

Peracchi, Silvia

Working Paper

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

CESifo Working Paper, No. 10070

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Peracchi, Silvia (2022) : The Migration Crisis in the Local News: Evidence from the French-Italian Border, CESifo Working Paper, No. 10070, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/267303>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

Silvia Peracchi

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

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

Abstract

This paper investigates the impact of local exposure to the migrant crisis on the local news market. Exploiting a narrow geographical setting, it explores a policy dating from June 2015, whereby French authorities introduced militarized controls at the Italian frontier. With the border controls in place, groups of migrants and asylum seekers who had planned to cross the border irregularly were pushed back to the Italian lands. With rejected migrants clustering at the border, natives residing along the Italian region were unevenly exposed to their settlement. Taking advantage of this unequal treatment as a natural experiment, this study uses novel data collected on the text and on the number of local news items for the border areas of Liguria, Italy, between 2011 and 2019. It documents that the backlog of migrants in the Italian border area was substantially mediatized: coverage of migration rose most in the most exposed municipalities. Conversely, anti-immigrant discourse in the news grew more in areas least directly in contact with the border. Exploring further this framing dimension, the bias effect turns out to be shaped by readers' demand and to be closely associated with local news penetration. Finally, this study documents that anti-immigrant slant and voting preferences share a similar broad direction, while a related broad pattern also appears in hate-crime records.

JEL-Codes: F220, L820, F500.

Keywords: media slant, EU borders, immigration, diff-in-diff.

Silvia Peracchi
Department of Economics and Management
DEM, University of Luxembourg
6 Rue Richard Coudenhove-Kalergi
1359 Luxembourg
silvia.peracchi@uni.lu

October 31, 2022

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 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 is part of the PhD thesis of the author, which is supported by the Luxembourg National Research Fund (FNR) (PRIDE15/10949242*). All outputs of this study are the sole responsibility of the author.

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, experienced 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; Djourelouva, 2020; Gentzkow, 2006). Despite this, media companies do not exogenously select the message to spread. When exploring the news market in particular, competition and readership beliefs are key factors in determining the supply of news information (Mullainathan and Shleifer, 2005).

This study examines these news production mechanisms and investigates them along two dimensions. It first poses the question: *How do local shocks in the presence of backlogged migrants impact the local news market?* Second, *Which news patterns, if any, match the local political economy?* 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 based on two choices: the volume and quality of discourse connected to migration.

The sudden shock of interest occurred in June 2015, when French local authorities systematically imposed militaristic border controls at the border with Italy. This policy aimed at and resulted in substantial push-backs of migrants to Italian lands, which, for morphological reasons, funnelled migrants predominantly to the coastal border into the town of Ventimiglia, Liguria (Italy). To better understand the geographic connection between France and Italy, see Figure 1. 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 geography of the region, showing how mountains make the coastal area more accessible.²

¹For 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).

²As a complement, an additional map realized by NGO La Cimade is reported in Figure A.1.1. 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 policy of border restrictions, many migrants in dire circumstances in Italy had taken advantage of this gateway in an effort to travel to their eventual destinations in other European nations.³

This study is grounded in a localized methodology that takes into account the unique geography of the imposition of border restrictions and the overall migration patterns of the area. It focuses on the local government units of the coastal border region of Liguria, and specifically on the border province, Imperia, and its adjacent province, Savona. For the identification strategy being used, this is essential. In this setting, locals' exposure to the migrant settlements is inversely related with how far they must commute to get to the border, where the incidents occurred. Based on that, distance can be used as a proxy for how municipalities are exposed directly to the border events. To better track changes in local information demand, local news is the information channel of interest within this local dimension. Given that the internet is a widely used technology in the region, this study focuses on online news.⁴ The analysis is based on a collection of about fifty 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"). Then, using Google Trends, articles are geo-localized and combined to create a municipality-month-year panel dataset that spans the years 2011 to 2019.

The impact of the settlement of migrants on local news is examined along two lines: 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 first dimension,

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.1.6 for internet usage rates over time in the region of interest, compared to all Italy).

the core outcome measure is a count of articles related to migration in the local news.⁵ For the second dimension, this study builds upon the literature on measuring 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).⁶

Specifically, this paper uses an algorithm to determine whether a local news article has anti-immigrant terminology 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 index of Groseclose and Milyo (2005), who compared congressional citations of US think tanks to citations of think tanks in newspapers, and the more sophisticated techniques of Gentzkow and Shapiro (2010) and Taddy (2013), which detected slant using the entire language in the Congress and in the news. In this paper, the local news slant is predicted using a body of national and regional newspapers with or without a known anti-immigrant orientation. This body constitutes the reference against which to compare the text in the local news pool. News articles in this pre-labeled dataset are carefully selected. Lacking comprehensive data on the text of the politicians, this enables the adoption of comparable machine learning algorithms by relying on rich text contents. This paper’s findings can be summarized as follows. The backlog of migrants led to a notable prominence of migration in the local media discourse. It caused a differential increase in news coverage that declines by 6% for every additional 10 minutes of commute time from the border. The share of anti-immigrant news, however, is observed to grow more in the municipalities further away after the migrants’ forced settlement at the border. For each additional 10 minutes increase in commute time from the border, anti-immigrant slant increases by 0.18 standard deviations, after the border closure. Further evidence proposes a set of mechanisms to rationalize this finding.

First, the role of the demand for news is found to be crucial. This finding emerges when comparing the results by giving more or less weight to the distribution of local news 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, the bias in the consensus is uniformly reflected in the news.

In this context, the direct exposure of locals to migrant settlement at the border may have increased opinion heterogeneity, pushing slant down in aggregate terms. This variability may have also been channeled by the area’s

⁵A second, less central measure is also presented: an index of news importance. This is taken from the patents literature (Kelly et al., 2021), and enables the capture of a perceptive description of text evolution over time. An *important* article is *impactful* (its text resembles future articles) and *novel* (its text is dissimilar from that in previous articles).

⁶Importantly, 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.

recorded presence of humanitarian help from both local and international NGOs (La Cimade, Oxfam, Caritas and Doctors Without Borders, to cite a few). The ideology remained more homogeneous the further away from the actual events, which contributed to the predominance of the anti-immigrant tone. It is profitable to accommodate beliefs since it allows local outlets to gain market from national outlets. This finding is also broadly in line with George and Waldfogel (2006)’s argument that local news may adjust their target, given national news penetration. The contact-versus-mere-exposure paradigm put forth in Steinmayr (2021) is another related perspective. Directly seeing the emergency may have led some of the local news consumers to empathize with the humanitarian difficulties, in the communities closest to the border. As remote municipalities lacked the direct contact element, a general negative reaction appeared to prevail. Accordingly, a graph of hate crime incidents that were measured à la Romarri (2020) demonstrates that hate crime trends increased more substantially in the farther province. Moreover, this framing shift is documented to matter for voting patterns for the general elections, which display a relative loss in left-wing support. According to studies such as Benesch et al. (2019); Djourelouva (2020) and Keita et al. (2021), news powerfully influences 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 A.2.1 for a set of these papers). The media component, key in this study, is touched on superficially in these references. Finally, while utilizing 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).

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 asylum migration. It also connects to a number of political science and political economics contributions that examine the evolution of the news market as well as 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 partisan 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 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 markets not only reflect but also shape personal preferences. Prat and Strömberg (2013) overview the 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).⁷ Importantly, 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 (2020) is the effective relationship between shocks on public attention and rescue efforts in the Mediterranean. Djourelouva (2020) 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. (2021) 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 (2020), the views of destination countries on immigration can be framed along two basic 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, some contributions instead observe a positive reception or more nuanced effects, found prevalently in cases where natives are exposed to newcomers with a repeated interaction. For this research, the 2010s were of specific interest, with the EU refugee crisis becoming a priority on the political agenda of EU countries. Table A.2.1 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

⁷See 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.

acceptance, which is broadly in accordance with the contact theory of social psychology (Allport, 1954; Pettigrew, 1998). 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 left-wing support.

This paper departs from these contributions in several aspects. First, the geographic dimension differs. The interest here zooms into the region near the French border, thus drawing attention to events taking place at an internal EU border. Cross-country borders within the EU area were important settings in the light of the migratory crisis,⁸ they have been receiving less institutional attention than the exterior EU borders.⁹ 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 in-depth investigation of this crucial yet underexplored information channel.

3 Context

3.1 Italy’s Migrant Crisis and the Controls at the French Border

During the 2010s, several undocumented migrants landed in Italy with the intention of rejoining family members and contacts in other European countries.¹⁰ In 2014, an allocation policy was implemented by the government to strengthen the pre-existing, insufficient reception system. However, several migrants attempted to continue the journey instead of having their status regularized (Capitani, 2018; Colombeau, 2020).¹¹ A typical gateway employed by migrants was located on the coastal border between France and Italy, in the north-western region of Liguria.¹² 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 Intemelio recorded two hundred migrants seeking refuge on the streets; nonetheless, these numbers

⁸see Cimade (2018).

⁹Frontex only collects data on illegal border crossings at the external EU frontiers. <https://frontex.europa.eu/we-know/migratory-map/>

¹⁰The 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.

¹¹Once landed in Southern Italy, migrants could submit an asylum application. As importantly mentioned in Campo et al. (2021), this procedure is remarkably lengthy.

¹²Others 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 the Ventimiglia region 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 and is still in effect as of 2022) with the intention of cooperating with Libyan coast guard authorities to prevent the boat lifts. Sourced by NGO La Cimade, Figure A.1.1 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 Intemelina collected a count of meals distributed to migrants. Figure A.1.2 in the Appendix reports a peak of 12500 migrants transiting in the area in July 2017. For comparison, Ventimiglia's population is approximately twenty-four thousand residents (Italian National Institute of Statistics [ISTAT] 2019). Despite measures being taken, the situation persisted over time. With some seasonality, migrant arrivals in Ventimiglia were also observed in 2018 and 2019.

3.2 Liguria and the Provinces of Interest

Given that the events of interest occurred at the western border, commuting distance from the border proxies the likelihood that locals will be subject to the migrants' settlement. The provinces are chosen for this study because of the morphology of the area of interest as well as from the intention to maintain an internally valid specification. This limited location enables greater control against confounding occurrences of other events linked to the migration crisis in Italy. Thus, the provinces of Imperia and Savona-the border and its neighbor respectively-are selected for analysis. The question of whether the results apply to more general circumstances arises from such a localized selection.

To better understand how generalizable this study is, Figures A.1.3 and A.1.4 depict the distributions of several socio-demographic variables for Italian provinces, estimated by kernel densities. The red and blue dashed vertical lines on the graphs indicate where Imperia and Savona, respectively, lie on the distribution. As shown in Figure A.1.3, Imperia has a relatively higher percentage of cross-border commuters. This aspect importantly highlights that greater border exposure is found in the areas of higher treatment levels. Regarding the other variables, the two provinces tend to appear comparable to the rest of Italy. This holds in terms of total shares of commuters, employment rates, skills 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.1.4 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 to the right, with Imperia experiencing higher right-wing incumbency than Savona.¹³ Besides these trends, a right-wing shift did appear in the

¹³For both provinces, 2013 constituted a relatively high success for the newly born anti-system, web-spread populist movement Movimento 5 Stelle (M5S).

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%. To better understand the role of some of these patterns, results section 6.3 allows for possible heterogeneous outcomes by initial political preferences, as well as by other socio-demographic characteristics.

Figure 2. Area of interest, Liguria



Notes: A reference map of the region of interest. A close-up view of the region of Liguria is on the left. In the west is the border town of Ventimiglia. The other labels designate the capitals of each province in the region: Imperia, Savona, Genova, the region's capital, and La Spezia. Imperia's province is in blue, while Savona's province is in green. Grey-colored provinces are not in sample. The map of Italy on the right shows Liguria highlighted in blue.

3.3 Treatment and Identification

The border shutdown 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 migrants trying to enter the UK from France. The area's topography 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 unpopular because it was arduous and perilous (Welandar, 2017).

It is important to note 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 in the area were in more or less direct contact with the border, depending on their geographic proximity. For instance, the top-left graph of Figure A.1.3 shows that more cross-border workers reside in the border province (Imperia, in red), compared to the province further away (Savona, in blue). Therefore, the identification strategy adopted here is based on the likelihood of natives to be exposed directly to the border events fading away the more distant their municipality is from the border town. The preferred

distance measure is commuting time, because it better captures actual proximity given geographical differences.

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. The degree of proximity interacted with the post-June 2015 time interval will determine the treated areas. The causal identification of this approach lies in the comparability of the municipalities across distance, and in the assumption that absent the border events, the municipalities in this study would have followed equal trends. To ensure this is the case, the empirical approach will explore the balance of the sample, allow for the presence of several covariates, and perform robustness checks with respect to the presence of different province-level time trends, or different forms of proximity to capture the degree of treatment. More details follow in the sections below.

4 Data and Measurement

The dataset was created by gathering a corpus of 57,589 migration-related internet articles from 2011 to 2019. This corpus represents most local online outlets present in the two provinces of interest for the time span considered. Then, with a month-year frequency, data at the article level are aggregated to a panel of municipalities. The procedure for gathering data and the text-based metrics are given in the paragraphs that follow. Further information is supplied in the Appendix sections A.4, A.5.2 and A.5.3, for the sake of concision.

4.1 News Data

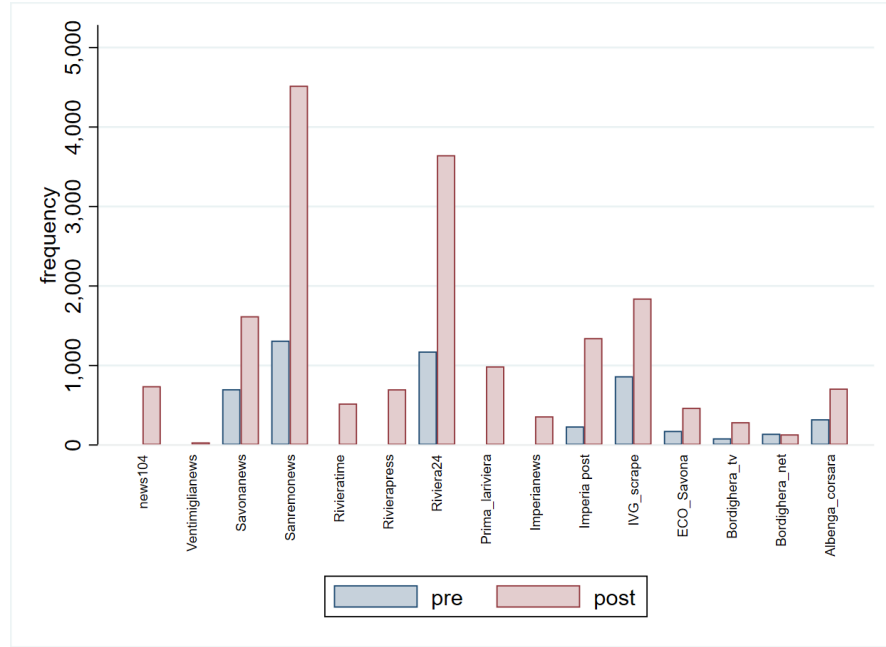
The news sample was collected with a web-crawling algorithm that individually scraped each relevant news outlet’s website. Articles were gathered for each local outlet under consideration if any of the stemmed variants of the following keywords appeared in the text: “migrant,” “immigrant,” “foreigner,” “non-EU citizen” (It., “extracomunitario”), or “displaced person” (It., “profugo”). Additionally, a second web-crawling algorithm extrapolated each website’s news count over time. Despite not formally capping the time period for data collection, 2011 to 2019 was chosen as the time period for the data due to restrictions on data availability. When choosing outlets, 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 at the municipality level. Following data collection, a geocoding process was used to map article-level data to municipalities. Working with this statistical unit enables the data to be matched with information on geography, politics, economics, and demographics.¹⁴ Since one text-based measure requires lagged or led information for its development, the final panel covers the period between January 2012 and December 2018 (details follow in this section). Simple Google searches of the kind, “Ventimiglia [or other Municipality in the area] news” were paired with Google Trends searches, restricted geographically in the Liguria region,¹⁵ to extrapolate the universe of the local online newspapers in the provinces. Google Trends Conveniently shows correlated term searches. For example, searching in Google Trends for

¹⁴Data inaccessibility excluded sources that had been discontinued at the time of data collection. To the author’s knowledge, however, their number is insignificant.

¹⁵The most granular geo-localization possible in Google Trends, for the area of interest.

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. Figure 3 displays the universe of the employed news outlets, distinguishing counts by before versus after June 2015. In sum, 57,589 articles were collected, of which 43,085 are from 2012, coming from a total of fifteen news sources. After inspection, articles matching exclusively with the keyword “foreigner” were identified to potentially capture aspects different than migration (many pertained to tourism). To clean the dataset from these unrelated news, articles that only matched with “foreigner” were analyzed through a latent Dirichlet allocation (LDA) approach (Blei et al., 2003; Hornik and Grün, 2011). Relating articles to one or more topics (at most 7), this procedure led to drop 9,580 articles from the original sample that were found pertaining to topics most likely unrelated to migration.¹⁶

Figure 3. Sample collection, text corpus by news source



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.

As mentioned above, each source’s collected articles included terms connected to migration. The process specifically extracted each article’s title, subtitle, and full text content, as well as certain metadata, including the author’s name (when provided), the article’s date, and a tag (if available). It’s important to note that not every local outlet considered serves the same areas equally. Although online news is, in principle, accessible without any geographic boundary, local outlets 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.

¹⁶Details on the procedure are found in the Appendix section A.5.1. It is advisable to read the appendix section after completing the reading on data and measurement, for a better understanding of the use of text as data.

Direct information on which local online outlets serve which areas is unobserved. However, a reasonably close proxy can be obtained by tracking Google searches for the names of the outlets over time and by municipality. These data are retrieved via Google Trends. The software returns i) a rating for each town based on the volume of searches for any term, such as “Sanremonews,” during a certain time interval. The scale for this rating, which ranges from 1 to 100, is normalized by 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 terms. This rating spans from 1 to 100 and is normalized across sources by taking a common comparison term.

With this knowledge, the measurement of news coverage of migration simply involves the log count of the number of articles that for each municipality returned a match within the given month-year. The exploration of agenda-setting 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.

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

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

1. *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 A.4. 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 A.4. 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.¹⁷ Details are presented in Appendix section A.4. Additionally, Google Trends searches for the

¹⁷A robustness test confirms the main findings without this imputation step.

