in Italy is a relatively recent phenomenon for many municipalities, and because immigration from several countries started only recently. In case of a zero immigrant share in the past for a particular nationality, there will never be new immigrants of that nationality artificially allocated in that city in subsequent years by the Bartik instrument, resulting in a positive share of immigrants instrumented with a zero share (see the appendix for an example). This would not be an issue if the number of cities with immigrants of given nationality did not change much over time, but this is not the case for Italy; or if the object was merely to construct an instrument for the total number of immigrants (see above), in which case it might have some predictive power.

I propose a solution that extends the logic of the Chain Migration hypothesis in a very intuitive way along geographic lines, according to what I call the *Spatial Migration Chains* hypothesis. The idea is that immigrants tend to settle *close to* the cities where other immigrants of the same nationality already live, and not just exactly there. Operationally, I construct a shift-share instrument by nationality with a share component that is not the actual share in a reference year in the past, but a predicted share based on a regression of the number of immigrants of that nationality in the reference year on a polynomial of the minimum distance between each city and the 5 main enclaves for immigrants of the same nationality. I then compute weighted average cultural distances in each municipality using these exogenous shares of immigrants by nationality.

[5] Since Italy is a developed country, the bigger the genetic distance between Italian natives and immigrants, the more likely it is for the immigrants to be unskilled and, because of their limited wage prospects, more likely to access welfare.

Table 3
Instrumenting cultural distance.

	City					
	Gen Dist	Imm/pop	Lang Dist	Imm/pop	Relig Dist	Imm/pop
Gen Dist instr	0.3606***	−0.0001***				
	(0.0316)	(0.00001)				
Lang Dist instr			0.3241***	−0.0043		
			(0.0409)	(0.0137)		
Relig Dist instr					0.5139***	−0.0444***
					(0.0298)	(0.0051)
Imm/pop instr	8.1674***	0.0063***	0.0038***	0.0063***	−0.0004	0.0063***
	(0.8856)	(0.0003)	(0.0004)	(0.0004)	(0.0004)	(0.0003)
R^2	0.135	0.728	0.428	0.728	0.338	0.729
F	106.8	180.3	93.2	206.1	148.9	231.6
	5 Km					
	Gen Dist	Imm/pop	Lang Dist	Imm/pop	Relig Dist	Imm/pop
Gen Dist instr	0.3787***	−0.0001***				
	(0.0321)	(0.00001)				
Lang Dist instr			0.3683***	0.0049		
			(0.0415)	(0.0109)		
Relig Dist instr					0.7939***	−0.0413***
					(0.0245)	(0.0049)
Imm/pop instr	3.3697***	0.0025***	0.0014***	0.0063***	−0.0002	0.0063***
	(0.4659)	(0.0001)	(0.0001)	(0.0003)	(0.0002)	(0.0003)
R^2	0.156	0.728	0.618	0.728	0.494	0.731
F	93.4	123.6	68.2	188.6	523.8	202.3

Notes: First stage regression results. Dependent variable is in columns. City refers to the variables in the municipality while 5 km to the average values among all municipalities within 5 km. Gen Dist is the weighted average genetic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Lang Dist is the weighted average linguistic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Relig Dist is the weighted average religious distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Imm/pop is the ratio of immigrants to total population (pct terms). Imm/pop instr is the shift-share instrument for the immigrants' ratio. Control variables included: Weighted average GDP per capita at purchasing power parity of the immigrants' origin country; Unemployment rate; real GDP adjusted for inflation; thefts and robberies 100k citizens; age dependency ratio; percentage of population above 65; average age; percentage of the population with: college or higher education, high school education and elementary or no education (excluded category: middle school education); percentage of the population regularly attending religious services; percentage of the population that does volunteer work; rate of dismissed firms in the year; rate of new firms in the year. See text for data sources. Municipality and election-by-region fixed effects included. Standard Errors clustered at the municipality level. 28523 observations for 7373 municipalities. *** significant at 1% level. * significant at 5% level. * significant at 10% level.

There are two main identification assumptions: the immigrants' settlement choices of the past are not affected by city covariates which influence current electoral outcomes, as in similar works that use shift-share instruments; the distances from the enclaves are exogenous to the determinants of political preferences. I comment extensively on the validity of both assumptions in the appendix, using a variety of data sources. In brief, I find that: migration chains are important in Italy; there has been a lot of political turnover in city elections among parties with different views on immigration, excluding the existence of "deep" determinants of the political preferences toward immigration; that culturally distant immigrants did not settle in cities with pro-immigration local governments only; that the minimum distance from the enclaves is uncorrelated with the immigrant share, with income per capita, with the size of the municipality and with the political orientation of the city mayor.

The details of the algorithm that I use to construct the instrument are spelled out in Section 4.2. Readers who are not interested in the technical details can skip directly to Section 5, where I discuss the results.

4.2. Spatial migration chains: The algorithm

The first step of the algorithm entails identifying the most important enclaves for all nationalities of origin in the reference year, 2004. I choose 2004 because[6] it is the first year for which detailed data on immigration by nationality are available.

[6] It would not have been ideal to use data for earlier dates, even if available, because immigration from many countries is a relatively recent phenomenon.

