MIGRATION CRISIS IN THE LOCAL NEWS: EVIDENCE FROM THE FRENCH-ITALIAN BORDER

Silvia Peracchi







Migration Crisis in the Local News: Evidence from the French-Italian Border. *

Silvia Peracchi, [†]

December 11, 2023

Abstract

The massive inflows of migrants across the Mediterranean has generated widespread political attention and backlash. This paper explores the impact of migrants' displacements at the EU's internal borders, due to militarized border push-backs and arising from the European migrant crisis in the 2010s. It investigates how these displacements affect both the local news market and the local political economy. To do so, it relies on a policy implemented in June 2015, whereby French authorities introduced militarized controls at the Italian frontier to redirect migrants and asylum seekers, originally intending to cross the border irregularly, back to the Italian territory. These dynamics created a quasi-experimental setting, where natives in the Italian region were unevenly exposed to pushed back migrants: those residing close to the French border experienced more directly the evolution of events. Using novel text and count data from local news in the interested areas of Liguria, Italy, between 2012 and 2019, this study finds that, following the border push-backs, media coverage of migration decayed with commuting distance to the border. Conversely, anti-immigrant discourse in the news exhibited a relative increase in areas least directly impacted by the border events. Exploring further this framing dimension, the results turn out to be shaped by readers' demand and to be closely associated with local news penetration. Finally, this study documents that voting preferences share a similar direction to news slant, while a related broad pattern also appears in hate-crime records.

Keywords: Media slant, EU borders, immigration, diff-in-diff

JEL classification: F22, L82, F50

^{*}I am thankful to M. Beine, Q. David, M. Fernandez Sanchez, C. Mahe, E. Murard, J. F. Maystadt, R. Poldermans, K. Tatsiramos, P. Van Kerm, S. Zanaj for precious comments and suggestions. I also wish to thank the participants and organizers of 12th European Meeting of the Urban Economics Association, the 2022 Seminar series on the Economics of Migration; the 15th RGS Doctoral Conference in Economics; the 2022 LISER Workshop on Globalization, Attitudes and Populism; the Sixth International Conference on Understanding Voluntary and Forced Migration; the International Conference of European Studies. This research project was part of the PhD thesis of the author, which was supported by the Luxembourg National Research Fund (FNR) (PRIDE15/10949242*). All outputs of this study are the sole responsibility of the author.

[†]IRES/LIDAM, UCLouvain. E-mail: silvia.peracchi@uclouvain.be

1 Introduction

Over the past decade, Europe has experienced an unprecedented rise in undocumented migration, with a remarkable peak between 2014 and 2016. Italy, in particular, saw a surge in arrivals of migrants and asylum seekers, departing from the Libyan coasts on the so-called "Central Mediterranean Route." In 2014 and 2015 arrivals along this route reached 170 thousand and 150 thousand individuals, respectively (European Border and Coast Guard Agency, henceforth "Frontex"). These events inflamed the public debate both nationally and at the EU level, leading to viral attention in the media discourse (Chouliaraki et al., 2017).¹ Several scholars attributed the widespread success of populist, anti-immigrant parties in various EU countries to the direct exposure of natives to the presence of, or proximity to, migrant inflows (Campo et al., 2021; Edo et al., 2019; Otto and Steinhardt, 2014). Moreover, the role of media has been identified as an important mechanism of diffusion and persuasion of attitudes and expressed political preferences (Bratti et al., 2020; DellaVigna and Kaplan, 2007; Djourelova, 2023; Gentzkow, 2006). Despite their persuasive potential, media companies do not select the message to spread at random. When exploring the news market in particular, competition and readership beliefs are key factors in determining the supply of information (Mullainathan and Shleifer, 2005).

This study is grounded in the context of the 2010s migration crisis and examines the formation of news along two dimensions. It first poses the question: *How do border controls and the resulting migrant displacement impact the local news characteristics?* Second, *Which news patterns, if any, match the local political economy?* In particular, this research relies on an exogenous shock in the number of migrants at the Italian border to examine how local news outlets changed their output. The shock of interest occurred in June 2015, when French local authorities systematically introduced militarized controls at the borders with Italy. This policy aimed at and resulted in substantial push-backs of migrants to Italian lands, which, for morphological reasons, funneled migrants predominantly to the coastal border into the town of Ventimiglia, Liguria (Italy). Figure 1 exemplifies the geographic connection between France and Italy. On the left, a physical map of France highlights how boundaries with Italy are shaped by the Alps. On the right, a physical map of Liguria outlines the regions' geography, showing how mountains make coastal areas the most accessible access points.²

 $^{^{1}}$ For a comprehensive review of the refugee crisis, with a focus on the reception policies from EU members, one can refer to Dustmann et al. (2017).

 $^{^{2}}$ As a complement, an additional map realized by NGO La Cimade is reported in Figure A.1. This map helps visualizing the dynamics of the border push-backs.



Figure 1. Physical maps of France and Liguria, Italy

Notes: On the left, a physical map of France. Source: Mapswire.com, https://mapswire.com/maps/france/. On the right, a physical map of Liguria. Source: Daniel Feher, https://www.freeworldmaps.net/europe/italy/images/liguria.jpg.

Before the border controls took place, many migrants located in Italy in precarious circumstances had taken advantage of the French-Italian gateway to reach other European destination countries.³ Once the policy was introduced, French authorities inspected migrants encountered along the entry points to France, and resettled those without a permit to stay in France to the Italian lands at the border (specifically, in the territory of the town of Ventimiglia). This study exploits the geography of border restrictions and focuses on the municipalities of two provinces situated in the vicinity of the border: Imperia, the closest, and its adjacent province, Savona. Although migration is a central theme in the national news landscape for the period considered, local news can still reflect variation in the salience of specific events and the related demand for information. For this reason, these are chosen as the preferred information output to analyze. Specifically, local news published online are considered, given the widespread internet availability and usage in the region.⁴ The analysis relies on a rich collection of about forty thousand articles from the quasi-universe of local news in the study areas. Articles are selected if they contain specific keywords, that is the stemmed versions of the words "migrant," "immigrant," "foreigner," "non-EU citizen" (It., "extracomunitario") or "displaced person" (It., "profugo"). Data are cleaned, then geo-localized via Google Trends and are combined to create a municipality-month-year panel dataset that spans the years 2012 to 2019.

With this selection, whether locals directly experienced the presence of resettled migrants depends (negatively) on how connected they are to the border, which in turn is a function of commuting time to get there. Therefore, proximity to the border, measured by the inverse of commuting distance, proxies the intensity of exposure to the border events, for individuals residing in the municipalities under study. The preferred specification is a difference-in-difference model that

³Other gateways included Bardonecchia, Como and Brennero (Capitani, 2018). Upon arrival in southern Italy, as described in Campo et al. (2021), migrants were subject to an allocation scheme created in 2014 to support the current receiving system of asylum petitions. Many intended to continue traveling, circumventing the bureaucratic process of status regularization in Italy (Capitani, 2018; Colombeau, 2020). For further information on how asylum seekers sorted in Europe, see also Aksoy and Poutvaara (2021).

⁴In the sampled period, the region of interest had a higher percentage of people using the internet for e-mail and news than the Italian average (see Figure A.6 for internet usage rates over time in the region of interest, compared to all Italy).

exploits the interaction between border proximity and a dummy for the period following June 2015. Several sensitivity checks are conducted to limit possible limitations of this type of analysis, as those listed in Callaway et al. (2021). A placebo replacing actual with random distances to other Italian cities is also added to the analysis.

The impact of the settlement of migrants on local news is examined along two key outcome variables: i) the coverage of migration and ii) the news' partisanship or "slant." These two important characteristics exist respectively as *agenda-setting* and *framing* in the taxonomy of the communication literature by Scheufele and Tewksbury (2007). For the second dimension, this study builds upon the literature on the measurement of partisanship in the news (Gentzkow and Shapiro, 2010; Groseclose and Milyo, 2005; Taddy, 2013), as well as advances in supervised machine learning for text analysis, to build an indicator of language slant. Throughout the paper, media bias and slant are used interchangeably, in line with the definitions of Gentzkow et al. (2015) for media bias—" systematic differences in the mapping from facts to news reports [...] which tend to sway naive readers to the right or left on political issues" (page 3)—and of Gentzkow and Shapiro (2010) for media slant-"the frequency with which newspapers use language that would tend to sway readers to the right or to the left on political issues" (page 36).⁵

Specifically, this paper uses an algorithm to determine whether a local news article has anti-immigrant inclination by comparing the similarity of its text with that of anti-immigrant versus more positively leaning national news outlets. This procedure builds upon the seminal works of Groseclose and Milyo (2005), comparing congressional citations of US think tanks with citations of those think tanks in newspapers, Gentzkow and Shapiro (2010) and Taddy (2013), which detected slant using the entire Congress and news language, using more sophisticated techniques. Here, classification of slant in local news is based on a carefully selected training sample, represented by a body of national and regional newspapers with well-known orientation. This approach enables the adoption of machine learning algorithms by relying on rich text contents.

The main findings can be summarized as follows. The backlog of migrants led to a notable prominence of migration in the local media discourse at the border. It caused a differential increase in news coverage that sharply decays with commuting time to the border. The share of anti-immigrant news, however, is observed to get relatively higher in the municipalities further away, after migrants' push-backs were introduced at the border. Further evidence proposes a set of mechanisms to rationalize this finding. First, demand for news is found to have a crucial role. This outcome emerges when comparing the results by giving more or less weight to the distribution of readership, which is proxied with Google Trends information. Additionally, where local news is found to gain market against national news, the slant effect is higher. According to the framework of Mullainathan and Shleifer (2005), readers have biased beliefs and media firms can accommodate these beliefs by slanting the information towards them. Where readers' opinions are more heterogeneous, their demand pushes slant in the two opposite directions, increasing news accuracy overall. Conversely, where consumers have homogeneous beliefs, bias is uniformly reflected in the news. In this context, direct exposure of locals to migrant settlement at the border appears to increase opinion heterogeneity, documented in voting patterns. These dynamics would push bias down in aggregate terms. This variability could have also been channeled by the area's recorded presence of humanitarian help from both local and international NGOs (La Cimade, Oxfam, Caritas and

⁵Importantly, the notion of agenda-setting can also be understood as a form of bias (Prat and Strömberg, 2013). This paper opts for the terminology of Gentzkow et al. (2015) and Gentzkow and Shapiro (2010) and treats media bias and slant as substitutes.

Doctors Without Borders, to cite a few). Ideology would remain more homogeneous the further away from the actual events, which contributed to the predominance of the anti-immigrant tone. The contact versus exposure paradigm put forth in Steinmayr (2021) would complement this explanation. Direct contact with the emergency might spark humanitarian reactions, in the communities closest to the border. As remote municipalities lacked a direct contact element, a general negative response appears to prevail. To further elaborate on these findings, a graph of hate crime incidents that were measured à la Romarri (2020) qualitatively demonstrates that hate crime trends increased more substantially in the farther province. According to studies such as Benesch et al. (2019); Djourelova (2023) and Keita et al. (2023), news powerfully influence readers' sentiments. This contribution departs from these works by letting supply shocks in media discourse be endogenous, and causally derived from shocks in migration presence. This result also relates to a rich and surging literature in migration economics that investigates how migrants' arrivals impact attitudes toward them (see Table B.1 for a set of these papers). The media component, key in this study, is touched on superficially in existing contributions. Finally, while relying on a novel context and original data, this work connects to a body of studies in political economics that use text-analysis techniques to analyze the dynamics of politically charged news content. A thorough review can be found in Gentzkow et al. (2019a). While analyzing a specific geographic context, the present study importantly generalizes to areas that experience border tensions in the context of migration. These phenomena are non negligible: in the European Union's context, internal borders have been object of political tensions between national governments in several occasions. This study sheds light on the political economy of internal borders. It urges for better border governance and cooperation, to ultimately prevent the rise of extremist views in areas that only receive partial and indirect information around migration. The rest of the paper proceeds as follows: the next section reviews the literature related to this contribution. Section 3 explores the setting and context of the study. Section 4 details the data and measurement, while the core analysis follows in section 5. Section 6 addresses the potential mechanisms and implications of the main results, continuing with a set of robustness tests. Finally, the last section provides concluding remarks.

2 Literature

This study adds to the empirical literature in migration economics that examines the consequences of asylum migration in developed economies. It also connects to a number of political science and political economics contributions on the market for information, and specifically to those investigating the impact of media exposure and media content on forming preferences and attitudes.

Related studies in media economics examine the determinants of news market outcomes. These contributions take into account the demand and supply of information in terms of quantity (coverage) and the type of discourse provided (slant). An extensive review of the theory is available in Gentzkow et al. (2015). Anderson and McLaren (2012) stress the role of mergers in news competition, while Gentzkow and Shapiro (2010) emphasize newspapers' incentives to tailor slant in the direction of the consumers' ideology. Focusing on local news, George and Waldfogel (2006), observe that national news penetration influences local news production. Focusing on the language of politics, Gentzkow et al. (2019b) propose a text-based measure of language partisanship in the US Congress that takes into account issues with finite-sample bias. The authors note partian differences in speech have been increasing over time in the American landscape, with a structural positive change sparked by the Republican takeover of Congress in the 1990s.

According to the communications literature, media coverage and slant are two key factors that influence how people perceive a subject (Scheufele and Tewksbury, 2007). The terms agenda-setting and framing were employed in the taxonomy for this body of literature to describe, respectively, i) the relation between the news coverage of an issue and its perceived importance by the mass audience and ii) the implications of different portrayals of a topic on audience perceptions. News delivery not only reflects but also shapes personal preferences. Prat and Strömberg (2013) overview research on the connection between the provision of information and political attitudes. Numerous studies uncover evidence of the media's persuasive effects on policymaking and political preferences by means of radio (DellaVigna et al., 2014; Strömberg, 2004), print (Gentzkow, 2009; Snyder Jr and Strömberg, 2010), TV (Campante et al., 2018; Facchini et al., 2017; Gentzkow, 2006) and the internet (Falck et al., 2014).⁶ Consequently, media can also impact migration-related outcomes. de Coulon et al. (2016) observe that higher coverage of migration and crime increases the return migration intentions of the involved minorities, as found for the case of Romanian immigrants in Italy, given state TV coverage of a crime committed by a Romanian immigrant in 2007. Exploring events involving the same migrant crisis as this paper, one of the conclusions of Battiston (2022) is the effective relationship between shocks on public attention and rescue efforts in the Mediterranean. Djourelova (2023) reveals that the Associated Press (AP) policy of forbidding the use of the term "illegal immigrant" resulted in milder views toward immigration among the AP's frequent readership. Keita et al. (2023) collect crime-related news content for Germany and take advantage of a local news outlet policy change, to find that consistently displaying criminals' origins, especially when offenders are natives, makes natives less opposed to immigration. In contrast to the work of these scholars, here the question is posed in reverse: news outlets are allowed to change their production endogenously and emphasis is placed on how migration-related events impact the news market.

Finally, this study relates to a growing body of research on how immigrants' presence or proximity affects on the local attitudes and preferences. In Alesina and Tabellini (Forthcoming), the views of destination countries on immigration can be framed along two intrinsic dimensions: an economic and a non-economic, cultural one. The authors note that while immigration inflows are often found to provoke a stronger right-wing backlash and conservative ideology, a number of contributions instead observe a positive reception or more nuanced effects, found prevalently in cases where natives are exposed to newcomers with a repeated interaction. The present research focuses on the 2010s, which saw the EU refugee crisis becoming a priority on the political agenda of EU countries. Table B.1 reviews the findings of some of the key contributions in this expanding literature, examining the links between immigration shocks and the political outcomes they influence.

Among these studies, Steinmayr (2021) focuses on Upper Austria and uses the Austria-Germany border to discriminate between municipalities that hosted asylum seekers and had frequent interaction with them and municipalities that only experienced their transit. The author posits that repeated interaction with the new minorities will promote native acceptance, which is broadly in accordance with the contact theory of social psychology (Allport, 1954; Pettigrew, 1998).

 $^{^{6}}$ See also the work of Iannelli et al. (2021) who consider several media channels and describe the dynamics in the distribution of sentiments towards immigration in Italy around the 2018 election campaign.

Conversely, exposure without contact discourages exchange and breeds hostile sentiments. The empirical findings show that there is this ambivalence in the case of Upper Austria. Campo et al. (2021) explore the effects of the Italian refugee crisis on voting for anti-immigrant parties within the Italian context. Focusing on the national dispersal policy, strongly reinforced in 2014 by the Italian Ministry of Internal Affairs, the authors find that anti-migrant voting, as measured by the difference between 2018 and 2013 election outcomes, concentrates in municipalities that hosted migrant centers. Similarly, Bratti et al. (2020) observe that proximity to municipalities receiving asylum seekers increased support for populist parties in the referendum results, thus reducing mainstream left-wing support.