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.1.5. 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.1.6 complements this information by plotting trends on internet usage in Liguria, compared with the rest of Italy, and based on ISTAT statistics.

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 Measures

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.¹⁸

The text-based measures are based on *bag-of-words* approaches.¹⁹ 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 ($tf - idf_{ij}$) will be the metric of interest for cell ij in the document term matrix. The recommended frequency statistic used in text analysis tasks is typically this measure. While mid-frequency words are often regarded as the most salient in this framework, alternative measurers like simple counts or frequencies tend to overvalue extremely rare and extremely common n-grams.

News Importance:

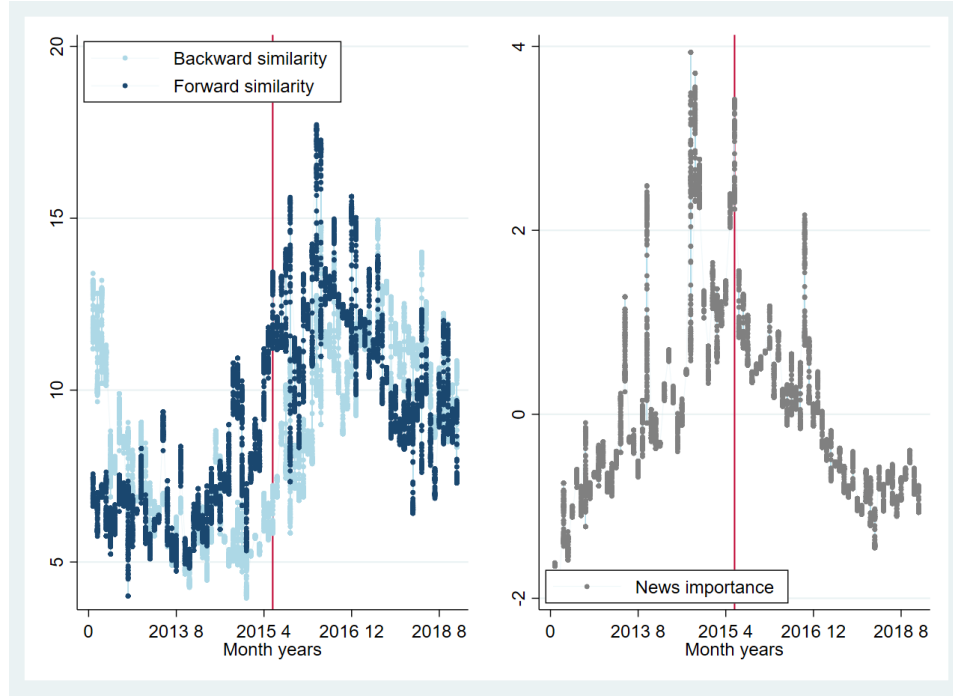
An index of news importance is computed from this information, with an approach borrowed from the patents’ literature (Kelly et al., 2021). This metric combines two indices: a *novelty* and an *impactfulness* indicator. While the former is based on the similarity of article i with past articles (higher *backward similarity* indicates lower novelty), the latter

¹⁸Text 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.

¹⁹These constitute a simple and standard 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.

relates to similarity of article i with future articles (higher *forward similarity* represents higher impactfulness). This measure is employed to explore how new content appeared geographically with the advent of the French border policy. The assumption made to adopt this measurement, which was validated for patents in Kelly et al. (2021), is that the text-based importance of patents is comparable to the text-based importance of news data. Figure 4 suggests this is the case: plotting the components of the news-importance variable—as well as the composite index over time, this measure accurately captures the shock of June 2015, appearing as a peak in the news importance time series. A detailed explanation of this procedure is presented in Appendix section A.5.2.

Figure 4. News novelty, impactfulness, and importance over time



Notes: Data visualization for backward similarity (the inverse of news novelty) and forward similarity (corresponding to impactfulness) in light and dark blue, respectively, in the graph on the left. In grey on the right side is news importance over time. The red line represents June 2015, the month of the border closure by the French authorities.

Framing:

A measure of media bias toward right-leaning,²⁰ anti-immigrant speech is used to evaluate the framing dimension in the local news language. This statistic is referred to as *media slant* in what follows.

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 uses text that has already been classified and labeled as 1 or 0 to create the

²⁰The rationale for focusing on right-wing partisanship 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, but more generally in the right-wing coalition).

training sample. 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). The data derive from a balanced collection of regional and national news articles that either clearly express or do not clearly demonstrate an anti-immigrant orientation. This information derives from both survey and anecdotal evidence, and details are given in Appendix section A.5.3. News items classified as 1 in the training sample belong to newspapers *La Verità* and *Il Giornale*, while text labeled 0 is extracted from *L’Unità News* and *Il Secolo XIX*. Articles from these sources are extracted in the same fashion as for the local news corpus.

Two alternative models are used to classify local news in the binary variable: 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). The two models are compared by cross-validating the training sample and comparing the predictions of the models with human predictions. The human predictions are obtained with *Amazon Mechanical Turk*. 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.

In the Appendix, Figure A.1.7 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 A.5.3 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.

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

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

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

5 Empirical Strategy

This study employs the baseline model:

$$Y_{mt} = \alpha_m + \delta_t + \beta_1 dist_m * Post_t + \Gamma X_{mt} + \epsilon_{mt} \quad (2)$$

where, for municipalities $m \in \{1, \dots, 132\}$ and month-year $t \in \{Jan2012, \dots, Dec2018\}$, 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 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 linearly with distance.²¹

X_{mt} is a vector of controls that vary with time and municipality. 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. All 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. General elections are held at the same date in the whole country. The time distance in months to these elections is therefore interacted with distance from the border, to make sure results are not driven by this alternative set of events. 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

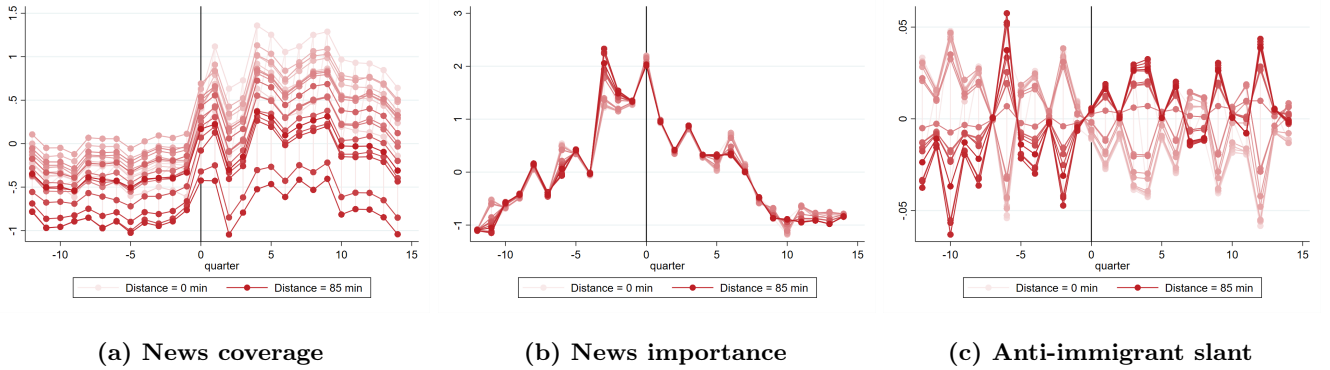
²¹For robustness, a set of extended results allows for different functional forms in the distance-based treatment.

2011 Census (sourced by ISTAT). These are included in the balance analysis as well as in a set of extended results. Table A.2.2 represents the summary statistics for these variables.

Balance-test statistics are presented in Figure A.1.8, to test for pre-existing differences in the covariates along commuting distance. The estimates result from an OLS specification on the pre-policy subsample, accounting for time fixed effects. All variables are standardized to ease comparability. The table indicates that some differential across distance existed in the period before the border closure for some explanatory variables. Higher vote shares for the right-wing coalition are found in the 2008 elections closer to the border. In these areas, the overall local news supply was higher and municipalities slightly smaller in area. Generally, distance to the border correlates positively with shares of older individuals, taxable income per capita, employment and the log of crime events, and negatively to the log of immigrant population stocks. 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.

Below, Figure 5 plots the evolution of Y_{mt} means over time, by commuting distance (shown in number of quarters of the year before and after the French border closure, which is time 0). From left to right, the plotted outcome is respectively news coverage, news importance, and anti-immigrant slant. 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: darker red lines indicate means from areas situated at a greater distance, while lighter tones pertain to means from areas closest to the border. The picture shows a clear switch in the link between distance and news coverage at the moment of the border events. A similar switch is observable to some extent for anti-immigrant slant, while importance levels are broadly more uniform.²²

Figure 5. Mean outcomes over time, by treatment



Notes: In this figure, means over time are plotted for the outcome variables Y_{mt} . In the

²²In panel c of the figure, slant means over time are plotted after partialling out monthly variation in the log count of total news articles. This is done to discard the fluctuation in the proportions of slanted news that is simply due to very low (or very high) quantities of news in one particular area.

5.1 Results

5.1.1 News Coverage and News Importance

To what extent did the border closure translate into a mediatic event? To answer this question, Table 2 presents the estimation results on news coverage on the top-that is, the log count measure of migrant-specific news items-and importance on the bottom-that is, the text-based measure of news similarity with previous versus future content.

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. All table columns also control for the time distance from the administrative and general elections. Population controls involve the log of population, population density (population per squared kilometer) and the share of population over sixty-five years of age, while migration controls are the log of immigration stocks and the logs of population inflows and outflows. Income per capita and crime are added in column 4. Finally, column 5 allows for the inclusion of a count of local outlets present in the municipality m at month-year t . This covariate is of particular interest for the coverage measure, as it allows exploring how much extent news coverage results are driven with vs across sources.

In all regressions, municipality fixed effects and month-year fixed effects account for the unobserved heterogeneity that is fixed over time, or municipality. At the top of the panel, the interaction coefficients of interest are negative and highly significant, pointing out increased news coverage closer to the border. The size of the coefficient stays stable across the different columns. An increase in commuting distance by 10 minutes corresponds to 5-6% lower coverage of migration in the aftermath of the border events.

Table 2. Baseline results: News coverage and importance

Dependent variable: Migrant news coverage (log)					
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.003** (0.001)
Adjusted R ²	0.94	0.94	0.94	0.94	0.97
Dependent variable: migrant news importance					
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	-0.002 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003** (0.001)
Observations	11088	11088	11088	11088	11088
Adjusted R ²	0.95	0.95	0.95	0.95	0.97
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
Migration controls			Yes	Yes	Yes
Income and crime				Yes	Yes
Tot. outlets					Yes

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 content importance index 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.

To dig into the effect's persistency, the left graph of Figure A.1.9a displays coefficients of an event-study specification, as in equation 3, where the effect is allowed to vary by quarter:

$$Y_{mt} = \alpha_m + \delta_t + \sum_{\tau=Q1,2012}^{Q4,2018} \beta_{\tau} I[\tau = t] * dist_m + \Gamma X_{mt} + \epsilon_{mt} \quad (3)$$

where m , t and ΓX_{mt} are defined as in equation 2 and the set of controls is the same as column 4 (i.e., the richest set, excluding the outlet counts, to avoid controlling for a media-related outcome), and the quarter before the events (Q1 2015) is the reference period. In the figure, a persistent higher coverage the closer to the border is identified. Despite some fluctuation, the size of the effects is relatively stable over time, with values similar to the baseline results. There's some significance in the pre-policy coefficients in this graph. As this could be a concern for the validity of the parallel trends assumption, an alternative event-study specification is presented on the right of subfigure a, where individual seasonal linear trends are added. With this addition, the parallel trends assumption appears maintained and a negative, persistent effect is still present.

News importance outcomes are presented on the bottom panel of Table 2. Similar to the results for coverage, news importance appears to correlate positively with the proximity to the events. However, the negative coefficient

on the interaction term is only significant at the 5% level in the specification of column 5, when controlling for the number of local outlets. Moreover, in terms of size, the effect is less relevant. Given that the dependent variable is standardized, an additional 10 minutes in commuting distance decreases the importance of news by at most 0.03 standard deviations. Figure A.1.9b presents the event-study specification for this outcome (with a control set equivalent to column 4) and does not show a significant impact. The figure displays no real persistence of these findings and further shows a significantly negative individual coefficient for one of the periods preceding the events. Inference on this outcome is then less robust to parallel trend violations and shall be taken with a grain of salt.

5.1.2 News Slant

Baseline results on anti-immigrant slant are presented in Table 3. The table structure follows closely the specifications of Table 2: controls are gradually added from the left to the right and all regressions include municipality and month-year fixed effects. 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 3 indicate that, as a result of the June 2015 events, the slant in local media got relatively higher in less-exposed, further areas. An increase in commuting distance by 10 minutes resulted in a significant rise of around 0.18 standard deviations in the proportion of slanted news. Table A.2.3 presents some robustness checks.

In the first division of Table A.2.3, the alternative classification method by Friedman et al. (2010) is employed.²³ 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 panels 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 table bottom, an alternative specification directly considers the aggregation of probability scores, rather than that of labels, as the outcome variable. Results are presented for both the benchmark index and for the index based on Friedman et al. (2010), leading to similar results.

Finally, an event-study specification is presented in Figure A.1.9c. On the left of panel c, the specification shows a significant policy pre-trend occurring 6 periods before the border events take place. Therefore, on the right the specification allows for individual time trends at quarter level, which appear to successfully capture this pre-trend effect. In this plot, all pre-policy coefficients are not significantly different from zero and the post-policy coefficients tend to have positive point estimates. While the coefficients on the right hand side of June 2015 display some fluctuation in their significance, the effect appears quite persistent, and only fading away in the latest months. Further sensitivity analyses for these results are examined in Section 6.

²³See also Friedman et al. (2009) detailing the R package employed.

Table 3. Baseline results: Anti-immigrant slant

	Dependent variable: Anti-immigrant slant				
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.019*** (0.004)	0.019*** (0.004)	0.019*** (0.004)	0.018*** (0.004)	0.018*** (0.004)
Observations	11088	11088	11088	11088	11088
Adjusted R ²	0.79	0.79	0.79	0.79	0.79
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
Migration controls			Yes	Yes	Yes
Income and crime				Yes	Yes
Tot. outlets					Yes

Notes: Baseline results for news anti-immigrant slant: results pertain to a reduced form diff-in-diff model outlined in Equation 2. The dependent variable is an index of right-wing slant for news. 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.

6 Mechanisms, Implications and Robustness

6.1 Mechanisms and Implications

Baseline results suggest the events of June 2015 increased the coverage of migration-related news in the most exposed areas, while the anti-immigrant slant increased relatively more in the municipalities further away. How sizeable are these effects? Focusing on the U.S.-Mexico border in California, Branton and Dunaway (2009) explore the correlation between Californian news coverage of immigration and distance to the borders, for the period 2004-2005. The authors find that an extra ten miles of distance to the border relates to a 2.6% lower coverage of Latino immigrant news. Comparing with this correlation, in the French-Italian setting a one-mile increase in distance results in roughly 8.7% lower coverage,²⁴ suggesting a strong salience for the border policy. 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 benchmarking, assuming that slant increases maximized profits means that an additional ten minutes commuting distance leads to a gain in news traffic by roughly 3.24%.²⁵

To further explore this set of results, this section attempts to distinguish the role of news demand in the slant results. A first analysis relies on alternative geocoding procedures to map the data at article level to a panel of municipalities. As anticipated in the data description, in the benchmark indicators news items are geocoded into municipalities given

²⁴In the present study, the maximum commuting distance within the first mile from the border is 14.6 mins.

²⁵The effect of a ten minute increase in distance after the policy is 0.18 standard deviations in the preferred specification of the baseline, with all controls added but the endogenous outlet count.

Google searches for each local news outlet. These searches are allowed to vary over time, implying baseline results contain a component of endogenous variation in the demand for local news sources over time. Two different approaches are used to address how results depend on this demand-side component. The estimates related to these approaches are presented in Table 4. In the table, column 1 displays the benchmark estimates for comparison. A first approach involves using only Google searches from before June 2015 to perform the geo-localization step. The resulting estimates for the slant indicator are presented in column 2 of Table 4. The size and significance of the coefficient of interest stays robust to this alternative specification, suggesting that the results are not merely driven by an endogenous change in the weights of search traffic across news sources over time. A second approach involves shutting down the demand component of these results, by proposing an alternative geocoding procedure: news items are assigned to municipalities according to the location of the headquarters of local outlets. Specifically, each outlet is allowed to cover municipalities within twenty kilometers (while not presented here, an alternative fifteen and twenty-five-kilometers range were also tested and produced similar findings). For all municipalities within this distance, articles from each local news source have the same weight. With this alternative construction, the slant effects presented in column 3 of Table 4 drop in size and lose significance. This suggests that the slant differential appears when articles with higher levels of news consumption are given more weight.