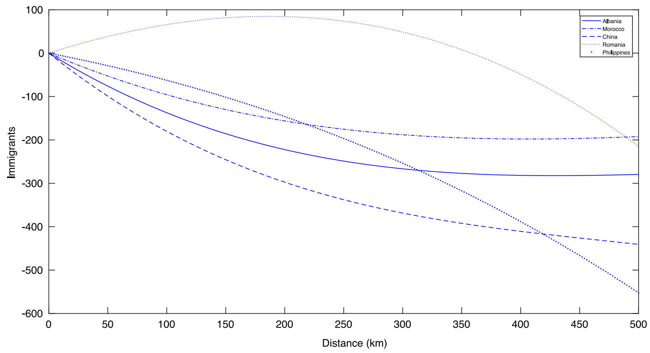

Fig. 3. Spatial migration chains. **Notes**: Empirically estimated relationship between the number of immigrants and the minimum distance between each city and the 5 most important enclaves for immigrants of the same nationality.

I focus on the 5 most important enclaves that, on average (over nationalities), account for 23% of all immigrants, although with significant differences.[7] The reason why I decided to focus on the first 5 enclaves is contingent on the predictive power of the final instrument. In particular, using too many enclaves poses the risk of losing the concept itself of enclaves, given that municipalities with a rather small absolute number of immigrants would qualify as such. On the other hand, considering just one or two towns is an oversimplification in case of several numerous immigrants' communities spread throughout the country. In both cases, the result would be a bad prediction of the number of immigrants based on the minimum distance and, therefore, a weak instrument.

The second step entails the computation of the minimum distance between each municipality and the 5 most important enclaves for all nationalities.

The third step is a regression of the number of immigrants of given nationality on a polynomial of the minimum distance computed at the previous step. The idea is that the decision of a new immigrant to settle in a given city depends on how far the city is from the closest enclave for immigrants of the same nationality. The regression, for each nationality k, is:

$$I_{j0}^{k} = P(h, d_j^k) + \eta_j \tag{4}$$

with:

$$d_j^k = min\{d_j^{k_1}, d_j^{k_2}, d_j^{k_3}, d_j^{k_4}, d_j^{k_5}\} \tag{5}$$

I_{j0}^{k} is the number of immigrants of nationality k in municipality j and reference year $t = 0$, $d_j^{k_z}$ is the distance between municipality j and the zth most important destination city for immigrants of nationality k, and $P(h, .)$ is a polynomial of order h. In the benchmark computations I used $h = 3$, but the results appeared robust to higher orders. In appendix, I discuss the results obtained upon the addition of further explanatory variables in this regression.

The fourth step is the computation of a predicted number of immigrants in each municipality in the reference year ($\hat{I}_{j0}^{k}$) based on the regression results. Fig. 3 shows the results graphically for the 5 most important origin countries: Albania, Romania, Morocco, China and Philippines. The graph shows that cities which are very far from the main enclaves are associated with a significantly smaller number of immigrants, although the settlement patterns for the cities which are closer to the enclaves depend on the nationality.

The above computation predicts a positive number of immigrants in (many) more cities as compared to where they actually settle. The fifth step is a correction for geographic over-dispersion using the relative number of cities with immigrants of given nationality in 2017. More formally, the predicted number of immigrants is corrected as follows:

$$\bar{I}_{j0}^{k} = \begin{cases} \hat{I}_{j0}^{k} & \text{prob} \quad \alpha_k \\ 0 & \text{prob} \quad (1 - \alpha_k) \end{cases} \tag{6}$$

where $\alpha_k = (J_T^k / J_T)$, with J_T^k equal to the number of cities with immigrants from country k in year $T = 2017$ and J_T equal to the total number of cities in 2017.[8]

The sixth step is the computation of the fraction of immigrants of nationality k in each municipality as follows:

$$\hat{F}_j^k = \frac{\bar{I}_{j0}^{k}}{\sum_{j=1}^{J} \bar{I}_{j0}^{k}} \tag{7}$$

[7] For instance, the cumulative share of immigrants from the Philippines in the 5 most important enclaves in 2004 is 57%. For Sri Lanka, this cumulative share is also quite high, 45%. Conversely, the cumulative share is only 20% for Romania, 12% for Russia and 9% for Senegal.

[8] Operationally, for each municipality, I draw a uniform random variable between 0 and 1 and I accept the predicted number of immigrants only if the value of the random variable is below α_k.

Table 4
Cultural distance and nationalist vote: Lega. OLS.