This paper departs from these contributions in several aspects. First, the geographic dimension differs. The interest here looks into an internal border of the European Union. Cross-country borders within the EU area were important settings in the light of the migratory crisis,⁷ however they have been receiving less institutional attention than the exterior EU borders.⁸ The current article also emphasizes as a critical outcome a better understanding of the media market, which is considered more superficially in the literature. Importantly, the focus on a localized setting results in an identification strategy that enables a causal in-depth investigation of this crucial yet underexplored information channel.

3 Context

3.1 Italy's Migrant Crisis and the Border Controls

During the 2010s, thousands of undocumented migrants landed in Italy with the intention of rejoining family members and contacts in other European countries. This surge was exacerbated by several political contingencies in developing countries, a key event being the "Arab spring" in the MENA regions. ⁹ In 2014, the Italian government enacted an allocation policy to strengthen a pre-existing, insufficient reception system for asylum migration. However, several migrants attempted to continue the journey instead of having their status regularized (Capitani, 2018; Colombeau, 2020).¹⁰ A typical gateway employed by migrants was located on the coastal border between France and Italy, in the north-western region of Liguria.¹¹ The geography of Liguria, a short strip of land, extends east and west (Figure 2 displays its localization). On its northern boundaries are the Maritime Alps and Ligurian Appennines, while Mediterranean Sea is on the South. Mountains delineate the French border region as well, making the coastal border town of Ventimiglia a crucial point of contact with France.

Following peaks in arrivals in southern Italy, the French local authorities introduced a militaristic border control in June 2015. This was done to push unauthorized migrants who attempted to reach France through the coastal gateway back to Italy, specifically to the border town of Ventimiglia. In the immediate wake of the policy's implementation, the local charity Caritas Intemelia recorded two hundred migrants seeking refuge on the streets; nonetheless, these numbers

⁷see Cimade (2018).

⁸Frontex only collects data on illegal border crossings at the external EU frontiers. https://frontex.europa.eu/we-know/migratory-map/ ⁹The international Organization for Migration (IOM) reports 62692 undocumented migrants entering Italy by sea in 2011, compared with

⁴⁴⁰⁶ and 9573 respectively in 2010 and 2009.

¹⁰Once landed in Southern Italy, migrants could submit an asylum application. As importantly mentioned in Campo et al. (2021), this procedure is remarkably lengthy.

¹¹Others were Bardonecchia, Como, Brennero (Capitani, 2018).

continued to rise during the summer months. Due to the occurrence of terrorist attacks, in late 2015 France declared a state of emergency, thus officializing the border closure. After a second peak in May 2016, several NGOs came in support of the emergency, and the Italian provincial authority opened a temporary Red Cross reception center. In May 2016, the Italian central government established a bus system to transport some of the migrants from Ventimiglia back to the major migration hubs in southern Italy. Nevertheless, Ventimiglia's border only saw a significant decline in arrivals in late 2017 as a result of a set of policies at the EU and national level which effectively decreased the number of migrants attempting the Central Mediterranean Route. In particular, Italy had signed a Memorandum of Understanding with Libya in February 2017 (which was later extended) with the intention of cooperating with Libyan coast guard authorities to prevent the boat lifts. Sourced by NGO La Cimade, Figure A.1 is a good summary of the time line and geography of the main events.

Structured data are lacking on the presence of migrants in Ventimiglia. However, in 2017 NGO Caritas Internelia collected a count of meals distributed to migrants. Figure A.2 in the Appendix reports a peak of 12500 migrants transiting in the area in July 2017. For comparison, Ventimiglia's population was approximately twenty-four thousand residents in 2019 (Italian National Institute of Statistics [ISTAT]). Despite measures being taken, the situation persisted over time. With some seasonality, migrant arrivals in Ventimiglia were also observed in 2018 and 2019.

Figure 2. Area of interest, Liguria



Notes: A reference map of the region of interest displays a close-up view of Liguria on the left side. Ventimiglia, a border town, is located in the west. Capital cities of each province in the region are labeled, including Imperia, Savona, Genova (the regional capital), and La Spezia. Imperia's province is shown in blue, and Savona's province is in green. Provinces in grey are not part of the sample. On the right side, there is a map of Italy with Liguria highlighted in blue.

3.2 External validity

To analyze the impact of migrant settlements occurring at the western borders in Italy, this study concentrates on two provinces at the border region, namely, Imperia and Savona. These two districts are located respectively at the border and at its immediate vicinity. This restriction aims at maintaining the claim of internal validity as robust as possible, for instance by minimizing the influence of other unobserved events related to the broader migration crisis in Italy. Commuting distance from the border town of Ventimiglia proxies the likelihood that locals are subject to the migrants' settlement. The strength of this proxy is due to the particular shape of the region, that makes the coastal border a crucial point of connection with France. This relationship would eventually fade away if very large and heterogeneous areas are considered. At the same time, it's important to investigate how the findings of this analysis can be also externally valid, and so applicable to more general contexts beyond these locations. Figures A.3 and A.4 compare the region in the sample with the distributions of several socio-demographic variables for all Italian provinces, estimated by kernel densities. The dashed vertical lines on the graphs indicate Imperia and Savona, in red and blue respectively. As shown in Figure A.3, Imperia has a relatively higher percentage of cross-border commuters. This aspect highlights that closeness to the border indicates a greater connection in terms of workers mobility, thus justifying the choice of commuting distance for the treatment. Regarding the other variables, the two provinces are comparable to the rest of Italy. This holds in terms of total shares of commuters, employment rates, human capital distributions of the employed population, shares of immigrant population. Regarding the main immigrant origins, the areas of study present higher shares of population from the American continent.

Figure A.4 depicts vote patterns from 2008 and 2013 for the Chamber of Deputies. With respect to the rest of Italy, the two provinces have comparable values of turnout and appear to lean more on the right of the political spectrum, with Imperia experiencing higher levels of right-wing incumbency than Savona.¹² Besides these trends, a right-wing shift did appear in the area of study in the post-migration crisis elections of 2018. The right-wing coalition obtained on average 28.97% of votes for the Chamber of Deputies in 2013, and this increased to 44.94% in 2018. Left-wing coalition votes decreased from 23.89% in 2013 to 18.08%. Appendix section C elaborates the political context of Italy in greater detail. To better understand the role of some of these patterns, results section 5.1.3 investigates how the main effects depend on initial conditions, such as political preferences, as well as other socio-demographic characteristics.

3.3 Treatment and Identification

The border closure in June 2015 deterred migrants from crossing over to France. In a number of reports, the location was dubbed as "Italian Calais" because of the similar bottleneck situation and precarious living conditions that were endured by those migrants trying to enter the UK by sea from France. The topography of the French-Italian border made it difficult for displaced people to locate easy alternative ways to cross the border. Although some immigrants thought of taking a different route to the north across the mountains, this option was relatively unpopular because it was arduous and perilous (Welander, 2017).

It is important to highlight that both before and after the border policy, migrants were transiting the entire sample region. Once the border closed, however, the border area was suddenly exposed to groups of displaced people in precarious conditions, sheltering in and around the town. At the same time, natives of the area were in more or less direct contact with the border, depending on the geographic proximity of their location. For instance, as discussed in

¹²For both provinces, 2013 constituted a relatively high success for the newly born anti-system, web-spread movement Movimento 5 Stelle (M5S).

the paragraph below, cross-border workers reside closer to the border. Therefore, the identification strategy adopted here is based on the likelihood of natives to be exposed directly to the border events. This likelihood fades away the more distant their municipality is from the border town. Given this, the preferred treatment measure is the inverse of commuting time, because it appropriately captures this actual proximity, reasonably more than aerial distance.

Data on commuting distances are sourced by ISTAT, which calculates drive times from the centroids of each municipality, considering different road types and assuming an ideal condition of zero traffic. Here, the degree of proximity interacted with the post-June 2015 time interval determines the treated areas.

The validity of this approach for causal inference hinges on the comparability of municipalities in terms of distance, assuming that, in absence of the border-related events, these municipalities would have followed parallel trends in the outcomes under study. To ensure this comparability, the empirical approach will start by assessing the balance of the sample. It will consider various covariates, conduct several robustness checks, it will look into pre-policy effects in an event-study graph. It will account for different province-specific time effects and will propose a placebo that exploits distances to other areas in Italy, extracted randomly. Further details are provided in the following sections.

4 Data and Measurement

Data collection began by web-crawling an initial corpus of 59,489 online local news articles (42,762 for the period 2012 to 2019). Specifically, this collection targeted news pieces that contain migrant-related keywords (details follow) and represents the quasi totality of local online outlets present in the two provinces of interest at the time span considered. To ensure data quality, a comprehensive data cleaning process filtered out articles captured by the keywords, but unrelated to migration. Subsequently, the corpus was aggregated at municipality, month-year level, a critical step for facilitating the geolocation of news events. More details on these aspects follow in this section. Further information is relegated to the Appendix sections D, and E, for the sake of concision.

4.1 News Data

The web-crawling process to collect news individually scraped each relevant news outlet's website. Articles were gathered if any of the stemmed variants of the following keywords appeared in the text: "migrant," "immigrant," "foreigner," "non-EU citizen" (It., "extracomunitario"), or "displaced person" (It., "profugo"). Additionally, a second web-crawling algorithm extrapolated the overall news count in the outlets over time. All local news sources that could be located somewhere within the region and time period of interest were selected. Importantly, this includes outlets that appeared after the policy went into effect. This choice was motivated by the goal to produce statistics representative of natives' exposure to news at the municipality level. ¹³ Figure 3 displays the names and counts of news outlets in the sample, with a distinction in the metric using blue for before June 2015 and red for after June 2015. To match data availability in the set of control variables, the final dataset covers the period between January 2012 and December 2019. In sum, 42,762 articles are present from 2012, coming from a total of fifteen distinct news sources. Google searches and Google

 $^{^{13}}$ Data inaccessibility excluded sources that had been discontinued at the time of data collection. To the author's knowledge, however, their number is negligible.

Trends were employed to extrapolate the universe of the local online newspapers in the provinces. Specifically, from an initial set of local news outlets, Google Trends would potentially reveal new outlet names in its correlated searches tool. For example, searching in Google Trends for the online outlet *Sanremonews*, would return correlated searches from users in the region for *Riviera24* and *Prima La Riviera*, which are outlets of interest as well. Articles matching exclusively with the keyword "foreigner" also captured aspects different than migration (for example tourism). Their content was analyzed through a latent Dirichlet allocation (LDA) approach (Blei et al., 2003; Hornik and Grün, 2011) and led to drop around 10 thousands unrelated articles.¹⁴





Notes: Distribution plot of the news in the sample, by news outlet. Blue bars represent data for the pre-treatment period while bars in red represent post-treatment frequencies.

An article's text is composed of its title, subtitle, and full body content. Some metadata was also extracted, including the author's name (when provided), the article's date, and a tag (if available). Local outlets do not serve the sample are equally. Although online news is, in principle, accessible without any geographic boundary, local news firms tend to serve a particular area by tailoring their content and advertising revenues to the readers within their coverage. This implies that the geographic distribution of each local outlet may vary, and it is crucial to know this information to determine which news is pertinent in each location.

The location of the local news markets is unobserved. However, as a proxy, this study uses Google Trends, by tracking Google searches for the names of the outlets over time and by municipality. Google Trends returns i) town ratings on the volume of searches for any term, such as "Sanremonews,", during a certain time interval. The scale ranges from 1 to 100 and is constructed by normalizing scores with respect to all Google searches combined. The platform also provides ii) a rating over time based on search traffic for any term, with the option to compare it to up to four other

¹⁴Details on the procedure are found in the Appendix section E.1.

terms. This rating spans from 1 to 100 and is normalized across sources by taking a common comparison term.

With this knowledge, news coverage of migration in baseline estimates simply implies the log count of articles matched to a municipality m in a given month-year t. The exploration of agenda-setting dynamics will use this as the baseline outcome measure. Each text-based measure Y_{mt} for municipality m at time t is resulting from a weighted sum of y_{imt} values for articles i associated with municipality m from Google Trends, as in equation 1.

$$Y_{mt} = \sum_{i \in N} w_{it} y_{imt} \tag{1}$$

Three different geo-localization techniques are applied to assign the distribution of weights w_i across municipalities and time:

- Google searches for each outlet, by municipality, varying in the before versus after time range are employed to
 match articles with the municipalities they serve. The construction of weights for each article to each municipality
 is detailed in Appendix section D. Importantly, measures derived with this matching reflect how demand (search
 traffic) evolves over time for these news items, both pre and post-border-closure events of June 2015. The baseline
 estimations adopt this measure.
- 2. Google searches for each outlet, by municipality, before June 2015 are employed to match articles with the municipalities they serve. The construction of weights for each article to each municipality is detailed in the Appendix section D. Importantly, measures derived with this matching reflect the demand (search traffic) for these articles only before border closure. This procedure is used in the extended analysis to check whether results are driven by a change in the distribution of demand for these outlets.
- 3. No Google Trends data. To shut down the demand component present in the procedures above, this alternative geocoding procedure assigns news articles to municipalities based on the location of the outlets' headquarters. Each outlet is allowed to cover municipalities within a distance of twenty kilometers (an alternative fifteen and twenty-five-kilometer range were also tested for robustness). The local news source's articles will all carry the same weight in all locations within this radius. This procedure is used in the extended analysis to check whether results are affected by articles-municipalities in which (pre-existing) demand is strongest.

Google Trends censors locations below a certain level of search traffic. To circumvent this issue, an imputation process that uses a first neighbor technique and information about commuting times (Italian Census 2011, ISTAT) extends the data.¹⁵ Details are presented in Appendix section D. Additionally, Google Trends searches for the two most popular sources—*Sanremonews* and *IVG*—are contrasted with searches for a set of alternative, reasonably popular internet search terms—"accident" (It., "incidente"), "pizzeria," "recipe" (It., "ricetta")—for the region of Liguria. This is done to make sure that the readership size for the local outlets collected is of economic relevance. The search comparisons are plotted over time in Figure A.5. The graph shows that local news is very much comparable to this set of search items, which suggests that the local readership is substantial. Figure A.6 complements this information by plotting trends on internet usage in Liguria, compared with the rest of Italy, and based on ISTAT statistics.

 $^{^{15}\}mathrm{A}$ robustness test confirms the main findings without this imputation step.

Some of the main indices of interest rely on the content of news, assessed with text-analysis techniques. The next section provides a detailed description.

4.2 Text-Based measure of framing

The document text for each article in the corpus is made up of the article's title, any headers or subtitles, and the text itself. The content is cleaned with pre-processing practices which are common in text-analysis applications.¹⁶

The text-based measure that follows is based on a supervised machine-learning technique of binary classification. Text is treated following a *bag-of-words* approach.¹⁷ These procedures entail creating a matrix that provides term frequencies by documents; these matrices are known as *document-term matrices*, and they are constituted by documents in their rows i and terms in their columns j.

In line with Gentzkow and Shapiro (2010), this study uses bi-grams and tri-grams (phrases of two and three words) rather than uni-grams (single words) as the terms of interest to calculate frequency statistics. Using combinations of more than one word usefully captures some word dependencies that would be otherwise discarded. The frequency of n-gram j in document i is specified as tf_{ij} =term frequency. For n-gram j, its Inverse document frequency= idf_{ij} is defined as the log of 1 over the share of documents containing n-gram j. The product of the term frequency and the inverse document frequency is termed $tf - idf_{ij}$ and is the metric used here. The advantage of this composite measure is that, unlike simple counts or frequencies, it avoids overly weighting extremely rare and extremely common n-grams.

A measure of media bias or slant toward right-leaning,¹⁸ anti-immigrant speech is used to evaluate the framing dimension in the local news language.

Articles in the corpus are classified, or labeled, into a dummy variable that takes a value of 1 if the classification procedure predicts the article to be anti-immigrant oriented, and 0 otherwise. This prediction is based on a supervised machine learning approach. This method relies on a training sample, that is formed by a pre-classified text. The training sample's text frequencies are then used to create a predictive model for the dataset's local news items (i.e., the unlabeled data). Training data derive from a balanced collection of regional and national news articles for which evidence exists that they express or not an anti-immigrant orientation. This evidence is a combination of survey and anecdotal knowledge, whose details are given in Appendix section E.2. News items classified as 1 in the training sample belong to newspapers *La Verità* and *Il Giornale*, while 0-labeled text is extracted from *L'Unità News* and *Il Secolo XIX*.

¹⁸The rationale for focusing on right-wing partianship comes from the context of interest. As several authors point out (Campo et al., 2021; Bratti et al., 2020), anti-immigration positions in Italy emerged from right-wing, populist parties (*Lega* and *Fratelli d'Italia* primarily, though these parties integrated the right-wing coalition during the years covered in this study).