All in all, these outcomes suggest that local-news slant is shaped by existing patterns in the consumption of news. They are not exclusively driven by switches in traffic across sources, nor do they derive from a mere supply channel.²⁶

Table 4. Extended results: Supply versus demand

	Dependent variable: Anti-immigrant slant		
	(1) Benchmark	(2) Pre-existing demand	(3) Supply
Border distance (min)			
# Post	0.020*** (0.005)	0.017*** (0.004)	0.004 (0.005)
Observations	5819	5819	5819
Adjusted R ²	0.80	0.82	0.46
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

Notes: Extended results on demand versus supply channels. Specifications are comparable to column 4 of Table 3. Standard errors are two-way clustered. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

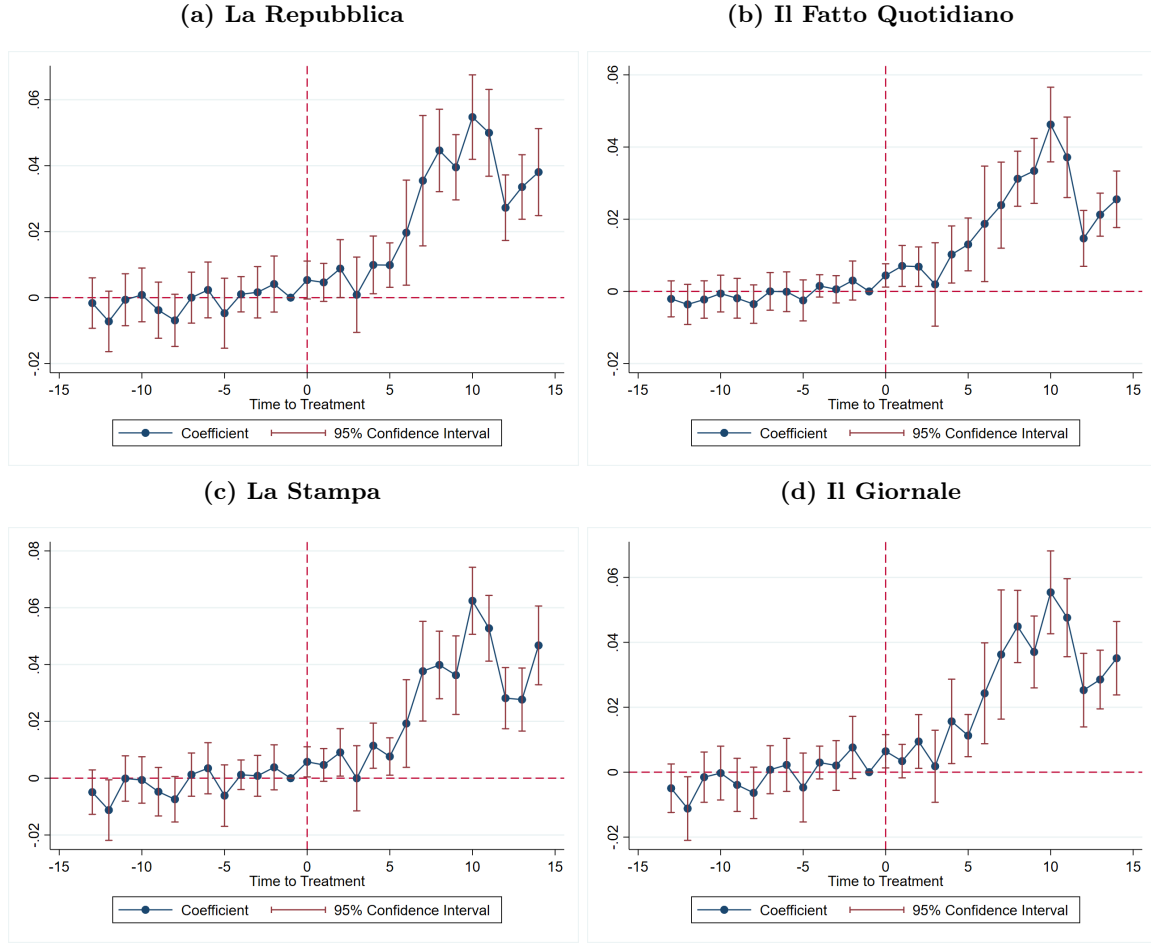
²⁶The number of municipalities covered in size is smaller when taking the supply-only approach. Therefore, N has been equalized in the table to this smaller sample, for better comparability.

George and Waldfogel (2006) suggest that local news outlets' position in the market depends on the presence of national news. In what follows, the effect of slant is investigated in relation to local versus national news penetration in the local market. These additional estimates rely on a new proxy variable, constructed with the use of Google Trends, to capture the relative traffic of a particular local source with respect to each of four main national newspapers: *La Repubblica*, *Il Fatto Quotidiano*, *La Stampa*, and *Il Giornale*.²⁷ For each local news in the sample, Google Trends monthly traffic for each local outlet a is obtained in comparison to the traffic for national newspaper b . Then the ratio a/b is computed. Data are then aggregated at the municipality level, as for the other media outcomes. The resulting measure indicates how many internet searches local news received in proportion to searches for national news. In this sense, it serves as a proxy for local news relative penetration vis à vis national outlets.²⁸ The four measures created represent local news penetration with respect to each of the four main national outlets considered and are plotted over time in Figure A.1.10.

²⁷This precise order follows a left-leaning to right-leaning scale, as from YouGov statistics.

²⁸See Askitas (2015) for a review on the scope of Google Trends data for detecting relevant socio-economic trends with high time-frequency, worldwide.

Figure 6. Migration-related articles, effect on anti-immigrant slant of articles per month.



Notes: Coefficients plot from an event-study 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.

Figure 6 reports results from an event-study specification as done for the media outcomes. Results suggest that, following the border controls, the relative demand differential for local news was increasingly more positive for the more distant areas. Further, this differential appears to increase over time. This finding suggests that changes in local news penetration and anti-immigrant slant took the same direction.

Building on this, Table 5 investigates whether and how results on the media-slant index relate to this penetration variable. In these estimates, the slant labels of the articles are now multiplied by the log of the penetration ratio a/b . This product allows to increase the weight of articles whose sources received higher relative penetration with respect to national news.

Column 1 reports the baseline coefficient for comparison. Columns 2 to 5 show the extended results for each of the national news considered. Both the outcome and the distance measure are standardized and the table bottom includes

the upper and lower bound of the 90% confidence interval for the coefficient of interest, to facilitate comparisons across specifications. The coefficient of interest appears significant and higher in size (from 0.018 in the baseline to 0.026-0.032 in the reweighted estimates). When exploring the confidence intervals, the baseline’s one overlaps with the one of column 2 and 4, while the confidence intervals of columns 3 and 5 are completely above the range of the benchmark. For two out of the four national outlets considered, the effect of the border policy is significantly higher when giving more weights to the most penetrated local outlets.

These results suggest as a possible mechanism that local news adapt their readers targeting in response to the market penetration of national news (as proposed in George and Waldfogel, 2006).

Table 5. Extended results: Relative penetration of local news

	Benchmark	Penetration weights			
	(1)	(2)	(3)	(4)	(5)
	Baseline	La Repubblica	La Stampa	Il Fatto Quotidiano	Il Giornale
Border distance # Post	0.018*** (0.004)	0.029*** (0.003)	0.031*** (0.003)	0.026*** (0.003)	0.032*** (0.003)
Observations	11088	11088	11088	11088	11088
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
Lower bound β_{-1} , 90% CI	0.0115	0.0232	0.0254	0.0214	0.0260
Upper bound β_{-1} , 90% CI	0.0252	0.0339	0.0367	0.0309	0.0376

Notes: Extended results, triple differences. Coefficients relate to the dependent variable *anti-immigrant slant*. The set of controls included is equivalent to Column 4 of the benchmark results. Columns 1 to 5 control for the relative penetration of local news, versus each of the four national outlets considered. Column 6 to 10 add all possible interactions between distance, penetration and the post-events dummy, to obtain a fully saturated model. Stars correspond to the following p-values: * $p < .10$, ** $p < .05$, *** $p < .001$.

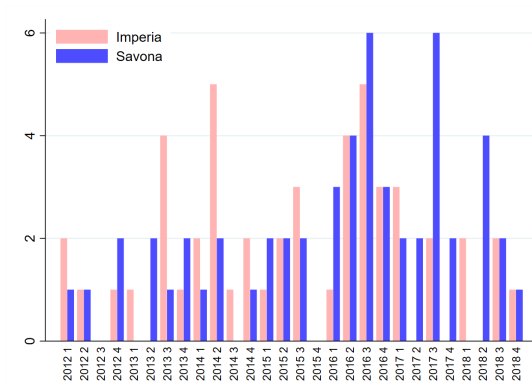
One further rationalization of this finding is that local news at the border are more subject to an audience that is also directly exposed to the events.²⁹ The heterogeneity of readers’ beliefs may impact newspaper slant choice, if, as in the model of Mullainathan and Shleifer (2005), readers seek confirmatory news, that reinforce their beliefs. Local demand for news at the border would then lead, on average, to increase aggregate levels of accuracy. Away from the policy area, ideological readers may gain interest in local news to reinforce their beliefs, thus allowing for local outlets

²⁹Gentzkow 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, possibly trusting more news that validate their views.

to penetrate the market.

Finally, results relate to the framework of Steinmayr (2021), who discusses the opposed implications of contact versus mere exposure of natives to migrants.³⁰ Figure 7 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 incumbency on hate crimes, employing a database collected by NGO Lunaria.³¹ Using the same data, Figure 7 depicts the time variation of hate crimes, aggregated by province. An emerging pattern reflects a growing incidence of hate crimes after border closure for the further province of Savona.³² In what follows, these overall results are compared with voting patterns, investigating the implications for political preferences.

Figure 7. Hate-crimes over time, by province



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

6.2 Voting Patterns

How do the outcomes found for local news match with aspects of the local political economy? To answer this question, this section explores shifts in voting patterns between the periods before and after the border-control policy.³³ Outcomes presented here derive from a first-differences estimation comparing the 2013 with the 2018 general elections' outcomes in the municipalities of interest.

The equation line is the following:

$$\Delta Y_{mc} = \beta_1 dist_m + \Gamma \Delta X_{mc} + \epsilon_{mc} \quad (4)$$

For municipalities $m \in \{1, \dots, M\}$, 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

³⁰This framework is based on the work on social psychology by Allport (1954). Such a channel is not mutually exclusive with the previous rationalizations.

³¹www.cronachediordinariorazzismo.org

³²Note 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.

³³See paragraph A.3 in the Appendix, for a description on the Italian political context.

treatment is defined as for the text-based results, with β_1 being the coefficient of interest. Table 6’s specifications all control for population (in logs), the total number of voters (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). Distance is standardized. Results on the shares of two explicitly anti-immigrant parties *Lega* and *Fratelli d’Italia* (*FDI*) are reported jointly in column 1, and separately in columns 2 and 3. No significant effect is found in this section. Results differ when shares are explored at coalition level, in columns 4 and 5. At a higher distance, the left-wing coalition lost preferences to the advantage of the right-wing coalition (which was also composed of these two parties). The framing of discourse in the local news matches with shifts in coalition voting. An increase of one standard deviation in commuting distance indicates 1 percentage points increase of right-wing preferences, and a 1 percentage points decrease in left-wing coalition votes. Specifically, there’s a relative loss in preferences for the Democratic Party, and a relative success for Berlusconi’s former party PDL. M5S also gained relative consensus. The bottom side of the table adds to these results, by allowing changes in the two main media variables to interact with distance, to capture potential mechanisms. Adding these covariates shows that the higher distance from the border positively relates with right-wing success in the 2018 elections relative to the ones held in 2013. A lower support for the left-wing coalition in areas away from the border is accentuated by greater level of slant, as shown from the negative and significant coefficient for the interaction between slant changes and commuting distance. Slant appears to be a salient mechanism through which the support for the Democratic Party lowered in municipalities further from the border, which favored a switch towards a more populist M5S.

To summarize, this evidence suggests that the border events provoked some changes in the local political economy, in terms of voting attitudes regarding the main coalitions, and the resulting news slant seems to disfavor the center-left.

These patterns share broadly the same direction as media-framing outcomes.

Table 6. Baseline results: Elections outcomes

Outcome: vote shares	Anti-immigration			Coalitions		Turnout	Other vote shares		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Lega, FDI	Lega	FDI	RW coalition	LW coalition	Turnout	PD	M5S	PDL, FI
Border dist	-0.006 (0.006)	-0.003 (0.006)	-0.004** (0.001)	0.010* (0.006)	-0.010** (0.004)	0.006* (0.003)	-0.009** (0.004)	0.019*** (0.006)	0.014** (0.005)
Observations	264	264	264	264	264	264	264	264	264
Outcome: vote shares	Anti-immigration			Coalitions		Turnout	Other vote shares		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Lega, FDI	Lega	FDI	RW coalition	LW coalition	Turnout	PD	M5S	PDL, FI
Border dist	0.001 (0.008)	0.004 (0.008)	-0.004* (0.002)	0.011* (0.006)	-0.036*** (0.005)	0.004 (0.004)	-0.034*** (0.004)	0.040*** (0.006)	0.009* (0.005)
Border dist #Diff slant	0.027 (0.021)	0.026 (0.020)	0.001 (0.006)	0.003 (0.019)	-0.096*** (0.014)	-0.008 (0.013)	-0.097*** (0.013)	0.082*** (0.017)	-0.021 (0.018)
Border dist #Diff coverage	-0.001 (0.009)	-0.002 (0.009)	0.001 (0.003)	-0.006 (0.006)	0.013* (0.008)	0.001 (0.005)	0.006 (0.006)	-0.002 (0.007)	-0.008 (0.007)
Observations	264	264	264	264	264	264	264	264	264
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

Notes: Results for voting outcomes. This table displays coefficients from a reduced form diff-in-diff model. On the bottom panel, anti-immigrant slant is added as a regressor. The dependent variable is the party/coalition vote shares. Election outcomes from this table pertain to Camera and Senato general elections of 2013 and 2018. In all columns, municipality fixed effects and year fixed effects are included. Standard errors are clustered at municipality level. Controls included are electorate population (in logs), population density, over 65 population shares, logs of immigration stocks, income per capita, population inflows and outflows, and crime events at province level. 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.

6.3 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 and a relatively higher increase in anti-immigrant slant in the least directly exposed municipalities. According to the results in the previous section, the framing dimension goes in the same direction as hate-crime rates and voting preferences. Table A.2.4 adds an heterogeneity study to these findings, to explore to which extent the effects differ across initial education levels (based on 2011 shares of University graduates, as sourced by census data), 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 (Census 2011). Column 1 is included for benchmark and is equivalent to column 4 of the baseline tables. The top of the table shows results on news coverage. Evidence suggests that the closest areas to the border receive more news coverage, especially in areas areas with lower altitude (these can be conceived as more urbanized areas, given the region's morphology). Right-wing-slant results

are also heterogeneous along altitude levels. In particular, the positive interaction coefficient is less strong if altitude is higher, possibly indicating that the stronger media effects are in the most urbanized municipalities, which lie on the coast. Regarding the other interactions, where historical support is higher for the left-wing coalition, the slant effect appears less pronounced. Results are relatively homogeneous across the other socio-demographic characteristics considered.

6.4 Robustness to Functional Forms for Distance and Parallel Trends Assumptions

Table A.2.5 shows robustness tests with respect to alternative functional forms for commuting distance. On the top side of the table, commuting distance enters log linearly, via the transformation $\log(x + 1)$, where adding one prevents the omission of Ventimiglia from the analysis. At the bottom, a quadratic term is introduced. The table broadly confirms the results obtained via the linear specification, both in terms of direction and significance. A 10% increase in commuting distance implies a decrease in news coverage by 1.6% and an increase in slant by 0.0478 standard deviations. The estimated decrease in news importance is not significant in this specification. The results on the quadratic effects do not point at significant non-linearities, leading to the linear specification for distance as the preferred form.

Finally, in Figure A.1.11, the linear 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. This approach is motivated by Callaway et al. (2021)'s considerations, for which average marginal effects from difference-in-difference estimations with continuous treatment estimated by two-way fixed effects may require stronger assumptions to hold, as well as producing weighted averages of the effects, with (positive) weights being sensitive to a set of conditions. 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.³⁴ Another robustness analysis involves testing the results to the inclusion of linear time trends that vary geographically. Table A.2.6 presents estimation results when linear time trends by municipality are accounted for. From the top to the bottom, the coefficients presented relate to coverage, importance and slant respectively. Overall, the results are quite comparable to the baseline estimations, with some negligible reduction in size for the coefficients in the coverage estimates; while coefficients for the slant measure effects increase slightly. The unobserved heterogeneity captured in the individual linear time trends signals a possible underestimation of the slant effect in the baseline specifications. The stronger effect found here may suggest, for instance, the presence of a measurement error that attenuates the baseline results. Overall, however, the main findings stay robust in size and significance.

³⁴David 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.

6.5 Other Robustness Tests

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.1.12, and A.1.13 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 A.1.14 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 A.2.7 shows that even when restricting the sample to the original non-imputed data, the direction of the results is maintained. Concerning size and significance, coverage results maintain the same sign, although the size of the effect is now diminished. The significance disappears, given higher standard errors. In this regard, note that the standard errors of every estimation are two-way clustered at both panel unit and at time unit, meaning the results are quite conservative. Reassuringly, slant coefficients are very much in line with the baseline estimates.

Another robustness exercise is displayed in Figure A.1.15. The figure presents possible alternatives to the article count measurement. In the baseline version, all articles are accounted for to explore a measure at the municipality level of news coverage, irrespective of Google Trends weights. This means that too much importance may be given to outlets that count very little for the search traffic in a municipality. To make sure that the agenda-setting results are robust to the presence or absence of very low traffic outlets, regressions in the figure exclude those articles that returned a search-traffic rating lower than $n \in \{5, 10, 15\}$ in Google Trends. To complement this alternative, the figure also shows regressions that exclude articles with very high search-traffic ratings ($n \in \{85, 90, 95, 100\}$). The graph presents coefficients and 90% confidence intervals, from the same diff-in-diff specification (column 4) in the baseline. Reassuringly, the graph suggests that the baseline findings are valid also under this alternative measurement.

Finally, a placebo test confirms that the predictive slant algorithm for articles is not producing a spurious result. A placebo training sample is built such that it is equal to the original training sample, but the predetermined 0 or 1 labels are now randomly shuffled. Then, slant predictions are made on the local news dataset as done for the original measure.³⁵ Table A.2.8 shows insignificant coefficients 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 economy. It provides new evidence that local shocks in migrant presence impact the local news discourse, both in news quantity

³⁵This procedure leads to a vast number of prediction probabilities to be exactly 1/2: these entries are randomly assigned a 0 or a 1.

(*agenda-setting*) and discourse (*framing*), though in opposite directions. The identification of these effects follows from a localized 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 understanding potential aspects of the evolution of the attitudes of citizens (Djourelouva, 2020; Keita et al., 2021) and language analysis can retrieve rich information on the political landscape that may even go beyond the observed electorate behavior (Gentzkow et al., 2019b).