	(1) City	(2) 5 km	(3) City	(4) 5 km	(5) City	(6) 5 km
Gen Dist	−0.0007***	−0.0009**				
	(0.0002)	(0.0004)				
Lang Dist			−0.9834	−1.2362		
			(0.6454)	(0.8485)		
Relig Dist					0.4476	−0.8222
					(0.5862)	(0.7355)
Imm/pop	5.2538**	14.0315***	4.7596**	13.4084***	4.5322**	13.3597***
	(2.3613)	(2.7334)	(2.3568)	(2.7188)	(2.3318)	(2.7249)
GDPPC	−0.0136	−0.0094	−0.0055	−0.0061	0.0019	−0.0094
	(0.0181)	(0.0176)	(0.0176)	(0.0174)	(0.0194)	(0.0179)
R^2	0.897	0.898	0.897	0.898	0.897	0.898

Notes: Dependent variable is the vote share for Lega. Gen Dist is the weighted average genetic distance between immigrants and natives (see text). Lang Dist is the weighted average linguistic distance between immigrants and natives (see text). Relig Dist is the weighted average religious distance between immigrants and natives (see text). GDPPC is the weighted average GDP per capita at purchasing power parity. Imm/pop is the ratio of immigrants to total population (pct terms), instrumented with a shift-share instrument (see text). In the column city, the above regressors and instruments are relative to each municipality, in column 5 km they are instead averaged over all cities within 5 km from each municipality. All regressions include: total population, unemployment rate, GDP adjusted for inflation, number of thefts and robberies 100) citizens, age dependency ratio, the percentage of population above 65, the average age, the percentage of the population with college (or higher)education, with high school education and with elementary or no education (excluded category: middle school education), the percentage of the population regularly attending religious services, the percentage of the population that does volunteer work, the rate of dismissed firms in the year and the rate of new firms in the year. See text for data sources. Municipality and election-by-region fixed effects are included. Standard Errors are clustered at the municipality level. 28523 total observations for 7373 municipalities. *** significant at 1% level. * significant at 5% level. * significant at 10% level.

These are the "share" components of the instrument, that allow an artificial redistribution of new entrants in subsequent years.

Given the potential endogeneity of the number of immigrants by nationality at the country level to the overall political situation, I used an alternative, "artificial shift" component equal to the number of emigrants from each origin country multiplied by the fraction of those immigrants that, in the past, chose Italy as destination:

$$N_t^k = D_{t-x}^k E_t^k \tag{8}$$

where E_t^k are all emigrants from country k in year t, including the ones that did not come to Italy, and D_{t-x}^k is the share of emigrants from country k that came to Italy in year $t-x$. With this procedure, I redistribute the number of new entrants that I observed if the same proportion of emigrants that came to Italy the distant past decided to immigrate to Italy each year. In other words, I use changes in the number of emigrants, which are determined by country-specific, exogenous, push shocks, to identify the empirical model, unlike previous works. The baseline year $t-x$ is 1995, the year when the Schengen treaty became effective for Italy.[9]

The last step of the algorithm entails attributing new entrants to municipalities according to the shares in a recursive fashion, which results in an instrument for the immigrant shares:

$$\hat{S}_{jt+1}^k = \frac{N_{t+1}^k \hat{F}_j^k + I_{jt}^k}{\sum_{k=1}^K (N_{t+1}^k \hat{F}_j^k + I_{jt}^k)} \tag{9}$$

where I_{jt}^k are the immigrants of nationality k in municipality j and year t from the previous stage of the recursive computation. Once I have exogenous shares, I can construct the instrument for the weighted average distance between immigrants and natives simply using, as weights, the instrumented shares rather than the actual ones:

$$\hat{D}_{jrt} = \sum_{k=1}^{K} \hat{S}_{jrt}^k D^k \tag{10}$$

For Eq. (2), I construct and instrument following the above procedure. For Eq. (3) I consider instead the average values delivered by the above procedure for all municipalities within $d = 5$ km.[10] Table 3 summarizes the first stage regression results for both empirical specifications, alongside the results for the shift-share instrument for the immigrant shares. The predictive power of the instruments is quite remarkable both for the immigrant share and for all distance metrics.

[9] I focus on the post-Schengen period because the treaty determined a significant shift in the composition by nationality of the incoming immigrants and, as already noted, going too far back in time implies a lot of zero shares of immigrants for many nationalities.

[10] An alternative empirical strategy would be to use the instrument for Eq. (3) in Eq. (2). Unfortunately the first stage regression results for these alternative specifications highlight a weak instrument problem, particularly for linguistic distance. Therefore I will not discuss such results.

Table 5
Cultural distance and nationalist vote: Lega. IV.

	(1) City	(2)	(3) 5 km	(4)	(5) City	(6) 5 km	(7) City	(8) 5 km
Gen Dist	0.0046***	0.0044***	0.0045**	0.0058**				
	(0.0017)	(0.0017)	(0.0023)	(0.0030)				
Lang Dist					13.5097**	26.3589***		
					(6.0303)	(7.9867)		
Relig Dist							−2.8201	−2.4734
							(2.4157)	(1.7727)
Imm/pop	−4.0253	−24.6227	−1.0534	−61.8782	−6.3413	−9.4204	2.1961	5.7875
	(5.2953)	(19.1384)	(12.1934)	(66.9799)	(6.1209)	(13.2713)	(5.0404)	(12.3102)
Gen*Imm		0.0531		0.1598				
		(0.0445)		(0.1644)				
GDPPC	0.0511*	0.0695**	0.0139	0.0392	0.0103	0.0167	−0.0433	−0.0204
	(0.0293)	(0.0332)	(0.0215)	(0.0341)	(0.0211)	(0.0211)	(0.0383)	(0.0205)
R^2	0.894	0.891	0.896	0.893	0.892	0.889	0.897	0.897
F (dist)	106.8	71.4	93.4	63.4	93.2	68.2	148.9	523.8
F (imm)	180.3	127.2	123.6	91.8	206.1	188.6	231.6	202.3