¹⁶Text is pre-processed with the removal of punctuation, numbers, and *stop words*-words that are frequent in texts and serve as building blocks for sentences, but do not add to the sentence content (e.g. *and*, *in*, *of*, *with*, *to*,). Then, words are stemmed to remove their endings (Bouchet-Valat, 2020). Upper case (proper) names are removed (except if at the beginning of a sentence), and all words are set to lower case.

¹⁷This approach constitutes a simple technique to represent a document as a function of its text. In the simplest version of this method, singular words and their frequency are the units of interest. The name *bag-of-words* derives from the fact that information on the words' ordering is discarded. As a straightforward extension, scholars considered replacing single words with pairs and triplets of words. To some extent, this complication allows tracking word combinations that better represent a sentence's meaning. See Gentzkow et al. (2019a) for a detailed review of these and alternative methodologies.

To collect migration-related articles from these sources, the same website-scraping method used for local news was adopted.

Two alternative binary-classification models are tested: i) a regularized logistic regression with elastic-net penalization and cross-validated tuning parameters (Friedman et al., 2009); ii) an inverse regression model by Taddy (2013). A cross-validation step evaluated the performance within the training sample and an additional step compared the model predictions models with human annotations. The hand labeling was performed by *Amazon Mechanical Turk* annotators. Validation results suggest similar predictive performance for the two alternative models, with a slight preference for the use of Taddy (2013)'s measure, which will be adopted in the baseline analysis. More details about the algorithm performance are presented in the appendix section E.3.

In the Appendix, Figure A.7 predictions resulting from this algorithm are explored. The figure reports frequent terms that return the lowest and highest predictive probabilities of containing anti-immigrant language. The expressions in orange at the top, which least predict anti-immigrant language, are mild expressions to refer to migrants (e.g. "migrating person," "migrating refugee") or terms related to migrants' rights (e.g. "fundamental rights," "protection system"). Terms in the most likely anti-immigrant news (in violet) relate, for example, to legal issues (e.g. "preliminary hearing," "plaintiff"), and expressions of concern on migrant flows (e.g. "uncontrolled immigration") or welfare (e.g. "layoff," "public debt"). Table 1 accompanies the figure with a small list from local headlines in the sample. Appendix section E.2 digs into the construction of this text-based measure in more detail and provides information on the validation procedure for this variable. Finally, as mentioned above in the geo-localization section, the classified articles are combined into a municipality-level metric that can be read as the proportion of migration-related local news that is classified as containing anti-immigrant language.

Title	Source	Period	Label
En., They sneak onto "commuter trains" to return to Italy after being expelled: two arrests	riviera24.it	Jan 2012	1
It., Si intrufolano sui "treni dei pendolari" per rientrare in Italia dopo l'espulsione: due arresti			
En., Giovanni Toti intervenes on the migrants matter: "We should not allow that Italy becomes a big	sanremonews.it	Jun 2015	1
refugee camp for Europe"			
It., Giovanni Toti interviene sulla questione migranti: "Non si può consentire che l'Italia diventi un grande			
campo profughi dell'Europa"			
En., Ventimiglia: police reports two migrants after a theft in Carrefour supermarket	Imperia Post	Nov 2018	1
It., Ventimiglia: la Polizia denuncia due migranti dopo un furto al supermercato Carrefour			
En., Call from the regional council to promote the diffusion of different cultures	bordighera.net	Jan 2012	0
It., Bando della Regione Liguria per promuovere la diffusione delle diverse culture			
En., Women, migration, and entrepreneurship, a congress at the Priamar	ivg.it	Mar 2015	0
It., Donne, migrazioni e imprenditoria un convegno al Priamar			
En., Ventimiglia: another 144 packed meals were delivered today by the Red Cross to migrants	sanremonews.it	May 2016	0
It., Ventimiglia: altri 144 pacchetti alimentari consegnati oggi dalla Croce Rossa ad altrettanti migranti			

Table 1. Anti-immigrant slant: Selected examples of the results

Notes: Examples from the predictions of Taddy (2013)'s model on the local news sample.

5 Empirical Strategy

This study adopts the following baseline model:

$$Y_{mpt} = \alpha_m + \delta_t + \beta_1 InverseDistance_m * Post_t + \Gamma X_{mpt} + \epsilon_{mpt}$$
(2)

where, for municipalities $m \in \{1, ..., 132\}$ located in province $p \in IM, SV$ and month-year $t \in \{Jan2012, ..., Dec2018\}$, we observe media outcomes Y_{mt} . Month-year and municipality fixed effects are included. β_1 constitutes the coefficient of interest and corresponds to the interaction term between the post-June-2015 dummy, and border proximity, measured with the inverse of commuting distance (in minutes) to the border area. In this sense, the degree of treatment corresponds to the likelihood of individuals living in municipality m to be directly exposed to the border policy and migrants' settlement. The degree of treatment decays with distance at an accelerating rate.¹⁹

 X_{mt} is a vector of controls that vary with time and municipality. To isolate the effect of distance from other geographic covariates, all specifications include an interaction between the post dummy and latitude, longitude and maximum latitude. In terms of news information, these involve the log of total news published in the local outlets, an average word count per outlet, and the count of outlets per municipality. Sourced by ISTAT, demographic variables are population (in logs), population density and the share of population above 65 years of age, immigrant population stocks (in logs), as well as inflows and outflows. Additional covariates are the average taxable income per capita, extracted by tax registers deriving from the Home Office and, varying at province-year level, a crime-rate variable. Some specifications also control for the time distance from the administrative and general elections. The dates and years of local elections are not uniform in the sample, thus a variable is created counting the months to the next election for each municipality. Further controls, only available cross-sectionally, are the unemployment rate, the rate of highly educated population, altitude levels, and superficial area, all collected from the 2011 Census (sourced by ISTAT). These are included in the balance analysis as well as in a set of extended results. Table B.2 presents the summary statistics for these variables.

Balance-test statistics are presented in Figure A.8, to look for pre-existing differences in the levels covariates and outcome variables along proximity to the border. The estimates result from an linear regression of the variable on inverse distance on the pre-policy subsample, accounting for time and province fixed effects. All variables are standardized to ease comparability. The table indicates that some differential in levels existed in the period before border closure for some variables. Population and mobility are higher closer to the border. Old-age shares are more frequent further from the border, as are maximum altitude and municipality's total area. Also the two outcomes of interest present slight but significant level differences. To circumvent potential issues arising from these pre-existing differences, observable characteristics that vary with time are included as controls in the model, while time-invariant covariates -unemployment, education, altitude and initial political preferences- are further explored in an extended analysis, to investigate any heterogeneous effects depending on their levels. It is crucial that level differences do not imply differential trends, which would harm the validity of diff-in-diff results. Several steps of the analysis explore these concerns. To start, Figure 4 plots the evolution of Y_{mt} means over time, by commuting distance (shown in quarters before and after the French

¹⁹More precisely, the magnitude of this decay is proportional to the square of distance. For robustness, a set of extended results allows for different functional forms in the distance-based treatment.

border closure, represented as 0 in the x-axis. News coverage appears on the left and anti-immigrant slant on the right. Means are computed for each time period and along spans of five minutes of commuting distance. The darker the line, the further the distance from the border. The picture delineates a clear switch in the link between distance and news coverage at the moment of the border events. A similar switch is also observable for anti-immigrant slant. These descriptive patterns highlight a change in the news measures, resulting from the border closure. This change is assessed more structurally in the econometrics sections that follow.



Figure 4. Mean outcomes over time, by treatment

this figure, means over time are plotted for the outcome variables Y_{mt} . The 0 level in the x-axis corresponds to the second quarter of 2015, which is the moment in which events took place.

5.1 Results

5.1.1 News Coverage

To what extent did the border controls translate into a sizeable media event? To answer this question, the top coefficients of Table 2 present the estimation results on news coverage - that is, the log count measure of migrant-specific news items.

In the table, controls are gradually added from left to right. News controls include the number of total news (in logs), as well as an average word count of the articles. These are included in all specifications. From Column 2 on, the specifications also control for the time distance from the administrative elections, together with population controls. These last involve the log of population, its density (population per squared kilometer) and the share of population over sixty-five years of age. Migration controls are added in Column 3 and involve the log of immigration stocks and the logs of population inflows and outflows. Crime rates at province level and income per capita are included from column 4. Column 5 includes the count of local outlets present in the municipality m at month-year t. The count of outlets could be a function of the formation of new demands and supplies of outlets as a response to the salient events at the border, therefore it may vary endogenously, but it is added in one specification because of its interest as a mechanism, as it allows exploring whether media effects are driven by changes in appearing sources versus changes occurring within sources. Finally Column 6 is an alternative specification, with controls equivalent to Column 4, which

adds province-specific time trends. This specification will be the preferred one to carry over in the extended results and robustness tests.

At the top of the panel, the interaction coefficients of interest are negative and highly significant, suggesting that news coverage increased relatively more in the border's vicinity. Regarding the size of the coefficient, we can interpret the output as follows. After the border events, increasing proximity to the border by 0.1 units corresponds to an increase in coverage by a factor of $e^{0.0984}$, or around 10%. Because proximity is measured with the inverse of commuting distance, the effects of interest are a function of commuting distance. Moving from 10 minutes to 5 minutes away from the border increases migration coverage by around 10%. Moving from 20 minutes away from 15 minutes away corresponds to a change of around 0.0167 proximity, which implies an increase in coverage by 1.65%. Such results indicates a significant effect of the settlement of migrants on migration coverage in the news, effect that decays rapidly with increasing distances.

To dig into the persistency of this impact, and to test for pre-existing linear trends, and event study is carried out and a plot of its output appears in Figure 5 a. The coefficients relate to the specification in equation 3, where the effect is allowed to vary by quarter:

$$Y_{mpt} = \alpha_m + \delta_t + \sum_{\tau=Q1,2012}^{Q4,2019} \beta_t I[\tau = t] * InverseDistance_m + \Gamma X_{mpt} + \rho_p * t + \epsilon_{mpt}$$
(3)

where m, p, t and ΓX_{mpt} are defined as in equation 2. The term $\rho_p * t$ captures province linear time trends. The set of controls is equivalent to Column 6 (i.e., the richest set, excluding the outlet counts, to avoid controlling for a media-related outcome), and taking the quarter before the events (Q1 2015) as the reference period. Importantly, with respect to Column 6, the event-study graphs additionally absorb out municipality-level seasonality. This term's addition allows to condition on seasonal variation that would otherwise give rise to some significant pre-treatment coefficients.

As Figure 5 a highlights, the effect of higher coverage closer to the border persists several periods after the border events. Some fluctuation appears, with some spikes in some quarters of 2016, and some fading in the most recent periods. Besides these fluctuations, the other coefficients appear similar in size to the baseline results.

5.1.2 News Slant

Baseline results on anti-immigrant slant are presented at the bottom of Table 2. As anticipated, the outcome of interest is the proportion of news items in municipality m that are predicted to be anti-immigrant. Estimation results of Table 2, Column 6, indicate that, as a result of June 2015's events, slant in local media got relatively higher in areas less directly involved, as located further away. An increase in proximity by 0.1 units corresponds to a decrease in slant by 0.2665 standard deviations. Again, increasing proximity, or inverse distance, means, for example, going from 10 to 5 minutes away. Going from 20 to 15 minutes away would imply reducing slant by 0.0444 standard deviations

Figure 5 b shows results from the event study specification as detailed for the media coverage results. In the plot, pre-policy coefficients are not significantly different from zero and the post-policy coefficients have negative point estimates. While the coefficients on the right hand side of June 2015 display some fluctuation in their significance, the effect appears relatively persistent, fading away in the latest periods. Several sensitivity analyses for baseline results are examined in Section 6.

Table 3 presents robustness checks with respect to some variations in the slant measurement.

In the first column of Table 3, the alternative classification method by Friedman et al. (2010) is employed.²⁰ In the baseline measure, a label of 1 is assigned to an article, if the predicted probabilities obtained in the predictive model are higher than 0.5. The second and third columns employ alternative threshold probability scores of 0.55 and 0.45 to assign slant labels to articles. Results from these specifications stay in line with the baseline, both in significance and size.

Proceeding to the right, an alternative specification directly considers probability scores, rather than labels, as the article-level outcome to aggregate at municipality level. Results are presented for both the benchmark index and for the index based on Friedman et al. (2010), leading to similar results. Figure 6 complements these robustness tests by showing estimated coefficients from slant classification based with probability matching. The variability is resulting from 1000 iterations. ²¹ In the graph, one point in each vertical segment corresponds to the estimated effect based on one out of 1000 regressions (each regression including a different iteration of the probability matching). When incorporating the classification error in this way, we find a similar visual pattern to the baseline event study. This suggests that results are robust to several measurement variations.

 $^{^{20}}$ See also Friedman et al. (2009) detailing the R package employed to classify news according to this method.

 $^{^{21}}$ According to this method, every article is assigned a number from a uniform distribution in the [0,1] interval. If the predicted probability is above this number, the article is classified as 1, and 0 otherwise. The procedure is then repeated M=1000 times and the average class across iterations the same article would converge to the predicted probability. While the baseline hard classification is the optimal decision-rule to map predicted probabilities into labels in terms of accuracy, this method allows to capture and carry over the uncertainty that results from the classification.

	Dependent variable: Migrant news coverage (log)						
	(1)	(2)	(3)	(4)	(5)	(6)	
Inverse distance (min)							
# Post	1.353^{**}	1.257^{**}	1.249^{**}	1.232^{**}	1.230^{**}	0.984^{**}	
	(0.433)	(0.425)	(0.424)	(0.420)	(0.425)	(0.383)	
Adjusted within \mathbf{R}^2	0.05	0.05	0.05 0.05		0.05	0.22	
		Dependen	t variable	: Anti-imm	igrant slan	t	
	(1)	(2)	(3)	(4)	(5)	(6)	
Inverse distance (min)							
# Post	-2.873**	-2.672*	-2.683*	-2.492*	-2.535*	-2.665**	
	(1.385)	(1.491)	(1.493)	(1.471)	(1.480)	(1.155)	
Observations	12636	12129	12129	12129	12129	12129	
Adjusted within \mathbf{R}^2	0.20	0.21	0.21	0.22	0.22	0.08	
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes	
Geo and News	Yes	Yes	Yes	Yes	Yes	Yes	
Months to elections		Yes	Yes	Yes	Yes	Yes	
Population controls		Yes	Yes	Yes	Yes	Yes	
Migration controls			Yes	Yes	Yes	Yes	
Income and crime				Yes	Yes	Yes	
Tot. outlets					Yes		
Prov. level trends						Yes	

Table 2. Baseline results: News coverage and Slant

Notes: Baseline results for news coverage and news importance: results pertain to a reduced form diff-in-diff model outlined in Equation 2. The dependent variable is the log of news counts in the top-panel, and a news slant at the bottom. Standard errors are two-way clustered at municipality and month-year level. Controls are gradually added from left to right. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

Figure 5. Event study, effect over time.



Notes: Coefficients plot from an event study specification. The outcome of interest is respectively news coverage in subfigure a, and anti-immigrant slant in b. These specifications control for province-level-time and seasonal trends. These inclusions allow to shut down the significance of some pre-policy coefficients, that is otherwise present for some period-specific coefficients.

	Dej	p. variable: Slant	Dep. variable : Predicted probabilities			
	(1)	(2)	(3)	(4)	(5)	
	Friedman et Al. 2010	Threshold 55	Threshold 45	Friedman et Al. 2010	Taddy (2013)	
Inverse distance (min)						
# Post	-2.587**	-2.837**	-2.798**	-2.827**	-3.127**	
	(1.168)	(1.233)	(1.214)	(1.227)	(1.300)	
Observations	12129	12129	12129	12129	12129	
Adjusted within \mathbb{R}^2	0.06	0.09	0.09	0.07	0.10	
Municipality FE	Yes	Yes	Yes	Yes	Yes	
Month-year FE	Yes	Yes	Yes	Yes	Yes	
Geo and News	Yes	Yes	Yes	Yes	Yes	
Months to elections	Yes	Yes	Yes	Yes	Yes	
Population controls	Yes	Yes	Yes	Yes	Yes	
Migration controls	Yes	Yes	Yes	Yes	Yes	
Income and crime	Yes	Yes	Yes	Yes	Yes	
Tot. outlets						
Prov. level trends	Yes	Yes	Yes	Yes	Yes	

Table 3. Extended results: Anti-immigrant slant

Notes: Extended results for news anti-immigrant slant : results pertain to a reduced form diff-in-diff model outlined in Equation 2. The set of controls is equivalent to the baseline's last column. Dependent variables vary by column and constitute a variation of the baseline slant index. Standard errors are two-way clustered at municipality and month-year level. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

Figure 6. Event study, slant effect over time, robustness to multiple predictions.