A rich data collection of articles allowed an investigation of two important factors of news production: *agenda-setting*, explored via indices of migration news coverage and news importance; and *framing*, analyzed via a measure of media bias, or slant, towards anti-immigrant discourse. Evidence suggests these border closure events translated into a mediatic shock: the municipalities closer to the event received more migrant-related news. Decreasing commuting distance by ten minutes resulted into 5-6% lower coverage. 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 can rely on higher existing news consumption. Highly demanded local news firms adjust more their content, possibly to accommodate an expected demand for slant which differs along a different exposure 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 some empathetic attitudes where 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 the left-wing coalition lost support. The 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."³⁶ Finally, this paper focused on local

³⁶Avviso 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.

news and employed standard text-analysis *bag-of-words* methods in the prediction of slant. An important area for future research may 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 F and Marco Tabellini (2020) “The Political Effects of Immigration: Culture or Economics?,” *Available at SSRN 3737621*.
- 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.
- Askatas, 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, G (2020) “Rescue on Stage: Border Enforcement and Public Attention in the Mediterranean Sea.”
- 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 Sobbrío (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.”
- 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.
- Djourelouva, Milena (2020) “Media persuasion through slanted language: Evidence from the coverage of immigration,” Economics Working Papers 1720, Department of Economics and Business, Universitat Pompeu Fabra, <https://ideas.repec.org/p/upf/upfgen/1720.html>.
- 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 (2018) “Not Welcome Anymore: The Effect of Electoral Incentives on the Reception of Refugees,” (7212), https://ideas.repec.org/p/ces/ceswps/_7212.html.
- 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.

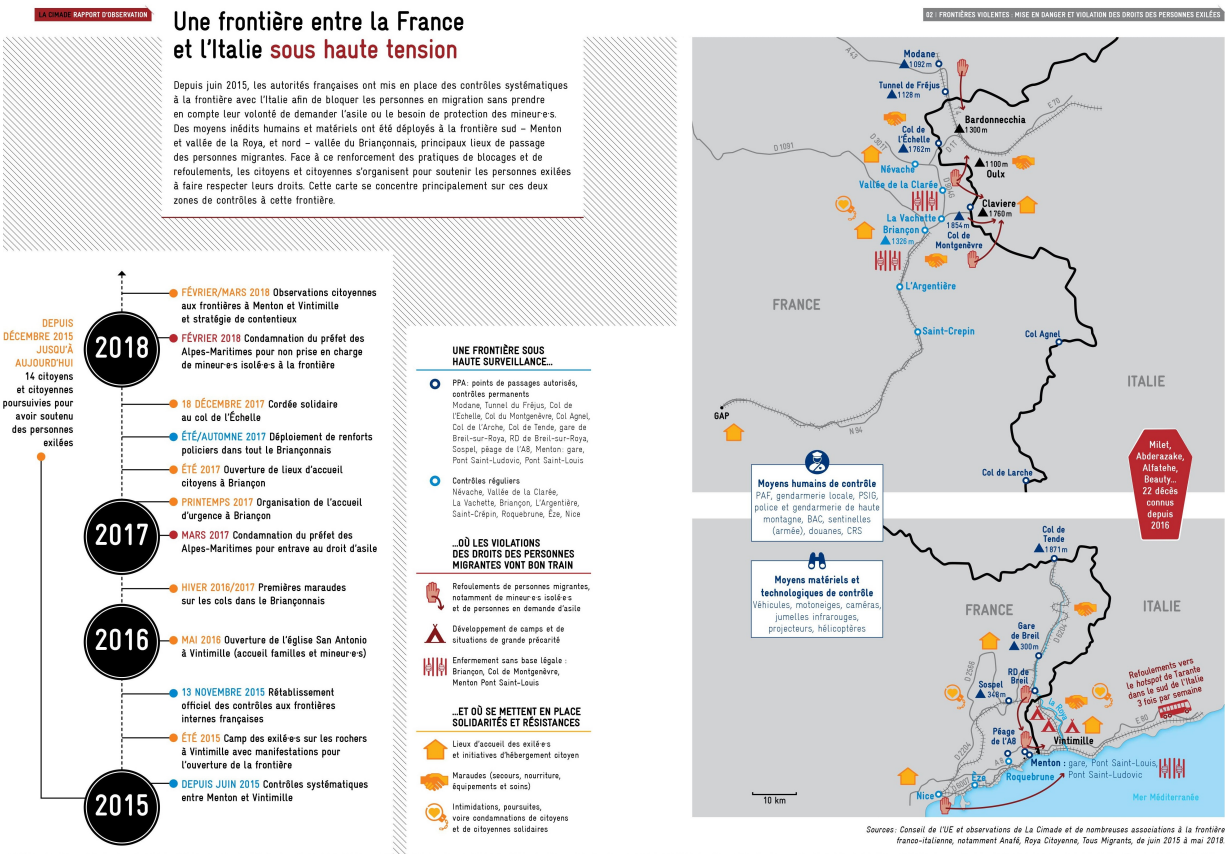
- Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone (2015) “Chapter 14 - Media Bias in the Marketplace: Theory,” in Anderson, Simon P., Joel Waldfogel, and David Strömberg eds. *Handbook of Media Economics*, 1 of Handbook of Media Economics, 623–645: North-Holland, <https://doi.org/10.1016/B978-0-444-63685-0.00014-0>.
- 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.
- 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 (2021) “The Usual Suspects: Offender Origin, Media Reporting and Natives’ Attitudes Towards Immigration,” (21004).
- Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy (2021) “Measuring Technological Innovation over the Long Run,” *American Economic Review: Insights*, 3 (3), 303–20, 10.1257/aeri.20190499.
- Lonsky, Jakub (2021) “Does immigration decrease far-right popularity? Evidence from Finnish municipalities,” *Journal of Population Economics*, 34 (1), 97–139.
- 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.
- 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,” *Evidence From Italy (December 6, 2020)*.
- 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.

A Appendix

A.1 Additional Figures

Figure A.1.1. Dynamics of the border closure



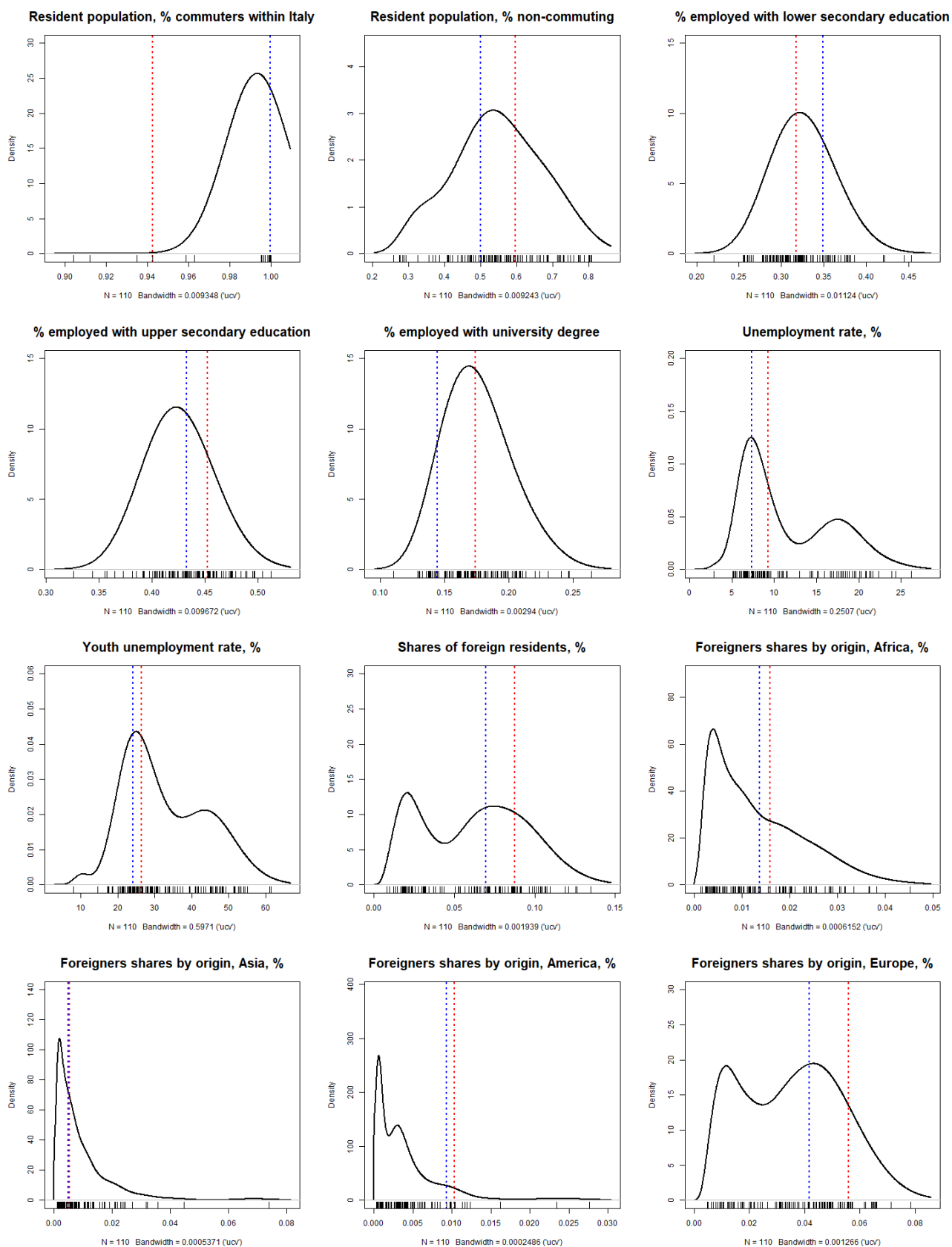
Notes: Figure A.1.1 depicts the border events. At the left, a timeline summarizes the main occurrences; at the right, a map identifies the area of interest. This image is courtesy of La Cimade (<https://www.lacimade.org/>).

Figure A.1.2. Meal counts distributed by Caritas in 2017



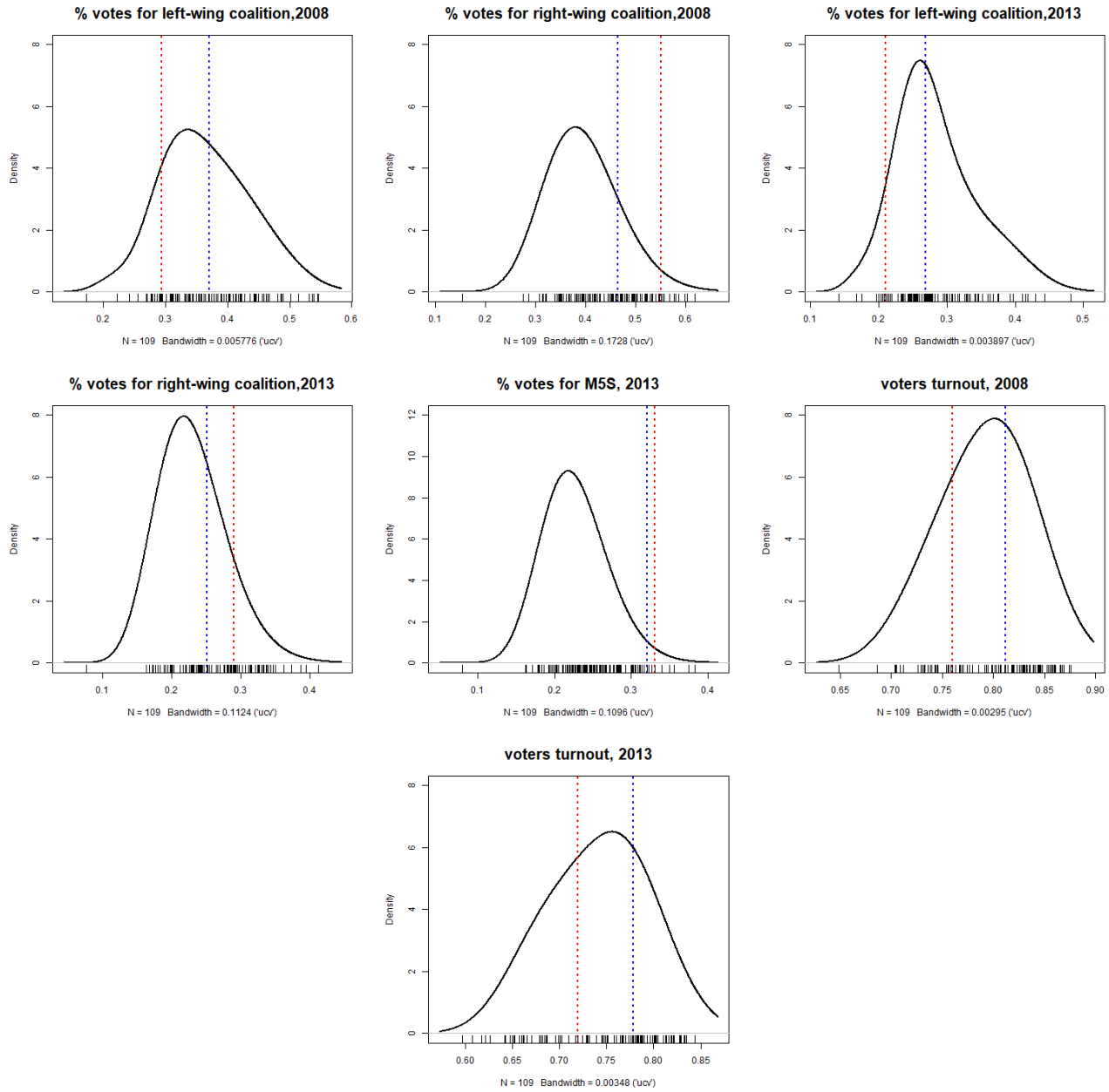
Notes: Stylized statistics on the presence of asylum seekers and migrants that were offered a meal by local NGO Caritas Intemelia. Credits for this graph go to Caritas Intemelia.

Figure A.1.3. 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).

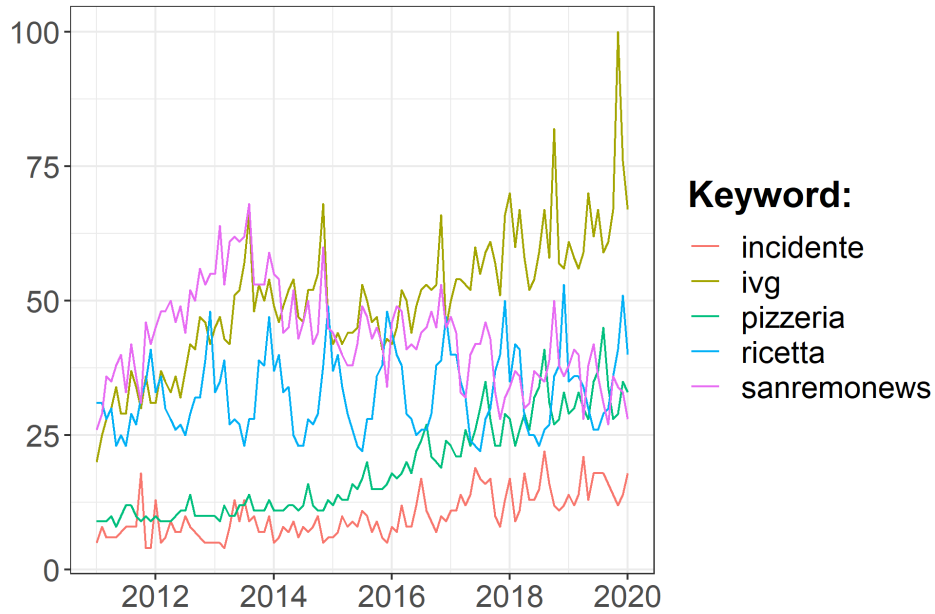
Figure A.1.4. Provinces of interest versus rest of Italy, continued



Notes: Continuation of Figure A.1.3.

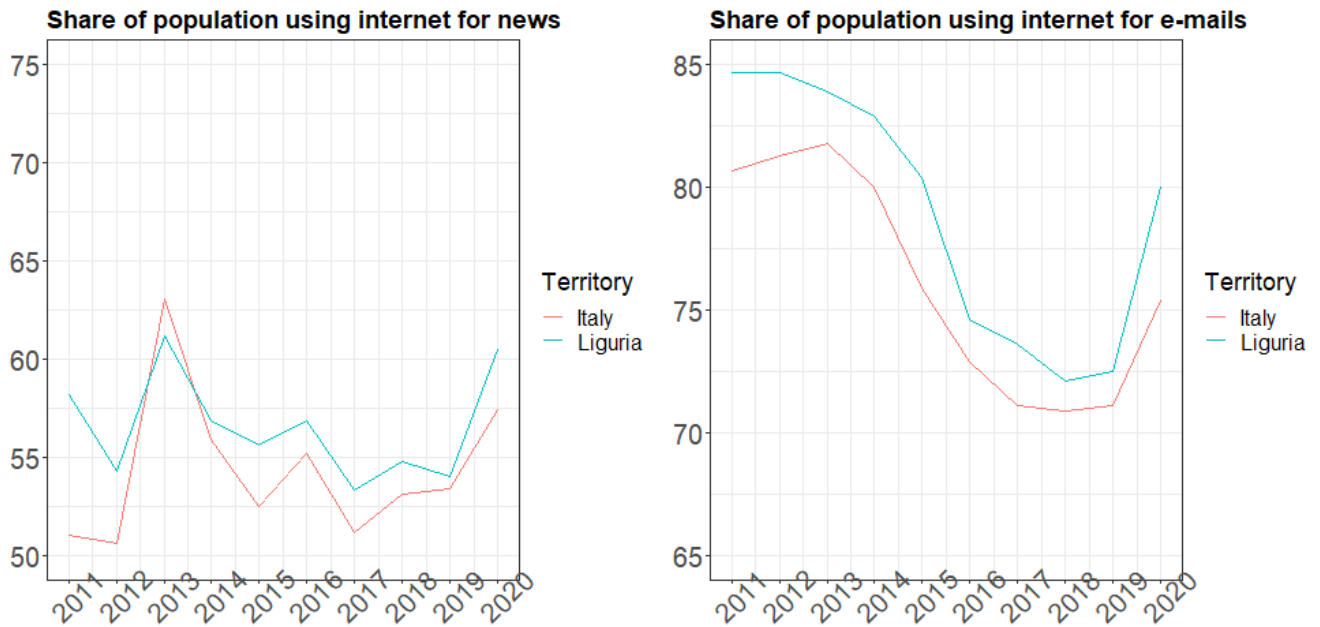
Figure A.1.5. Local news readership

Local news vs other searches



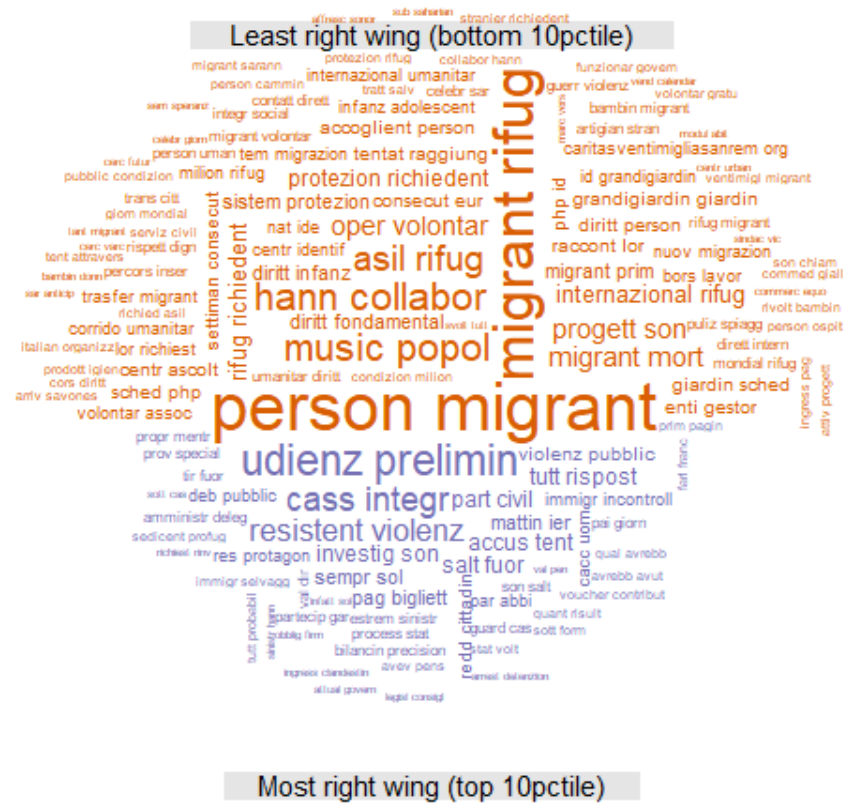
Notes: Comparison of Google searches for two of the main local news outlets, versus other internet searches.