Notes: Dependent variable is the vote share for Lega. Gen Dist is the weighted average genetic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Lang Dist is the weighted average linguistic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Relig Dist is the weighted average religious distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Imm/pop is the ratio of immigrants to total population (pct terms), instrumented with a shift-share instrument (see text). Gen*Imm is the interaction term between genetic distance and the ratio of immigrants to population. In the column city, the above regressors and instruments are relative to each municipality, in column 5 km they are instead averaged over all cities within 5 km from each municipality. All regressions include: GDP per capita at purchasing power parity of the immigrants' origin country (GDPPC), total population, unemployment rate, GDP adjusted for inflation, number of thefts and robberies 100k citizens, age dependency ratio, the percentage of population above 65, the average age, the percentage of the population with college (or higher)education, with high school education and with elementary or no education (excluded category: middle school education), the percentage of the population regularly attending religious services, the percentage of the population that does volunteer work, the rate of dismissed firms in the year and the rate of new firms in the year. See text for data sources. Municipality and election-by-region fixed effects are included. Standard Errors are clustered at the municipality level. F (dist) is the first stage F statistics for the cultural distance variable. F (imm) is the first stage F statistics for the immigrants' share. 28514 total observations for 7372 municipalities. *** significant at 1% level. * significant at 5% level. * significant at 10% level.

5. Results

Table 4 reports the reference OLS regression results for both empirical models with Lega as dependent variable. Genetic and linguistic distance are negatively associated with the vote share for Lega. Both results illustrate the endogeneity of cultural distance: culturally more distant immigrants avoid settling in cities with more nationalists, thereby inducing a negative correlation between cultural distance and votes for Lega. Similar results hold for XR coalition (details available upon request).

The main results of the paper are reported in Tables 5 and 6, where the vote shares for Lega (Table 5) and for the artificial XR coalition (Table 6), are regressed on the immigrant share and on the average cultural distance between immigrants and natives, controlling for municipality fixed effects, election-by-region fixed effects and for a wide set of covariates, including the weighted average GDP per capita of the origin country that allows to isolate the effect of cultural traits from the development stage of the home country.

Genetic distance is positively and significantly associated with the vote share for nationalist parties. A rather extreme thought experiment is helpful to give a sense of the magnitude of the results. Suppose that, in a given municipality, all immigrants are from Spain and, thus, both culturally and genetically very close to Italians (genetic distance equal to 60). Suppose that those immigrants are suddenly replaced by an equal number of immigrants from Kenya, which are culturally and genetically more distant (genetic distance equal to 2212). The results predict an increase of the vote share for Lega by 10 percentage points and for the XR artificial coalition by 12 percentage points. Increased linguistic distance between immigrants and natives is also positively associated with more votes for nationalist parties. I perform a similar thought experiment to gauge the magnitude of the effect implied by the regression results, but this time replacing immigrants from France (linguistic distance 0.56) with immigrants from Ghana (linguistic distance 1). The predicted vote share increase for Lega (XR) is 5.8 (8.3) percentage points. As for religious distance, there results highlight the absence of an empirical relationship with the vote share nationalist parties. The results are very similar when looking at averages over all municipalities within 5 km. This suggests that the results are robust to the potential selection of immigrants into municipalities based on unobservable characteristics that might induce a bias in the previous results.

As for the share of immigrants, I do not find any robust empirical relationship with the votes for nationalist parties once the proper measure of cultural distance between immigrants and natives is accounted for.

Table 6
Cultural distance and nationalist vote: Extreme-Right (XR). IV.

	(1) City	(2)	(3) 5 km	(4)	(5) City	(6) 5 km	(7) City	(8) 5 km
Gen Dist	0.0057***	0.0055***	0.0087***	0.0087***				
	(0.0022)	(0.0023)	(0.0029)	(0.0034)				
Lang Dist					18.9215***	25.6392***		
					(7.5507)	(8.9902)		
Relig Dist							2.6568	1.9228
							(3.0861)	(1.8857)
Imm/pop	−2.4749	−14.3583	0.1277	1.4941	−6.8289	−1.7112	5.1059	12.4945
	(6.1952)	(21.9793)	(13.9938)	(72.7307)	(7.2251)	(14.8897)	(5.5514)	(13.8174)
Gen*Imm		0.0306		−0.0036				
		(0.0524)		(0.1784)				
GDPPC	0.0682*	0.0789**	0.0371	0.0365	0.0214	0.0197	0.0331	0.0072
	(0.0364)	(0.0394)	(0.0259)	(0.0371)	(0.0258)	(0.0241)	(0.0465)	(0.0238)
R^2	0.861	0.859	0.896	0.861	0.856	0.857	0.864	0.864
F (dist)	106.8	71.4	93.4	63.4	93.2	68.2	148.9	523.8
F (imm)	180.3	127.2	187.4	134.6	206.1	188.6	231.6	202.3