Notes: Coefficients plot from an event study specification based on 1000 regressions. The outcome of interest is anti-immigrant slant. In each regression, articles are classified based on probability matching. These specifications control for individual time and seasonal trends as for the baseline event-study graphs.

5.1.3 Heterogenous Results by Time Invariant Controls

Baseline results suggest that local shocks in migrants' presence created a rise in news coverage in the most exposed areas while relatively augmenting anti-immigrant slant in the municipalities least directly exposed. Table B.3 adds an heterogeneity study to these findings, to explore how effects differ across initial education levels (based on 2011 shares of University graduates, initial incumbent ideology (i.e., vote-shares for the left-wing coalition in the general elections of 2008), altitude levels (i.e., maximum altitude reached in the municipality), and initial unemployment rates. Column 1 is included for benchmark and is equivalent to the rightmost of the baseline columns. The top of the table shows results on news coverage. Evidence suggests that the closest areas to the border receive more news coverage, and that this effect is stronger in areas with higher altitude (lower altitude municipalities can be conceived as urbanized, dense areas, given the region's morphology) and in areas with lower incumbency of left-wing parties. Slant results are also heterogeneous along altitude levels. The negative effect of proximity to the border on slant is more pronounced in areas with higher altitude, higher left-wing incumbency and higher education level. Results are relatively homogeneous across the other characteristics considered.

6 Mechanisms, Implications and Robustness

6.1 Mechanisms and Implications

6.1.1 Mechanisms

Are these results comparable to existing estimates? Though literature is scarce, Branton and Dunaway (2009) look at a broadly similar relation between distance to the border and media outcomes, and focus on the U.S.-Mexico border in California. The authors correlate Californian news coverage of immigration and distance to the borders, for the period 2004-2005, and find that an extra ten miles of distance to the border relates to a 1.98% lower coverage of Latino immigrant news. Comparing the present study with this correlation, and assuming 10 miles can be traveled in 20 minutes, baseline results suggest a much steeper slope, that decays to comparable values from around 23 minutes away from the border onward. The authors also find distance to correlate negatively with coverage of negative aspects of immigration. By exploiting a quasi-experimental setting, the present study finds an opposite pattern, which suggests the importance to consider shocks in the salience of distance, besides the exploration of general tendencies. In the US setting, Gentzkow and Shapiro (2010) predict an 18 % loss in circulation of an average (national) newspaper if it were to deviate by 1 standard deviation from the profit-maximizing level of slant. Taking this reference as a crude benchmark, assuming that slant adjustments result from news firms maximizing profits, not adjusting slant to the border events could reduce readers' interest by roughly 4.5% for every 0.1 units of border proximity, in this crude comparison.

To further investigate these dynamics, this section continues on with an exploration of readers' demand for ideology and its role in the observed slant results.

In the benchmark indicators, news items are mapped into municipalities following Google searches for local outlets, as described in the data section. As Google searches data vary over time, the baseline geocoding includes a component of endogenous variation in the demand for local news sources.

Here, two alternative geocoding procedures are proposed to map article data into municipalities. The related estimations are presented in Table 4. In the table, Column 1 displays benchmark estimates for comparison.

The first alternative considers only Google searches from before June 2015, to geolocalize news. The resulting estimates for the slant indicator are presented in Column 2.

The size and significance of the coefficient of interest drops. This suggests that, at least in part, slant's effects are driven by changes in local news demand after June 2015. A second alternative involves shutting down the demand component in geocoding news: here, news items are assigned to municipalities according to where outlets' headquarters are located. Specifically, each outlet is allowed to cover municipalities within twenty kilometers (while not presented here, an alternative fifteen and twenty-five-kilometers buffer were also tested and produced similar findings). Crucially, for all municipalities within a same buffer, articles from each local news source have the same weight, regardless of readers' traffic.

Some remote municipalities are not located in any buffer. They are dropped as they're not assigned any news, leading to a smaller sample size. For the sake of comparability, the table fixes the observations from this subset for all regressions. With this supply-only alternative, the slant effects presented in Column 3 of Table 4 significantly drop in size and lose significance. This set of results suggest that local-news slant is shaped by the dynamics of demand for news consumption.

	Dependent variable: Anti-immigrant slant				
	(1) Benchmark Estimates	(2) Pre-existing demand	(3) Supply (Headquarters		
Border distance (min)					
# Post	-1.108*	-0.789	-0.304		
	(0.592)	(0.675)	(0.938)		
Observations	6367	6367	6367		
Adjusted \mathbb{R}^2	0.06	0.21	0.10		
Municipality FE	Yes	Yes	Yes		
Month-year FE	Yes	Yes	Yes		
News controls	Yes	Yes	Yes		
Months to elections	Yes	Yes	Yes		
Population controls	Yes	Yes	Yes		
Migration controls	Yes	Yes	Yes		
Income and crime	Yes	Yes	Yes		
Prov. level trends	Yes	Yes	Yes		

Table 4. Extended results: Supply versus demand

Notes: Extended results on demand versus supply channels. Standard errors are two-way clustered. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

In George and Waldfogel (2006)'s framework, local news outlets' positioning in the market depends on the incumbency of national news firms. Below, the effect of slant is allowed to vary with the penetration of local news relative to national news in the local market. To represent this penetration, a new proxy variable, constructed with the use of Google Trends keyowrd comparisons, captures the relative traffic of a local source with respect to one of four mainstream national newspapers: La Repubblica, Il Fatto Quotidiano, La Stampa, and Il Giornale. In particular, for each news item j, local penetration is proxied by ratio $\frac{a_j}{b_j}$, where a_j is Google Trends monthly traffic score for the local outlet that published j, relative to national news traffic b_j . Article-level data are then aggregated at the municipality level following the same weights used for the baseline media outcomes. The resulting measure indicates how much online interest local news received in proportion to searches for national news.²² A visualization of the four measures created one for each national news outlets compared - appears in Figure A.9.

 $^{^{22}}$ See Askitas (2015) for a review on the scope of Google Trends data for detecting relevant socio-economic trends with high time-frequency, worldwide.



Figure 7. Migration-related articles, effect on anti-immigrant slant of articles per month.

(a) La Repubblica

(b) Il Fatto Quotidiano

Notes: Coefficients plot from an event study specification. Local news penetration is the outcome of interest. From the top-left to the bottom-right, the national outlets of reference for the measure are, respectively: La Repubblica, Il Fatto Quotidiano, La Stampa, Il Giornale.

Figure 7 reports results from an event study specification that employs these local-news penetration measures as outcomes of interest. Results suggest that, after the border controls, the relative demand differential for local news initially increased more in the interested areas, close to the border, where coverage of migration in the news was higher, but over time it shifted further from the border, following the same direction of slant.

Building on this finding, Figure 8 plots results from a triple-interaction specification, with the penetration ratio as outcome, and with the proximity times post treatment interacted with slant. The figure shows marginal effects of border proximity on penetration, by levels of slant in the pre and post-event periods.

Following the border closure, we observe that the effect of distance on relative penetration grows with the levels of anti-immigrant slant. Specifically, it turns significantly negative as long as slant is not extremely low and increases in magnitude where slant is higher. The link between slant, distance to the border and local-news penetration appears to reflect a mechanism of content and reader targeting by local news. Further from the border, readers do not experience events directly. They form (or hold) beliefs that differ from those of readers at the border, and look for their views to be matched in the news. Local news react by adjusting their offer to satisfy this demand. By adjusting slant, they appear to gain market from national news, which are less able to accommodate specific areas. Content adjustments by local news based on reader targeting and national news coverage are also documented in George and Waldfogel (2006), for the US.

Figure 8. Effects on penetration by levels of slant



Notes: marginal effects from a triple-differences specification. The local-news penetration proxy is the outcome of interest. From the top-left to the bottom-right, the national outlets of reference for the measure are, respectively: La Repubblica, Il Fatto Quotidiano, La Stampa, Il Giornale.

News at the border are subject to audience that also directly observes the events.²³ Close to the events, slant is lower. This finding may indicate a uniformly favorable response of natives (then reflected in the news), or alternatively a mixed ideological response. Along this second interpretation, readers' heterogeneity in beliefs can lower newspaper

 $^{^{23}}$ Gentzkow et al. (2018) consider a model of trust in information sources where individuals both receive information from news and from direct exposure to an event, and decide on the information accuracy depending on their own ideology, allowing them to trust more news that validate their views.

bias, if, as in the model of Mullainathan and Shleifer (2005), readers seek confirmatory news, that reinforce their beliefs, and news outlets accommodate them. In the authors' framework, readers diversity is a crucial condition to reach overall accuracy, because biased news exist to accommodate readers in both ideological directions, and so diversity in the readers' beliefs lead to news whose bias directions neutralize in aggregate. Away from the policy area, beliefs are less bounded by reality, which is distant and thus less observed. By accommodating this bias, local news succeed in penetrating the market.

In accordance with this interpretation, Steinmayr (2021) discusses the opposite implications of contact versus mere exposure of natives to migrants.²⁴ Consistently with the contact-theory framework, natives would have responded on average more favorably closer to the border. Anecdotally, several NGOs were present at the border in the aftermaths of the borders' closure. This could have been a further stimulus for a (relatively) positive response in the area. Figure 9 is in line with the contact-theory mechanism. It reports a count measure of hate crimes in the two provinces of interest. This figure builds on Romarri (2020) who investigates the impact of far-right mayors on hate crimes, employing a database collected by NGO Lunaria.²⁵ Using the same data, Figure 9 depicts the time variation of hate crimes, aggregated by province. An emerging pattern reflects that, past 2015, hate crime prevalence grew relatively more in the province of Savona, which is further from the border.²⁶ In what follows, these overall results are compared with voting patterns, investigating the implications for political preferences.





Notes: Time series of hate crimes, compiled by database Lunaria, used in Romarri (2020).

6.1.2 Voting Patterns

How do the outcomes in local news match with aspects of the local political panorama? To answer this question, the present section explores how migrants' push-backs influenced voting patterns, comparing the periods before and after the border-control policy.²⁷ Outcomes presented here derive from a first-differences estimation, where differences are

 $^{^{24}\}mathrm{This}$ framework is based on the work on social psychology by Allport (1954).

 $^{^{25} {\}rm www.cronachediordinariorazzismo.org}$

 $^{^{26}}$ Note the count variable in the period of interest has low variation. This lack of heterogeneity motivates the choice of not running a regression for this outcome.

²⁷See Appendix C for a description of the Italian political context.

computed between 2018 and 2013, years in which general elections were held.

The equation line is the following:

$$\Delta Y_{mc} = \beta_1 InverseDistance_m + \beta_2 InverseDistance_m * Slant_m + \beta_3 InverseDistance_m * Coverage_m + \Gamma \Delta X_{mc} + \epsilon_{mc}$$

$$\tag{4}$$

For municipalities $m \in \{1, ..., M\}$, we observe election outcomes ΔY_{mc} , corresponding to the difference in vote shares for a party or coalition in 2018 with respect to 2013, for each of the parliament chambers c: {Senato, Camera}. The treatments of interest are a combination of distance to the border (with corresponding coefficient β_1), and its interactions with the average from 2015 to the election year of media outcomes $Slant_m$ and $Coverage_m$ (with corresponding coefficients β_2 and β_3 , respectively). Table 5's specifications all control for population (in logs), population density the share of the population over 65, the log of immigrant stocks, income per capita, population inflows, outflows, and crime events (at province level), plus the geographic controls that are present in the baseline. The key covariates are standardized, to ease interpretation. Results on the shares of two most explicitly anti-immigrant parties *Lega* and *Fratelli d'Italia (FDI)* are reported jointly in Column 1, and separately in Columns 2 and 3.

At lower proximity, slant correlates positively with vote share changes for FDI, at the expense of the previous main party of the right-wing coalition PDL, and with voter turnout. Closer to the border areas, FDI gains success, more so if slant is higher. However, the left-wing coalition also gains popularity in the vicinity of the border events, at a higher pace with higher coverage and slant levels Coverage of migration appears to positively affect turnout and FDI shares in a similar fashion as slant.

This evidence suggests that migrants' push-back did not provoke an homogeneous positive response the closer to the events, but rather sparked heterogeneous voting reactions. From one side they raised anti-immigrant voting, on the other side, the opposite ideology grew as well. In light of the news-related findings, this heterogeneous response might have been a key mechanism to push aggregate slant down in the vicinity of the push backs, relative to areas least directly involved, that had a more uniform, relatively more negative response.

Outcome: vote shares	Anti-immigration		Coalitions		Turnout	Other vote shares			
	(1) Lega, FDI	(2) Lega	(3) FDI	(4) RW coalition	(5) LW coalition	(6) Turnout	(7) PD	(8) M5S	(9) PDL, FI
Inverse dist	-0.026	-0.110	0.084^{**}	-0.121	-0.037	-0.034	-0.002	-0.073	-0.078
	(0.102)	(0.097)	(0.037)	(0.100)	(0.064)	(0.060)	(0.049)	(0.111)	(0.083)
RW Slant	0.092	0.060	0.031^{**}	-0.068	0.068	0.103**	0.057	0.008	-0.159***
Inverse dist	(0.083)	(0.078)	(0.013)	(0.075)	(0.052)	(0.035)	(0.046)	(0.058)	(0.041)
#Average slant	0.005	-0.074	0.079^{**}	-0.045	0.214^{**}	-0.009	0.139^{*}	-0.086	-0.063
Inverse dist	(0.108)	(0.104)	(0.031)	(0.113)	(0.084)	(0.071)	(0.071)	(0.132)	(0.096)
#Average coverage	0.017	-0.015	0.032	0.020	0.233^{**}	0.006	0.140^{**}	-0.047	-0.020
	(0.095)	(0.094)	(0.023)	(0.122)	(0.082)	(0.067)	(0.069)	(0.119)	(0.110)
Coverage	0.067	0.031	0.036^{**}	-0.061	0.057	0.085^{**}	0.027	0.023	-0.133***
	(0.063)	(0.060)	(0.014)	(0.058)	(0.043)	(0.032)	(0.037)	(0.052)	(0.035)
Observations	262	262	262	262	262	262	262	262	262
Adj. R squared	0.184	0.221	0.315	0.137	0.283	0.220	0.300	0.083	0.187
Mean Outcome	0.239	0.217	0.022	0.154	-0.065	-0.020	-0.066	-0.058	-0.082
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

 Table 5. Baseline results: Elections outcomes

Notes: Results for voting outcomes. This table displays coefficients from a first differences model. The dependent variable constitutes the party or coalition's vote shares. Election outcomes from this table pertain to general elections of 2013 and 2018. Standard errors are clustered at municipality level. The set of controls is in first-differences and includes: population (in logs), population density, over 65 population shares, logs of immigration stocks, income per capita, population inflows and outflows, and crime prevalence (in logs) at province level, plus the same geographic controls added in the baseline. Acronyms read as follows: FDI is *Fratelli d'Italia*, RW and LW are respectively "right-wing" and "left-wing". PD, M5S and PDL, FI stand for *Democratic Party, Movimento 5 Stelle*, and *Popolo della Libertà, Forza Italia*, respectively. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

6.2 Sensitivity to identification assumptions and other robustness

6.2.1 Identification and Placebos

The analysis presented so far accounts for several violations of the parallel trends. All specifications from the last Column in the baseline onward include province-specific time effects. Estimates also absorb out unobserved heterogeneity that is captured by possibly confounding geographic characteristics; namely latitude, longitude and altitude, interacted with the post-events dummy. The addition of several covariates after studying the balance of the sample is also a measure to improve comparability. Event-study graphs with this specification do not point at significant pre-policy effects. As a robustness to the baseline specification of fixed effects, Table B.4 adds municipality-level time trends to the preferred columns of the study and displays coefficients in line with the main findings. A concern, however, relates to the presence of a continuous treatment in the context of a diff-in-diff methodology. Callaway et al. (2021) observe that when the treatment is continuous, estimating the average causal response (ACR) of a treatment dose d by two-way fixed effects (TWFE) presents some limitations. As a side note, this analysis is spared from further complications that also arise when the treatment timing is staggered. Here instead, the border closure occurred in one

specific period, impacting the whole sample at the same time. A first limitation in the present analysis involves the weights that TWFE estimates assign to each dose. If the goal is to obtain a weighted average of the ACR, with weights matching the real dose distribution, TWFE estimates tend to give too much importance to mean-dose observations, thus under-representing the distribution tails. To address this issue, a battery of sensitivity checks looks at several alternative functional forms of distance, including binary distance specifications. As these different measurements change how the distance (dose) distribution looks like, one can observe how sensitive are TWFE parameters to the use of inverse distance. An additional issue of TWFE estimates relates to the parallel trends assumption, which is required to hold at every treatment-dose comparison. If this stronger assumption doesn't hold, TWFE estimates would include a selection bias term arising from the different paths that different doses would have followed in a counterfactual state of nature. The policy of border push-backs was introduced by the French authorities, which respond to different governments and jurisdictions than the sampled areas. Distance to the border would hardly reflect a selection into treatment for native in the Italian areas. The use of distance induces a reduced-form estimate of the direct exposure to migrants, without the element of selection into exposure. However, distance to the border may also proxy several unrelated factors that may harm the full comparability of municipalities, beyond those that have been included in the controls, and even if the provinces in sample were targeted specifically to ensure that observations were more homogeneous along several factors. To challenge the validity of using proximity to the border as an appropriate proxy for treatment intensity, a placebo analysis is performed as follows. In the baseline specification, the actual inverse distance to the border town of Ventimiglia is replaced by the inverse distance to a set of Italian towns taken randomly (outside the North West region of Italy, to avoid an excessive collinearity between distance measures). Geographic coordinates for Italian towns, villages and cities derive from OpenStreetMap (2023). A random 10% sample of the towns' list is retained for the placebo. Commuting times are calculated with the R package osrmTable (Mark Padgham et al., 2017). A placebo regression is repeated for every extracted town. The histogram of Figure 10 plots the distribution of resulting p-values on the left. On the right of the figure, coefficients' values are plotted. A 4% of the random distances leads to a significant estimate, indicating that baseline estimates do not over reject the null-hypothesis for the coefficient of interest. The continuous treatment adopted in the baseline therefore appears to well-identify the analyzed impact.