Figure A.1.6. Internet usage for news consumption



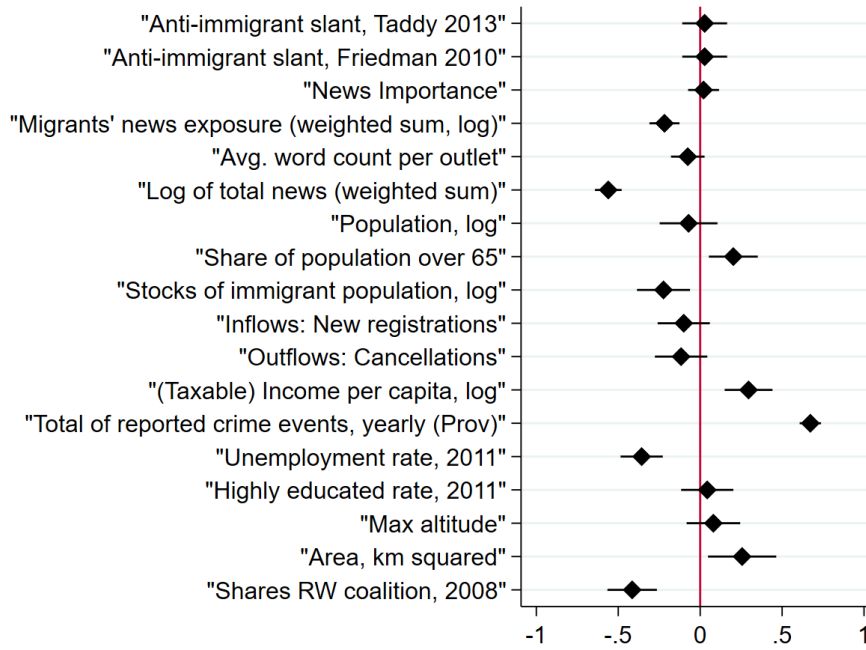
Notes: Share of population that uses internet, sourced by ISTAT. On the left, the share of overall population that uses internet for news. On the right, the proportion of population that uses internet for e-mail.

Figure A.1.7. Comparison cloud. Terms with highest vs. lowest right-wing affinity



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.

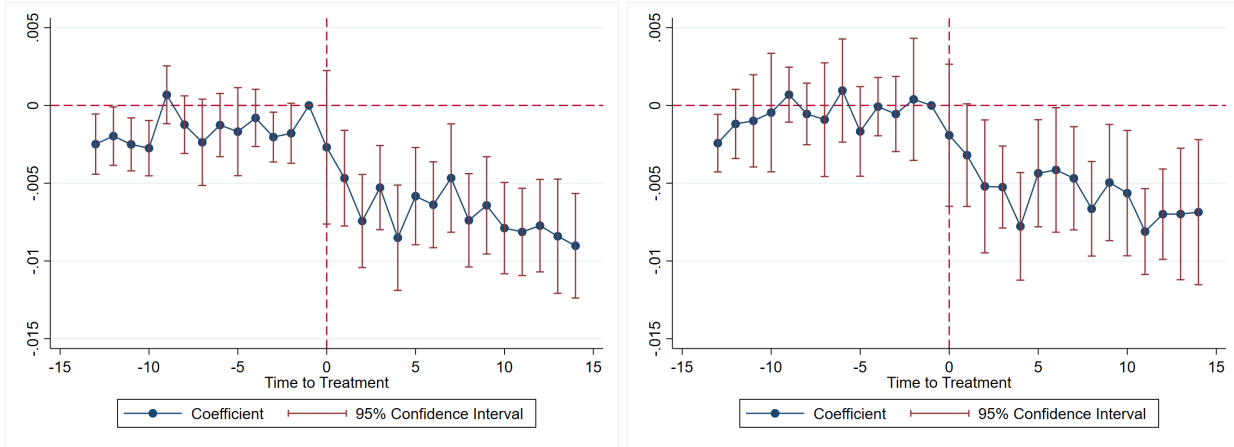
Figure A.1.8. Balance check



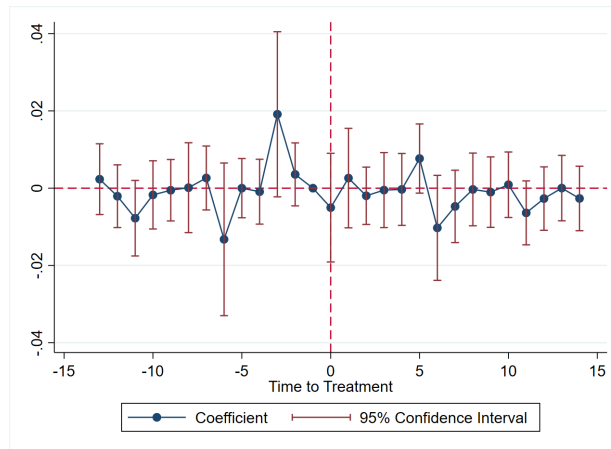
Notes: Balance check: Significance of distance in explaining the pre-June 2015 covariates. Estimates pertain to an OLS estimation of the covariate of interest on commuting distance to the border for the pre-policy subsample, accounting for time fixed effects (except for the census sourced, time-invariant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

Figure A.1.9. Event-study, effect over time.

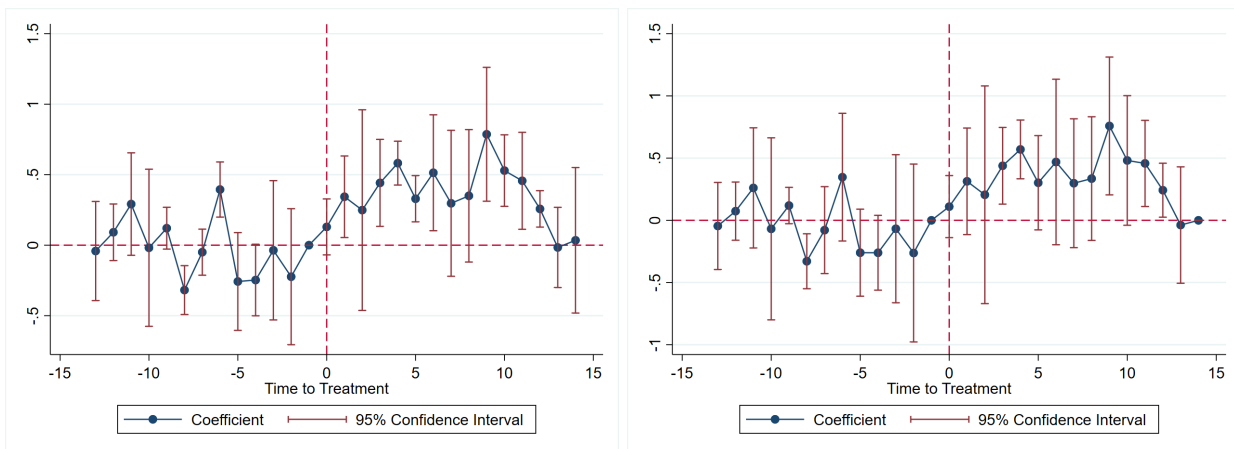
(a) Coverage



(b) Importance

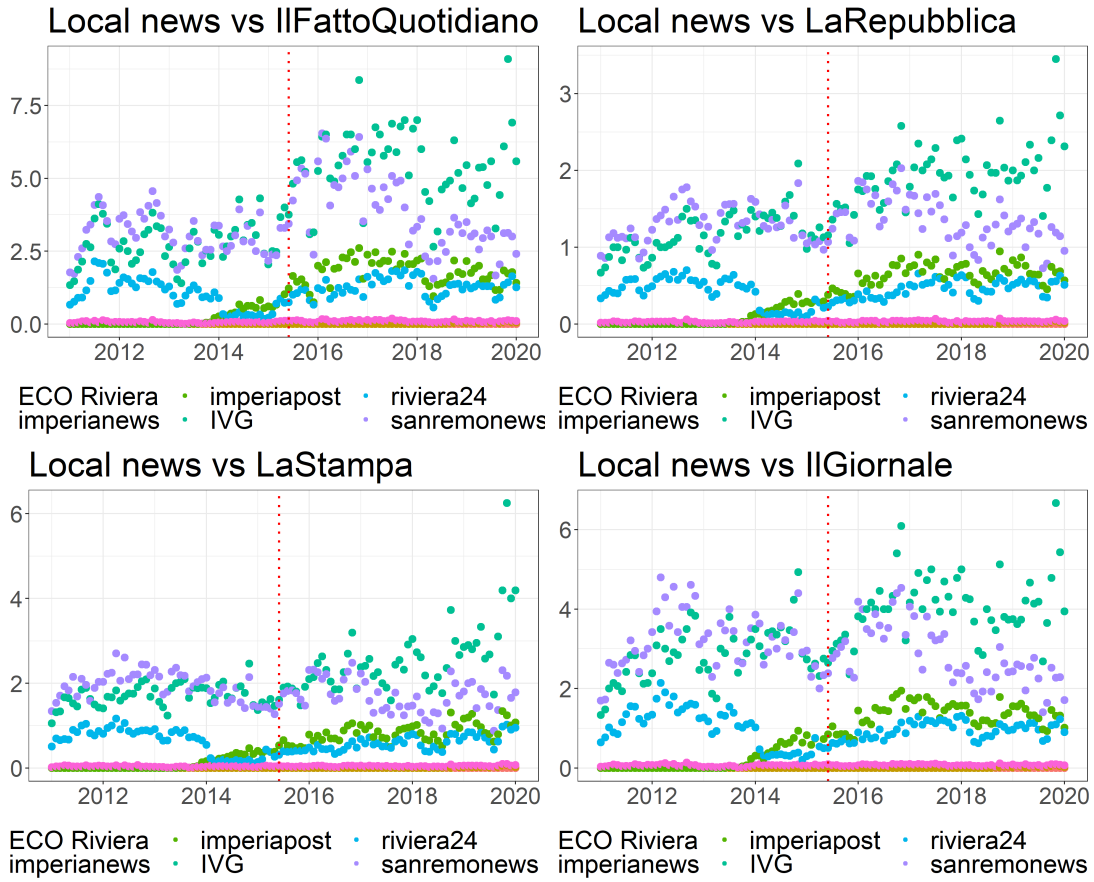


(c) Slant



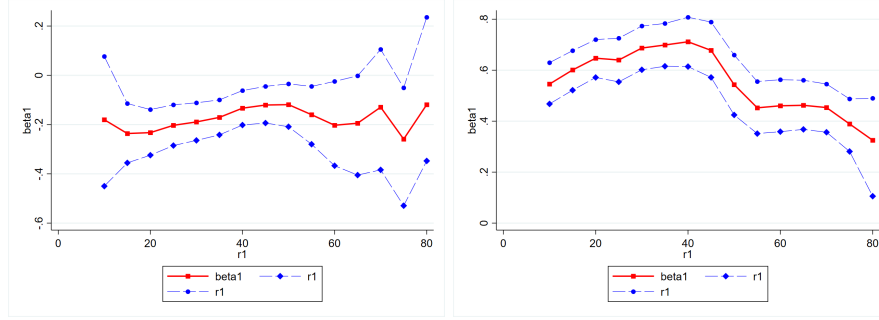
Notes: Coefficients plot from an event-study specification. The outcome of interest is respectively news coverage in subfigure a, news importance in b, and anti-immigrant slant in c. In a and c, due to the presence of some significant pre-policy coefficients on the left, the right-hand side shows alternative specifications that control for individual time trends.

Figure A.1.10. Local news searches vis à vis national news.



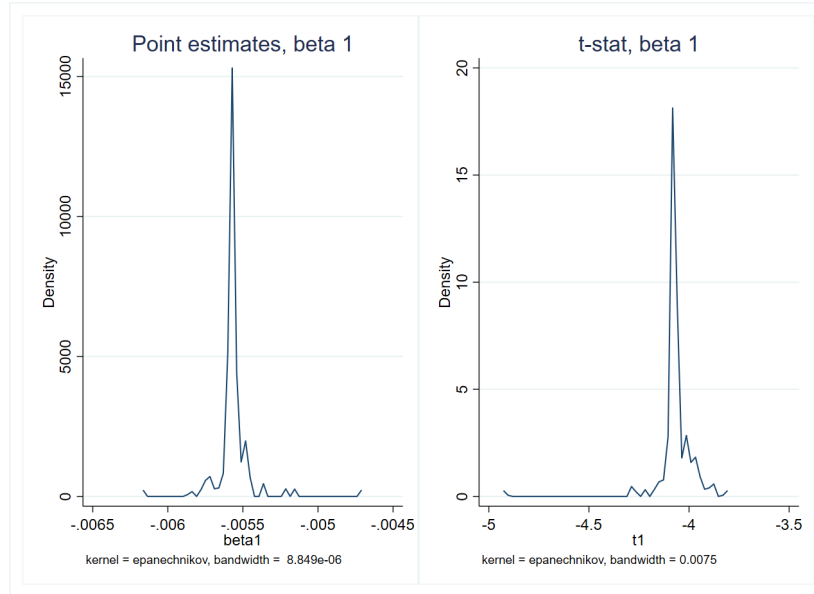
Notes: Proxy for local-news penetration. Values over time. This proxy is calculated as the ratio between internet searches for the national outlet i (\in Il Fatto Quotidiano, La Repubblica, La Stampa, Il Giornale) and local sources s (each appearing as different colors in the graphs), as retrieved via 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.1.11. β_1 robustness with alternative distance thresholds



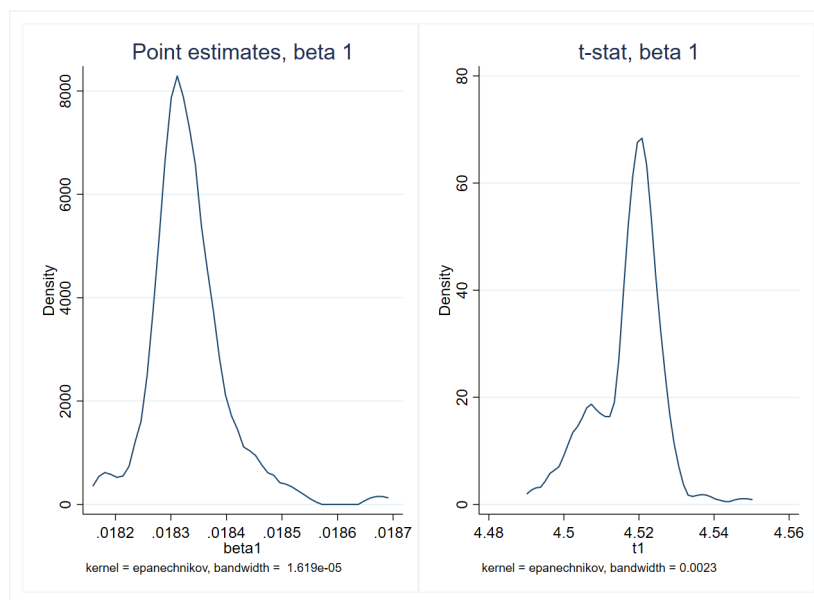
Notes: This figure presents the distribution of estimation results from a alternative specification that considers a binary treatment for distance: on the x-axis the threshold level of commuting distance up to which units are considered to be treated: e.g. for distance= n , treatment is defined as 1 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.

Figure A.1.12. Coefficients robustness on news coverage: Leave-one-municipality-out estimates



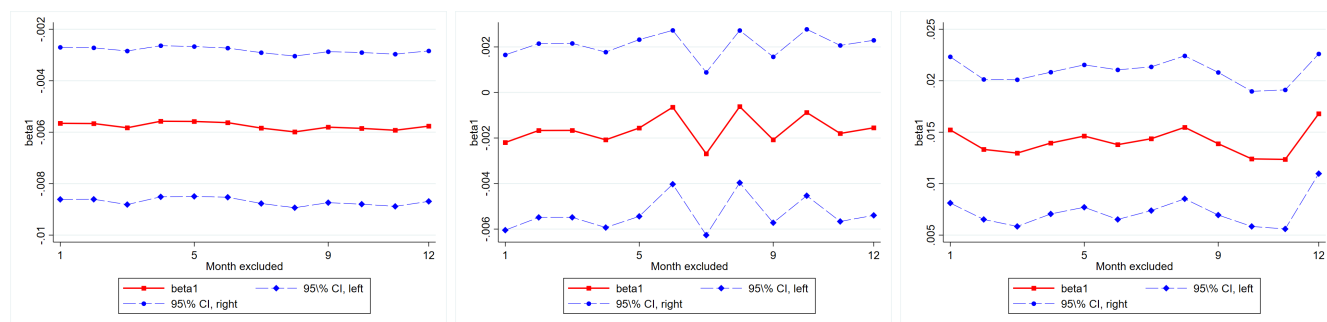
Notes: This figure presents the distribution of estimation results for β_1 , where, at each iteration, one municipality is left out of the sample. The reference is the baseline results on news coverage (Table 2) with the specification including all control variables. In the figure, the left-hand side depicts the distribution of point estimates for β_1 . The t-stat distribution for these estimates is on the right.