Notes: Dependent variable is the vote share for the Extreme-right (XR) artificial coalition (see text). Gen Dist is the weighted average genetic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Lang Dist is the weighted average linguistic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Relig Dist is the weighted average religious distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Imm/pop is the ratio of immigrants to total population (pct terms), instrumented with a shift-share instrument (see text). Gen*Imm is the interaction term between genetic distance and the ratio of immigrants to population In the column city, the above regressors and instruments are relative to each municipality, in column 5 km they are instead averaged over all cities within 5 km from each municipality. All regressions include: GDP per capita at purchasing power parity of the immigrants' origin country (GDPPC), total population, unemployment rate, GDP adjusted for inflation, number of thefts and robberies 100k citizens, age dependency ratio, the percentage of population above 65, the average age, the percentage of the population with college (or higher)education, with high school education and with elementary or no education (excluded category: middle school education), the percentage of the population regularly attending religious services, the percentage of the population that does volunteer work, the rate of dismissed firms in the year and the rate of new firms in the year. See text for data sources. Municipality and election-by-region fixed effects are included. Standard Errors are clustered at the municipality level. F (dist) is the first stage F statistics for the cultural distance variable. F (imm) is the first stage F statistics for the immigrants' share. 28514 total observations for 7372 municipalities. *** significant at 1% level. * significant at 5% level. * significant at 10% level.

Weighted average GDP per capita differences between immigrants and natives are positively associated with nationalist votes, although the coefficient is not significant in all regressions. Among the other controls, There is negative relationship between income per capita and nationalist vote and a positive relationship between unemployment and nationalist vote. Moreover, there is a negative relationship between firm creation and nationalist vote and a positive relationship between firm destruction and nationalist vote. Finally, thefts and robberies are positively associated with nationalist votes.

I also found that all of the above results are robust to the inclusion of an interaction term between the share of immigrants and the cultural distance between immigrants and natives. In those regressions, the interaction term itself is not significant. When all distances are included together, genetic and linguistic distances are still, respectively, positively and negatively significantly associated with the vote share for nationalist parties, while religious distance is not.

The results are similar when excluding small municipalities below 1 thousand individuals and big municipalities above 500 thousand individuals. When restricting the sample to southern Italy, the coefficient on genetic distance becomes not statistically significant, while that on linguistic and religious distance is significant for XR only. When restricting to Norther regions, the results resemble what I found for the whole sample.

I also tried running a regression similar to Barone et al. (2016) in my sample and with my empirical specification, which entails the exclusion of cultural distance variables and the inclusion, together with municipality fixed effects, of region-by-election fixed effects, rather than election only, and a wide set of control variables. The instrument for the immigrant share is a canonical shift-share, constructed using the shares by nationality in 2004 to redistribute immigrants' flows. Despite all differences, the main result by Barone et al. (2016) still stands: more immigrants are associated with more votes for center-right parties (see the appendix for details on the construction of the center-right coalition). Thus the reason why I find that the immigrant share does not explain nationalist vote is not because of the different sample, of the different fixed effects, of the different control variables or of the different base year, but because of the inclusion of cultural distance. Moreover, I did not find any statistical relationship between the immigrant share and the votes for Lega and XR, which proposed the harshest anti-immigration agenda.

Summarizing, the empirical evidence shows that the cultural distance between immigrants and natives is positively and statistically significantly associated with nationalist vote, although not as a result of religious differences, and that the immigrant share does not explain electoral outcomes once the cultural distance between immigrants and natives is accounted for. Thus the

Table 7
Cultural distance and nationalist vote. IV. Sample split.

	Δ Gen Dist>0		*Δ* Gen Dist<0	
	Lega (1)	XR (2)	Lega (3)	XR (4)
Gen Dist	0.0182** (0.0086)	0.0244** (0.0116)	0.0036 (0.0025)	0.0034 (0.0034)
Imm/pop	−19.0138 (12.2989)	−22.3874 (16.0181)	3.5522 (6.7397)	4.1713 (8.1506)
GDPPC	0.1982** (0.0911)	0.2631** (0.1193)	−0.0392 (0.0358)	−0.0292 (0.0451)
R^2	0.851	0.804	0.908	0.877
F (dist)	22.6	22.6	42.5	42.5
F imm	89.9	89.9	146.9	146.9
Obs	15418	15418	13068	13068
Cities	4013	4013	3352	3352

Notes: Dependent variable in columns (1) and (3) is the vote share for Lega, in columns (2) and (4) the vote share for the XR artificial coalition (Extreme-right). Gen Dist is the weighted average genetic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Imm/pop is the ratio of immigrants to total population (pct terms), instrumented with a shift-share instrument (see text). GDPPC is the weighted average GDP per capita at purchasing power parity of the immigrants' origin country. All regressions include: total population, unemployment rate, GDP adjusted for inflation, number of thefts and robberies 100k citizens, age dependency ratio, the percentage of population above 65, the average age, the percentage of the population with college (or higher)education, with high school education and with elementary or no education (excluded category: middle school education), the percentage of the population regularly attending religious services, the percentage of the population that does volunteer work, the rate of dismissed firms in the year and the rate of new firms in the year. See text for data sources. Municipality and election-by-region fixed effects are included. Standard Errors are clustered at the municipality level. F (dist) is the first stage F statistics for the cultural distance variable. F (imm) is the first stage F statistics for the immigrants' share. In columns (1), (2) and (3), the sample includes only municipalities with a genetic distance increase between the first and the last election-year. In columns (4), (5) and (6) the sample includes instead only the municipalities with a genetic distance decrease. Obs is the total number of observations. Cities is the total number of municipalities. *** significant at 1% level. * significant at 5% level. * significant at 10% level.

increased cultural distance between immigrants and natives is one of the determinants of the increased vote share for nationalist parties in Italy.