Figure 10. Placebo coefficients



6.2.2 Other Robustness Tests

Table B.5 shows robustness tests with respect to alternative functional forms for commuting distance as a measure for treatment intensity. On the top side of the table, standardized commuting distance enters linearly. In the mid section, distance is transformed with the expression log(x + 1). At the bottom, a quadratic term is introduced. The table broadly confirms results obtained via the linear specification, both in terms of direction and significance. A 1 standard deviation increase in commuting distance in the post period reduces coverage by around 11 percent and increases slant by 0.312 standard deviations. In Column 2, a 1% increase in commuting distance results in dropping news coverage by 0.07% and in rising slant by 0.0015 standard deviations. In Column 3, the effect of distance (times post) on coverage stays negative until more than one standard deviation above the mean, while the effect on slant stays positive until more than 2 standard deviations above average.

Finally, in Figure A.10, the inverse distance specification is replaced by a set of binary treatment diff-in-diff regressions, in which a binary treatment equals 1 if commuting distance is higher than distance n, in the x-axis. The y-axis plots the values of the diff-in-diff coefficient, with p-values and confidence intervals obtained with wild-cluster bootstraps. The overall take away of the graph is that, despite some size fluctuations, results are robust in sign and significance to the linear distance coefficients.²⁸

A second battery of robustness tests is performed to allow the possible exclusion of a municipality (i.e., a panel unit). Because the set of municipalities under study is restricted to a small sample, a test is conducted to make sure that the results on the three outcomes are not driven by the presence of a single municipality. Figures A.11, and A.12 reproduce both betas and associated t-statistics for all possible permutations of such exclusions. The figures indicate that both sign and significance of the effects are maintained in all these alternatives. On a similar ground, Figure

 $^{^{28}}$ David et al. (2018) suggest that different approaches in addressing the scale of analysis may induce conflicting results on the drivers of individual attitudes and political preferences. These sensitivity checks around the measurement of exposure would at least partially address these concerns.

A.13 summarizes results when a month is excluded from the estimation. Results persist in sign and significance to each of these possible alternatives. A related concern involves the use of an imputation method to extend the original geographic distribution of news, retrieved by Google Trends. Note without imputing the sample substantially shrinks to 38 municipalities. Table B.6 shows that even when restricting the sample to the original non-imputed data, the direction of the results is maintained. Concerning size and significance, there's some fluctuation in size, but results maintain the same sign. The significance disappears, given higher standard errors. Reassuringly, slant coefficients are very much in line with the baseline estimates.

Finally, a placebo test confirms that the predictive slant algorithm for articles is not producing a spurious result. The test runs as follows. A training sample is built such that it is equal to the original training sample, but the predetermined 0-1 labels are now randomly shuffled. Then, slant predictions are made on the local news dataset as done for the original measure. The process is repeated to construct 1000 placebo measures of slant.²⁹ Figure A.14 plots the distribution of resulting p-values on the left. On the right of the figure, coefficients' values are plotted. The picture shows insignificant results for the placebo-slant measure. This confirms that it is the original training sample, rather than a random label assignment, that leads to significant results in the baseline.

7 Conclusion

This paper examined how local news markets can be shaped by salient events in the context of the recent European migration crisis and documents a relationship between news-market outcomes and trends in the local political landscape. It provides new evidence that local shocks in migrant presence impact discourse in local news, both in terms of quantity (*agenda-setting*) and quality (*framing*), though in opposite directions. The identification of these effects follows from a specific geographic and historical setting. A local shock in migrant presence at the French-Italian border in June 2015 emerged from the establishment of a militarized push-back system by the French authorities. The coastal border had previously constituted an important gateway for precarious migrants, intending to leave Italy for other European destinations.

Understanding how news speech evolves is key to unravel on aspects related to the evolution of citizens' attitudes (Djourelova, 2023; Keita et al., 2023) and language analysis can retrieve rich information on the political landscape beyond what observed electoral voting can capture (Gentzkow et al., 2019b). For example, Campante et al. (2018) observe how internet spreading in Italy influenced dissatisfied voters. Their behavior changed over time and only ultimately resulted into voting for the anti-elite Five-Star Movement.

A rich data collection of articles allowed investigating news outputs. Evidence suggests the border push-back events translated into a salient shock in the media: municipalities closer to the event received more migrant-related news, and coverage decays rapidly with distance.

The share of anti-immigrant news, however, was observed to get relatively higher in the municipalities further away from the border. The study proposes some rationalizations of this finding. Results are found to be driven by articles that received more news consumption. Highly demanded local news firms adjusted their content, possibly to

 $^{^{29}}$ This procedure leads to a vast number of prediction probabilities to be exactly 1/2: these entries are randomly assigned a 0 or a 1.

accommodate an expected demand for slant which differs along a different exposure levels to the real events. Local vis-à-vis national news search traffic is taken as a relative measure of penetration in the analysis. The relative interest in local news turns out to change in the same direction as slant, after the border policy. Additionally, in the least exposed areas, slant is the strongest where local news gain more market. If local news readers close to the border become more heterogeneous in terms of migration views, local demand for anti-immigrant voicing may be counterbalanced, at the border, by the presence of favourable counterparts. This mechanism can also relate to Steinmayr (2021)'s contact versus exposure paradigm. Lower slant levels in the areas closest to the border may reflect the presence of empathetic attitudes as more contact occurred with the migrants' precarious situation, whilst municipalities further away, lacking this contact component, responded on average more negatively. This line of thought is matched by: i) overall trends in hate crime events, measured à la Romarri (2020) and ii) vote shares differentials for the general elections, where higher heterogeneity in the support appears in the vicinity of the border. Baseline results prove robust to several sensitivity checks on the specification of the model and the measurement of the main variables.

While focusing on a local setting, the policy relevance of this contribution is broader. This study relates to the importance of international coordination in migration policies (Cimade, 2018; Giordani and Ruta, 2013): as an example of cross-country tensions in the political debate on internal borders policies, a provocative tweet of anti-immigrant party FDI leader Giorgia Meloni stated: "Important notice to migrants! France, Germany and Luxembourg have declared that closing ports violates international law and that rescue cannot be criminalized. Head to their *borders*, you will be received with open arms by these welcoming states."³⁰ Finally, this paper focuses on local news and employs standard text-analysis *bag-of-words* methods in the prediction of slant. An important area for future research will involve the exploration of other relevant media and the use of promising, evolving techniques to possibly capture even deeper mappings between text nuances and ideology.

Bibliography

- Aksoy, Cevat Giray and Panu Poutvaara (2021) "Refugees' and irregular migrants' self-selection into Europe," *Journal of Development Economics*, 152, 102681, https://doi.org/10.1016/j.jdeveco.2021.102681.
- Alesina, Alberto and Marco Tabellini (Forthcoming) "The Political Effects of Immigration: Culture or Economics?" Journal of Economic Literature, 10.1257/jel.20221643.
- Allport, Gordon W (1954) "The nature of prejudice," Reading: Addison-Wesley, 129.
- Altındağ, Onur and Neeraj Kaushal (2020) "Do refugees impact voting behavior in the host country? Evidence from Syrian refugee inflows to Turkey," *Public Choice*, 1–30.
- Anderson, Simon P and John McLaren (2012) "Media Mergers and Media Bias with Rational Consumers," Journal of the European Economic Association, 10 (4), 831–859.

³⁰Avviso importante ai migranti! Francia, Germania e Lussemburgo hanno dichiarato che chiudere i porti viola il diritto internazionale e che il soccorso non può essere criminalizzato. Dirigetevi verso le loro frontiere, sarete accolti a braccia aperte da questi Stati accoglienti. Source: https://twitter.com/giorgiameloni/status/1145015881195540483, posted in June 2019, retrieved in October 2021.

Askitas, Nikolaos (2015) "Google search activity data and breaking trends," IZA World of Labor (206).

- Barone, Guglielmo, Alessio D'Ignazio, Guido de Blasio, and Paolo Naticchioni (2016) "Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior," *Journal of Public Economics*, 136, 1–13.
- Battiston, Giacomo (2022) "Rescue on Stage: Border Enforcement and Public Attention in the Mediterranean Sea," "Marco Fanno" Working Papers 0292, Dipartimento di Scienze Economiche "Marco Fanno", https://ideas.repec. org/p/pad/wpaper/0292.html.
- Benesch, Christine, Simon Loretz, David Stadelmann, and Tobias Thomas (2019) "Media coverage and immigration worries: Econometric evidence," Journal of Economic Behavior & Organization, 160, 52–67.
- Blei, David M, Andrew Y Ng, and Michael I Jordan (2003) "Latent Dirichlet Allocation," Journal of Machine Learning Research, 3 (Jan), 993–1022.
- Bouchet-Valat, Milan (2020) "Package 'SnowballC'," R package version 0.7. 0.
- Branton, Regina P and Johanna Dunaway (2009) "Spatial Proximity to the U.S.-Mexico Border and Newspaper Coverage of Immigration Issues," *Political Research Quarterly*, 62 (2), 289–302.
- Bratti, Massimiliano, Claudio Deiana, Enkelejda Havari, Gianluca Mazzarella, and Elena Claudia Meroni (2020) "Geographical proximity to refugee reception centres and voting," *Journal of Urban Economics*, 120.
- Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna (2021) "Difference-in-Differences with a Continuous Treatment," arXiv preprint arXiv:2107.02637.
- Campante, Filipe, Ruben Durante, and Francesco Sobbrio (2018) "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation," Journal of the European Economic Association, 16 (4), 1094–1136.
- Campo, Francesco, Sara Giunti, and Mariapia Mendola (2021) "The Refugee Crisis and Right-Wing Populism: Evidence from the Italian Dispersal Policy," IZA DP No. 14084.
- Capitani, Guilia (2018) "Nowhere But Out: The failure of France and Italy to help refugees and other migrants stranded at the border in Ventimiglia," 10.21201/2018.2708.
- Chen, Song Xi (2000) "Probability Density Function Estimation Using Gamma Kernels," Annals of the Institute of Statistical Mathematics, 52 (3), 471–480.
- Chouliaraki, Lilie, Myria Georgiou, Rafal Zaborowski, and WA Oomen (2017) "The European 'migration crisis' and the media: a cross-European press content analysis."
- Cimade, L (2018) "Dedans, Dehors: Une Europe Qui S'enferme," Report, Cimade, France, June.
- Colombeau, Sara Casella (2020) "Crisis of Schengen? The effect of two 'migrant crises' (2011 and 2015) on the free movement of people at an internal Schengen border," *Journal of Ethnic and Migration Studies*, 46 (11), 2258–2274, 10.1080/1369183X.2019.1596787.

- de Coulon, Augustin, Dragos Radu, and Max Friedrich Steinhardt (2016) "Pane e Cioccolata: The Impact of Native Attitudes on Return Migration," *Review of International Economics*, 24 (2), 253–281.
- David, Quentin, Jean-Benoit Pilet, and Gilles Van Hamme (2018) "Scale Matters in Contextual Analysis of Extreme Right Voting and Political Attitudes," *Kyklos*, 71 (4), 509–536.
- DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya (2014) "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia," American Economic Journal: Applied Economics, 6 (3), 103–32.
- DellaVigna, Stefano and Ethan Kaplan (2007) "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics*, 122 (3), 1187–1234.
- Dinas, Elias, Konstantinos Matakos, Dimitrios Xefteris, and Dominik Hangartner (2019) "Waking Up the Golden Dawn: Does Exposure to the Refugee Crisis Increase Support for Extreme-Right Parties?," *Political Analysis*, 27, 1–11, 10.1017/pan.2018.48.
- Djourelova, Milena (2023) "Persuasion through Slanted Language: Evidence from the Media Coverage of Immigration," American Economic Review, 113 (3), 800–835, 10.1257/aer.20211537.
- Durante, Ruben, Paolo Pinotti, and Andrea Tesei (2019) "The Political Legacy of Entertainment TV," American Economic Review, 109 (7), 2497–2530.
- Dustmann, Christian, Francesco Fasani, Tommaso Frattini, Luigi Minale, and Uta Schönberg (2017) "On the economics and politics of refugee migration," *Economic Policy*, 32 (91), 497–550.
- Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm (2019) "Refugee migration and electoral outcomes," The Review of Economic Studies, 86 (5), 2035–2091.
- D'Alimonte, Roberto (2013) "The Italian elections of February 2013: the end of the Second Republic?" Contemporary Italian Politics, 5 (2), 113–129.
- Edo, Anthony, Yvonne Giesing, Jonathan Öztunc, and Panu Poutvaara (2019) "Immigration and electoral support for the far-left and the far-right," *European Economic Review*, 115, 99–143.
- Facchini, Giovanni, Anna Maria Mayda, and Riccardo Puglisi (2017) "Illegal immigration and media exposure: evidence on individual attitudes," IZA Journal of Development and Migration, 7 (1), 1–36.
- Falck, Oliver, Robert Gold, and Stephan Heblich (2014) "E-lections: Voting Behavior and the Internet," American Economic Review, 104 (7), 2238–65.
- Flaxman, Seth, Sharad Goel, and Justin M Rao (2016) "Filter Bubbles, Echo Chambers, and Online News Consumption," Public Opinion Quarterly, 80 (S1), 298–320.
- Friedman, Jerome, Trevor Hastie, Rob Tibshirani et al. (2009) "glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models," R package version, 1 (4), 1–24.

- Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2010) "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33 (1), 1.
- Gamalerio, Matteo and Margherita Negri (2023) "Not welcome anymore: the effect of electoral incentives on the reception of refugees," *Journal of Economic Geography*, 23 (4), 901–920, 10.1093/jeg/lbad002.

Gentzkow, Matthew (2006) "Television and Voter Turnout," The Quarterly Journal of Economics, 121 (3), 931–972.