Figure A.1.13. Coefficients robustness on anti-immigrant slant: Leave-one-municipality-out estimates



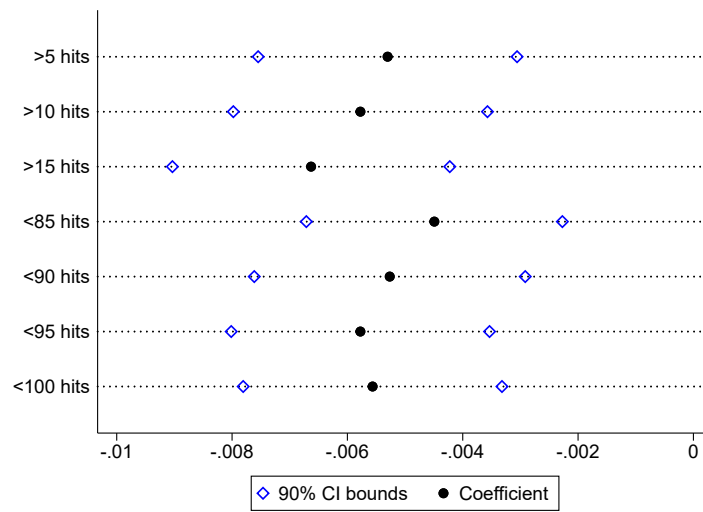
Notes: This figure presents the distribution of estimation results for β_1 , where, at each iteration, one municipality is left out of the sample. The reference are baseline results on anti-immigrant slant (Table 3) with the specification including all control variables. In the figure, the left-hand side depicts the distribution of point estimates for β_1 . The t-stat distribution for these estimates is on the right.

Figure A.1.14. 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. From left to right, graphs pertain respectively to estimations on news coverage, on news importance and on anti-immigrant slant.

Figure A.1.15. Migration-coverage robustness: Dropping very low, very high hits



Notes: This figure presents regression results from alternative counting of migrant-related news. Each row represents a different regression result for the coefficient of interest, comparable to the baseline estimates (Table 2, column 4). The rows exclude articles that returned a rating lower than 5, 10, or 15, or higher than 85, 90, 95, or 100, in Google Trends search traffic. 90% confidence bounds are shown in blue, while the point estimates are displayed in black. Their value is on the horizontal axis.

A.2 Additional Tables

Table A.2.1. Further contributions

Study	Setting	Results	Heterogeneity of the results
Dinas et al. (2019)	Greece (2012-15)	Municipalities/townships w/ asylum seekers \Rightarrow (+) anti-immigrant voting	channels are (-) support for center-right size, (+) turnout.
Campo et al. (2021)	Italy (2013-18)	Municipalities w/ asylum seekers centers \Rightarrow (+) anti-migrant voting	stronger for (-) municipality size, (+) immigrants' stocks, (-) education level.
Bratti et al. (2020)	Italy (2013-18)	(+) proximity to hosting municipalities \Rightarrow (+) support for populist parties.	stronger for (-) municipality size, (-) per capita incomes, (-) prior left-wing orientation.
Edo et al. (2019)	France (1988-2017)	(+) immigration \Rightarrow (+) support for far-right.	Driven by low-educated immigrants from non-Western countries.
Dustmann et al. (2019)	Denmark (1986-98)	(+) refugee shares \Rightarrow (+) vote for right-leaning, anti-immigration parties.	Urban (weakest /opposite effect) vs rural populations (strongest effect).
Halla et al. (2017)	Austria (1970s-2013)	Inflow of immigrants into a community \Rightarrow (+) votes for the FPÖ party.	Channel of "compositional amenities". Communities with (+) immigration flows present lower public resources for natives (public day-care, school proximity).
Barone et al. (2016)	Italy (2001-2008)	(+) immigration \Rightarrow (+) center-right coalition votes.	Effect is heterogeneous by municipality size. Main channels: cultural diversity, labor market competition(education based), public services competition (migrants' fertility), political competition (closeness of votes share of first and second party).
Altındağ and Kaushal (2020)	Turkey(2012-2016)	Syrian-refugee presence produces polarized attitudes towards migrants between Erdoğan's AKP party supporters and opposers.	No impact on AKP voting shares.
Steinmayr (2021)	Upper Austria (2009-2015)	Opposite effects: exposure to asylum seekers \Rightarrow (+) far right support; contact (sustained interaction) \Rightarrow (-) far-right votes	Qualitative evidence: channel is reduced prejudice after contact.
Otto and Steinhardt (2014)	Hamburg city districts (1987-1998)	Immigration \Rightarrow (+) far-right, anti-immigrant voting.	Channels: attitudes are shaped by non-economic determinants and welfare state considerations.
Lonsky (2021)	Finland (2006-2015)	(+) share of foreign citizens in municipalities \Rightarrow (-) the Finns Party's vote share by 3.4 percentage points.	Mechanisms:(+) voter turnout. In line with contact theory, results holds only in municipalities with greatest immigration levels.
Gessler et al. (2021)	Hungary (2014-2018)	Settlements with highest refugee transit \Rightarrow (+) 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 \Rightarrow (-) 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.

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.

Table A.2.2. Summary Statistics

	Mean	Standard Deviation	N	Min	Max
Anti-immigrant slant, Friedman 2010	0.001	1.001	11064	-2.984	3.884
Anti-immigrant slant, Taddy 2013	0.000	1.001	11064	-3.040	3.497
Anti-immigrant slant (predictions)	0.000	1.001	11064	-2.813	2.594
Anti-immigrant slant (predictions), Taddy 2013	0.001	1.001	11064	-3.542	2.960
News Importance	-0.001	1.000	11064	-1.925	3.935
Migrants' news exposure (weighted sum, log)	5.368	0.626	11064	2.833	7.212
Log of total news (weighted sum)	7.607	0.204	11064	7.134	8.312
Avg. word count per outlet	382.300	54.446	11064	239.235	567.256
Population, log	7.159	1.323	11064	4.691	11.032
Population over 65, log	5.903	1.290	11064	3.466	9.789
Population density (population per sq. km)	236.716	343.183	11064	4.759	1922.043
Share of population over 65	0.289	0.051	11064	0.170	0.477
Stocks of immigrant population, log	4.671	1.383	11064	0.000	8.765
Inflows: New registrations	1.713	1.300	11064	0.000	7.650
Outflows: Cancellations	1.674	1.251	11064	0.000	5.832
(Taxable) Income per capita, log	9.952	0.130	11064	9.664	10.723
Commuting distance to Ventimiglia, minutes	45.744	20.072	11064	0.000	83.460
Total of reported crime events, yearly (Prov)	9.470	0.165	11064	9.210	9.740
Unemployment rate, 2011	0.079	0.027	11064	0.013	0.159
Share of university educated, 2011	0.074	0.025	11064	0.025	0.167
Max altitude	955.381	428.416	11064	157.000	2166.000
Area, sq. km	19.981	18.037	11064	1.289	100.658
Total voters, 2008	2369.059	5354.037	11064	67.000	39405.000
Top party is RW, 2008	0.841	0.366	11064	0.000	1.000
Shares LW coalition, 2008	0.322	0.069	11064	0.181	0.544
Top party is RW, 2018	0.425	0.494	11064	0.000	1.000
Shares RW coalition, 2018	0.464	0.068	11064	0.286	0.741

Notes: Summary statistics for the sample of interest.

Table A.2.3. Extended results: Anti-immigrant slant

Dep. variable: Anti-immigrant slant, Friedman et Al. 2010					
	(1)	(2)	(3)	(4)	(5)
	Diff-in-diff, reduced form	Diff-in-diff, reduced form	Diff-in-diff, reduced form	Diff-in-diff, reduced form	Diff-in-diff, reduced form
Border distance (min)					
# Post	0.017*** (0.004)	0.017*** (0.004)	0.017*** (0.004)	0.018*** (0.004)	0.018*** (0.004)
Adjusted R ²	0.77	0.77	0.77	0.77	0.77
Dep. variable: Anti-immigrant slant, threshold 55					
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.018*** (0.004)	0.018*** (0.004)	0.018*** (0.004)	0.018*** (0.004)	0.018*** (0.004)
Adjusted R ²	0.74	0.74	0.74	0.74	0.74
Dep. variable: Anti-immigrant slant, threshold 45					
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.021*** (0.004)	0.021*** (0.004)	0.021*** (0.004)	0.021*** (0.004)	0.021*** (0.004)
Adjusted R ²	0.78	0.78	0.78	0.78	0.78
Dep. variable: Anti-immigrant slant, pred. pr., Taddy (2013)					
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)
Adjusted R ²	0.76	0.76	0.76	0.76	0.76
Dep. variable: Anti-immigrant slant, pred. pr., Friedman et Al. 2010					
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.021*** (0.004)	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)	0.022*** (0.004)
Observations	11088	11088	11088	11088	11088
Adjusted R ²	0.78	0.78	0.78	0.78	0.78
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
Migration controls			Yes	Yes	Yes
Income and crime				Yes	Yes
Tot. outlets					Yes

Notes: Extended results for news anti-immigrant slant : results pertain to a reduced form diff-in-diff model outlined in Equation 2. The dependent variables vary by section and constitute a variation of the baseline index employed in Table 3. 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.

Table A.2.4. Extended results: Heterogeneity

	News coverage (log)				
	(1) Benchmark results	(2) Education, 2011	(3) LW vote shares, 2008	(4) Altitude (max)	(5) Unemployment rates, 2011
Border distance (min)					
# Post	-0.006*** (0.001)	-0.004* (0.002)	-0.004** (0.002)	-0.007*** (0.002)	-0.005*** (0.002)
Triple Interaction		-0.001 (0.001)	-0.000 (0.001)	0.001* (0.001)	-0.000 (0.000)
Adjusted R ²	0.94	0.94	0.94	0.94	0.94
	Anti-immigrant slant				
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.018*** (0.004)	0.018*** (0.004)	0.021*** (0.004)	0.020*** (0.005)	0.018*** (0.004)
Triple Interaction		0.000 (0.000)	-0.001* (0.000)	-0.001* (0.001)	0.001 (0.000)
Observations	11088	11088	11088	11088	11004
Adjusted R ²	0.79	0.79	0.79	0.79	0.79
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

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 Column 4 of the benchmark results. 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.

Table A.2.5. Extended results: Distance functional forms

	Log of distance		
	(1)	(2)	(3)
	News coverage (log)	News importance	Anti-immigrant slant
Border distance (min), log # Post	-0.161*** (0.031)	-0.057 (0.045)	0.478*** (0.132)
Adjusted R ²	0.94	0.95	0.79
	Quadratic distance		
	(1)	(2)	(3)
	News coverage (log)	News importance	Anti-immigrant slant
Border distance, # Post	-0.006 (0.005)	-0.001 (0.002)	0.025*** (0.005)
Border distance (min), squared # Post	0.000 (0.000)	-0.000 (0.000)	-0.000* (0.000)
Observations	11088	11088	11088
Adjusted R ²	0.94	0.95	0.79
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

Notes: Extended results, Different functional forms for the distance-based treatment. Results compare to column 4 of the baseline tables 2 and 3. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

Table A.2.6. Extended results: Time trends at municipality level

	News coverage (log)				
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	-0.005** (0.002)	-0.005** (0.002)	-0.005** (0.002)	-0.005** (0.002)	-0.004** (0.001)
Adjusted R ²	0.98	0.98	0.98	0.98	0.98
	News importance				
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	-0.006 (0.005)	-0.006 (0.005)	-0.006 (0.005)	-0.006 (0.005)	-0.006 (0.005)
Adjusted R ²	0.95	0.95	0.95	0.95	0.95
	Anti-immigrant slant				
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.020** (0.008)	0.020** (0.008)	0.020** (0.008)	0.020** (0.008)	0.019** (0.008)
Observations	11088	11088	11088	11088	11088
Adjusted R ²	0.80	0.80	0.80	0.80	0.80
Municipality FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
Municipality time trends	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
Migration controls			Yes	Yes	Yes
Income and crime				Yes	Yes
Tot. outlets					Yes

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.

Table A.2.7. Robustness to the exclusion of the imputed sample

Dep. variable: News coverage (log)					
	(1) Diff-in-diff, reduced form	(2) Diff-in-diff, reduced form	(3) Diff-in-diff, reduced form	(4) Diff-in-diff, reduced form	(5) Diff-in-diff, reduced form
Border distance (min) # Post	-0.002 (0.003)	-0.001 (0.003)	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.002)
Adjusted R ²	0.95	0.95	0.95	0.96	0.96
Dep. variable: News importance					
	(1)	(2)	(3)	(4)	(5)
Border distance (min) # Post	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)
Adjusted R ²	0.94	0.94	0.94	0.94	0.94
Dep. variable: Anti-immigrant slant					
	(1)	(2)	(3)	(4)	(5)
Border distance (min) # Post	0.020*** (0.004)	0.020*** (0.004)	0.020*** (0.004)	0.019*** (0.005)	0.019*** (0.005)
Observations	2987	2987	2987	2987	2987
Adjusted R ²	0.78	0.78	0.78	0.79	0.79
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
Migration controls			Yes	Yes	Yes
Income and crime				Yes	Yes
Tot. outlets					Yes

Notes: Robustness test of the main results against the use of the non-imputed 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.

Table A.2.8. Extended results: Placebo predictions for slant

	Dependent variable: Anti-immigrant slant				
	(1)	(2)	(3)	(4)	(5)
Border distance (min)					
# Post	0.004	0.004	0.004	0.004	0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Observations	11088	11088	11088	11088	11088
Adjusted R ²	0.69	0.69	0.69	0.69	0.69
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
Migration controls			Yes	Yes	Yes
Income and crime				Yes	Yes
Tot. outlets					Yes

Notes: Extended results, baseline estimates are tested to the use of a placebo outcome, created by applying the slant classification algorithm to a training sample where 0 and 1 labels are shuffled randomly. Stars correspond to the following p-values: * p < .10, ** p < .05, *** p < .001.

A.3 The Political Context in Italy

As of 2021, Italy is 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 in the early 90s, caused by a series of scandals that were famously termed *Tangentopoli* (Bribeville). 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 setting 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 (Campante et al., 2018). This political movement is generally associated with a broadly populist stance (Bratti et al., 2020; Campo et al.,

2021). Within 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, 2018; Romarri, 2020): Lega and Fratelli d'Italia (FDI).

Lega's party was founded in 1991 as Lega Nord and was originally 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. In the 1992's general elections, the party obtained a vote share of around 8%. More recently, in the 2013 general elections, it joined Berlusconi's coalition, and obtained little above 4% in each of the two chambers. After the 2013 elections, Matteo Salvini assumed leadership of the party and abandoned the Northist stance and maintaining a strong anti-immigrant rhetoric. Lega's Manifesto of 2018 expressly takes a stand against the NGO operators in the Mediterranean Sea and maintains that immigration should be based on cultural proximity. Salvini's party obtained well above 17% of the vote in the 2018 general elections.

Fratelli d'Italia is a nationalist party that emerged in 2012 with a separation from the center-right coalition. It has been led by Giorgia Meloni since 2014, it has a conservativist, nationalist ideology of post-fascist inspiration. It takes a well-defined anti-receptionist position with a strong anti-islamic rhetoric. 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.

A.4 Geocoding News

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 specific to 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. Google Trends serves as a proxy to circumvent this issue, as it provides data 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.³⁷. There are two alternative time-ranges considered: January 2011-January 2015 and June 2015-January 2020. A broad enough time-range allows Google Trends to provide more uncensored information for municipalities with less traffic. In fact, the narrower the time interval, the less detailed data are on the platform, given that it does not

³⁷<https://support.google.com/trends/answer/4365533?hl=en>

display information below a minimum search-traffic rate. For 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 this resulting metrics be called $rank1_{ism\tau}$: the Google Trends' rating of source s over the sum of all s for m , $\tau \in \{2010 - 2015; 2015 - 2020\}$. Figures A.4.2 and A.4.3 show Google Trends search results for the sources with most search traffic, in the time-ranges 2010-2015 and 2015-2020, respectively.

ii) To allow for more granular time-variation, data on source traffic over time *for the whole region* is retrieved, using a benchmark source for reference (the sample source IVG is used as reference, reaching 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 $rank2_{ismt}$, i.e., the Google Trends rating of source s at month-year t .

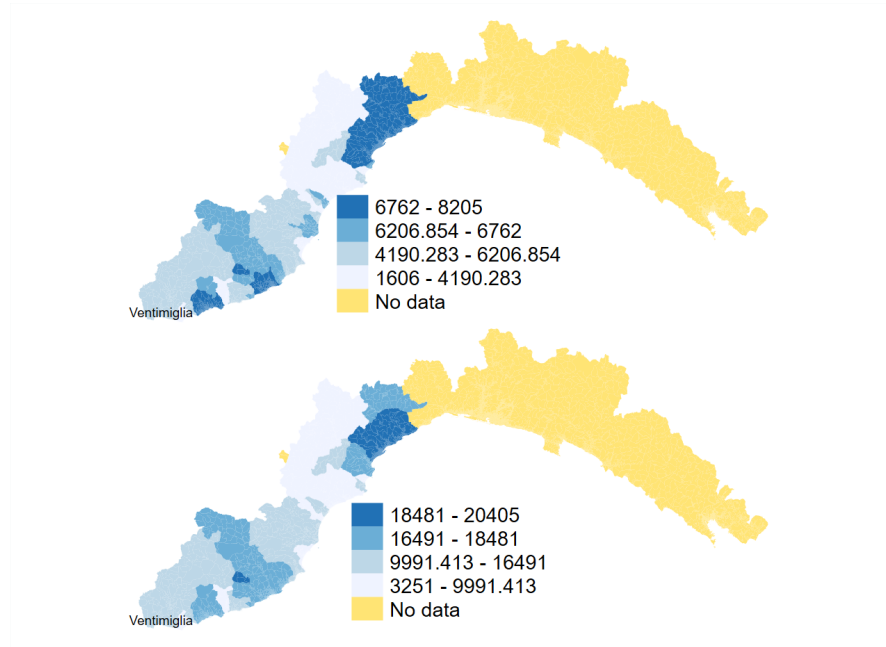
Importantly, news sources with less traffic did not return municipality level information from Google Trends. Their weight is zero in the benchmark analysis.