6. Robustness and extensions

To see if the interpretation of the results is correct, I split the sample according to the change in genetic distance between the last and the first election, and I run the regression separately for the municipalities with an increase in genetic distance and for the municipalities with a decrease. The results are summarized in Table 7. The coefficient on genetic distance is much bigger in the sample of municipalities with a genetic distance increase, while it is small and not statistically significant in the sample of municipalities with a decrease. In particular, the coefficient on Lega is 0.0182 in the sample with an increase, 4 times as big as in the baseline regression, while the coefficient on XR 0.0244, 4.3 times bigger. The conclusion is that the effect of cultural distance on nationalist vote is driven by what happened in the municipalities with a genetic distance increase. As for linguistic distance, I tried doing the same exercise, but the regression results in the sample with a linguistic distance increase are imprecisely estimated due to the small number of observations (933 municipalities only).

The results are robust when considering, as dependent variable, the vote share for an alternative artificial coalition composed by Lega and by more moderate, center-right, political parties (the full list is available in the appendix). The only notable difference is that, in the regression with religious distance, which is itself not significantly associated with the vote share, the coefficient on the immigrant share is positively and significant.

I performed the analysis looking at alternative measures of genetic distance. First, I tried an alternative aggregation at the country level based on plurality groups rather than on weighted averages, obtaining exactly the same results. Then, I considered the alternative microsatellite measures of relatedness between populations (Pemberton et al., 2013), which are based on the similarity between DNA tracts (microsatellites) rather than genes (genetic markers). The results were robust (details can be found in the appendix). For linguistic proximity, I tried the alternative measure based on linguistic trees as computed by Fearon (2003). Similarly to the religious distance measure, the idea is to classify languages according to families to then compute the relative number of common nodes for each pair of languages. The results for the XR artificial coalitions were robust, although the result for Lega were not (details can be found in the appendix).

Table 8
Cultural distance and nationalist vote: Lega. IV, robustness.

	City	2.5 km	5 km	7.5 km	City	2.5 km	5 km	7.5 km
Gen Dist	0.0046***	0.0034**	0.0045**	0.0074**				
	(0.0017)	(0.0018)	(0.0023)	(0.0031)				
Lang Dist					13.5097**	22.1159***	26.3589***	63.4654***
					(6.0303)	(7.1765)	(7.9867)	(19.0171)
Imm/pop	−4.0253	−1.7229	−1.0534	−1.2434	−6.3413	−10.0968	−9.4204	−24.9483
	(5.2953)	(6.8106)	(12.1934)	(17.2036)	(6.1209)	(8.2084)	(13.2713)	(20.9019)
GDPPC	0.0511*	0.0248	0.0139	0.0121	0.0103	0.0246	0.0167	0.0215
	(0.0293)	(0.0256)	(0.0215)	(0.0201)	(0.0211)	(0.0224)	(0.0211)	(0.0238)
R^2	0.894	0.895	0.896	0.895	0.892	0.882	0.889	0.872
F (dist)	106.8	105.4	93.4	112.9	93.2	82.2	68.2	38.9
F (imm)	180.3	183.4	123.6	191.3	206.1	193.7	188.6	184.1
cities, med	1	0.75	3.61	8.12	1	0.75	3.61	8.12
cities, min	1	0	0	0	1	0	0	0
cities, max	1	8	22	41	1	8	22	41
dist (km)	0	1.85	3.52	5.14	0	1.85	3.52	5.14

Notes: Dependent variable is the vote share for Lega. Gen Dist is the weighted average genetic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Lang Dist is the weighted average linguistic distance between immigrants and natives (see text), instrumented with artificial immigrant shares based on the *Spatial Migration Chain* hypothesis (see text). Imm/pop is the ratio of immigrants to total population (pct terms), instrumented with a shift-share instrument (see text). In the column city, the above regressors and instruments are relative to each municipality. In columns 2.5 km, 5 km and 7.5 km they are instead averaged over all cities within, respectively, 2.5, 5 and 7.5 km from each municipality. cities, med is the median number of neighboring cities for each distance. cities, min is the minimum number of neighboring cities for each distance. cities, max is the maximum number of neighboring cities for each distance. dist is the average distance, in km, of those cities (see text for more stats). All regressions include: GDP per capita at purchasing power parity of the immigrants' origin country (GDPPC), total population, unemployment rate, GDP adjusted for inflation, number of thefts and robberies 100k citizens, age dependency ratio, the percentage of population above 65, the average age, the percentage of the population with college (or higher)education, with high school education and with elementary or no education (excluded category: middle school education), the percentage of the population regularly attending religious services, the percentage of the population that does volunteer work, the rate of dismissed firms in the year and the rate of new firms in the year. See text for data sources. Municipality and election-by-region fixed effects are included. Standard Errors are clustered at the municipality level. F (dist) is the first stage F statistics for the cultural distance variable. F (imm) is the first stage F statistics for the immigrants' share. 28514 total observations for 7372 municipalities. *** significant at 1% level. * significant at 5% level. * significant at 10% level.