- (2009) "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions," *American Economic Journal: Applied Economics*, 1 (2), 35–52.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019a) "Text as data," *Journal of Economic Literature*, 57 (3), 535–74.
- Gentzkow, Matthew and Jesse M Shapiro (2006) "Media bias and reputation," *Journal of Political Economy*, 114 (2), 280–316.
- Gentzkow, Matthew and Jesse M. Shapiro (2010) "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, 78 (1), 35–71.
- Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone (2015) "Chapter 14 Media Bias in the Marketplace: Theory," in Anderson, Simon P., Joel Waldfogel, and David Strömberg eds. *Handbook of Media Economics*, 1 of Handbook of Media Economics, 623–645: North-Holland, https://doi.org/10.1016/B978-0-444-63685-0.00014-0.
- Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy (2019b) "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech," *Econometrica*, 87 (4), 1307–1340.
- Gentzkow, Matthew, Michael B Wong, and Allen T Zhang (2018) "Ideological Bias and Trust in Information Sources," Unpublished Manuscript.
- George, Lisa M. and Joel Waldfogel (2006) "The "New York Times" and the Market for Local Newspapers," *The American Economic Review*, 96 (1), 435–447, http://www.jstor.org/stable/30034376.
- Gessler, Theresa, Gergő Tóth, and Johannes Wachs (2021) "No country for asylum seekers? How short-term exposure to refugees influences attitudes and voting behavior in Hungary," *Political Behavior*, 1–29.
- Giordani, Paolo E and Michele Ruta (2013) "Coordination failures in immigration policy," Journal of International Economics, 89 (1), 55–67.
- Groseclose, Tim and Jeffrey Milyo (2005) "A measure of media bias," *The Quarterly Journal of Economics*, 120 (4), 1191–1237.
- Halla, Martin, Alexander F Wagner, and Josef Zweimüller (2017) "Immigration and voting for the far right," Journal of the European Economic Association, 15 (6), 1341–1385.
- Hangartner, Dominik, Elias Dinas, Moritz Marbach, Konstantinos Matakos, and Dimitrios Xefteris (2019) "Does Exposure to the Refugee Crisis Make Natives More Hostile?" American Political Science Review, 113 (2), 442–455.
- Hornik, Kurt and Bettina Grün (2011) "topicmodels: An R package for fitting topic models," *Journal of Statistical Software*, 40 (13), 1–30.
- Iannelli, Laura, Bianca Biagi, and Marta Meleddu (2021) "Public opinion polarization on immigration in Italy: the role of traditional and digital news media practices," *The Communication Review*, 24 (3), 244–274, 10.1080/10714421. 2021.1972720.
- Keita, Sekou, Thomas Renault, and Jérôme Valette (2023) "The Usual Suspects: Offender Origin, Media Reporting and Natives' Attitudes Towards Immigration," *The Economic Journal*, uead059, 10.1093/ej/uead059.
- Lonsky, Jakub (2021) "Does immigration decrease far-right popularity? Evidence from Finnish municipalities," *Journal* of Population Economics, 34 (1), 97–139.
- Mark Padgham, Bob Rudis, Robin Lovelace, and Maëlle Salmon (2017) "osmdata," Journal of Open Source Software, 2 (14), 305, 10.21105/joss.00305.
- Mayda, Anna Maria, Giovanni Peri, and Walter Steingress (2022) "The Political Impact of Immigration: Evidence from the United States," *American Economic Journal: Applied Economics*, 14 (1), 358–89, 10.1257/app.20190081.
- Mullainathan, Sendhil and Andrei Shleifer (2005) "The Market for News," American Economic Review, 95 (4), 1031–1053.
- OpenStreetMap (2023) "Italy's Sub Regions retrieved from https://download.geofabrik.de/europe/italy.html on 2023-05-03," https://www.openstreetmap.org.
- Otto, Alkis Henri and Max Friedrich Steinhardt (2014) "Immigration and election outcomes-Evidence from city districts in Hamburg," *Regional Science and Urban Economics*, 45, 67–79.
- Pettigrew, Thomas (1998) "Intergroup Contact Theory," Annual Review of Psychology, 49, 65–85, 10.1146/annurev. psych.49.1.65.
- Prat, Andrea and David Strömberg (2013) "The Political Economy of Mass Media," Advances in Economics and Econometrics: Theory and Applications, Proceedings of the Tenth World Congress of the Econometric Society.
- Romarri, Alessio (2020) "Do Far-Right Mayors Increase the Probability of Hate Crimes? Evidence From Italy," December, 10.2139/ssrn.3506811.
- Scheufele, Dietram A and David Tewksbury (2007) "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models," *Journal of Communication*, 57 (1), 9–20.
- Snyder Jr, James M and David Strömberg (2010) "Press Coverage and Political Accountability," Journal of Political Economy, 118 (2), 355–408.

- Steinmayr, Andreas (2021) "Contact versus Exposure: Refugee Presence and Voting for the Far Right," The Review
 of Economics and Statistics, 103 (2), 310-327, https://ideas.repec.org/a/tpr/restat/v103y2021i2p310-327.
 html.
- Strömberg, David (2004) "Radio's Impact on Public Spending," The Quarterly Journal of Economics, 119 (1), 189-221.
- Taddy, Matt (2013) "Multinomial inverse regression for text analysis," Journal of the American Statistical Association, 108 (503), 755–770.
- Tebaldi, Mauro (2014) "Il potere di esternazione del capo dello Stato nella Seconda Repubblica," *Comunicazione Politica*, 14 (3), 485–506.
- Welander, Marta (2017) "In Dangerous Transit: Filling information gaps relating to refugees and displaced people in Ventimiglia, Italy,"Technical report, Denise Charlton and Associates, Refugee Rights Europe, Refugee Youth Service.
- Zou, Hui and Trevor Hastie (2005) "Regularization and Variable Selection via the Elastic Net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (2), 301–320.

Appendix

A Additional Figures



Figure A.1. Dynamics of the border closure

Notes: Figure A.1 depicts the border events. At the left, a timeline summarizes the main occurrences; at the right, a map identifies the area of interest. This image is courtesy of La Cimade (https://www.lacimade.org/).



Figure A.2. Meal counts distributed by Caritas in 2017

Notes: Stylized statistics on the presence of asylum seekers and migrants that were offered a meal by local NGO Caritas Intemelia. Credits for this graph go to Caritas Intemelia.



Figure A.3. Provinces of interest versus rest of Italy, 2011

Notes: These graphs show where Imperia (red dotted line) and Savona (blue dotted line) lie within the distribution of Italian provinces on a set of socio-demographic characteristics. Data are sourced from ISTAT, and belong to the Census 2011 dataset. Density graphs are recovered from a univariate kernel density estimation that uses gamma kernel functions as in Chen (2000), where such kernel choice accounts for the positive range of the data. Bandwidth selection is based on unbiased cross-validation (UCV).



Figure A.4. Provinces of interest versus rest of Italy, continued

Notes: Continuation of Figure A.3.



Figure A.5. Local news readership

Notes: Comparison of Google searches for two of the main local news outlets, versus other internet searches.

Figure A.6. Internet usage for news consumption



Notes: Share of population that uses internet, sourced by ISTAT. On the left, the share of overall population that uses internet for news. On the right, the proportion of population that uses internet for e-mail.

Figure A.7. Comparison cloud. Terms with highest vs. lowest right-wing affinity



Most right wing (top 10pctile)

Notes: Exploration of the predictions resulting from the baseline classification approach. Prediction probabilities are employed to subset the sampled articles into least likely (bottom 10% of the assigned probability, at the top, in orange) and most likely (top 10% of the assigned probability, at the bottom, in violet) anti-immigrant.



Figure A.8. Balance check

Notes: Balance check: Significance of distance in explaining the pre-June 2015 covariates. Estimates pertain to an OLS estimation of the covariate of interest on commuting distance to the border for the pre-policy subsample, accounting for time fixed effects (except for the census sourced, time-invariant covariates). For comparability, all variables are standardized. Standard errors are two-way clustered: by municipality and month-year.



Figure A.9. Local news searches vis à vis national news.

Notes: Proxy for local-news penetration. Values over time. This proxy is calculated as the ratio between internet searches for the national outlet i (\in Il Fatto Quotidiano, La Repubblica, La Stampa, Il Giornale) and local sources *s* (each appearing as different colors in the graphs), as retrieved via the Google Trends platform. Local sources whose Google Trends search did not report any result were assigned a value of zero and do not appear in the graphs.



Figure A.10. β_1 robustness with alternative distance thresholds

Notes: This figure presents the distribution of estimation results from a alternative specification that considers a binary treatment for distance: on the x-axis the threshold level of commuting distance up to which units are considered to be treated: e.g. for distance=n, treatment is defined as 1 if municipality m is at distance >n. The specifications are comparable to Column 4 of the baseline table. Threshold values are plotted on the x-axis and the relative point estimate (red), with 95% confidence interval (blue, dotted lines) constitute the y-coordinates. On the left, graphs pertain to news-coverage. On the right, they refer to anti-immigrant slant estimations. P-values and confidence intervals are obtained with wild cluster-bootstraps based on 2000 repetitions. Clustering is one-way, on panel units.





Notes: This figure presents the distribution of estimation results for β_1 , where, at each iteration, one municipality is left out of the sample. The reference specification is the baseline's preferred column. In the figure, the left-hand side depicts the distribution of point estimates for β_1 . The t-statistics distribution for these estimates is on the right.

Figure A.12. Coefficients robustness on anti-immigrant slant: Leave-one-municipality-out estimates



Notes: This figure presents the distribution of estimation results for β_1 , where, at each iteration, one municipality is left out of the sample. The reference specification is the baseline's preferred column. In the figure, the left-hand side depicts the distribution of point estimates for β_1 . The t-statistics distribution for these estimates is on the right.



Figure A.13. Coefficients robustness: Leave-one-month-out estimates

Notes: This figure presents the distribution of estimation results for the coefficient of interest. At each iteration, one month is left out of the estimation sample. The reference estimates are baseline specifications including all control variables. Threshold values are plotted on the x-axis and the relative point estimate (red), with 95% confidence interval (blue, dotted lines) constitute the y-coordinates. Graphs pertain to estimations on news coverage and anti-immigrant slant respectively on the left and on the right.





Placebo estimates for slant. Slanted news are classified by shuffling randomly the labels in the training sample. The distribution of p-values resulting from 1000 shuffling iterations is on the left, while coefficients appear on the right.

B Additional Tables

Study	Setting	Results	Heterogeneity of the results
Dinas et al. (2019) (see also a related study in Hangartner et al. (2019))	Greece (2012-15)	Municipalities/townships w/ asylum seekers \implies (+) anti-immigrant vot- ing	channels are (-) support for center-right size, (+) turnout.
Campo et al. (2021)	Italy (2013-18)	Municipalities w/ asylum seekers centers \implies (+) anti-migrant voting	stronger for (-) municipality size, (+) immi- grants' stocks, (-) education level.
Bratti et al. (2020)	Italy (2013-18)	(+) proximity to hosting municipal- ities \implies (+) support for populist parties.	stronger for (-) municipality size,(-) per capita incomes, (-) prior left-wing orientation.
Edo et al. (2019)	France (1988-2017)	(+) immigration \implies (+) support for far-right.	Driven by low-educated immigrants from non-Western countries.
Dustmann et al. (2019)	Denmark (1986–98)	(+) refugee shares \implies (+) vote for right-leaning, anti-immigration par- ties.	Urban (weakest /opposite effect) vs rural pop- ulations (strongest effect).
Halla et al. (2017)	Austria (1970s-2013)	Inflow of immigrants into a commu- nity \implies (+) votes for the FPÖ party.	Channel of "compositional amenities". Com- munities with (+) immigration flows present lower public resources for natives (public day- care, school proximity).
Barone et al. (2016)	Italy (2001-2008)	(+) immigration \implies (+) center- right coalition votes.	Effect is heterogeneous by municipality size. Main channels: cultural diversity, labor mar- ket competition(education based), public ser- vices competition (migrants' fertility), polit- ical competition (closeness of votes share of first and second party).
Altındağ and Kaushal (2020)	Turkey(2012–2016)	Syrian-refugee presence produces po- larized attitudes towards migrants be- tween Erdoğan's AKP party support- ers and opposers.	No impact on AKP voting shares.
Steinmayr (2021)	Upper Austria (2009- 2015)	Opposite effects: exposure to asylum seekers \implies (+) far right support; contact (sustained interaction) \implies (-) far-right votes	Qualitative evidence: channel is reduced prej- udice after contact.
Otto and Steinhardt (2014)	Hamburg city dis- tricts (1987-1998)	Immigration \implies (+) far-right, anti- immigrant voting.	Channels: attitudes are shaped by non- economic determinants and welfare state con- siderations.
Lonsky (2021)	Finland (2006=2015)	(+) share of foreign citizens in munic- ipalities \implies (-) the Finns Party's vote share by 3.4 percentage points.	Mechanisms:(+) voter turnout. In line with contact theory, results holds only in munici- palities with greatest immigration levels.
Gessler et al. (2021)	Hungary (2014-2018)	Settlements with highest refugee tran- sit \implies (+) votes against EU refugees quotas (2016 referendum).	Election outcomes suggest vote-share shift from right to far-right. Spillover results show travel distance drastically decreases the effect.
Mayda et al. (2022)	US (1990-2010)	(+) share of high-skilled immigrants \implies (-) Republican votes share, vice versa for the low-skilled immigrants.	No heterogeneity of the effects with respect to immigrants' origins. In line with labour market concerns, stronger negative results are found in rural, low-skilled counties.

Table B.1. Further contributions

Notes: Complement to the literature review. This table summarizes a set of contributions on the effect of immigration on attitudes and political preferences of natives.

	Mean	Standard Deviation	Ν	Min	Max
Anti-immigrant slant, Friedman 2010	-0.013	0.978	12636	-3.420	2.364
Anti-immigrant slant, Taddy 2013	-0.003	1.001	12636	-2.496	3.132
Anti-immigrant slant (predictions)	-0.004	0.986	12636	-2.667	2.517
Anti-immigrant slant (predictions), Taddy 2013	0.006	1.002	12636	-2.415	2.911
Migrants' news exposure (log)	3.667	0.428	12636	2.079	5.067
Log of total news	7.600	0.238	12636	7.132	8.581
Avg. word count per outlet	467.108	66.805	12636	291.500	713.776
Population, log	7.155	1.325	12636	4.691	11.032
Population over 65, log	5.904	1.291	12636	3.466	9.789
Population density (population per sq. km)	235.963	342.012	12636	4.611	1922.043
Share of population over 65	0.291	0.051	12636	0.170	0.477
Stocks of immigrant population, log	4.673	1.386	12636	0.000	8.765
Inflows: New registrations	1.718	1.294	12636	0.000	7.650
Outflows: Cancellations	1.681	1.247	12636	0.000	5.832
(Taxable) Income per capita, log	9.957	0.131	12636	9.664	10.723
Commuting distance to Ventimiglia, minutes	45.755	20.075	12636	0.000	83.460
Total of reported crime events, yearly (Prov)	9.454	0.161	12636	9.210	9.740
Unemployment rate, 2011	0.079	0.027	12636	0.013	0.159
Share of university educated, 2011	0.074	0.025	12636	0.025	0.167
Max altitude	954.985	428.289	12636	157.000	2166.000
Area, sq. km	19.985	18.042	12636	1.289	100.658
Total voters, 2008	2370.596	5355.497	12636	67.000	39405.000
Top party is RW, 2008	0.840	0.366	12636	0.000	1.000
Shares LW coalition, 2008	0.322	0.069	12636	0.181	0.544
Top party is RW, 2018	0.425	0.494	12636	0.000	1.000
Shares RW coalition, 2018	0.464	0.068	12636	0.286	0.741

Table B.2. Summary Statistics

Notes: Summary statistics for the sample of interest.

	News coverage (log)				
	(1) Benchmark results	(2) Education, 2011	(3) LW vote shares, 2008	(4) Altitude (max)	(5) Unemployment rates, 2011
Inverse distance (min)					
# Post	0.984^{**}	0.826*	0.603	0.445^{**}	0.229
	(0.383)	(0.443)	(0.501)	(0.188)	(0.688)
Triple Interaction		4.391	2.884	0.002^{***}	9.635
		(6.527)	(3.021)	(0.001)	(8.320)
Adjusted within \mathbb{R}^2	0.22	0.22	0.22	0.25	0.22
			Anti-immigrant	slant	
	(1)	(2)	(3)	(4)	(5)
Inverse distance (min)					
# Post	-2.665**	-1.631	-1.298	-0.975^{**}	-0.907
	(1.155)	(1.157)	(1.162)	(0.448)	(1.674)
Triple Interaction		-28.712**	-10.342*	-0.007***	-22.531
		(13.541)	(6.122)	(0.001)	(17.854)
Observations	12129	12129	12129	12129	12040
Adjusted within \mathbb{R}^2	0.08	0.09	0.09	0.11	0.08
Municipality FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
News controls	Yes	Yes	Yes	Yes	Yes
Months to elections	Yes	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes	Yes
Migration controls	Yes	Yes	Yes	Yes	Yes
Income and crime	Yes	Yes	Yes	Yes	Yes
Prov. level trends	Yes	Yes	Yes	Yes	Yes

Table B.3. Extended results: Heterogeneity

Notes: Coefficients relate to the dependent variable *news coverage* at the top, and *anti-immigrant slant* at the bottom. The set of controls included is equivalent to the rightmost of the benchmark table's columns. Column 1 is included for comparison and indicates benchmark results. Column 2 allows for an interaction with education levels (as of 2011). Column 3 considers initial political ideology (as from 2008's elections), Column 4 and 5 include, respectively, the average altitude of the municipality and the levels of unemployment as of 2011. Standard errors are two-way clustered. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

	Dep. variable:		
	(1)	(2)	
	News coverage (\log)	Anti-immigrant slant	
Inverse distance (min)			
# Post	1.204**	-2.319*	
	(0.454)	(1.183)	
Observations	12129	12129	
Adjusted within \mathbf{R}^2	0.12	0.05	
Municipality FE	Yes	Yes	
Municipality time-trends	Yes	Yes	
Month-year FE	Yes	Yes	
Geo and News	Yes	Yes	
Months to elections	Yes	Yes	
Population controls	Yes	Yes	
Migration controls	Yes	Yes	
Income and crime			

Table B.4. Time trends at municipality level

Notes: Extended results, linear time trends interacted with municipality dummies. The table structure is comparable to the baseline estimates. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