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 i) its importance returned by the cross-municipality rating and ii) 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 $w_{ismt} = \frac{rank1_{ism\tau} * rank2_{ismt}}{\sum_{i \in m, t} (rank1_{ism\tau} * rank2_{ismt})}$.

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 work or study in municipality m_2 . If the percentage of commuters among the residents is less than 25%, data are not imputed.³⁸ 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 A.4.1. Table A.6.1 lists all 132 municipalities for which news data has been geo-localized in both the periods before and after June 2015.

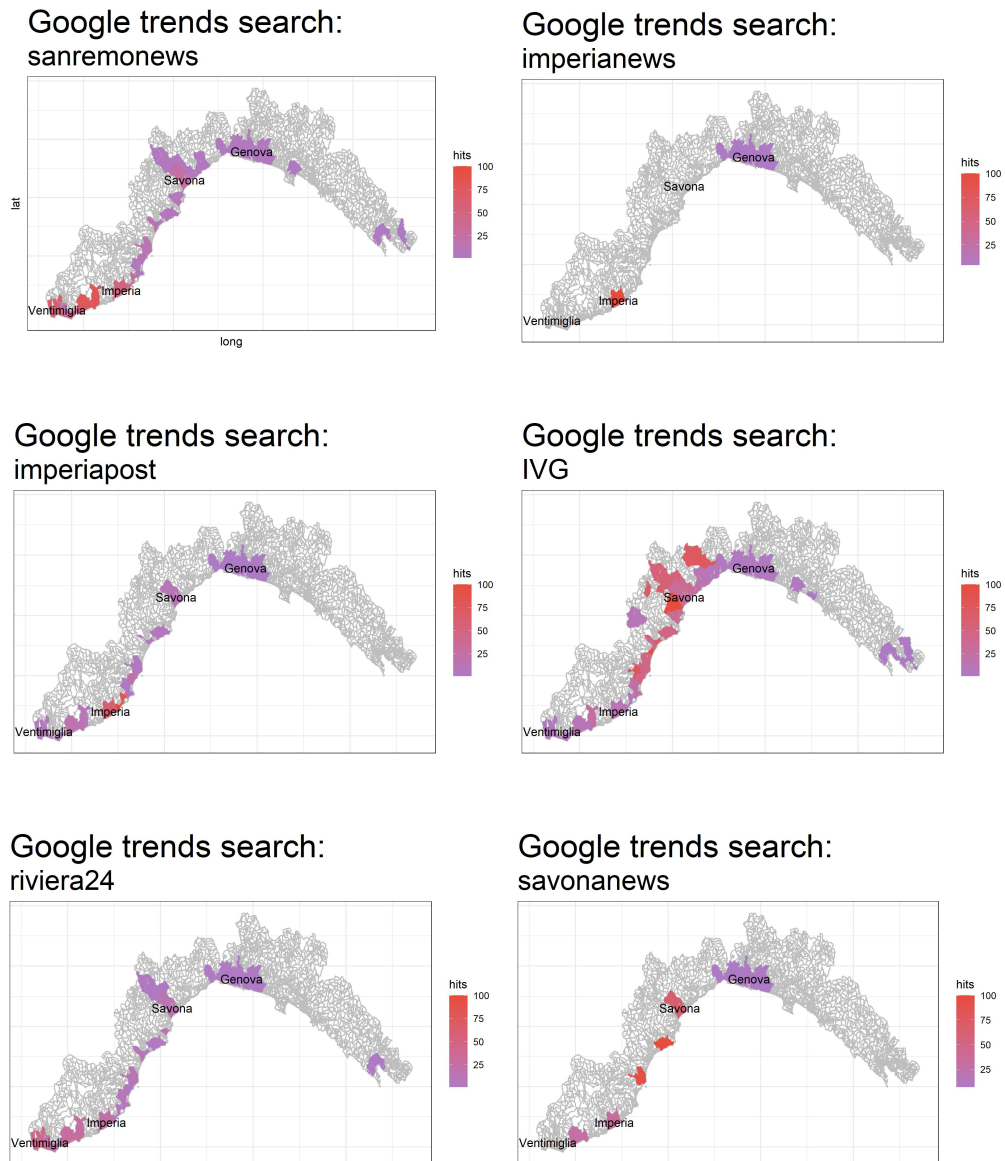
³⁸The rationale is that the higher the level of commuters, the more likely the news at the commuting destination may reach the commuters.

Figure A.4.1. News sample geographic distribution, before and after June 2015



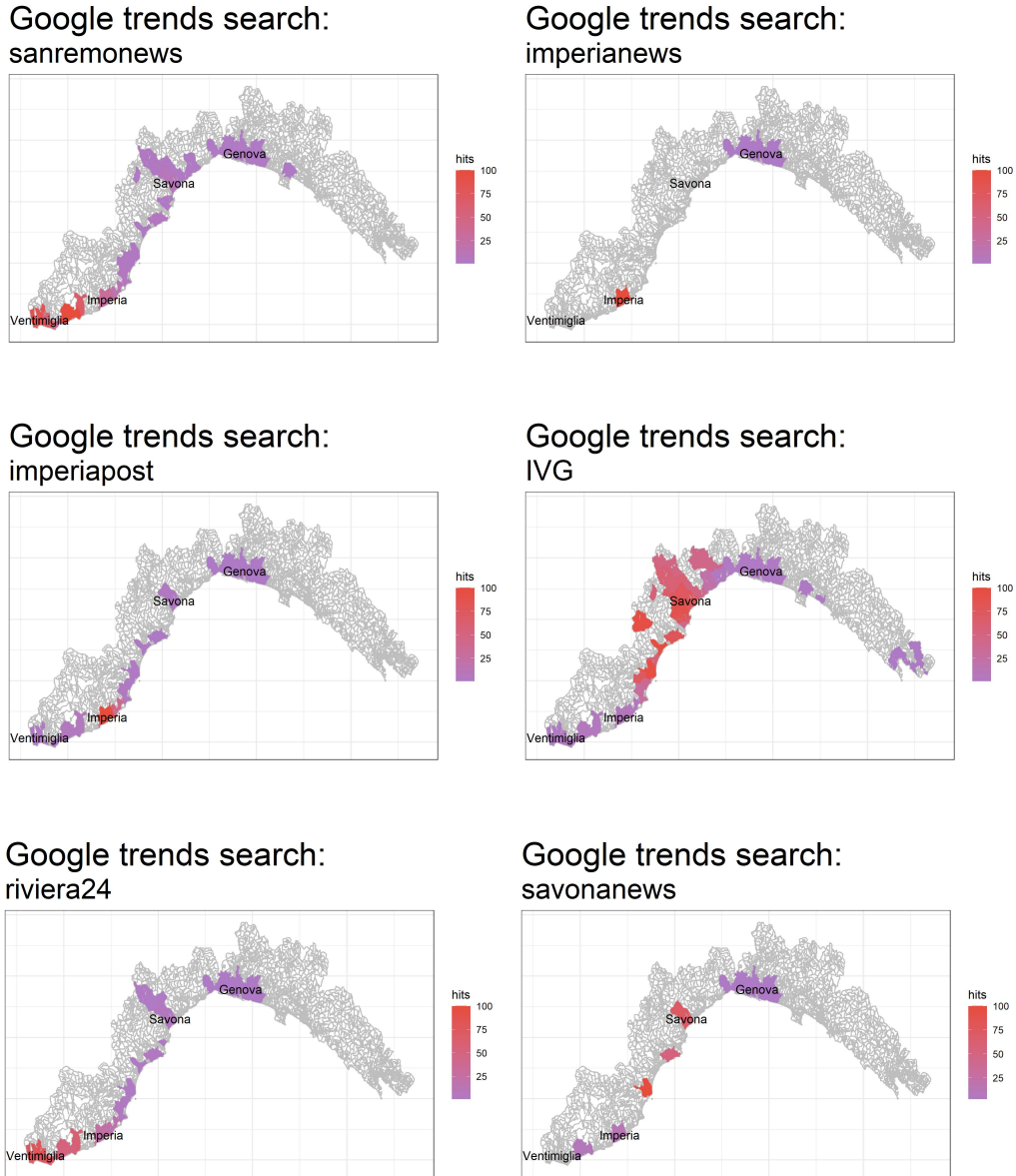
Notes: Data visualization of the news sample. On the top map are data related to the pre-treatment period, while the post treatment period is illustrated on the bottom. News coverage patterns result from the geo-localization process described in the paragraph and pertain to the provinces of Imperia and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

Figure A.4.2. 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 A.4.3. 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.

A.5 Text-Based Measures

A.5.1 News cleaning

To obtain the final sample of migration-related news, 25,315 articles extracted that only matched the keyword “foreigner” were further screened. This step was implemented to avoid the presence of articles that would only be tourism-related and captured unintentionally with the use of this keyword. To clean this part of the corpus, a topic model was constructed via a Latent Dirichlet Allocation (DTA), following the seminal contribution of Blei et al. (2003) and the R-package

topcmodels 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 application³⁹. Let a corpus be formed of $i \in \{1, \dots, N\}$ documents, containing text that belongs to a dictionary of $j \in \{1, \dots, M\}$ 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:

$$\beta \sim \text{Dirichlet}(\delta) \quad (5)$$

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

$$\theta \sim \text{Dirichlet}(\alpha) \quad (6)$$

For every j term,

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

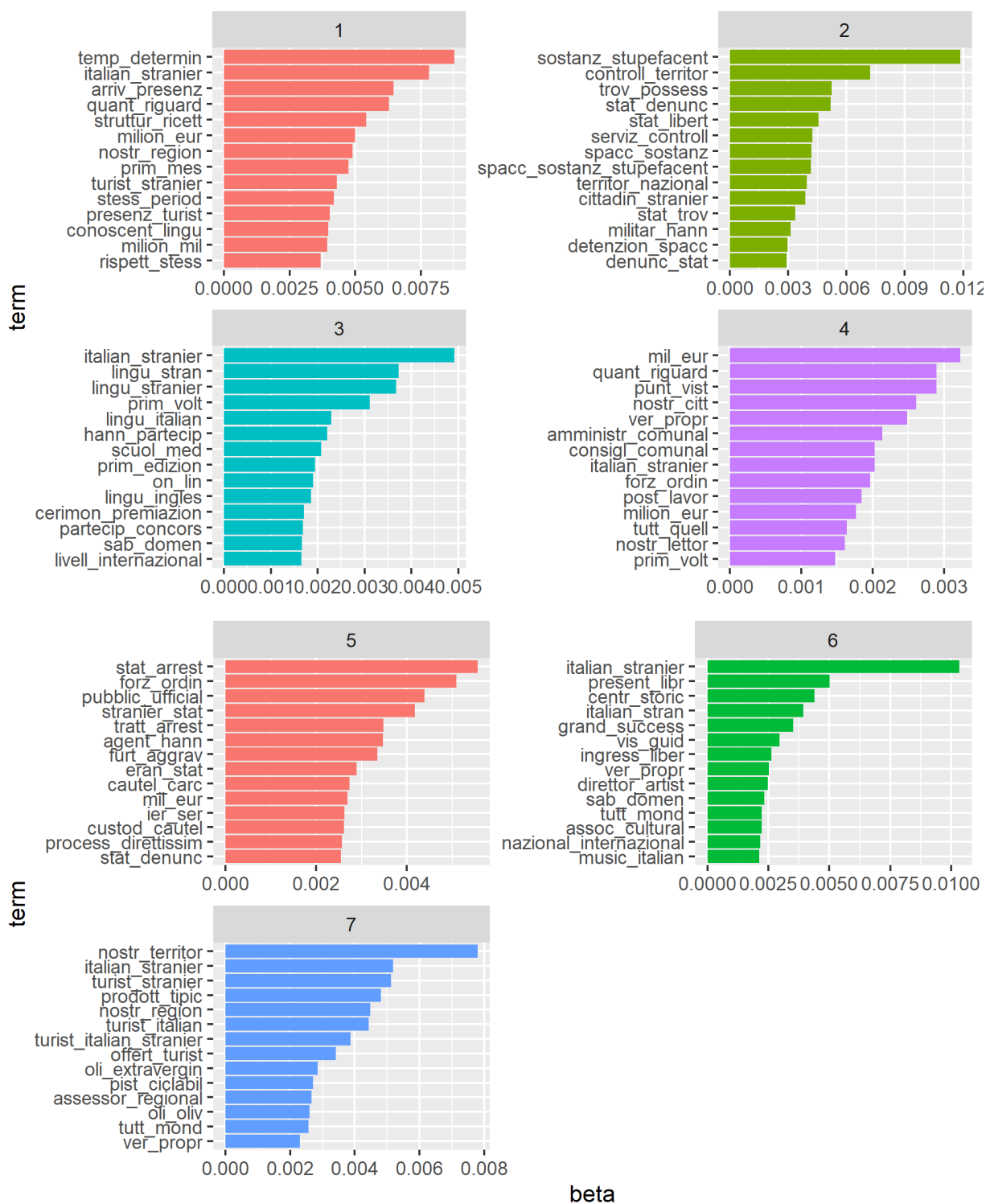
The estimation methodology follows from a variational expectation-maximization (VEM) algorithm, where the alpha parameter is estimated (see Hornik and Grün (2011) for details). To summarize the resulting topic distribution, a plot of the most recurring terms for each topic is presented below in figure A.5.1. 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:	turism 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.

³⁹A higher κ appeared to produce redundant topics

Figure A.5.1. Document-term matrix, sample of frequent terms.



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

A.5.2 News Importance

By drawing a parallel from Kelly et al. (2021)’s index of patent importance, a measure of news importance is constructed as the ratio of two indices: a novelty measure and a measure of impactfulness. Both these indicators are built given text similarities of the articles in the sample, where similarity is a function of term frequencies in the texts. The original measure is applied by Kelly et al. (2021) in the patent literature, as a new indicator of patent quality. Higher quality for patents is described as greater novelty-low similarity with previously published patents-and greater impactfulness-high similarity with future patents. This text-based measure for patents was observed to correlate strongly with market value indicators, which the authors interpret as a sign of high reliability. Closely following this patent-related approach, the same measure is constructed in this study, where, instead of patents, the corpus of text is constituted by news.

In this way, the evolution of news content can be explored over time and across distance levels. The end goal is investigating whether and how novel and impactful content spread as a consequence of the border policy. The only divergence from Kelly et al. (2021) here involves the time frequency, as this setting adopts smaller time spans, to accommodate the higher frequency of news publications, relative to patents. Importantly, the development of this metric involves first building pairwise similarity values between each pair of documents, and second aggregating these values into a document-level index.

In an $N \times M$ matrix of documents $i \in \{1, \dots, N\}$ and terms $j \in \{1, \dots, M\}$, the term-frequency, inverse document-frequency is defined as

$$tf - idf_{ij} = tf_{ij} \times idf_{ij} \quad (7)$$

A higher $tf - idf_{ij}$ level indicates that term j is generally frequent in document i , but is not very much frequent in most other documents, thus suggesting it is an informative term for the document’s content.

In the alternative version presented by Kelly et al. (2021), time dimensionality is introduced in this indicator, and the *backward-idf* is defined as the ratio between log frequency of all documents dated before document i , over the log frequency of all these prior documents that contain term j . Time is considered here at month-year level.

$$Bidf_{ij} = \frac{\log(\text{number of documents prior to } i)}{\log(1 + \text{number of documents prior to } i \text{ containing term } j)} \quad (8)$$

This measure thus represents the history of the term frequency up to the time of document i . Given documents’ pair $(i; i')$, their similarity will be based on the following modified $tf - idf$:

$$tf - Bidf_{ijt} = tf_{ij} \times Bidf_{ijt} \text{ where } t = \min(t_i, t_{i'}) \quad (9)$$

Vector $tf - Bidf_{it}$ will then have size W equal to the set of terms j present in both i and i' . This vector is normalized to have unit length, and proximity between i and i' is measured by cosine similarity. This computation is typically used in unsupervised learning clustering methods such as spherical K-means. Compared to other proximity measures, such as the Euclidean distance, cosine similarity is conveniently independent of document length and corresponds to the dot product between the two normalized vectors. It ranges between 0 and 1, 0 occurring when documents have no overlapping terms and 1 when two documents use the same words in the same proportion.

$$\text{Cosine similarity}_{i,i'} = \rho_{i,i'} = \frac{tf - Bidf_{it}}{\|tf - Bidf_{it}\|} \cdot \frac{tf - Bidf_{i't}}{\|tf - Bidf_{i't}\|} \quad (10)$$

It is important to stress that the sample of interest contains news as old as 2010, and that the time unit considered is the month-year. For each of these metrics, a sliding window is constructed such that for month-year t , articles dating at most one year prior (i.e., $t-12$) and at most one year after (i.e., $t+12$) are employed. With this methodology, measuring similarity at time t requires lagged data availability. For this reason, the first month-year for which this indicator is available for analysis is January 2012, as the previous 12 months will be used as lags in the measure construction. Similarly, data on this index would end in 2018, as 2019 data are used as leads in the construction. Kelly et al. (2021) define *novel* patents as those whose content is most distinct from their predecessors, meaning they rely less on existing contributions. In the context of news, an article featuring novel content will be less similar to previous discourse. *Backward similarity* is defined as:

$$BS_{i',12} = \sum_{i \in \mathbf{B}_{i',12}} \rho_{i,i'} \quad (11)$$

Where $\mathbf{B}_{i',12}$ denotes the set of documents

dating prior to i' , given a maximum of $T=12$ lags considered. *Impactful* news items are those that are most influential, which would be reflected in sharing content with future content. In this sense, *Forward similarity* is defined as:

$$FS_{i',12} = \sum_{i \in \mathbf{F}_{i',12}} \rho_{i,i'} \quad (12)$$

Where $\mathbf{F}_{i',12}$ denotes the set of documents

dating posterior to i' , given a maximum of $T=12$ leads considered. Finally, news importance $q_{i'}$ combines forward and backward similarity to identify novel and impactful news:

$$q_{i'} = \frac{FS_{i',12}}{1 + BS_{i',12}} \quad (13)$$

Which is equivalent to a ratio of *forward similarity* over *backward similarity*⁴⁰. Figure 4 represents the evolution over time of the three computed measures: novelty, impactfulness (on the left) and importance (on the right). The red line displayed in the graph defines the periods before and after June 2015. As an interesting visual pattern, both similarity measures feature higher levels in the post period. A peak in news importance appears during the month of June 2015. This higher level is suggestive of the reliability of this measure in identifying unprecedented shocks in news content.