I considered two alternative values of d, respectively 2.5 and 7.5 km from the main municipality. The results are summarized in Table 8 for the case of Lega, together with the range and median number of cities within each distance and with the average distance. In all cases, the results are robust.

As an additional way to measure the changed composition of the stock of immigrants, I computed an immigrant fractionalization index. The logic is similar to Alesina et al. (2003), but, instead of using the population composition by ethnicity, I used the shares by nationality to compute the index. More formally, the immigrants fractionalization index H_{jt} is:

$$H_{jt} = 1 - \sum_{k=1}^{K} (S_{jt}^{k})^2 \tag{11}$$

where S_{jt}^{k} is the shares of immigrants from country k in municipality j and election t. To avoid endogeneity, I instrumented the shares S_{jt}^{k} with the same strategy that I used to instrument cultural distance, using, in particular, the shares $\hat{S}_{jt}^{k}$ (see Section 4.2). I found a positive and significant relationship between immigrants' fractionalization and nationalist vote, but the effect is quantitatively small. In particular, 1 std deviation increase of fractionalization increases the vote share for Lega (XR) by 0.9 (1.5) pct points (further details can be found in the appendix). One possible explanation of this result is that more fragmented immigrant communities are more prone to unrest and conflict. Another possibility is the fear of multiculturalism. I also tried including the fractionalization index in the main regressions, without any significant change in the results.

An additional result, is that the genetic and religious distance between immigrants and natives are negatively related to voters' turnout, although linguistic distance in not (see the appendix for details).

Finally I also studied the relationship between the votes for Movimento Cinque Stelle and immigration, since its populist propaganda might also have benefited from immigration. Since they competed in the last two elections only, I restricted the sample accordingly. I did not find any statistically significant relationship.

7. Conclusion

Previous studies identified a positive relationship between immigration increases and the votes for nationalist parties, in countries such as, among others, Italy, Denmark, Austria and France, although there has not been enough research yet on which factors matter the most in this relationship. In this paper I provide an encompassing study that disentangles the effects of different dimensions of immigration, focusing on cultural differences between immigrants and natives. To identify them, I provide an enhanced Bartik instrument that can deal with an increasing territorial diffusion of immigration and an increasing number of nationalities. The main result is that the increased cultural distance between immigrants and natives, and not the change in their number, is the main driving force behind the electoral success of nationalist parties. This electoral success, in turn, might also determine the implementation of more stringent anti-immigration and asylum policies. The relationship between the identity of the migrants and the type of anti-immigration policies implemented remains however unclear, and so it is the influence of the electoral system. Moreover, it is also unclear whether the election of anti-immigrants representatives might foster discrimination, for instance in the access to welfare. Both topics deserve future investigations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Compliance with Ethical Standards

No funding was received for this research.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2021.103781.