	Linear distance		
	(1)	(2)	
	News coverage (\log)	Anti-immigrant slant	
Linear distance # Post	-0.112***	0.312***	
//	(0.018)	(0.060)	
	()	(*****)	
Adjusted within R ²	0.27	0.13	
	Log d	istance	
	(1)	(2)	
	News coverage (log)	Anti-immigrant slant	
Log distance,			
# Post	-0.074***	0.151**	
	(0.016)	(0.053)	
Adjusted within \mathbb{R}^2	0.106	0.073	
	Quadratic distance		
	(1)	(2)	
	News coverage (log)	Anti-immigrant slant	
Distance,			
# Post	-0.070***	0.234^{***}	
	(0.017)	(0.051)	
Linear distance, squared $\#$ Post	0.027**	-0.050**	
	(0.009)	(0.018)	
Observations	12218	12218	
Adjusted within R ²	0.29	0.13	
Municipality FE	Yes	Yes	
Month-year FE	Yes	Yes	
Geo and News	Yes	Yes	
Months to elections	Yes	Yes	
Population controls	Yes	Yes	
Migration controls	Yes	Yes	
Income and crime	Yes	Yes	

Table B.5. Extended results: Distance functional forms

Notes: Extended results, Different functional forms for the distance-based treatment. Results compare to the rightmost columns of the baseline results. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

	Dep. variable:		
	(1)	(2)	
	News coverage (\log)	Anti-immigrant slant	
Inverse distance (min)			
# Post	0.300***	-1.679^{***}	
	(0.086)	(0.360)	
Observations	3214	3214	
Adjusted within \mathbb{R}^2	0.20	0.04	
Municipality FE	Yes	Yes	
Month-year FE	Yes	Yes	
Geo and News	Yes	Yes	
Months to elections	Yes	Yes	
Population controls	Yes	Yes	
Migration controls	Yes	Yes	
Income and crime	Yes	Yes	
Tot. outlets			

Table B.6. Robustness to the exclusion of the imputed sample

Notes: Robustness test of the main results against the use of the nonimputed sample only. The dependent variables are the three media outcomes of interest. Standard errors are two-way clustered at municipality and month-year level. Controls are gradually added from left to right. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

=

C The Political Context in Italy

As of 2021, Italy was formed by 7,904 municipalities (source: ISTAT), which constitute the lowest level of government. Municipalities are organized in 107 territorial bodies which correspond to 83 provinces, 14 metropolitan areas and 10 otherwise denominated entities, with an equivalent broad role. Provinces are then grouped into 20 regions. At the national level, the Italian Parliament is bicameral, composed of the Chamber of Deputies and Senate of the Republic. Elections of their members are held every five years.

The recent Italian political landscape is rooted in a major political crisis which took place during the early 90s, sparked by a series of scandals that were famously dubbed *Tangentopoli* (Bribeville). Specifically, a set of investigations uncovered extensive corruption systems that implicated entrepreneurs and politicians in power, for both major political parties at the time (the Italian Socialist Party and the Christian Democrats). In this setting, the entrepreneur Berlusconi, although involved in some connected investigations, gained influence and formed a center-right coalition which allowed him to become the Prime Minister for the first time in 1994. This set of scandals marked the end of the so-called First *Republic* and the beginning of the *Second Republic* in 1994, which period saw the rise of a bipolar landscape with two major coalitions: the center-left and center-right. Several authors mark the 2013 elections as defining the end of this Second Republic (D'Alimonte, 2013; Durante et al., 2019; Tebaldi, 2014). In 2013 in particular, comedian Beppe Grillo brought his newly created Five Stars Movement to the elections, obtaining an unanticipated, relative success. Five Star gained significant traction among an unsatisfied, unrepresented electorate, and internet diffusion was found to have a major role in the movement's ascent (Campante et al., 2018). This political movement was considered taking a broad populist stance by several sources (Bratti et al., 2020; Campo et al., 2021). Conditional on this scenario, and with the migration crisis peaking in 2014-2015. Italy saw the rise of two right-wing parties that maintained a strong nativist. anti-immigrant rhetoric (Bratti et al., 2020; Campo et al., 2021; Gamalerio and Negri, 2023; Romarri, 2020): Lega and Fratelli d'Italia (FDI).

Lega's party was founded in 1991 as a separatist party that broadly aimed at a secession of the more productive Northern Italy from the Center-South and therefore supported political and fiscal federalism. The party's vote shares oscillated from around 8% in 1992's general elections, to 4% in 2013, after joining Berlusconi's coalition, to above 17% in 2018, under ne party head Matteo Salvini. The new leader abandoned the Northist stance and maintained a strong anti-immigrant rhetoric. Lega's Manifesto of 2018 condemned NGO operators in the Mediterranean Sea and stated that immigration should be based on cultural proximity.

Fratelli d'Italia is a nationalist party that emerged in 2012 from a separation from the center-right coalition. Giorgia Meloni took the lead since 2014. The party's identity is associated with a conservative, nationalist ideology of post-fascist heritage. On migration issues, it took an anti-reception position. FDI had been part of the same center-right coalition as Lega in the 2013 general elections. While failing to reach a 2% vote share in the 2013, 2018's general elections saw the party's electorate increase more than two fold, exceeding 4% of the vote. In the elections of 2022, the party's success sparked: it was the most voted in the right-wing coalition, leading Giorgia Meloni to become prime minister.

D Geocoding News in the Baseline

Key to the empirical strategy of this paper, each local news outlet considered must be assigned to a local area it serves. Online newspapers are in principle readable by anyone regardless of their locations, but local outlets tend to provide content on traffic, events, and advertisements, targeting a particular area they establish to cover. Coverage data are not directly available: we do not observe individuals' consumption of local news in the sample of interest. Therefore, Google Trends serves as a proxy to circumvent this issue, as it provides information on how specific terms are searched on Google by municipality. Using this information, exposure to local news for each municipality is approximated with a two-steps procedure.

1. Google Trends is used to obtain search metrics for each news source across municipalities. It provides:

i) a rating of users' traffic by municipality, scaled from 1 to 100, given a time-range. The platform normalizes search results by location and time range to avoid artificially giving too much weight to areas with overall high search volume. The resulting scaled outcome is therefore a measure relative to all searches on all topics.³¹. There are two time ranges considered in i): January 2011-January 2015 and June 2015-January 2020. Google Trends censores information for municipalities where traffic is too low. As a consequence, selecting a broad enough time range limits this censoring to some extent. To give a rating example, Sanremonews was mostly searched in the municipalities of Taggia (Google Trends rating =100) and Sanremo (Google Trends rating =75) for the 2011-2015 period. Let's call procedure i)'s resulting metrics $rank1_{ism\tau}$: the Google Trends' rating of source s over the sum of all s for $m, \tau \in \{2010 - 2015; 2015 - 2020\}$. Figures D.2 and D.3 show Google Trends search results for the sources with most search traffic, in the time-ranges 2010-2015 and 2015-2020, respectively.

ii) To better capture time-variation, data on source traffic over time for the whole region is retrieved, using a benchmark source for reference (i.e. outlet IVG, which reaches the highest relative traffic in the series period). For each month-year period, the traffic for a source is rated again on a scale of 1-100. The resulting cross-time metric is $rank_{2ist}$, i.e., the Google Trends rating of source s at month-year t.

Importantly, news sources with less traffic did not return municipality level information from Google Trends. Their weight is zero in the benchmark analysis.

To recap, i) varies by municipality and ii) varies by time (month-year). Combining this information, the weight of article *i* for municipality *m* will be the product of a) its importance returned by the cross-municipality rating and b) its importance returned by the time-varying rating. Weights are normalized to add up to 1 for every panel-month-year unit. More precisely: the *weight* w_{ismt} of article *i*, published in source *s*, for municipality *m*, month-year *t* equals $w_{ismt\tau} = \frac{rank1_{ism\tau}*rank2_{ist}}{\sum_{i \in m, t}(rank1_{ism\tau}*rank2_{ist})}$.

2. In a second step, a simple imputation method expands on the initial proxy to a set of municipalities not appearing in Google Trends. Municipality m_1 's information is imputed from its nearest neighbor municipality m_2 , where neighborhood is established based on commuting patterns. Data on commuting patterns at municipality level for 2011 is sourced by ISTAT, with information on how many individuals residing in municipality m_1 commute to

 $^{^{31} \}rm https://support.google.com/trends/answer/4365533? hl=en$

work or study in municipality m_2 . If the percentage of commuters among the residents is less than 25%, data are not imputed.³² Otherwise, information for municipality m_1 is proxied with the average across commuting destinations $m_k \in K$, where the set K is such that the difference between the share of all commuters and the share of commuters for which data on m_k is available is no more than 20%. Information for m_1 is imputed as the weighted average of information from observed js, where importance weights are the commuting shares. This procedure is reiterated and leads to data as in Figure D.1. Table A.1 lists all 132 municipalities for which news data has been geo-localized in both the periods before and after June 2015.





Notes: Data visualization of the news sample. On the top map are data related to the pre-treatment period, while the post treatment period is illustrated on the bottom. News coverage patterns result from the geo-localization process described in the paragraph and pertain to the provinces of Imperia and Savona, which are respectively the closest and second closest provinces to the French border in the region of Liguria.

 $^{^{32}}$ The rationale is that the higher the level of commuters, the more likely the news at the commuting destination may reach the commuters.



Figure D.2. Local news sources, Google trends data. Pre-treatment







Notes: Local news sources and geographic interest of users, based on Google Trends data for the period ranging 2010 to 2015.



Figure D.3. Local news sources, Google Trends data. Post-treatment

Notes: Local news sources and geographic interest of users, based on Google trends data for the period ranging June 2015 to 2020.

E Text-Based Measures

E.1 News cleaning

To obtain the final sample of migration-related news, extracted articles that only matched with keyword "foreigner" were further screened. This step was implemented to avoid the presence of articles that would be, for instance, tourism-related and captured unintentionally with the use of this keyword. To clean this part of the corpus, a topic model was constructed via Latent Dirichlet Allocation (DTA), following the seminal contribution of Blei et al. (2003) and the R-package topcimodels by Hornik and Grün (2011). Topic modeling is a methodology that allows to classify documents into categories along an unsupervised approach. LDA is arguably the most common of such methodologies, and involves a Bayesian mixture model that treats each article as a mix of a set of κ topics, and each topic as a mix of ω words. Differently from a clustering method that would strictly assign each article to one single group, here each news item is allowed to be assigned to several topics, with some proportion. The number of topics must be fixed a priori, it was set to 7 in this application³³. Let a corpus be formed of $i \in \{1, ...N\}$ documents, containing text that belongs to a dictionary of $j \in \{1, ...M\}$ terms. Each document i is formed by a subset of M_j terms.

Following Hornik and Grün (2011), the model specification assumes the following generative process:

The term distribution β is determined for every topic by:

$$\beta \sim Dirichlet(\delta) \tag{5}$$

The proportions θ of the topic distribution for the document i are:

$$\theta \sim Dirichlet(\alpha)$$
 (6)

For every j term,

- Choose a topic $z_j \sim Multinomial(\theta)$
- Choose a word j from a multinomial probability distribution conditioned on topic $z_j : p(j|z_j, \beta)$.

The estimation methodology follows from a variational expectation-maximization (VEM) algorithm, where the alpha parameter is estimated (see Hornik and Grün (2011) for details). To summarize the resulting topic distribution, a plot of the most recurring terms for each topic is presented below in Figure E.1. From this visualization, the 7 groups can be identified as relating to the following subjects:

topic 1: labor market (with turism)
topic 2: crime, drugs
topic 3: events, competitions, cerimonies
topic 4: local governance
topic 5: crime prosecution
topic 6: cultural events
topic 7: tourism sector

As a decision rule, if the probability for document i to relate to topic 1, 3, 6 or 7 is above 90%, that document is excluded as these topics are the least likely related to migration (keep in mind that no other migration-related keyword other than "foreigner" is found for these articles). This process excludes 9,580 articles from the initial sample of news.

 $^{^{33}\}mathrm{A}$ higher κ appeared to produce redundant topics



Figure E.1. LDA model, frequent terms by group.

Notes: Tables of most frequent terms in each topic distribution.

E.2 Media Slant

Among the prominent contributions in text-based measures of media slant, Groseclose and Milyo (2005) constitute a seminal example. The authors use US Congressional citations to estimate the political positions of think tanks, and then use their mentions to predict media partisanship in twenty news outlets. Gentzkow and Shapiro (2010) build on this method and obtain media slant indices for 433 outlets, by enlarging the set of terms from sole think tank citations, to a broader set of text occurrences, selected in terms of relevance via a χ^2 statistic. Given this feature selection, the authors predict slant via a two-step supervised generative model. The broad idea behind Groseclose and Milyo (2005) and Gentzkow and Shapiro (2010) is to use text issued from politicians' discourse, whose affiliation is known, as a training set to predict the ideology of newspapers, unknown a priori but inferable from how similar it is to the political speeches. More recently, Taddy (2013) improves on these methodologies by proposing a model in which a multinomial inverse regression reduces the dimensionality of the predictors (terms) to a univariate score. A forward regression step then predicts the response (slant labels in this context), given this reduced information. In this application, Taddy (2013) is benchmarked against an alternative elastic-net penalized logistic regression (Friedman et al., 2009, 2010) and found to perform very similarly in a validation step. To construct the training set, this study relies on a set of four news sources whose slant (or lack thereof) is taken as evident. This constitutes a difference from Groseclose and Milvo (2005) and Gentzkow and Shapiro (2010), who employ politicians' speeches as their predetermined, labeled sample. Employing news to train the model guarantees that they belong to the same journalistic dimension as the unlabeled, target news set. Other than this comparability argument, a motivation for this choice is data availability. Use of Italian political speeches would present data availability limitations, further complicated by the complexity of a fractionalized political scenario that does not directly reduce to the more tractable bipolar *Democrats-Republicans* setting adopted in the US studies. A discussion of the Italian political landscape is detailed in appendix C.

The classification proposed here involves the construction of a simple binary variable, where a value of 1 corresponds to the detection of similarity with anti-immigrant discourse and 0 to the lack of it. The *training sample* is constructed as follows.

Anti-immigrant labeled news items in the training sample derive from national newspapers La Verità and ll Giornale. The former outlet was founded in 2016 and openly takes an anti-immigrant stand. This is documented, for instance, in the online section discussing the migration crisis context, entitled Chronicles of Invasion (it. Cronache dell'Invasione). According to survey statistics diffused by the market research platform YouGov,³⁴ this news outlet is considered right-wing by 81% of respondents. Of these, 17% classify it as far-right, the highest percentage among all news outlets in the list. 12% of people classify it as center oriented or non-oriented, while the remainder considers it a left-wing source. Data from La Verità are collected from its online platform by considering all news that is included into the Chronicles of Invasion (it. Cronache dell'Invasione) section, and complement this extraction with migrant-related news by a similar keywords-search fashion as for my main dataset. The latter source ll Giornale shares similar Yougov reviews with La Verità and also takes an evidently critical rethoric against migration.³⁵ Data from the ll Giornale are

³⁴https://it.yougov.com/news/2019/05/03/giornata-mondiale-della-liberta-di-stampa/

 $^{^{35} \}mathrm{An}\ \mathrm{headline}\ \mathrm{example}\ \mathrm{from}\ \mathrm{February}\ 2022:\ https://www.ilgiornale.it/news/politica/carola-leroina-fuori-legge-2007541.html:\ ``Carola-leroina-fuori-legge-2007541.html:\ ``C$

collected from the online platform by considering all news from the same keyword searches used for the local news sample. These news articles take label value 1.

Non right-wing labeled news in the training sample derive from two sources. One is the website L'Unità news, self-defined as a not-for-profit news aggregator of antifascist and anti-racist information. The website also hosts the archives for and shares the name with the discontinued newspaper L'Unità founded by Antonio Gramsci in 1924 as the official news outlet for the Italian Communist Party. News published in this online platform are extracted, following the keywords-search fashion as for the main dataset. The second source is the online version of regional newspaper Il Secolo XIX. This outlet was founded in Genova, Liguria in 1886. Since 2015, the outlet had been affiliated with the national source La Stampa and since 2017 is part of editorial group GEDI, which includes, among others, the daily La Repubblica and the weekly L'Espresso.³⁶. News collection for Il Secolo XIX follows the same keyword-based steps. Articles coming from these two sources are labeled with a 0 in the binary slant measure.