A.5.3 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)

⁴⁰Adding 1 to avoid having zero at the denominator

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 slightly better 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 Milyo (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 limitation, 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 A.3.

The classification proposed here involves the construction of a simple binary variable, where a value of 1 corresponds to bias toward 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 *Il 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,⁴¹ 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 considered by the researchers. 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 *Il Giornale* shares similar YouGov statistics to *La Verità* and also takes an evidently critical rhetoric against migration.⁴² Data from the *Il Giornale* are 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

⁴¹<https://it.yougov.com/news/2019/05/03/giornata-mondiale-della-liberta-di-stampa/>

⁴²An headline example from February 2022: <https://www.ilgiornale.it/news/politica/carola-leroina-fuori-legge-2007541.html>: “Carola 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.”

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*.⁴³ 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.⁴⁴

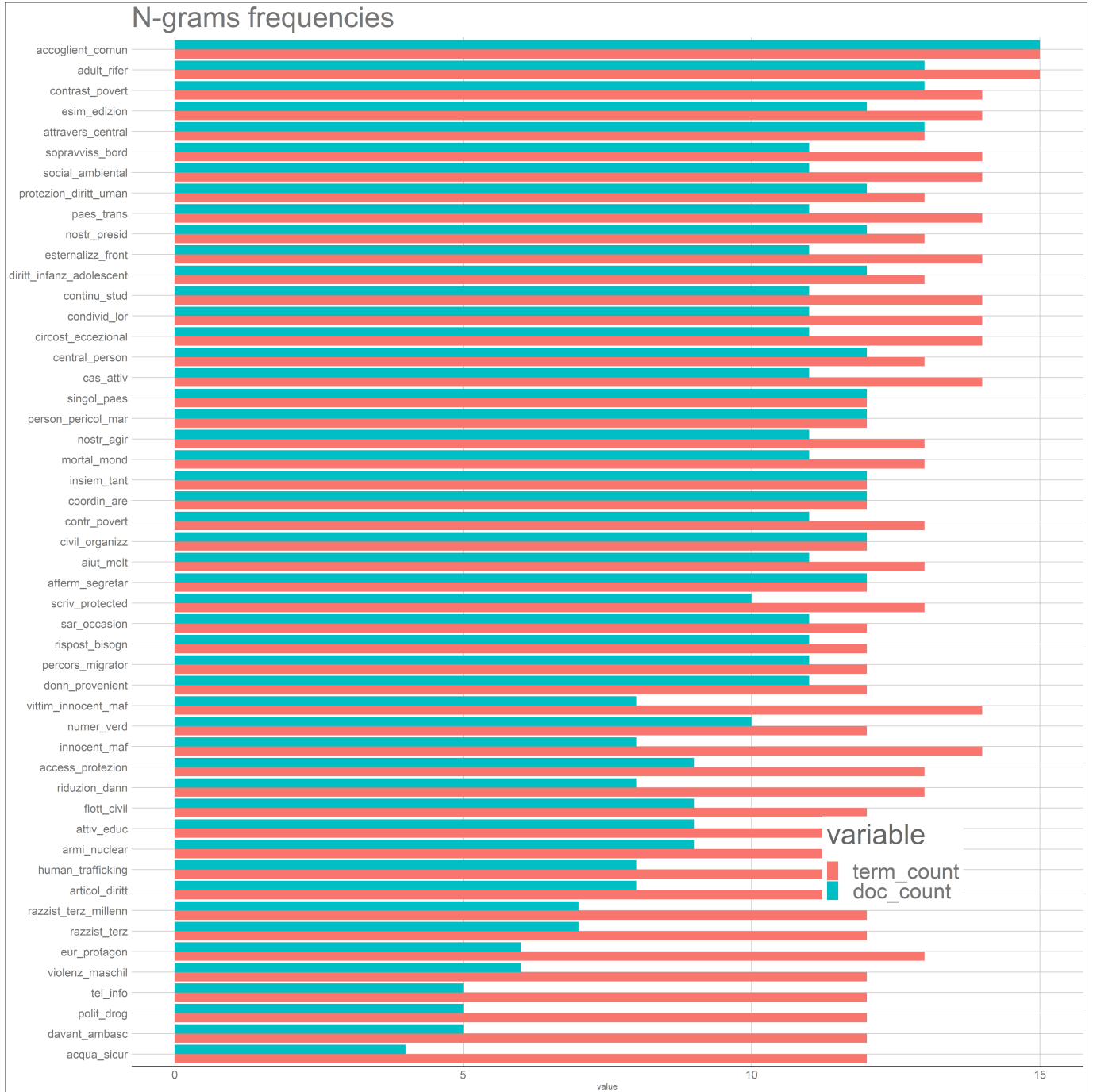
For every article in the training set, a document-term matrix is constructed as described in section 4.2, with rows constituting the training set articles i and columns constituting the n-grams j ($n \in 2, 3$) present in the articles. Preprocessing steps are performed via the R package *text2vec*⁴⁵. 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 A.5.2 summarizes the frequencies of some top frequent n-grams.

⁴³According 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.

⁴⁴The two measures share the same prediction in roughly 86.7% of instances.

⁴⁵<https://CRAN.R-project.org/package=text2vec>

Figure A.5.2. 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 regression⁴⁶ is the penalized negative binomial log-likelihood function, given N sample size and p

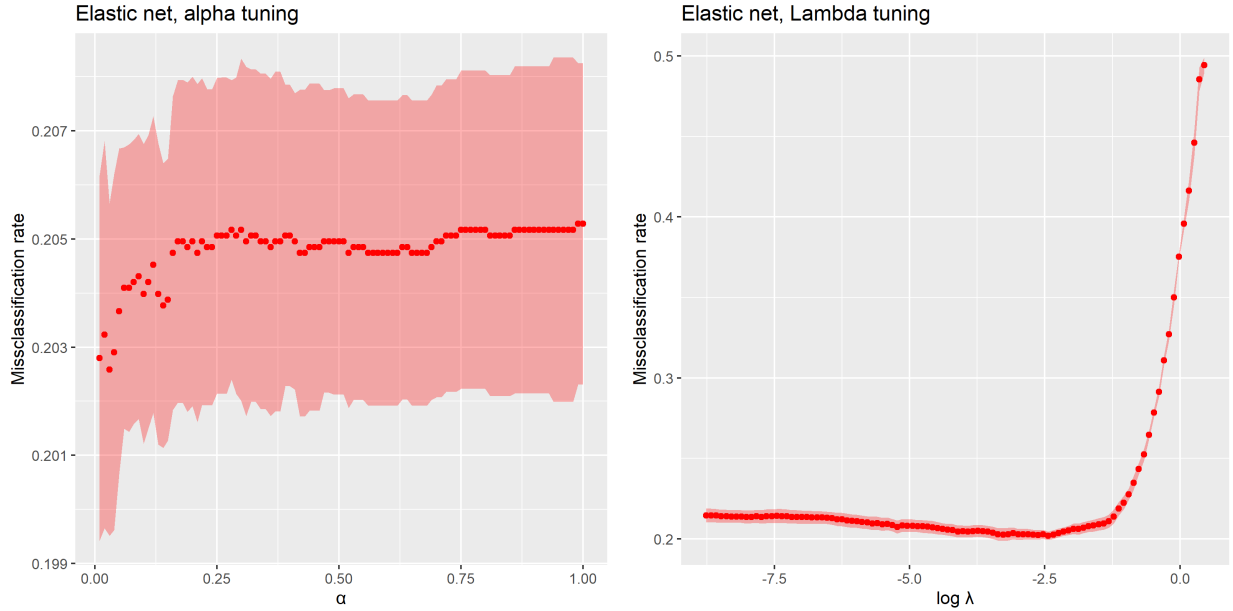
⁴⁶Given an observed binary outcome $Y \in \{0, 1\}$, the logistic model can be written as $\frac{p(Y=1|X=x)}{p(Y=0|X=x)} = \beta_0 + x^T \beta$.

parameters:

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} - \left[\frac{1}{N} \sum_{i=1}^N y_i (\beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right] + \lambda [(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1] \quad (14)$$

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 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 degeneracies 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 A.5.3 displays the levels of α (on the left) and λ (on the right) that minimize the missclassification rate (on the y axis).

Figure A.5.3. Tuning parameters for the Elastic net.



Notes: Cross-validation results. Tuning parameters appear on the x-axes. On the left, α constitutes the level of mixing between lasso and ridge. On the right, λ constitutes the severity of the penalty component. On the y-axis, the missclassification rate is used as a measure of prediction quality.

The binary classification procedure 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).

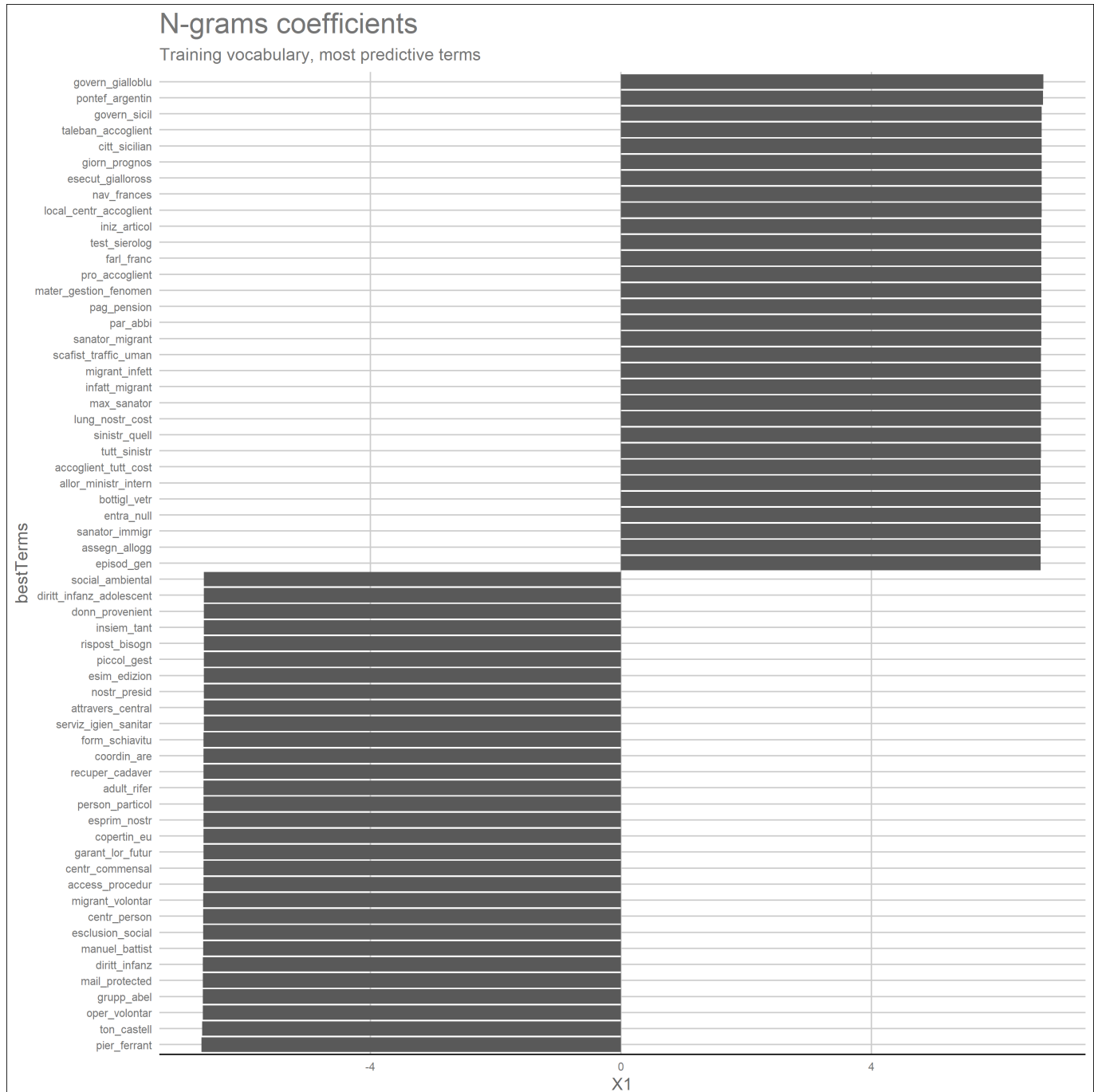
In Taddy (2013)'s procedure, for each document i , equipped with term counts $\mathbf{x}_i = [x_{1i}, \dots, x_{Ji}]'$, the n -gram total counts for labels $y \in \{0, 1\}$ are defined as $x_y = \sum_{i: y_i = y} \mathbf{x}_i$. Given row sums $m_i = \sum_{j=1}^J m_{ij}$, the multinomial inverse

regression would then be:

$$\mathbf{x}_y = MN(\mathbf{q}_y, m_y) \text{ with } q_{yj} = \frac{\exp[\alpha_j + y\phi_j]}{\sum_{l=1}^J \exp[\alpha_l + y\phi_l]}, j = \{1, \dots, J\}, y \in \{0, 1\} \quad (15)$$

Where MN is a j-dimensional multinomial distribution with size $m_y = \sum_{i:y_i=y} m_i$ and probabilities $q_y = [x_{y1}, \dots, x_{yJ}]'$. Under a set of conditions, a sufficient reduction score for frequency $\mathbf{f}_i = \mathbf{x}_i/m_i$ is defined as $\mathbf{z}_i = \boldsymbol{\phi}'\mathbf{f}_i, \implies y_i \perp\!\!\!\perp \mathbf{x}_i$. 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 A.5.4 represent some predictive terms, with associated scores. Terms with negative values in the x-axis predict slant negatively, while those with positive values are positive predictors of slant.

Figure A.5.4. Most predictive coefficients, Taddy 2013.



Notes: Examples of terms used in Taddy (2013) classification. Positive x values indicate the term impacts the slant prediction positively. Vice versa, negative values in the x-axis mean the term is a negative predictor of the slant index.

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 migrant-related 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 A.4.

A.5.4 Media Slant Validation and Exploration

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

To establish the prediction precision for the labelling process, a 10 fold cross-validation procedure was performed on both the baseline and the alternative anti-immigrant slant index. Table A.5.1 below presents cross-validation results.

As a further measure of validation, results were also compared with human classification of roughly 1000 entries. This sample was extracted randomly from the main corpus, with source-based stratification. The same text was analyzed by two individuals from the population of the platform *Amazon Mechanical Turk*. Having two respondents instead of one for the same task allowed for possible variation in the response. At the same time, this variation, though minimal, allows to partly account for subjective biases in the response. Interestingly, this procedure revealed some evidence that human classification of slant is likely heavily influenced by predetermined ideology: 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 indirettamente, 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 più 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 può 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), può 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. In case one the news author writing contains some ideological inclination; in case two the subject of the article involves a person's opinion (or facts) that contains a political tendency. For the scope of this study a broad definition of media bias is adopted in constructing the measures. 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; Djourelouva (2020) who focuses on the banned use of the expression "illegal immigrant").

Comparisons of the slant indices with human classification is displayed in Table A.5.1.

Generally, the two measures appear to be fairly comparable: the accuracy for Taddy's measure (i.e., the proportion of correct classifications over the total sample size) is estimated to be 0.786 with a standard deviation of 0.0016. The elastic-net measure has estimated accuracy of 0.777 with a standard deviation of 0.006, suggesting a slightly lower precision estimate though accompanied by slightly more stability (smaller variability). When compared to the human classification, accuracy is not a good performance indicator, as labels are unbalanced for the 1000 entries: the majority

are labeled as 0.⁴⁷ Therefore, the Younden-J metric (1-specificity -sensitivity) is reported, alongside a sensitivity and specificity. Sensitivity corresponds to the share of 1-labeled instances, over the ‘true’ 1 valued entries. Specificity refers to the share of 0-assigned observations, over the total of true zeros. All in all, these statistics appear to favor the use of Taddy (2013)’s index over the Elastic-net, as a benchmark indicator. They also point to a better performance in terms of correctly identifying the lack of slant, rather than its presence. This may suggest that the media-slant outcomes may suffer from an attenuation bias and can therefore be conceived as a lower bound.

Table A.5.1. Cross-validation results

	<i>Cross validated</i>		<i>Out of sample</i>			<i>Human classification</i>		
Model	Mean accuracy	st. dev.	Accuracy	Specificity	Sensitivity	Younden-J	Specificity	Sensitivity
Taddy (2013)	0.786	0.016	0.739	0.646	0.826	0.239	0.761	0.472
Elastic Net	0.777	0.006	0.733	0.652	0.826	0.214	0.758	0.456

Notes: Performance results from the two generative models. Performance in the left section is based upon a 10-fold cross validation, done manually for Taddy (2013). Package *cv.glmnet* performs cross validation automatically. In the middle, an out of sample prediction is based on a test-set left aside from the training sample. The right-most columns include comparisons with human classification of roughly 1000 entries.

A.6 Municipalities in Sample

Table A.6.1. List of municipalities in the baseline panel

Municipalities:	Airole; Alassio; Albenga; Albisola; Albissola; Altare; Andora; Apricale; Aquila D’arroschia; Armo; Arnasco; Aurigo Badalucco; Bajardo; Balestrino; Bardineto; Bergeggi Boissano; Bordighera; Borghetto D’arroschia; Borghetto Santo Spirito; Borgio Verezzi; Borgomaro; Bormida Cairo Montenotte; Calice Ligure; Calizzano; Camporosso; Caravonica; Carcare; Carpasio Casanova Lerrone; Castel Vittorio; Castebianco; Castellaro; Castelvechio Di Rocca Barbena Celle Ligure; Cengio; Cerialle; Ceriana; Cervo; Cesio Chiusanico; Chiusavecchia; Cipressa; Cisano Sul Neva; Civezza; Cosio D’arroschia; 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; Osliglia; Ospedaletti Pallare; Perinaldo; Piana Crixia; Pietra Ligure; Pietrabrana; 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; Stellanella 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.

⁴⁷human classification returned 73% of 0s; Taddy (2013)’s measure returned 68% and the elastic-net model returned 67%