References

Alesina, Alberto, Devleeschauwer, Arnaud, Easterly, William, Kurlat, Sergio, Wacziarg, Romain, 2003. Fractionalization. J. Econ. Growth 8 (2), 155–194.
Barone, Guglielmo, D'Ignazio, Alessio, Blasio, Guido De, Naticchioni, Paolo, 2016. Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives political preferences. J. Public Econ. 136 (1), 1–86.
Borjas, George J., 2003. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. Q. J. Econ. 118 (4), 1335–1374.
Bratti, Massimiliano, Deiana, Claudio, Havari, Enkelejda, Mazzarella, Gianluca, Meroni, Elena Claudia, 2017. What are You Voting for? Proximity to Refugee Reception Centers and Voting in the 2016 Italian Constitutional Referendum. IZA DP, IZA, Bonn, 11060.
Brunner, Beatrice, Kuhn, Andreas, 2018. Immigration, cultural distance and natives' attitudes towards immigrants: Evidence from swiss voting results. Kyklos 71 (1), 28–58.
Card, David, 2001. Immigrant inflows, native outflows, and the local market impacts of higher immigration. J. Labor Econ. 19 (1), 22–64.
Card, David, Dustmann, Christian, Preston, Ian, 2012. Immigration, wages, and compositional amenities. J. Eur. Econ. Assoc. 10, 78–119.
Cattaneo, Cristina, Fiorio, Carlo V., Peri, Giovanni, 2015. What happens to the careers of European workers when immigrants take their jobs?. J. Hum. Resour. 50 (3), 655–693.
Cavalli-Sforza, Luca L., Menozzi, Paolo, Piazza, Alberto, 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, NJ.
Desmet, Klaus, Breton, Michel Le, Ortuno-Ortin, Ignacio, Weber, Sholmo, 2011. The stability and breakup of nations: A quantitative analysis. J. Econ. Growth 16 (2), 183–213.
Dinas, Elias, Matakos, Konstantinos, Xefteris, Dimitrios, Hangartner, Dominik, 2019. Refugee crisis increase support for extreme-right parties?. Political Anal. 27, 244–254.
Dustman, Christian, Vassiljeva, Kristine, Damm, Anna Piil, 2016. Refugee Migration and Electoral Outcomes. CREAM Discussion Paper CPD 19/16, Center for research and Analysis of Migration, University College London, London, UK.
Dustmann, Christian, Preston, Ian, 2001. Attitudes to ethnic minorities, ethnic context and location decisions. Econ. J. 111 (470), 353–373.
Dyen, Isidore, Kruskal, Joseph B., Black, Paul, 1992. An indoeuropean classification: A lexicostatistical experiment. Trans. Am. Phil. Soc. 82, 1–132.
Edo, Anthony, Giesing, Yvonne, Öztunc, Jonathan, Poutvaara, Panu, 2019. Immigration and electoral support for the far-left and the far-right. Eur. Econ. Rev. 115 (1), 99–143.
Facchini, Giovanni, Mayda, Anna Maria, 2009. Individual attitudes towards immigrants: Welfare-state determinants across countries. Rev. Econ. Stat. 91 (2), 295–314.
Fearon, James D., 2003. Ethnic and cultural diversity by country. J. Econ. Growth 8 (2), 195–222.
Gerdes, Christer, Wadensjo, Eskil, 2008. he Impact of Immigration on Election Outcomes in Danish Municipalities. IZA Discussion Paper 3586, Institute for the Study of Labor, Bonn, Germany.
Guiso, Luigi, Herrera, Helios, Morelli, Massimo, Sonno, Tommaso, 2019. Global crises and populism: the role of eurozone institutions. Econ. Policy 34, 95–139.
Hainmueller, Jens, Hopkins, Daniel J., 2014. Public attitudes toward migration. Annu. Rev. Political Sci. 17, 225–249.
Halla, Martin, Wagner, Alexander F., Zweimuller, Josef, 2017. Immigration and voting for the far right. J. Eur. Econ. Assoc. 15 (6), 1341–1385.
Harmon, Nikolaj, 2018. Immigration, ethnic diversity and political outcomes: Evidence from Denmark. Scand. Econ. 120 (4), 1043–1074.
Inglehart, Ronald, Baker, Wayne, 2000. Modernization, cultural change, and the persistence of traditional values. Am. Sociol. Rev. 65 (1), 19–51.
Mayda, Anna Maria, 2006. Who is against immigration? A cross-country investigation of individual attitudes toward immigrants. Rev. Econ. Stat. 88, 510–530.
Mayda, Anna Maria, Peri, Giovanni, Steingrass, Walter, 2016. Immigration to the U.S.: A Problem for the Republicans or the Democrats? NBER working paper no. 21941.
Mecham, R. Quinn, Fearon, James, Laitin, David, 2006. Religious Classification and Data on Shares of Major World Religions. Stanford University.
Mendez, Ildefonso, Cutillas, Isabel M., 2014. Has immigration affected spanish presidential elections results?. J. Popul. Econ. 27 (1), 135–171.
Ottaviano, Gianmarco I.P., Peri, Giovanni, 2006. The economic value of cultural diversity: Evidence from US cities. J. Econ. Geography 6, 9–44.
Otto, Alkis Henry, Steinhardt, Max Friedrich, 2014. Immigration and election outcomes — Evidence from city districts in Hamburg. Reg. Sci. Urban Econ. 45, 67–79.
Pemberton, Trevor J., DeGiorgio, Michael, Rosenberg, Noah A., 2013. Population structure in a comprehensive genomic data set on human microsatellite variation. G3-Genes/Genomes/Genetics 3, 903–919.
Sherif, Muzafer, 1953. Groups in Harmony and Tension. Harper, New York, NJ.
Slotwinski, Michaela, Stutzer, Alois, 2019. The deterrent effect of an anti-minaret vote on foreigners' location choices. J. Popul. Econ. 32, 1043–1095.
Spolaore, Enrico, Wacziarg, Romain, 2009. The diffusion of development. Q. J. Econ. 124 (2), 469–529.
Spolaore, Enrico, Wacziarg, Romain, 2015. Ancestry, language and culture. In: Ginsburgh, Victor, Weber, Shlomo (Eds.), The Palgrave HandBook of Economics and Language. Palgrave Macmillan, London.
Steinmayr, Andreas, 2016. Exposure to Refugees and Voting for the Far-Right: (Unexpected) Results from Austria. IZA Discussion Paper 9790, Institute for the Study of Labor, Bonn, Germany.
Tabellini, Marco, 2020. Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. Rev. Econ. Stud. 87 (1), 454–486.
Vertier, Paul, Viskanic, Max, 2018. Dismantling the Jungle: Migrant Relocation and Extreme Voting in France CeSifo Working Paper Series No. 6927.