Via this procedure, the training set is constituted by 5926 0-labeled and 5293 1-labeled instances, which will serve to classify the main news dataset. This study proposes two classification approaches: i) a penalized maximum likelihood logistic model with elastic-net regularization (Zou and Hastie, 2005; Friedman et al., 2010), with 10-fold cross-validated tuning parameters λ (the penalty parameter) and α (the level of mixing between lasso and ridge, with $\alpha = 1 \implies$ full lasso). This is compared with ii) the alternative Taddy (2013)'s methodology.³⁷

For every article in the training set, a document-term matrix is constructed as described in section 4.2, with rows constituting the training set articles *i* and columns constituting the n-grams *j* ($n \in 2, 3$) present in the articles. Preprocessing steps are performed via the R package *text2vec*³⁸. The vocabulary of n-grams is pruned via a feature-selection based on the χ^2 statistic of Gentzkow and Shapiro (2010). For an idea of the matrix content, Figure E.2 summarizes the frequencies of some top frequent n-grams.

 37 The two measures share the same prediction in roughly 86.7% of instances.

 $^{38} \rm https://CRAN.R\mbox{-}project.org/package=text2vec}$

Rackete, the hero of the radical-chic, rightwing world, reveals that she was even more Taliban than the reception of the German NGO Sea Watch."

³⁶According to the aforementioned YouGov report, La Stampa is perceived as center-oriented by 33% of respondents and left-wing by 39%. La Repubblica is considered left-wing leaning by 79% of the interviewed.



Figure E.2. Document-term matrix, sample of frequent terms.

Notes: Frequency table of most frequent terms from the training sample.

In Friedman et al. (2010)'s method, the document-term matrix cells will contain the metric $tf - idf_{ij}$. the objective function for logistic regression³⁹ is the penalized negative binomial log-likelihood function, given N sample size and p parameters:

$$min_{(\beta_0,\beta)\in R^{p+1}} - \left[\frac{1}{N}\sum_{i=1}^{N}y_i(\beta_0 + x_i^T\beta) - log(1 + e^{(\beta_0 + x_i^T\beta)})\right] + \lambda\left[(1 - \alpha)||\beta||_2^2/2 + \alpha||\beta||_1\right]$$
(7)

On the right, the penalty severity is determined by the tuning parameter λ . This regularization is a convenient approach for models with a high number of parameters, which may suffer from over-fitting problems. The elastic-net mixes between two penalty functions: the ridge-regression penalty ($\alpha = 0$) and the lasso-regression penalty ($\alpha = 1$). The motivation behind this mixing lies on the aim to exploit the qualities of each approach, while overcoming their limitations if considered singularly. lasso allows for a better parsimony of the model as it shrinks some parameters to zero. In case of high correlations between predictors, however, lasso may omit relevant information, as would only

³⁹Given an observed binary outcome $Y \in 0, 1$, the logistic model can be written as $\frac{p(Y=1|X=x)}{p(Y=0|X=x)} = \beta_0 + x^T \beta$.

select one coefficient, ignoring the rest. Friedman et al. (2010) suggest that $\alpha = 1 - \epsilon$ for small $\epsilon > 0$ may already be a convenient solution to avoid lasso degeneracy and conserve its sparsity property. In this study, hyper-parameters λ and α are chosen from a 10-fold cross-validation procedure on the training sample. Figure E.1 displays the levels of α (on the left) and λ (on the right) that minimize the misclassification rate (on the y axis).

Table E.1. Tuning parameters for the Elastic net.

alpha	lambda
0.100000	0.008216

Notes: Cross-validation results. Best tuning parameters.

Binary classification returns probability scores, where a probability corresponds to the estimated degree of likelihood that the article is labeled as 1 (= right-wing). In the benchmark, for a probability score higher than 0.5, the article is assigned the value 1 (some variation on this threshold is considered in the extended results).

In Taddy (2013)'s procedure, for each document i, equipped with term counts $\mathbf{x}_i = [x_{1i}, ..., x_{Ji}]'$, the n-gram total counts for labels $y \in \{0, 1\}$ are defined as $x_y = \sum_{i:y_i=y} \mathbf{x}_i$. Given row sums $m_i = \sum_{j=1}^J m_{ij}$, the multinomial inverse regression would then be:

$$\mathbf{x}_{y} = MN(\mathbf{q}_{y}, m_{y}) \text{ with } q_{yj} = \frac{exp[\alpha_{j} + y\phi_{j}]}{\sum_{l=1}^{J} exp[\alpha_{l} + y\phi_{l}]}, \ j = \{1, ..., J\}, y \in \{0, 1\}$$
(8)

Where MN is a j-dimensional multinomial distribution with size $m_y = \sum_{i:y_i=y} m_i$ and probabilities $q_y = [x_{y1}, ..., x_{yJ}]'$. Under a set of conditions, a sufficient reduction score for frequency $\mathbf{f}_i = \mathbf{x}_i/m_i$ is defined as $\mathbf{z}_i = \phi' \mathbf{f}_i$, $\implies y_i \perp \mathbf{x}_i$. Given this score, the forward regression would simply be a univariate model, which in the present case, take logistic form: $E[Y] = \frac{1}{e^{\beta_0 + \beta_1 z_i}}$. The author proposes a so-called gamma-lasso algorithm to find a joint maximum a posteriori estimate of both coefficients and their prior scale, by maximizing the log-likelihood of the model with a lasso penalty, with λ_j possibly varying with j. Further details on this estimation are in the author's contribution. Although the two alternative models produced very similar labeling, a set of validation steps leads to the adoption of Taddy (2013)'s measure as the baseline slant index. Figure E.3 represent some predictive terms, with associated scores.



Figure E.3. Most predictive coefficients, Taddy 2013.

Notes: Examples of terms used in Taddy (2013) classification. Positive x values indicate higher term impacts in absolute terms on the slant predictions.

Given this binary classification of documents, average indicators are computed at month-year level for each municipality. These aggregate averages then can be interpreted as a proportion of right-slanted news out of migrantrelated news to which a particular municipality was exposed. Importantly, given the news geo-localization described in the previous section, a document (news piece) can be read in more municipalities. When constructing these averages, importance weights were used according to the Google Trends based information, as detailed in section D.

E.3 Media Slant Validation and Exploration

This section explores the validation procedures undertaken for the construction of the anti-immigrant slant index.

To maximize the prediction accuracy for the labelling process, models were cross-validated on the pre-labeled sample. A 10 fold cross-validation procedure was performed on both the baseline and the alternative anti-immigrant slant index. Table E.2 below presents cross-validation results. Generally, the two measures appear to be fairly comparable: the elastic-net measure appears in the first row and has an estimated accuracy of 0.85. The accuracy for Taddy's measure, in the second row (i.e., the proportion of correct classifications over the total sample size) is estimated to be 0.87. Both models present higher levels of specificity than sensitivity. Both values are above 80% for both measures. All in all, these statistics appear to favor the use of Taddy (2013)'s index over the Elastic-net, as a benchmark indicator.

Table E.2. C	Cross-validation	results
--------------	------------------	---------

model	$^{\rm cv}$ _acc	cv_acc_sd	cv_{sens}	cv_{fscore}	cv_spec
Elnet	0.85	0.01	0.83	0.84	0.86
Taddy	0.87	0.01	0.85	0.87	0.89

Notes: Performance results from the two classification models, based upon a 10-fold cross validation, done manually for Taddy (2013). Columns present, from left to right, the following statistics: crossvalidated accuracy (cv_acc) and its standard deviation (cv_acc_sd); sensitivity (cv_sens), F-score (cv_fscore) and specificity (cv_spec).

As a further measure of validation, results were also compared with human classification of roughly 900 random entries. This sample was extracted from the main corpus, with source-based stratification. The same text was analyzed by two *Amazon Mechanical Turk* annotators. Having multiple annotators for one task is key to capture response variation, which may also reflect subjective biases in the response. Interestingly, this procedure revealed some evidence that human classification of slant is likely heavily influenced by predetermined ideology in a general crowd: human classifiers did not agree on slant in 27 % of the instances. Below, the full instructions entered in the platform are provided.

ORIGINAL TEXT:

Instructions: Per favore, legga il testo qui sotto elencato. Le si chiederà di interpretare se il contenuto della news ha un'accezione politica. Recentemente, il dibattito politico in Italia si è incentrato molto sul tema della migrazione. In generale, i partiti a destra sostengono un'ideologia critica riguardo ai recenti sbarchi dei richiedenti asilo, e del loro arrivo in Italia, affermando per esempio problemi sicurezza emergono per i cittadini. Il centro sinistra si è visto più propenso all'accoglienza. Le viene domandato se l'articolo afferma, anche generalmente/indirettamente/secondariamente, il tema della migrazione. Per esempio se l'articolo riporta un furto ad opera di individui stranieri presenti in Italia, il tema della migrazione, per quanto indirettemente, è presente. Se l'articolo parla soltanto di eventi cultura o di turismo, in quel caso non affronta il tema della migrazione. Le sarà poi chiesto di valutare se da chi ha scritto l'articolo si evince un tono vicino ad un'ideologia politica (e quale). Per esempio: -se l'autore sceglie il termine "nullafacenti" per descrivere dei cittadini stranieri/ migranti, si evince un tono critico sulla migrazione, e quindi maggiore vicinanza al discorso di destra. -Se l'articolo giudica importante o necessario un evento di accoglienza organizzato nel Comune di riferimento, questo tono è piu' vicino al discorso di sinistra. Infine le viene domandato se il SOGGETTO dell'articolo, di cui il testo parla, ha un'inclinazione ideologica e quale. Per esempio, se il testo menziona la decisione di un sindaco di non approvare un progetto di accoglienza nel nome di una maggiore sicurezza dei cittadini, il soggetto dell'articolo si categorizza di un'ideologia più a destra. (NOTA:La vicinanza politica dell'autore puo' differire da quella del soggetto del testo). Un'ultima sezione le chiede un breve commento sulla sua scelta, basta la citazione di un punto del testo, una parola chiave o una sua breve frase che confermi che le risposte sono inserite a seguito di una sua valutazione.

Testo dell'articolo:[...]

Questo articolo parla, anche generalmente, di migrazione? Risponda con SI o NO.

I fatti descritti nell'articolo sono delineati neutralmente o con un'accezione ideologica? Risponda con N se neutralmente, e con O se pensa esista del contenuto ideologico.

Il tono di questo articolo ha una somiglianza con il discorso politico di destra/centro-destra, di sinistra/centro-sinistra o nessuno dei due? Risponda con una delle tre opzioni: DESTRA; SINISTRA o NEUTRO.

Questo articolo menziona l'opinione o i fatti di una persona (un politico o cittadino o un intervistato) di destra/centro-destra, di sinistra/centro-sinistra o nessuno dei due? Risponda con una delle tre opzioni: DESTRA; SINISTRA o NEUTRO.

Indichi una frase o parola del testo che motivi la sua scelta. Basta una breve risposta. Se l'articolo non parla di migrazione (risposta 1), puo' inserire NOMIG.

ENGLISH VERSION:

Instructions: Please read the text below.

You will be asked to interpret if the news content has a political meaning. Recently, the political debate in Italy has focused a lot on the issue of migration. In general, the parties on the right support a critical ideology regarding the recent landings of asylum seekers, and their arrival in Italy, stating for example security problems arise for citizens. The center left has seen more inclination to their reception. You are asked if the article affirms, even generally / indirectly / secondarily, the theme of migration. For example, if the article reports a theft by foreign individuals present in Italy, the theme of migration, albeit indirectly, is present. If the article speaks only of cultural or tourism events, in that case it does not address the issue of migration. You will then be asked to evaluate if the writer of the article reveals a tone close to a political ideology (and which one). For example: -if the author chooses the term "deadbeat" to describe foreign citizens / migrants, there is a critical tone on migration, and therefore greater proximity to the right-wing discourse. -If the article judges a welcome event organized in the relevant Municipality to be important or necessary, this tone is closer to the left-wing discourse.

Finally, you'll be asked if the SUBJECT of the article, of which the text speaks, has an ideological inclination and which one. For example, if the text mentioned a mayor's decision not to approve a reception project in the name of greater citizen safety, the subject of the article categorizes with a more right-wing ideology. (NOTE: The political closeness of the author may differ from that of the subject of the text).

A last section asks you for a brief comment on your choice, you just need to quote a point in the text, a keyword or a short sentence from it to confirm that the answers are inserted after your evaluation.

Article text:[...]

Does this article also talk about migration, even generally? Answer with YES or NO.

Are the facts described in the article outlined neutrally or with an ideological meaning? Answer with N if neutral, and with O if you think there is an ideological content.

Does the tone of this article bear a resemblance to right / center-right, left / center-left political discourse, or neither? Answer with one of the three options: RIGHT; LEFT or NEUTRAL.

Does this article mention the opinion or facts of a person (a politician or citizen or an interviewee) on the right / center-right, left / center-left or both? Answer with one of the three options: RIGHT; LEFT or NEUTRAL.

Indicate a sentence or word of the text that motivates your choice. A short answer is enough. If the article does not talk about migration (answer 1), you can enter NOMIG.

The questionnaire allowed for a different answer for two potentially different cases. 1) the author of the article wrote with some ideological inclination; 2) the subject of the article involves a person's opinion (or facts) with a political tendency. For the scope of this study a broad definition of media bias is adopted: no distinction between the two cases is investigated, and giving voice to third persons with a political orientation is considered on the same level of writing with that orientation (recall, as broadly similar approaches, Groseclose and Milyo (2005) who define media-slant based on think-tank citations; Gentzkow and Shapiro (2006) estimating informative words via a data-driven approach, Taddy (2013)'s refinement; Flaxman et al. (2016) who propose a non text-based measure, based on the geographic distribution of readership and voters rather than on text; Djourelova (2023) who focuses on the banned use of the expression "illegal immigrant").

Slant indices and human classification agreed in 66% of cases. Figure E.4 plots the distribution of the baseline classification's predicted probabilities, against the evaluation of human annotators. This last, takes values 1 if at least an annotator coded the article as "right-wing", 0 if the article was coded neutral (or the two annotators gave opposite marks), -1 if the annotator assigned the article to be "left-wing" oriented. Both neutral and left-wing marked articles have their higher mass at low predicted probabilities.



Notes: Densities of predicted probabilities in the benchmark classification (density functions), against human classification (y-axis factors).

All in all, this results point at a general validity of the adopted measures for slant. While the general behavior is as expected, measurement error exists, and several approaches to circumvent it are taken in the main analysis.

F Municipalities in Sample

Table A.1. List of municipalities in the baseline panel

Municipalities:	Airole; Alassio; Albenga; Albisola; Albissola; Altare; Andora; Apricale; Aquila D'arroscia; Armo; Arnasco; Aurigo
	Badalucco; Bajardo; Balestrino; Bardineto; Bergeggi
	Boissano; Bordighera; Borghetto D'arroscia; Borghetto Santo Spirito; Borgio Verezzi; Borgomaro; Bormida
	Cairo Montenotte; Calice Ligure; Calizzano; Camporosso; Caravonica; Carcare; Carpasio
	Casanova Lerrone; Castel Vittorio; Castelbianco; Castellaro; Castelvecchio Di Rocca Barbena
	Celle Ligure; Cengio; Ceriale; Ceriana; Cervo; Cesio
	Chiusanico; Chiusavecchia; Cipressa; Cisano Sul Neva; Civezza; Cosio D'arroscia; Cosseria; Costarainera
	Dego; Diano Arentino; Diano Castello; Diano Marina; Diano San Pietro; Dolceacqua; Dolcedo; Erli; Finale Ligure
	Garlenda; Giustenice; Giusvalla; Imperia; Isolabona; Laigueglia; Loano; Lucinasco
	Magliolo; Mallare; Mendatica; Millesimo; Mioglia; Molini Di Triora; Montalto Ligure; Montegrosso Pian Latte; Murialdo; Nasino; Noli
	Olivetta San Michele; Onzo; Orco Feglino; Ortovero; Osiglia; Ospedaletti
	Pallare; Perinaldo; Piana Crixia; Pietra Ligure; Pietrabruna; Pieve Di Teco; Pigna; Plodio
	Pompeiana; Pontedassio; Pontinvrea; Pornassio; Prelà; Ranzo; Rezzo; Rialto; Riva Ligure; Roccavignale; Rocchetta Nervina
	San Bartolomeo Al Mare; San Biagio Della Cima; San Lorenzo Al Mare; Sanremo; Santo Stefano Al Mare; Sassello; Savona
	Seborga; Soldano; Spotorno; Stella; Stellanello
	Taggia; Terzorio; Testico; Toirano; Tovo San Giacomo; Triora; Vado Ligure; Vallebona; Vallecrosia; Varazze; Vasia
	Vendone; Ventimiglia; Vessalico; Vezzi Portio; Villa Faraldi; Villanova D'albenga; Zuccarello

Notes: Final list of Municipalities in the panel, resulting from the news geo-localization.

INSTITUT DE RECHERCHE ÉCONOMIQUES ET SOCIALES

Place Montesquieu 3 1348 Louvain-la-Neuve

ISSN 1379-244X D/2023/3082/21



